

# 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. In this text, we present the fundamental concepts of database management. These concepts include aspects of database design, database languages, and database-system implementation.

This text is intended for a first course in databases at the junior or senior undergraduate, or first-year graduate, level. 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 bank enterprise. 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. Details of particular database systems are discussed in Part 9, “Case Studies.”

In this, the fifth edition of *Database System Concepts*, we have retained the overall style of the prior editions while evolving the content and organization to reflect the changes that are occurring in the way databases are designed, managed, and used. We have also taken into account trends in the teaching of database concepts and made adaptations to facilitate these trends where appropriate. Before we describe the content of the book in detail, we highlight some of the features of the fifth edition.

- **Earlier coverage of SQL.** Many instructors use SQL as a key component of term projects (see our Web site, [www.db-book.com](http://www.db-book.com), for sample projects). In order to give students ample time for the projects, particularly for universities and colleges on the quarter system, it is essential to teach SQL as early as possible. With this in mind, we have undertaken several changes in organization:
  1. Deferring the presentation of the entity-relationship model to Part 2, entitled “Database Design.”
  2. Streamlining the introduction of the relational model by deferring coverage of the relational calculus to Chapter 5, while retaining coverage of the relational algebra in Chapter 2.
  3. Devoting two early chapters to SQL. Chapter 3 covers basic SQL features including data definition and manipulation. Chapter 4 covers more advanced features, including integrity constraints, dynamic SQL, and procedural constructs. New material in this chapter includes expanded coverage of JDBC, procedural constructs in SQL, recursion in SQL, and new features from SQL:2003. The chapter also includes a short overview of authorization; detailed coverage of authorization is deferred to Chapter 8.

These changes allow students to begin writing SQL queries early in the course, and gain familiarity with the use of database systems. This also allows students to develop an intuition about database design that facilitates the teaching of design methodology in Part 2 of the text. We have found that students appreciate database-design issues better with this organization.

- **A new part (Part 2) that is devoted to database design.** Part 2 of the text contains three chapters devoted to the design of databases and database applications. We include here a chapter (Chapter 6) on the entity-relationship model that includes all of the material from the corresponding chapter of the fourth edition (Chapter 2), plus several significant updates. We also present in Chapter 6 a brief overview of the process of database design. Instructors who prefer to begin their course with the E-R model can begin with this chapter without loss of continuity, as we have strived to avoid dependencies on any prior chapter other than Chapter 1.

Chapter 7, on relational design, presents the material covered in Chapter 7 of the fourth edition, but does so in a new, more readable style. Design concepts from the E-R model are used to build an intuitive overview of relational design issues, in advance of the presentation of the formal approach to design using functional and multivalued dependencies and algorithmic normalization. This chapter also includes a new section on temporal issues in database design.

Part 2 concludes with a new chapter, Chapter 8, that describes the design and development of database applications, including Web applications, servlets, JSP, triggers, and security issues. In keeping with the increased need to secure software from attacks, coverage of security has been significantly expanded from the fourth edition.

- **Thoroughly revised and updated coverage of object-based databases and XML.** Part 3 includes a heavily revised chapter on object-based databases that emphasizes SQL object-relational features, replacing the separate chapters on object-oriented and object-relational databases from the fourth edition. Some of the introductory material on object-orientation which students are familiar with from earlier courses has been removed, as have syntactic details of the now defunct ODMG standard. However, important concepts underlying object-oriented databases have been retained, including new material on the JDO standard for adding persistence to Java.

Part 3 includes also a chapter on the design and querying of XML data, which is significantly revised from the corresponding chapter in the fourth edition. It includes enhanced coverage of XML Schema and XQuery, coverage of the SQL/XML standard, and more examples of XML applications including Web services.

- **Reorganized material on data mining and information retrieval.** Data mining and online analytic processing are now centrally important uses of databases—no longer only “advanced topics.” We have, therefore, moved our coverage of these topics into a new part, Part 6, containing a chapter on data mining and analysis along with a chapter on information retrieval.
- **New case study covering PostgreSQL.** PostgreSQL is an open-source database system that has gained enormous popularity in the past few years. In addition to being a platform on which to build database applications, the source code can be studied and extended in courses that emphasize database internals. A case study of PostgreSQL is therefore added to Part 9, where it joins three case studies that appeared in the fourth edition (Oracle, IBM DB2, and Microsoft SQL Server). The latter three case studies have been updated to reflect the latest versions of the respective software.

The coverage of topics not listed above, including transaction processing (concurrency and recovery), storage structures, query processing, and distributed and parallel databases are all updated from their fourth-edition counterparts, though their overall organization is relatively unchanged. The coverage of QBE in Chapter 5 has been revised, removing syntactic details of aggregation and updates that do not correspond to any actual implementation, while retaining the key concepts behind QBE.

## Organization

The text is organized in nine major parts, plus three appendices.

- **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 banking enterprise consisting of multiple bank branches. This example

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 5). Chapter 2 introduces the relational model of data, covering basic concepts as well as the relational algebra. The chapter also provides a brief introduction to integrity constraints. Chapters 3 and 4 focus on the most influential of the user-oriented relational languages: SQL. While Chapter 3 provides a basic introduction to SQL, Chapter 4 describes more advanced features of SQL, including how to interface between a programming language and a database supporting SQL. Chapter 5 covers other relational languages, including the relational calculus, QBE, and Datalog.

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 6 through 8). Chapter 6 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 7 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 7.1 through 7.3 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 8 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: Object-Based Databases and XML** (Chapters 9 and 10). Chapter 9 covers object-based databases. The chapter describes the object-relational data model, which extends the relational data model to support complex data types, type inheritance, and object identity. The chapter also describes database access from object-oriented programming languages.

Chapter 10 covers the XML standard for data representation, which is seeing increasing use in the exchange and storage of complex data. The chapter also describes query languages for XML.

- **Part 4: Data Storage and Querying** (Chapters 11 through 14). Chapter 11 deals with disk, file, and file-system structure. A variety of data-access techniques are presented in Chapter 12, including hashing and B<sup>+</sup>-tree indices. Chapters 13 and 14 address query-evaluation algorithms and query optimization. These chapters provide an understanding of the internals of the storage and retrieval components of a database.
- **Part 5: Transaction Management** (Chapters 15 through 17). Chapter 15 focuses on the fundamentals of a transaction-processing system, including transaction atomicity, consistency, isolation, and durability, as well as the notion of serializability.

Chapter 16 focuses on concurrency control and presents several techniques for ensuring serializability, including locking, timestamping, and optimistic (validation) techniques. The chapter also covers deadlock issues.

Chapter 17 covers the primary techniques for ensuring correct transaction execution despite system crashes and disk failures. These techniques include logs, checkpoints, and database dumps.
- **Part 6: Data Mining and Information Retrieval** (Chapters 18 and 19). Chapter 18 introduces the concept of a data warehouse and explains data mining and online analytical processing (OLAP), including SQL:1999 support for OLAP and data warehousing. Chapter 19 describes information-retrieval techniques for querying textual data, including hyperlink-based techniques used in Web search engines.

Part 6 uses the modeling and language concepts from Parts 1 and 2, but does not depend on Parts 3, 4, or 5. It can therefore be incorporated easily into a course that focuses on SQL and on database design.
- **Part 7: Database-System Architecture** (Chapters 20 through 22). Chapter 20 covers computer-system architecture, and describes the influence of the underlying computer system on the database system. We discuss centralized systems, client-server systems, parallel and distributed architectures, and network types in this chapter.

Chapter 21, on parallel databases, explores a variety of parallelization techniques, including I/O parallelism, interquery and intraquery parallelism, and interoperation and intraoperation parallelism. The chapter also describes parallel-system design.

Chapter 22 covers distributed database systems, revisiting the issues of database design, transaction management, and query evaluation and optimization, in the context of distributed databases. The chapter also covers issues of system availability during failures and describes the LDAP directory system.
- **Part 8: Other Topics** (Chapters 23 through 25). Chapter 23 covers performance benchmarks, performance tuning, standardization and application migration from legacy systems.

Chapter 24 covers advanced data types and new applications, including temporal data, spatial and geographic data, multimedia data, and issues in the management of mobile and personal databases.

Finally, Chapter 25 deals with advanced transaction processing. Topics covered include transaction-processing monitors, transactional workflows, electronic commerce, high-performance transaction systems, real-time transaction systems, long duration transactions, and transaction management in multi-database systems.

- **Part 9: Case Studies** (Chapters 26 through 29). In this part we present case studies of four leading database systems, including PostgreSQL, Oracle, IBM DB2, and Microsoft SQL Server. These chapters outline unique features of each of these systems, and describe their internal structure. They provide a wealth of interesting information about the respective products, and help you see how the various implementation techniques described in earlier parts are used in real systems. They also cover several interesting practical aspects in the design of real systems.
- **Online Appendices.** Although most new database applications use either the relational model or the object-relational model, the network and hierarchical data models are still in use in some legacy applications. For the benefit of readers who wish to learn about these data models, we provide appendices describing the network and hierarchical data models, in Appendices A and B respectively; the appendices are available only online (<http://www.db-book.com>).

Appendix C describes advanced relational database design, including the theory of multivalued dependencies, join dependencies, and the project-join and domain-key normal forms. This appendix is for the benefit of individuals who wish to study the theory of relational database design in more detail, and instructors who wish to do so in their courses. This appendix, too, is available only online, on the Web page of the book.

## The Fifth Edition

The production of this fifth edition has been guided by the many comments and suggestions we received concerning the earlier editions, by our own observations while teaching at Yale University, Lehigh University, and IIT Bombay, and by our analysis of the directions in which database technology is evolving.

Our basic procedure was to rewrite the material in each chapter, bringing the older material up-to-date, adding discussions on recent developments in database technology, and improving descriptions of topics that students found difficult to understand. As in the fourth edition, each chapter has a list of review terms that can help readers review key topics covered in the chapter. Most chapters also have a tools section at the end of the chapter that provides information on software tools related to the topic of the chapter. We have also added new exercises and updated references.

In the fifth edition, we have divided the exercises into two sets: **practice exercises** and **exercises**. The solutions for the practice exercises are publicly available on the Web page of the book. Students are encouraged to solve the practice exercises on their own, and later use the solutions on the Web page 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. We have marked several sections as advanced, using the symbol \*\*. These sections may be omitted if so desired, without a loss of continuity. Exercises that are difficult (and can be omitted) are also marked using the symbol “\*\*.”

It is possible to design courses by using various subsets of the chapters. We outline some of the possibilities here:

- Sections of Chapter 4 from Section 4.6 onward may be omitted from an introductory course.
- Chapter 5 can be omitted if students will not be using relational calculus, QBE or Datalog as part of the course.
- Chapters 9 (Object-Based Databases), 10 (XML), and 14 (Query Optimization) can be omitted from an introductory course.
- Both our coverage of transaction processing (Chapters 15 through 17) and our coverage of database-system architecture (Chapters 20 through 22) consist of an overview chapter (Chapters 15 and 20, respectively), followed by chapters with details. You might choose to use Chapters 15 and 20, while omitting Chapters 16, 17, 21, and 22, if you defer these latter chapters to an advanced course.
- Chapters 18 and 19, covering data mining and information retrieval, can be used as self-study material or omitted from an introductory course.
- Chapters 23 through 25 are suitable for an advanced course or for self-study by students.
- The case-study Chapters 26 through 29 are suitable for self-study by students.

Model course syllabi, based on the text, can be found on the Web home page of the book (see the following section).

## Web Page and Teaching Supplements

A Web home page for the book is available at the URL:

<http://www.db-book.com>

The Web page contains:

- Slides covering all the chapters of the book
- Answers to the practice exercises
- Laboratory material
- The three appendices
- An up-to-date errata list
- Supplementary material contributed by users of the book

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](mailto:customer.service@mcgraw-hill.com). In the United States, you may call 800-338-3987. The McGraw-Hill Web page for this book is

<http://www.mhhe.com/silberschatz>

## Contacting Us and Other Users

We have endeavored to eliminate typos, bugs, and the like from the text. But, as in new releases of software, bugs probably remain; an up-to-date errata list is accessible from the book's home page. We would appreciate it if you would notify us of any errors or omissions in the book that are not on the current list of errata.

We would be glad to receive suggestions on improvements to the books. We also welcome any contributions to the book Web page that could be of use to other readers, such as programming exercises, project suggestions, online labs and tutorials, and teaching tips.

Email should be addressed to [db-book@cs.yale.edu](mailto:db-book@cs.yale.edu). Any other correspondence should be sent to Avi Silberschatz, Department of Computer Science, Yale University, 51 Prospect Street, P.O. Box 208285, New Haven, CT 06520-8285 USA.

We also provide a mailing list through which users of our book can communicate among themselves and with us, and receive updates on the book and other related information. The list is moderated, so you will not receive junk mail on the list. Please follow the mailing list link from the book's home page to subscribe to the mailing list.

## Acknowledgments

This edition has benefited from the many useful comments provided to us by the numerous students who have used the prior four editions. In addition, many people



have written or spoken to us about the book, and have offered suggestions and comments. Although we cannot mention all these people here, we especially thank the following:

- Hani Abu-Salem, DePaul University; Jamel R. Alsabbagh, Grand Valley State University; Ramzi Bualuan, Notre Dame University; Zhengxin Chen, University of Nebraska at Omaha; Jan Chomick, SUNY Buffalo University; Qin Ding, Penn State University at Harrisburg; Frantisek Franek, McMaster University; Shashi K. Gadia, Iowa State University; William Hankley, Kansas State University; Randy M. Kaplan, Drexel University; Mark Llewellyn, University of Central Florida; Marty Maskarinec, Western Illinois University; Yiu-Kai Dennis Ng, Brigham Young University; Sunil Prabhakar, Purdue University; Stewart Shen, Old Dominion University; Anita Whitehall, Foothill College; Christopher Wilson, University of Oregon; Weining Zhang, University of Texas at San Antonio; who served as reviewers of the book and whose comments helped us greatly in formulating this fifth edition.
- Anastassia Ailamaki, Sailesh Krishnamurthy, Spiros Papadimitriou, and Bianca Schroeder (Carnegie Mellon University) for writing the appendix describing the PostgreSQL database system.
- Hakan Jakobsson (Oracle), for the appendix on the Oracle database system,
- Sriram Padmanabhan (IBM), for the appendix describing the IBM DB2 database system.
- Sameet Agarwal, José A. Blakeley, Thierry D’Hers, Gerald Hinson, Dirk Myers, Vaqar Pirzada, Bill Ramos, Balaji Rathakrishnan, Michael Rys, Florian Waas and Michael Zwilling (all of Microsoft) for the appendix on the Microsoft SQL Server database system. José Blakeley also for coordinating and editing Chapter 29, and César Galindo-Legaria, Goetz Graefe, Kalen Delaney, and Thomas Casey (all of Microsoft) for their contributions to the previous edition of the Microsoft SQL Server chapter.
- Chen Li and Sharad Mehrotra for providing material on JDBC and security that helped update and extend Chapter 8.
- Valentin Dinu, Goetz Graefe, Bruce Hillyer, Chad Hogg, Nahid Rahman, Patrick Schmid, Jeff Storey, Prem Thomas, Liu Zhenming, and particularly N. L. Sarda for their feedback, which helped us prepare the fifth edition.
- Rami Khouri, Nahid Rahman, and Michael Rys for feedback on draft versions of chapters from the fifth edition.
- Raj Ashar, Janek Bogucki, Gavin M. Bierman, Christian Breimann, Tom Chappell, Y. C. Chin, Laurens Damen, Prasanna Dhandapani, Arvind Hulgeri, Zheng Jiaping, Graham J. L. Kemp, Hae Choon Lee, Sang-Won Lee, Thanh-Duy Nguyen, D. B. Phatak, Juan Altmayer Pizzorno, Rajarshi Rakshit, Greg Riccardi, N. L. Sarda, Max Smolens, Nikhil Sethi, and Tim Wahls for pointing out errors in the fourth edition.

- Marilyn Turnamian, whose excellent secretarial assistance was essential for timely completion of this fifth edition.

The publisher was Betsy Jones. The sponsoring editor was Kelly Lowery. The developmental editor was Melinda D. Bilecki. The project manager was Peggy Selle. The executive marketing manager was Michael Weitz. The marketing manager was Dawn Bercier. The cover illustrator and cover designer was JoAnne Schopler. The freelance copyeditor was George Watson. The freelance proofreader was Judy Gantenbein. The designer was Laurie Janssen. The freelance indexer was Tobiah Waldron.

This edition is based on the four previous editions, so we thank once again the many people who helped us with the first four editions, including R. B. Abhyankar, Don Batory, Phil Bernhard, Haran Boral, Paul Bourgeois, Phil Bohannon, Robert Brazile, Yuri Breitbart, Michael Carey, Soumen Chakrabarti, J. Edwards, Christos Faloutsos, Homma Farian, Alan Fekete, Shashi Gadia, Jim Gray, Le Gruenwald, Eitan M. Gurari, Ron Hitchens, Yannis Ioannidis, Hyoung-Joo Kim, Won Kim, Henry Korth (father of Henry F.), Carol Kroll, Gary Lindstrom, Irwin Levinstein, Ling Liu, Dave Maier, Keith Marzullo, Fletcher Mattox, Sharad Mehrotra, Jim Melton, Alberto Mendelzon, Hector Garcia-Molina, Ami Motro, Bhagirath Narahari, Anil Nigam, Cyril Orji, Meral Ozsoyoglu, Bruce Porter, Jim Peterson, K. V. Raghavan, Krithi Ramamritham, Mike Reiter, Odinaldo Rodriguez, Mark Roth, Marek Rusinkiewicz, Sunita Sarawagi, N. L. Sarda, S. Seshadri, Shashi Shekhar, Amit Sheth, Nandit Soparkar, Greg Speegle, Dilys Thomas, and Marianne Winslett.

Marilyn Turnamian and Nandprasad Joshi provided secretarial assistance for the fourth edition, and Marilyn also prepared an early draft of the cover design for the fourth edition. Lyn Dupré copyedited the third edition and Sara Strandtman edited the text of the third edition. Nilesh Dalvi, Sumit Sanghai, Gaurav Bhalotia, Arvind Hulgeri, K. V. Raghavan, Prateek Kapadia, Sara Strandtman, Greg Speegle, and Dawn Bezziner helped to prepare the instructor's manual for earlier editions. The new cover is an evolution of the covers of the first four editions. The idea of using ships as part of the cover concept was originally suggested to us by Bruce Stephan.

Finally, Sudarshan would like to acknowledge his wife, Sita, for her love and support and son Madhur for his love. Hank would like to acknowledge his wife, Joan, and his children, Abby and Joe, for their love and understanding. Avi would like to acknowledge Valerie for her love, patience, and support during the revision of this book.

A. S.  
H. F. K.  
S. S.