Vol. 33 No. 12　　情　報　処　理　　Dec. 1992

解　説

データベースプロセッサ

データベース演算処理装置 DBE†

松田　進††　井上　栄†††
島川和典††††　山田朝彦††††

1. はじめに

近年の情報量の急速な増大にともない，大規模なデータを高速に処理できることへの要求が高まってきている．単なるデータ量の増大のみならず，同時に，情報相互間の関連も一層複雑化しており，関係の深いデータを有機的に関連づけて，高速に検索する必要性も出てきている．このような大量データの処理には，大別してソート処理と，関係データベース (RDB) 処理とがある．ソート処理では，数十万件〜数百万件のレコードを一度に並べ替えることが必要となるため，いかにしてディスクと主記憶間のデータ転送と並べ替え処理とを並列化するかがポイントとなる．

一方，RDB 処理においては，その代表的演算である結合演算と選択演算の高速化技術がキーとなる．一般に，非定型問合せでは，索引を利用できないために，データベース全体をサーチする必要があり，特に，情報相互の関連付けでは，複数のテーブル（リレーション）全部をアクセスして，負荷の重い結合演算を実行する必要がある．その結果，ディスクと主記憶間のデータ転送の増大，CPU の負荷の増大，さらに主記憶の大量消費が起こることとなり，最終的に，ホストの負荷が増大し，システム全体のスループットを著しく低下させている．

このような背景のなか，著者らは，(財)新世代コンピュータ技術開発機構 (ICOT) 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[1)〜3)]．データベース演算処理装置 (DBE) は，この研究の成果をもとに，TP ホストコンピュータ TP 90/70 モデルのソート及び関係演算処理専用の付加プロセッサとして製品化したものである[4)]．本稿では，DBE の機能概要，設計目標，ハードウェア構成，ソフトウェア構成，性能評価，利用技術について解説し，最後にまとめを述べる．

2. DBE の機能概要

前述のように，大量データの一括処理には，ソート・マージ処理や，索引の利用できない非定型業務があげられるが，これらは必ずしも常にバッチ処理的に行われるわけではない．たとえば，複雑な条件によるアドホックな問合せや，いろいろな観点からの統計値の問合せは，条件や対象データを臨機応変に設定しつつ，会話的に実行されることが多い．しかし，このような全データベース探索を必要とする処理は，システムの全体性能を大きくダウンさせる原因となっている．

DBE は，このような大量データの一括探索処理の高速化を目的として開発したものである．ソート・マージ及び RDB の選択・射影・結合などを，並列ハードウェアソータを軸として，高速に処理することができる．前者のソート・マージについては，ホスト計算機上のソートユーティリティの機能を，また後者の RDB については，同じく RDB/V と呼ぶ関係データベース管理システム (RDBMS) の機能を，DBE 上に実装したものである．

ホストの提供するソート機能は，通常のソートやマージのほかに，対象データを一つのテーブルに見立てた場合の，RDB の演算機能を含んでい

† A Database Operation Processor: DBE by Susumu MATSUDA (Computer Hardware Development Dept., Ome Works, Toshiba Corp.), Sakae INOUE (First Computer Software Development Dept., Ome Works, Toshiba Corp.), Kazunori SHIMAKAWA and Asahiko YAMADA (Advanced Computer Architecture Dept., Information Systems & Engineering Laboratory, Toshiba Corp.).

†† (株)東芝青梅工場コンピュータハードウェア開発部
††† (株)東芝青梅工場第一コンピュータソフトウェア開発部
†††† (株)東芝情報処理・機器技術研究所開発第七部

る．それら RDB の機能には，たとえば，複合条件による選択処理や，レコード内のフィールドの並べ替え，フィールド長の変更，フィールドの削除といった射影処理のほかに，特定のフィールドでグループ化し，グループごとに集計する機能もある．また，ソートの対象となるファイル形式には，順編成，相対編成，索引編成などがある．DBE は，これらの機能を完全にサポートしている．

一方，RDBMS である RDB/V は，データベース言語である SQL (JIS X 3005-1990)[5] に準拠した，SQL/V と，COBOL を親言語とした埋め込み SQL をサポートしている．表-1 は，DBE の持つ主な RDB 機能をまとめたものである．以下に，DBE サポートの特徴を述べる．

（1） DBE とホストの機能分担

DBE は，SQL/V と埋め込み SQL の機能を，ホストと連携・分担して処理する．DBE とホストの機能分担の例をあげると，埋め込み SQL のカーソルのオープン時の作業テーブルの生成は，DBE が行い，カーソルによる検索は，1 レコードずつの処理であるため，ホストが行う．他の例では，条件による選択処理において，索引が使用可能ならば，ホストが選択し，そうでなければ，DBE が選択するという，一種の最適化も行っている．

（2） DBE の高い独立性

DBE は，ホストからのコマンドにより，演算を実行し，その結果を，ホストの指定したディスク上の作業テーブルに格納する．ホストは，その作業テーブルを入力し，次の演算を実行する．このように，DBE のホストからの独立性は高く，作業テーブルを介することにより，一つの SQL 文の処理を完遂する．DBE の独立性により，主記憶を大量に使用することがないため，ホスト側の RDB 制御プログラム (RDBCS) のバッファの管理が，ソフトウェア処理時の方式とまったく同じでよく，結果として，オーバヘッドのない効率的な処理が可能である．さらに，ホストの CPU の負荷も最小限におさえることができる．

（3） DBE の接続透過性

DBE のホストへの接続の有無は，ユーザアプリケーションに対して透過的である．RDB/V は，SQL 文を DBE で実行するかどうかの判断を実行時に動的に行う．これは，DBE 専用の SQL オブジェクトは生成せずに，ソフトウェア処理用の SQL オブジェクトがそのまま DBE 処理に適用できる方式を採用したことによる．したがって，DBE を利用するかどうかの判断を，RDBCS に任せるのであれば，埋め込み SQL のソースプログラムの再コンパイルは不要である．しかし，アプリケーションによっては，特定の SQL 文のみを DBE で実行させたい場合も考えられる．このため，DBE の使用レベルを，ジョブ単位，アクセス対象のデータベース単位，SQL 単位に設定できるようにしている．

（4） 非整合読み込み

大量データの統計処理は，非整合読み込みと呼ぶ同時実行制御方式により，DBE を利用して効率的に行うことができる．非整合読み込みは，更新中のデータベースでも，共用を許すものである．一般に，統計処理は非定型な問合せとして行われることが多く，データベース全体を処理しなければならないため，ホストの負荷が大きい．しかし，DBE を非整合読み込みで使用することにより，ホストの負荷の軽減を図ることができるうえ，高速な検索処理が可能となる．

（5） 索引生成

索引は，定型問合せに頻繁に使用される．索引生成には，前処理としてソートが必要であるが，これは，DBE を利用して，高速に実行することができる．また，定型問合せでは，挿入，削除，変更などの更新処理も多く行われるため，索引ファイルの定期的なメンテナンスが性能上重要となってくる．このような場合にも，索引の再作成を，DBE を利用して高速に行うことができる．

表-1 DBE の主要な RDB 機能一覧

主要機能	SQL/V, 埋め込み SQL の要素
選択	比較述語，BETWEEN 述語，NULL 述語，IN 述語
結合	比較述語
準結合	副問合せの結果との比較述語，IN 述語
射影	問合せ指定の選択リストの列指定
ソート	ORDER BY
グループ化	GROUP BY
ユニーク化	DISTINCT
カウント	COUNT (*)
索引生成	CREATE INDEX

（6） 疑似テーブル

RDB/V は，システムのデータ管理機能の提供する各種の編成のファイルをテーブルとして疑似的に定義し，処理する機能を持っている．これにより，RDB ファイルと通常のファイルとの共存を実現している．DBE は，RDB/V のサポートするすべての編成のファイルを処理できる．

3. 設計目標

DBE の設計目標を，第1に，ホストとの接続において独立性を高くすること，第2に，すべてのディスク装置上のデータのアクセスを可能とすることに置いた．これは，ホスト CPU と主記憶の負荷を，可能なかぎり最小におさえるためである．このために，DBE を入出力チャネルとディスク装置の間に接続し，ホストとは独立してディスクアクセスができる接続形態を採用した．この意味で，DBE はバックエンド型プロセッサとみることができる．

第3の設計目標は，DBE 内でのデータ処理において，並行性を十分に引き出すことである．このために，DBE とホスト/ディスク装置間の入出力処理，演算処理，全体制御を，それぞれ並行して実行できるように，三つのマイクロプロセッサ (MPU) を内蔵させた．

第4の設計目標は，大量のデータを一括して処理できるようにすることである．このために，32MB から最大 512MB の大容量共有メモリを実装可能とした．共有メモリを超えるデータ量に対しては，ホストが DBE に作業ファイルを与えて処理を行う，拡張処理方式をサポートした．

4. ハードウェア構成

DBE のハードウェアは，入出力処理用の MPU (EIP)，ソータ・関係演算部（演算部），選択処理及び演算部の制御を行う MPU (ECAM)，ホストから DBE に送られるコマンド (DBE コマンド）を解析し，DBE 全体の実行制御を行う MPU (ECP)，大容量共有メモリから構成される (図-1)．演算部は，演算対象キーフィールドのみをレコードから切り出し，データ型の変換を行い，それに共有メモリ上のレコードアドレスをレコード識別子（長さは4バイト）として付加したものを入力する，キー切り出し方式を採用しており，これに

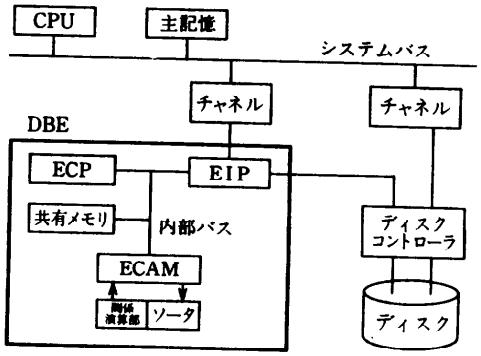


図-1 DBE のハードウェア構成

より，大量のレコードを一度に処理できる．演算部からの出力は，結果レコード識別子の集合であり，それをもとに，レコードを再構成することにより，最終の演算結果が得られる (図-2)．結合演算の場合には，レコード識別子の組が出力されるので，そこから新しいレコードを生成する．演算部のソータは，18段のソートセルからなり，関係演算プロセッサが，最終段のソートセルとして機能するので，全体では19段である．各セルの最大許容キー長は，12，28，…，$2^{S+3}-4$ (演算開始セル段数 $S=1, 2, \cdots, 19$) である．一回に処理できるレコード数は，ソートで 2^{19} 件，結合と制約（準結合）の関係演算で 2^{18} 件である．選択演算は，ECAM がソフトウェア処理として実行するので，件数の上限はない．

5. ソフトウェア構成

DBE は，ホスト上の2種類のソフトウェアにより制御される．一つはソートユーティリティプログラムであり，もう一つは RDBCS である．これらのソフトウェアは，DBE コマンドを用いて DBE を制御する．本章では，この二つのソフトウェアの構成と DBE の係わり，及び DBE の制御マイクロプロセッサである ECP 内のソフトウェア構成について述べる．

まず，ホストのソフトウェア構成について述べる (図-3)．ソートユーティリティプログラムの場合は，処理の単位はソートあるいはマージであるので，DBE に1回で DBE コマンドを発行している．これに対して，RDBCS の場合には，SQL を解析し中間言語に展開した後，中間言語単位でソフト処理か DBE 処理かを動的に判断している

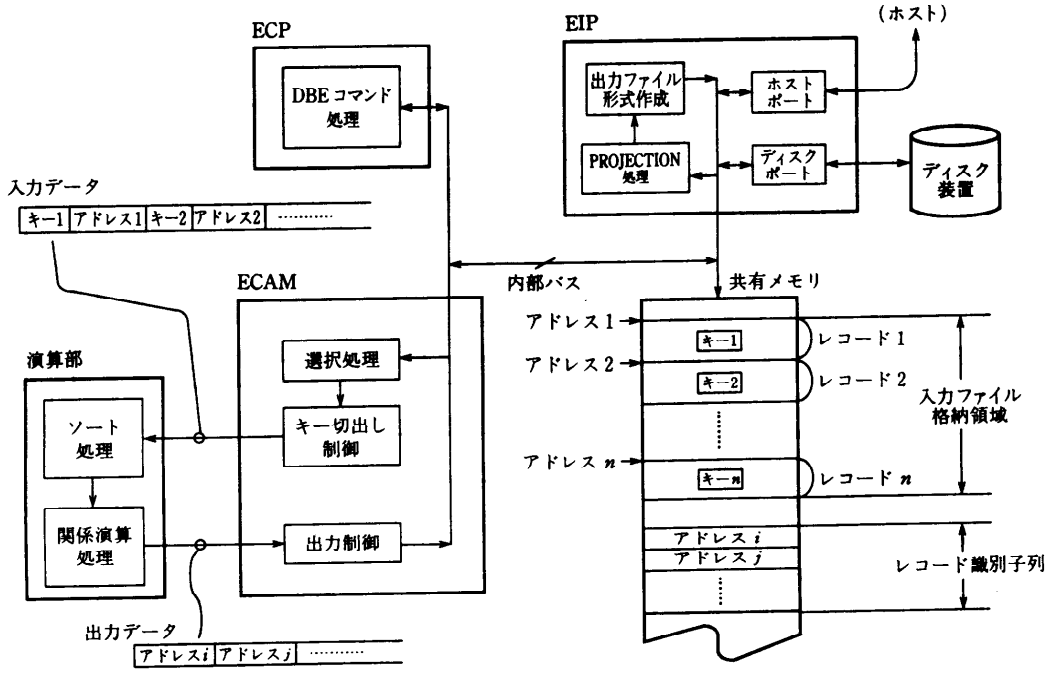
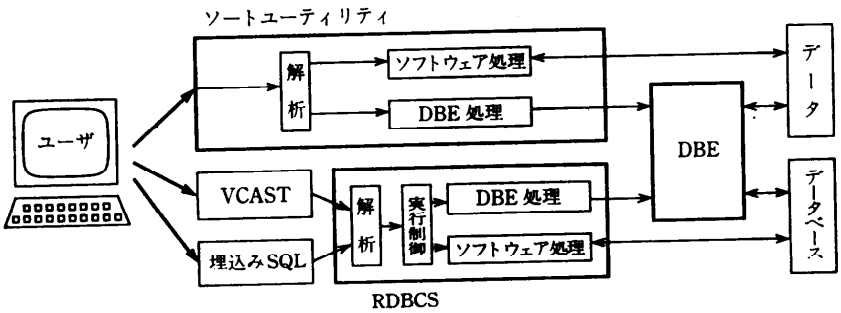


図-2 DBE の処理メカニズム



VCAST : RDB 用簡易ユーザインタフェース
埋込み SQL : プログラムに埋め込まれた SQL

図-3 ホストのソフトウェア構成

表-2 DBE コマンド一覧

	処　理	DBE コマンド
ソート	複数ファイルの一括入力	一括 READ コマンド
	並び替えとファイルの出力	SORT & PROJECTION コマンド
関係演算	テーブルの入力	READ コマンド
	探索条件中の選択	AND/OR 条件付きの SELECT コマンド
	〃　結合	JOIN コマンド (選択条件を同時に指定可能)
	副問合せ結果による選択	RESTRICT コマンド
	並び替え	SORT コマンド
	グループ化	SORT コマンド
	ユニーク化	ユニーク指定の SORT コマンド
	カラムの抽出とテーブルの出力	PROJECTION コマンド
	カウント	カウント指定の各コマンド
	索引の生成	SORT コマンド

ので，一つの SQL に対して複数個の DBE コマンドが必要となる．表-2 は，DBE コマンドの一覧を示したものである．DBE コマンドには，これ以外にも演算処理の中断や資源解放など，いくつかの運用性を高めるための制御コマンドがある．

次いで，ECP のソフトウェア構成に基づき，それぞれのモジュールと機能を処理順序に従って説明する（図-4）．ホスト上のソートユーティリティプログラムあるいは RDBCS から送られた DBE コマンドは，EIP を通して受け取る．その際の処理をホストインタフェースモジュールが行う．受け取った DBE コマンドがソート用であればソート制御モジュールが，RDB 用であれば RDB 制御モジュールが DBE を制御する．ソート処理も RDB 処理もデータ量やソータ段数，共有メモリ容量など DBE 内資源を要因として，複数の処理方式から選択的に実行する必要がある．その処理方式を表-3 に示す．各制御モジュールによって処理方式が決定されると，対応するモジュールにより，実際の処理が開始される．これらの処理モジュールは，入出力インタフェースモジュールを介して，ディスクからデータを入力し，ECAM インタフェースモジュールを介して，ソートや関係演算の実行を ECAM に指令する．ディスクへの出力は，ECAM からの処理結果に基づき，出力データを生成し，再び入出力インタフェースモジュールを介して行う．モニタは，メモリ管理や割り込み管理のほかに，上記モジュールのスレッド化による並行動作の制御を行う．

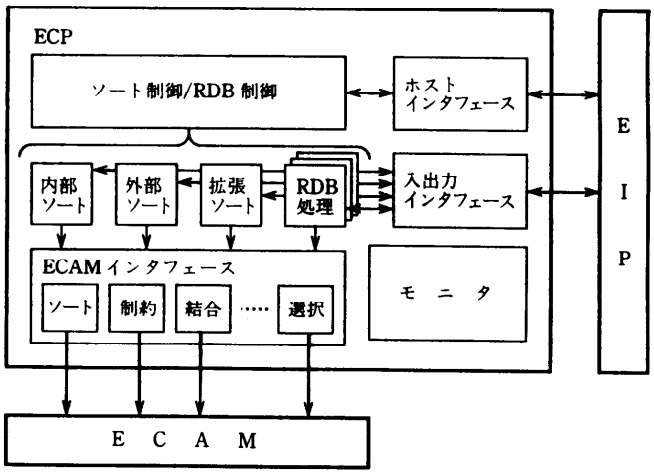


図-4　ECP のソフトウェア構成

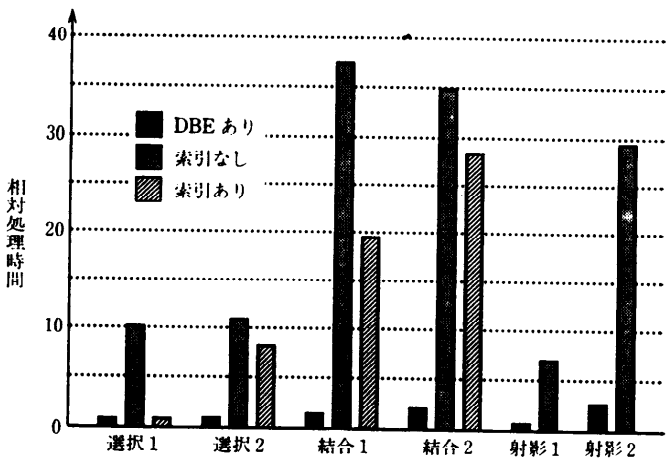
表-3　DBE の処理方式

処理方式	処理方式の説明
通常処理	処理対象及び結果の総データ量が共有メモリのサイズを超えない場合の処理
・内部処理	総レコード数と総キー長の組み合わせが演算部のハードウェア定数を超えずに 1 回で処理できる場合の処理
・外部処理	ハードウェア定数を超えるために ECP によるマージが必要となる処理
拡張処理	処理対象及び結果の総データ量が共有メモリのサイズを超える場合の処理（作業ファイルを使用する）

6. 性能評価

（1）　ソート処理

DBE でのソート処理とホストの最上位モデルでのソフトウェアによるソート処理の性能比は，通常処理時で 3～5 倍，拡張処理時で 1.5～2 倍の高速化がはかれた．



選択 1：10 万件→1,000 件＝1,000 件
選択 2：10 万件→1 万件＝1 万件
結合 1：(10 万件→1 万件)×10 万件＝1 万件
結合 2：1 万件×(10 万件→1 万件)×(10 万件→1 万件)＝1 万件
射影 1：10 万件→1 万件（5 つのカラムを射影しユニーク化）
射影 2：1 万件→1 万件（全カラムを射影しユニーク化）
注）→は選択，×は結合，＝は結果を示す

図-5　DBE とホストの関係演算性能比

(2) 関係演算処理

拡張ウィスコンシンベンチマーク[6),7)]による関係演算の性能は，"索引なし" のホスト処理と比較して，選択で 11 倍以上，結合で 16 倍以上，射影で 10 倍以上が得られた (図-5)．"索引あり" のホスト性能は，選択率が小さいときほど高いが，逆に大きくなるにつれて低下していく．このことは，選択率がある値を超えるような処理では，DBE を使用したほうが相当に効果的であることを示している．

7. 利用技術

一般に，データベースの応用システムには2種類ある．一つは勘定系のシステムであり，もう一つは情報系のシステムである．勘定系のシステムは，たとえば顧客管理，在庫管理，販売管理といった従来からの定型業務が主体であり，高トラフィックなオンライントランザクション処理から大量データのバッチ処理に至るまで，その処理形態はさまざまである．これに対して，情報系のシステムの主要な目的は，戦略的かつ効率的な企業経営のための情報を，タイムリに提供することにある．このような意思決定支援のための情報を得るには，勘定系のデータベースに対して，統計的な分析処理などを施すことが必要となってくる．すなわち，非定型な問い合わせを高速に行うことが要求されてくる．

このような状況において，勘定系のシステムでの大量データの一括処理や，情報系のシステムでの統計処理は，データベース規模の成長とともに，ホスト計算機にとって非常に負荷の高いものとなりつつある．このような高負荷な処理に，DBE を利用することで，ホスト計算機本体の負荷を軽減するとともに，処理の高速化を可能とすることができる．特に，情報系のシステムでは，索引を予期して付けることができないため，データベース全体をサーチせざるをえない．このような場合に，DBE は非常に有効となる．

8. おわりに

DBE の機能概要，設計目標，ハードウェア/ソフトウェア構成，性能評価，利用技術について述べた．DBE の最大の特徴は，DBE とホスト間の演算コマンドインタフェースの機能単位を大きく設定し，ホストからの独立性の高いバックエンド型の DB プロセッサを実現したことにある．このような高いホスト独立性は，DBE が，演算対象のデータをディスクから直接入力し，演算結果を直接ディスクに出力することにより，達成している．高いホスト独立性の利点は，ホストの CPU 負荷の低減はむろんのこと，主記憶の使用率の極小化に加えて，RDBMS のバッファ管理を DBE 専用のものとする必要がないことにある．主記憶上のデータを直接アクセスし，結果を再び主記憶上に格納する方式の DB プロセッサでは，RDBMS に対して，複雑なバッファ管理のアルゴリズムが要求される．しかし，DBE では，ソフトウェアによる問合せ処理の方式とまったく同じで済んでいる．これにより，RDBMS のオーバヘッドの最小化を可能としている．今後は，さらなる一括検索処理の高機能化を目指していく予定である．

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) 松田，東郷，島川，岩崎：データベース演算処理装置のアーキテクチャ，情報処理学会研究会報告 91-ARC-90, Vol. 91, No. 86 (1991).
5) データベース言語 SQL JIS X 3005-1990，日本規格協会 (1990).
6) Bitton, D., Dewitt, D. J. and Turbyfill, C.: Benchmarking Database System—A Systematic Approach, Proc. VLDB (1983).
7) Dewitt, D. J.: A Performance Analysis of the Gamma Database Machine, 1988 ACM 0-89791-268-3/88/0006/0350.

(平成4年5月27日受付)

松田 道（正会員）

昭和23年生．昭和46年九州大学工学部通信工学科卒業．同年東京芝浦電気(株)（現(株)東芝）入社．同社青梅工場にて小型計算機，分散処理計算機のハードウェア開発に従事．

井上 栄（正会員）

昭和35年生．昭和57年千葉大学理学部数学科卒業．昭和60年東京都立大学大学院理学研究科数学専攻修士課程修了．同年，(株)東芝に入社．以来，分散処理計算機の基本ソフトウェア開発に従事．データ管理関係を主に担当．

島川 和典（正会員）

昭和26年生．昭和45年島根県立浜田高等学校普通科卒業．同年東京芝浦電気(株)（現(株)東芝）入社．同社青梅工場，情報処理・機器技術研究所において，データベース管理システム，データベースマシンの研究・開発に従事．

山田 朝彦（正会員）

昭和32年生．昭和56年国際基督教大学教養学部理学科卒業．昭和61年上智大学大学院理工学研究科数学専攻博士後期課程修了．同年，(株)東芝に入社．以来，情報処理・機器技術研究所においてコンパイラ及びデータベース処理エンジンの研究・開発に従事．