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

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

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。

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。

前　　言

　　20 世纪末，以计算机和通信技术为代表的信息科学和技术对世界经济、科技、军事、教育和文化等产生了深刻影响。信息科学技术的迅速普及和应用，带动了世界范围信息产业的蓬勃发展，为许多国家带来了丰厚的回报。

　　进入 21 世纪，尤其随着我国加入 WTO，信息产业的国际竞争将更加激烈。我国信息产业虽然在 20 世纪末取得了迅猛发展，但与发达国家相比，甚至与印度、爱尔兰等国家相比，还有很大差距。国家信息化的发展速度和信息产业的国际竞争能力，最终都将取决于信息科学技术人才的质量和数量。引进国外信息科学和技术优秀教材，在有条件的学校推动开展英语授课或双语教学，是教育部为加快培养大批高质量的信息技术人才采取的一项重要举措。

　　为此，教育部要求由高等教育出版社首先开展信息科学和技术教材的引进试点工作。同时提出了两点要求，一是要高水平，二是要低价格。在高等教育出版社和信息科学技术引进教材专家组的努力下，经过比较短的时间，第一批引进的 20 多种教材已经陆续出版。这套教材出版后受到了广泛的好评，其中有不少是世界信息科学技术领域著名专家、教授的经典之作和反映信息科学技术最新进展的优秀作品，代表了目前世界信息科学技术教育的一流水平，而且价格也是最优惠的，与国内同类自编教材相当。

　　这项教材引进工作是在教育部高等教育司和高教社的共同组织下，由国内信息科学技术领域的专家、教授广泛参与，在对大量国外教材进行多次遴选的基础上，参考了国内和国外著名大学相关专业的课程设置进行系统引进的。其中，John Wiley 公司出版的贝尔实验室信息科学研究中心副总裁 Silberschatz 教授的经典著作《操作系统概念》，是我们经过反复谈判，做了很多努力才得以引进的。William Stallings 先生曾编写了在美国深受欢迎的信息科学技术系列教材，其中有多种教材获得过美国教材和学术著作者协会颁发的计算机科学与工程教材奖，这批引进教材中就有他的两本著作。留美中国学者 Jiawei Han 先生的《数据挖掘》是该领域中具有里程碑意义的著作。由达特茅斯学院 Thomas Cormen 和麻省理工学院、哥伦比亚大学的几

位学者共同编著的经典著作《算法导论》，在经历了 11 年的锤炼之后于 2001 年出版了第二版。目前任教于美国 Massachusetts 大学的 James Kurose 教授，曾在美国三所高校先后 10 次获得杰出教师或杰出教学奖，由他主编的《计算机网络》出版后，以其体系新颖、内容先进而倍受欢迎。在努力降低引进教材售价方面，高等教育出版社做了大量和细致的工作。这套引进的教材体现了权威性、系统性、先进性和经济性等特点。

教育部也希望国内和国外的出版商积极参与此项工作，共同促进中国信息技术教育和信息产业的发展。我们在与外商的谈判工作中，不仅要坚定不移地引进国外最优秀的教材，而且还要千方百计地将版权转让费降下来，要让引进教材的价格与国内自编教材相当，让广大教师和学生负担得起。中国的教育市场巨大，外国出版公司和国内出版社要通过扩大发行数量取得效益。

在引进教材的同时，我们还应做好消化吸收，注意学习国外先进的教学思想和教学方法，提高自编教材的水平，使我们的教学和教材在内容体系上，在理论与实践的结合上，在培养学生的动手能力上能有较大的突破和创新。

目前，教育部正在全国 35 所高校推动示范性软件学院的建设和实施，这也是加快培养信息科学技术人才的重要举措之一。示范性软件学院要立足于培养具有国际竞争力的实用性软件人才，与国外知名高校或著名企业合作办学，以国内外著名 IT 企业为实践教学基地，聘请国内外知名教授和软件专家授课，还要率先使用引进教材开展教学。

我们希望通过这些举措，能在较短的时间，为我国培养一大批高质量的信息技术人才，提高我国软件人才的国际竞争力，促进我国信息产业的快速发展，加快推动国家信息化进程，进而带动整个国民经济的跨越式发展。

教育部高等教育司
二〇〇二年三月

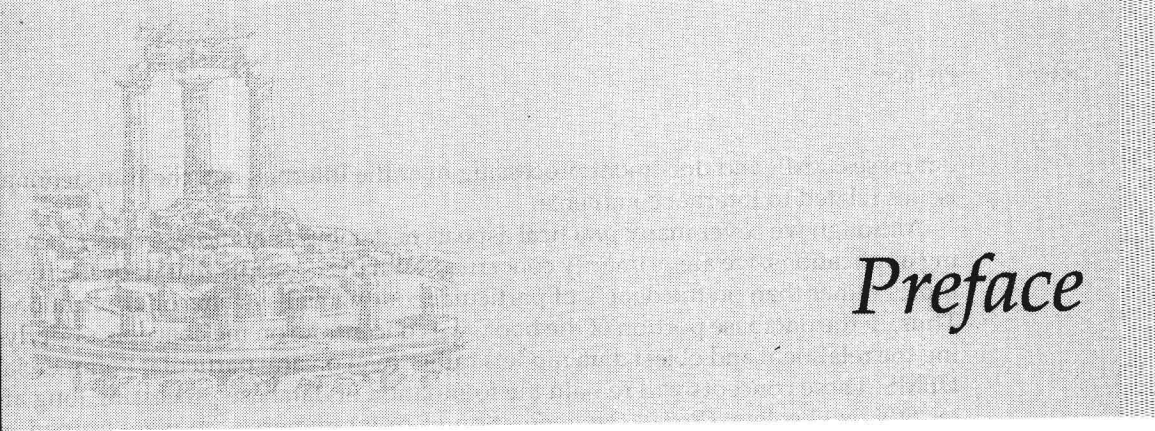*To my wife, Rhoda, my children, and my grandchildren.*   P.M.L.

To my wife, Edie, my children, and my grandchildren.   A.J.B.

To Lora, children, and parents.   M.K.

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 r elational 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^+ 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 per haps 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 per haps 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.
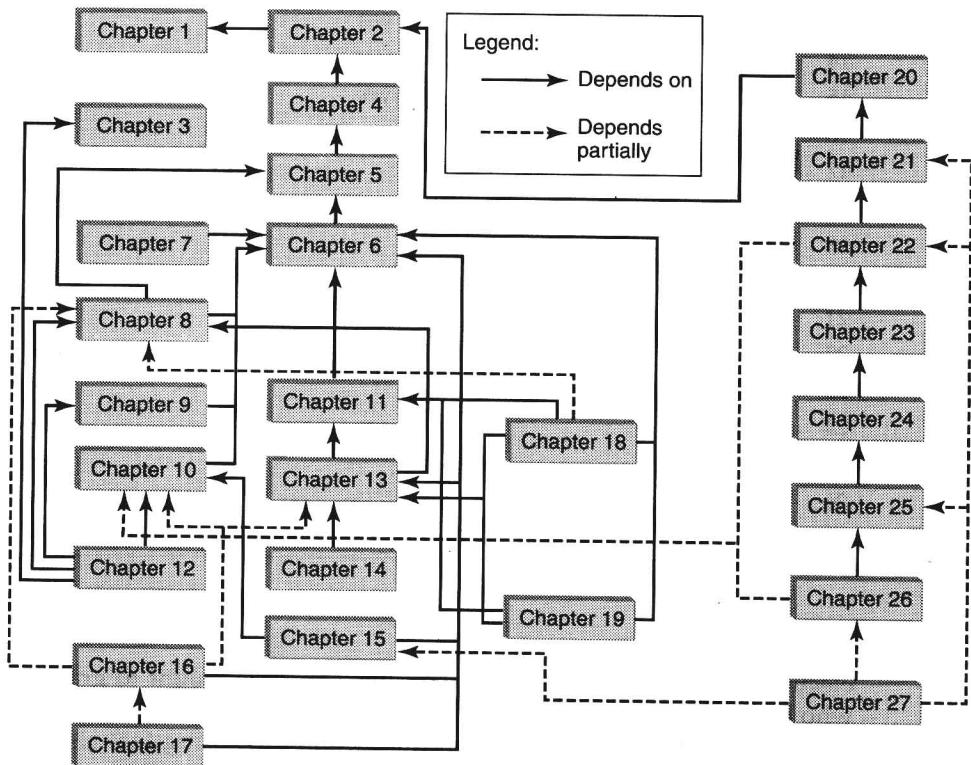
SUPPLEMENTS

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*. The solutions manual and Power-Point 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 Beeri, The Hebrew University

Rick Cattel, Sun Microsystems

Jan Chomicki, SUNY Buffalo

Henry A. Etlinger, Rochester Institute of Technology

Leonidas Fegaras, University of Texas at Arlington

Alan Fekete, University of Sidney

Johannes Gehrke, Cornell University

Jiawei Han, Simon Fraser Universtiy

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

DATABASES AND TRANSACTION PROCESSING

An Application–Oriented Approach