

教育部高等教育司推荐  
国外优秀信息科学与技术系列教学用书

# 数据库与事务处理

——面向应用的方法

(影印版)

## DATABASES AND TRANSACTION PROCESSING

An Application-Oriented Approach

■ Philip M. Lewis  
Arthur Bernstein  
Michael Kifer



高等教育出版社  
Higher Education Press



Pearson Education  
出版集团

教育部高等教育司推荐  
国外优秀信息科学与技术系列教学用书

# 数据库与事务处理

## ——面向应用的方法

(影印版)

## DATABASES AND TRANSACTION PROCESSING

**An Application-Oriented Approach**

Philip M. Lewis  
Arthur Bernstein  
Michael Kifer



高等教育出版社



Pearson Education 出版集团

图字: 01-2002-3780 号

**Databases and Transaction Processing: An Application-Oriented Approach, First Edition**

Philip M. Lewis, Arthur Bernstein, Michael Kifer

本书封面贴有 Pearson Education 出版集团激光防伪标签, 无标签者不得销售。

English reprint edition copyright ©2002 by PEARSON EDUCATION NORTH ASIA LIMITED and HIGHER EDUCATION PRESS. (Databases and Transaction Processing: An Application-Oriented Approach from Addison-Wesley Publishing Company, Inc.'s edition of the Work)

**Databases and Transaction Processing: An Application-Oriented Approach, 1e** by Philip M. Lewis, Arthur Bernstein and Michael Kifer, Copyright ©2002.  
All Rights Reserved.

Published by arrangement with the original publisher, Pearson Education, Inc., publishing as Addison-Wesley Publishing Company, Inc.

This edition is authorized for sale only in the People's Republic of China (excluding the Special Administrative Regions of Hong Kong and Macau).

**图书在版编目(CIP)数据**

数据库与事务处理: 面向应用的方法 / (美) 刘易斯 (Lewis, P. M.), (美) 伯恩斯坦 (Bernstein, A.), (美) 基弗 (Kifer, M.) 著. —影印本. —北京: 高等教育出版社, 2002.10

ISBN 7-04-011396-1

I. 数... II. ①刘...②伯...③基... III. 数据库系统 - 高等学校 - 教材 - 英文 IV. TP311.13

中国版本图书馆 CIP 数据核字(2002)第 081697 号

数据库与事务处理——面向应用的方法(影印版)

Philip M. Lewis, Arthur Bernstein, Michael Kifer

出版发行 高等教育出版社  
社 址 北京市东城区沙滩后街 55 号  
邮政编码 100009  
传 真 010-64014048  
经 销 新华书店北京发行所  
印 刷 北京二〇七工厂  
开 本 787×1092 1/16  
印 张 65.75  
字 数 1 620 000

购书热线 010-64054588  
免费咨询 800-810-0598  
网 址 <http://www.hep.edu.cn>  
<http://www.hep.com.cn>  
版 次 2002 年 10 月第 1 版  
印 次 2002 年 10 月第 1 次印刷  
定 价 65.00 元

本书如有缺页、倒页、脱页等质量问题, 请到所购图书销售部门联系调换。

版权所有 侵权必究



# *Preface*

Database and transaction processing systems occupy a central position in our information-based society. Virtually every large system with which we interact in our daily lives has a database at its core. The systems range from those that control the most trivial aspects of our lives (e.g., supermarket checkout systems) to those on which our lives depend (e.g., air traffic control systems). Over the next decades, we will become increasingly dependent on the correctness and efficiency of these systems.

We believe that every computer scientist and information systems professional should be familiar with the theoretical and engineering concepts that underlie these systems. These are the people who will be designing, building, maintaining, and administering these highly complex systems.

This book is intended to be a text for any of the following courses in a computer science or technically oriented information systems curriculum:

- An introductory undergraduate or graduate course in databases
- An undergraduate or graduate course in transaction processing for students who have had an introductory course in databases
- An advanced undergraduate or a first graduate course in databases for students who have had an introductory course in databases

If only one course is to be taught covering both databases and transaction processing, the instructor can select material related to both topics.

Rather than focusing on how to build the database management system itself, our approach focuses on how to build applications. We believe that many more students will be implementing applications than will be building DBMSs. We believe that placing databases in the context of transaction processing accentuates this emphasis, since transactions provide the mechanism that applications use to access databases. Furthermore, we include substantial material describing the languages and APIs used by transactions to access a database, such as embedded SQL, ODBC, and JDBC.

While the book thoroughly covers conventional topics: relational databases, SQL, and the ACID properties of transactions it also provides a very substantial treatment of less conventional and newer issues, such as object and object relational

databases, XML and document processing over the Internet, and the transactional issues related to Internet commerce.

Although we cover many practical aspects of database and transaction processing applications, we are primarily concerned with the concepts that underlie these topics rather than on the details of particular commercial systems or applications. Thus, in the database portion of the book, we concentrate on the concepts underlying the relational and object data models rather than on any particular commercial DBMS. These concepts will remain the foundation of database processing long after SQL is obsolete. (Recall the generation of programmers who were trained in COBOL and found it extremely difficult to learn any other language.) In a similar way, in the transaction processing portion of the book, we concentrate on the concepts underlying the ACID properties and the technical issues involved in their implementation, rather than on any particular commercial DBMS or TP monitor.

To enhance students' understanding of the technical material, we have included a case study of a transaction processing application, the Student Registration System, which is carried through the book. While a student registration system can hardly be considered glamorous, it has the unique advantage that all students have interacted with such a system as users. More importantly, it turns out to be a surprisingly rich application, so we can use it to illustrate many of the issues in database design, query processing, and transaction processing.

A unique aspect of the book is a presentation of the software engineering concepts required to implement transaction processing applications, using the Student Registration System as an example. Since the implementations of many information systems fail because of poor project management and inadequate software engineering, we feel that these topics should be an important part of the student's education. Our treatment of software engineering issues is brief, as many students will take a separate course in this subject. However, we believe that they will be better able to understand and apply that material when they see it presented in the context of an information system implementation. Since the courses that use this text at Stony Brook are not software engineering courses, we do not cover this material in class. Instead, we ask the students to read it and require that they use good software engineering practice in their class projects. We do cover in class those aspects of the Student Registration System that illustrate important issues in databases and transaction processing.

## OVERVIEW

---

There is sufficient material in the book for three one-semester courses. The first half of the book is a text for a first course on databases. For students who have completed such a course, the second half of the book concentrates on transaction processing and advanced topics in databases. At Stony Brook, we offer both an undergraduate (introductory) and a more advanced (graduate) database course as well as an undergraduate and graduate version of the transaction processing course.

The book is divided into parts so that the instructor can more easily organize the material. We have included a Chapter Dependency Chart to make it easier to design customized courses.

### **Part I: Introduction**

Chapters 1 through 3 contain introductory material for a first course in databases. Chapter 1 serves as general introduction. Chapter 2 briefly covers SQL and the ACID properties of transactions. By introducing this basic material early, we are able to remove some constraints on the order of presentation of topics discussed later.

Chapter 3 begins our discussion of the Student Registration System and the software engineering concepts appropriate for its implementation. In particular, we discuss requirements and specification documents and the use of application generators to design graphical user interfaces. In the introductory database course at Stony Brook, we ask the students to read this material but do not cover it in class lectures. However, at this point in the course, we start the class project by asking students to write a Specification Document.

### **Part II: Database Management**

Chapters 4 through 15 constitute the core of a first course in databases. Some of the topics covered are:

- The concept of a relation and the DDL features of SQL, including automatic constraint checking
- The Entity-Relationship Model and schema design, including methods (and their limitations) for converting E-R diagrams to relational schemas
- Relational algebra, calculus, and the DML features of SQL with particular attention to the semantics of complex SQL queries through relational algebra and calculus
- Functional dependencies and normalization, including algorithms for decomposing a schema into 3NF, BCNF, and 4NF
- Triggers and active databases, including triggers in SQL:1999
- The inclusion of SQL statements in a program written in a conventional programming language, including embedded SQL, dynamic SQL, ODBC, JDBC, and SQLJ. The recently standardized language for stored procedures, SQL/PSM, is also discussed
- Physical organization of the data and indexing, including B<sup>+</sup> trees, ISAM, and hash indices
- Query processing and optimization, including algorithms for computing selections and joins, and methods for estimating the cost of query plans

Software engineering issues, as applied to the Student Registration System, are integrated throughout these chapters. In Section 5.7, we present the database design

for the system, including the E-R Diagram and relation schema. Chapter 12 presents material on Design Documents, Test Plan Documents, and project planning that is needed to complete the system. In Section 12.6, we present the detailed design and part of the Java/JDBC program for one of the transactions in the system.

Chapter 15 summarizes some of the material on transaction processing from later chapters. It can be used to enrich a database course if time permits.

### **Part III: Advanced Topics in Databases**

Chapters 16 through 19 contain materials that can become part of an advanced database course. Such a course includes all the chapters in this part plus Chapter 27. In our experience, lack of time prevents one or more sections of Chapters 7, 8, 9, 10, 11, and 14 from being covered in a first database course, so this material can also be included. The topics covered in Part III include

- Object and object-relational databases, including the conceptual model, ODMG databases, object-relational extensions of SQL:1999, and CORBA
- Database aspects of document processing on the Web, including a detailed discussion of XML Schema, XPath, XSLT, and XQuery
- Distributed databases, including heterogeneous and homogeneous systems, multidatabases, fragmentation, semijoins, global query optimization, query design, and distributed database design
- Online analytic processing and data mining, including star schemas, the CUBE and ROLLUP operators, associations, and classification

### **Part IV: Transaction Processing**

Chapters 20 through 27, together with portions of Chapters 9 and 10, contain material for a one-semester course in transaction processing. Many of the examples in these chapters refer to the Student Registration System design developed in Chapters 3 and 12 and in Section 5.7. We ask the students to read this material as appropriate.

Chapter 20 contains a detailed description of the ACID properties of transactions. Chapters 21 and 22 describe a variety of transaction models and the architecture of transaction processing systems in a distributed and heterogeneous client/server environment. Some of the topics covered are

- Models of transactions, including savepoints, chained transactions, transactional queues, nested and multilevel transactions, distributed transactions, multidatabase systems, and workflow systems
- Architectures for transaction processing systems, including, client-server organizations for both centralized and distributed databases, two-tiered and three-tiered architectures, TP monitors, and transaction managers. Transactional remote procedure call and peer-to-peer communication together with their use in organizing a transaction processing system are discussed

- Implementation of transaction architectures and models in transaction processing applications on the Internet

Chapters 23 through 26 describe how the ACID properties of atomicity, isolation, and durability are implemented in both centralized and distributed systems. Some of the topics covered are

- Concurrency controls for abstract databases, including strict two-phase locking, optimistic concurrency controls, timestamp-ordered concurrency controls, concurrency controls for object databases, and locking protocols to implement the different models of transactions
- Concurrency controls for relational databases, including locking protocols for the different isolation levels, examples of correct and incorrect schedules at each isolation level, granular locking, index locking, and multiversion concurrency controls, including SNAPSHOT isolation
- Logging and recovery, including write-ahead logs, dumps, and checkpoints
- Distributed transactions, including the two-phase commit protocol, global serialization, global deadlocks, and synchronous and asynchronous algorithms for managing replicated data

Chapter 27 covers security and Internet commerce. Some of the topics covered are

- Symmetric and asymmetric encryption, digital signatures, blind signatures, and certificates
- The Kerberos protocol for authentication and key distribution
- Internet protocols, including the SSL Protocol for authentication and session encryption, the SET Protocol for secure transactions, electronic cash protocols, and protocols that guarantee goods atomicity, certified delivery, and money atomicity

Our goals in Chapters 20 through 27 are

- To make students aware of the architecture of the transaction processing systems with which their transactions must interact, so that they can better evaluate the features offered by competing system vendors
- To describe the costs, measured in system resources and performance, involved in implementing the ACID properties of transactions
- To describe techniques to decrease these costs for example, granular locks, indices, denormalization, and table fragmentation.
- To describe situations in which an application will execute correctly even though isolation is not total for example, transactions that might execute correctly at an isolation level less stringent than `SERIALIZABLE`

We have also included an Appendix, which covers certain system issues that are important for understanding parts of the text. These include the ideas behind modular systems and encapsulation, the basics of the client/server architecture, multiprogramming and threads, and the basics of interprocess communication. Instructors might choose to present some of this material if the students have not covered it in previous courses.

## DEPENDENCIES AMONG CHAPTERS

---

To help instructors tailor the material to the needs of their courses, we have identified certain sections that can be skipped without disrupting the flow of material in the corresponding chapters. Such sections are marked with an asterisk. Even though material covered in these sections is sometimes referenced in other chapters, these references can be ignored. In addition, exercises that are more difficult than the rest are marked with one or two asterisks, depending on the difficulty.

The text can be used in a number of different ways depending on the goals of the course. To provide some guidance to the instructor, the table on the next page shows the chapters that might be included in five different courses that address different student populations and attempt to emphasize different aspects of the subject. In this table, **yes** means that all parts of the chapter should be covered by the lectures. **Parts** means that the instructor can select only parts of the material presented in the chapter. **Read** means that the chapter can be given as a reading assignment to the students.

Column 1 marks chapters that would be covered in a slow-paced introductory database course. In such a course, for instance, only parts of Chapter 8 on normalization theory might be included perhaps only the introductory sections. Similarly, only some parts of Chapter 10, on various ways in which SQL can be combined with a host language, might be covered perhaps only one approach, the one required for the course project.

Columns 2 and 3 outline two more intensive introductory database courses. Column 2 expands the material covered in the introductory course in the direction of database applications, while Column 3 describes a more theory-oriented version of the course. It provides a more in-depth coverage of the normalization theory, foundations of query languages, and query optimization at the expense of the application-oriented material in Chapter 10. Although we have characterized this material as theory-oriented, we might also have characterized it as system-oriented, because it covers issues involved in the design of a DBMS. The courses in these two columns are the ones we teach in our undergraduate program at Stony Brook (depending on the interests of the instructor).

Column 4 describes an advanced database course. The course might start by reviewing or filling in material that the instructor judges the students might not have covered in a prerequisite, introductory database course. Such material would probably be found in Chapters 7, 8, 9, 10, and 14. The body of the course then continues with advanced database topics and some material on transactions in

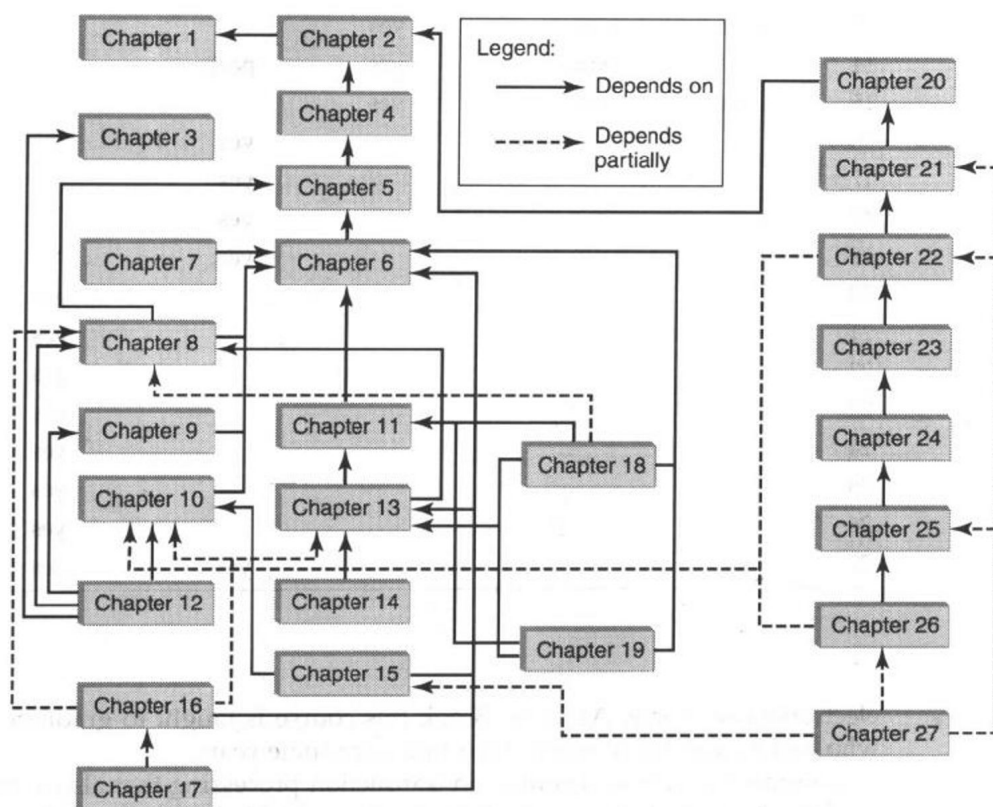
| Chapter | Courses  |                 |           |             |              |
|---------|----------|-----------------|-----------|-------------|--------------|
|         | DB/Intro | DB/Applications | DB/Theory | DB/Advanced | Transactions |
| 1       | yes      | yes             | yes       |             | yes          |
| 2       | yes      | yes             | yes       |             |              |
| 3       | read     | read            |           |             |              |
| 4       | yes      | yes             | yes       |             |              |
| 5       | yes      | yes             | yes       |             |              |
| 6       | yes      | yes             | yes       |             |              |
| 7       |          | parts           | yes       | parts       |              |
| 8       | parts    | parts           | yes       | parts       |              |
| 9       |          | yes             | yes       | yes         | yes          |
| 10      | parts    | yes             |           | parts       | yes          |
| 11      | yes      | yes             | yes       |             |              |
| 12      | read     | read            |           |             |              |
| 13      | parts    | yes             | yes       |             |              |
| 14      |          | parts           | yes       | parts       |              |
| 15      | yes      | yes             | yes       |             |              |
| 16      |          |                 |           | yes         |              |
| 17      |          |                 |           | yes         |              |
| 18      |          |                 |           | yes         |              |
| 19      |          |                 |           | yes         |              |
| 20      |          |                 |           |             | yes          |
| 21      |          |                 |           |             | yes          |
| 22      |          |                 |           |             | yes          |
| 23      |          |                 |           |             | yes          |
| 24      |          |                 |           |             | yes          |
| 25      |          |                 |           |             | yes          |
| 26      |          |                 |           |             | yes          |
| 27      |          |                 |           | yes         | yes          |

electronic commerce. At Stony Brook this course is taught to graduate students who had a database course in their undergraduate years.

Column 5 describes a course on transaction processing that also assumes that students have had an introductory database course as a prerequisite. At Stony Brook, we teach both an undergraduate and a graduate version of this course. The material on transaction processing is supplemented with related material that

might not have been covered in the student's prerequisite database course, for example, some material from Chapters 9 and 10.

For further fine-tuning of courses, the chapter dependency diagram that follows can be of help. The figure identifies two kinds of dependencies. Solid arrows indicate that one chapter depends on much of the material presented in another chapter, except for the sections marked as optional. Dotted arrows indicate weak dependency, which means that only a few concepts developed in the prerequisite chapter are used in the dependent chapter, and those concepts can be covered quickly. The dependencies involving Chapter 27 are a special case. It can be taught either at the end of a transaction processing course, in which case it depends on Chapters 21, 22, and 25, or at the end of a database course, in which case it depends on Chapter 15.



## SUPPLEMENTS

---

In addition to the text, the following supplementary materials are available to assist instructors:

- Online PowerPoint presentations for all chapters
- Online PowerPoint slides of all figures
- An online solution manual containing solutions for the exercises
- Any additional references, notes, errata, homeworks, exams, and so forth, that we think might be of interest to our readers

For more information on obtaining these supplements, please check online information for this book at [www.aw.com/cssupport](http://www.aw.com/cssupport). The solutions manual and PowerPoint presentations are available only to instructors through your Addison-Wesley sales representative.

## ACKNOWLEDGMENTS

---

We would like to thank the reviewers, whose comments and suggestions significantly improved the book:

Suad Alagic, Wichita University  
 Catriel Beerli, The Hebrew University  
 Rick Cattell, Sun Microsystems  
 Jan Chomicki, SUNY Buffalo  
 Henry A. Etlinger, Rochester Institute of Technology  
 Leonidas Fegaras, University of Texas at Arlington  
 Alan Fekete, University of Sydney  
 Johannes Gehrke, Cornell University  
 Jiawei Han, Simon Fraser University  
 Peter Honeyman, University of Michigan  
 Vijay Kumar, University of Missouri–Kansas City  
 Jonathan Lazar, Towson University  
 Dennis McLeod, University of Southern California  
 Rokia Missaoui, University of Quebec in Montreal  
 Clifford Neuman, University of Southern California  
 Fabian Pascal, Consultant  
 Sudha Ram, University of Arizona  
 Krithi Ramamritham, University of Massachusetts–Amherst and IIT Bombay

Andreas Reuter, International University in Germany, Bruchsal

Arijit Sengupta, Georgia State University

Munindar P. Singh, North Carolina State University

Greg Speegle, Baylor University

Junping Sun, Nova Southeastern University

Joe Trubicz, Consultant

Vassilis J. Tsotras, University of California, Riverside

Emilia E. Villarreal, California Polytechnic State University

We would also like to thank the following people who were kind enough to provide us with additional information and answers to our questions: Don Chamberlin, Daniela Florescu, Jim Gray, Pankaj Gupta, Rob Kelly, and C. Mohan.

Several people taught out of beta versions of the book and made useful comments and suggestions: David S. Warren and Radu Grosu. Joe Trubicz served not only as a reviewer when the manuscript was complete, but provided critical comments on early versions of many of the chapters.

A number of students were very helpful in reading and checking the correctness of various parts of the book: Ziyang Duan, Shiyong Lu, Guizhen Yang, and Yan Zhang.

Many thanks to the staff of the Computer Science Department at Stony Brook, and in particular Kathy Germana, who helped make things happen at work.

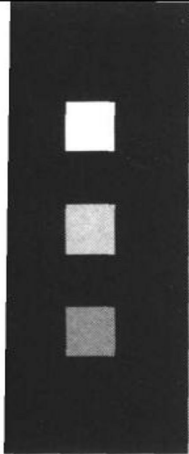
We would particularly like to thank Maite Suarez-Rivas, our editor at Addison-Wesley, who played an important role in shaping the contents and approach of the book in its early stages and throughout the time we were writing it. We would also like to thank the various staff members of Addison-Wesley, who did an excellent job of editing and producing the book: Katherine Harutunian, Pat Mahtani, Diane Freed, Regina Hagen, Paul Anagnostopoulos, and Jacqui Scarlott.

Last, but not least, we would like to thank our wives, Rhoda, Edie, and Lora, who provided much needed support and encouragement while we were writing the book.



# **DATABASES AND TRANSACTION PROCESSING**

## **An Application-Oriented Approach**



# *Contents*

Preface

xxiii

## **PART One** **1**

---

### **Introduction**

#### **1 Overview of Databases and Transactions** **3**

- 1.1 What Are Databases and Transactions? 3
- 1.2 Features of Modern Database and Transaction Processing Systems 5
- 1.3 Major Players in the Implementation and Support of Database and Transaction Processing Systems 9
- 1.4 Decision Support Systems OLAP and OLTP 10
- 1.5 Exercises 12

#### **2 A Closer Look** **15**

- 2.1 Case Study: A Student Registration System 15
- 2.2 Introduction to Relational Databases 15
- 2.3 What Makes a Program a Transaction? 21
- 2.4 Bibliographic Notes 26
- 2.5 Exercises 26

|           |                                                               |           |
|-----------|---------------------------------------------------------------|-----------|
| <b>cs</b> | <b>3 Case Study: Starting the Student Registration System</b> | <b>29</b> |
| 3.1       | Software Engineering Methodology                              | 29        |
| 3.2       | Requirements Document                                         | 31        |
| 3.3       | Requirements Analysis New Issues                              | 38        |
| 3.4       | Application Generators                                        | 40        |
| 3.5       | GUIs and Objects                                              | 40        |
| 3.6       | Events and Procedures                                         | 44        |
| 3.7       | Accessing Databases and Executing Transactions                | 46        |
| 3.8       | Specifying the Student Registration System                    | 48        |
| 3.9       | Specification Document                                        | 49        |
| 3.10      | Bibliographic Notes                                           | 50        |
| 3.11      | Exercises                                                     | 50        |

## PART Two 53

---

### Database Management

|          |                                  |           |
|----------|----------------------------------|-----------|
| <b>4</b> | <b>The Relational Data Model</b> | <b>55</b> |
| 4.1      | What Is a Data Model?            | 55        |
| 4.2      | The Relational Model             | 58        |
| 4.2.1    | Basic Concepts                   | 59        |
| 4.2.2    | Integrity Constraints            | 62        |
| 4.3      | SQL Data Definition Sublanguage  | 69        |
| 4.3.1    | Specifying the Relation Type     | 70        |
| 4.3.2    | The System Catalog               | 70        |
| 4.3.3    | Key Constraints                  | 70        |
| 4.3.4    | Dealing with Missing Information | 71        |
| 4.3.5    | Semantic Constraints             | 72        |
| 4.3.6    | User-Defined Domains             | 75        |
| 4.3.7    | Foreign Key Constraints          | 76        |
| 4.3.8    | Reactive Constraints             | 78        |
| 4.3.9    | Database Views                   | 81        |
| 4.3.10   | Modifying Existing Definitions   | 82        |
| 4.3.11   | SQL-Schemas                      | 84        |
| 4.3.12   | Access Control                   | 85        |

4.4 Bibliographic Notes 87

4.5 Exercises 87

## 5 Database Design I: The Entity-Relationship Model 89

5.1 Conceptual Modeling with the E-R Approach 89

5.2 Entities and Entity Types 90

5.3 Relationships and Relationship Types 94

5.4 Advanced Features of the E-R Approach 99

5.4.1 Entity Type Hierarchies 99

5.4.2 Participation Constraints 103

5.5 A Brokerage Firm Example 106

5.6 Limitations of the E-R Approach 110

■ 5.7 Case Study: A Design for the Student Registration System 114

5.8 Bibliographic Notes 119

5.9 Exercises 119

## 6 Query Languages I: Relational Algebra and SQL 123

6.1 Relational Algebra: Under the Hood of SQL 124

6.1.1 Basic Operators 124

6.1.2 Derived Operators 133

6.2 The Query Sublanguage of SQL 140

6.2.1 Simple SQL Queries 140

6.2.2 Set Operations 146

6.2.3 Nested Queries 149

6.2.4 Aggregation over Data 154

6.2.5 A Simple Query Evaluation Algorithm 159

6.2.6 More on Views in SQL 161

6.2.7 The Null Value Quandary 166

6.3 Modifying Relation Instances in SQL 167

6.4 Bibliographic Notes 172

6.5 Exercises 172