

データ・ベース管理システム最近の動向

株式会社ニエック総合研究所 小林功哉

個々に開発された単一のプログラムを、総合的な大規模プログラムに統合する必要性や、人間機械率としての情報システムのニーズが強く意識されるようになったり、これらのシステム開発の土台としてのデータ・ベース・マネジメント・システムの重要性が提唱されるようになったのはおぼろげに旧聞とでは無い。先行的な実験システムとして、1962年よりMETRE社の手によるADAMの開発が挙げられるが、ソフトウェア商品としてユーザーに提供されたのは1963年にIBM709の在来レポート・ジェネレータQPACの一般化としてGEの作ったIDSの第一版である。しかし、このころに今日のデータ・ベース・マネジメント・システムとしての意味が一般に感じとられるわけでは無い。同じころ、UNIVACにIM-RADSの開発計画があり、その仕様書には多くの優れたアイデアが盛り込まれていたが、ほとんど注目を集めなかった。

1967年、ASIS (American Society of Information Science) の前身であるADI (American Documentation Institute) の年報 Annual Review of Information Science and Technology に Maryland 大学の J. Minker, Anerbach, J. Sable の両氏により、File Organization and Data Management の項が掲載され、当時開発済み、あるいは開発中のデータ・ベース・マネジメント・システムのレビューがなされた。この種のソフトウェアを纏めて論じたものとしてはほほいめこのものがあつたと思う。1968年から1969年にかけて、IDS, IMS, GIS, TDMS, MARK II などシステムがほぼ今日の姿で発表され、才在 COOASYL システム発表会の冲上版レポート A Survey of the Generalized Data Base Management Systems も発表された。このころからデータ・ベース・マネジメント・システムの名も定着したようである。もうやく草刈りとなつた MIS 論議と関連して、データ・ベース・マネジメント・システムの必要性が強く論じられ、オペレーティング・システムや言語プロセサと並んで、計算機に不可欠のソフトウェアとされるようになったのは此以降である。

今日の汎用データ・ベース・マネジメント・システムの開発には二つの異なる動向があつたように思われる。その第一のものはむしろ単なるファイル統合であつた。過去に開発された単一の個別プログラムを単一の業務処理の在来の一連のプログラムを、相互の関連を考慮して一貫したシステムとし、大規模なソフトウェア・パッケージに改造することが盛んに行われてきたようになった。とくにハードウェアとして純粋のファイル媒体の復達がこの傾向に拍車をかけた。同じく作られた一貫システムの中で、データの重複や更新時点の相違によるデータ間の矛盾を避けるため、個別プログラムで扱われていたファイルを統合したものがデータ・ベースと名づけられたのである。ファイルの統合に伴って、単一プログラムの中のファイル処理を行なう処理としたものがあつた。複雑なファイル制御の処理もあつたが、ほとんどはオペレーティング・システムのデータ・マネジメント機能はこれらの要求を満たすに貧弱であつた。この在来オペレーティング・システムの機能の拡大を計つたものが汎用データ・ベース・マネジメント・システム

開発の第一の動機があり、今日のいわゆるホスト言語システムの発展を促した。ホスト言語システムでは複雑なファイル制御のサブルーチンを整備し、これを既存の言語プロセサに移植することによって、統合ファイルを持つ大規模システムを比較的容易に作れる。統合システム作成のプログラムマの工数を大いに減らすことが可能になるわけである。

一方、経営情報システム、技術情報システムなど、いわゆる情報システムといふ言葉の中には、第一の動機にない意味合いがあるかと思われる。実際これらの情報システムは同じ機械系ではなく、人間と機械の間の類型的インタラクティブを前提とする一統の人間機械系の構築を目指しているかと思われる。人工知能に対する挑戦が、形式論と意味論の互いの壁などによってなされる困難に直面し、ロマン認識、学習課程などの領域でフロンティアの試行の階層が実際に認識されてくる今日、同じ機械系に対する問題解決の階層を越えざるを得ない。計算機のアルゴリズムの機械性と、これに對比される人間のヒューリスティックな機械性を協調させ、設計作業など高度の知的活動と必要とする分野の問題解決を計るのがCADの基本理念であり、これがMIT流のサイエニフィックの目的とするところであろう。経営情報システムなどもこの範疇で考えられるべきかと思われる。人間機械系としての情報システムを考へるとき、その機械性をインフォーマション・マネジメント・システムと告げよう。

インフォーマション・マネジメント・システムの持つべき機能は、人間の必要とする要求にも応じて即時に必要情報を与えることにあり、そのための情報の蓄積、制御された記憶のアルゴリズムの蓄積、そして優れたマン・マシン・インタフェイスと必要とする。とくに、いかなる形の情報をも蓄積し、いかなる標準條件に於いても情報の検索が可能であり、容易な更新手段を備えたデータ・ベース・マネジメント機能が必要である。これが、汎用データ・ベース・マネジメント・システム開発の第一の動機と告げよう。このように要請に充ちるものとして過去にレポート・プログラム、シエネシーがあった。いわゆるセルフ・コンテンツ・システムはRPGの延長線上に考へられる。

これらの要請に充ちるため、各計算機メーカーやソフトウェア・ハウスが競ってデータ・ベース・マネジメント・システムを開発した。国内でも日本電気のCESS、富士通のRAPID、日立のSELDAM など多くのシステムが作られた。残念ながら、これらのシステムの間には、過去の経験をもとにした便宜的考慮のみを以て作られたものがあり、多少なりとも理論的考慮を払ったものは少ない。そのため、その機能は一長一短で汎用性に乏しい。ホスト言語システムについては、その性格がいろいろのものが実用に供され、CODASYL Data Base-Task Group によるCOBOLの拡張の提案の基に標準化の動きがある。機能面で今は納得し難い点もある。とくにデータの論理的構造と物理的構造の分離が中途半端であるなどの不満もあるが、汎用性、標準性の二面で現在これに優るシステムはない。CODASYL提案はUNIVACの年によりDMS1100としてソフトウェア化されているが、COBOL同様いずれ他のメーカーも追随することになる。しかしながら、セルフ・コンテンツ・システムの多くは未だ実験段階にとどまり、応用範囲もきわめて狭い領域に限られてしまっている。

データ・ベース・マネジメント・システムの開発の動機は相違、開発過程での重点の置き方の差などから、初期のシステムの性能は千差万別で、性能の比較などの議論は必ず多くの混乱を生じた。前述したCODASYLシステム委員会レポートのレポートは、まず一般的なシステム性能を抽出し、そのリストに従っていくつかのシステムの性能を記述したもので、その意味では画期的なものであった。しかし、現時点ではホスト言語システムとセルフ・コンテインド・システムの区別も明確になさず、その他の点でも多くの曖昧なところが残っていた。CODASYLシステム委員会はその後この仕事を続け、1971年には改訂版とて「*Feature Analysis of Generalized Data Base Management Systems*」が発表された。このレポートではホスト言語システム、セルフ・コンテインド・システムの区別も明確なものとなつたが、記述の基礎となるデータ構造やデータ処理のモデルは、あまり洗練されたものと考えられる。

データ・ベースや、データ・ベースの上での情報処理についてより数学的表現を考へ、システムの定性的評価により客観的な基礎を身之ると同時に、新しいシステム開発時に合理的な設計指針を与えることを考へたのが筆者の仕事で、1962年CODASYLの言語構造グループが提起したInformation Algebraを拡張し、集合論的モデルを構成した。結果はData Base Management Systems - A Theory and a Practiceに収められている。筆者のモデルに沿って開発された日本ユニバーク総合研究所のFORIMSはFORTRANをホスト言語とするホスト言語性能と、一般ユーザのための1/4型、リモート型のシステム言語を採り、複雑なデータ構造を容易に扱えるなど、特異な位置を占める。

上記モデルでは属性値、記録、ファイル、あるいはアーク属性値、アーク記録、構造、そしてデータ・ベースなどの概念を、計算機とは無関係に論理的に定義し、さらに計算機記憶装置上での物理的表現を考へる。物理的表現は結局連合の表現方法を考へることに帰着し、いくつかの連合表現法をどう組合せるかの問題となる。一方情報処理は記録関数の計算、記録の集合としてのファイルの交換と集合演算に分解され、非手続的に記述された処理を手順処理に翻訳する問題となる。

このモデルに対して、中間的な記録やアーク、ファイル、構造などの概念の導入を避け、値とその間の連合だけを扱ってデータ構造のモデルを組立てたのがIBM San Jose 研究所のE.F. Coddのモデルであり、上記モデルとは異つたデータ処理要素を扱って情報処理を組立てている。このモデルはRelational Data Base と呼ばれ、主としてACMのSIGDの誌とつてあるSIGFIDETを中心にとりあげられ、1972年のCOMSでも多くの論文が発表された。このモデルはより人間の頭脳のデータ処理方式に似ていると考えられる。QAシステムなど柔軟なデータ表現やデータ処理能力の要求されるものにはこのプロローグが本筋があるように思われる。このモデルでは情報と情報処理の論理的構造に重点が置かれ、物理的表現は向後の問題である。いくつかのimplementationの試みもあるが、十分に成功したとは言えない。モデル自身も多少洗練の余地があるように思われる。

いずれにせよ適当なモデルを土台に、概念と用語の整理を行うことが、データ、ベース、マネジメント、システムの技術の共通の理論と合理的な評価のためには必要と思われる。また論理的な構造と物理的表現を分け立りとは別して論ずることも重要である。前者はシステムの定性的評価のためには必要である。前者はあくまでも計算機のハードウェアに依存するものであり、計算機に付随した技法があり、遂に計算機の発達に伴って変革を受けるものである。逆に、論理的構造を物理的表現に直す問題についての定量的な要素が、新しいハードウェア技術の発達を促す可能性もある。

COEDASYL System Committee - a Chairman である E.H. Sibley の居る Michigan 大学の Industrial and Operational Engineering には R. Teichroew を中心に ISDOS (Information System Design and Optimization System) プロジェクトが進められている。情報システムの開発の手順を標準化しようという大掛かりな構想で、データ、ベース、マネジメントもその一環として論じられている。データ、ベース間の変換、data translation にも手を付けられている。一連の意味あるリポートが表されたり、将来の期待されるプロジェクトがある。