HZ BOOKS
华章教育

Mc Graw Hill Education

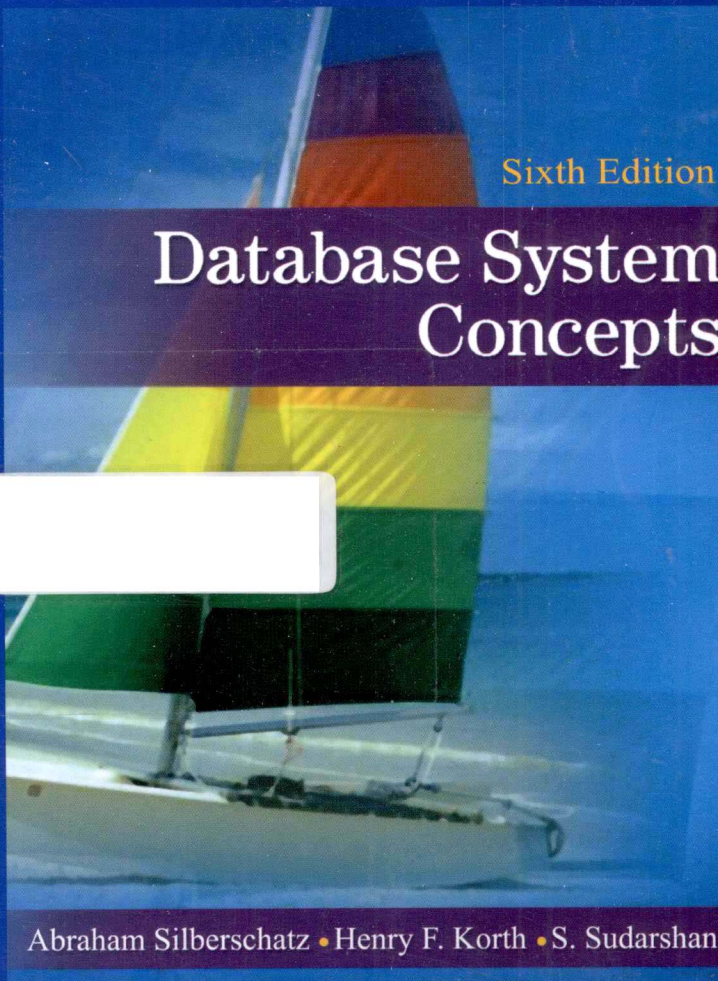
经 典 原 版 书 库

数据库系统概念

（美）Abraham Silberschatz 耶鲁大学
Henry F. Korth 利哈伊大学
（印）S. Sudarshan 印度理工学院 著
杨冬青 北京大学 改编

（英文精编版·第6版）

Database System Concepts (Sixth Edition)



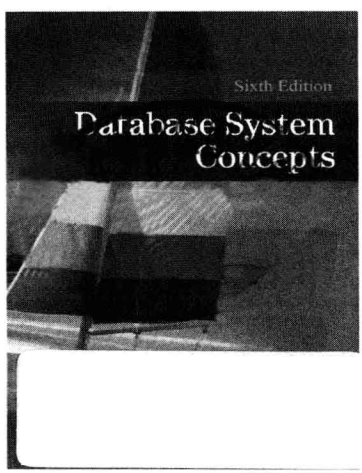
机械工业出版社
China Machine Press

经 典 原 版 书 库

数据库系统概念

（英文精编版·第6版）

Database System Concepts (Sixth Edition)



（美）Abraham Silberschatz 耶鲁大学 Henry F. Korth 利哈伊大学 （印）S. Sudarshan 印度理工学院 著 杨冬青 北京大学 改编

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

数据库系统概念（英文精编版·第6版）/（美）西尔伯沙茨（Silberschatz, A.）等著；杨冬青改编. —北京：机械工业出版社，2012.11
（经典原版书库）
书名原文：Database System Concepts, Sixth Edition

ISBN 978-7-111-40086-8

I. 数… II. ①西… ②杨… III. 数据库系统－教材－英文 IV. TP311.13

中国版本图书馆CIP数据核字（2012）第245209号

本书版权登记号：图字：01-2011-5829

机械工业出版社（北京市西城区百万庄大街22号 邮政编码 100037）
责任编辑：迟振春
北京京师印务有限公司印刷
2013年1月第1版第1次印刷
150mm×214mm·24.25印张
标准书号：ISBN 978-7-111-40086-8
定 价：69.00元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：(010) 88378991 88361066 投稿热线：(010) 88379604
购书热线：(010) 68326294 88379649 68995259 读者信箱：hzjsj@hzbook.com

出版者的话

文艺复兴以降，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域取得了垄断性的优势；也正是这样的传统，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭示了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切。这对计算机教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短的现状下，美国等发达国家在其计算机科学发展的几十年间积淀和发展的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起到积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章公司较早意识到“出版要为教育服务”。自1998年开始，我们就将工作重点放在了遴选、移译国外优秀教材上。经过多年的不懈努力，我们与Pearson，McGraw-Hill，Elsevier，MIT，John Wiley & Sons，Cengage等世界著名出版公司建立了良好的合作关系，从他们现有的数百种教材中甄选出Andrew S. Tanenbaum，Bjarne Stroustrup，Brain W. Kernighan，Dennis Ritchie，Jim Gray，Afred V. Aho，John E. Hopcroft，Jeffrey D. Ullman，Abraham Silberschatz，William Stallings，Donald E. Knuth，John L. Hennessy，Larry L. Peterson等大师名家的一批经典作品，以“计算机科学丛书”为总称出版，供读者学习、研究及珍藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力襄助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作；而原书的作者也相当关注其作品在中国的传播，有的还专程为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近两百个品

种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍。其影印版“经典原版书库”作为姊妹篇也被越来越多实施双语教学的学校所采用。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证。随着计算机科学与技术专业学科建设的不断完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都将步入一个新的阶段，我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方法如下：

华章网站：www.hzbook.com
电子邮件：hzjsj@hzbook.com
联系电话：（010）88379604
联系地址：北京市西城区百万庄南街1号
邮政编码：100037

华章科技图书出版中心

改编者序

数据库系统是对数据进行存储、管理、处理和维护的软件系统，是现代计算环境中的一个核心成分。随着计算机硬件、软件技术的飞速发展和计算机系统在各行各业的广泛应用，数据库技术的发展尤其迅速，引人注目。有关数据库系统的理论和技术是计算机科学技术教育中必不可少的部分。《数据库系统概念》是一本经典的、备受赞扬的数据库系统教科书，其内容由浅入深，既包含数据库系统的基本概念，又反映数据库技术新进展。本书被国际上许多著名大学采用，并多次再版。

我们先后将本书的第3版、第4版、第5版和第6版译成中文，由机械工业出版社分别于2000年、2003年、2006年和2012年出版发行。国内许多大学采用《数据库系统概念》作为本科生和研究生数据库课程的教材或主要教学参考书，收到了良好的效果。

我们基于《数据库系统概念》第5版进行了改编，保留其中的基本内容，压缩或删除了一些高级内容，形成了该书的本科教学版，其目的是使它更适合本科生的数据库课程使用。该本科教学版由机械工业出版社于2008年出版发行，被国内许多高校采用作为本科生数据库课程的教材或主要教学参考书。

现在我们又基于《数据库系统概念》第6版进行了改编工作，希望它能够成为一本效果更好、更实用的本科生数据库课程的教材。

本书的前9章是最基本的内容，讲述数据库系统的基本概念，包括对数据库系统的性质和目标的综述，对关系数据模型和关系语言的介绍，对数据库设计过程、关系数据库理论以及数据库应用设计和开发的详细讨论。第10至12章介绍了数据库系统实现的核心技术，包括数据存储管理、查询处理和事务管理。第13至16章是高级话题，介绍了数据仓库和数据挖掘，新型的数据库系统——基于对象的数据库和XML数据库，以及与高级应用开发相关的性能调整、性能基准程序、

标准化等内容。

本书可作为大学本科数据库概论课程的教材或主要参考资料，教师可以选择重点讲授前 12 章，并介绍第 13 至 16 章中的部分内容。

限于改编者水平，改编中疏漏和错误在所难免，欢迎批评指正。

杨冬青

2012 年 10 月于北京大学

Preface

Database management has evolved from a specialized computer application to a central component of a modern computing environment, and, as a result, knowledge about database systems has become an essential part of an education in computer science.

This book is adapted from the sixth edition of Database System Concepts, for a first course in databases at the junior or senior undergraduate level. In this text, we present the fundamental concepts of database management. These concepts include aspects of database design, database languages, and database-system implementation. In addition to basic material for a first course, the text contains advanced material that can be used for course supplements, or as introductory material for an advanced course.

We assume only a familiarity with basic data structures, computer organization, and a high-level programming language such as Java, C, or Pascal. We present concepts as intuitive descriptions, many of which are based on our running example of a university. Important theoretical results are covered, but formal proofs are omitted. In place of proofs, figures and examples are used to suggest why a result is true. Formal descriptions and proofs of theoretical results may be found in research papers and advanced texts that are referenced in the bibliographical notes.

The fundamental concepts and algorithms covered in the book are often based on those used in existing commercial or experimental database systems. Our aim is to present these concepts and algorithms in a general setting that is not tied to one particular database system.

In this adapted text of the sixth edition of *Database System Concepts*, we have retained essential contents of *Database System Concepts*, condensed or deleted some of the advanced materials. The purpose is to make this book more suitable for a first course in databases at the undergraduate level. Now we describe organization of this book in brief.

The text is organized in five major parts.

- Overview (Chapter 1). Chapter 1 provides a general overview of the nature and purpose of database systems. We explain how the concept of a database

system has developed, what the common features of database systems are, what a database system does for the user, and how a database system interfaces with operating systems. We also introduce an example database application: a university organization consisting of multiple departments, instructors, students, and courses. This application is used as a running example throughout the book. This chapter is motivational, historical, and explanatory in nature.

- Part 1: Relational Databases (Chapters 2 through 6). Chapter 2 introduces the relational model of data, covering basic concepts such as the structure of relational databases, database schemas, keys, schema diagrams, relational query languages, and relational operations. Chapters 3, 4, and 5 focus on the most influential of the user-oriented relational languages: SQL. Chapter 6 covers the formal relational query languages: relational algebra, tuple relational calculus, and domain relational calculus.

  The chapters in this part describe data manipulation: queries, updates, insertions, and deletions, assuming a schema design has been provided. Schema design issues are deferred to Part 2.
- Part 2: Database Design (Chapters 7 through 9). Chapter 7 provides an overview of the database-design process, with major emphasis on database design using the entity-relationship data model. The entity-relationship data model provides a high-level view of the issues in database design, and of the problems that we encounter in capturing the semantics of realistic applications within the constraints of a data model. UML class-diagram notation is also covered in this chapter.

  Chapter 8 introduces the theory of relational database design. The theory of functional dependencies and normalization is covered, with emphasis on the motivation and intuitive understanding of each normal form. This chapter begins with an overview of relational design and relies on an intuitive understanding of logical implication of functional dependencies. This allows the concept of normalization to be introduced prior to full coverage of functional-dependency theory, which is presented later in the chapter. Instructors may choose to use only this initial coverage in Sections 8.1 through 8.3 without loss of continuity. Instructors covering the entire chapter will benefit from students having a good understanding of normalization concepts to motivate some of the challenging concepts of

functional–dependency theory.

Chapter 9 covers application design and development. This chapter emphasizes the construction of database applications with Web-based interfaces. In addition, the chapter covers application security.

- Part 3: Data Storage, Querying and Transaction Management (Chapters 10 through 12). Chapter 10 introduces physical storage devices briefly; describes how to map records to files, and then to bits on disks; and describes several types of indices used in database systems. Chapter 11 describes how queries are processed, presents algorithms for implementing individual operations, and then describes the process of query optimization. Chapter 12 describes the fundamentals of a transaction-processing system, including atomicity, consistency, isolation, and durability; describes several concurrency-control techniques that help implement the isolation property; and describes the recovery management component of a database, which implements the atomicity and durability properties.
- Part 4: Advanced Topics (Chapters 13 through 16). Chapter 13 introduces the concepts and methodology of data warehousing and data mining. Chapter 14 covers object based databases, describes the object-relational data model, and this chapter also describes database access from object-oriented programming languages. Chapter 15 covers the XML standard for data representation, which is seeing increasing use in the exchange and storage of complex data, this chapter also describes query languages for XML. Chapter 16 covers advanced issues in application development, including performance tuning, performance benchmarks, database-application testing, and standardization.

We keeps the same as the sixth edition of *Database System Concepts*, the exercises are divided into two sets: practice exercises and exercises. The solutions for the practice exercises are publicly available on the Web site of the book. Students are encouraged to solve the practice exercises on their own, and later use the solutions on the Web site to check their own solutions. Solutions to the other exercises are available only to instructors (see "Instructor's Note" below, for information on how to get the solutions).

Instructor's Note

The book contains both basic and advanced material, which might not be

covered in a single semester.

The first 12 chapters are basic materials. For an introductory course, the instructor can choose to teach materials in Chapter 1 to Chapters 12, and introduce some materials in Chapter 13 to Chapter 16.

Web Site and Teaching Supplements

A Web site for the book is available at the URL: http://www.db-book.com. The Web site contains:

- Slides covering all the chapters of the book.
- Answers to the practice exercises.
- The five appendices.
- An up-to-date errata list.
- Laboratory material, including SQL DDL and sample data for the university schema and other relations used in exercises, and instructions for setting up and using various database systems and tools.

The following additional material is available only to faculty:

- An instructor manual containing solutions to all exercises in the book.
- A question bank containing extra exercises.

For more information about how to get a copy of the instructor manual and the question bank, please send electronic mail to customer.service@mcgraw-hill.com. In the United States, you may call 800-338-3987. The McGraw-Hill Web site for this book is http://www.mhhe.com/silberschatz.

Contents

Chapter 5 Advanced SQL

Chapter 6 Formal Relational Query Languages

PART TWO ■ DATABASE DESIGN

Chapter 7 Database Design and the E-R Model

Chapter 8 Relational Database Design

Chapter 9 Application Design and Development

PART THREE ■ DATA STORAGE, QUERYING, AND TRANSACTION MANAGEMENT

Chapter 10 Data Storage and Data Access

Chapter 11 Query Processing and Query Optimization

Chapter 12 Transaction Management

PART FOUR ■ ADVANCED TOPICS

Chapter 13 Data Warehousing and Mining

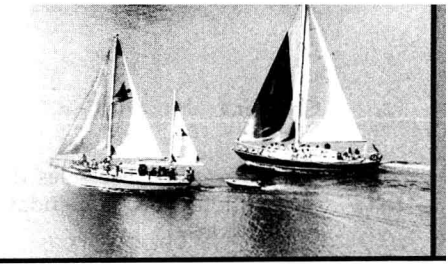
Chapter 14 Object-Based Databases

Chapter 15 XML

Chapter 16 Advanced Application Development

Bibliography 729

CHAPTER 1



Introduction

A database system (DBS) is a collection of interrelated data and a set of programs to access those data. The collection of data, usually referred to as the database, contains information relevant to an enterprise. The primary goal of a DBS is to provide a way to store and retrieve database information that is both *convenient* and *efficient*.

Database systems are designed to manage large bodies of information. Management of data involves both defining structures for storage of information and providing mechanisms for the manipulation of information. In addition, the database system must ensure the safety of the information stored, despite system crashes or attempts at unauthorized access. If data are to be shared among several users, the system must avoid possible anomalous results.

Because information is so important in most organizations, computer scientists have developed a large body of concepts and techniques for managing data. These concepts and techniques form the focus of this book. This chapter briefly introduces the principles of database systems.

1.1 Database-System Applications

Databases are widely used. Here are some representative applications:

- *Enterprise Information*
  - *Sales*: For customer, product, and purchase information.
  - *Accounting*: For payments, receipts, account balances, assets and other accounting information.
  - *Human resources*: For information about employees, salaries, payroll taxes, and benefits, and for generation of paychecks.
  - *Manufacturing*: For management of the supply chain and for tracking production of items in factories, inventories of items in warehouses and stores, and orders for items.

- *Online retailers*: For sales data noted above plus online order tracking, generation of recommendation lists, and maintenance of online product evaluations.

- *Banking and Finance*
  - *Banking*: For customer information, accounts, loans, and banking transactions.
  - *Credit card transactions*: For purchases on credit cards and generation of monthly statements.
  - *Finance*: For storing information about holdings, sales, and purchases of financial instruments such as stocks and bonds; also for storing real-time market data to enable online trading by customers and automated trading by the firm.

- *Universities*: For student information, course registrations, and grades (in addition to standard enterprise information such as human resources and accounting).

- *Airlines*: For reservations and schedule information. Airlines were among the first to use databases in a geographically distributed manner.

- *Telecommunication*: For keeping records of calls made, generating monthly bills, maintaining balances on prepaid calling cards, and storing information about the communication networks.

As the list illustrates, databases form an essential part of every enterprise today, storing not only types of information that are common to most enterprises, but also information that is specific to the category of the enterprise.

Over the course of the last four decades of the twentieth century, use of databases grew in all enterprises. In the early days, very few people interacted directly with database systems, although without realizing it, they interacted with databases indirectly—through printed reports such as credit card statements, or through agents such as bank tellers and airline reservation agents. Then automated teller machines came along and let users interact directly with databases. Phone interfaces to computers (interactive voice-response systems) also allowed users to deal directly with databases—a caller could dial a number, and press phone keys to enter information or to select alternative options, to find flight arrival/departure times, for example, or to register for courses in a university.

The Internet revolution of the late 1990s sharply increased direct user access to databases. Organizations converted many of their phone interfaces to databases into Web interfaces, and made a variety of services and information available online. For instance, when you access an online bookstore and browse a book or music collection, you are accessing data stored in a database. When you enter an order online, your order is stored in a database. When you access a bank Web site and retrieve your bank balance and transaction information, the information is retrieved from the bank's database system. When you access a Web site, informa-