

リレーショナル データベース入門

—データモデル・SQL・管理システム—

増永良文 著

サイエンス社

INFORMATION
& COMPUTING

007.6
R47

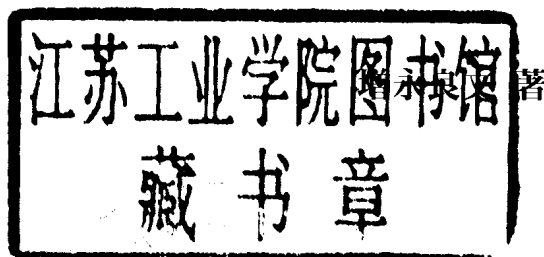
230457



Information & Computing-43
日文 701638122

リレーショナル データベース入門

—データモデル・SQL・管理システム—



サイエンス社

著者略歴

増 永 良 文
ます なが よし ふみ

1970年 東北大学大学院工学研究科博士課程
電気及通信工学専攻修了

現 在 図書館情報大学図書館情報学部教授
工学博士

Information & Computing=43

リレーショナルデータベース入門

ーデータモデル・SQL・管理システムー

1991年 1月25日 © 初 版 発 行

1996年 4月25日 初版第10刷発行

著 者 増永良文 発行者 森平勇三

印刷者 篠倉正信

製本者 宮下菊男

発行所 株式会社 サイエンス社

〒151 東京都渋谷区千駄ヶ谷1丁目3番25号

〔営業〕 ☎(03)5474-8500代 振替00170-7-2387

〔編集〕 ☎(03)5474-8600代

FAX ☎(03)5474-8900

印刷 株式会社ディグ 製本 徳住製本所

〈検印省略〉

本書の内容を無断で複写複製することは、著作者および出版社の権利を侵害することがありますので、その場合にはあらかじめ小社まで許諾をお求め下さい。

ISBN 4-7819-0594-3

PRINTED IN JAPAN

サイエンス社のホームページのご案内
<http://www.bekkoame.or.jp/~saiensu>
ご意見・ご要望は
saiensu@lib.bekkoame.or.jp まで。

まえがき

古人は「世界は大地，空気，水そして火の四つの要素からなる」と考えた。コンピュータはハードウェア，ソフトウェア，ヒューマンウェアそしてデータベースからなる。人が他の動物と大きく異なる点の一つは火を自由に使える点にあった。それにより食物を調理し，寒さをしのぎ，様々な物を加工することができ，文明を生み出した。

社会は情報化 (computerization) の一途をたどっている。それは人の好むと好まざる，望むと望まざるにかかわらず進行している。そのような中で，人は単に生き残るのではなく，さらなる豊かさを求めている。それを可能とする切札は，ハードウェアやソフトウェアではなく，データベースをいかに使いこなすことができるかにかかっているのではなからうか。

これまでのコンピュータ世界はアルゴリズム中心であった。コンピュータは大量，複雑な計算を瞬時にこなす道具として重宝がられた。しかし，現代は人間的確な判断を下せるための情報のもととなる大量で複雑なデータ群をうまく組織化し，効率良く管理しておくことをコンピュータに求めている。まさしく，世の中はデータ中心の世界に移行したということであろう。

誤解があっては困るが，データを重視する視点はコンピュータの歴史と共にあった。たとえば IBM システム 360 ではデータ管理のもとで，データはファイルとして管理された。また世界で初めてのデータベース管理システムは 1963 年に General Electric 社から IDS という商品として発表されている。しかしデータベースが空気や水でなく，「火」としての役割を担えることができるようになったのは，1970 年の IBM サンホセ研究所のコッド (E. F. Codd) による「リレーショナルデータベース」の提案によると筆者は考える。リレーショナルデータベースが実用に供されるようになったのは 1980 年代に入ってからだが，それが有する平明性，高いデータ独立性，非手続的データベース言語等の特長は，ソフトウェアの生産性を著しく向上させ，情報資源管理に道をひらき，誰もが

データベースをアクセスできる環境を提供することとなった。

ご縁があってサイエンス社刊の隔月発行雑誌“Computer Today”に、1987年9月号から1989年11月号にかけて、計14回「リレーショナルデータベース入門」を連載させていただいた。このたびそれをまとめて、単行本として出版して下さるという。連載では第1回目をデータベースの全般的入門にあて、第2回から8回までをリレーショナルデータモデルの解説に、第9回から最終回（第14回）までをリレーショナルデータベース管理システムの解説にあてた。本書ではそれぞれに対応して、第1章から第14章の構成になっている。また、同じく著者がComputer Today 1985年3月号に掲載した解説「分散型データベースシステム」に若干手を加えて、それを第15章とした。分散型リレーショナルデータベース管理システムを垣間見ることができればと思う。

単行本とすべく、今、連載を一括してながめてみると、冗長であったり、書き足らなかつたり、意を尽くしていないところも多い。これはひとえに著者の力不足の現われである。データベースの重要性に免じてお許しを乞いたい。

末筆ながら、これまで著者をご指導、ご鞭達下さった様々の方々、そしてComputer Today 誌編集長富澤昇氏とスタッフの皆様に、心から感謝の意を表したい。

平成2年11月

著者しるす

目 次

まえがき	i
1 データベースとは何か	
——リレーショナルデータベースの誕生——	
1.1 本章のはじめに	1
1.2 データベースとは——料理と器の譬え——	3
1.3 データと情報	3
1.4 データモデル——実世界記述の記号系——	8
1.5 リレーショナルデータベースの意義	12
1.6 本章のおわりに	13
2 リレーショナルデータモデル	
2.1 本章のはじめに	16
2.2 徹底的にフォーマルなリレーショナルデータモデル	17
2.3 リレーション	18
2.4 属性名とリレーション名	20
2.5 リレーションスキーマ	21
2.6 主 キー	22
2.7 外部キー	25
2.8 一貫性制約	27
2.9 第一正規形	28
2.10 リレーショナルデータベーススキーマ	31
2.11 本章のおわりに	32

3	データ操作言語とリレーショナル代数(1)	
	——集合演算——	
3.1	本章のはじめに	33
3.2	リレーショナルデータ操作言語の基本概念	34
3.3	リレーショナル代数とリレーショナル論理	36
3.4	リレーショナル代数	38
3.4.1	リレーショナル代数の枠組	38
3.4.2	リレーショナル代数の諸演算	39
3.5	本章のおわりに	44
4	データ操作言語とリレーショナル代数(2)	
	——特有な演算——	
4.1	本章のはじめに	45
4.2	リレーショナル代数の諸演算	46
	——四つのリレーショナル代数に特有の演算——	
4.3	リレーショナル代数表現	56
4.4	本章のおわりに	58
5	リレーショナルデータベース設計の基礎理論(1)	
	——リレーションの情報無損失分解——	
5.1	本章のはじめに	59
5.2	第一正規形リレーションでの更新時異状	60
5.3	更新時異状の解消	62
5.4	リレーションの情報無損失分解	64
5.5	多値従属性	67
5.6	本章のおわりに	70

6 リレーショナルデータベース設計の基礎理論(2)

——リレーションの正規化理論——

6.1	本章のはじめに	72
6.2	関数従属性	73
6.3	関数従属性の公理系	75
6.4	第二正規形	78
6.5	第三正規形	79
6.6	ボイス-コード正規形	84
6.7	第四正規形	87
6.8	本章のおわりに	89

7 データベース言語 SQL(1)

——基本機能——

7.1	本章のはじめに	90
7.2	リレーショナルデータベース言語の標準化	91
7.3	SQL による質問	92
7.4	単純質問	94
7.5	本章のおわりに	98

8 データベース言語 SQL(2)

——高度な質問指定機能——

8.1	本章のはじめに	99
8.2	SQL の質問指定の基本型	99
8.3	結合質問	100
8.4	入れ子型質問	103
8.5	SQL のリレーショナル完備性	105
8.6	更新操作	106
8.7	PLI/SQL	107
8.8	WHAT vs. HOW	109
8.9	本章のおわりに	111

9	データベース管理システムの基本アーキテクチャ ——三層スキーマ構造——	
9.1	本章のはじめに	112
9.2	ANSI/X3/SPARC の DBMS の三層スキーマ構造	113
9.3	三層スキーマの意義	117
9.4	ビ ュ ー	121
9.5	本章のおわりに	124
10	リレーショナル DBMS の事例研究 ——INGRES——	
10.1	本章のはじめに	125
10.2	INGRES の開発経緯	126
10.3	INGRES と UNIX	128
10.4	INGRES のシステムアーキテクチャ	129
10.5	INGRES のプロセス構造	132
10.6	DBMS の三大機能	133
10.7	リレーショナル DBMS vs. ネットワーク DBMS	135
10.8	本章のおわりに	136
11	リレーショナル DBMS の質問処理(1) ——質問処理とは何か——	
11.1	本章のはじめに	137
11.2	リレーションの実装	138
11.3	質問処理——その基本スキーム——	143
11.4	結合質問の処理	145
11.5	本章のおわりに	149

12	リレーショナル DBMS の質問処理 (2)	
	——質問処理の最適化——	
12.1	本章のはじめに	150
12.2	質問の処理コスト	150
12.3	質問処理の最適化	155
12.4	INGRES の質問処理系	158
12.5	本章のおわりに	166
13	リレーショナル DBMS のトランザクション管理 (1)	
	——障害時回復——	
13.1	本章のはじめに	167
13.2	トランザクション	167
13.3	障害時回復	170
13.4	本章のおわりに	177
14	リレーショナル DBMS のトランザクション管理 (2)	
	——同時実行制御——	
14.1	本章のはじめに	178
14.2	同時実行制御の必要性	178
14.3	スケジュールの直列化可能性	182
14.4	ロッキング	186
14.5	本章のおわりに	192
15	分散型データベースシステム	
15.1	本章のはじめに	193
15.2	リレーショナルアプローチの適合性	194
15.3	分散型データベースシステムとその研究・開発	195
15.4	System R*	197
15.5	本章のおわりに	200

あ と が き	201
参 考 文 献	202
索 引	208

1 データベースとは何か

——リレーショナルデータベースの誕生——

1.1 本章のはじめに

データベース (database) という術語も最近は広く社会に受け入れられてきたようである。流行語で終ることはないだろう。これはデータベースがハードウェア、ソフトウェア、ヒューマンウェアと並んでコンピュータ、あるいは情報処理システムの中心要素として広く認識されてきたという証拠であろう。コンピュータは単に計算を人と比べて桁違いに高速にこなせる機械から、実世界のデータの収集、加工、蓄積、提供、利用に係わる全てのフェーズを自動化し、効率の良いその管理・運用を可能とする切札としての機械へと変身しているのである。情報立国を国是としなければならない日本にとって、そのような意味で、データベースとそれに係わる技術の整備と振興は現在一番大事な事柄の一つであろう。数多くのデータベースを外国に依存している現状は大変危険であり、一方データベースの構築・管理・運用技術が、たとえば自動車や半導体製造技術のように、世界をリードできる技術に育っているかというところではない。まずは、広く人々が、“データベースとは何か”を理解することが改めて必要ではないかと考えられる。

さてデータベースに係わる諸事象を1960年代初頭から、通時的に追ってみてみると、一つの大きな潮流のあることに気付く。それは1963年に米国のGE社（ジェネラルエレクトリック）で、バックマン（C. W. Bachman）のもとで開発されたデータベース管理システムIDS（Integrated Data Store）に始まり、1970年に米国のIBM社のコッド（E. F. Codd）の提案したデータのリレーショ

ナルモデルに基づく、現在のリレーショナルデータベースへいたる流れである。

別の観点からは、**ビジネスデータ処理**用データベースに始まり今日の**非ビジネスデータ処理**用データベースの構築と管理・運用への流れである。ここにビジネスとは庶務、財務、商務、工務等のいわゆる事務一般であり、ビジネスデータとして具体的には従業員データ、顧客データ、在庫データ、売掛・買掛データ等がある。一方非ビジネスデータとはオフィスオートメーションやファクトリオートメーションに必要なデータで、具体的には文書(ドキュメント)、設計図や地図等の図面、X線写真やランドサット(LANDSAT, 人工衛星)イメージのような画像、あるいは音(音声、音響)等のデータである。

なお、ビジネスデータ処理用データベースから非ビジネスデータ処理用データベースへの展開は、ビジネス領域で確かめられたリレーショナルデータベースの幅広い適用性を非ビジネス領域に敷衍しようとするものであることに注意したい。確かにビジネスデータ処理は **CODASYL** (The CO^ンference on DA^タ SY^スtems Languages, データシステムズ言語協議会)のDBTG (Data Base Task Group, データベース作業班)が1960年代の後半に入って作業を開始し、1971年春に報告したDBTG **ネットワーク (network) データモデル**の果たしてきた役割抜きに語ることはできないことは十分承知している。しかし、データベースの流れは確実にリレーショナルに向っている。これはリレーショナルデータベースが

- モデルの平明性
- 高度な論理的・物理的データ独立性の達成
- データ操作の非手続性
- 分散型データベースへの適合性

等ほとんど全ての点で優っているからである。最近 IBM 社が自社の**ハイアラキカル (hierarchical, 階層) データモデル**のデータベース管理システムIMSを特定応用向け専用とし、DB2というリレーショナルデータベース管理システムを汎用として位置付ける戦略に出たことはむべなるかなである。

つまり、我々がデータベースを理解しようとするとき、リレーショナルアプローチを中心にすえてそれを行なうのが的確である。そこで本書では、リレーショナルデータベースの誕生、モデル論、管理・運用技術、利用技術、そして、

その展望について、解説を行ないたい。

1.2 データベースとは——料理と器の譬え——

人々は様々な意味をこめてデータベースという術語を使っているのではなかろうか。ある人は情報検索システムを、別の人は勤務先の従業員とか顧客ファイル群を、はたまた別の人は DB2 とか UNIFY とかの製品を。

さてデータベースという術語の原義はデータの基地とでもいったようなことだったのだろうけれど、現在は多義語のようで、基本的に二つの意味を持って使われているようである。使い手が明確にその違いを意識して使っていれば問題はないのだが、あいまいなまま使用しているのではないかと思われる場合もあるので、一言述べる。

この二つの関係は料理とそれを盛る器の関係に似ている。この譬えでは、データベースの原義は料理にあたる。一方、近頃これもデータベースといっているような器のほうは、専門家は**データベース管理システム** (Database Management System, **DBMS** と略記) といって厳格に区別してきたものである。派生して、料理を器と一緒にしてデータベースといっている場合も少なくないようである。これは第三の意味になろう。図 1.1 (次ページ) に管理されるデータベースとそれを管理するデータベース管理システム、そしてデータベースを利用する者の関係を示す。なお、ユーザには大別して二種類のユーザ——一つはオンラインでデータベースを検索・更新しようとする**エンドユーザ** (end user), もう一つはバッチモードで行なおうとする**応用プログラム** (application program) 群——がある。

1.3 データと情報

データベースのことを語る上で、データとは何か、はっきりさせておく必要があると思う。とくに日本では高度情報化社会といった、よく考えれば一体何をもってその定義とするのかよく分からないままにでもそれを好んで使うという国民性なのか、データと情報はほとんど明確な概念区分なしに使われている

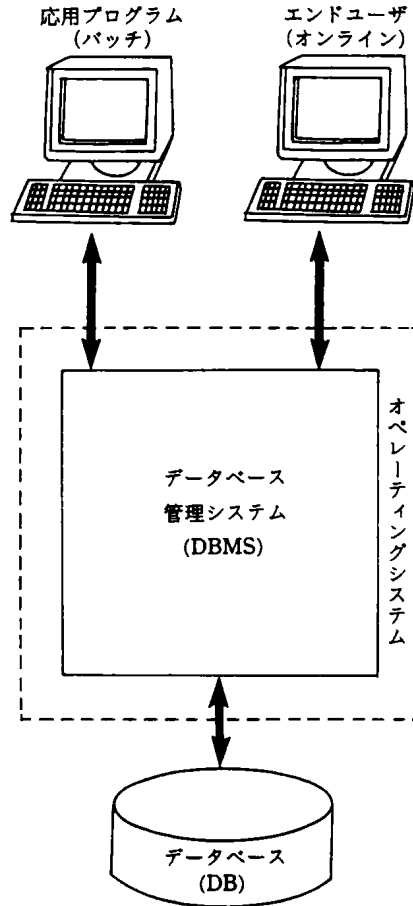


図1.1 データベースとデータベース管理システム

のではないかと思える。

データは英語で書けば data, 一方, 情報は information であって, 両者の概念は明らかに異なる。にもかかわらず日本ではデータというべきところを情報といたり記したりしていることが多くあるようであり, また誰もそれをとがめないようである。

データは実験や観察, 調査や捜査, 検査や検診, あるいは様々な活動や営為の結果得られる文字や数値の並び, 少し抽象度をあげれば, 記号 (symbol) の

集まりであって、それ以上でもそれ以下でもない。データは表やグラフ、図、あるいは写真や画像、模型、音等の表現をとることもある。

一方**情報**は、データの**受け手**、あるいはユーザの存在を前提として成り立つ概念である。つまり、データはその受け手に情報を与えることもあり、そうでないこともある。その関係を図1.2に示す。つまり、データはその受け手によりまず意味解釈される。これは辞書、構文規則、意味規則を用いてなされよう。

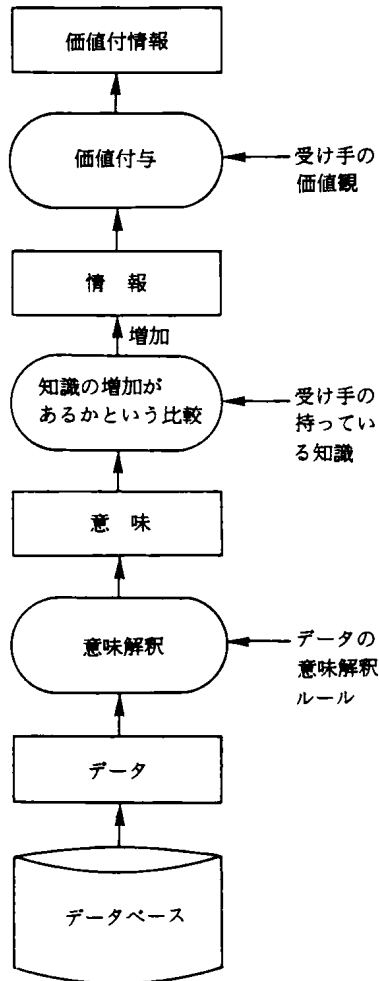


図1.2 データと情報

リレーショナルデータモデルの誕生

さて、データベースの概念を根底から変えてしまった男がいた。それがリレーショナルデータベースの生みの親である Edgar F. Codd (テッド コッドと呼ぶ。Ted は Edgar の縮小形。Robert を Bob という類。) 博士だ。博士の顔写真を写真 1 に示す。コード博士はイギリスのオックスフォード大学で数学を専攻、学士、修士の称号を授与されたのち、米国ミシガン大学の計算機および通信科学科に入学し、博士の称号を授与されている。1949年に米国 IBM 社に入社、汎用多重プログラミング機能を有する最初のオペレーティングシステムの開発を率いたりした後、オートマトン理論の研究を行ない、1968年には Academic Press 社から“*Cellular Automata*”という単行本を出版している。彼のこの分野での業績はあまねく知られている。筆者も1960年代後半はオートマトン理論の研究を行なっていたので、後にコード博士の有名なリレーショナルデータモデルを提案した最初の論文“A Relational Model of Data for Large Shared Data Banks,” *Communications of the Association for Computing Machinery*, Vol. 13, No. 6, pp. 377-387 (1970年6月) に接したとき、この論文の著者の E. F. Codd 氏と、あのオートマトンの E. F. Codd 氏が同一の Codd 氏であることを納得するのに、しばしの時間と苦痛をおぼえたほどである。(上記論文表題で Databases とワンワードはおろか、Data Bases ともならず、Data Banks となっているところが、時

写真1 エドガーF. コッド博士