Databases Illuminated

Second Edition

Catherine Ricardo, Iona College

JONES & BARTLETT
LEARNING

World Headquarters
Jones & Bartlett Learning
40 Tall Pine Drive
Sudbury, MA 01776
978-443-5000
info@jblearning.com
www.jblearning.com

Jones & Bartlett Learning Canada
6339 Ormindale Way
Mississauga, Ontario L5V 1J2
Canada

Jones & Bartlett Learning International
Barb House, Barb Mews
London W6 7PA
United Kingdom

Jones & Bartlett Learning books and products are available through most bookstores and online booksellers. To contact Jones & Bartlett Learning directly, call 800-832-0034, fax 978-443-8000, or visit our website, www.jblearning.com.

Substantial discounts on bulk quantities of Jones & Bartlett Learning publications are available to corporations, professional associations, and other qualified organizations. For details and specific discount information, contact the special sales department at Jones & Bartlett Learning via the above contact information or send an email to specialsales@jblearning.com.

Production Credits
Publisher: Cathleen Sether
Senior Acquisitions Editor: Timothy Anderson
Senior Editorial Assistant: Stephanie Sguigna
Production Director: Amy Rose
Senior Marketing Manager: Andrea DeFronzo
Associate Marketing Manager: Lindsay White
V.P., Manufacturing and Inventory Control: Therese Connell
Composition: Northeast Compositors, Inc.
Cover and Title Page Design: Kristin E. Parker
Cover Image: © ErickN/ShutterStock, Inc.
Printing and Binding: Malloy, Inc.
Cover Printing: Malloy, Inc.

Library of Congress Cataloging-in-Publication Data
Ricardo, Catherine M.
Databases illuminated / Catherine M. Ricardo. — 2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-4496-0600-8 (casebound)
1. Database management. I. Title.
QA76.9.D3R514 2011
005.74—dc22
2010052808

6048
Printed in the United States of America
15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Dedication

To my beloved husband, Henry

Contents

Preface

Purpose of This Book

Since the publication of the first edition of this book in 2004, databases have become an even more central part of the framework of modern computer systems, and the study of database systems, design, and management is an essential part of the education of computer science and information science students. A database course should provide a strong theoretical background, practice in database design, and experience creating and developing a working database. This book is designed to help students integrate theoretical material with practical knowledge, using an approach that has a firm theoretical basis applied to practical database implementation.

Structure

Theoretical foundations are presented early and the concepts are used repeatedly throughout the book, including the chapters that deal with implementation. Logical database design is given full consideration. The entity-relationship model is introduced early and then mapped to the relational model. Relational normalization is studied in detail, and many examples of the normalization process are discussed. The object-oriented model is presented using UML as a vehicle for logical design. The enhanced entity-relationship model is presented, and mapped to both relational and object-relational models. XML and the semi-structured data model are introduced in Chapter 13. A continuing example of a university database is incorporated throughout the text, to illustrate concepts and techniques and to provide both continuity and contrast. Other

examples are presented as needed. Purely relational, object-relational, and object-oriented database systems are described and used for implementation of the examples. Details of database management systems are described so that students can learn the specifics of these real-life systems, down to the physical implementation level. Microsoft® Access® is used initially in the examples, but Oracle® is introduced as the material is developed. However, the examples are suitable for use with any relational or object-relational DBMS. InterSystems Caché® DBMS is used to illustrate object-oriented databases.

Purpose of the Sample Project and Student Projects

The various approaches are integrated in a sample project that is a unique feature of this book. Starting at the end of the first chapter, students can visit the Student Companion Website to see a sample project that is developed as the book progresses. This project is independent of the examples that are presented in the chapter text or in exercises. The design begins with a description of the application: a database needed by an art gallery. At the end of the first chapter, students see how the information needs are specified. After the students study planning techniques in Chapter 2, they see how to create a complete user-oriented data dictionary and how other planning tools could be used. When they learn about logical models, they see the step-by-step development of an E-R diagram at the end of Chapter 3. They see the steps involved in mapping the diagram to the relational model after Chapter 4, and normalizing the model after Chapter 6. Complete details of creating and manipulating a purely relational Oracle database using SQL are presented for the relational model after Chapter 5. The E-R diagram is expanded to a complete EE-R diagram and mapped to an object-relational database after Chapter 8, using the object-relational features of Oracle. The EE-R diagram is transformed into a UML diagram after Chapter 7, which includes the design and creation of an object-oriented database using ODL and also InterSystems Caché DBMS. The sample project is extended after Chapter 12 to show a distributed design that might be used in a client-server environment or a true distributed environment. Details of creating a simple website for the art gallery are provided after Chapter 13 using PL/SQL, JDBC, Caché, and XML. Thus, every important technique of planning, design, and implementation is illustrated using real-life systems. The sample pro-

ject section is always connected to continuing student projects, which require students to emulate the steps of the sample.

From my experiences assigning similar projects while teaching database courses, I have found that students learn most effectively by creating a working database and developing their own continuing projects to incorporate new concepts, and that they benefit by seeing samples as they progress. Such realistic experiences provide insights that no amount of lecturing can produce. Later chapters deal with topics of database security, concurrency control, recovery techniques, query optimization, distributed databases, social and ethical issues, data warehouses, and data mining. Online appendices cover physical data organization, the network model, the hierarchical model, and other specialized material.

Learning Features

The writing style is conversational. Each chapter begins with a statement of learning objectives. Examples and applications are presented throughout the text. Illustrations are used both to clarify the material and to vary the presentation. Exercises, including lab exercises where appropriate, appear at the end of each chapter and on the Student Companion Website, with solutions provided in the Instructor's Guide. The sample project on the Student Companion Website is an important part of the text, providing an application of the material just presented. The student projects that follow the sample are introduced after the first chapter, and students are expected to choose one, or have one assigned, and to develop that project to parallel the sample as they progress. Student projects may be done individually or in a group. Access, Oracle, and Caché implementations of the example databases used in the text, and of the sample project, are available on the book's website; solutions for the student projects, since they are intended to be used as assignments, are not included there. Resources for student laboratory exercises are also available on the website. Chapter summaries are included in the text to provide a rapid review or preview of the material and to help students understand the relative importance of the concepts presented. The Instructor's Guide contains Microsoft® PowerPoint® presentations for each chapter, solutions for the student projects, copies of figures, full statements of objectives for each chapter, alternative student projects, quizzes, chapter tests, comprehensive examinations for multiple chapters, and solutions to exercises.

Audience

The material is suitable for junior or senior computer science majors or information science majors with a good technical background. Students should have completed at least one year of programming, including data structures. The book could also be used as an introductory database text for graduate students or for self study.

Mapping to Curriculum Guidelines

Although this book was based on the author's experiences, it fits the ACM-IEEE Curriculum Guidelines well. It covers the topics and supports the learning objectives listed in the IS 2010 Curriculum Guidelines for IS 2010.2, Data and Information Management. It also supports the guidelines of ACM Computer Science Curriculum 2008, An Interim Revision of CS2001. This revision retains the basic structure and guidelines of the 2001 recommendations, while including some new trends that have emerged since the first report. These include more attention to security, concurrency, and net-centric computing, all of which are addressed in this text. CC2001 provides a model database course, CS270T, for the topic-based curriculum. The course includes three areas, Human-Computer Interaction (HCI), Information Management (IM), and Social and Professional Issues (SP). The units included in CS270T are listed below, along with the corresponding chapters in the book.

IM1 Information models and systems (3 core hours)—Chapters 1, 2

IM2 Database systems (3 core hours)—Chapter 2

IM3 Data modeling (4 core hours)—Sections 2.7, 13.4; Chapters 3, 4, 7, 8, Appendices B, C

IM4 Relational databases (5 hours)—Chapter 4

IM5 Database query languages (4 hours)—Sections 4.5, 7.5, 8.7; Chapters 6, 11, 13; Appendices B, C

IM6 Relational database design (4 hours)—Chapter 6

IM7 Transaction processing (3 hours)—Chapter 10

IM8 Distributed databases (3 hours)—Chapters 12, 13

IM9 Physical database design (3 hours)—Appendix A

HCI Foundations of human–computer interaction (2 core hours)—Chapter 14

SP6 Intellectual property (3 core hours)—Chapter 14

SP7 Privacy and civil liberties (2 core hours)—Chapters 9, 14

Elective topics (1 hour)—Chapter 15

In addition to the text's coverage, the sample project and student projects also provide practice in data modeling, database design, human–computer interaction, relational databases, database query languages, distributed databases, web-based databases, and physical database design. Some aspects of security, privacy, and civil liberties are discussed in the sample project, and similar issues will arise and should be treated in the student projects.

Acknowledgments

Many people have offered useful comments, advice, and encouragement during the writing of this book. I am grateful to students and colleagues throughout the world who used the first edition and provided comments that have helped shape this book, especially Mary Courtney, Pace University and Stephen D'Alessio and Smiljana Petrovic, Iona College.

For his encouragement and support for this project, I am grateful to Tim Anderson, Senior Acquisitions Editor at Jones & Bartlett Learning. I would also like to thank the editorial and production staff at Jones & Bartlett Learning, especially Amy Rose.

Finally I would like to thank my family, especially my husband, Henry, for his love, patience, and understanding and for his active help in critiquing and proofreading the manuscript.