解　説

データベースプロセッサ

データベースアシスト DBA†

工　藤　哲　郎††

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[1),2)].

本文では，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[3)]．ハードウェア化を実現するうえでは，幾つかのデータベース機能が想定できるが，今回は一般的で使用頻度が高く，効果が期待できるサーチ関連機能の実現について述べる．

図-1 に DBA のシステムにおける位置付けを示す．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，一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．ただし，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などを解決する必要がある．

† The Data Base Assist by Tetsuro KUDO (File System Division, Fujitsu Limited).

†† 富士通(株)ファイルシステム事業部

2. 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，選抜，射影及び集約の三つである．これらの機能は，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，ページ形式は最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．表-1 に今回実現を試みたデータベース機能を示す．

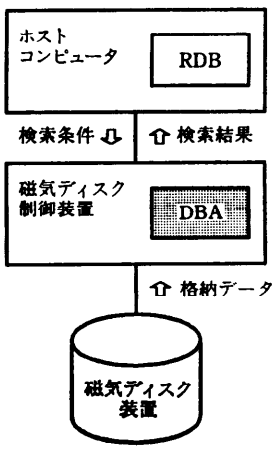

図-1　システム構成

表-1 DBA のデータベース機能

述語比較	=, <, >, ≠, ≤, ≥
データ形式	文字, INTEGER, DECIMAL, NUMERIC, FLOAT
選択条件	述語の AND/OR による任意の論理式
ヌル処理	カラム不在処理, NULL タグ処理
カラム抽出	指定された任意のカラム抽出
集計処理	COUNT(*), COUNT(項目), SUM(項目)

3. ソフトウェア・インタフェース

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などを表す情報が必要となる．したがって，あらかじめ RDB から受け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．

述語判定を行うためには，比較対象カラムごとに，比較演算子，比較データ，比較データタイプ，位取りなど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 による論理式をなんらかの形で 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 (Conjunct Normal Format)/DNF (Disjunct Normal Format) 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．

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．なお，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．

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．

4. 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[4]一般の I/O 処理及び DBA を使用しない処理では，DA (Device Adaptor)⇒SS (Shared Storage)⇒CA (Channel Adaptor) を経由してホストコンピュータ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．

DBA の内部は，図-3 に示すように複数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にはサーチ処理を実施すべき入力データ及びカラム情報がシステムバスを経由して格納される．出力ページバッファにはサーチ処理結果である出力データが格納され，システムバスを経由して外部に送出される．また，出力ページ以外にもセレクション情報，プロジェクション情報及び集計結果が格納される．解析ロジックではバッファに格納された入力ページについて解析処理を行い，レコード単位及びカラム単位に分解された後，抽出ロジックへ送出される．抽出ロジックで

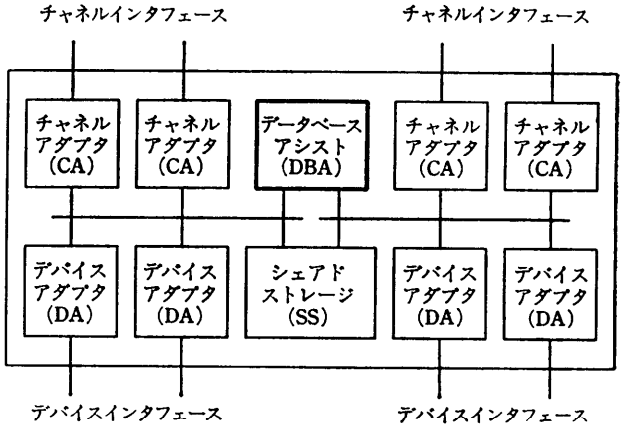

図-2 磁気ディスク制御装置の内部構成

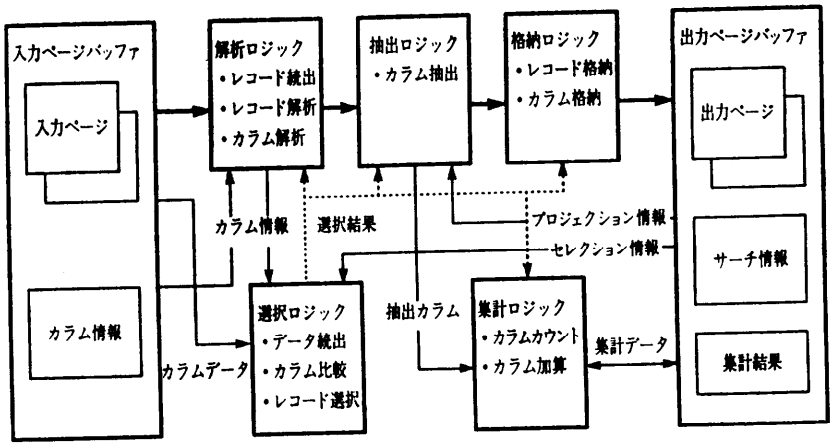

図-3 DBA 内部構成

は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．格納ロジックで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バッファに格納し，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により新たなページを作成する．選択ロジックで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．集計ロジックでは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．

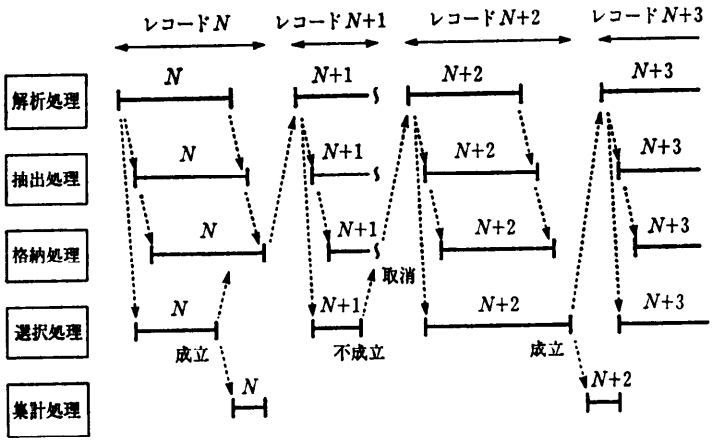

図-4 DBA 内部動作シーケンス例

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，集計といったおのおの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．

5. 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．ただし，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) &　(1)
(a3=XXX)

(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

図-5 は，(2)の式を実行形式に変換した表であり，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，少なくともある一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，一つ一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．ただし，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など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．

図-6 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

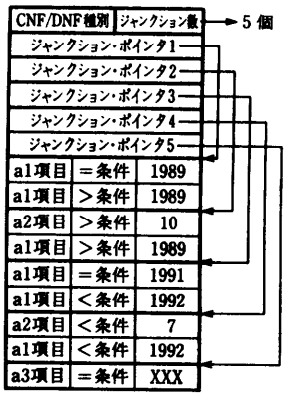

図-5 条件判定実行表（ソフトウェア用）

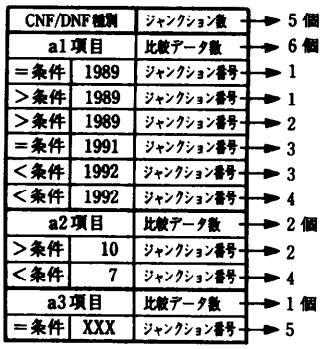

図-6 条件判定実行表（ハードウェア用）

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一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，本方式のほうが速いケースもある．また，最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，一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．

6. 性能評価結果

表-2 は約34 MIPSの汎用コンピュータシステムにおいて，2万件のデータから2千件を抽出した場合のDBAありとなしの性能測定結果を示したものである．DBAなしの場合には，チャネルビジー率はJOBの多重度が上がるに従って上昇

表-2 性能測定結果

測定項目	DBAの有無	JOB多重度				
		1	2	4	6	8
チャネルビジー率 (%)	なし	12.4	25.9	41.9	48.9	55.3
	あり	0.5	2.3	4.9	5.5	5.9
CPU処理時間 (msec)	なし	9.2	39.0	78.4	117.1	156.7
	あり	3.7	7.4	14.5	21.8	29.1
データベースJOB実行時間 (msec)	なし	131.1	135.3	163.2	208.3	252.5
	あり(1個)	107.3	108.8	123.7	165.5	218.6
	あり(2個)	107.4	108.5	111.1	123.4	152.4
一般JOB実行時間 (msec)	なし	—	220.7	241.4	281.5	327.4
	あり(1個)	—	166.3	173.2	183.8	224.8
	あり(2個)	—	165.9	168.1	173.9	188.5

するが，DBA を使用した場合には，ほとんど変化はみられない．また，DBA 使用時の CPU 処理時間は DBA なしの場合に比べて約 1/5 程度となるという結果が得られた．これらはデータ転送量及び CPU 負荷の軽減という DBA の目的を立証するものである．データベース JOB の実行時間は，1 DBA の場合に約 20%，DBA を二つ搭載した場合には最大 40% 短縮される．さらに，データベース JOB と一般の JOB を混在して実行させた場合に，一般 JOB の I/O 処理時間も改善されるという結果が得られた．

上記と同様の評価環境下で，ソフトウェアによる検索時間と DBA の内部処理時間を比較したところ，DBA のほうが常に 4～6 倍の性能を出しているとの数値を得られた．そして，その性能差は条件式が増えるほど顕著となる傾向にある．DBA のマシンサイクル速度は CPU の 1/10 以下にもかかわらず，これだけの結果を得られた点は専用ハードウェア化の効果であると言える．

7. まとめ

DBA のシステム性能は，当初の目標を満足するものであり，現在の汎用コンピュータシステムで十分効果を発揮すると言える．また，今回は比較的大規模なシステムにおける評価を行ったが，CPU 性能がそれほど高くない中規模以下のコンピュータシステムでは，DBA の効果は一層顕著になると言える．

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，さまざまな条件下における性能評価を実施することにより，システム性能をさらに向上させるためのポイントを解明していく必要がある．

参考文献

1) 喜連川，中野：データベース・マシン，bit 臨時増刊，Vol. 21，No. 4，(1989 年 3 月)．
2) 金子：磁気ディスク高性能化技法，電子情報通信学会，技術研究報告 DE 89-36，(1989 年 12 月)．
3) 工藤：データベース処理のハードウェア化，情報処理学会，研究報告 91-ARC-90-6 (1991年10月)．
4) 小池他：F 6427 H 磁気ディスクサブシステム雑誌 FUJITSU，Vol. 42，No. 1 (1991 年 1 月)．

(平成 4 年 6 月 29 日受付)

工藤 哲郎

1958 年生．1981 年上智大学理工学部電気電子工学科卒業．同年富士通(株)入社．以来，同社ファイルシステム事業部に所属し，磁気ディスクコントローラ及び磁気ディスクキャッシュなどの開発に従事．