

データベース演算処理装置のソート処理方式

岩崎孝夫*, 山田広佳*, 井上栄**

(株) 東芝 *情報処理・機器技術研究所 ** 青梅工場

1. はじめに

ソート処理および関係データベース（以下RDB）の検索系処理を高速に実行する専用ハードウェアであるデータベース演算処理装置（以下DBE）を開発した[1]。DBEをTPホストコンピュータTP90/70モデルに接続することで、ソート処理の場合、ソフトウェア実行に対し3～5倍の高速化と負荷分散が図れる。

DBEはソートやRDB検索処理を高速に行うハードウェアと3つのマイクロプロセッサから構成される。各マイクロプロセッサはそれぞれの役割を効率よく処理するため、密結合ながら各々に適した独立のモニタにより制御される。これによりソート処理およびRDBの検索系処理を高速に実行できる。本稿ではDBEのソート処理方式ならびに性能向上効果について報告する。

2. ソート機能概要

DBEでサポートするソート機能には通常のソートの他、レコードの複合条件による選択、レコードの再構成、ファイル属性の異なる複数ファイルの入力（マージではない）、入力ファイルへの出力などがある。

図1にソフトウェアでの各機能と処理手順について示す。



図1 ソートの処理手順（ソフトウェア処理）

この中で、前処理1は論理的に削除されているレコードの処理であり、前処理2は特種な形式をしたレコードに含まれる制御フィールド部分を切り落とす処理である。＜同一キー値処理＞はソートキーによるユニーク処理またはRDB問合せのGROUP化ありのSUM関数と同様の処理を指す。＜選択処理＞は複合条件をAND/ORにより結合し、取捨選択を行う機能である。

基本的に、ソフトウェアによるソートと同等の機能をDBEでサポートする必要がある。同等の機能をサポートするためのDBEでの処理手順について図2に示す。

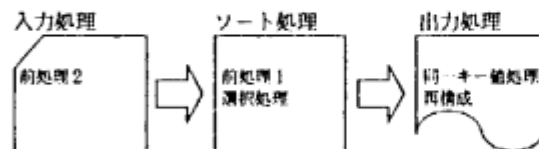


図2 ソートの処理手順（DBE処理）

この他、ソフトウェア処理によるソートにはレコード番号を出力する機能や、＜同一キー値処理＞において入力順や逆入力順に出力する機能がある。したがって、DBEでも同等の機能をサポートするため、図中の入力処理でレコード番号の付加をしたり、出力処理でレコードを逆順に並べ替えたりする処理が必要となる。

3. ソート処理方式

DBEでのソートには、DBEのハードウェア定数とデータ量により3つのレベルがある。一つめは、全データをハードウェアソータで一度にソートする内部ソート。二つめは、複数ファイルを入力するときや、ハードウェアソータで一度にソートできず、マージ処理を必要とする外部ソート。そして、全データを内部バッファに取り込めないため、ワークファイルを必要とする拡張ソートである。

3.1 内部ソート

内部ソートは入力処理、ソート処理と出力処理からなる。これらの処理はスレッドとして実現されているため並行動作が可能で、バッファのやりとりを除けば非同期に動作する。この時の内部の動作状況のタイムチャートを図3に示す。このことから、内部ソートはデータの入出力時間により全体の処理時間が決まることがわかる。

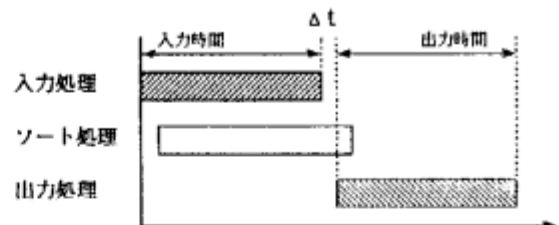


図3 内部ソートのタイムチャート

3.2 外部ソート

外部ソートはソートフェーズとマージフェーズからなる。

ソートフェーズはハードウェアソータで一度にソートできる件数毎のストリングに全データを分割し、そのストリング毎にソートを行うフェーズである。ソートフェーズには、入力処理とソート処理がある。このとき、内部ソート同様に、入力処理とソート処理が並行に動作する。したがって、ソートフェーズはほぼデータの入力時間で完了する。

マージフェーズはソートフェーズで作成したn本のソートストリングをマージして出力するフェーズである。マージフェーズには、マージ処理と出力処理がある。マージ処理はnウェイ

マージを行うため、ソートフェーズで作成したストリング本数により1バッファ分の出力レコードを作成するための性能が異なる。このDBEで採用した方式では、 n 本のストリングから1レコードをマージするために $\log_2(n)$ 回の比較が必要である。したがって、1バッファ分の出力レコード数を N とすると、出力すべきバッファを作成するために

$$O(N \cdot \log_2(n))$$

の処理時間を要する。一方、出力処理が1バッファ分の出力に $O(N)$ の時間を要する。このことは、マージフェーズでは、マージ処理と出力処理が並行動作しているものの、ストリング本数が増えると図4で示すように、出力処理に比べマージ処理の時間が長くなる場合があることを示している。したがって、外部ソートではレコード数が多くストリング数が増えると、入力時間+出力時間ではソートが完了できなくなる場合がある。

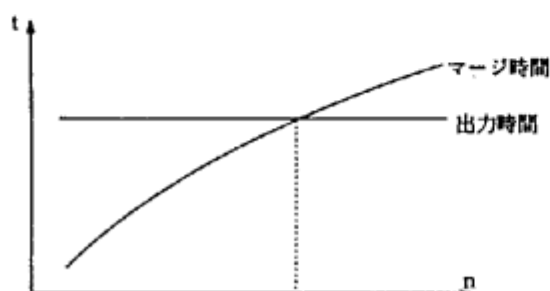


図4 1バッファ分のマージ時間と出力時間

3.3 拡張ソート

拡張ソートもソートフェーズとマージフェーズからなる。

ソートフェーズは一度にソートできる件数毎のストリングに全データを分割し、そのストリング毎にソートを行うフェーズである。しかし、外部ソートとは異なり、入力データを内部バッファに蓄えることができないため、ストリング毎にワークファイルに出力する必要がある。したがって、ソートフェーズでは入力処理、ソート処理と出力処理がある。拡張ソートのソートフェーズでも入力とソートは並行に動作するため、ソートフェーズの処理時間は入力時間+出力時間の総和となる。

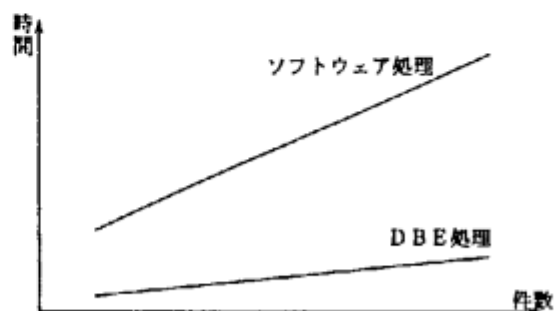
マージフェーズはワークファイル上のソートストリングを入力しながらマージを行い出力するフェーズである。ソートストリングの全データを内部バッファに入力することはできない。したがって、入力処理はまずストリング毎に2バッファずつのレコードを読み込んでおく。以降は、マージによりバッファが空になったストリングのデータを読み込むという要求充足型のサービスを行う。ソートフェーズの入力と異なり、マージフェーズでの入力はワークファイル空間でのランダムアクセスが必要となる。しかし、この2バッファ方式と要求充足型サービスにより極力ランダムアクセスによるオーバーヘッドを吸収している。この入力処理と並行にマージ処理と出力処理を外部ソートの場合と同様に処理を行う。

以上のように拡張ソートではワークファイルの入出力に伴うのと同じデータのソートを外部ソートで行う場合に比べおおよそ2倍の処理時間を必要とする。

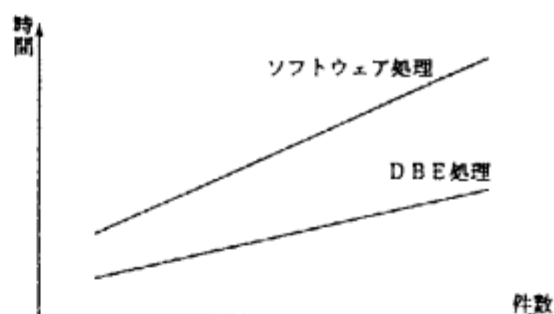
4. 効果

図5にDBEでのソート処理とホストの最上位モデルでのソフトウェアによるソート処理との性能比を示す。図5-1は内部ソートと外部ソートの場合を示し、図5-2は拡張ソートの場合を示す。

このグラフが示すように内部・外部ソート時で3~5倍、拡張ソート時で1.5倍~2倍の高速化が図れる。また、DBE使用時はソート処理の高速化だけではなく、ホストのCPU負荷がほぼ0となり、システム全体のスループットの向上に寄与する。



5-1 内部・外部ソート



5-2 拡張ソート

図5 ソフト処理とDBE処理とのソート性能

5. おわりに

本稿では、DBEを用いてソート機能を実現する場合の処理方式ならびに性能向上効果について述べた。DBEは、ホストでのソート機能と同等の機能をホストのCPUに負荷を掛けずに提供することができる。

今後は更なる性能向上とソート処理以外への適用について検討・実現を図りたい。

謝辞

本データベース演算処理装置は(財)新世代コンピュータ技術開発機構の研究成果をもとに製品化したものである。

参考文献

- [1] 松田 他：データベース演算処理装置のアーキテクチャ
情報処理学会第43回全国大会、1M-2(1991)