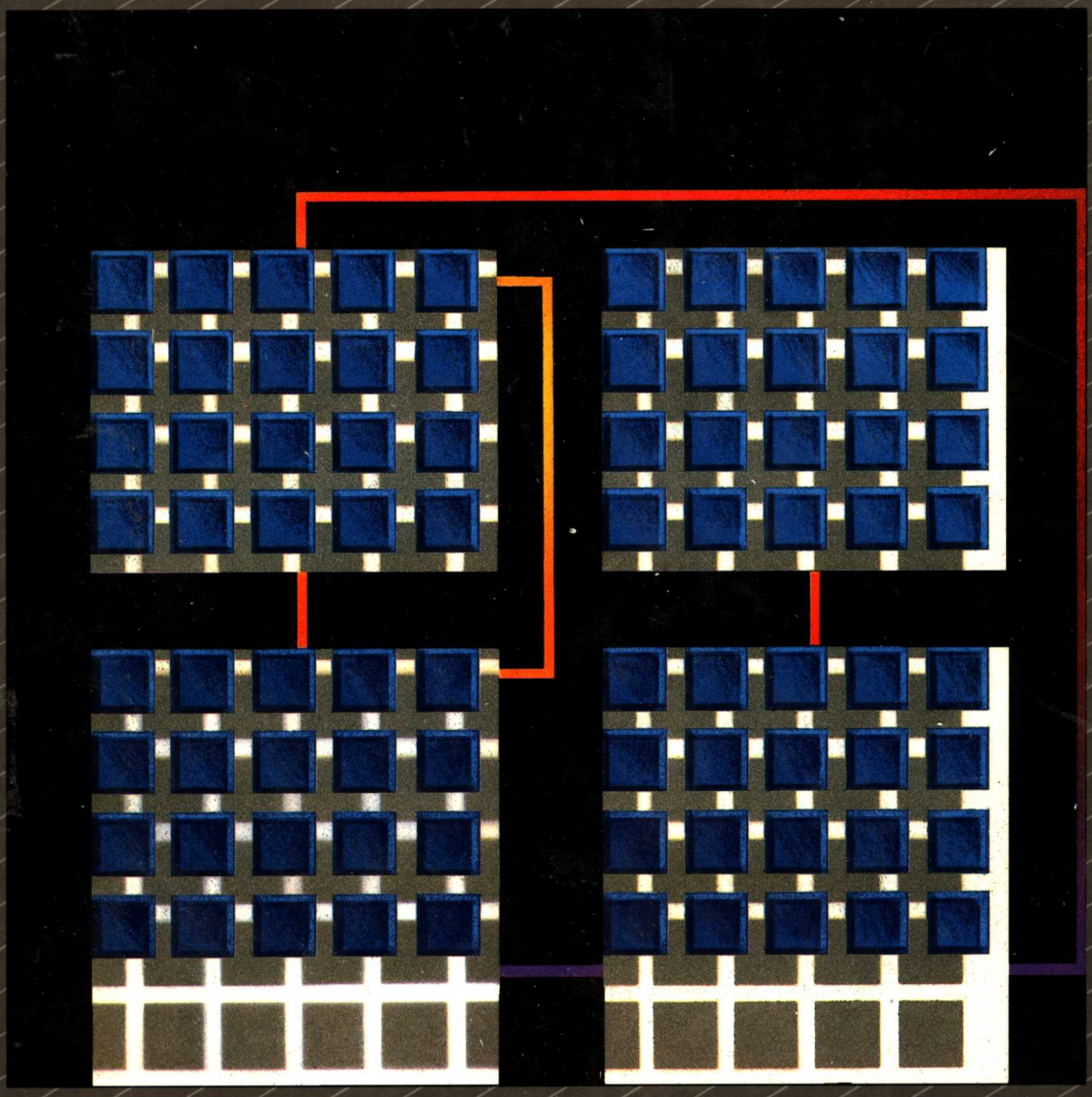
Database Management: Theory and Application



JOHN C. SHEPHERD

Database Management

Theory and Application

John C. Shepherd
Duquesne University

IRWIN

Homewood, IL 60430
Boston, MA 02116

Senior sponsoring editor: *Lawrence E. Alexander*
Developmental editor: *Rebecca J. Johnson*
Project editor: *Waivah Clement*
Production manager: *Carma W. Fazio*
Designer: *Image House*
Artist: *Benoit Design*
Compositor: *J.M. Post Graphics, Corp.*
Typeface: *10/12 Palatino*
Printer: *R.R. Donnelley & Sons Company*

Library of Congress Cataloging-in-Publication Data

Shepherd, John C.
Database management : theory and application / John C. Shepherd.
p. cm.
ISBN 0-256-07829-7
1. Data base management. I. Title
QA76.9.D3S534 1990
005.74—dc20

8927933
CIP

Printed in the United States of America
1 2 3 4 5 6 7 8 9 0 DO 7 6 5 4 3 2 1 0

PREFACE

INTENDED AUDIENCE FOR THIS BOOK

Database Management: *Theory and Application* is appropriate for an applied programmer/analyst–oriented course in database management. It is recommended that students have had at least one COBOL course, as this is the language used in the examples. It is also beneficial if students have already developed sequential and random access programs, which provide a basis for comparing the traditional file-processing methods with the database approach.

This book is best suited to those courses emphasizing database application programming from an applied but thorough approach. It covers each of the three data models (relational, network, and hierarchical) in great detail and describes the theoretical basis for each.

For those courses that take a more managerial, database administration approach, Parts I, II, and VI, which deal with database and design concepts, will be of most benefit. Parts III, IV, and V discuss the data models, beginning with the theoretical foundation of each model, and contain numerous programming examples.

PURPOSE OF THIS BOOK

Today's database environment is changing. For the first time, there is a preferred database methodology and language: the relational approach and Structured Query Language (SQL). However, as recently as early 1989, there were more than three times as many commercial sales of network and hierarchical database packages. Some current texts ignore these models or cover them in just a single chapter. Thus, students may not learn how to develop applications with the network and hierarchical database products used by more than 15,000 organizations.

Using many years of experience as a DBA and many more as a programmer using a database management system (DBMS), I provide intensive coverage of actual commercial packages so that students can use the text as a reference after graduation. For example, instead of merely saying that a program should check for error codes after invoking the DBMS, I provide actual values to be checked. When I worked in this field prior to entering the teaching profession, no books on the market offered appropriate coverage of commercial packages. When faced with the task of developing a program that used a DBMS, students either had to read vendors' reference manuals or attend training sessions in order to gain

even a rudimentary knowledge about how to write such programs. This book attempts to fill that vacuum.

FEATURES

Specific features of this text include:

- *A carefully constructed foundation for each of the three data models.* Part I, which discusses the theory behind database management, builds a foundation for the three parts devoted to the data models. Chapter 3 on data structures is particularly complete. All of the database management systems discussed in this book use a data structure for their implementation. By studying data structures, students are better prepared to understand how the packages work, how they are implemented, and why they must be used in the manner described.
- *At least two chapters devoted to each model.* The first chapter opening each model-related part introduces the concepts behind the model and how to define the database. The second chapter shows how to manipulate the resulting database. The relational model will eventually be so dominant that current texts might be able to successfully get away with eliminating a discussion of the network and hierarchical models. The early-1990 reality, however, is that thousands of organizations are still using applications written to take advantage of database systems built on these two models. Parts IV and V provide thorough coverage of the CODASYL network and hierarchical models.
- *Pedagogical symmetry for each data model.* In addition to the traditional sections on data storage and retrieval, the chapters devoted to each model contain sections on data integrity, data backup and recovery, and transaction processing. This enables students to compare and contrast all three models.
- *Integrated microcomputer coverage.* The hierarchical and network parts contain an integrated section, while micro-based versions of relational packages are discussed in Chapter 14.
- *Skills-oriented approach.* This book includes more programming examples for each model than any current text. Whenever possible, I have tried to illustrate more than just a few lines of code. It is vital that students be able to use a DBMS after they graduate. Only by seeing many complete examples can this be accomplished.
- *Fourth-generation language coverage.* One of the database trends I have observed from my consulting work is the tendency of organizations and software developers to use SQL in concert with a fourth-generation language (4GL). Chapter 8 on embedded SQL and Chapter 14 on data efficiencies discuss how SQL can be used in such an environment. While most database applications are written in COBOL, the relational approach requires new languages to fully exploit its power.

- *Managerial coverage.* The human element in database management cannot be ignored. The roles of the database administrator (DBA), programmer, analyst, and end user are clearly defined throughout the text. For example, Part II is oriented toward the database administration function, and Part VI discusses where the database market is headed. Finally, each of the three data model–related parts includes thorough coverage of the model's concepts in a program-free context.
- *Up-to-date information.* Writing a text about a topic as dynamic as database is quite a task. Over the course of a couple of years, a text goes through several drafts. In many subject areas, changes to texts from draft to draft are often aimed at correcting grammatical, organizational, and conceptual errors. This wasn't the case with this book. Each new draft had to be updated to include new topics. For example, the DB2 coverage was updated for the final time less than one month after the most recent IBM version was released, and the discussions about database machines, distributed databases, and 4GLs reflect research and developments that were less than a month old when I wrote the last draft.
- *Chapter-opening cases.* Many chapter-opening cases illustrate what *not* to do or how the concepts discussed in the chapter could have been used to avoid the problem presented in the case. In this text, each case includes a series of discussion questions to help prepare students for the chapter that follows. Two of the cases are integrated throughout the text so that students can compare problem solutions using a package from each data model.
- *Chapter objectives.* After each opening case, the chapter proceeds with behavioral objectives that students should be able to achieve after reading the chapter. Very few texts in this market provide chapter objectives for students to test themselves.
- *An integrated case.* Based on over seven years of consulting in the hospital industry, I developed an integrated case study, Community Hospital, which demonstrates how to implement concepts discussed in the particular chapter. Several of the case episodes demonstrate a complete COBOL program that illustrates how a DBMS is used. By placing this material in a case format, I could illustrate all the "messy" details without getting bogged down in the chapters. The case approach also offered a unique opportunity to show how an application might look if programmed using a package from each data model.
- *Solved problems.* Most chapters contain numerous problems and accompanying solutions. This technique has proven an effective learning device, as it lets students see how to apply a concept to solve an actual problem. In many cases, the same problems are used in different chapters so that students can see how the problem is solved for each model.
- *Annotated bibliography.* The bibliography at the end of the text contains many references, several of them annotated.

ORGANIZATION

Part I introduces students to the database approach, data, data relationships, and data structures—the mechanisms for implementing data relationships. Because the structures are stored on disk drives, instructors might find it useful to supplement Part I with Appendixes A and B.

Part II covers the design of a database. Chapter 4 presents an approach for developing the conceptual design, a model of the organizations' entities and the relationships among them. The major thrust of this chapter is a discussion of normalization, a technique originally designed for the relational approach that is applied to all of the data models. This approach has seldom been taken, but it makes the chapter applicable regardless of the data model being covered in class. Chapter 5 shows students how to map the conceptual model to one compatible with the organization's DBMS and how to physically implement the transformed design using a data structure.

Part III covers the relational model, probably the most important model in this book. Chapter 6 discusses relational algebra. This method for manipulating relational databases has proven an effective way for students to grasp the concept of set-at-a-time versus the more conventional record-at-a-time processing to which they have become accustomed. This chapter also includes a highly comprehensive discussion of Query-By-Example, an IBM-developed approach that lets end users easily manipulate a relational database. Chapters 7 and 8 discuss Structured Query Language, first in an interactive mode and then embedded in COBOL and 4GL programs. Learning SQL before studying other database management languages can be somewhat misleading. What SQL can do in one statement, the network and hierarchical models would require more than twenty. It may be beneficial to assign the parts out of sequence, saving the easiest for last, so that students can better appreciate the power and benefits of the relational approach.

Parts IV and V discuss the network and hierarchical models, respectively. Both models had their origins in the 1960s and are significantly more complex than the relational approach. Part IV uses the industry standard DBMS developed by CODASYL as the basis for the discussion. The Community Hospital continuing case shows how Cullinet's IDMS/R package, the market leader for the network model, differs from the standard.

Part V discusses IBM's hierarchical package, IMS. As of early 1989, there were still more than twice as many IMS installations as relational ones, so these chapters are particularly important.

Part VI is devoted to management of the database. Chapter 13 discusses the duties and responsibilities of the database administrator (DBA), the individual usually responsible for database management. This chapter also returns to some topics addressed in Chapter 1: concurrency control and backup and recovery. Chapter 14 illustrates some of the current trends in database management, including how organizations are organizing their data for more efficient access.

By distributing the data to the end users, data communications costs as well as response time can be reduced. Another way to gain efficiency is to use a dedicated machine that performs the data access operations, a device usually referred to as a *database machine.* The final technique for improving the efficiency of data retrieval is the use of a 4GL. Two such packages are discussed in the chapter.

ACKNOWLEDGMENTS

I wish to acknowledge the assistance given to me by Cullinet and IBM, both of which provided me with reference materials and answered my questions relating to their database packages. I also want to thank ALCOA for assisting me with the IMS portion of the text. Finally, I wish to express my gratitude to the dedicated reviewers of this text, without whom this book would not have been possible:

James Benjamin
Southern Illinois University
at Edwardsville

Ralph Bisland
University of Southern Mississippi

Norman Brammer
Colorado State University

Helen Casey
Sam Houston State University

William Cornette
Southwest Missouri State

Ralph Duffy
North Seattle Community College

Carl Friedman
University of the District of Columbia

Alden Lorents
Northern Arizona University

Laurie MacDonald
Bryant College

Don Musselman
James Madison University

Richard Ramirez
Arizona State University

Shannon Scanlon
Henry Ford Community College

Steven Schindler
Kent State University

Glenn Smith
James Madison University

Marguerite Summers
Murray State University

Elizabeth Unger
Kansas State University

Susan White
Catonsville Community College

Peggy Wingo
Richland College

John C. Shepherd

CONTENTS