

# データベース システムの動向と評価

## Trends and Developments in Data Base Systems

データの共有と即時的な情報要求への対応を基本とするデータベース システムは、多様な情報処理システムの形態において、競合的な理念と技法を提供するまでに成長した。システムを構築するためには、データベース、データ コミュニケーション、データ アドミニストレーション及びエンド ユーザー言語などの多面的な基本機能が必要とされる。

我々は一般動向調査とユーザーの体験的要求に基づいて、データベース マネジメント システム PDM 及び ADM を開発した。また実用性の高い性能評価技法の提供をはじめとして、事例に基づいたシステム設計技法の蓄積を進めている。

酒井博敬\* *Hiroataka Sakai*  
 鮫島隆展\* *Takanobu Samejima*  
 松岡 潤\*\* *Hiroshi Matsuoka*

### 1 緒 言

データ処理部門の管理者によって共通的に認識されているデータベースとは、「定常業務処理プログラムにも、即時的なマネジメント ニーズに対しても使用できる共有化されたランダム アクセス ファイル」である<sup>(1)</sup>。データベース システムは、情報の通用性(currency)を求めるオンライン化の傾向とともに、その適用領域も実施レベルの定常業務処理から、計画レベルの戦略システムへの拡大が指向されつつある。システムの設計や運用、あるいは今後の改善や拡張に対する関心の増大とともに、基本ソフトウェアについても多面的な能力が求められるようになった。ここではデータベース システムに対するソフトウェア面からの我々のアプローチについて述べる。なお基本的な背景については、R.L.Nolanの著書<sup>(1)</sup>中に優れた記述がある。

### 2 一般動向

データベース システムは、多様な情報処理システムの形態において、競合的な理念と技法を提供するまでに成長した。企業におけるデータベース システムの導入は、その背景に業務機能間及び組織階層間のデータ統合化によるMIS (Management Information System) 指向や、システムの変化 成長に対する柔軟性の追求などがあるにせよ、現実には巨大化したファイルの維持問題や、良質データの適時な提供を求めるユーザー要望の顕在化を解決することが直接の動機となっている。

データベース システムのもたらすインパクトは、特にシステムの設計と運用における技術的及び管理的な問題に集中しているといえよう<sup>(2)</sup>。この領域は、システム性能の評価と調整、システム障害への対応、データ定義情報の管理、データ機密保護の維持、エンド ユーザー言語機能の提供などに関するものである。これらの諸問題はシステム運営の責任を負うデータ アドミニストレータの優れた能力を要求するとともに、ユーザーのためのサービス手段の提供への要望を顕在化させている。

システム負荷の観点からは、データの関連性や独立性が重視された結果として、データベース アクセスへの負荷が、端末からのメッセージ ハンドリングにかかるトランザクション負荷と同程度以上のウエイトを占めるようになる。安定した

システムの成長を目指すならば、システム利用環境も、バッチからデータのオンライン収集・問合せへ、更にデータベースのオンライン更新のレベルへと段階的な推移をたどることが望ましいとされている。この過程でデータベース負荷とトランザクション負荷のトレード オフを考えた性能評価と調整問題の解決が要求される。

データベース システムの構築においては、多面的な技術及び管理能力が要請されるが、この能力を提供又は支援する基本的なソフトウェアは、機能面から次の四つの分野に分けられる。

#### (1) データベース機能

データベース マネジメント システム (Data Base Man-

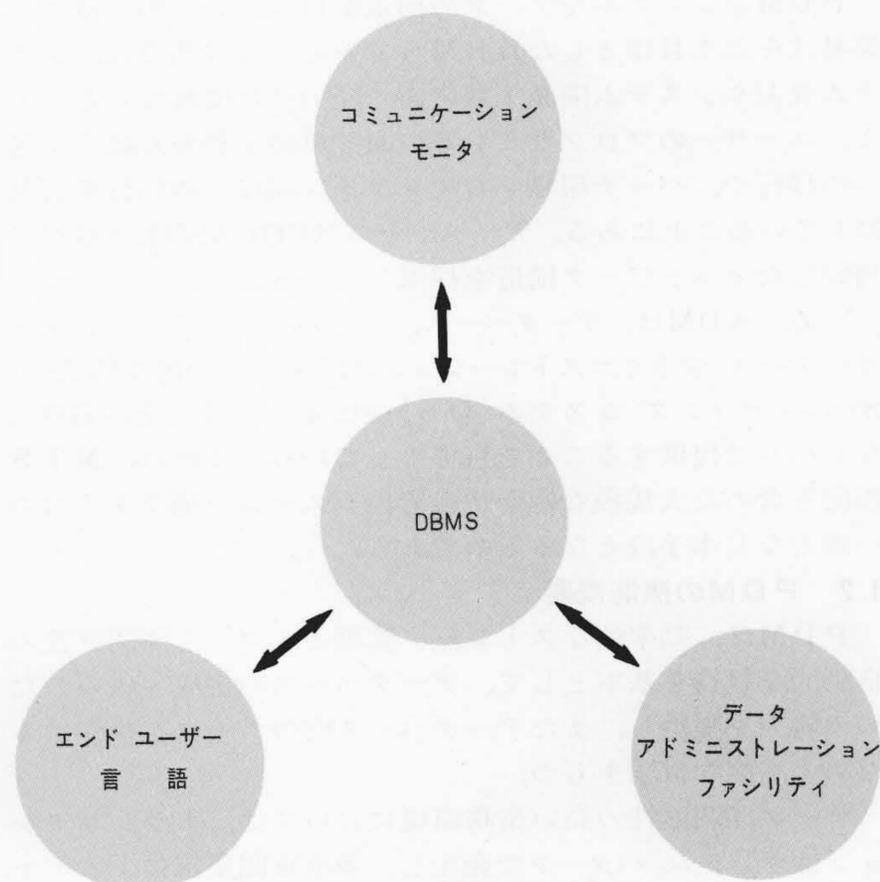


図1 データベース システム基本機能の分類 DBMSを中心とし、これとインターフェースをもつ他の三つの機能要素が集まって基本機能を構成する。

\* 日立製作所ソフトウェア工場 \*\* 日立製作所システム開発研究所

agement System, 以下DBMSと略す) がカバーする最も基本的な分野

#### (2) データ コミュニケーション機能

オンライン環境でデータベース アクセスを可能にするコミュニケーション モニタリング機能

#### (3) データ アドミネストレーション機能

データベース システムの運用・維持のための管理機能システムの定義・維持ユーティリティや、性能評価、データ定義情報管理並びにデバッグ及びテストなどのためのツールが必要とされる。

#### (4) エンド ユーザー機能

システム能力を広く利用するためのエンド ユーザー言語  
これら4種の基本機能に対する関心は、ユーザー経験の成熟につれて(1)から(4)の分野へ漸次移行しつつあると言えよう。また、これら基本ツールの利用に際して特に注意すべき問題は、上記(1)のDBMSに対する(2), (3), (4)の各機能のインタフェースであろう(図1参照)。

このような動向を背景として、我々はデータベース及びデータ コミュニケーション機能を提供する基本プログラムを開発した。また、システム性能の測定及び評価技法をはじめ、データ アドミネストレーション用各種サービス手段の開発を進め、更に豊富なデータベース システム構築事例に基づいたシステム設計技法の蓄積を行なっている。

### 3 日立製作所のデータベース マネジメント システム

#### 3.1 PDMとADMの目標

我々はそれぞれ異なる特性をもつ二つのタイプのデータベース マネジメント システムPDM(Practical Data Manager)<sup>(3)</sup>及びADM(Adaptable Data Manager)<sup>(4)</sup>を開発し、アプリケーションの実情に合わせて選択的に利用できるようにしている。

PDMはシステムやデータの構造が目に見える形で扱える簡易性を基本目標としたDBMSである。この利点は、システム資源やシステム開発工数の面で経済性に優れていることと、ユーザーのプログラミングの面で単純業務から統合業務への移行や、バッチ環境からオンライン環境への移行を容易にしていることにある。データベースはCODASYL仕様<sup>(6)</sup>に類似したネットワーク構造を採用している。

一方、ADMは、データベース、データ コミュニケーション及びデータ アドミネストレーションに関する統合的な機能を、オペレーティング システム(OS)の拡張として完全に論理的なレベルで提供することを目標としている。これは、MIS指向を含めた大規模な総合情報管理システムを構築するための強力な基本手段となるものである。

#### 3.2 PDMの機能概要

PDMは、効率的なストレージ管理とレコード間関連性の自動的な維持を基本として、データベースの創成・維持のための能力を提供し、またデータベース内のデータを操作するための高度な言語をもつ。

データ間関連性の高い業務環境においては、トランザクションはランダム パターンで発生し、多重質問要求によるデータ アクセスが増える。DBMSはこのようなアクティビティ環境に適合するように設計されたとき、最大の効果が得られるであろう。

PDMはマスタ及びバリアブルの二つのタイプのデータ セットをもつ。マスタ データ セットは安定した記述的データを蓄積し、ランダムマイジング手法によってダイレクトにアク

セスされる。バリアブル データ セットは受注量や在庫量のような変化しやすいデータを含み、関連するレコードを2方向のポインタで結んだ多重リスト編成をもつ。データベースは二つのタイプのデータ セットをネットワークに統合したもので、各データ セットをエントリとして扱い、また他のどのデータ セットへも関連づけることができる。

PDMは、データベースの構造を定義するためのデータベース定義言語と、ホスト プログラミング言語と結合してデータの蓄積、検索、追加、削除などの操作を指令するためのデータ操作言語とを基本成分として構成される。ホスト プログラミング言語にはCOBOL, PL/1, FORTRAN 及びアセンブラを使うことができる。

データベースの修正が必要なときは、変更部分だけについて再定義と再構成を行なえばよい。例えば、アプリケーション プログラムを変更したり再コンパイルすることなく、フィールド タイプやレコード タイプを追加あるいは削除したり、また新しいデータ セットをデータベースに追加したりして、既存のデータ セットにリンクさせることができる。

ハードウェア資源管理の面では、レコードのメイン メモリ内処理、入・出力バッファ プールの共用、ディスク ストレージの自動再編成などの技法によって高い性能を保持している。

オンライン データベース システムへの拡張に際しては、TMS(Transaction Management System)やTCS(Transaction Control System)などの日立製作所のコミュニケーション モニタと結合することができる。

PDMはHITAC 8000シリーズにおけるNDOS(New Disk Operating System), EDOS, EDOS-MSO(Extended Disk Operating System-Multistage Operation)及びHITAC MシリーズにおけるVOS(Virtual Storage Operating System)1, VOS 2, VOS 3の各OSのもとで稼動する。

#### 3.3 ADMの機能概要

ADMは、データベースとアプリケーション プログラム及び端末とアプリケーション プログラムの独立性を達成することを目標として、データベース機能、データ コミュニケーション機能及びデータ アドミネストレーション機能を、オペレーティング システムとアプリケーション プログラム間のインタフェースとして提供する。

はじめの二つの機能は、“CALL”ファンクションと呼ばれる共通的なインタフェースを介して、COBOL, PL/1, アセンブラなどのホスト プログラミング言語で書かれたアプリケーション プログラムによって利用される。データ アドミネストレーション機能はシステムの運用管理のために、システムやデータベースの構成、機密保護の維持、バックアップと回復などを行なう定義、維持及び制御機能を統合したものである。これらの機能を利用するために、データ アドミネストレータに対して、資源管理用のコマンドや定義・維持用の各種ユーティリティ群が用意されている(図2参照)。

データベース レコードはセグメントを構成要素とする多重レベルの階層構造をもち、更に階層トリー内、あるいは階層トリー間のセグメントを関係づけることによってネットワーク構造が導入される。ストレージ構造は、シーケンシャル及びダイレクトの二つのタイプの編成が基本となるが、それぞれにインデックスを付けることもでき、合計四つの異なる編成とアクセス法が提供される。また端末メッセージは、メッセージ セグメントを単位として扱われ、アプリケーション プログラムとのインタフェースもデータベース セグメントの場合と共通な形式をもつ。

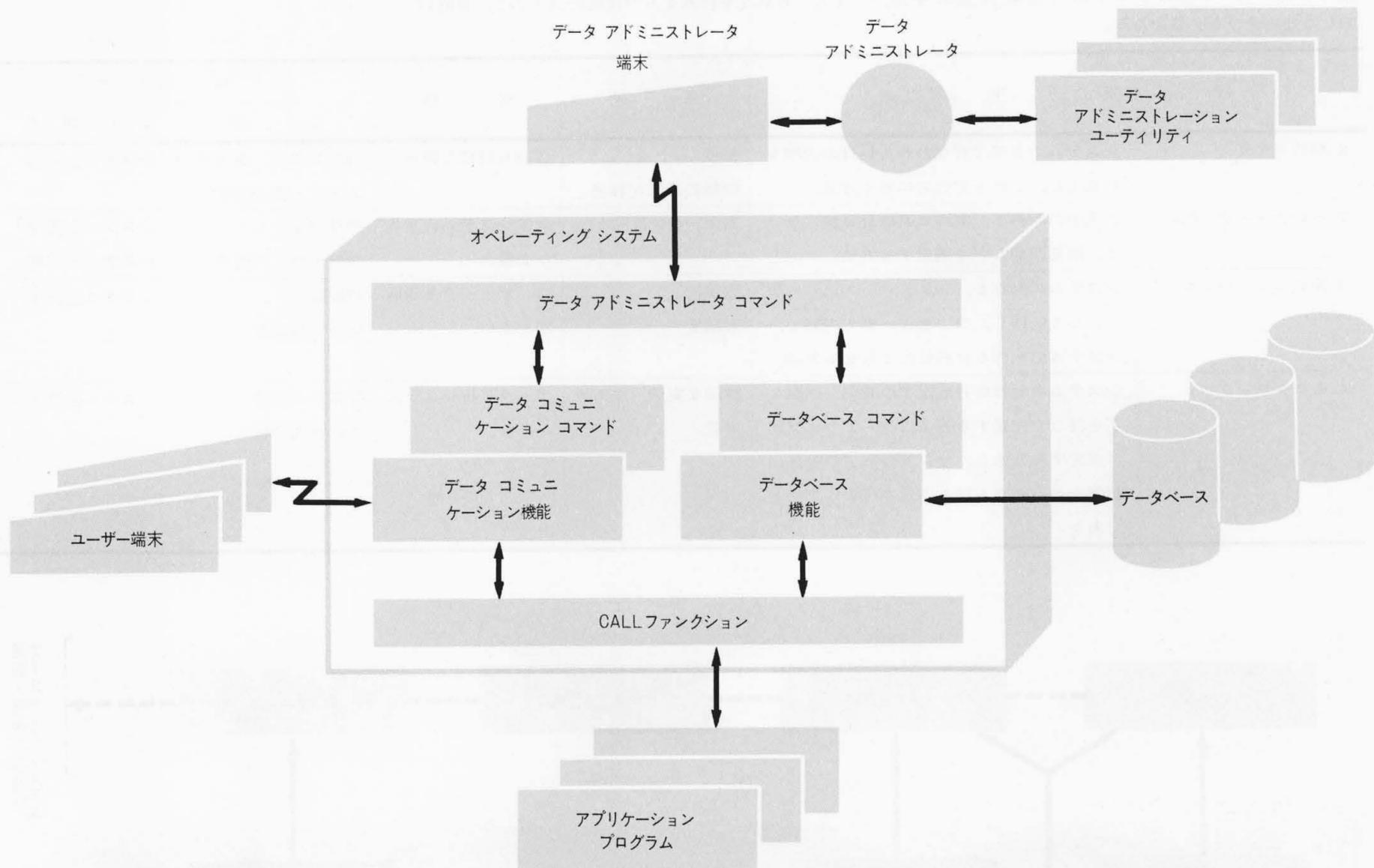


図2 ADMの機能構成 各機能成分はオペレーティング システムの延長上に位置づけられ、アプリケーション プログラムに対して共通的なインタフェースを提供する。

システムの利用環境にはバッチとオンラインがある。それぞれのシステムは、モジュール化されたADM成分の最適な組合せとして得られる。オンライン環境ではシステム破壊を防ぐために、ADMの制御モジュールとアプリケーションプログラムは、メモリ保護機構によって分離されたそれぞれ異なるパーティションにロードされる。

またシステム資源利用の平均化を図るために、データベース、端末、メモリ、タスクなどの割当ては動的にスケジュールされる。

ADMはHITAC 8000シリーズのEDOS-MSO及びHITAC MシリーズのVOS 2, VOS 3の各オペレーティング システムのもとで稼動する。

## 4 性能評価

### 4.1 性能評価の必要性と方法

データベース システムの性能にはシステムを構成する多様な要素が複雑に影響するため、性能測定、評価及び調整が特に重要である。データベース システムの性能の主要内容はスループットと応答時間である。前者は単位時間当たりの処理トランザクション件数で、オンライン、バッチ双方に共通するものであり、後者はオンライン システムにおいて、端末からメッセージを投入してからその応答が返されるまでの所要時間である。

性能評価の目的には次のものがある。

- (1) データベース システム計画時での性能検討
- (2) システム調整

(1)は、システム方式の適正化やハードウェア及びソフトウ

ェア構成要素の選定など、システムの方針を大きく左右する問題である。また(2)は、より良い性能を引き出すために、システム資源の配分やシステム パラメータを調整するものであり、システム詳細設計時及びシステム運用時に行なわれる。

一般に用いられる性能評価の手法には、表1に示す幾つかのものがある。それぞれ独特な長所はあるが、短所もあり、一つで万能とはいえない。実際上は、これらを目的に応じて相補的に併用すべきであろう。

### 4.2 性能評価システム

データベース システム実現の各段階において要求される性能評価に対して、表1に示した各手法を実施するためのソフトウェア及びハードウェアからなる性能評価システムを図3に示す。我々はPDM, ADMに即した性能評価システムの整備・開発を行なっているが、その一部を以下に例示する。

図4はADMのシミュレーション モデルASSESS (ADM System Simulation and Evaluation Support System) の構成概略図である。同図に示されるように、入力データとして業務プログラム モデルを必要とする理由は、業務プログラムの内容がシステムへの負荷の大きな要因であることによる。このASSESSの備えている業務プログラム モデル記述言語を用いると、COBOLなどによって業務プログラムを記述する場合の1/4以下の労力で記述できる。本モデルは、システム性能に影響を与えると考えられるほとんどあらゆる要因、例えば、マルチ タスク数の影響、データ構造の影響、データベース入・出力バッファ エリア サイズの影響、各種インデックスの主メモリ常駐化の効果などをシミュレーション結果に反映できる<sup>(5)</sup>。シミュレーション結果の一部を図5に示す。

表1 データベースシステムの性能評価の手法 それぞれ異なる特長をもつ性能評価手法を、目的に応じて使い分ける必要がある。

評価手法	長所	短所	主な用途	
			目的	評価時点
1. 解析モデル	簡便 概括的比較が容易。	詳細な評価に難点。	性能見積り、システム方式・構成選定	システム計画時
2. ベンチマークテスト	実データが得られる。	対象システムの実在が必要。	機種選定 ソフトウェア選定	システム計画時 システム設計時
3. シミュレーション	定量的 高精度	コンピュータ使用時間が長い。	性能確認 システム調整	システム設計時
4. モニタリング	詳細な実データ入手可能。	データ分析システムが必要。	システム調整 性能要因分析	システム運用時

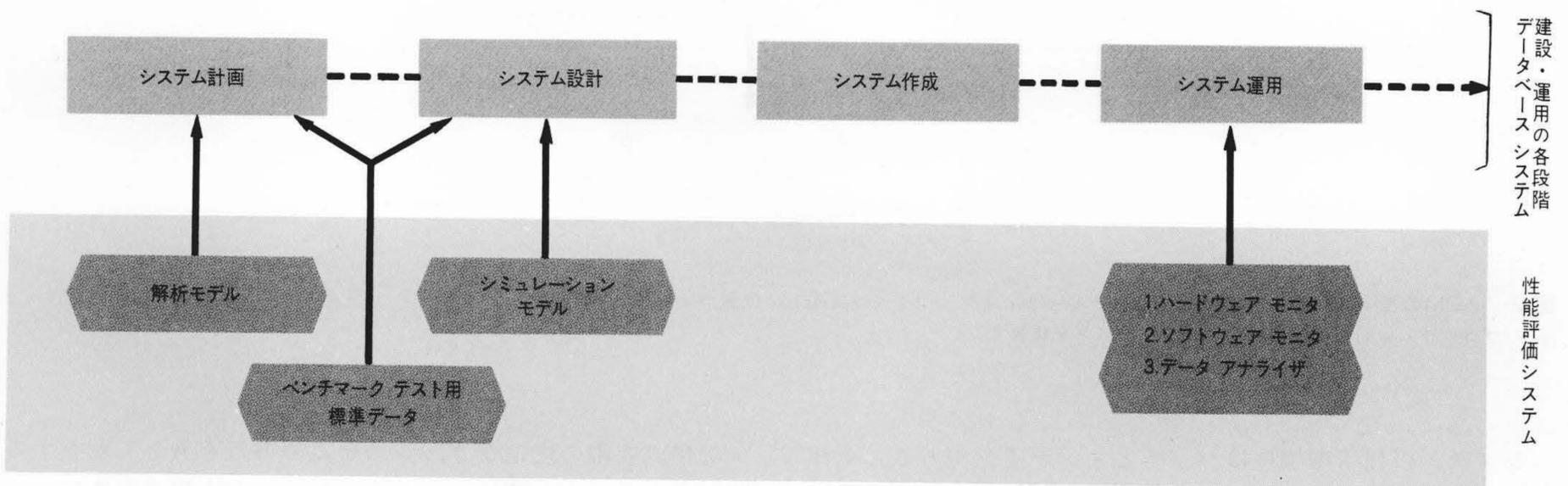


図3 性能評価システム システムの計画から運用までの各段階で、各種の性能評価ツールを相補的に利用することができる。

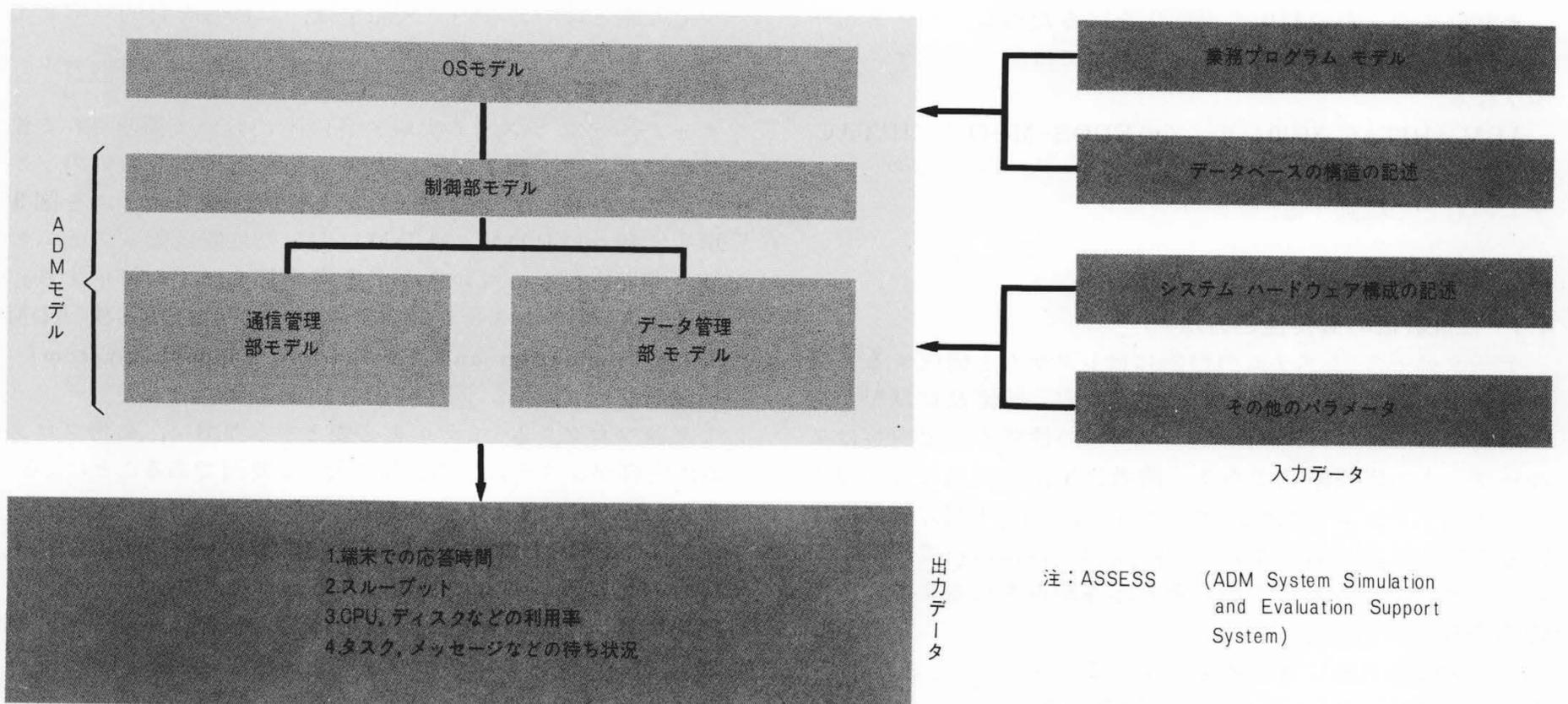


図4 ASSESSの構成と機能 業務プログラム、データベース構造、ハードウェア構成などの記述を入力すると、シミュレーション結果としてシステム性能に関するデータが出力される。

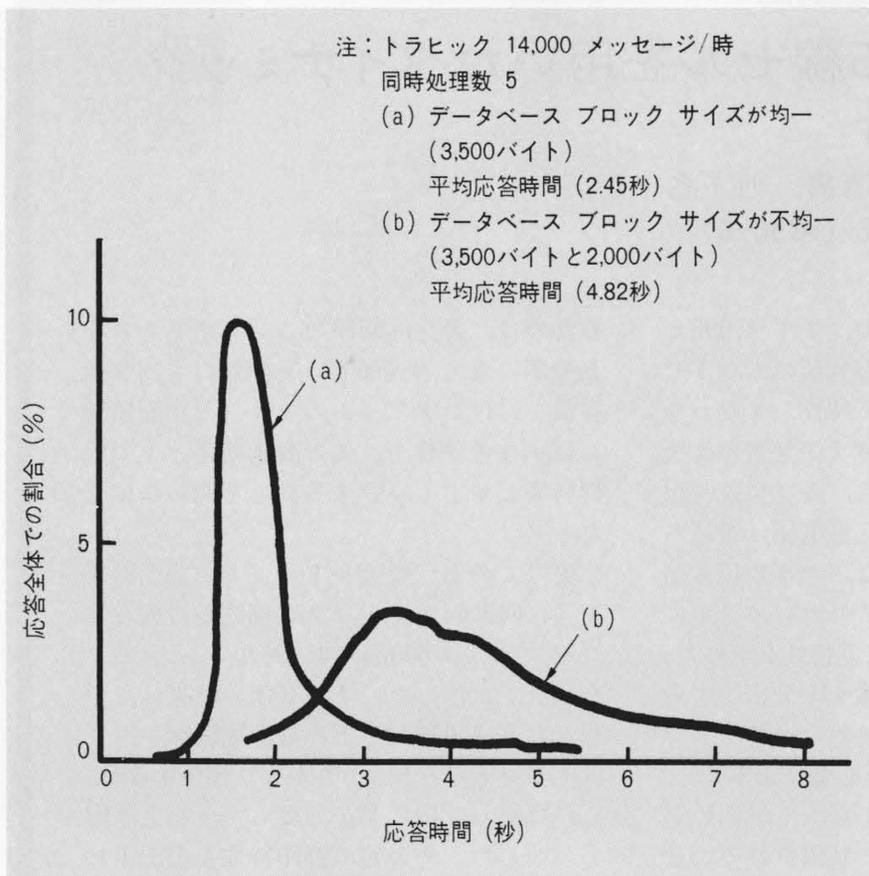


図5 ASSESSによる応答時間分布のシミュレーション結果  
データベース ブロック サイズが不均一な場合(b)は、均一な場合(a)に比べて、平均応答時間が2倍になっている。精度は90~95%である。

この図はデータベース構造に関する二つのシステム条件(a)、(b)においてシステムの応答時間がどのように分布するかを表わしたものである。ハードウェア モニタを用いて、実際のシステムの動作を測定し、これによって本モデルのシミュレーション結果の精度を調べたところ、相対誤差は5~10%程度であった。

次に、PDMのマスター データ セットへデータを追加する過程について、PDM解析モデルPEAL(PDM Performance Estimator by Analytical Method)によって分析を行なった一部を示す。PDMのマスター データ セットは、アドレスオーバーラップを起こしたデータに対して特設エリアをもたないランダムマイジング法を用い、着目データ ブロック中の未使用エリアを極力利用することによって入・出力の回数を少なくする方法をとっている。このようなファイル形態においては、データベースの収容率(収容可能データ量に対する既収容データ量の比)の関数として、PDMの処理量の統計的性質が表わされる。図6は、データを追加する際にPDMがデータベースからブロックを読み込む回数を表わしたものである。収容率の各種の値において、解析モデルと実測値との差が極少であることを示している。PEALを用いることにより、業務プログラムから複雑に組み合わされて出されるデータベース処理要求の処理時間の算定が可能である。

データベース システムの運用段階において、効率の良い運用を保持するために、多種の運用評価プログラムが必要とされる。図3に示したデータ アナライザとしては、モニタリングで得られるデータのアナライザのほかに、システムの各種状態のアナライザがあり、その例としてMDSS (Master Data Set Search), VDSS (Variable Data Set Search)がある。これらは、PDMのデータベース内の状況を出力する機能をもっている。一般にデータベースは使用を重ねると、物理的に削除・変更の痕跡が累積し、処理効率が低下するので、適当な時期にデータベースの再編成が必要となる。

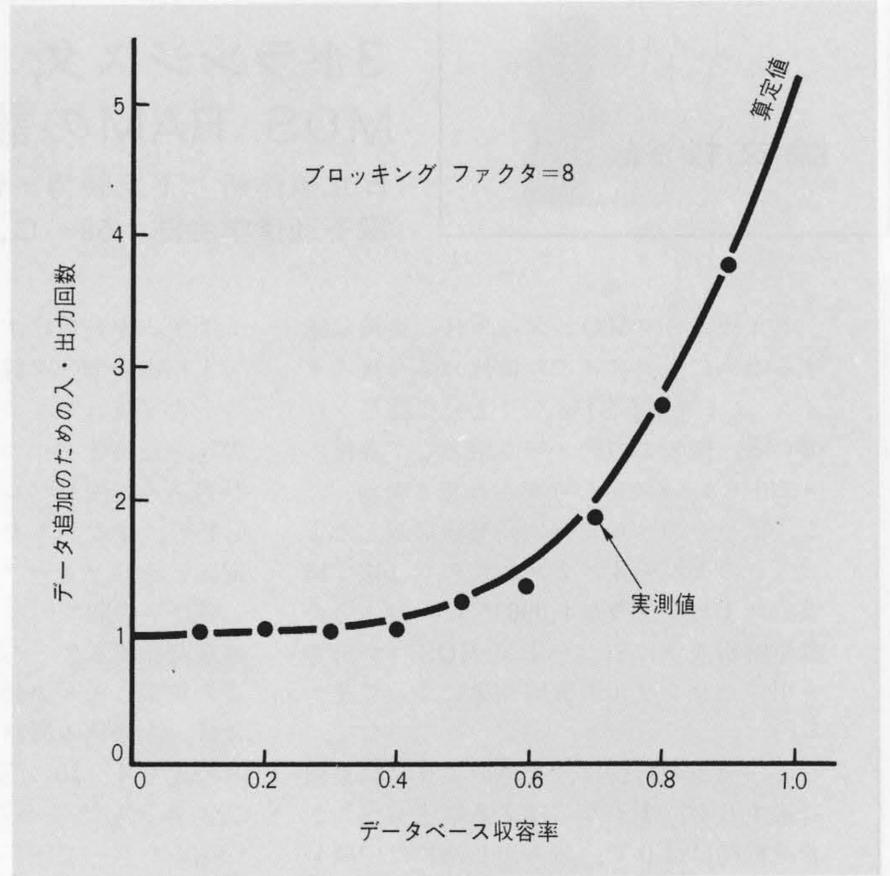


図6 PEALによる入・出力回数算定結果 データベース収容率 0.6以上では、入・出力回数が急増する傾向を示している。算定値の精度は約90%である。

MDSS, VDSSは、データベース内のデータ間のチェーンの長さ、異なるシリンダにまたがるチェーンの割合などの統計値並びに収容率及びシノニム発生率を短時間で出力し、再編成時期の判断のための資料を提供する。

## 5 結 言

我々はPDM及びADMの使用実績から得られたユーザーの要求に基づき、両システムの性能、障害処理能力及び操作性の向上に重点をおいて、その改善を続けている。性能評価の面では、シミュレータ、ソフトウェア モニタ、解析モデルなど、多様な評価ツールの併用による総合的な性能評価技法が蓄積されつつある。更に、PDMとADMを実際のシステム構築に適用するに当たって開発された、テスト及びプロダクション用サブシステム、データベース ドキュメント化ツールなどのサービス ファシリティが整備されてきており、これらを駆使した汎用的なシステム設計技法の開発も着実に進捗しつつある。

## 参考文献

- (1) R.L.Nolan, "Computer data bases: the future is now", Harvard Business Review 51, 5, 98 (1973)
- (2) R.G.Canning, "Problem areas in data management", EDP ANALYZER 12, 3 (1974)
- (3) 日立製作所, 「EDOS/EDOS-MSO データ・マネジメント・システムPDM (Practical Data Manager) 概説」(1973)
- (4) 日立製作所, 「EDOS-MSO データ・マネジメント・システムADM (Adaptable Data Manager) 概説」(1974)
- (5) F.Nakamura, I.Yoshida and H.Kondo, "A simulation model for data base system performance evaluation", AFIPS conference proceedings 44, 459 (1975)
- (6) Data Description Language Committee, CODASYL Data Description Language Journal of Development 1973, The Conference on Data Systems Languages (1973)

## 3トランジスタ, 2.5線セルを用いたダイナミック MOS RAMの設計

日立製作所 下東勝博・伊藤清男, 他1名  
電子通信学会誌 58-C, 327 (昭50-6)

ダイナミックMOSメモリは高集積に適するために、コアメモリに代わる今後のメインメモリ用LSIメモリとして注目されている。従来この種メモリ用として各種のメモリセルタイプが考案されてきたが、ここでは、そのうちで特に高集積に適した1ビットのセルを3トランジスタ, 2.5線で構成した1ビット及び4,096ビットメモリの動作解析並びにSiゲートN-MOSプロセスを用いたサンプルの実験結果について述べる。

このセルの特長は、上述のように高集積に適するが、(1)セルに加える電圧レベルが非選択時には0V, 読み出し時には中間レベルである3~5V, 書込み時には12Vと3値レベルをとらねばならぬこと、(2)書込み時に周辺回路を構成する書込み回路とセルがレシオとなるために消費電力の増大とセルへの書込み電圧が低いことなどの難点が予想される。そこで、これらの難点に伴

う問題点を定量化するため、まず4,096ビット相当のデータ線容量を負荷にもつ1ビットのメモリセルの書込み特性、すなわち書込み回路とセルのコンダクタンスの比及び書込み電圧レベルの関係、並びに読み出し特性、すなわちセルの記憶電圧(書込み電圧)をパラメータにした、セルの読み出し電圧(中間レベル)とアクセスタイムに直接関係するデータ線の放電特性を求めた。この結果、セルの動作領域を十分広くするには、(1)書込み回路とセルのコンダクタンスの比を4~10の間にとる必要があること、(2)読み出し特性から中間レベルが $2V_{th}$ ( $V_{th}$ :しきい電圧)程度に制限されるのでデータ線の放電時間が遅い。高速にするには高感度センスアンプの採用が不可欠である。(3)データ線の放電時間は中間レベルと $V_{th}$ の差の電圧に敏感なため、これらの電圧を厳しく制御する必要があることなどを明らかにした。これらの解析結果を確認す

るために、書込み回路のコンダクタンスを5種類に変え、4,096ビット相当のデータ線容量を付けた1ビットのセル(セル面積 $38 \times 42 \mu^2$ )を試作し、この評価結果と上記解析結果とがよく一致することを明らかにした。

更にこのようなセル1ビットの評価結果を基に $64 \times 64$ のマトリクス構成から成るアクセスタイム600nsの4,096ビットメモリを設計、試作した。その評価の結果は、予想どおり4,096ビットの電源電圧マージンは1ビットのそれよりもかなり縮小するが、これはチップ内の $V_{th}$ のばらつきが主原因と判明した。その他の動作特性も上記1ビットの特性とよく一致することが分かった。その他、ダイナミックメモリに特有な情報保持時間特性などについても述べ、このセルの実用化上の問題点を明確にした。

## 周波数領域での階層的パターン整合法による漢字認識

日立製作所 中野康明・安田道夫, 他1名  
電子通信学会論文誌 58-D, 2, 94 (昭50-2)

印刷漢字の自動認識のためのアルゴリズムとして、これまでに周辺分布のスペクトルの利用、階層的パターン整合法、対判定加重相関法の3手法が提案され、教育漢字を対象とする認識実験によってその有効性が確認されている。これら3手法は、それぞれ、(1)処理が単純である。(2)認識精度を落とさずに全体の情報処理量を削減できる。(3)最終判定が非常に精密にできる。という長所を有している。

本論文では、これら3手法の長所を総合した手法として、周波数領域での階層的パターン整合法(以下、HPMFと略す)を提案する。

HPMFは基本的には階層的パターン整合法である。すなわち、粗い解像のパターンを用いた整合により、対象カテゴリーを減少させ、この減少された対象カテゴリーについて密な解像のパターンを用いて精密な最終判定を行なうものであるが、これら

の処理を空間周波数領域で行なうことにより、各層ごとに異なる標準パターンを用意する必要がなく、またすべての処理を一次的なものに還元している。

文字パターンを、

$$x(i, j), \quad 0 \leq i, j \leq N-1$$

で表わす。その二次元離散的フーリエ変換は、

$$X(k, l) = \sum_i \sum_j x(i, j) W^{ik+jl}$$

で表わされる。 $X(k, l)$ の各成分をスペクトル成分と呼ぶ。

二次元パターンの処理でよく現われるたたみこみ演算は、周波数領域では成分ごとの積で表わされるので、一次的なプロセッサで処理できる利点を得られる。

二つのパターンの間の位置ずれを考慮した相関は、

$$y(s, t) = \sum_k \sum_l W^{-(sk+tl)} X_1(k, l) X_2^*(k, l)$$

で表わされる。ここで、添字1, 2はパタ

ーンの区別を、\*は共役複素数を、s, tは二つのパターン間の相対位置ずれ補正量を表わす。

分類、すなわち候補カテゴリーの選出においては、相関をとる周波数範囲を低周波領域に限定し、最終判定において周波数範囲を拡大することにより、階層的パターン整合が周波数領域で行なわれる。

本手法を計算機シミュレーションにより当用漢字1,850字を対象として4号明朝体タイプ活字の認識に適用した。対象サンプルをビジコンで撮像し、2値に量子化してからサンプリングピッチ0.1mmで標準化した。1文字カテゴリーについて各2サンプル、合計7,400サンプルについて認識実験を行ない、誤認識率0%、リジェクト率0.05%という結果を得た。

本手法により、漢字認識はアルゴリズムとしては既に実用上要求される水準に達したといえる。