Fundamentals of Database Systems

FUNDAMENTALS OF DATABASE SYSTEMS

Ramez Elmasri
Department of Computer Science
University of Houston

Shamkant B. Navathe
Database Systems Research and Development Center
Department of Computer and Information Sciences
University of Florida

The Benjamin/Cummings Publishing Company, Inc.
Redwood City, California • Fort Collins, Colorado
Menlo Park, California • Reading, Massachusetts • New York
Don Mills, Ontario • Wokingham, U.K. • Amsterdam • Bonn
Sydney • Singapore • Tokyo • Madrid • San Juan

Sponsoring Editor: Alan R. Apt
Associate Editor: Mark McCormick
Production Editor: Mary B. Shields
Text and Cover design: Hal Lockwood
Copy Editor: Mary Prescott
Composition: Graphic Typesetting Service/Coordinator: Sharon Squires
Typesetter: Rebecca Herren

Cover Art: "Sandstone". Original hand pulled limited edition serigraph by Tetsuro Sawada. Image size, 35" × 22-3/4". Edition size, 60.
With thanks to Galerie de Metropolitan, La Jolla, CA. Exclusive Sawada publisher and distributor: Buschelen/Mowatt Fine Arts Ltd. 111-1445 West Georgia Street, Vancouver, Canada V6G 2T3.

The basic text of this book was designed using the Modular Design System, as developed by Wendy Earl and Design Office Bruce Kortebein.

Library of Congress Cataloging-in-Publication Data

Elmasri, Ramez.
Fundamentals of database systems.

Bibliography: p. 751
Includes indexes.
1. Data base management. 2. Data base design. I. Navathe, Sham. II. Title.
QA76.9.D3E57 1989 005.74 88–35001

ISBN 0-8053-0145-3

EFGHIJ-MU- 93210

The Benjamin/Cummings Publishing Company, Inc.
390 Bridge Parkway
Redwood City, California 94065

Preface

The purpose of this book is to introduce the fundamental concepts necessary for the design, use, and implementation of database systems. Our presentation stresses the fundamentals of database modeling and design, the languages and facilities provided by database management systems, and the techniques for implementing database systems. The book is meant to be used as a textbook for a one or two semester course in database systems at the junior, senior, or graduate levels, and as a reference book. We assume that readers are familiar with elementary programming and data structuring concepts, and have some exposure to basic computer organization. To keep the book as self-contained as possible we discuss a few elementary topics that some readers may be familiar with; for example, the discussion on characteristics of disk storage devices in Chapter 4 and the summary of tree data structures necessary for the understanding of indexing in Chapter 5.

We have chosen to start the book with the presentation of the concepts at both ends of the database spectrum—the conceptual modeling concepts and the physical file storage techniques. We believe that these concepts are essential to achieving a good understanding of database systems. For students who have already taken a course on file organization techniques, parts of Chapters 4 and 5 could be assigned as reading material to review file organization concepts. Chapter 3, which covers conceptual modeling using the Entity-Relationship (ER) model, provides an important conceptual understanding of data. However, it may be left out, or covered later if the instructor so wishes.

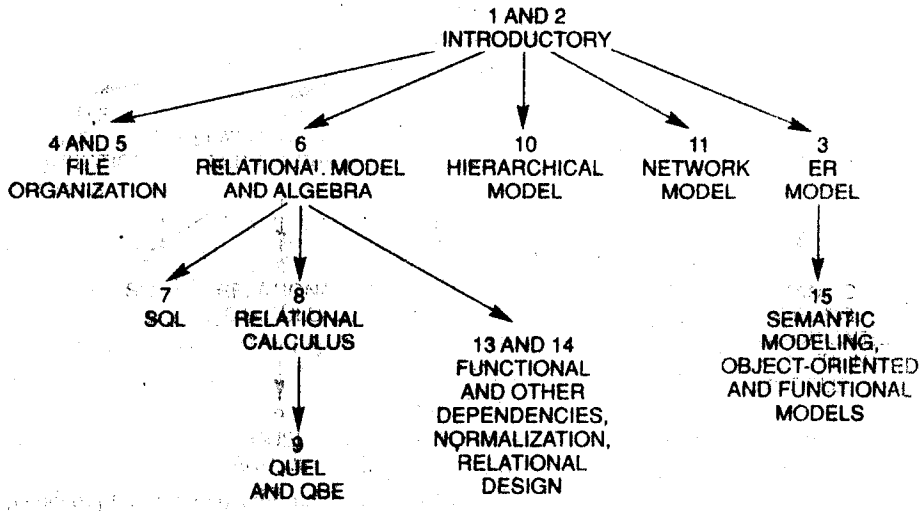
Key Features of the Book

We would like to point out several key features of *Fundamentals of Database Systems*. These are the following:

1. Coverage of data models: We offer comprehensive coverage of the relational, network, and hierarchical data models. In addition, we include in-depth presenta-

tions of the ER model (Chapter 3) and semantic modeling concepts (Chapter 15). We also discuss object-oriented and functional data modeling concepts in Chapter 15. In Chapter 12, we present the similarities and differences among the classical data models. We also show how an ER schema can be mapped to relational, network, or hierarchical schemas, and continue the mapping procedure in Chapter 15 to include advanced semantic modeling concepts. We have tried to keep the individual chapters as self-contained as possible. This modular organization will enable instructors to select the chapters they wish to cover, and to cover them in their preferred order. In particular, the ER model and the three data models presented in Part II—the relational, network, and hierarchical models—can be taught in any order. The instructor may choose not to cover some of these data models at all without sacrificing continuity.

2. Examples: We use the same example of a COMPANY database throughout the book so the reader can compare the different approaches using the same database application. The COMPANY database was structured in a way to illustrate the fine points of the various data models and query languages. The same set of queries is demonstrated with different languages, and queries are labelled for easy cross referencing. Several additional example databases are introduced in the exercises.
3. Flexibility: Because it is very difficult to cover all the data models in depth in a single course, we present each model individually without reference to the other models. Individual instructors may cover their preferred models in their favorite order. In addition, file organization concepts may be covered in depth early on in a course, left out entirely, or presented later. The simplified dependency chart below shows which chapter sequences can be covered following the introductory chapters 1 and 2. (Later chapters are not included on the dependency chart to maintain simplicity.)



Dependency Chart for chapters up to Chapter 15.

4. Concepts before systems: We cover the conceptual aspects of each major approach to database management before we give examples of real database management systems. This approach is pedagogically sound because systems are often in a state of flux whereas basic concepts are more stable. We have found that our students grasp the concepts better when they are presented independently of real systems, which often have their own idiosyncracies. Chapter 23 provides coverage of several commercial DBMSs, including a relational, a hierarchical, a network, and an object-oriented system. In Chapter 23, we discuss the similarities and differences between the concepts presented earlier and their implementation in the actual systems. This material can be integrated into the course wherever the instructor feels is appropriate.
5. Coverage of database design techniques and DBMS system implementation concepts: Database design techniques, both theoretical and practical, are covered in Part III. Implementation concepts and techniques for query optimization, concurrency control, recovery, security, and integrity are covered in Part IV. The functions of the DBMS catalog/data dictionary are discussed in Chapter 17. The concept of a database transaction is introduced in Chapter 19.
6. State-of-the-art coverage: Coverage of recent advances in database systems is included. Semantic modeling and object-oriented concepts are presented in Chapter 15. Distributed databases are covered in Chapter 21. Chapter 22 includes a discussion of a number of recent advances in database technology and applications, such as expert database systems, knowledge bases, multi-media databases, databases for engineering design, and office information systems.
7. Class testing and comprehensive reviewing: The notes and various drafts upon which this book is based have been class tested over the past three years at both the University of Houston and University of Florida. The text has been thoroughly reviewed by numerous database systems experts who provided many suggestions that contributed to major improvements over the course of three drafts of the book.

Contents of *Fundamentals of Database Systems*

PART I describes the basic concepts necessary for a good understanding of database design and implementation. The first two chapters introduce databases, their typical users, and DBMS concepts and architecture. In Chapter 3, we discuss the conceptual design of databases using the concepts of the Entity-Relationship (ER) model. Chapters 4 and 5 show how database files are organized at the physical storage level. Chapter 4 describes the primary methods of organizing files of records on disk; Chapter 5 describes indexing techniques for files, including B-tree and B^+-tree data structures.

PART II describes the data models and languages used in the majority of current commercial database systems. These models are the relational, network, and hierarchical data models. The material on the relational model is covered in several chapters—Chapters 6 through 9—because it is becoming the preferred model, both in industry and for formalizing database issues in academic research. We cover in detail the formal rela-

tional algebra and calculus, as well as a number of significant commercial relational languages that have been implemented—SQL, QUEL, and QBE. The hierarchical and network data models are covered in Chapters 10 and 11, respectively, independently of specific DBMSs. Chapter 12 compares the data models and corresponding DBMSs based on their modeling concepts, languages, integrity constraints, and storage structures, and also shows how to convert the conceptual design of a database schema in the ER model into relational, network, or hierarchical schemas.

PART III covers database design. First, we cover the formalisms, theory, and algorithms developed for relational database design in Chapters 13 and 14. This material includes functional and other types of dependencies and normal forms for relations. Step by step intuitive normalization is presented in Chapter 13, and formal relational design algorithms are given in Chapter 14. Chapter 15 discusses data abstraction and semantic data modeling concepts, and briefly compares these to knowledge representation techniques. The ER model is extended to incorporate these ideas, leading to the enhanced-ER (EER) data model. The concepts presented include subclasses, specialization, generalization, and categories. We also describe the object-oriented and functional approaches to data modeling. Chapter 16 presents an overview of the different phases of the database design process for medium-sized and large organizations, and also discusses physical database design issues pertinent to relational, network, and hierarchical DBMSs.

PART IV discusses techniques used in the implementation of database management systems (DBMSs). Chapter 17 describes implementation of the DBMS catalog, which is a vital part of any DBMS. Chapter 18 presents the techniques used for processing and optimizing queries specified in a high-level database language. Chapter 19 introduces the concept of a transaction and discusses concurrency control and recovery techniques for multi-user DBMSs. Chapter 20 discusses techniques for specifying and maintaining database security constraints and semantic integrity constraints.

PART V includes two chapters. In Chapter 21, we discuss distributed databases, where the database and the DBMS are distributed over many sites connected via a communication network. With powerful workstations and high speed communication networks, truly distributed databases are becoming viable. Chapter 22 surveys the trends in database technology and includes discussions of several emerging database applications and technologies, including expert database systems, databases for computer-aided design applications, multi-media databases, and office information systems.

Finally, PART VI describes some characteristics of several representative commercial DBMSs. The DBMSs surveyed include IBM's DB2 relational system, the IMS hierarchical system, and Cullinet's IDMS network system. We also present features of the Vbase DBMS as an example of an object-oriented system.

Guidelines for Using *Fundamentals of Database Systems*

There are many different ways to teach a database course. The chapters in Part I and II can be used in an introductory course on database systems in the order they are given or in the preferred order of each individual instructor. Selected chapters may be left out, and the instructor can add other chapters from the rest of the book, depending on the emphasis of the course. For an emphasis on system implementation techniques, selected

chapters from Part IV can be used. For an emphasis on database design, chapters from Parts III can be used. Examples to illustrate the use of specific systems can be included from Part VI.

For a single-semester course, based on this book, some chapters can be assigned as reading material. Chapters 4, 5, 12, 16, 17, and 22 can be considered for such reading assignments. The book can also be used for a two-semester sequence. The first course, "Introduction to Database Systems," at the sophomore, junior, or senior level, could cover most of Chapters 1 to 13. The second course, "Database Design and Implementation Techniques," at the senior or first year graduate level, can cover the remaining chapters, and any chapters left out from the first course. Part V can serve as introductory material for additional topics the instructor may wish to cover. Chapters from Part VI can be used selectively in either semester, and material describing the DBMS at the local institution can be covered in addition to the material in the book.

Acknowledgements

It is a great pleasure for us to acknowledge the assistance and contributions of a large number of individuals to this effort. First and foremost, we owe a great deal to our editor Alan Apt for his persistent personal interest in the project from its inception and his constant encouragement. He was able to solicit comments from an eminent array of database professionals, and provided crucial feedback that influenced the contents and organization of the book.

Many individuals have reviewed the outlines and various drafts of the manuscript for *Fundamentals of Database Systems*. We are very grateful to Scott Downing, Dennis Heimbigner, Julia Hodges, Yannis Ioannidis, Jim Larson, Per-Ake Larson, Rahul Patel, David Stemple, and Kyu-Young Whang, who have reviewed one or more drafts of the manuscript. They provided us with many helpful and insightful comments that influenced the final manuscript. In particular, we are indebted to Jim Larson who provided pages of constructive criticism on every chapter. Dennis Heimbigner's comments contributed to the final contents of several chapters, and in particular to the order of presentation of relational algebra operations in Chapter 6. Kyu-Young Whang's comments on Chapters 18 and 19 were very helpful. We would like to acknowledge the influence of Rafi Ahmed, Scott Downing, and Julia Hodges on Chapter 8, and Yannis Ioannidis on Chapter 9.

In addition to our main reviewers, we would like to thank several persons who reviewed our proposed outline or drafts of individual chapters of the book. These include Don Batory, Dennis McLeod, Nicholas Roussopoulos, Michael Stonebraker, and Frank Tompa. We are also grateful to Tim Andrews of Ontologic, and Vic Ghorpadey, Sam Greenlaw, and Jeff Scherb of Cullinet who reviewed the material on their respective commercial systems in Chapter 23. Any bugs that remain in the manuscript are, of course, the responsibility of the authors.

Ramez Elmasri used the drafts of various chapters numerous times as class notes at the University of Houston for the undergraduate database course COSC 3380 (and its predecessor course COSC 4320). In addition, selected chapters were used in the graduate

course COSC 6340. We are grateful to the students' comments on the manuscript. Unfortunately, we did not keep track of the names of students who provided feedback on the earlier drafts of the book. We would like to mention Christopher Lewis, Xing-Fang Lin, Srinivas Pyda, Ed Smelko, and Yu-Tung Wu, who looked at more recent drafts. Ramez Elmasri would also like to acknowledge the influence of students at the Institute for Retraining in Computer Science, where he taught the database course for two summers. Discussions with the students, particularly Joe Bergin and Wadi Juredeini, clarified some of the concepts presented in Chapters 13 and 14.

Sham Navathe used drafts of various chapters for his course CIS 6120 at the University of Florida. Students taking the course provided very useful feedback. We would particularly like to mention Subhasis Das, Mike Dybevick, Seong Geum, C.K. Luo, Jackie Morie, David Noden, Terry Sullivan, Mridul Tandon, and J.J. Yu. Many graduate students working with Navathe provided technical and editorial comments during the later revisions. We particularly thank Arunkumar Balaraman, Ajay Budhraja, Aloysius Cornelio, Dinesh Desai, Sunit Gala, Sunil George, Stephan Grill, K. Kamalakar, Sumant Pal, Sangam Pant, Bhavani Ravichandran, Prabhakar Satya, Yogesh Sridhare, and Arun Thakore.

Rajeev Kumar helped immensely in manuscript preparation for Chapters 22 and 23. Sharon Grant provided assistance in manuscript preparation and secretarial help throughout. Mohan Pillalamarri helped in index preparation. John Gibbs provided precious assistance with writing style during the early stages of the project. We would also like to thank Colleen Dunn, Mark McCormick, Mary Shields, and Mary Ann Telatnick of Benjamin Cummings Publishing Company.

Finally, Ramez Elmasri would like to thank his parents for their continuous encouragement and support, and his wife Amalia and children Ramy, Riyad, and Katrina who endured through the course of development of the book. Sham Navathe would like to thank his grandparents and parents for their encouragement and support throughout his career, and his wife Aruna and children Manisha and Amol for their sacrifice and patience during the project.

R.E.
S.B.N.

Contents