データベースプロセッサ DIAPRISM（2）データ管理方式

4K－8

道下学　郡光則　安藤隆朗　藤森敬悟　石井篤
三菱電機　情報通信システム開発センター

1. はじめに

DIAPRISM は、明細レベルのテーブルに対する多次元的な集計処理を特徴とするデータベースシステムである[1]。カラムワイズかつ負荷平準なデータ転送を実現するデータ管理方式「ブロック化転置ファイル」を用いることにより、検索速度を向上させることができた。本稿ではこのデータ管理方式について報告する。

2. DIAPRISM の構成とデータフロー

DIAPRISM でのデータフローは図1の通りである[2]。DIAPRISM/DSF は H/W ソータ DIAPRISM/SS を用いながら外部 DB から入力したデータを結合し、明細データを作成する。この明細データ若しくは逐次入力される明細レコードを、当データ管理は明細レベルのテーブルに格納しておく。検索時、データ管理は問合せ要求に応じてディスクより明細データを入力し、I/O プロセッサボード DIAPRISM/SP へデータを送る。この際、演算を除く射影処理を行う。DIAPRISM/SP は演算及び選択処理を行う（CPU が代行することもある）。DIAPRISM/AQL はソート・マージ・集計を行い、問合せ結果を出力する。

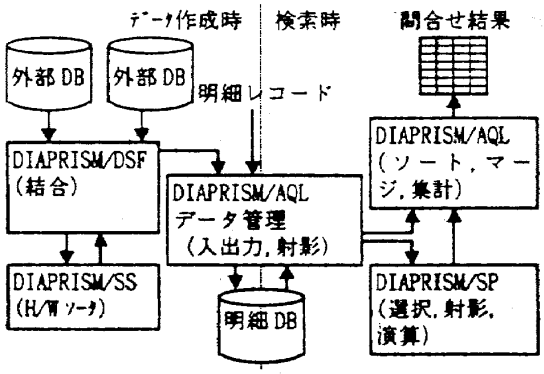


図1. DIAPRISM でのデータフロー

3. 解決すべき課題

明細レベルのテーブルの特性として
・履歴蓄積データであるため多件数
・非正規化により1テーブルが多属性
・分析用の集計検索では、参照対象の属性が少数かつ非定型、件数が多数

なる傾向があり、検索高速化のためにはデータ転送量削減と効率的な並列処理を実現する必要がある。

従来型のアプローチとして indexed file, transposed file が挙げられる。前者は特定属性に焦点をあてた事前の最適化が必要であり、非定型検索では効果が薄い。後者で複数ディスクへの垂直分割を施したとしても、検索時に参照する属性の組合わせによっては転送量が特定ディスクに偏る場合があり、非定型検索に対する効果は不安定である。

4. ブロック化転置ファイル

ブロック化転置ファイルによる配置方式と、これを利用した検索方式を下記に述べる。

配置方式

(A1) 1テーブルのレコードをN件毎に分割する。分割先の数をPとする。要素数Nの一次元レコード配列がP個作成されたことになる。

(A2) レコード内をFバイト単位のM個の物理フィールドに分割する。Fバイトを超える属性は複数の物理フィールドに格納する。(A1)の各レコード配列は物理フィールドのN×M二次元配列になる。

(A3) レコード内の同一オフセットにある物理フィールドをN個集める。この要素数Nの物理フィールド配列を「ブロック」と呼ぶ。ブロックを物理フィールド個数分集める。この要素数Mのブロック配列を「ページ」と呼ぶ。(A2)のN×M二次元配列はM×N二次元配列に転置されたことになる。

(A4) ページをラウンドロビン等により複数ディスクに分配する。1ディスク内に分配された分は、一次元ページ配列として格納する。これを「エクステント」と呼ぶ。

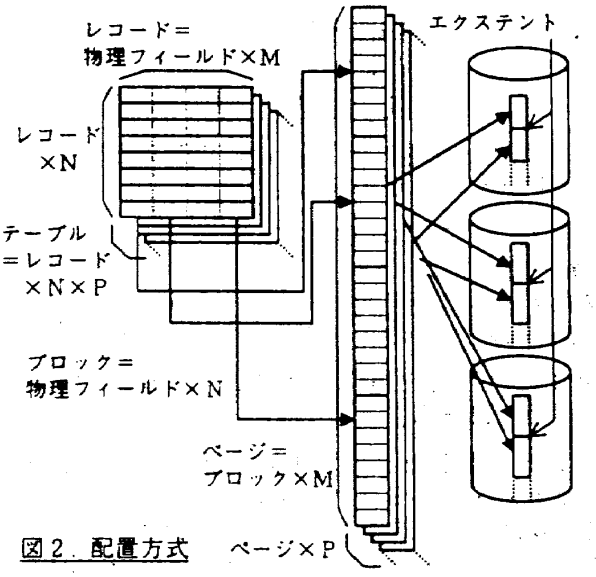


図2. 配置方式

(A1)～(A4)の手順で配置することにより、データの局所性及びデータ量に関して次の特徴を持つ。

・同一属性の物理フィールドは、（N個毎に）エクステント上隣接して置かれる。

Database Processor DIAPRISM(2) Data Management
M.Michishita, M.Kori, T.Ando, K.Fujimori, A.Ishii.
Mitsubishi Electric Co. Information & Communication Systems Development Center
5-1-1 Ofuna Kamakura Kanagawa, 247 Japan

・複数の物理フィールドに分割された属性は、エクステント上隣接したブロックに置かれる。
・１レコード分のデータは、エクステント上１ページの範囲内に置かれる。
・どの属性のデータも、全エクステント即ち全ディスクへほぼ均等量分配される。

検索方式

(B1) １ページの内、参照される属性に該当するブロックのみ読込む。演算を除く射影処理を実行したことになる。
(B2) １レコードに該当する物理フィールドは、入力された各ブロック内の同一オフセットにある。これを対象に演算・選択処理を行う。
(B3) １エクステント内で、先頭ページから最終ページまで(B1)～(B2)を繰り返す。
(B4) 複数のエクステントに対して、(B3)を並列に要求する。全ページに関してソート・マージ・集計処理を行う。

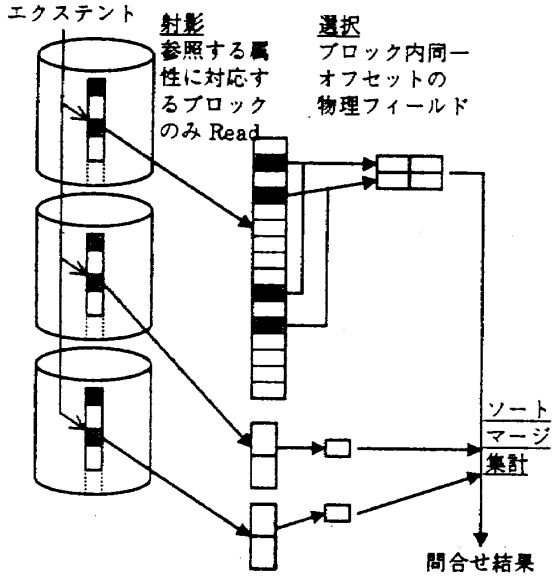


図３．検索方式

ブロック化転置ファイルによる効果として、次の点を見込むことが出来る。

・カラムワイズによる、各ディスク当たりのデータ転送量の削減
・複数ディスクへの負荷平準化 (DIAPRISM/SP による並列処理と併せて、スケーラブルに性能向上)

更に次の工夫を施すことにより、追加効果を見込むことが出来る。

・１ブロックがセクタ境界に合致する様な、F，Nの値を採用する。パディングにより、レコード内の属性の位置をFバイト境界に揃える。
→カラムワイズなデータ転送効率の向上を見込む。

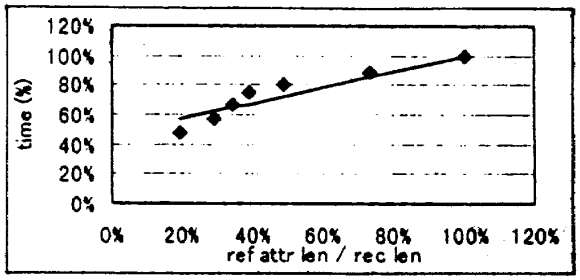
・レコード長に応じてNの値を変動させることにより、１ページのサイズを制限可能にする。
→入出力バッファ用メモリ使用効率の向上を見込む。

・ＯＳ (WindowsNT) のファイルシステムが１ファイル内の連続したデータを隣接したセクタに配置することを前提とし、１エクステントを１ディスク内の１ファイルとして実装した。
→複数ブロックに跨る属性の読込みは、隣接するセクタからの読込みとなる。

5. 性能評価

カラムワイズなデータ転送の効果を評価するために、被参照属性データ長／レコード長の比率に対する、所要検索時間の変化を測定した。
H/W 構成は PentiumPro200MHz×４、DIAPRISM/SP×２、ディスク 4.3G×８。データとして TPC-D[3] の LINEITEM (SF=10)を使用した。問合せは下記 SQL 文を用い、全件該当となる検索条件を WHERE 句に AND で付加することによって被参照属性の増減を指定した。

```
SELECT COUNT(*),
 SUM(L_EXTENDEDPRICE*L_DISCOUNT) AS REVENUE
FROM LINEITEM
WHERE L_SHIPDATE >= DATE '19940101'
  AND L_SHIPDATE < DATE '19950101'
  AND L_DISCOUNT BETWEEN 0.04-0.01
      AND 0.04+0.01 AND L_QUANTITY < 24
```



グラフは、Ｘ軸に「被参照属性データ長／レコード長」をとり、Ｙ軸に「検索時間／Ｘ＝１００％時の所要検索時間」をとったものである。
検索時間は、データ長比率に伴って変動する入力時間と、固定的に費やす集計時間等からなると考えられる。グラフより参照属性の減少に伴って検索時間が短縮化されていることがわかる。

6. おわりに

本稿は DIAPRISM で採用したデータ管理方式「ブロック化転置ファイル」について報告した。今後は当方式をベースに応用技術を発展させ、更なる性能向上や運用性向上を図っていきたい。

参考文献

[1] 佐藤他「多次元明細データベース DIAPRISM/AQL の概要」情報処理学会第 55 回大会 2AD-6
[2] 鹿島他「データベースプロセッサ DIAPRISM(1) I/O プロセッサ制御方式」情報処理学会第 57 回大会 4K-07
[3] "TPC BENCHMARK™ D (Decision Support) Standard Specification Revision 1.3.1", Transaction Processing Performance Council (TPC), San Jose, CA 95112, USA