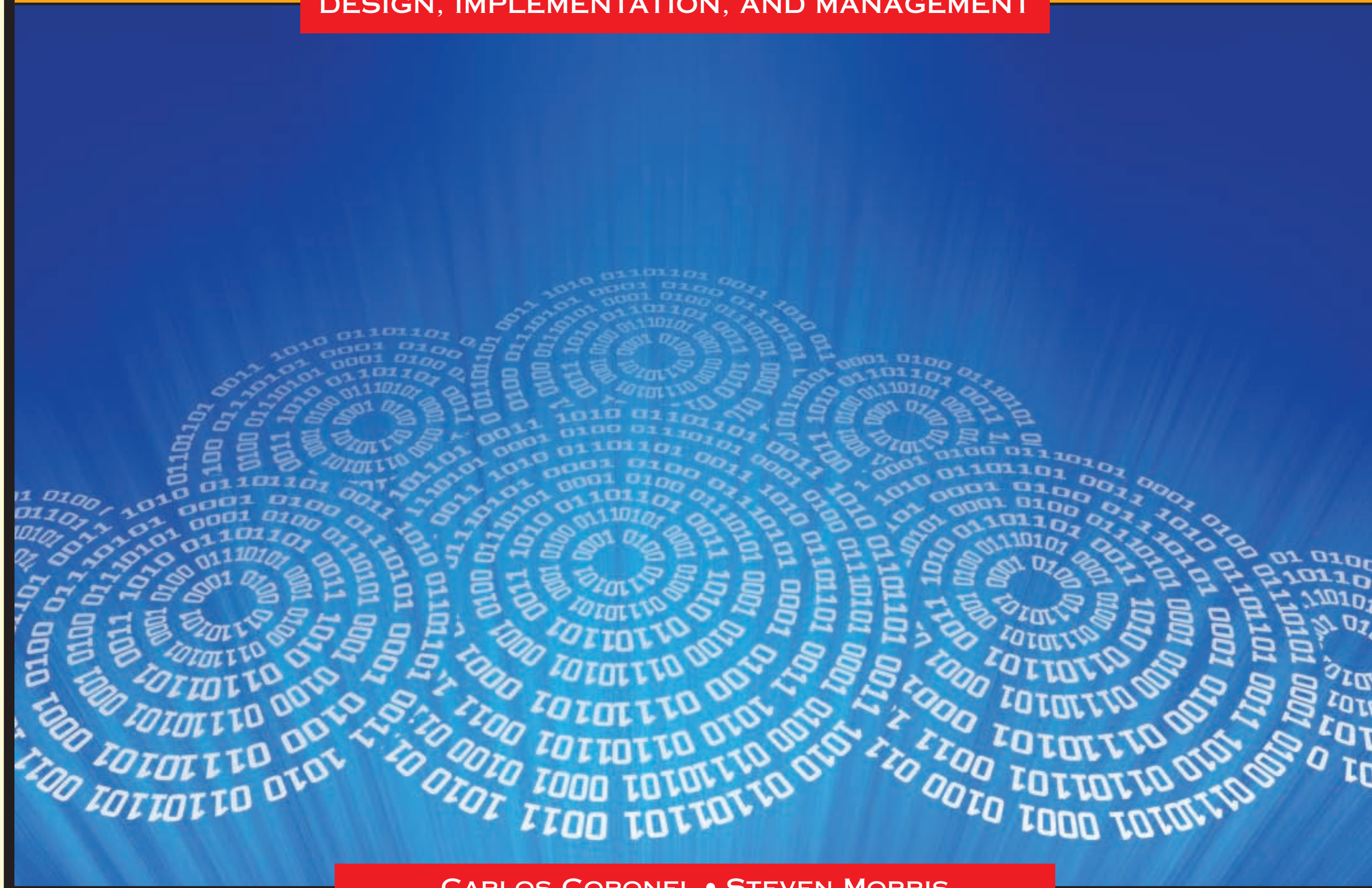
Coronel / Morris

Database Systems

Design, Implementation, and Management 11e

DATABASE SYSTEMS

DESIGN, IMPLEMENTATION, AND MANAGEMENT



CARLOS CORONEL • STEVEN MORRIS

CENGAGE Learning®

Australia • Brazil • Mexico • Singapore • UnitedKingdom • UnitedStates

Database Systems: Design, Implementation, and Management, Eleventh Edition
Carlos Coronel and Steven Morris

Product Director: Joe Sabatino

Product Manager: Clara Goosman

Senior Content Developer: Kate Mason

Content Developer: Anne Merrill

Product Assistant: Brad Sullender

Marketing Manager: Erik La Scola

Marketing Coordinator: Elizabeth Murphy

Design and Production Service: GEX Publishing Services

Manufacturing Planner: Ron Montgomery

Rights Acquisition Specialist: Christine Myaskovsky

Cover image(s): ©iStockphoto.com/derrrek

For product information and technology assistance, contact us at
Cengage Learning Customer & Sales Support, 1-800-354-9706

For permission to use material from this text or product,submit all requests online at **www.cengage.com/permissions.**
Further permissions questions can be e-mailed to
permissionrequest@cengage.com

Library of Congress Control Number: 2013954240

Student Edition:

ISBN-13: 978-1-285-19614-5

ISBN-10: 1-285-19614-7

Instructor Edition:

ISBN-13: 978-1-285-19618-3

ISBN-10: 1-285-19618-X

Cengage Learning
200 First Stamford Place, 4th Floor
Stamford, CT 06902
USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at:
www.cengage.com/global

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

For your course and learning solutions, visit
www.cengage.com.

Purchase any of our products at your local college store or at our preferred online store **www.cengagebrain.com.**

Instructors: Please visit **login.cengage.com** and log in to access instructor-specific resources.

Printed in the United States of America
1 2 3 4 5 6 7 17 16 15 14 13

Dedication

DEDICATION

To the treasures in my life: To Victoria, for 23 wonderful years. Thank you for your unending support, for being my angel, my sweetie, and most importantly, my best friend. To Carlos Anthony for the many good times you have given us; you show us the way. Thank you for your words of wisdom, contagious happiness, and for bringing us shining days. You are still young; your best times are still to come. To Gabriela Victoria, who is the image of brilliance, beauty, and faithfulness. Thank you for being the sunshine in my cloudy days. To Christian Javier, whose endless energy and delightful smiles keep us going every day. Thank you for being the youthful reminder of life's simple beauties. To my parents, Sarah and Carlos, thank you for your sacrifice and example. To all of you, you are all my inspiration. "TQTATA."

Carlos Coronel

To Pamela, from high school sweetheart through 24 years of marriage, the beautiful love of my life who has supported, encouraged, and inspired me. More than anyone else, you are responsible for whatever successes I have achieved. To my son, Alexander Logan, whose depth of character is without measure. You are my pride and joy. To my daughter, Lauren Elizabeth, whose beauty and intensity take my breath away. You are my heart and soul. Thank you all for the sacrifices you have made that enabled me to pursue this dream. I love you so much more than I can express. To my mother, Florence Maryann, and to the memory of my father, Alton Lamar, who together instilled in me the desire to learn and the passion to achieve. To my mother-in-law, Connie Duke, and to the memory of my father-in-law, Wayne Duke, who taught me to find joy in all things. To all of you, with all my love, I dedicate this book.

Steven Morris

FOR PETER

To longtime colleague and friend, Peter Rob: Your drive and dedication to your students started this book. Your depth of knowledge, attention to detail, and pursuit of excellence made it succeed. Your patience and guidance continue to light our path. It is our sincere hope that as we move forward, we can continue to live up to your standard. Enjoy your retirement, my friend; you have surely earned it.

Carlos Coronel and Steven Morris

BRIEF CONTENTS

The following appendixes are included online at cengagebrain.com

The CourseMate Site for this text is available at *www.cengagebrain.com*.

TABLE OF CONTENTS

PART I DATABASE CONCEPTS

PART II DESIGN CONCEPTS

TABLE OF CONTENTS

PART III ADVANCED DESIGN AND IMPLEMENTATION

TABLE OF CONTENTS

PART IV ADVANCED DATABASE CONCEPTS

TABLE OF CONTENTS

PART V DATABASES AND THE INTERNET

PART VI DATABASE ADMINISTRATION

The following appendixes are included on the Instructor and Student Companion Sites or at the CourseMate site for this text, *www.cengagebrain.com*.

The CourseMate Site for this text is available at *www.cengagebrain.com*.

PREFACE

An eleventh edition is a milestone that few textbooks achieve. We wrote the first edition of this book because we wanted to explain the complexity of database systems in a language that was easy for students to understand. Over the years, we have maintained this emphasis on reaching out to students to explain complex concepts in a practical, approachable manner. This book has been successful through ten editions because the authors, editors, and the publisher paid attention to the impact of technology and to adopter questions and suggestions. We believe that this eleventh edition successfully reflects the same attention to such factors.

In many respects, rewriting a book is more difficult than writing it the first time. If the book is successful, as this one is, a major concern is that the updates, inserts, and deletions will adversely affect writing style and continuity of coverage. The combination of superb reviewers and editors, plus a wealth of feedback from adopters and students of the previous editions, helped make this new edition the best yet.

PREFACE

CHANGES TO THE ELEVENTH EDITION

In this eleventh edition, we have added some new features and reorganized some coverage to provide a better flow of material. Aside from enhancing the already strong coverage of database design, we have made other improvements in the topical coverage. Here are a few of the highlights:

- Expanded relational algebra coverage with formal definitions and notations
- Updated Business Vignettes showing the impact of database technologies in the real world
- Updated coverage of cloud data services
- Expanded coverage of Big Data and related Hadoop technologies
- Added coverage of data visualization
- SQL coverage expanded to include MySQL databases
- Improved readability and overall visual appeal

This eleventh edition continues to provide a solid and practical foundation for the design, implementation, and management of database systems. This foundation is built on the notion that, while databases are very practical, their successful creation depends on understanding the important concepts that define them. It's not easy to come up with the proper mix of theory and practice, but the previously mentioned feedback suggests that we largely succeeded in our quest to maintain the proper balance.

THE APPROACH: A CONTINUED EMPHASIS ON DESIGN

As the title suggests, ***Database Systems: Design, Implementation, and Management*** covers three broad aspects of database systems. However, for several important reasons, special attention is given to database design.

- The availability of excellent database software enables people with little experience to create databases and database applications. Unfortunately, the "create without design" approach usually paves the road to any number of database disasters. In our experience, many database system failures are traceable to poor design and cannot be solved with the help of even the best programmers and managers. Nor is better DBMS software likely to overcome problems created or magnified by poor design. Even the best bricklayers and carpenters can't create a good building from a bad blueprint.

- Most vexing problems of database system management seem to be triggered by poorly designed databases. It hardly seems worthwhile to use scarce resources to develop excellent database management skills merely to use them on crises induced by poorly designed databases.

- Design provides an excellent means of communication. Clients are more likely to get what they need when database system design is approached carefully and thoughtfully. In fact, clients may discover how their organizations really function once a good database design is completed.

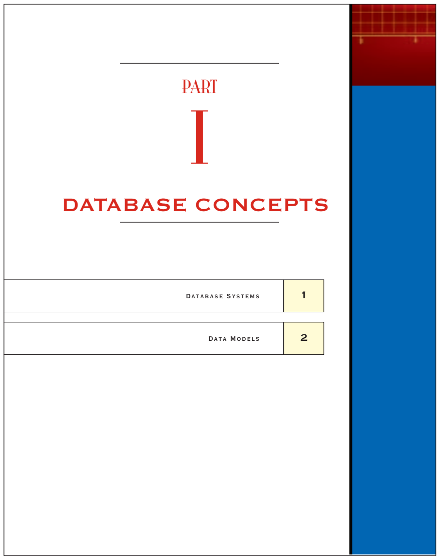
- Familiarity with database design techniques promotes understanding of current database technologies. For example, because data warehouses derive much of their data from operational databases, data warehouse concepts, structures, and procedures make more sense when the operational database's structure and implementation are understood.

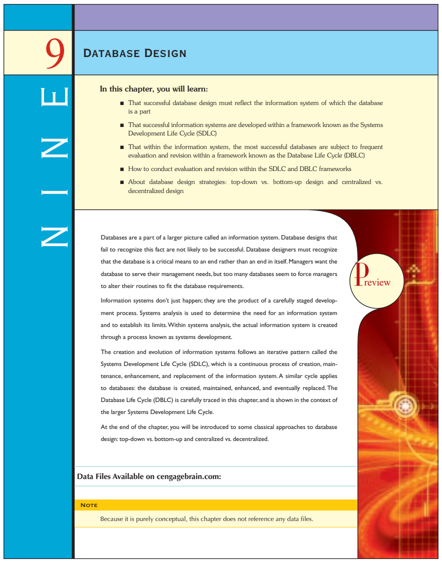
Because the practical aspects of database design are stressed, we have covered design concepts and procedures in detail, making sure that the numerous end-of-chapter problems and cases are sufficiently challenging so students can develop real and useful design skills. We also make sure that students understand the potential and actual conflicts between database design elegance, information requirements, and transaction processing speed. For example, it makes little sense to design databases that meet design elegance standards while they fail to meet end-user information requirements. Therefore, we explore the use of carefully defined trade-offs to ensure that the databases meet end-user requirements while conforming to high design standards.

TOPICAL COVERAGE

The Systems View

The book's title begins with ***Database Systems***. Therefore, we examine the database and design concepts covered in Chapters 1–6 as part of a larger whole by placing them within the systems analysis framework of Chapter 9. Database designers who fail to understand that the database is part of a larger system are likely to overlook important design requirements. In fact, Chapter 9, Database Design, provides the map for the advanced database design developed in Appendixes B and C. Within the larger systems framework, we can also explore issues such as transaction management and concurrency control (Chapter 10), distributed database management systems (Chapter 12), business intelligence and data warehouses (Chapter 13), database connectivity and web technologies (Chapter 14), and database administration and security (Chapter 15).




9 DATABASE DESIGN

NINE

In this chapter, you will learn:

- That successful database design must reflect the information system of which the database is a part
- That successful information systems are developed within a framework known as the Systems Development Life Cycle (SDLC)
- That within the information system, the most successful databases are subject to frequent evaluation and revision within a framework known as the Database Life Cycle (DBLC)
- How to conduct evaluation and revision within the SDLC and DBLC frameworks
- About database design strategies: top-down vs. bottom-up design and centralized vs. decentralized design

Preview

Databases are a part of a larger picture called an information system. Database designs that fail to recognize this fact are not likely to be successful. Database designers must recognize that the database is a critical means to an end rather than an end in itself. Managers want the database to serve their management needs, but too many databases seem to force managers to alter their routines to fit the database requirements.

Information systems don't just happen; they are the product of a carefully staged development process. Systems analysis is used to determine the need for an information system and to establish its limits. Within systems analysis, the actual information system is created through a process known as systems development.

The creation and evolution of information systems follows an iterative pattern called the Systems Development Life Cycle (SDLC), which is a continuous process of creation, maintenance, enhancement, and replacement of the information system. A similar cycle applies to databases: the database is created, maintained, enhanced, and eventually replaced. The Database Life Cycle (DBLC) is carefully traced in this chapter, and is shown in the context of the larger Systems Development Life Cycle.

At the end of the chapter, you will be introduced to some classical approaches to database design: top-down vs. bottom-up and centralized vs. decentralized.

Data Files Available on cengagebrain.com:

NOTE

Because it is purely conceptual, this chapter does not reference any data files.

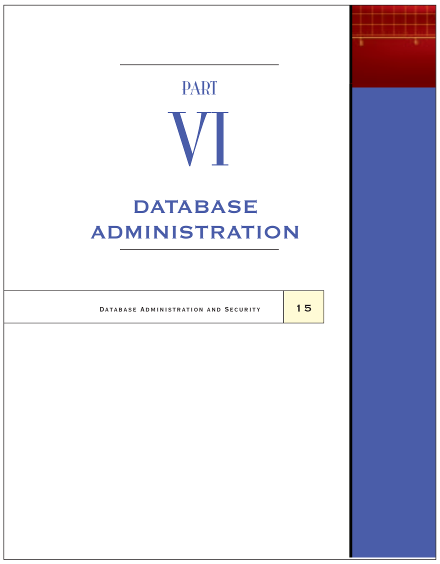
Database Design

The first item in the book's subtitle is ***Design***, and our examination of database design is comprehensive. For example, Chapters 1 and 2 examine the development and future of databases and data models, and illustrate the need for design. Chapter 3 examines the details of the relational database model; Chapter 4 provides extensive, in-depth, and practical database design coverage; and Chapter 5 explores advanced database design topics. Chapter 6 is devoted to critical normalization issues that affect database efficiency and effectiveness. Chapter 9 examines database design within the systems framework and maps the activities required to successfully design and implement the complex, real-world database developed in Appendixes B and C. Appendix A, Designing Databases with Visio Professional: A Tutorial, provides a good introductory tutorial for the use of a database design tool.

Because database design is affected by real-world transactions, the way data are distributed, and ever-increasing information requirements, we examine major database features that must be supported in current-generation databases and models. For example, Chapter 10, Transaction Management and Concurrency Control, focuses on the characteristics of database transactions and how they affect database integrity and consistency. Chapter 11, Database Performance Tuning and Query Optimization, illustrates the need for query efficiency in a world that routinely generates and uses terabyte-sized databases and tables with millions of records. Chapter 12, Distributed Database Management Systems, focuses on data distribution, replication, and allocation. In Chapter 13, Business Intelligence and Data Warehouses, we explore the characteristics of databases that are used in decision support and online analytical processing. Chapter 14, Database Connectivity and Web Technologies, covers the basic database connectivity issues in a web-based data world, development of web-based database front ends, and emerging cloud-based services.

Implementation

The second portion of the subtitle is ***Implementation***. We use Structured Query Language (SQL) in Chapters 7 and 8 to show how databases are implemented and managed. Appendix M, Microsoft Access Tutorial, provides a quick but comprehensive guide to implementing an MS Access database. Appendixes B and C demonstrate the design of a database that was fully implemented; these appendixes illustrate a wide range of implementation issues. We had to deal with conflicting design goals: design elegance, information requirements, and operational speed. Therefore, we carefully audited the initial design in Appendix B to check its ability to meet end-user needs and establish appropriate implementation protocols. The result of this audit yielded the final design developed in Appendix C. The special issues encountered in an Internet database environment are addressed in Chapter 14, Database Connectivity and Web Technologies, and in Appendix J, Web Database Development with ColdFusion.



Management

The final portion of the subtitle is ***Management***. We deal with database management issues in Chapter 10, Transaction Management and Concurrency Control; Chapter 12, Distributed Database Management Systems; and Chapter 15, Database Administration and Security. Chapter 11, Database Performance Tuning and Query Optimization, is a valuable resource that illustrates how a DBMS manages data retrieval. In addition, Appendix N, Creating a New Database Using Oracle 12c, walks you through the process of setting up a new database.

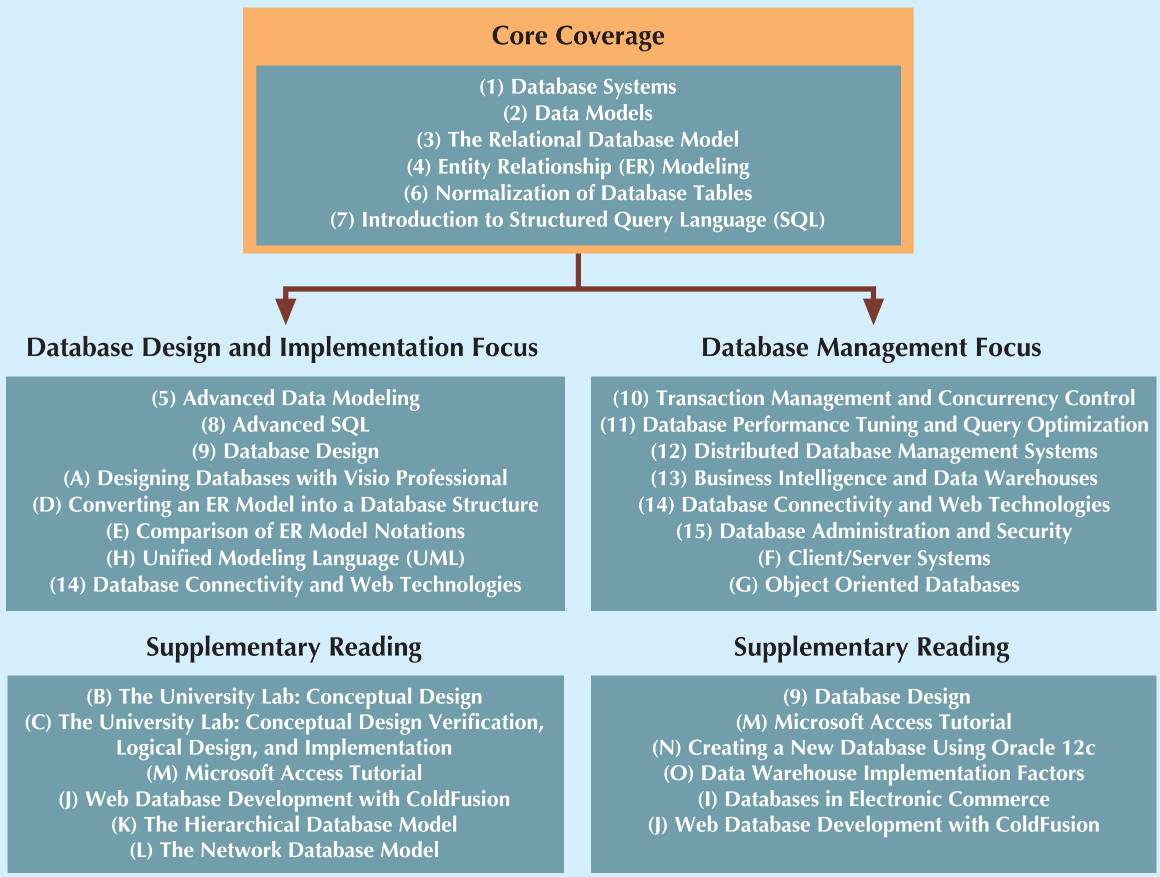
TEACHING DATABASE: A MATTER OF FOCUS

Given the wealth of detailed coverage, instructors can “mix and match” chapters to produce the desired coverage. Depending on where database courses fit into the curriculum, instructors may choose to emphasize database design or database management. (See Figure 1.)

The hands-on nature of database design lends itself particularly well to class projects for which students use instructor-selected software to prototype a system they design for the end user. Several end-of-chapter problems are sufficiently complex to serve as projects, or an instructor may work with local businesses to give students hands-on experience. Note that some elements of the database design track are also found in the database management track, because it is difficult to manage database technologies that are not well understood.

The options shown in Figure 1 serve only as a starting point. Naturally, instructors will tailor their coverage based on their specific course requirements. For example, an instructor may decide to make Appendix I an outside reading assignment and Appendix A a self-taught tutorial, and then use that time to cover client/server systems or object-oriented databases. The latter choice would serve as a gateway to UML coverage.

FIGURE 1



Business Vignettes highlight topics in a real-life setting.

The Relational Evolution and Beyond

Business Vignette

For over 40 years, relational databases (RDBMS) have dominated the database market – with Oracle maintaining a leading position. In 2012, Gartner, the leading information technology research company, reported that Oracle holds a greater share of the RDBMS revenue than its four closest competitors combined.[1]

Researchers, journalists, and business leaders agree, however, that the industry now stands at the brink of another database revolution. New developments have given rise to new database models. The Internet has witnessed the explosion of unstructured and semi-structured data, such as online journals, blogs, and videos. Gartner predicts that unstructured data will account for 80 percent of the new data uploaded to the Internet.[2] Although relational databases meet rigorous standards for data integrity and consistency, they do not scale unstructured data as well as new database models, such as NoSQL.[3] As a result, social networking sites, such as Twitter and Facebook, that are home to semi-structured data and do not require high levels of data consistency and integrity have adopted noSQL databases.

The number of data transactions has also skyrocketed in recent years. Within a decade, businesses will have to analyze 50 times more data files that are 75 times as large as they are today.[4] Companies need databases that can process this "Big Data" so that they can obtain timely and accurate business intelligence. As a result, many Big Data companies have turned to new data models. When Visa needed

Online Content boxes draw attention to material at *www.cengagebrain.com* for this text and provide ideas for incorporating this content into the course.

Online Content

The databases used in each chapter are available at *www.cengagebrain.com*. Throughout the book, Online Content boxes highlight material related to content at *www.cengagebrain.com*.

Notes highlight important facts about the concepts introduced in the chapter.

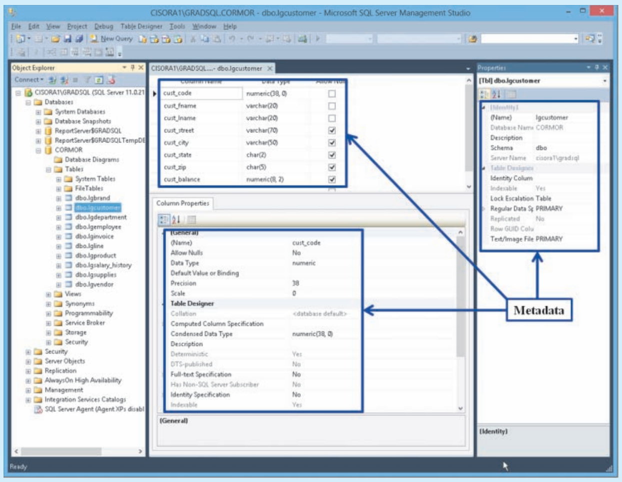
Note

Data that display data inconsistency are also referred to as data that lack data integrity. **Data integrity** is defined as the condition in which all of the data in the database are consistent with the real-world events and conditions. In other words, data integrity means that:

- Data are *accurate*—there are no data inconsistencies.
- Data are *verifiable*—the data will always yield consistent results.

A variety of **four-color figures**, including ER models and implementations, tables, and illustrations, clearly illustrate difficult concepts.

FIGURE 1.10 **Illustrating metadata with Microsoft SQL Server Express**



Cengage Learning © 2015

TEXT FEATURES

SUMMARY

- The ERM uses ERDs to represent the conceptual database as viewed by the end user. The ERM's main components are entities, relationships, and attributes. The ERD also includes connectivity and cardinality notations. An ERD can also show relationship strength, relationship participation (optional or mandatory), and degree of relationship (unary, binary, ternary, etc.).
- Connectivity describes the relationship classification (1:1, 1:M, or M:N). Cardinality expresses the specific number of entity occurrences associated with an occurrence of a related entity. Connectivities and cardinalities are usually based on business rules.
- In the ERM, an M:N relationship is valid at the conceptual level. However, when implementing the ERM in a relational database, the M:N relationship must be mapped to a set of 1:M relationships through a composite entity.

A robust **Summary** at the end of each chapter ties together the major concepts and serves as a quick review for students.

KEY TERMS

algorithms	data mart	explanatory analytics
attribute hierarchy	data mining	extraction, transformation, and loading (ETL)
business intelligence (BI)	data warehouse	fact table
cube cache	decision support system (DSS)	facts
dashboard	dimension tables	governance
data analytics	dimensions	key performance indicators (KPIs)
data cube	drill down	

An alphabetic list of **Key Terms** summarizes important terms.

REVIEW QUESTIONS

1. What two conditions must be met before an entity can be classified as a weak entity? Give an example of a weak entity.
2. What is a strong (or identifying) relationship, and how is it depicted in a Crow's Foot ERD?
3. Given the business rule "an employee may have many degrees," discuss its effect on attributes, entities, and relationships. (*Hint:* Remember what a multivalued attribute is and how it might be implemented.)
4. What is a composite entity, and when is it used?
5. Suppose you are working within the framework of the conceptual model in Figure Q4.5.

Review Questions challenge students to apply the skills learned in each chapter.

PROBLEMS

1. Given the following business rules, create the appropriate Crow's Foot ERD.
   a. A company operates many departments.
   b. Each department employs one or more employees.
   c. Each of the employees may or may not have one or more dependents.
   d. Each employee may or may not have an employment history.
2. The Hudson Engineering Group (HEG) has contacted you to create a conceptual model whose application will meet the expected database requirements for the company's training program. The HEG administrator gives you

Problems become progressively more complex as students draw on the lessons learned from the completion of preceding problems.

COURSEMATE

Cengage Learning's *Database Systems* CourseMate brings course concepts to life with interactive learning, study, and exam preparation tools that support the printed textbook. Watch student comprehension soar as your class works with the printed textbook and the textbook-specific website. CourseMate goes beyond the book to deliver what you need! Learn more at *www.cengage.com/coursemate*.

Engagement Tracker

How do you assess your students' engagement in your course? How do you know students have read the material or viewed the resources you've assigned? How can you tell if your students are struggling with a concept? With CourseMate, you can use the included Engagement Tracker to assess student preparation and engagement. Use the tracking tools to see progress for the class as a whole or for individual students. Identify students at risk early in the course. Uncover which concepts are most difficult for your class. Monitor time on task. Keep your students engaged.

Interactive Teaching and Learning Tools

CourseMate includes interactive teaching and learning tools:

- Quizzes
- Crossword puzzles
- Flashcards
- Videos
- PowerPoint presentations
- and more

These assets enable students to review for tests, prepare for class, and address the needs of their varied learning styles.

Interactive eBook

MindTap Reader is Cengage Learning's re-imagination of the traditional eBook, specifically designed for assimilating content and media assets in a fully online – and often mobile – reading environment. MindTap Reader combines thoughtful navigation ergonomics, advanced annotation support and interactivity through the placement of inline documents and media assets. These capabilities create an engaging reading experience further enhanced through tightly integrated web-apps (e.g. text-to-speech, note-taking, utilities) delivering a learning tool that drives immediacy, relevancy and engagement for today's learners.

Appendixes

Fifteen online appendixes provide additional material on a variety of important areas, such as using Microsoft® Visio® and Microsoft® Access®, ER model notations, UML, object-oriented databases, databases and electronic commerce, and Adobe® ColdFusion®.

Database, SQL Script, and ColdFusion Files

The online materials for this book include all of the database structures and table contents used in the text. For students using Oracle®, MySQL and Microsoft SQL Server™, SQL scripts are included to help students create and load all tables used in the SQL chapters (7 and 8). In addition, all ColdFusion scripts used to develop the web interfaces in Appendix J are included.

ADDITIONAL FEATURES

INSTRUCTOR RESOURCES

Database Systems: Design, Implementation, and Management, Eleventh Edition, includes teaching tools to support instructors in the classroom. The ancillary material that accompanies the textbook is listed below. They are available on the web at *www.cengage.com*.

Instructor's Manual
The authors have created this manual to help instructors make their classes informative and interesting. Because the authors tackle so many problems in depth, instructors will find the *Instructor's Manual* especially useful. The details of the design solution process are shown in the *Instructor's Manual*, as well as notes about alternative approaches that may be used to solve a particular problem.

SQL Script Files for Instructors
The authors have provided teacher's SQL script files to let instructors cut and paste the SQL code into the SQL windows. (Scripts are provided for Oracle, MySQL, and MS SQL Server.) The SQL scripts, which have all been tested by Cengage Learning, are a major convenience for instructors. You won't have to type in the SQL commands, and the use of the scripts eliminates typographical errors that are sometimes difficult to trace.

ColdFusion Files for Instructors
The ColdFusion web development solutions are provided. Instructors have access to a menu-driven system that lets teachers show the code as well as its execution.

Databases
For many chapters, Microsoft® Access® instructor databases are available that include features not found in the student databases. For example, the databases that accompany Chapters 7 and 8 include many of the queries that produce the problem solutions. Other Access databases, such as the ones that accompany Chapters 3, 4, 5, and 6, include implementations of the design problem solutions to let instructors illustrate the effect of design decisions.. In addition, instructors have access to all the script files for Oracle, MySQL, and MS SQL Server so that all the databases and their tables can be converted easily and precisely.

Cengage Learning Testing Powered by Cognero
A flexible, online system that allows you to:

- Author, edit, and manage test bank content from multiple Cengage Learning solutions
- Create multiple test versions in an instant
- Deliver tests from your LMS, your classroom or wherever you want

Start right away!
Cengage Learning Testing Powered by Cognero works on any operating system or browser.

- No special installs or downloads needed
- Create tests from school, home, the coffee shop – anywhere with Internet access

What will you find?

- Simplicity at every step. A desktop-inspired interface features drop-down menus and familiar, intuitive tools that take you through content creation and management with ease.
- Full-featured test generator. Create ideal assessments with your choice of 15 question types (including true/false, multiple choice, opinion scale/likert, and essay). Multi-language support, an equation editor and unlimited metadata help ensure your tests are complete and compliant.
- Cross-compatible capability. Import and export content into other systems.

PowerPoint® Presentations

Microsoft PowerPoint slides are included for each chapter. Instructors can use the slides in a variety of ways—for example, as teaching aids during classroom presentations or as printed handouts for classroom distribution. Instructors can modify these slides or include slides of their own for additional topics introduced to the class.

Figure Files

Figure files for solutions are presented in the Instructor's Manual to allow instructors to create their own presentations. Instructors can also manipulate these files to meet their particular needs.

ACKNOWLEDGMENTS

Regardless of how many editions of this book are published, they will always rest on the solid foundation created by the first edition. We remain convinced that our work has become successful because that first edition was guided by Frank Ruggirello, a former Wadsworth senior editor and publisher. Aside from guiding the book's development, Frank also managed to solicit the great Peter Keen's evaluation (thankfully favorable) and subsequently convinced PK to write the foreword for the first edition. Although we sometimes found Frank to be an especially demanding taskmaster, we also found him to be a superb professional and a fine friend. We suspect Frank will still see his fingerprints all over our current work. Many thanks.

A difficult task in rewriting a book is deciding what new approaches, topical coverage, and changes to depth of coverage are appropriate for a product that has successfully weathered the test of the marketplace. The comments and suggestions made by the book's adopters, students, and reviewers play a major role in deciding what coverage is desirable and how that coverage is to be treated.

Some adopters became extraordinary reviewers, providing incredibly detailed and well-reasoned critiques even as they praised the book's coverage and style. Dr. David Hatherly, a superb database professional who is a senior lecturer in the School of Information Technology, Charles Sturt University–Mitchell, Bathhurst, Australia, made sure that we knew precisely what issues led to his critiques. Even better for us, he provided the suggestions that made it much easier for us to improve the topical coverage in earlier editions. Dr. Hatherly's recommendations continue to be reflected in this tenth edition. All of his help was given freely and without prompting on our part. His efforts are much appreciated, and our thanks are heartfelt.

We also owe a debt of gratitude to Professor Emil T. Cipolla, who teaches at St. Mary College. Professor Cipolla's wealth of IBM experience turned out to be a valuable resource when we tackled the embedded SQL coverage in Chapter 8.

Every technical book receives careful scrutiny by several groups of reviewers selected by the publisher. We were fortunate to face the scrutiny of reviewers who were superbly qualified to offer their critiques, comments, and suggestions—many of which strengthened this edition. While holding them blameless for any remaining shortcomings, we owe these reviewers many thanks for their contributions:

Mubarak Banisaklher, Bethune Cookman University

David Bell, Pacific Union College

Yurii Boreisha, Minnesota State University Moorhead

Laurie Crawford, Franklin University

Mel Goetting, Shawnee State University

Jeff Guan, University of Louisville

William Hochstettler, Franklin University

Laurene Hutchinson, Louisiana State University, Baton Rouge

Nitin Kale, University of Southern California, Los Angeles

Gerald Karush, Southern New Hampshire University

Michael Kelly, Community College of Rhode Island

Timothy Koets, Grand Rapids Community College

Klara Nelson, The University of Tampa

Chiso Okafor, Roxbury Community College

Brandon Olson, The College of St. Scholastica

James Reneau, Shawnee State University

Julio Rivera, University of Alabama at Birmingham

Ruth Robins, University of Houston Downtown

Samuel Sambasivam, Azusa Pacific University

Paul Seibert, North Greenville University

Ronghua Shan, Dakota State University

Andrew Smith, Marian University

Antonis Stylianou, University of North Carolina Charlotte

Brian West, University of Louisiana at Lafayette

Nathan White, McKendree University

ACKNOWLEDGMENTS

In some respects, writing books resembles building construction: When 90 percent of the work seems done, 90 percent of the work remains to be done. Fortunately for us, we had a great team on our side.

- We are deeply indebted to Deb Kaufmann for her help and guidance. Deb has been everything we could have hoped for in a development editor and more. She brought fresh eyes and a new perspective that meshed wonderfully with our vision for the new edition.

- After writing so many books and ten editions of *this* book, we know just how difficult it can be to transform the authors' work into an attractive product. The production team, both at Cengage Learning (Matt Hutchinson) and GEX Publishing Services (Marisa Taylor), have done an excellent job.

- We also owe Kate Mason, our Content Developer, special thanks for her ability to guide this book to a successful conclusion. Kate's work touched on all of the publication bases, and her managerial skills protected us from publishing gremlins that might have become a major nuisance. Also, her skills in dealing with occasionally cranky authors far exceed those of any diplomat we can think of. And did we mention that Kate is, quite simply, a delightful person? We would also like to thank Anne Merrill, Content Developer, who competently shepherded the text in its later stages.

We also thank our students for their comments and suggestions. They are the reason for writing this book in the first place. One comment stands out in particular: "I majored in systems for four years, and I finally discovered why when I took your course." And one of our favorite comments by a former student was triggered by a question about the challenges created by a real-world information systems job: "Doc, it's just like class, only easier. You really prepared me well. Thanks!"

Special thanks go to a very unique and charismatic gentleman. For over 20 years Peter Rob has been the driving force behind the creation and evolution of this book. This book originated as a product of his drive and dedication to excellence. For over twenty years, he was the voice of *Database Systems*, and the driving force behind its advancement. We wish him peace in his retirement, time with his loved ones, and luck on his many projects.

Last, and certainly not least, we thank our families for their solid support at home. They graciously accepted the fact that during more than a year's worth of rewriting, there would be no free weekends, rare free nights, and even rarer free days. We owe you much, and the dedications we wrote are but a small reflection of the important space you occupy in our hearts.

Carlos Coronel and Steven Morris

PART I

DATABASE CONCEPTS

Business Vignette

The Relational Evolution and Beyond

For over 40 years, relational databases (RDBMS) have dominated the database market – with Oracle maintaining a leading position. In 2012, Gartner, the leading information technology research company, reported that Oracle holds a greater share of the RDBMS revenue than its four closest competitors combined.[1]

Researchers, journalists, and business leaders agree, however, that the industry now stands at the brink of another database revolution. New developments have given rise to new database models. The Internet has witnessed the explosion of unstructured and semi-structured data, such as online journals, blogs, and videos. Gartner predicts that unstructured data will account for 80 percent of the new data uploaded to the Internet.[2] Although relational databases meet rigorous standards for data integrity and consistency, they do not scale unstructured data as well as new database models, such as NoSQL.[3] As a result, social networking sites, such as Twitter and Facebook, that are home to semi-structured data and do not require high levels of data consistency and integrity have adopted noSQL databases.

The number of data transactions has also skyrocketed in recent years. Within a decade, businesses will have to analyze 50 times more data files that are 75 times as large as they are today.[4] Companies need databases that can process this "Big Data" so that they can obtain timely and accurate business intelligence. As a result, many Big Data companies have turned to new data models. When Visa needed to process two years of credit transactions, the company turned to Hadoop to process 70 billion transactions, thereby reducing processing time from one month to 13 minutes.[5] Another technology that is fast gaining ground is in-memory databases, which store data in main memory rather than secondary storage. In-memory databases take advantage of parallel computing and multicore processors to deliver faster business intelligence.[6]

SAP's market share grew by an impressive 48 percent in 2012 (compared to Oracle's 18 percent) and the company attributes this growth to the SAP HANNA In-Memory Appliance.[7]

Additionally, cloud computing provides on-demand access to clusters of computers over the Internet at a reasonable cost. Cloud computing enables the widespread adoption of powerful databases in organizations that could not previously afford the high cost of massive server infrastructure.[8]

All these technology advances are revolutionizing the database market. However, the relational model is not standing idly by. Today, most large relational database vendors provide in-memory database and cloud capabilities. Oracle has invested heavily in research and has acquired many smaller companies in order to expand its capabilities. At the end of 2011, it acquired RightNow Technologies for $1.5. The deal gave Oracle a leading position in cloud customer service.[9]

The relational model has a history of overcoming challenges from other emerging database technologies, which is why the overwhelming majority of existing systems still rely on relational databases. It remains to be seen if the relational model will evolve once again and adapt to these new challenges. The race is on.

[1] Mark Fontecchio, "Oracle the clear leader in $24 billion RDBMS market," *IT Knowledge Exchange*, April 12, 2012, *http://itknowledgeexchange.techtarget.com/eye-on-oracle/oracle-the-clear-leader-in-24-billion-rdbms-market/*.

[2] Sameer Nori, "Are You Ready for the Growth in Unstructured Information?" *ATTIV/O Active Intelligence,* March 7, 2013, *http://www.attivio.com/blog/55-industry-insights/1217-are-you-ready-for-the-growth-in-unstructured-information.html*.

[3] Guy Harrison, "10 things you should know about NoSQL databases," *TechRepublic*, August 26, 2010, *http://www.techrepublic.com/blog/10things/10-things-you-should-know-about-nosql-databases/1772*.

[4] David Rosenbaum, "That New Big Data Magic," *CFO*, August 29, 2011, *http://www3.cfo.com/article/2011/ 8/analytics_that-new-big-data-magic*.

[5] David Strom, "NoSQL: Breaking free of structured data," *IT World*, June 9, 2011, *http://www.itworld.com/data-centerservers/172477/nosql-breaking-free-structured-data*.

[6] Mike Simons, "SAP Cofounder Hasso Plattner Promises in-Memory Database Revolution," *CIO*, May 14, 2009, *http://www.cio.com/article/492832/SAP_Cofounder_Hasso_Plattner_Promises_in_Memory_ Database_Revolution*.

[7] Mark Fontecchio, "Oracle the clear leader in $24 billion RDBMS market," *IT Knowledge Exchange*, April 12, 2012, *http://itknowledgeexchange.techtarget.com/eye-on-oracle/oracle-the-clear-leader-in-24-billion-rdbms-market/*.

[8] Joshua Greenbaum, "A Revolution Threatens the Relational Database," *IT Managment*, May 13, 2008, *http://itmanagement.earthweb.com/columns/entad/article.php/3746191*.

[9] Larry Dignam, "Oracle acquires RightNow for $1.5 billion, aims turrets at Salesforce.com," *ZDNet,* October 24, 2011, *http://www.zdnet.com/blog/btl/oracle-acquires-rightnow-for-1-5-billion-aims-turrets-at-salesforce-com/61681*.

1

ONE

Database Systems

In this chapter, you will learn:

- The difference between data and information.
- What a database is, the various types of databases, and why they are valuable assets for decision making.
- The importance of database design.
- How modern databases evolved from file systems.
- About flaws in file system data management.
- The main components of the database system.
- The main functions of a database management system (DBMS).

Preview

Good decisions require good information that is derived from raw facts. These raw facts are known as data. Data are likely to be managed most efficiently when they are stored in a database. In this chapter, you will learn what a database is, what it does, and why it yields better results than other data management methods. You will also learn about various types of databases and why database design is so important.

Databases evolved from computer file systems. Although file system data management is now largely outmoded, understanding the characteristics of file systems is important because file systems are the source of serious data management limitations. In this chapter, you will also learn how the database system approach helps eliminate most of the shortcomings of file system data management.

Data Files Available on cengagebrain.com:

Chapter	Available formats: MS Access	Oracle	MS SQL	MySQL	Questions and Problems	Available formats: MS Access	Oracle	MS SQL	MySQL
Ch01_Text	✓	✓	✓	✓	Ch01_Problems	✓	✓	✓	✓
Ch01_Design_Example	✓	✓	✓	✓					

1.1 Why Databases?

Imagine trying to operate a business without knowing who your customers are, what products you are selling, who is working for you, who owes you money, and to whom you owe money. All businesses have to keep this type of data and much more; just as importantly, they must have those data available to decision makers when necessary. It can be argued that the ultimate purpose of all business information systems is to help businesses use information as an organizational resource. At the heart of all of these systems are the collection, storage, aggregation, manipulation, dissemination, and management of data.

Depending on the type of information system and the characteristics of the business, these data could vary from a few megabytes on just one or two topics to terabytes covering hundreds of topics within the business's internal and external environment. Telecommunications companies, such as Sprint and AT&T, are known to have systems that keep data on trillions of phone calls, with new data being added to the system at speeds up to 70,000 calls per second![1] Not only do these companies have to store and manage immense collections of data, they have to be able to find any given fact in that data quickly. Consider the case of Internet search staple Google. While Google is reluctant to disclose many details about its data storage specifications, it is estimated that the company responds to over 91 million searches per day across a collection of data that is several terabytes in size. Impressively, the results of these searches are available almost instantly.

How can these businesses process this much data? How can they store it all, and then quickly retrieve just the facts that decision makers want to know, just when they want to know it? The answer is that they use databases. Databases, as explained in detail throughout this book, are specialized structures that allow computer-based systems to store, manage, and retrieve data very quickly. Virtually all modern business systems rely on databases; therefore, a good understanding of how these structures are created and their proper use is vital for any information systems professional. Even if your career does not take you down the amazing path of database design and development, databases will be a key component underpinning the systems that you use. In any case, you will probably make decisions in your career based on information generated from data. Thus, it is important that you know the difference between data and information.

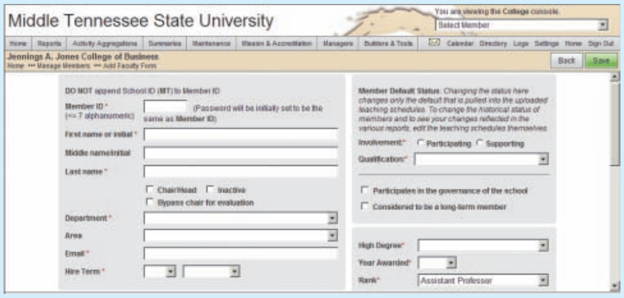
1.2 Data vs. Information

To understand what drives database design, you must understand the difference between data and information. **Data** are raw facts. The word raw indicates that the facts have not yet been processed to reveal their meaning. For example, suppose that a university tracks data on faculty members for reporting to accrediting bodies. To get the data for each faculty member into the database, you would provide a screen to allow for convenient data entry, complete with drop-down lists, combo boxes, option buttons, and other data-entry validation controls. Figure 1.1(a) shows a simple data-entry form from a software package named Sedona. When the data are entered into the form and saved, they are placed in the underlying database as raw data, as shown in Figure 1.1(b). Although you now have the facts in hand, they are not particularly useful in this format. Reading through hundreds of rows of data for faculty members does not provide much insight into the overall makeup of the faculty. Therefore, you transform the raw data into a data summary like the one shown in Figure 1.1(c). Now you can get quick answers to questions such as "What percentage of the faculty in the Information Systems (INFS) department are adjuncts?" In this case, you can quickly determine that 20 percent of the INFS faculty members are adjunct faculty. Because graphics can enhance your ability to quickly extract meaning from data, you show the data summary pie chart in Figure 1.1(d).

[1] "Top Ten Largest Databases in the World," *Business Intelligence Lowdown,* February 15, 2007, *http://www.businessintelligencelowdown.com/2007/02/top_10_largest_.html.*

FIGURE 1.1 Transforming raw data into information

a) Data entry screen



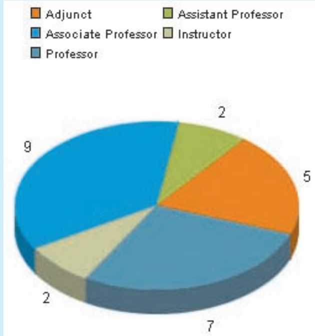
b) Raw data

Id	LastName	MidName	FirstName	DeptCode	Office	Email	Rank	HireYear	Degree
1	Washingtor	A.	George	MGMT	N135	gwashington@mtsu.edu	Professor	2001	Ph.D.
2	Adams		John	FIN	N313	jadams@mtsu.edu	Professor	1984	Ph.D.
3	Jefferson	L.	Thomas	ECON		tjefferson@mtsu.edu	Instructor	2002	M.B.A.
4	Madison	D.	James	FIN	N236	jmadison@mtsu.edu	Associate Professor	1994	Ph.D.
5	Monroe	N.	James	ACCT	N411	jmonroe@mtsu.edu	Assistant Professor	1995	Ph.D.
6	Adams	Q.	John	ACCT	N418	jqadams@mtsu.edu	Associate Professor	1989	Ph.D.
7	Jackson	C.	Andrew	ECON	N303	ajackson@mtsu.edu	Associate Professor	1999	Ph.D.
8	Van Buren	T.	Martin	FIN	N306	mvanburen@mtsu.edu	Professor	1988	Ph.D.
9	Harrison	R.	William	MKTG	N118	wharrison@mtsu.edu	Professor	1994	Ph.D.
10	Tyler	M.	John	MGMT		Jtyler@mtsu.edu	Assistant Professor	2000	Ed.D.
11	Polk		Cheryl	MKTG	N143	cpolk@mtsu.edu	Associate Professor	2002	Ph.D.
12	Taylor	G.	Zachary	ACCT	N415	ztaylor@mtsu.edu	Associate Professor	1996	Ph.D.
13	Fillmore		Millard	JCB	N219	mfillmore@mtsu.edu	Professor	1992	Ph.D.
14	Pierce	A.	Franklin	MKTG	N359	pfranklin@mtsu.edu	Instructor	2005	M.B.A.
15	Buchanan	T.	James	MGMT	N146	jbuchanan@mtsu.edu	Associate Professor	1996	D.B.A.
17	Lincoln	W.	Larry	MGMT	N150	llincoln@mtsu.edu	Associate Professor	1996	Ph.D.
18	Johnson		Andrew	ISYS	N360	ajohnson@mtsu.edu	Professor	1987	Ph.D.
19	Grant		Katie	MKTG	N120	kgrant@mtsu.edu	Assistant Professor	1989	D.B.A.
20	Rutherford		Hayes	ACCT	N408	hrutherford@mtsu.edu	Professor	1992	Ph.D.
21	Grefield	T.	Denise	ACCT		dgerfield@mtsu.edu	Assistant Professor	2018	Ph.D.
22	Arthur		Emily	ACCT	N413	earthur@mtsu.edu	Associate Professor	2003	J.D.
23	Cleveland	G.	Robert	ACCT	N401	rcleveland@mtsu.edu	Associate Professor	1997	Ph.D.
24	Harrison	X.	Patricia	BULA	N406	pharrison@mtsu.edu	Associate Professor	2001	J.D.
25	McKinley	B.	Priscilla	ISYS	N363	pmckinley@mtsu.edu	Adjunct	1994	M.S.
26	Roosevelt	F.	Hillary	MGMT	N104	hroosevelt@mtsu.edu	Associate Professor	2002	Ph.D.
27	Wilson		Laura	BCEN	N448	lwilson@mtsu.edu	Professor	1992	Ph.D.
28	Harding		Warren	MKTG	N114	wharding@mtsu.edu	Professor	1984	Ed.D.
29	Coolidge		Calvin	ECON	N316	ccoolidge@mtsu.edu	Professor	1975	Ph.D.
30	Hoover		Lisa	MGMT		lhoover@mtsu.edu	Adjunct	1978	M.B.A.
31	Truman		Betty	ACCT	N416	btruman@mtsu.edu	Professor	1971	Ed.D.
32	Johnson		Robert	BCEN	N240	rjohnson@mtsu.edu	Professor	2001	Ph.D.

c) Information in summary format

Rank	COUNT	%/INFS	TOT/COL	%/COL. TOT.	%/COL. FAC.
Adjunct	5	20.00%	23	21.74%	3.27%
Assistant Professor	2	8.00%	28	7.14%	1.31%
Associate Professor	9	36.00%	37	24.32%	5.88%
Instructor	2	8.00%	18	11.11%	1.31%
Professor	7	28.00%	47	14.89%	4.58%

d) Information in graphical format



Information is the result of processing raw data to reveal its meaning. Data processing can be as simple as organizing data to reveal patterns or as complex as making forecasts or drawing inferences using statistical modeling. To reveal meaning, information requires *context*. For example, an average temperature reading of 105 degrees does not mean much unless you also know its context: Is this reading in degrees Fahrenheit or Celsius? Is this a machine temperature, a body temperature, or an outside air temperature? Information can be used as the foundation for decision making. For example, the data summary for the faculty can provide accrediting bodies with insights that are useful in determining whether to renew accreditation for the university.

Keep in mind that raw data must be properly *formatted* for storage, processing, and presentation. For example, dates might be stored in Julian calendar formats within the database, but displayed in a variety of formats, such as day-month-year or month/day/year, for different purposes. Respondents' yes/no responses might need to be converted to a Y/N or 0/1 format for data storage. More complex formatting is required when working with complex data types, such as sounds, videos, or images.

In this "information age," production of accurate, relevant, and timely information is the key to good decision making. In turn, good decision making is the key to business survival in a global market. We are now said to be entering the "knowledge age."[2] Data are the foundation of information, which is the bedrock of **knowledge**—that is, the body of information and facts about a specific subject. Knowledge implies familiarity, awareness, and understanding of information as it applies to an environment. A key characteristic of knowledge is that "new" knowledge can be derived from "old" knowledge.

[2] Peter Drucker coined the phrase "knowledge worker" in 1959 in his book *Landmarks of Tomorrow.* In 1994, Esther Dyson, George Keyworth, and Dr. Alvin Toffler introduced the concept of the "knowledge age."

Let's summarize some key points:

- Data constitute the building blocks of information.
- Information is produced by processing data.
- Information is used to reveal the meaning of data.
- Accurate, relevant, and timely information is the key to good decision making.
- Good decision making is the key to organizational survival in a global environment.

Timely and useful information requires accurate data. Such data must be properly generated and stored in a format that is easy to access and process. And, like any basic resource, the data environment must be managed carefully. **Data management** is a discipline that focuses on the proper generation, storage, and retrieval of data. Given the crucial role that data play, it should not surprise you that data management is a core activity for any business, government agency, service organization, or charity.

1.3 Introducing the Database

Efficient data management typically requires the use of a computer database. A **database** is a shared, integrated computer structure that stores a collection of the following:

- End-user data—that is, raw facts of interest to the end user.
- **Metadata**, or data about data, through which the end-user data are integrated and managed.

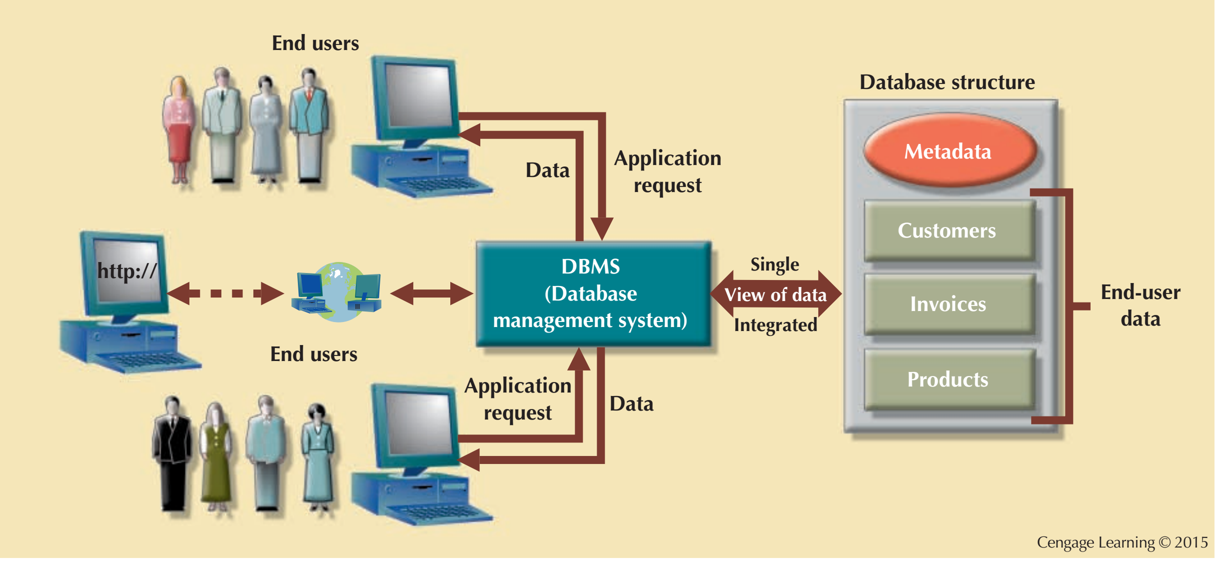
The metadata describe the data characteristics and the set of relationships that links the data found within the database. For example, the metadata component stores information such as the name of each data element, the type of values (numeric, dates, or text) stored on each data element, and whether the data element can be left empty. The metadata provide information that complements and expands the value and use of the data. In short, metadata present a more complete picture of the data in the database. Given the characteristics of metadata, you might hear a database described as a "collection of *self-describing* data."

A **database management system (DBMS)** is a collection of programs that manages the database structure and controls access to the data stored in the database. In a sense, a database resembles a very well-organized electronic filing cabinet in which powerful software (the DBMS) helps manage the cabinet's contents.

1.3.1 Role and Advantages of the DBMS

The DBMS serves as the intermediary between the user and the database. The database structure itself is stored as a collection of files, and the only way to access the data in those files is through the DBMS. Figure 1.2 emphasizes the point that the DBMS presents the end user (or application program) with a single, integrated view of the data in the database. The DBMS receives all application requests and translates them into the complex operations required to fulfill those requests. The DBMS hides much of the database's internal complexity from the application programs and users. The application program might be written by a programmer using a programming language, such as Visual Basic.NET, Java, or C#, or it might be created through a DBMS utility program.

FIGURE 1.2 The DBMS manages the interaction between the end user and the database



Having a DBMS between the end user's applications and the database offers some important advantages. First, the DBMS enables the data in the database *to be shared* among multiple applications or users. Second, the DBMS *integrates* the many different users' views of the data into a single all-encompassing data repository.

Because data are the crucial raw material from which information is derived, you must have a good method to manage such data. As you will discover in this book, the DBMS helps make data management more efficient and effective. In particular, a DBMS provides these advantages:

- *Improved data sharing*. The DBMS helps create an environment in which end users have better access to more and better-managed data. Such access makes it possible for end users to respond quickly to changes in their environment.
- *Improved data security*. The more users access the data, the greater the risks of data security breaches. Corporations invest considerable amounts of time, effort, and money to ensure that corporate data are used properly. A DBMS provides a framework for better enforcement of data privacy and security policies.
- *Better data integration*. Wider access to well-managed data promotes an integrated view of the organization's operations and a clearer view of the big picture. It becomes much easier to see how actions in one segment of the company affect other segments.
- *Minimized data inconsistency*. **Data inconsistency** exists when different versions of the same data appear in different places. For example, data inconsistency exists when a company's sales department stores a sales representative's name as Bill Brown and the company's personnel department stores that same person's name as *William G. Brown*, or when the company's regional sales office shows the price of a product as $45.95 and its national sales office shows the same product's price as $43.95. The probability of data inconsistency is greatly reduced in a properly designed database.
- *Improved data access*. The DBMS makes it possible to produce quick answers to ad hoc queries. From a database perspective, a **query** is a specific request issued to the DBMS for data manipulation—for example, to read or update the data. Simply put, a query is a question, and an **ad hoc query** is a spur-of-the-moment question. The DBMS sends back an answer (called the **query result set**) to the application. For example, when dealing with large amounts of sales data, end users might want quick answers to questions (ad hoc queries). Some examples include the following:
  - What was the dollar volume of sales by product during the past six months?

- What is the sales bonus figure for each of our salespeople during the past three months?
- How many of our customers have credit balances of $3,000 or more?

- *Improved decision making*. Better-managed data and improved data access make it possible to generate better-quality information, on which better decisions are based. The quality of the information generated depends on the quality of the underlying data. **Data quality** is a comprehensive approach to promoting the accuracy, validity, and timeliness of the data. While the DBMS does not guarantee data quality, it provides a framework to facilitate data quality initiatives. Data quality concepts will be covered in more detail in Chapter 15, Database Administration and Security.
- *Increased end-user productivity*. The availability of data, combined with the tools that transform data into usable information, empowers end users to make quick, informed decisions that can make the difference between success and failure in the global economy.

The advantages of using a DBMS are not limited to the few just listed. In fact, you will discover many more advantages as you learn more about the technical details of databases and their proper design.

1.3.2 Types of Databases

A DBMS can be used to build many different types of databases. Each database stores a particular collection of data and is used for a specific purpose. Over the years, as technology and innovative uses of databases have evolved, different methods have been used to classify databases. For example, databases can be classified by the number of users supported, where the data are located, the type of data stored, the intended data usage, and the degree to which the data are structured.

The number of users determines whether the database is classified as single user or multiuser. A **single-user database** supports only one user at a time. In other words, if user A is using the database, users B and C must wait until user A is done. A single-user database that runs on a personal computer is called a **desktop database**. In contrast, a **multiuser database** supports multiple users at the same time. When the multiuser database supports a relatively small number of users (usually fewer than 50) or a specific department within an organization, it is called a **workgroup database**. When the database is used by the entire organization and supports many users (more than 50, usually hundreds) across many departments, the database is known as an **enterprise database**.

Location might also be used to classify the database. For example, a database that supports data located at a single site is called a **centralized database**. A database that supports data distributed across several different sites is called a **distributed database**. The extent to which a database can be distributed and the way in which such distribution is managed are addressed in detail in Chapter 12, Distributed Database Management Systems.

Both centralized and decentralized (distributed) databases require a well-defined infrastructure (hardware, operating systems, network technologies, etc.) to implement and operate the database. Typically, the infrastructure is owned and maintained by the organization that creates and operates the database. But in recent years the use of cloud databases has been growing in popularity. A **cloud database** is a database that is created and maintained using cloud data services, such as Microsoft Azure or Amazon's AWS. These services, provided by third-party vendors, provide defined performance measures (data storage capacity, required throughput, and availability) for the database, but do not necessarily specify the underlying infrastructure to implement it. The data owner does not have to know, or be concerned about, what hardware and software is being used to support their database. The performance capabilities can be renegotiated with the cloud provider as the business demands on the database change. For example, during the 2012 presidential election in the United States, the Obama campaign used a cloud database hosted on infrastructure capabilities purchased from Amazon. The campaign did not have to buy, install, configure, or maintain any hardware, operating systems, or network devices. It simply purchased storage and processing capacity for its data and applications. As the demands on the database increased, additional processing and storage capabilities could be purchased as needed.

In some contexts, such as research environments, a popular way of classifying databases is according to the type of data stored in them. Using this criterion, databases are grouped into two categories: general-purpose and discipline-specific databases. **General-purpose databases** contain a wide variety of data used in multiple disciplines—for example,