Modern Database Management

THIRTEENTH EDITION

Jeffrey A. Hoffer

V. Ramesh

Heikki Topi

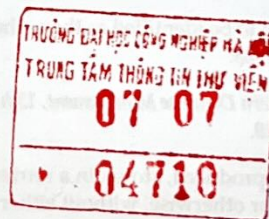**THIRTEENTH EDITION**

GLOBAL EDITION

MODERN DATABASE MANAGEMENT

Jeffrey A. Hoffer
University of Dayton

V. Ramesh
Indiana University

Heikki Topi
Bentley University

Pearson

BRIEF CONTENTS

PREFACE

This text is designed for introductory courses in database management. Such a course is usually required as part of an information systems curriculum in business schools, computer technology programs, and applied computer science departments. The Association for Information Systems (AIS), the Association for Computing Machinery (ACM), and the International Federation of Information Processing Societies (IFIPS) curriculum guidelines (e.g., IS 2010 and MSIS 2016) all outline this type of database management course or the competencies a student completing the course is expected to have. Previous editions of this text have been used successfully for more than 35 years at both the undergraduate and graduate levels as well as in management and professional development programs.

WHAT'S NEW IN THIS EDITION?

This 13th edition of *Modern Database Management* updates and expands materials in areas undergoing rapid change as a result of improved managerial practices, database design tools and methodologies, and database technology. Later, we detail changes to each chapter. The themes of this 13th edition reflect the major trends in the information systems field and the skills required of modern information systems graduates. The most important changes are as follows:

- The book has been restructured in several important ways. Chapter 7 on databases in applications now also includes segments on transaction integrity, designing multi-user solutions, and application level security, bringing these important perspectives together with their context. The revised chapter on physical database design and database infrastructure (new Chapter 8) includes also coverage of database security, backup and recovery, cloud-based database solutions, and other essential database infrastructure topics. This new comprehensive structure on physical design and infrastructure is now placed after the SQL chapters. The new version of Chapter 9 integrates material on data warehousing and data integrity in a conceptually natural pairing. Recognizing the way in which analytics capabilities rely on all types of data management solutions, Chapter 11, on analytics and implications, is now separate from Chapter 10, on big data. Finally, Chapter 12 brings together data and database administration with data quality, emphasizing the essential connections between the three.
- The part structure of the book has been redesigned to be fully aligned with the new chapter structure.
- We have introduced a new overarching framework (Figure 1-5), which gives our readers a clearer overview of structure of the book and its core topic areas. The framework communicates clearly the increasing importance of informational systems (divided into Analytics–Data Warehousing and Analytics–Big Data) in addition to this book's traditional strength of transactional systems.
- Given the continued and still increasing interest in big data and analytics, we have continued to expand content in this area. The book has now separate chapters on big data technologies (Chapter 10) and analytics (Chapter 11). In addition to general coverage of NoSQL and Hadoop technologies, Chapter 10 provides also detailed examples of MongoDB, Pig, and Hive. Chapter 11 includes extended coverage of R, Python, and Apache Spark—all essential technologies for analytics professionals that allow a link between analytics and data management architectures.
- We emphasize the increasing importance of cloud-based database solutions, mobile technologies, and agile development throughout the book.
- Chapter 1 now better recognizes the broad range of enterprise level applications data management solutions enable and support, including enterprise systems, data warehouses, and data lakes.

- Chapter 7 on databases in applications now includes an extensive example demonstrating the use of Python in the context of database-driven applications.
- The instructor's manual will have more material to support the case Forondo Artist Management Excellence that was introduced in the 12th edition.

In addition to the new topics covered, specific improvements to the textbook have been made in the following areas:

- Every chapter went through significant edits to streamline coverage to ensure relevance with current technologies and eliminate redundancies.
- The entire book has been edited so that its language clearly reflects its focus on the readers as learners instead of authors as teachers
- End-of-chapter material (review questions, problems and exercises, and/or field exercises) in every chapter has been revised with new and modified questions and exercises.
- We continued to update the figures in several chapters to reflect the changing landscape of technologies that are being used in modern organizations.
- The Web Resources section in each chapter was updated to ensure that students have information on the latest database trends and expanded background details on important topics covered in the text.
- The book continues to be available through VitalSource, an innovative e-book delivery system, and as an electronic book in the Kindle format.

Also, we continue to provide on the student Companion Web site several custom-developed short videos that address key concepts and skills from different sections of the book. These videos, produced by the textbook authors, help students learn difficult material by using both the printed text and a mini-lecture or tutorial. Videos have been developed to support Chapters 1 (introduction to database), 2 and 3 (conceptual data modeling), 4 (normalization), and 6 and 7 (SQL). Look for special icons on the opening page of these chapters to call attention to these videos, and go to **www.pearsonglobaleditions.com** to find these videos.

FOR THOSE NEW TO *MODERN DATABASE MANAGEMENT*

Modern Database Management has been a leading text since its first edition in 1983. In spite of this market leadership position, some instructors have used other good database management texts. Why might you want to switch at this time? There are several good reasons:

- One of our goals, in every edition, has been to lead other books in coverage of the latest principles, concepts, and technologies. See what we have added for the 13th edition in "What's New in This Edition?" In the past, we have led in coverage of object-oriented data modeling and UML, Internet databases, data warehousing, and the use of CASE tools in support of data modeling. For the 13th edition, we continue this tradition by continuing to expand and improve coverage of big data and analytics, focusing on what every database student needs to understand about these topics.
- While remaining current, this text focuses on what leading practitioners say is most important for database developers. We work with many practitioners, including the professionals of the Data Management Association (DAMA) and The Data Warehousing Institute (TDWI), leading consultants, technology leaders, and authors of articles in the most widely read professional publications. We draw on these experts to ensure that what the book includes is important and covers not only important entry-level knowledge and skills but also those fundamentals and mind-sets that lead to long-term career success.
- In the 13th edition of this highly successful book, material is presented in a way that has been viewed as very accessible to students. Our methods have been refined through continuous market feedback for more than 35 years as well as through our own teaching. Overall, the pedagogy of the book is sound, and we believe that the new framework that we introduced in Chapter 1 will further strengthen our students'

understanding of the big picture of data management. We use many illustrations that help make important concepts and techniques clear. We use the most modern notations. The organization of the book is flexible, so you can use chapters in whatever sequence makes sense for your students. We supplement the book with data sets to facilitate hands-on, practical learning and with new media resources to make some of the more challenging topics more engaging.

- Our text can accommodate structural flexibility. For example, you may have particular interest in introducing SQL early in your course. Our text makes this possible. First, we cover SQL in depth, devoting two full chapters to this core technology of the database field. Second, we include many SQL examples in early chapters. Third, many instructors have successfully used the two SQL chapters early in their course. Although logically appearing in the life cycle of systems development as Chapters 5 and 6, part of the implementation section of the text, many instructors have used these chapters immediately after Chapter 1 or in parallel with other early chapters. Finally, we use SQL throughout the book, for example, to illustrate Web application connections to relational databases in Chapter 7 and online analytical processing in Chapter 11.
- We have the latest in supplements and Web site support for the text. See the supplement package for details on all the resources available to you and your students.
- This text is written to be part of a modern information systems curriculum with a strong business systems development focus. Topics are included and addressed so as to reinforce principles from other typical courses, such as systems analysis and design, networking, Web site design and development, MIS principles, and application development. Emphasis is on the development of the database component of modern information systems and on the management of the data resource. Thus, the text is practical, supports projects and other hands-on class activities, and encourages linking database concepts to concepts being learned throughout the curriculum the student is taking.

SUMMARY OF ENHANCEMENTS TO EACH CHAPTER

The following sections present a chapter-by-chapter description of the major changes in this edition. Each chapter description presents a statement of the purpose of that chapter, followed by a description of the changes and revisions that have been made for the 13th edition. Each paragraph concludes with a description of the strengths that have been retained from prior editions.

PART I: THE CONTEXT OF DATABASE MANAGEMENT

Chapter 1: The Database Environment and Development Process

This chapter discusses the role of databases in organizations and previews the major topics in the remainder of the text. The primary change to this chapter has been the introduction of a new integrated data management framework (Figure 1-5) and supporting text accompanying it. This framework recognizes the increasing importance of the *informational* systems in addition to the traditional focus of this book on *transactional* systems. After presenting a brief introduction to the basic terminology associated with storing and retrieving data, the chapter presents a well-organized comparison of traditional file processing systems and modern database technology. The chapter then introduces the core components of a database environment. It then goes on to explain the process of database development in the context of structured life cycle, prototyping, and agile methodologies. The chapter also discusses important issues in database development, including management of the diverse group of people involved in database development and frameworks for understanding database architectures and technologies (e.g., the three-schema architecture). Reviewers frequently note the compatibility of this chapter with what students learn in systems analysis and design classes. A brief history of the evolution of database technology, from pre-database files to modern object-relational technologies, is presented. The chapter also provides

an overview of the range of database applications that are currently in use within organizations—personal, multi-tier, and enterprise applications. The explanation of enterprise databases includes databases that are part of enterprise resource planning systems and data warehouses. The chapter concludes with a description of the process of developing a database in a fictitious company, Pine Valley Furniture. This description closely mirrors the steps in database development described earlier in the chapter. The first chapter provides an introduction to the FAME case, which then continues through the book until Chapter 8.

PART II: DATABASE ANALYSIS AND LOGICAL DESIGN

Chapter 2: Modeling Data in the Organization

This chapter presents a thorough introduction to conceptual data modeling with the entity-relationship (E-R) model. The chapter title emphasizes the reason for the E-R model: to unambiguously document the rules of the business that influence database design. Specific subsections explain in detail how to name and define elements of a data model, which are essential in developing an unambiguous E-R diagram. The chapter continues to proceed from simple to more complex examples, and it concludes with a comprehensive E-R diagram for the Pine Valley Furniture Company. In the 13th edition, we have provided six new problems and exercises; these new exercises present some more modern situations, such as Internet of Things applications for databases. A variety of other problems and exercises as well as review questions have been changed to emphasize important topics of the chapter. Appendix A provides information on different data modeling tools and notations.

Chapter 3: The Enhanced E-R Model

This chapter presents a discussion of several advanced E-R data model constructs, primarily supertype/subtype relationships. As in Chapter 2, problems and exercises have been revised, with three new exercises and several building on or extending the new exercises from Chapter 2. The third part of the new FAME case is presented in this chapter. The chapter continues to present thorough coverage of supertype/subtype relationships and includes a comprehensive example of an extended E-R data model for the Pine Valley Furniture Company.

Chapter 4: Logical Database Design and the Relational Model

This chapter describes the process of converting a conceptual data model to the relational data model, as well as how to merge new relations into an existing normalized database. It provides a conceptually sound and practically relevant introduction to normalization, emphasizing the importance of the use of functional dependencies and determinants as the basis for normalization. Concepts of normalization and normal forms are extended in Appendix B. The chapter features a discussion of the characteristics of foreign keys and introduces the important concept of a nonintelligent enterprise key. Enterprise keys (also called surrogate keys for data warehouses) are emphasized as some concepts of object-orientation have migrated into the relational technology world. New problems and exercises are included that draw upon the new problems and exercises from Chapters 2 and 3 for relational modeling and normalization. The chapter continues to emphasize the basic concepts of the relational data model and the role of the database designer in the logical design process.

PART III: DATABASE IMPLEMENTATION AND USE

Chapter 5: Introduction to SQL

This chapter (Chapter 6 in 12th edition) presents a thorough introduction to the SQL used by most DBMSs (SQL:1999) and introduces the changes that are included in the latest standards (SQL: 2011 and SQL:2016). This edition adds coverage of the new features of SQL:2016, including row pattern recognition, JSON support, and extended analytical

capabilities. The new edition also clarifies coverage of SQL data types and, overall, makes it easier to move from relational design in Chapter 4 directly to database implementation without the material on physical database design (now in Chapter 8). The coverage of SQL is extensive and divided between this chapter and Chapter 6. This chapter includes examples of SQL code, using mostly SQL:1999 and SQL:2016 syntax, as well as some Oracle 12c and Microsoft SQL Server syntax. Some unique features of MySQL are mentioned. In this edition, coverage of views has been moved to Chapter 6. Chapter 5 explains the SQL commands needed to create and maintain a database and to program single-table queries. Five review questions and 13 problems and exercises have been added to the chapter or modified extensively. The chapter continues to use the Pine Valley Furniture Company case to illustrate a wide variety of practical queries and query results.

Chapter 6: Advanced SQL

This chapter (Chapter 7 in 12th edition) continues the description of SQL, with a careful explanation of multiple-table queries, transaction integrity, data dictionaries, dynamic and materialized views, triggers and stored procedures (the differences between them are now more clearly explained), and embedding SQL in other programming language programs. All forms of the OUTER JOIN command are covered. Standard SQL (with an updated focus on SQL:2016) is also used. The revised version of the chapter includes now thorough coverage of views and the purposes for which they are used, including their role in enabling security and privacy solutions. This chapter illustrates how to store the results of a query in a derived table, the CAST command to convert data between different data types, and the CASE command for doing conditional processing in SQL. Emphasis continues on the set-processing style of SQL compared with the record processing of programming languages with which the student may be familiar. The section on routines has been revised to provide clarified, expanded, and more current coverage of this topic. The material of transaction integrity, has, however been moved to Chapter 7, where it most naturally belongs. The chapter continues to contain a clear explanation of subqueries and correlated subqueries, two of the most complex and powerful constructs in SQL. At the end, the chapter discusses material that is new to this chapter: data dictionary facilities (in practice, using SQL to understand the structure of the database) and recent extensions and enhancements to SQL. Chapter review material has been updated with 13 new problems and exercises and three new review questions.

Chapter 7: Databases in Applications

This chapter (Chapter 8 in 12th edition) provides a modern discussion of the concepts of client/server architecture and applications, middleware, and database access in contemporary database environments. The chapter has been structurally significantly modified to provide additional clarity, including the integration of material on a two-tiered architecture into the section on three-tiered architecture. In addition to a revised example of writing a Java web application, there is an entire new section—including an extensive and detailed example—on writing Web applications with Python, a widely used general purpose programming language that has become very popular in analytics. Sections on transaction integrity, concurrent access, and application level data security have been revised and moved to this chapter to provide additional conceptual clarity. Material on cloud computing has been moved to Chapter 8 on database infrastructure. Review questions and problems and exercises have been updated.

Chapter 8: Physical Database Design and Database Infrastructure

This chapter (Chapter 5 in the 12th edition) describes the steps that are essential in achieving an efficient database design, with a strong focus on those aspects of database design and implementation that are typically within the control of a database professional in a modern database environment. In addition, several new topics on database infrastructure have been integrated into this chapter to improve the structural clarity of the book, including data dictionaries and repositories, general database software security features, and database backup and recovery. A revised and extended section on cloud-based database infrastructure completes the chapter. Overall, the chapter emphasizes ways to

improve database performance, with references to specific techniques available in Oracle and other DBMSs to achieve this goal. The discussion of indexes includes descriptions of the types of indexes that are widely available in database technologies as techniques to improve query processing speed. Appendix C provides excellent background on fundamental data structures for programs of study that need coverage of this topic. The chapter continues to emphasize the physical design process and the goals of that process. Review questions and problems and exercises have been updated and extended based on the new structure and content of the chapter.

PART IV: ADVANCED DATABASE TOPICS

Chapter 9: Data Warehousing and Data Integration

This chapter describes the basic concepts of data warehousing, the reasons data warehousing is regarded as critical to competitive advantage in many organizations, and the database design activities and structures unique to data warehousing. The most important change of this chapter is the integration of material on data integration (formerly in Chapter 10 in the 12th edition) into it. This change strengthens the readers' ability to understand the essential role of data integration in data warehousing (particularly in ETL and other aspects of data preparation), and it clarifies the structure of the book. Topics covered in this chapter include alternative data warehouse architectures and the dimensional data model (or star schema) for data warehouses. In this edition, additional attention is given to cloud-based implementation of data warehouses. Throughout the chapter, several details have been updated to ensure technical correctness. Operational data store and independent, dependent, and logical data marts are defined. The chapter includes multiple new and revised review questions and problems and exercises.

Chapter 10: Big Data Technologies

This chapter incorporates big data infrastructure material from Chapter 11 in the 12th edition, significantly expanding it and making it more directly applicable with substantial detailed descriptive examples of MongoDB (the most popular NoSQL database) and Pig (scripting language and task automation environment for Hadoop) and Hive (an SQL-like declarative language for querying data stored in Hadoop). This new version of the material gives the students a much more practical, hands-on sense of the purposes for which these well-known tools can be used and how they can serve the goals of big data management. The chapter also includes several new problems and exercises based on these environments. Overall, the chapter helps the readers understand how big data technologies have expanded the possibilities for analytics-driven innovation through advanced informational systems that are pushing boundaries further in terms of volume, velocity, and variety of data while paying continuous attention to value and veracity of big data.

Chapter 11: Analytics and its Implications

Chapter 11 offers integrated coverage of analytics, including descriptive, predictive, and prescriptive analytics. It is based on material on analytics in the big data and analytics chapter in the 12th edition, expanding it with comprehensive new sections on R, Python, and Apache Spark and bringing in material on analytical functions in SQL. The discussion on analytics is linked not only to the coverage of big data but also the material on data warehousing in Chapter 9 and the general discussion on data management in Chapter 1 (as indicated in the new framework in Chapter 1). The chapter also covers approaches and technologies used by analytics professionals, such as on-line analytical processing, data visualization, business performance management and dashboards, data mining, and text mining. Finally, the chapter integrates the coverage of big data and analytics technologies to the individual, organizational, and societal implications of these capabilities. Review questions on the new material have been added.

Chapter 12: Data and Database Administration with Focus on Data Quality

This chapter presents a thorough discussion of the importance and roles of data and database administration and describes a number of the key issues that arise when these functions are performed. This chapter emphasizes the changing roles and approaches of data and database administration, with a renewed and strengthened emphasis on data quality. The chapter both discusses essential characteristics of high-quality data and the mechanisms that organizations need to put in place to enable data quality improvement. Data governance, data availability, and master data management are also covered. The chapter continues to emphasize the critical importance of data and database management in managing data as a corporate asset.

Chapter 13: Distributed Databases

This chapter—available on the book's Web site—reviews the role, technologies, and unique database design opportunities of distributed databases. The objectives and trade-offs for distributed databases, data replication alternatives, factors in selecting a data distribution strategy, and distributed database vendors and products are covered. This chapter provides thorough coverage of database concurrency access controls. Many reviewers have indicated that they are seldom able to cover this chapter in an introductory course, but having the material available is critical for advanced students or special topics.

Chapter 14: Object-Oriented Data Modeling

This chapter presents an introduction to object-oriented modeling using Object Management Group's Unified Modeling Language (UML). This chapter has been carefully reviewed to ensure consistency with the latest UML notation and best industry practices. UML provides an industry-standard notation for representing classes and objects. The chapter continues to emphasize basic object-oriented concepts, such as inheritance, encapsulation, composition, and polymorphism. As with Chapter 13, Chapter 14 is available on the textbook's Web site.

APPENDICES

In the 13th edition three appendices are available on the book's Web site and are intended for those who wish to explore certain topics in greater depth.

Appendix A: Data Modeling Tools and Notation

This appendix addresses a need raised by many readers—how to translate the E-R notation in the text into the form used by the CASE tool or the DBMS used in class. Specifically, this appendix compares the notations of CA ERwin Data Modeler r9.7, Oracle SQL Data Modeler 4.2, SAP Sybase PowerDesigner 16.6, and Microsoft Visio Professional 2016. Tables and illustrations show the notations used for the same constructs in each of these popular software packages.

Appendix B: Advanced Normal Forms

This appendix presents a description (with examples) of Boyce-Codd and fourth normal forms, including an example of BCNF to show how to handle overlapping candidate keys. Other normal forms are briefly introduced. The Web Resources section includes a reference for information on many advanced normal form topics.

Appendix C: Data Structures

This appendix describes several data structures that often underlie database implementations. Topics include the use of pointers, stacks, queues, sorted lists, inverted lists, and trees.