計算機アーキテクチャ　90－6
（１９９１．１０．７）

データベース処理のハードウェア化

工藤　哲郎

富士通株式会社

汎用コンピュータシステムにおいて，リレーショナル・データベース処理をハードウェア化するための技術について述べる。磁気ディスク制御装置内に，サーチ処理を実行するハードウェア機構を設け，ホストコンピュータの負荷軽減及びデータ転送量の削減を図ったものである。実現にあたっては，従来ソフトウェアにて処理されていたデータベース構造を，直接ハードウェアロジックによって処理することを可能にした。また，複数の処理を並列に実行するとともに，ハードウェア化に適した選択条件判定方式を採用することにより，データベース処理の高速化を実現した。

A HARDWARE FOR DATABASE PROCESSING

Tetsuro Kudo

Fujitsu Limited

The method to make the relational database processing hardware in the general purpose computer system is described. We installed the hardware mechanism which executed search operation in the magnetic disk controller to achieved the load reduction of the host computer and the reduction of the amount of the data transfer. In the method, the database structure so far processed by software can be directly processed by the hardware logic. Futher, we adopted the parallel execution of two or more processings and a suitable selection condition judgment method for hardware. As a result, the speed-up of the processing with hardware was achieved.

1. まえがき

近年，ホストコンピュータ上で行われていたデータベース処理を，専用のハードウェアによって実施することにより，ホストコンピュータの負荷軽減及び処理の高速化を図ることが注目されている。

本稿では，ホストコンピュータとデータベースが格納されている磁気ディスク装置との間に，専用のデータベースアシスト機構（以下，ＤＢＡ）を設け，リレーショナルデータベース処理をハードウェア化するための技術を紹介する。

ハードウェア化を実現する上では，いくつかのデータベース機能が想定出来るが，今回は一般的で使用頻度の高く，効果が期待出来るサーチ関連機能の実現について述べる。

2. システム構成

図１にＤＢＡのシステムにおける位置付けを示す。

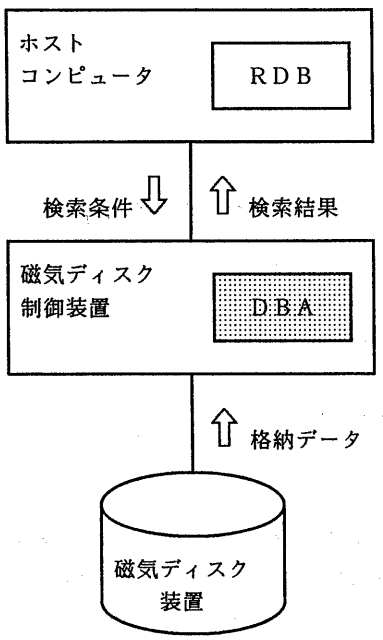

図１　システム構成

ＤＢＡは，ホストコンピュータと磁気ディスク装置の間に位置する磁気ディスク制御装置に搭載され，ホスト上に常駐するリレーショナルデータベース・プログラム（以下，ＲＤＢ）と連携して動作する。

磁気ディスク制御装置内にデータベース処理機構を搭載することは，１つの磁気ディスクサブシステムにおいて，一般のＩ／Ｏ処理，ＤＢＡを使用したデータベース処理及びＤＢＡを使用しないデータベース処理を効率良く混在制御することが可能となる。また，独立した装置を設ける場合に比べて，外部インタフェース制御部を初めとする物理資源の共用が可能になるとともに，物理スペース面でのメリットが大きいと言える。

但し，一般のＩ／Ｏ処理との同時制御を行うという観点より，制御の複雑さ及び一般処理への悪影響等を解決する必要がある。

3. ハードウェア化機能

今回ハードウェア化の対象としたサーチ関連機能は以下の３つである。

- 選択：条件に合致するレコードの選択
- 射影：レコードからの指定カラムの抽出
- 集約：統計的集計を行い，その結果を通知

上記機能は，磁気ディスク装置から読み出される大量のデータをオンザフライで順次処理することが可能であるため，ハードウェア化には適した機能であると言える。また，ホストコンピュータに送出する際に，膨大なデータ量を大幅に削減することが可能となり，システム効率上のメリットも大きい。

データベースの形式は，ＤＢＡの有無に係わらずデータベース構造の共通化を図るために，特にハードウェア化を意識した構造とはしなかった。

したがって，ページ形式はもっとも一般的なスロット方式を採用し，レコード形式は，ＩＤや長さを表す固定部を先頭に，固定長部，可変長部，さらに，各カラムにＩＤを有するカラムワイズ部により構成される構造とした。

(1) 述語比較

述語判定は，“＝”，“＜”，“＞”及びこれらの否定である“≠”，“≧”，“≦”を基本とし，文字タイプでは部分文字列比較（％），任意文字指定（＿）及びＥＳＣＡＰＥ文字をサポートした。

(2) データ形式

データ形式については，以下のものをサポート対象とした。

・文字形式（１，２バイト）
・ＩＮＴＥＧＥＲ形式
・ＤＥＣＩＭＡＬ形式
・ＮＵＭＥＲＩＣ形式
・ＦＬＯＡＴ形式

(3) 選択条件

選択条件判定は，述語のＡＮＤ／ＯＲによる任意の論理式にもとづいて判定する。

(4) ナル処理

目的のカラムが存在しない場合やＮＵＬＬタグ値がゼロで無い場合には，カラムをナルと見なした上，述語比較及びカラム抽出を実施する。

(5) カラム抽出

カラム抽出では，指定された任意のカラムを抽出し，新たなレコードを作成する機能をサポートした。

(6) 集計処理

集計処理については，取り合えずもっとも効果の期待出来る以下の機能を実現した。

・ＣＯＵＮＴ（＊）
・ＣＯＵＮＴ（項目）
・ＳＵＭ（項目）

但し，ＦＬＯＡＴ形式については，精度の問題があり，ＳＵＭ（項目）の実現は見送った。

4. ソフトウェア・インタフェース

ハードウェアにてデータベース処理を実施するためのソフトウェアとのインタフェースを下記に述べる。

(1) カラム情報

サーチ処理を実施する場合には，ハードウェアにおいてデータベースをカラム単位まで解読する必要がある。レコード単位の処理はページ内に格納されたレコードポインタを参照することにより可能であるが，カラム単位の処理を実施するためには，カラム属性，データ形式及びデータ長等を表す情報が必要となる。

したがって，予めＲＤＢから受取った複数のカラム情報を装置内に格納しておき，データベース読出し時に，対応するカラム情報と照合することによって，カラム単位の識別を可能とした。

(2) セレクション情報

述語判定を行うためには，比較対象カラム毎に，比較演算子，比較データ，比較データタイプ，位取り等の情報が必要である。さらに，選択条件判定のためには，述語のＡＮＤ／ＯＲによる論理式を何らかの形でＲＤＢより受け取る必要がある。今回は，ＤＢＡ処理の依頼時に，セレクション関連の情報を受け取ることとした。

論理式の条件判定は，ハードウェアに適した高速処理方式を検討した結果，事前にＣＮＦ／ＤＮＦ形式に展開されたものを受け取りそれをハードウェアにて処理する方式を採用した。

(3) プロジェクション情報

カラムの抽出処理を実施するためには，セレクション情報同様，ＲＤＢより事前に抽出条件を受け取る必要がある。

固定長部及び可変長部については，格納順に並べられたビットマップにより抽出条件を指定し，カラムワイズ部については，抽出対象となるカラムＩＤを指定することとした。

尚，集計処理の対象カラム指定は，インタフェース情報量を削減する目的で，プロジェクション情報を併用することとした。

(4) 処理結果通知

ＤＢＡの有無に係わらず，ＲＤＢ側での処理の共通化を図るために，サーチ結果の通知形式は，格納時と同一な構造とした。したがって，ＤＢＡ側は処理結果を，スロット方式により構成されるページ形式にまとめた後，ページ単位にＲＤＢへ送出する。

5. ハードウェア構成

図２にディスク制御装置の内部構成図を示す。チャネルアダプタ（ＣＡ）はホストとのコミュニケーションを分担するモジュールであり，デバイスアダプタ（ＤＡ）は磁気ディスク装置を制御するモジュールである。

シェアドストレージ（ＳＳ）はキャッシュ等に使用される大容量メモリであり，カラム属性等を定義するカラム情報はＳＳ内に保持されている。

一般のＩ／Ｏ処理及びＤＢＡを使用しない処理では，ＤＡ⇨ＳＳ⇨ＣＡを経由してホストコンピュータへデータを送出するが，ＤＢＡ使用時にはＳＳに格納されたデータがＤＢＡにおいて一度加工された後，ホストコンピュータへ送出される。

一つのディスク制御装置において，同時に複数のデータベース処理が実行可能なように，本ディスク制御装置では複数のＤＢＡが搭載可能な構造をとっている。

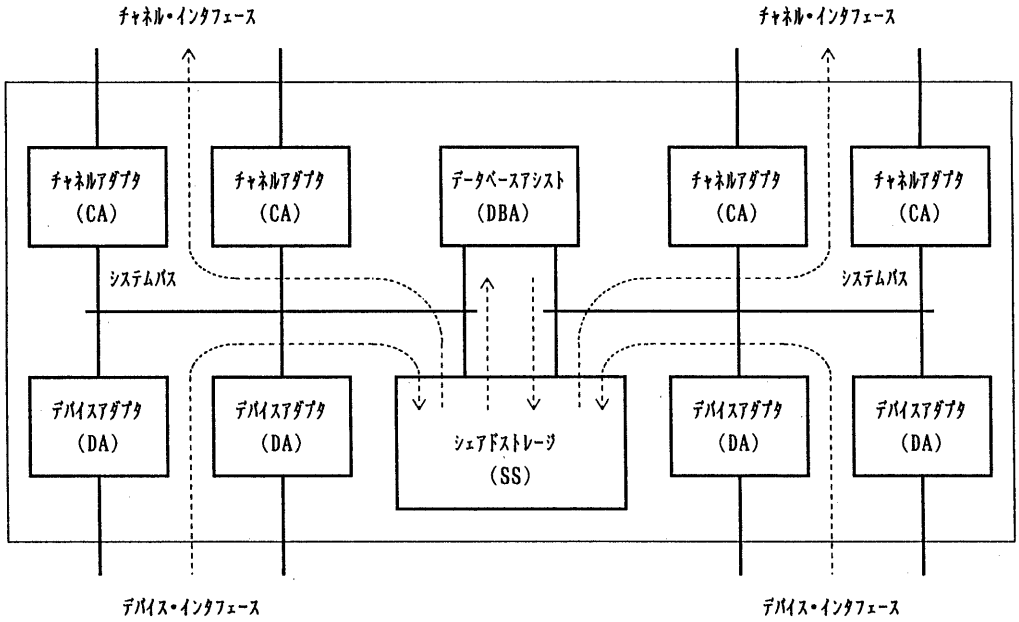

図２　ディスク制御装置内部構成

ＤＢＡの内部は，図３に示すように以下の論理ブロックにより構成されている。

- 入力ページバッファ
- 出力ページバッファ
- 解析ロジック
- 抽出ロジック
- 格納ロジック
- 選択ロジック
- 集計ロジック

(1) 入力ページバッファ

サーチ処理を実施すべき入力データが，システムバスを経由してページ単位に格納されるバッファであり，最大６４Ｋバイト分の入力ページが格納可能である。

入力ページ以外にも入力データに対応したカラム情報が最大６４Ｋバイト格納される。

(2) 出力ページバッファ

サーチ処理結果である出力データがページ単位に格納されるバッファであり，最大６４Ｋバイト分の出力ページが格納可能である。

出力ページ以外にも対象となるサーチ処理に対応したセレクション情報，プロジェクション情報及び集計結果が格納される。

(3) 解析ロジック

解析ロジックは，バッファ内のカラム情報をもとに，バッファに格納された入力ページについてレコード及びカラムの解析処理を行う。解析された入力ページはレコード単位，カラム単位に分解され抽出ロジックへ送出される。

また，カラム内の実データが格納されているアドレス等，カラム比較処理に必要となるカラム関連情報を選択ロジックへ送出する。

(4) 抽出ロジック

抽出ロジックは，出力ページバッファ内のプロジェクション情報をもとに，解析ロジックより受け取ったカラムの中から抽出対象となるカラムを選択し，格納ロジックへ送出する。

(5) 格納ロジック

格納ロジックは，抽出ロジックより受け取ったカラムを順次出力ページに格納し，新たなレコードを作成する。

さらに，選択ロジックより通知されたレコードの選択結果をもとに，選択対象となるレコードで新たなページを出力ページバッファ内に作成する。

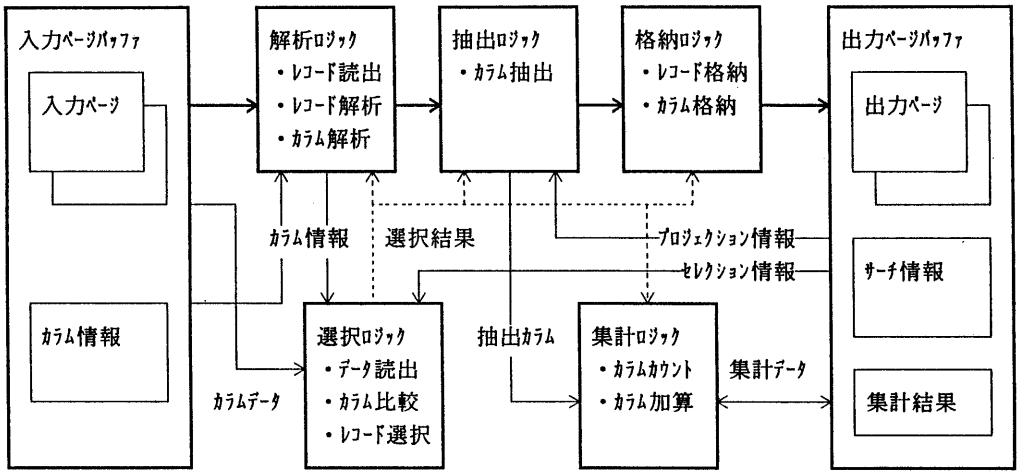

図３　ＤＢＡ内部構成

(6) 選択ロジック

選択ロジックは，出力ページバッファ内のセレクション情報と解析ロジックより受け取ったカラム関連情報をもとに，入力ページ中のカラム内データの比較処理を行い，レコードの選択結果を他のロジックへ通知する。

(7) 集計ロジック

選択対象となったレコードについて，カラム数のカウント処理及びカラム内容の加算処理を行い，集計結果を出力ページバッファ内に格納する。

図4にＤＢＡ内部の動作シーケンス例を示す。解析，抽出，格納，選択，集計といった各々の処理動作は，レコード単位に並列に実行される。

本例では，レコードN及びN＋2は選択条件を満足したレコードであり，格納処理及び選択処理が完了するのを待って，次のレコード処理へ移行する。レコードN＋1は選択条件を満足しないレコードであり，処理途中に選択対象外であることが判明した場合には，そのレコードに対する全ての処理を速やかに中断して次のレコード処理へ移行する。また，集計処理は，抽出処理及び選択処理の完了後，動作が開始される。

各ロジック部は，物理的に３つのＬＳＩにより構成され，システムバスとのインタフェース制御ロジックや全体動作を管理するマイクロプロセッサを含めて，１枚のＰＣＢ（Printed Circuit Board）に搭載されている。

表１に各ロジック部のＬＳＩテクノロジーを示す。

表１　ＬＳＩテクノロジー

プロセス		富士通製ＣＭＯＳ－ＬＳＩ（０．８μ，メタル２層）
ゲート数	解析ﾛｼﾞｯｸ部	２３，８９４ゲート
	抽出ﾛｼﾞｯｸ部＋格納ﾛｼﾞｯｸ部	２２，１０１ゲート
	選択ﾛｼﾞｯｸ部＋集計ﾛｼﾞｯｸ部	４８，８２７ゲート＋２ＫビットＲＡＭ
パッケージ		２５６ピン，ＰＧＡ

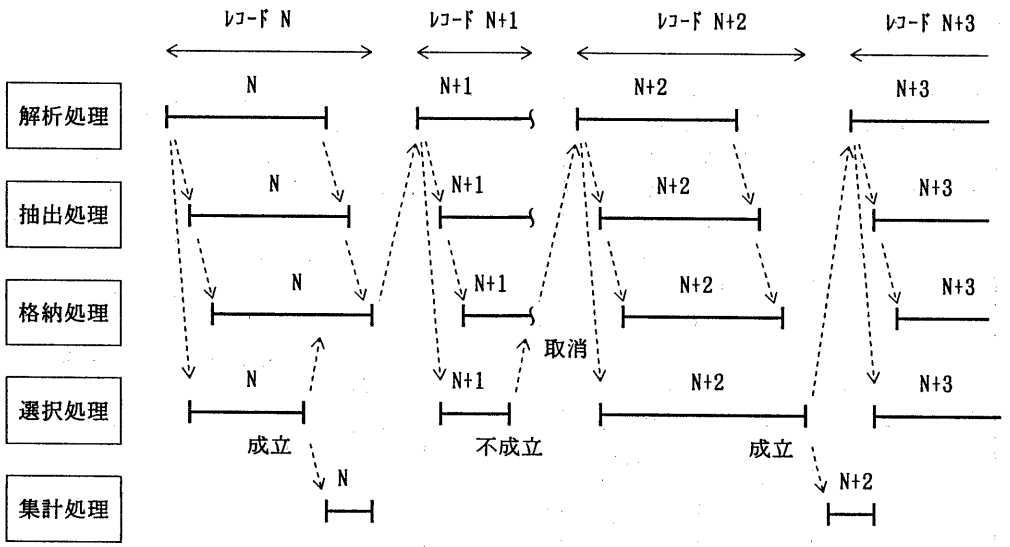

図４　ＤＢＡ内部動作シーケンス例

6. 選択条件判定方式

レコードの選択処理では，ハードウェア回路による高速処理を実現するために，論理式を事前にＣＮＦ／ＤＮＦ形式に展開したものを処理する方式を採用した。但し，一般的にソフトウェアで使用されている方式は，ハードウェアには不向きである。本章では，ハードウェア化に適したレコードの選択条件判定方式について述べる。

(1)は簡単な条件式の例であり，(1)をＣＮＦ形式に展開したものが(2)である。

```
(a1=1989 & a2>10 or a1>1989) &
(a1=1991 & a2<7  or a1<1992) &      ┐
(a3=XXX )                           ┘ (1)
```

```
(a1=1989 or a1>1989) & ..... ①      ┐
(a2>10   or a1>1989) & ..... ②      │
(a1=1991 or a1<1992) & ..... ③      ┘ (2)
(a2<7    or a1<1992) & ..... ④
(a3=XXX ) ..... ⑤
```

表２は，(2)の式を実行形式に変換した表であるが，ソフトウェアによる処理を前提とした一般的な構造である。ＣＮＦ形式では，少なくともある１つのジャンクションについての条件が不成立の場合に，全体が条件不成立となる。したがって，ソフトウェアでは，１つ１つのジャンクションを順次処理していく方式が一般的である。

但し，この場合，項目の出現が順不同となり，前のカラムに戻って処理をする必要が出てくる。したがって，これをハードウェアで処理するためには，各カラムの格納位置を保存しておき，再度同一カラムを読み出す等の処理が必要となる。これはハードウェアの物量及び高速化の観点より，ハードウェア化には適していないと言える。

表２　条件判定実行表（ｿﾌﾄｳｪｱ用）

CNF/DNF種別	ｼﾞｬﾝｸｼｮﾝ数	-> 5
ｼﾞｬﾝｸｼｮﾝ･ﾎﾟｲﾝﾀｰ1		
ｼﾞｬﾝｸｼｮﾝ･ﾎﾟｲﾝﾀｰ2		
ｼﾞｬﾝｸｼｮﾝ･ﾎﾟｲﾝﾀｰ3		
ｼﾞｬﾝｸｼｮﾝ･ﾎﾟｲﾝﾀｰ4		
ｼﾞｬﾝｸｼｮﾝ･ﾎﾟｲﾝﾀｰ5		
a1項目	＝条件	1989
a1項目	＞条件	1989
a2項目	＞条件	10
a1項目	＞条件	1989
a1項目	＝条件	1991
a1項目	＜条件	1992
a2項目	＜条件	7
a1項目	＜条件	1992
a3項目	＝条件	XXX

表３　条件判定実行表（ﾊｰﾄﾞｳｪｱ用）

CNF/DNF種別		ｼﾞｬﾝｸｼｮﾝ数	-> 5	
a1項目		比較ﾃﾞｰﾀ数		6
＝条件	1989	ｼﾞｬﾝｸｼｮﾝ番号	-> 1	
＞条件	1989	ｼﾞｬﾝｸｼｮﾝ番号	-> 1	
＞条件	1989	ｼﾞｬﾝｸｼｮﾝ番号	-> 2	
＝条件	1991	ｼﾞｬﾝｸｼｮﾝ番号	-> 3	
＜条件	1992	ｼﾞｬﾝｸｼｮﾝ番号	-> 3	
＜条件	1992	ｼﾞｬﾝｸｼｮﾝ番号	-> 4	
a2項目		比較ﾃﾞｰﾀ数		2
＞条件	10	ｼﾞｬﾝｸｼｮﾝ番号	-> 2	
＜条件	7	ｼﾞｬﾝｸｼｮﾝ番号	-> 4	
a3項目		比較ﾃﾞｰﾀ数		
＝条件	XXX	ｼﾞｬﾝｸｼｮﾝ番号	-> 5	

表３は，ハードウェアによる処理を前提とした実行形式表である。本表は，同一項目に関するものをまとめた構造をとっている。この構造のもとでは，項目順に比較処理を行い，その結果を対応したジャンクションに反映していくことになる。そして，ＣＮＦの場合，全ジャンクションのＡＮＤ条件を常に監視し選択条件を判定する。

本方式では，ＣＮＦ形式の場合，全てのジャンクションの条件が成立した場合にのみ，全体の条件が成立となる。したがって，本来のＣＮＦ／ＤＮＦ方式の目的とは異なってくる。しかしながらジャンクション内は少なくとも１つの条件が成立すれば十分であるため，条件式によっては，本方式の方が速いケースもある。また，もっとも重要なポイントは，本方式では，格納順にカラムを処理していくことが可能になるとともに，１つのカラムについて同時に複数の比較処理を実行することも可能になる点である。

7. ハードウェア性能

ＤＢＡのデータ処理能力は，最低限磁気ディスク装置のデータ転送速度を満足する必要がある。現在，一般的な大型磁気ディスク装置の最大データ転送速度は４．５ＭＢ／Ｓであるが，磁気ディスク装置のテクノロジィーの進歩，ディスクアレイを初めとする新しいアーキテクチャの確立により，データ転送速度は飛躍的に向上する可能性がある。したがって，処理速度は平均１０ＭＢ／Ｓ以上を目標値とした。また，十分な処理能力があれば，１つのＤＢＡで複数の処理を時分割的に実施することも可能になる。

以下に，ＤＢＡ単体としての性能評価結果を示す。表４は使用したデータベースの仕様であり，表５は，表４のデータベースをもとにした処理速度である。

測定結果からは，プロジェクション条件は処理速度には影響しないことが判る。これは，サーチ条件に係わらず，全カラムについて解析処理を実施するためである。処理速度に影響する条件は，実際に比較されるカラムの数である。今回の評価では意図的に全ての比較対象カラムに対する処理が完了しないと全体の条件が判定できない設定としたが，実際には全ての比較対象カラムを処理する必要が有るケースは少ないため，実際の処理速度は表５に示した結果よりも若干向上すると言える。

ＤＢＡ内部での処理速度は，セレクションの条件によって大きく左右されるが，一般的なサーチ条件下では，９～１２ＭＢ／Ｓの範囲内に収まると言える。

表４　評価用データベース仕様

ページ長	４Ｋバイト
ページ数	５
レコード数	１０（ページ当たり）
カラム数	１００（レコード当たり）
データ形式	文字，数値混在

表５　処理速度測定結果

サーチ条件		測定結果	
ﾌﾟﾛｼﾞｪｸｼｮﾝ対象ｶﾗﾑ	ｾﾚｸｼｮﾝ対象ｶﾗﾑ	処理時間（msec）	処理速度（MB/s）
全カラム	条件無し	０．９７	２０．６
３カラム	条件無し	０．９９	２０．２
全カラム	５カラム	１．６８	１１．９
３カラム	５カラム	１．６５	１２．１
３カラム	10カラム	２．１５	９．３
３カラム	全カラム	３．６４	５．５

8. まとめ

ＤＢＡの単体性能は，当初の目標値をクリアしているが，システムとして十分な性能値を得られるかが，今後のポイントである。また，同一装置内における一般Ｉ／Ｏ動作との混在動作時や複数のＤＢＡ動作時の性能評価も必要である。

ハードウェア化については，今後一層大規模なＬＳＩが使用可能になるにつれて，機能拡張も可能となってくるであろう。また，ハードウェアの弱点である論理変更の困難性については，今後，プログラマブル・ゲートアレー等を活用することにより，データベース構造や機能の変更に対しても，柔軟に対応していくことが可能である。