

データベースの超初級

Hukanzen

平成30年11月8日

目 次

第 1 章	はじめに	2
第 2 章	データベースとは	3
2.1	データベース	3
2.1.1	データモデル	3
2.1.2	データベース管理システム	3
2.1.3	データベースとファイルの違い	4

第1章 はじめに

本書は若造も良いところの、未熟ポンコツ学生が教科書¹の内容を削りつつ、自分なりにわかりやすく記述する。そのため、当然間違いや、歯抜けもあるだろうが、その時は、指摘してくれると有難い²。本書の用途としては、自分の振り返りや後輩等への教育、データベースに対するイメージ付け程度である。そのため、本書を理解したからと言って、データベースが完璧なわけではないし、本書がさっぱりわからないからと言って、データベースから遠ざかったりしないで欲しい。特に、リレーショナルデータベースなどは、数学の集合論に基づいて作成されている。そのため、当たり前のことだが、集合論に対する理解や知識が、最低限必要となってくる。しかし、高専3年や大学1年生といった、若造は、集合論の概念はなかなかハードルが高く、存在自体知らない場合もある。そこで、本書では極力数学的な知識を必要とせず、表や図を用いて説明を行い、補助的に数学を用いる程度に留める。もし、数学などを通じた理論を必要として、本書を見つけたのなら、即刻引き返して欲しい。おそらく、数学とデータベースの関連に気付いている人にとっては、まったくの無価値であろう。

繰り返すが、本書は、あくまで、データベースに対する印象づけや、感覚を養う程度であり、理論に関しては、全く役に立たないことを御理解頂きたい。あと、もうちょっと本章増やす。

¹問題があれば、教えてください(震え)。

²当て字をよく使うので、お見苦しければ、指摘してください。

第2章 データベースとは

2.1 データベース

データベースとは、コンピュータ内に構築された「実世界の写し絵」である。実世界のものを、データモデル¹を用いて、データモデリング²することで、データベースを作成することができる。そして、データベース管理システムを用いて、データを管理³する。

用語としての「データベース」は主に、2つの意味を持つ。

- コンテンツとしてのデータの格納庫
- Oracle, MySQL, PostgreSQL などの、データベース管理システム

しかし、それをいうと、Google 等で検索できるインターネットの情報は、地球規模のデータベースになってしまう。原義を考えると、データの基地なので、ケースバイケースで読み取ろう。

2.1.1 データモデル

現在、データモデルには、大きく分けると3つ存在する。これらは世代ごとに順になっている。

第1世代 ネットワーク・データモデル、ハイアラキカル・データモデル

第2世代 リレーショナル・データモデル⁴

第3世代 オブジェクト指向・データモデル

一般的には、リレーショナル・データモデルに基づいた、リレーショナル・データベース⁵が広く使われおり、講義で行う内容もリレーショナル・データベース中心であることが多い。

第1世代は、1960年代当時のコンピュータ事情に即して、考案された。親子集合とレコード型の要素を用いており、レコードとレコードは、ポインタで関連付けられる。そのため、細かなデータを得る場合は、繰り返し親から子に辿る必要があり、非効率である。第2世代は、徹底的にフォーマルなデータモデルであり、数学の集合論に基づいて作成されてる。全てのデータを表で管理し、リレーショナル代数やリレーショナル論理を用いて、データ集合を得る。従って、ポインタやレコードの実装技術などとは、全く無関係である。なお、用語として、行を示すことを「レコード」と呼ぶことがあるが、恐らく間違った使い方である。第3世代は、CAD等のエンジニアリングデータやマルチメディアデータ、文書や図版が入り混じった文書データなどの部品展開構造を持つ、複合オブジェクトを表現するのに向いている。

以上より、実世界では同じ内容でも、データモデルの違いにより、異なって表現される。本来は、データベースに格納するデータに合わせて、それに適したデータモデルを使うべきである。しかしながら、一般的には、リレーショナル・データモデルで全て解決させてしまうことが多い。

2.1.2 データベース管理システム

データベース管理システム⁶とは、データベースに存在するデータを「管理」するシステムである。商用、オープンソースでは、Oracle, SQL Server, PostgreSQL, MySQL, Maria DBなどを指す。通常、OS⁷の力を借りて、その上で稼働するミドルウェアである。基本的に、次に示す3つの機能を有する。

¹記号系

²実世界のデータ化する過程

³取得、挿入、更新など

⁴関係モデル

⁵関係データベース

⁶DataBase Management System

⁷Operating System

- メタデータ管理機能
- 質問処理機能
- トランザクション管理機能

メタデータ管理機能とは、「データのデータ」としてのメタデータを管理する。データベースのデータ構造などを管理するデータでありそのデータの管理を行う。ユーザと DBMS そのものに対して用いられる。ユーザに対して中にどのようなデータがどのように格納されているかを示す。DBMS に対しては、自分が管理しているデータの種類、サイズ、インデックス、アクセス件など、基本的な情報提供に不可欠なものである。質問処理機能とは、文字通り、ユーザやアプリケーションが要求する「質問」に対し、処理を行う。特にリレーショナル・データベースでは、SQL に代表される、質問言語を用いる。トランザクションとは、DBMS に対するアプリケーションレベルの仕事の単位である。障害時回復、同時実行制御の 2 つの機能を有しており、一貫性の保持や、複数人の同時実行を行う。

2.1.3 データベースとファイルの違い

データベースに比較して、プログラムごとにデータをファイル保存する手法がある。C 言語ならば、FILE 構造体を用いて、fprintf(), fscanf() 関数などで、ファイルに対して入出力を行える。世の中には、馬鹿げた話であるが、中規模で大量のデータのやり取りを行う場合でも、頑なにファイル保存から離れない開発者も存在する。データというのは、組織体にとって、共有資源であり、管理し、様々な場所に使用することに価値がある。しかし、この場合だと、データはプログラムに従属する形になり、無秩序になりがちである。また、データ形式の違いから、同じデータを複数保存したり、プログラムによるアクセスのタイミングから矛盾が発生することがある。また、近年は、Microsoft Excel などの高機能な表計算ソフトの登場により、表計算ソフトをデータ保存に用いている場合も有る。この場合、一見、完璧に一元管理が可能に見えるが、異なるファイルに対しては、同様に矛盾が発生しやすく、一元管理がなされないことが多い。

データベースに関しては、DBMS を用いることで、決まった形にデータを保存することが可能である。また、全てのプログラムから 1 つのデータベースにアクセスすることで、重複や矛盾を排除し、データを一元管理することができる。データは共有資源にすることに意味があり、様々なプログラムにさせることは必須である。