TM-1119他

情報処理学会 第43回全国大会
論文集

October, 1991

ICOT

Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan

(03)3456-3191～5
Telex ICOT J32964

Institute for New Generation Computer Technology

TM1119	データベース演算処理装置のアーキテクチャ	松田　進、東郷　一生、島川　和典、岩崎　孝夫
TM1120	データベース演算処理装置の関係演算処理方式	島川　和典、山田　朝彦、佐藤　祐治、外尾　博紀、天野　慎一
TM1121	データベース演算処理装置を用いた問い合わせ処理方式	菊地　哲男、箕田　稔、三友　雄司、道家　まり
TM1122	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
TM1127	PIM/p におけるパイプライン制御方式	安里　彰、木村　通秀、篠木　剛、服部　彰

データベース演算処理装置のアーキテクチャ

松田進*，東郷一生*，島川和典**，岩崎孝夫**

（株）東芝　*青梅工場　**情報処理・機器技術研究所

1．はじめに

近年の高度情報化社会の著しい進展に伴い，あらゆる種類の情報を有機的に統合して管理するためにデータベース処理技術が必要不可欠となってきている．データベース処理では，データ量の増大に対する「ソート処理の高速化」，及びデータの取扱いが容易な「関係データベース（以下ＲＤＢ）」への需要がますます増大しており，その結果これらを高速に処理する専用ハードウェアへの期待が大きい．

当社は，(財)新世代コンピュータ技術開発機構の発足以来，同機構からの再委託研究の一環として，「ソート及び関係代数演算を高速に処理するハードウェア」の開発に従事してきたが［1］［2］［3］，今回その成果に基づき，高速な「ソート及び関係代数演算処理」を行うことができるデータベース演算処理装置（以下ＤＢＥ）を製品化したので［4］［5］［6］，そのアーキテクチャの概要を報告する．

2．既存技術の説明

前記再委託研究で試作した関係データベースエンジン（以下エンジン）の概要は以下のようなものである［3］．

2．1　試作システムの構成

試作システムの構成を図1に示す．本試作システムでは，エンジンはホストシステムのチャネル部に接続した．

エンジンでの処理が必要なときは，磁気ディスクから読み出されたデータは入力チャネルを通してエンジンへ送られ，そこで加工されて，出力チャネルを通してホストへ転送される．

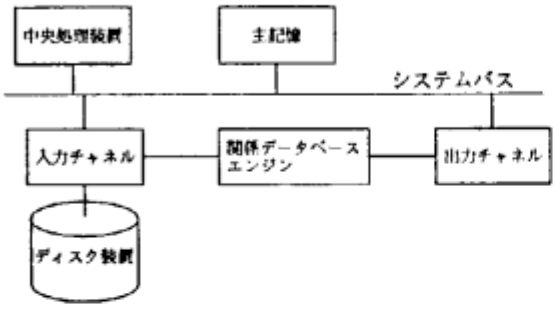

図1　試作システムの構成

2．2　エンジンの構成及び特徴

エンジンの構成を図2に示す．入力チャネルを通して送られてきた対象データは，レコードバッファへ格納される．

同時に前処理部は，「演算に必要なキーのみを切り出し，レコード識別番号を付加してソータ，関係代数演算部へ送出する」という処理をレコード毎に行う．

ソータは，複数のソートセルをカスケードに接続して構成されている．各ソートセルは，２ウェイマージソートアルゴリズムを採用している．

関係代数演算部は，Ｓｅｌｅｃｔｉｏｎ，Ｊｏｉｎ等の関係代数演算を実行するモジュールである．

演算の結果は，レコード識別番号の列で出力されるので，それをもとに元のレコードを読みだし最終の結果を得る．

以上のように，本エンジンはレコードデータから，キー部のみを取り出し，これにレコード識別番号を付加して演算部に流す「キー切り出し方式」を特徴としていた．

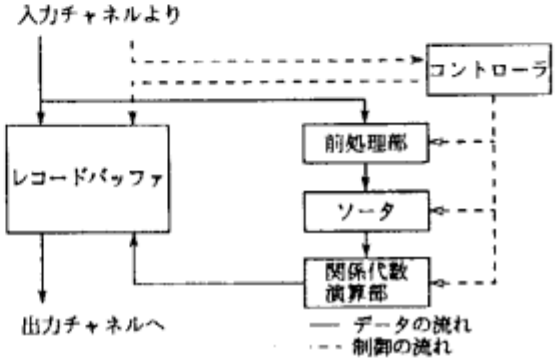

図2　関係データベースエンジンの構成

3．ＤＢＥの内部アーキテクチャ

ＤＢＥは上記エンジンをベースに開発したものである．

関係データベースエンジンでは，サポートするファイル形式を特定していたので，前処理部，結果の再構成部等はハードワイアードロジックで組まれていたが，ＤＢＥでは商用のシステムに適合させるために，以下のような変更を施した．

3．1　前処理部のマイクロプロセッサ化

本装置が対象とするファイル形式は，順編成，索引編成，ＲＤＢ専用の統合編成等多種にわたるので，ハードウェアのみで「キー切り出し方式」をサポートするのは効率が悪い．

従って，本装置ではディスクから読み出されたデータはいったんレコードバッファへ格納し，その後で「キー切り出し専用のマイクロプロセッサ」により，キー部とレコード識別番号（レコードバッファ上のレコードの先頭アドレス）を抽出し，演算部へ送出する．

3．2　再構成部のマイクロプロセッサ化

ＲＤＢをサポートするときは，以下のような状況が想定されるので，レコード識別番号から単純にもとのレコードを取り出すだけではすまない．

・入力ファイルと出力ファイルのファイル編成が異なる．

Architecture of Database Processor

Susumu MATSUDA†, Kazuo TOGO†, Kazunori SHIMAKAWA†† and Takao IWASAKI††

TOSHIBA Corp. †OME WORKS, ††INFORMATION SYSTEMS ENGINEERING LABORATORY

・projectionが指示されたときは，「指定されたカラムの組み合わせ」で出力レコードを再構成する必要がある．

従って，本装置では出力ファイルの指定情報をもとに「再構成専用のマイクロプロセッサ」により，レコードバッファ上のレコードデータと演算結果であるキー識別番号から出力ファイルの再構成処理を行う．

3．3　レコードバッファの共有メモリ化

上記のようにすると，レコードバッファ上の同一データを複数のプロセッサモジュールがアクセスする必要が生じるので，レコードバッファを共有メモリとし，この共有メモリと各プロセッサモジュールを内部バスにより接続し，各プロセッサモジュールの共通のアドレス空間上に共有メモリを配置した．

3．4　ハードウェア構成

図3に本装置のハードウェア構成を示す．各モジュールの機能分担は以下のようになっている．

（1）　ＥＩＰ

マイクロプロセッサを内蔵し，ホストとの物理レベルのインタフェースを制御する．同時にディスクコントローラとも接続し，ディスクとの入出力を制御する．また，出力ファイルの再構成処理も行う．

（2）　ＥＣＰ

ホストからのコマンドを解析し，ＥＩＰ／ＥＣＡＭを制御して要求された機能を実現する．本モジュールもマイクロプロセッサを内蔵する．

（3）　ＥＣＡＭ

マイクロプロセッサを内蔵し，共有メモリ上に入力されたレコードデータからキー部を切り出し，ソータ，関係代数演算部へ送出する．このとき，キーデータの内部データ形式への変換も行う．また，ソータ，関係代数演算部から出力された演算結果を共有メモリ上へ転送する．

（4）　ＰＳＯＭ

複数のソートセルからなるパイプラインソータであり，各ソートセルは2ウェイマージソートアルゴリズムによる．本装置では最大１８段まで接続可能である．

（5）　ＰＲＡＭ

ＲＤＢにおける関係代数演算を行うモジュールである．またソートセルの最終段としても機能する．

（6）　ＥＢＤＭ

ディスクから入力されたレコードデータ，ＥＣＡＭから出力される演算結果等を格納する大容量の共有メモリである．本装置では最大５１２ＭＢまで実装できる．

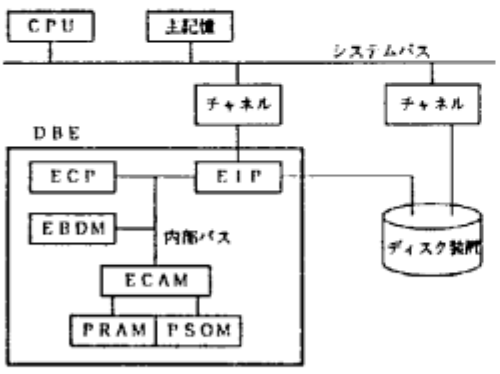

図3　ＤＢＥのハードウェア構成

4．ホスト接続方式

ＤＢＥをホストコンピュータへ接続するときの形態としては，下記が考えられる．

a．システムバス接続方式
　独立装置として，ホストコンピュータのシステムバスに接続する．

b．コントローラ内蔵方式
　ディスクコントローラ部に内蔵する．

c．チャネル接続方式
　チャネルに接続される独立装置として，全ディスクコントローラと接続する．

システムバス接続方式では，システム上の全データを処理の対象とできるが，データアクセス時にホストシステムの介在を必要とするので，データアクセスに要する性能がホストシステムの性能の制限をうけ，専用ハードウェアの性能が充分に出せないおそれがある．

コントローラ内蔵方式では，専用ハードウェアを内蔵するコントローラ下のデータへのアクセスは，ホストシステムの性能の制限を受けずにアクセスすることができるが，別コントローラに接続されたデバイス上のデータへのアクセスが難しくなる．

チャネル接続方式では，全コントローラと接続する必要があるためハード量は増大するが，データアクセスの性能，自由度とも優れている．

以上から，ＤＢＥでは機能，性能を重視して，チャネル接続方式を採った．

5．おわりに

ＤＢＥの内部アーキテクチャとホスト接続方式について述べた．

ＤＢＥを用いることによりＲＤＢ処理の高速化が図れ，従来性能上の理由でＲＤＢが適用されなかったケースでも，ＲＤＢを採用することが期待できる．

参考文献

［1］Sakai,H.,Iwata,K.,Shibayama,S.,Abe,M and Itoh,H.:
Development of Delta as a First Step to a Knowledge Base Machine, in Sood, A.K. and Qureshi, A.H.(eds.),
Database Machine Modern Trends and Applications, pp.159-181,
Springer-Verlag, Berlin(1986).

［2］岩田，神谷，酒井，柴山，伊藤，村上：
関係データベース処理エンジンのソータの試作と評価，情報処理学会論文誌，Ｖｏｌ．２８，Ｎｏ．７，ｐｐ．７４８－７５７（１９８７）．

［3］伊藤，島川，東郷，松田，伊藤，大場：
可変長レコード用関係データベース処理エンジンの試作とソート処理性能の評価，情報処理学会論文誌，Ｖｏｌ．３０，Ｎｏ．８，ｐｐ．１０３３－１０４４（１９８９）．

［4］岩崎　他：
データベース演算処理装置のソート処理方式，情報処理学会第４３回全国大会，１Ｍ－３（１９９１）

［5］島川　他：
データベース演算処理装置の関係演算処理方式，情報処理学会第４３回全国大会，１Ｍ－４（１９９１）

［6］菊地　他：
データベース演算処理装置を用いた問合せ処理方式，情報処理学会第４３回全国大会，１Ｍ－５（１９９１）