

## データベース・システムの研究開発動向\*

西野 博 二\*\*

### 1. はじめに

データベース database という術語が、文献等によく見られるようになったのは、1960年代の半ばを過ぎてからである。同じ頃に、同義語もしくは、ややニュアンスの違った意味で、データバンク data bank という用語も使われていたが、最近ではほとんど使われなくなっている。

1960年代に入ってから、計算機が大規模な事務データ処理に使われるようになるにつれて、データ・ファイルの量が膨大となり、その取り扱いが大きな問題となってきた。まず、最初に起こった大きな変化は、ユーザがファイルの物理的な構造を正確に知らなくても、ファイルのデータが取り扱えるような種々のアクセス法が開発されたことである。

さらに、似かよったファイルを業務（応用プログラム）ごとに作成し、保守するのは、明らかに重複であり、無駄である。このようなファイル群を有機的に関連づけて、ファイルを共通化し、作成と保守の作業を集中化して、重複を避け、無駄を省くという切実な要望が起ってきた。

データベースとは、このような目的の下に、「種々の業務（またはユーザ）に利用できるように統合化された共用ファイルである」と定義される。

このような目的を達成するために、データベースに要請される必要条件には、次のものがある。

- (1) 各種の業務に適用可能であるためには、それ等の応用プログラムから独立であることが必要である。これをデータ独立 data independence といい、データの論理的構成からの独立、物理的配置からの独立等のレベルがある。
- (2) 個人のファイルと違って、共用のファイルであるから、正しいデータであることがより保証

され、より信頼性の高いことが必要である。これを、データ保全 data integrity という。

- (3) 多数のユーザが共用するためには、データベースの検索、修正、追加等の取り扱いが、共通であり、しかも便利なユーザ・インタフェースを持たねばならない。これは、データベース用語の問題であって、後述するように、データ記述言語 data description [definition] language (DDL), およびデータ操作言語 data manipulation language (DML) の二つがある。

- (4) また、多数のユーザが使用することによって、データベースが不当にそこなわれないような措置を必要とする。これはデータ保護 data protection, さらに広い意味で data security の問題である。

このような諸条件を満たすために、データベースを管理運営するためのシステムが、データベース管理システム data base management system (DBMS) である。最近ではさらに略して、データベース・システムともいわれる。

このように、データベースに関連する用語や技術は、事務データ処理の分野に起源を持つものである。一方従来から情報検索と称する分野でも、種々の研究開発が行われてきた。最近では、両者は異なった背景の下に生れてきたが、同じ目的を持つ全く同一の技術分野になったと理解されている。

### 2. 研究開発の歴史

データベース・システムの動向を、手っとり早く理解するために、ここではまず、図-1 (次頁参照) に示すように、研究開発の歴史を幾つかの流れに分けて、簡単に追ってみよう。

1960年代初期から、ファイル処理に関する種々の先駆的な研究開発があったが、最もよく知られており、かつ現在我々が持っている概念での最初の DBMS とされているのは、GE 社が 1963 年に開発した IDS

\* Research & Development Directions in Data Base Management Systems by Hiroji NISHINO (Electrotechnical Laboratory)

\*\* 電子技術総合研究所

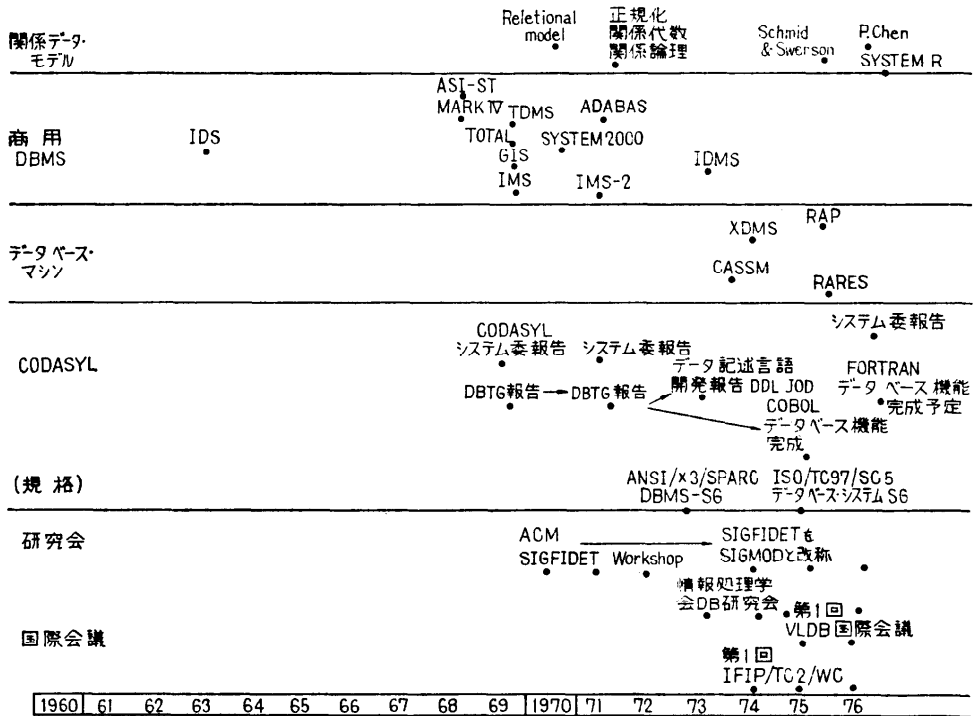


図-1 データベース・システムの研究開発の歴史

(Integrated Data Store) である。IDSは以後のDBMS開発に大きな影響を与えた。

1960年代の後期から1970年にかけて、多くの商用DBMSが開発された。図-1はそれ等の全部を網羅してはいないが、この時期のDBMSについては、1969年及び1971年に、CODASYLのシステム委員会が公刊した報告書<sup>1)</sup>に、詳細な分析結果が述べられている。この報告書に記載されたDBMSは、CODASYLが提案した仕様も含めて、GIS, Mark IV, NIPS/FFS, TDMS, UL/I, JOD COBOL, DBTG, IDS, IMS, SC-1等の10システムであった。

1970年代に入ってから、これ等の商用DBMSの普及が始まり、本格的な実用期に入ったとみられる。DBMSの販売統計では、1960年代の後期から、販売実績は年々指数関数的な増加を示しているといわれる。現在、商用DBMSとして、よく利用されているものは、TOTAL, IMS, IDS, IDMS, SYSTEM 2000, ADABAS, Mark IV等である。

ここで注目すべきことは、CODASYLのシステム委員会の報告書に述べられていたシステムで、現在でも生残しているのは、IMS, IDS, Mark IVの3シス

テムにすぎないことである。あるものは消え去り、あるものはその後の発展と拡張により、全くシステムとしての形を変えている。このことは、1970年の前後に、商用DBMSの急激な発展があったということ、またそれに伴い激的な自然淘汰の競争が行われたことを、物語っている。

一方、データベース・システムについて、CODASYLが果たした役割も極めて顕著である。前述のシステム委員会の報告と並行して、データベースのタスクグループ(DBTG)は、1969年、1971年とデータベース用言語の仕様を提案して、1973年には、データベースの記述言語 data definition language (DDL) の仕様、1975年にはCOBOLを親言語とするデータベース機能の言語仕様を完成させた。この間の経緯ならびに詳細な内容については、本特集号でも述べられるはずであるが、本学会のデータベース言語研究委員会の報告が、本年6月号の会誌に掲載されている。

COBOLは、実用言語として最も普及している言語であるから、COBOLのDBMSは、この分野での大きな勢力になろうとしている。前述した商用DBMSにも、DBTG提案の原型とみなされるIDSやDBTG

提案の影響を受けている TOTAL, IDMS 等があり、今後開発される商用 DBMS に、DBTG 提案が大きな影響を及ぼすことは確実であろう。このような情勢から、規格化の動きも活潑になりつつある。既に 1972 年秋には、ANSI/X3/SPARC で DBMS の研究グループが編成され、国際的にも ISO/TC 97/SC 5 に、研究グループが設立され、1975 年から活動を始めている。

今一つの大きな流れは、関係データモデル relational data model に関するものである。データモデルに関する提案は、後述するように種々のものがあるが、1970 年に E. F. Codd によって提案された関係データモデルは、数学的な関係の定義を理論的基礎としてモデルを展開したもので、多くの支持者を集めている。関係データモデルに関しては、数多くの理論的論文や、実験的な試みがあるが、最近 IBM 社も、関係データモデルによる本格的な DBMS として、System R を開発している。

ここ数年の極めて顕著な傾向は、DBMS を専用プロセッサで実現しようとする研究が、活潑になったことである。これには、DBMS を親計算機のソフトウェアから分離して、ミニコンピュータで実現しようとする単純な発想から出発した XDMS のようなものから、磁気ディスクのトラックごとにマイクロ・コンピュータを付設して、トラック上のデータを並列処理することによって、DBMS の画期的な性能向上を狙った RAP のようなものがある。また、両者を両極端として、その折衷や変形の提案や、開発も進行しており、このような問題向きマシンを総称して、データベース・マシンといている。

データベース・マシンの研究開発は、今後の DBMS の発展に大きな影響を及ぼすことが予想され、関心を集めている。また、これと関連して、主記憶に連想機能のような論理機能を付加しようとする連想メモリ、連想プロセッサ、知能メモリ等の古くからある発想を実現する試みも、依然として跡をたたない。IMSAI 社の商品名と混同されるが、磁気ディスクのデバイス・レベルで強力な論理機能を付加したものを、知能ディスクということもある。

### 3. データモデルと DBMS

#### 3.1 データモデル

データベース・システムを構築するに際して、まず問題になるのは、ユーザからみて、データベースに関

する現実の世界を、どのようにモデル化(形式化)するか、ということである。このようなデータモデルとしては、従来から様々のものが提案され、モデルとして認知されているものは、20 を超えるといわれる。

このうち、よく知られているものを列挙すれば、前述した DBTG モデル、関係データモデルの他、IMS 等のように階層的データ構造を土台とした階層構造モデル、Senko の提案した DIAM II モデル(1974 年)、Abrial の提案したデータ・セマンティックス(1974 年)、自然言語の意味解析に使われているセマンティックス・ネットワークを土台としたセマンティックス・ネットワーク・モデル(Roussopoulos 及び Mylopoulos, 1975 年)、関係データモデルの発展としての Entity-Relationship モデル(Chen, 1976 年)、拡張集合論モデル(Childs, 1974 年)等々があり、わが国でも小林氏の情報代数モデル(1975 年)の提案がある。

データモデルは、データに含まれる意味をユーザがどのように把えるかという、いわばユーザのデータ観ともみられるが、データモデルの論理的な実現手段であるデータ構造の性格を強く反映している。いささか異論もあるが、データ構造としては、(1)階層構造(または木構造)、(2)網構造(またはネットワーク構造)、(3)関係構造の三つが、代表的なものとみられる。

階層的データ構造を採用したものは、IMS が代表的であり、網構造を採用したものには、IDS, IDMS, DBTG 提案等がある。関係構造は、数学、特に集合論に基礎を置くものの代表として挙げたもので、Senko の binary association や Child の拡張集合論に基づくもの等も、この型に属する。

これ等の代表的なデータモデルや、対応するデータ構造については、本特集号で記述される予定であるが、モデルの段階で、これ等を比較し評価することは難しい。現実の DBMS には、これ等のモデルを組み合わせたものすらあり、DBMS の性能はその土台となるモデルに依存することもさることながら、それ以上にその実現方法の問題でもあるからである。

#### 3.2 DBMS の構成要素

DBMS のシステム構成要素としては、次の 4 要素を考えるのが普通である。(1)データ記述言語(DDL)、(2)データ操作言語(DML)、(3)アクセス法(AM)ならびにアクセス・ルーチン、(4)データベース管理者(DBA)。

かつて、応用プログラムごとに、固有のファイルが

作られていた時代には、ファイルの論理的構成とデータの記述は、COBOLのDATA DIVISIONにみられるように、応用プログラムの中に組み込んで、支障がなかった。しかし、データベースが共用ファイルとして、多数の応用プログラムから使用されるようになると、データベースの論理的構成と物理的配置は、応用プログラムとは別の所で、集中的に管理運営されねばならない。これを実行するのが、データベース管理者の役割である。

したがって、データ記述言語のひとつの側面は、データベース管理者に対するインタフェースを持っている。データ記述言語のもう一つの側面は、データベース全体で定義された論理的構成を、個々の応用プログラムで、固有の解釈をするためのインタフェースである。最近では、data directory/dictionary (DD/D) のような実現手法も用いられている。

データ操作言語は、実際のユーザに最も身近なインタフェースを提供するものである。これには大別して、システムとして自己完結型の独立方式のものと、親言語から使われるものの2種類がある。独立方式のデータ操作言語には、Mark IVのように、言語というよりも、ユーザはパラメータを設定するだけという簡単なものもあり、定型的な業務に普及している。しかし、ユーザがきめの細かいプログラムを作成するには、親言語方式のものが適している。最近、幾つかのよく使われている言語で、データベース機能を増強する動きが目立っており、FORTRANを親言語とするデータベース言語の仕様設定が進んでいるのも、その一例である。

データベースを使用するユーザには、応用プログラムを作るプログラマと、プログラムは作らないで、ただデータベースの内容を知りたいだけのユーザがいる。後者のようなユーザに対しては、より使い易いデータ操作言語を提供する必要がある。前述のMark IVのように、パラメータを設定すれば済むようなDBMSがよく使われているのも、このような要望に答えているためである。最近では、このようなユーザのために、できるだけ自然言語に近い問い合わせ言語を提供しようとする試みが多い。これは最近の人工知能研究の分野で、自然言語による質問応答システムが活潑に研究開発され、大きな成果をあげつつあることから、これ等の成果を実際に採用しようとする試みである。

データ記述言語や、データ操作言語は、DBMSの外側の部分で、ユーザとのインタフェースを提供するも

のであるが、DBMSの内部には、データの論理的構成を物理的配置に変換するためのアクセス・ルーチンがあり、このインタフェースを提供するものが、アクセス法である。物理的なデータ独立は、これ等のアクセス法を利用することによって、達成される。

DBMSには、その他、障害が発生した場合の回復の措置、データ保全やデータ保護のための有効な手段、データの冗長性や矛盾を減らす工夫等、様々の現実的な要請を満たすことも重要である。また、機能面の充実のみならず、システムの効率をいかに向上するかもまた、大きな課題であり、前述したデータベース・マシンの研究開発には、この面からの大きな期待が、かけられている。

1960年代から1970年代の初頭にかけて開発され実用化された各種の商用DBMSは、その後の新しいデータモデルの提案、新しいDBMS機能の設定と、その実現手段の研究開発等を背景にして、新しい世代のDBMSが今後、続々と登場して来る機運にある。

#### 4. 新たに研究開発を志す人のために

新しいデータベース・システムの研究開発は、いまや現在の情報処理技術の中心的な問題の一つになってきた。いささか蛇足であるが、これからデータベースの研究開発を志す人のために、参考文献や関連する会議、研究会等について図-1を参照して解説してみよう。

データベースに関する文献は、最近ではおびただしい数に上っているが、初学者のための入門書としては、J. Martin<sup>2)</sup>と、C. J. Date<sup>3)</sup>の著書が、よく知られている。J. Martinは、実時間処理システムの解説書でわが国でも広く知られた著者であり、今度の解説書も要領よくまとまっている。但し、最近の極めて新しい動向は収録されていないのが残念で、その意味では、C. J. Dateの著書の方が内容はやや新しい。また、本学会でデータベースの特集を企画したのと、期せずして一致して、ACMのComputing Surveyでも、今年の3月号に解説の特集<sup>4)</sup>を出している。本年3月に電子工業振興協会の主催で「データベース管理システムの研究動向」と題して、MITのM. Hammer教授の講演会が5日間行われ、その報告が同協会から出ている。これも最近のトピックスを判り易く解説している。

専門的な論文は、Comm. ACMやProc. NCC等、情報処理関係の学会誌や大会予稿等に散見されるが、今年からACMではデータベースの論文誌(Transac-

tion on Database Systems) を創刊した。また、1971年から ACM の SIGFIDET の研究会でデータベースの研究集会が行われており、この研究会は 1974 年に SIGMOD (Management of Data) と名称を変えたが、年次ごとの研究会の proceedings が出ている。ちなみに、SIGBPP (Business Data Processing) の季刊機関誌の表題もデータベースであるが、これは必ずしもデータベースのことばかり掲載されている訳ではない。

また、IFIP の TC2 では、1974 年にデータベース管理についての国際研究集会を開催し、これが契機となって、WG 2.6 というデータベースの研究会が設立された。以後、毎年研究集会を開いている。国際会議としては、VLDB (Very Large Data Base) の第 1 回会議が 1975 年に開催され、毎年開催されることが予定されている。

わが国では、1973 年に当学会にデータベース研究会が設立され、隔月に研究集会を開いている。なお、学会誌ではないが、Pergamon Press から出ている Information Systems という表題の季刊誌も、1975 年から創刊しているデータベースの国際専門誌である。

## 5. おわりに

データベースは、社会、経済、産業、研究等の各分野で、益々大規模で高度なものが要請されつつある。大規模の尺度として、磁気テープ 500 巻を 1 トンと呼ぼうとする提案もあり、このような大規模なデータベースを効率よく管理し、ユーザが使い易い DBMS の

開発の重要性は、強調しすぎるということはないであろう。

本特集号は、最近のデータベースに関する研究開発のほとんど全分野が、わが国では最も知しつしている専門家によって執筆され、研究開発の全貌を掴むことが出来ると同時に、わが国での実際的な適用例も併せて紹介している。この特集号を契機として、本誌の読者に、データベースに関する理解がより浸透し、わが国のこの分野の研究開発がさらに活潑になれば幸である。

本稿を草するに当たり、本特集号の執筆者の多くの方々（それは、データベース研究会の幹事の方々でもある）から、直接または間接に有益な示唆を受けた。特に、植村俊亮氏には、研究開発の歴史に関するデータの提供等、多大の助力を得た。撰筆するに当り感謝する次第である。

## 参考文献

- 1) CODASYL Systems Committee: Feature Analysis of Generalized Data Base Management Systems, ACM (1976-05).
- 2) J. Martin: Computer Data-Base Organization, Prentice-Hall (1975).
- 3) C. J. Date: An Introduction to Database Systems, Addison-Wesley (1975).
- 4) ACM Computing Survey, Special Issue: Data-Base Management Systems, Vol. 8, No. 1 (1976-03).

(昭和 51 年 8 月 9 日受付)