

データベースプロセッサ DIAPRISM(1) I/O プロセッサ制御方式

4K-7

○鹿島理華* 郡光則* 三角武嗣** 清水英弘* 高山茂伸*

*三菱電機 情報通信システム開発センター **三菱電機 情報システム製作所

1. はじめに

DIAPRISM は、明細レベルのテーブルに対する多次元集計処理を特徴とするデータベースシステムである^[1]。データウェアハウスに代表されるような非定型な全件検索、集計処理を行う場合、演算処理能力よりもデータ供給能力がボトルネックとなっていることが多い。通常は処理規模が増加するとともにディスクの数を増やし並列化を図る。しかし、入出力バスやメモリシステムの転送能力の限界から、ディスクの数に比例した性能が得られなくなることが多い。このような問題を解決するには、従来は高価な並列機やクラスタを必要としていた。

我々はデータベースプロセッサ DIAPRISM において、I/O プロセッサ技術を用い上記の問題を解決した。本方式は WindowsNT、PC サーバ、I₂O といった標準的プラットフォーム上に実現しており、きわめて低コストにその効果をあげている。

本稿では、この I/O プロセッサ制御方式の概要を報告する。

2. ハードウェア概要

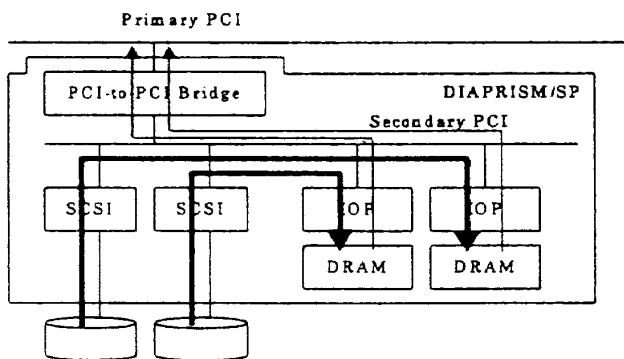


図1 ハードウェア構成図(1)

図1に I/O プロセッサ・ボード DIAPRISM/SP のハードウェア構成図を示す。DIAPRISM/SP は、PC サーバ内蔵の PCI ボードとして実装されている。

Database Processor DIAPRISM(1) I/O Processor Control

R.Kashima*, M.Kori*, K.Misumi**, H.Shimizu*, S.Takayama*

*Mitsubishi Electric Co. Information & Communication Systems Development Center

5-1-1 Ofuna Kamakura Kanagawa, 237 Japan

**Mitsubishi Electric Co. Information Systems Engineering Center

325 Kamimachiya Kamakura Kanagawa, 247 Japan

複数の IOP(I/O Processor)と複数の SCSI コントローラから構成される。IOP と SCSI コントローラはセカンダリ PCI バスに接続される。プライマリ PCI バスには、PCI-to-PCI ブリッジを介し接続される。

図2に示すように、DIAPRISM/SP はプライマリ PCI バスに複数枚接続可能である。

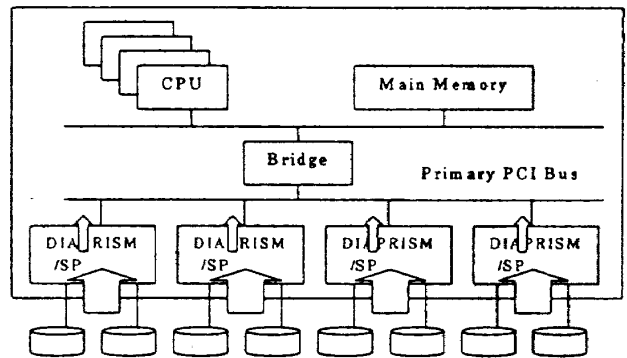


図2 ハードウェア構成図(2)

3. ソフトウェア概要

図3に DIAPRISM/SP を用いたデータベース管理システム DIAPRISM/AQL のソフトウェア構成図を示す。

IOP の制御には I₂O(Intelligent I/O)規格^[2]で規定されたインタフェースを利用している。ホスト側 OS には WindowsNT、IOP 上の IRTOS(I₂O Real Time OS)には IxWorks を採用している。

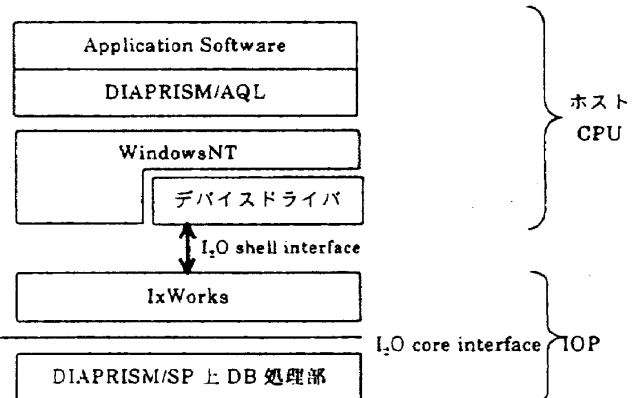


図3 ソフトウェア構成図

4. 高速化アーキテクチャ

DIAPRISM/SP の制御方式を下記に述べる。

4.1. デバイス間直接転送

DIAPRISM/AQL では、入力ファイルをセカンダリ PCI バスを介しディスクから IOP のローカルメモリへ、デバイス間で直接転送する。この転送は主記憶を経由しない。つまり、性能のネックとなっていたプライマリ PCI バスにはデータが流れない。

4.2. データベース処理

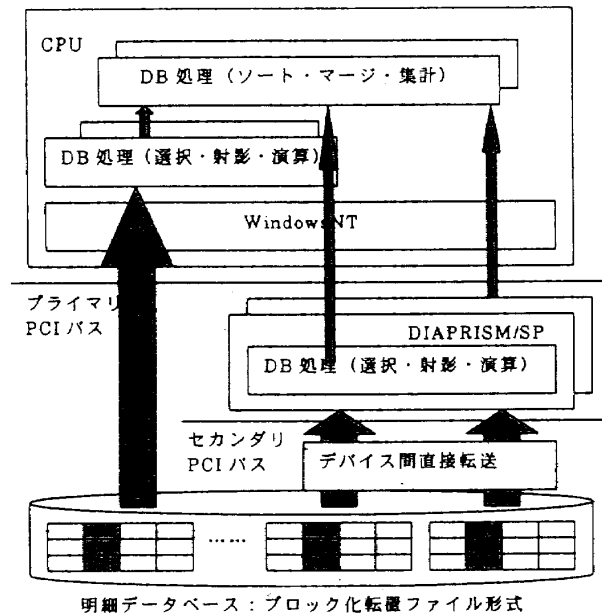


図4 動作概要

DIAPRISM/AQL では、ブロック化転置ファイル形式^[3]を採用している。これは一般に長いレコード長を持つことの多い明細データベースから、レコードの必要な部分のみをディスクから読み出すことを目的としている。DIAPRISM/SP 上のデータベース処理部は、転置処理を行いつつ転送したデータに、データベースの前処理にあたる選択・射影・演算を行いデータ量を削減する。

4.3. 並列処理

前述のデータベース前処理は個々の IOP で独立に並列処理される。また、複数の DIAPRISM/SP が接続されていた場合、図2のハードウェア構成が示すように、ディスクから DIAPRISM/SP 上のメモリへのデバイス間直接転送は、個々の DIAPRISM/SP 上の SCSI コントローラとセカンダリ PCI バスを介するため、互いに他の DIAPRISM/SP の影響を受けない。

また、DIAPRISM/AQL は、ハードウェアとソフトウェアとの協調並列処理方式を採用している。システム中の DIAPRISM/SP の負荷状態により、IOP 上の選択・射影・演算処理はソフトウェア・

スレッドでも実行可能である。

5. 性能

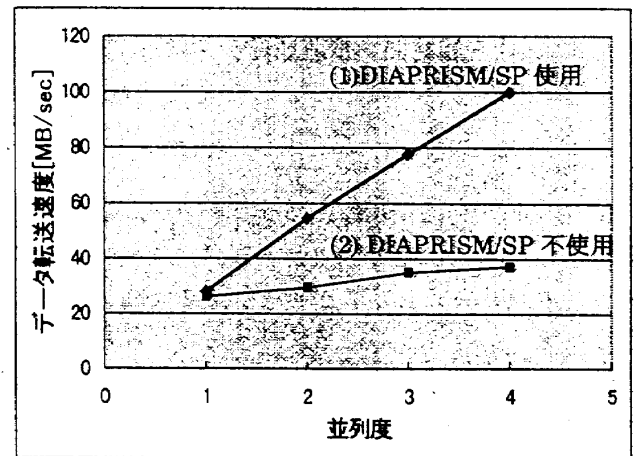


図5 IOPの効果

本 I/O プロセッサ制御方式を評価するために、評価用のソフトウェアを使用して、1レコード 2000 バイト、レコード選択率 1/16 の検索集計を実施した。図5に並列度の増加に伴う、ディスクからのデータ読み出しのデータ転送速度の変化を示す。測定に使用した PC サーバは CPU が PentiumPro(200MHz)×2、チップセットは 450GX。並列度1あたり、(1)は DIAPRISM/SP 1枚、ディスク5本、(2)は SCSI コントローラ1個、ディスク5本という構成。

DIAPRISM/SP を使用しない場合(2)、入出力バスやメモリシステムの転送能力の限界のため性能は頭打ちとなっているが、DIAPRISM/SP 使用時(1)は、DIAPRISM/SP の高速化アーキテクチャにより、プライマリ PCI バス・トラフィックが大幅に軽減され、ディスクの数に比例したスケーラビリティが得られている。以上の結果は、本 I/O プロセッサ制御方式の有効性を示している。

6. おわりに

データベースプロセッサ DIAPRISM で採用した I/O プロセッサ制御方式についてその概要を報告した。今後は、更なる高速性を目指して、機能強化を進めていく予定である。

<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] Intelligent I/O(I₂O) Architecture Specification, Version 1.5(March 1997), I₂O SIG
- [3] 道下他「データベースプロセッサ DIAPRISM(2) データ管理方式」情報処理学会第57回大会 4K-08