



連載

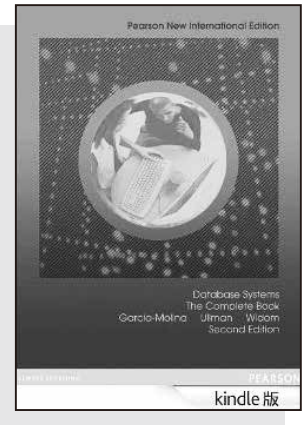
ビブリオ・トーク
-私のオススメ-

… 西澤 格 ((株)日立製作所)

Database Systems : The Complete Book

Hector Garcia-Molina, Jeffrey D. Ullman and Jennifer Widom 著
Pearson Education Limited (印刷版 : 2008, Kindle 版 : 2013),
1152p., ISBN : 9781292024479

[US 原書] ISBN : 9780131873254 US\$202.00 出版元 : ピアソン



紹介する本

今回紹介するのは, Hector Garcia-Molina, Jeffrey D. Ullman and Jennifer Widom による “Database Systems : The Complete Book” である. 私の手元にある紙の本は, 2001 年の Prentice Hall 版であるが, Complete と銘打っているだけあって, データベースシステムおよびデータ処理の関連トピックを網羅的にカバーしており, 1,000 ページを超えていた.

この Prentice Hall 版は私が 2002 年に Stanford 大学に滞在していた際に, 受け入れ元の Jennifer Widom 教授にサインを入れていただいた思い出のある本なのだが, 現在では入手が難しいようである. そのため, 今回のビブリオ・トーク執筆に伴い, 新版 (といっても少々古いが…) を購入し, 再度目を通した. 新版は新しいトピックをカバーするいくつかの章が追加され, さらにボリュームアップしている.

著者の紹介と本書の位置づけ

本書の著者らは米 Stanford 大学の計算機科学科の教授であるとともに, データ管理系の研究者としてインパクトのある論文を多数発表されている. 研究室運営に加えて, 学生の指導, 企業との連携で非常に多忙な中, この内容とボリュームの本を書きあげていく生産性に驚かされる. 特に Ullman 教授は, データベース以外にもコンパイラ, 計算理論などの名著を多数執筆されている.

本書は, 教授らのデータ管理系の授業の講義内容をベースとして執筆されている. 当時の講義は教授らの用意したスライド, 板書をベースに進められ, 直接この本を参照していたわけではないが, 本

書はその内容から, 教科書と呼ぶのが適切と感じる. Ullman 教授の Web ページ (2015 年 1 月 31 日現在の URL は <http://infolab.stanford.edu/~ullman/dscb.html>) には本書に関する補助資料として, 講義でのプロジェクト紹介, 本書の練習問題の回答, 講義で使用されたスライドがまとめられている.

本書の内容紹介

データベースシステム, 特に関係データベースシステムは, 1960 年代から積み上げられた理論に裏打ちされた, 最も高度かつ実用的なソフトウェアの 1 つである. 本書は, データベースシステムの理論, 実装, 関連トピックの非常に広い範囲をカバーしており, 大きく 5 部 23 章からなる.

まず第 1 部では関係データベースシステムの基礎となる関係モデルの基礎, 関係演算, 関係データベースの設計理論が説明される. 続く第 2 部では, 関係モデルに対する抽象プログラミング言語として, 関係代数と関係論理が, そしてデータベースシステムのデファクトの問合せ言語である SQL が詳細に説明される. 第 3 部では, 半構造データモデル, その 1 つの実装としての XML, および XML のためのプログラミング言語として XPath, XQuery が紹介される. そして第 4 部では, データベースシステムの実装に関する数多くのトピックが説明される. そして最後の第 5 部では情報統合, および検索エンジン, ストリームデータ処理などの関連トピックが紹介される.

今回は, この中から第 2 部の 5 章で説明される Datalog と, 第 4 部のデータベースシステムの実装の 2 つを紹介する.

Datalog

ここでは厳密な定義は紹介しないが、第2部の5章で紹介されている Datalog (database logic) は、if-then ルールで構成される宣言的な問合せ記述言語（どうやって処理するかでなく、処理で得られる結果の条件を書く言語）である。たとえばスキーマが Movies (title, year, length, genre) として定義される Movies データベースから、100分以上の上映時間の映画のタイトルと公開年を抽出する LongMovie (t, y) は、以下のように定義される。

```
LongMovie (t, y) ←
```

```
  Movies (t, y, l, g) AND l >= 100
```

ルールは矢印の後が満足されると、矢印の前が成り立つと読む。すなわち、ifが矢印の後、thenが矢印の前になる。私は、演繹データベースをかじった際に Datalog に触れ、少し表現がおかしいかもしれないがデータとルールが統合された非常に美しいモデルに感銘を受けた。

本書では Datalog は、SQL では複雑になる再帰クエリをとってもシンプルに書くことができると控えめに紹介されており、ほかには21章の情報統合のルールの記述への適用例が紹介されているにとどまっている。現在では、ネットワークの記述やプログラム解析への適用例も報告されているので、興味をお持ちいただけたならぜひ読んでみていただきたい。

データベースシステムの実装

本書では、第4部のデータベースシステムの実装が450ページ超と最もボリュームが大きい。ここでは、二次記憶管理、インデックス構造、クエリ実行機構、バッファ管理、クエリコンパイラ、トランザクション処理機構が説明される。ここを読めばシンプルなデータベース管理システム（以下、DBMS）が開発できる程度の記述レベルである。

Stanford 滞在中に、第4部でカバーされている Widom 教授の講義を聴講した。講義では、各学生が3カ月の間に DBMS を実装する。ある日の講義でインデックス機構が説明されたら、その週の課題はインデックス機構の実装となり、それぞれの学生は提出期限までにコードをシステムに登録する。そして登録されたソースコードを TA (ティーチングアシスタント) がテストし、実装が評価される。そして DBMS 全体を実装した後、それぞれが実装した DBMS でベンチマークを実行する。それぞれの課題に加えて、ベンチマークの優勝者には高い評価が与えられていた。

学生一人ひとりが短い期間で DBMS を実装していくのを目の当たりにして、当時非常に驚き、かつその効率的な教育システムに衝撃を受けたことを思い出す。データベースシステムの実装について書かれた本は少なく、本パートだけをとっても本書を読む価値は高いと考える。

オズメの読み方

本書は1,100ページを超え、かつ翻訳は出ていない。英語は平易かつ例も多いため比較的読みやすいが、くまなく全部を読もうとすると途中でめげてしまう恐れがある。そこで私は、まず目次を見て自分の興味を持った章、もしくは研究や業務に関連する章をざっと読み、さらに深く学びたい場合には元論文にあたる、リファレンス的な読み方をお勧めしたい。

(2015年2月2日受付)

西澤 格 (正会員) itaru.nishizawa.cw@hitachi.com

東京大学大学院工学系研究科博士課程修了 (1996)、博士 (工学)、日立製作所中央研究所、米 Stanford 大学客員研究員を経て、2013年より日立製作所 情報・通信システム社 IT プラットフォーム事業本部主任技師、ストリームデータ処理システムの開発に従事。