4K－9

データベースプロセッサ DIAPRISM (3)
高速ハードウェアソータ

東 辰輔、中野 孝、佐久間 孝夫
三菱電機（株） 情報通信システム開発センター

1　はじめに

高速ハードウェアソータDIAPRISM/SSを開発した。DIAPRISM/SS は当社 GREO-1F[1]に実装されていたパイプラインマージソータを発展させたもので、小型化・大容量化および高速化を目的として開発した。本稿では、DIAPRISM/SS で採用したアルゴリズム、アーキテクチャ、ハードウェアおよび性能測定結果について報告する。

2　ソートアルゴリズム

一般にパイプラインマージソータは、n個の Kウェイマージソートプロセッサを一次元接続し（図1)、ソートプロセッサ S_iはそれぞれ K^{i-1} レコードからなる K 本のソート済みレコード列（ストリング）をマージして K^i レコードからなる 1 本のストリングを出力する。ソートプロセッサ S_i は $K-1$本目までの入力ストリングを格納するためにローカルメモリ L_iを有する。L_iは前段のローカルメモリ L_{i-1}の K 倍の容量をもつ。

この時、最大 K^n件のレコードのソートが可能となる。また最終段のローカルメモリ L_nの容量が M バイトの場合は、$\frac{K}{K-1}M$ バイトのデータ量をソートすることができる。

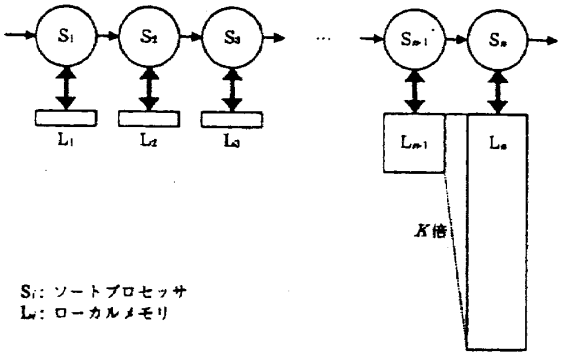

図1　Kウェイパイプラインマージソート

同じだけの一括ソート可能レコード件数とソート容量を実現するには、ソートプロセッサ数およびメモリ素子数の点からウェイ数 Kが多い方が有利であるのは明らかである。一方、Kが大きくなるとソートプロセッサのハードウェア規模が増大する。そこでDIAPRISM/SS では8ウェイパイプラインマージソートアルゴリズムを採用した。

8 ウェイパイプラインマージソートでは、図 2に示すように、例えばソートプロセッサ S_2は 8 レコードからなるストリングを順次入力し、8 本のストリングをマージして 64 レコードからなるストリングを出力する。

なお、Kウェイパイプラインマージソートにおいて Nレコードのソートにかかる所要時間は、図 2からも分かる通り理論上 $2N+\log_K N-1$ に比例する。

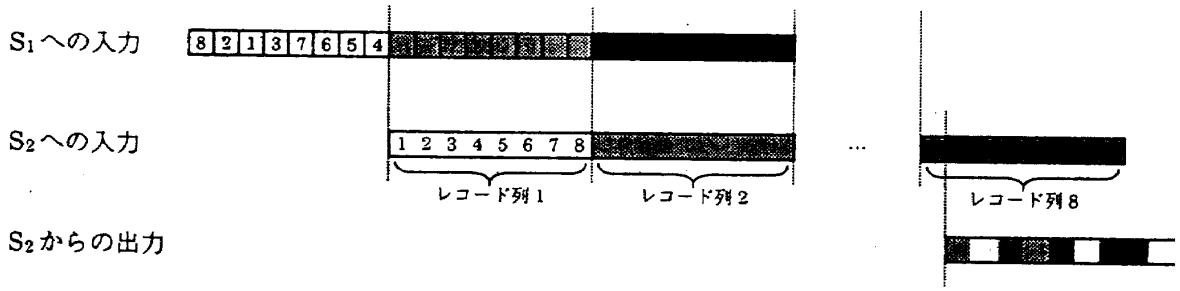

図2　8 ウェイマージソートの動作

Database Processor DIAPRISM (3)　High-Speed Hardware Sorter
Shinsuke Azuma, Takashi Nakano, Takao Sakuma
Mitsubishi Electric Corp. Information & Communication Systems Development Center

3　ソートプロセッサアーキテクチャ

図3にソートプロセッサのブロック図を示す。前段から送られて来たデータはバッファを介してローカルメモリに格納される。比較部はローカルメモリより8本のストリングをバッファに読み込み、マージして後段に送り出す。以下、このプロセッサの特徴を述べる。

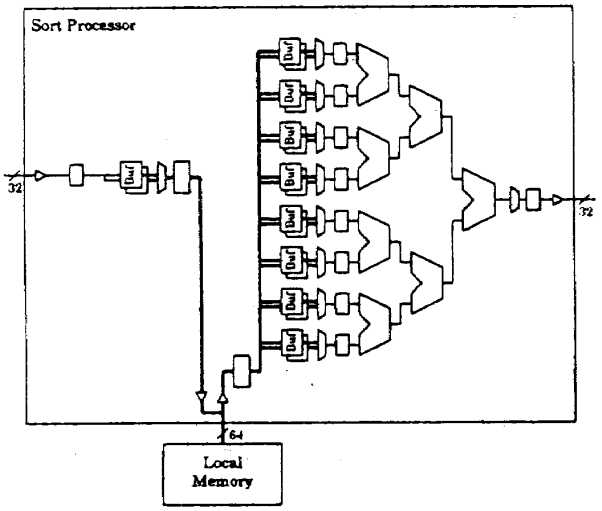

図3　ソートプロセッサのブロック図

① 比較器を7個トーナメント状に接続し、8ウェイから最小のものを選択する。この時、前回勝ち抜けたレコード（champion）の後続のレコードとchampionに直接負けた3レコードの計4レコードのみを比較することにより次の champion を決定する。すなわち定常的に 4 レコードのみの比較を行なうため、ローカルメモリからの読み込みのボトルネックを低減することができる。

② 動的あるいは静的に 1 から 8 の任意の数のストリングをマージできる。これにより、どのようなレコード長に対してもストリング長を自由にチューニングすることができ、ローカルメモリを有効活用することが可能となる。

③ ローカルメモリの読み書きは内部バッファを介して 256 バイト単位のバーストで行なう。これによりメモリアクセスのオーバヘッドを低減する。

④ ローカルメモリを接続せずに内部のバッファだけを使用することも可能である。これは、初段のようにストリング長が短い場合には、メモリアクセスのオーバヘッドを回避することができ性能面で有効である。

⑤ ローカルメモリは、論理的には 256 バイト長のページがリスト構造で連結された構成となっている。ソートプロセッサはリストの先頭と末尾を保持しており、新たな領域を必要とする場合はリスト先頭のページを確保し、不要になったページはリスト末尾に連結することによりメモリ管理を行なっている。

⑥ ローカルメモリとして、EDO DRAM および SDRAM（Synchronous DRAM）をサポートし、高速アクセスを可能とする。

4　ハードウェアソータの諸元

表1および表2に DIAPRISM/SS およびソートプロセッサ LSI の諸元を示す。

DIAPRISM/SS は8ウェイマージソートアルゴリズムを採用して実装面積を小さくし、さらに PCI インタフェースを設けたことにより PC サーバへの内蔵が可能となった。

表1　DIAPRISM/SS の諸元

接続インタフェース	PCI 2.1 準拠
ボードサイズ	PCI ボード（10cm×31cm）×2 枚
ソートプロセッサ数	8 石
一括ソートレコード件数	1,600 万件
ソート容量	560M バイト
レコード長	4～32760 バイト

表2　ソートプロセッサの諸元

LSI プロセス	CMOS 0.35μm エンベデッドアレイ
ゲート数	91K ゲート＋41K ビット RAM
パッケージ	320 ピン BGA（Ball Grid Array）

5　性能

レコード長 100 バイト、キー先頭 10 バイトのランダムデータ 500 万件をソートした時の「主記憶（元データ）→DIAPRISM/SS→主記憶（ソート結果）」の所要時間は、12.5 秒であった。なお PC サーバの CPU は PentiumII（300MHz）、チップセットは 440LX、OS は Windows NT で測定した。

6　おわりに

高速ハードウェアソータ DIAPRISM/SS について報告した。今後は射影、選択などの機能追加を検討するとともに、PC サーバ本体のバスがボトルネックとなるのを回避するべくアーキテクチャを検討する予定である。

参考文献

[1] 山崎他「データベースプロセッサ GREO-1F ハードウェアソータ」、情報処理学会第 53 回全国大会、1996