C 473

BIBLIOTHEQUE DU CERIST

Data Base Management Systems

Dionysios C. Tsichritzis
Frederick H. Lochovsky

University of Toronto

ACADEMIC PRESS New York San Francisco London

A Subsidiary of Harcourt Brace Jovanovich, Publishers

BIBLIOTHEQUE DU CERIST

COPYRIGHT © 1977, BY ACADEMIC PRESS, INC.
ALL RIGHTS RESERVED.
NO PART OF THIS PUBLICATION MAY BE REPRODUCED OR TRANSMITTED IN ANY FORM OR BY ANY MEANS, ELECTRONIC OR MECHANICAL, INCLUDING PHOTOCOPY, RECORDING, OR ANY INFORMATION STORAGE AND RETRIEVAL SYSTEM, WITHOUT PERMISSION IN WRITING FROM THE PUBLISHER.

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1

Library of Congress Cataloging in Publication Data

Tsichritzis, Dionysios C.
Data base management systems.

(Computer science and applied mathematics series)
Bibliography: p.
1. Data base management. I. Lochovsky, Frederick H., joint author. II. Title.
QA76.9.D3T74 001.6'442 76-13952
ISBN 0-12-701740-2

PRINTED IN THE UNITED STATES OF AMERICA

BIBLIOTHEQUE DU CERIST

CONTENTS

Preface ix
Acknowledgments xiii
Notes to the Instructor xv

Part 1 DBMS CONCEPTS

Chapter 1 **Information Systems and Data Base Management Systems**

1.1 Information Systems 3
1.2 Data Processing Systems 9
1.3 Data Base Management Systems 20
1.4 Concluding Remarks 28
Exercises 29

Chapter 2 **Data Models**

2.1 Introduction 32
2.2 Relationships 34
2.3 The Network Data Model 38
2.4 The Relational Data Model 43
2.5 Normalization 45
2.6 Concluding Remarks 55
Exercises 55

Chapter 3 **Data Languages**

3.1	Introduction	58
3.2	Data Selection	60
3.3	Network Selection	64
3.4	Relational Algebra	69
3.5	Concluding Remarks	78
	Exercises	80

Chapter 4 **DBMS Facilities**

4.1	Schemas and Subschemas	82
4.2	Interfaces	83
4.3	Language Facilities	86
4.4	Data Independence	93
4.5	Concluding Remarks	95
	Exercises	97

Chapter 5 **Hierarchical Systems**

5.1	Hierarchical Data Model	99
5.2	Hierarchical Data Languages	105
5.3	Examples of Applications	119
5.4	Concluding Remarks	133
	Exercises	134

Chapter 6 **Network Systems**

6.1	Network Data Model	136
6.2	CODASYL DBTG Proposal	139
6.3	Network Data Languages	147
6.4	Examples of Applications	168
6.5	Concluding Remarks	180
	Exercises	183

Chapter 7 **Relational Systems**

7.1	Relational Data Model	185
7.2	Relational Data Languages	186
7.3	Examples of Applications	198
7.4	Concluding Remarks	212
	Exercises	213

Chapter 8 **Implementation Considerations**

8.1	Searching	215

BIBLIOTHEQUE DU CERIST

8.2 Inverted Files 218
8.3 Network Implementation 221
8.4 Hierarchical Implementation 232
8.5 Relational Implementation 241
8.6 Data Base Machines 245
Exercises 247

Chapter 9 **Operational Requirements**

9.1 Performance 251
9.2 Integrity 253
9.3 Security 255
9.4 Concurrency 257
9.5 Utilities 261
9.6 Concluding Remarks 263
Exercises 263

Part 2 DBMS EXAMPLES

Chapter 10 **Introduction** 267

Chapter 11 **Information Management System (IMS)**

11.1 Data Model 270
11.2 Data Definition Facility 272
11.3 Data Manipulation Facility 273
11.4 Storage Organizations 277
11.5 Example Application 281
11.6 Concluding Remarks 289

Chapter 12 **MRI System 2000**

12.1 Data Model 291
12.2 Data Definition Facility 291
12.3 Data Manipulation Facility 294
12.4 Example Application 297
12.5 Concluding Remarks 298

Chapter 13 **Integrated Database Management System (IDMS)**

13.1 Data Model 300
13.2 Data Definition Facility 301
13.3 Data Manipulation Facility 303
13.4 Example Application 307
13.5 Concluding Remarks 314

BIBLIOTHEQUE DU CERIST

Chapter 14 **Total**

14.1 Data Model 316
14.2 Data Definition Facility 316
14.3 Data Manipulation Facility 318
14.4 Example Application 321
14.5 Concluding Remarks 328

Chapter 15 **Adaptable Data Base System (ADABAS)**

15.1 Data Model 329
15.2 Data Definition Facility 329
15.3 Data Manipulation Facility 331
15.4 Example Application 335
15.5 Concluding Remarks 337

Appendix **Data Base Exercises and Games**

A.1 Introduction 339
A.2 Exercises 340
A.3 Games 346

References 364
Index 375

BIBLIOTHEQUE DU CERIST

BIBLIOTHEQUE DU CERIST

PREFACE

Data base management systems or DBMS's have come into vogue in recent years. Their number and their usage have increased tremendously in a very short time. Many organizations have looked to a DBMS as a panacea for their data processing problems. Unfortunately, the claims made for and about DBMS's have sometimes exceeded their capabilities. In addition, it is very difficult for some organizations to take advantage of this new technology immediately. The rapid growth in the usage of DBMS's has created a shortage of personnel trained in the effective use of such systems. As a result, many organizations have found that instead of solving their data processing problems, DBMS's have sometimes created new ones [Cox, 1974].

As a first step toward resolving this situation, the user, or potential user, of a DBMS should have a firm grasp of DBMS concepts. These concepts include an understanding of the nature of data and data processing. They also deal with basic DBMS organizations and their capabilities and limitations. The first part of this book covers DBMS concepts. In Chapter 1, we examine briefly the history of data processing. We show how the need evolved for a unified approach to data management and how DBMS's help achieve this goal. The organization of data and the meaningful representation of data relationships via data models are presented in Chapter 2. Chapter 3 discusses data access via data languages in the context of the data models presented in Chapter 2. Chapter 4 discusses, in general, the facilities provided by DBMS's. Chapters 5, 6, and 7 describe respectively hierarchical, network, and relational systems. The description includes the respective data model, some data languages, and some example applications that demonstrate their use. Chapter 8 surveys implementation techniques. Finally, Chapter 9 discusses some operational requirements of DBMS's. With this background a user should be better able to evaluate various DBMS's and the claims made about them.

As a second step, the prospective user should acquire some knowledge of the current state of the art in DBMS's. Commercial DBMS's should be investigated and their different features evaluated. The second part of this book serves this purpose by presenting a survey of various DBMS's. In this manner the user is not required to struggle with many sets of manuals containing completely different terminologies, attitudes, and approaches. In addition, the prospective user can appreciate how the abstract concepts discussed in the first part of the book are mapped into concrete system features.

As a third step, it is beneficial for the user to get some real "hands on" experience with a DBMS. Such experience is very valuable for alleviating a prospective user's "mental block" with respect to information systems in general and DBMS's in particular. In the Appendix we discuss a teaching environment that provides this type of experience.

Finally, the knowledgeable user may participate in the design, implementation, or maintenance of a real DBMS. However, we do not expect all our readers to go through this final step.

This book is written mainly for the DBMS user or prospective user. It is not a design manual for DBMS's nor a feature analysis of existing DBMS's. We hope to present the reader with an overall view of DBMS requirements, capabilities, and existing features. We have tried to present data base management as a unified subject and not as a collection of unrelated approaches and techniques. For this reason, we present some general, unifying concepts in Chapters 2–4, before we discuss the different approaches in Chapters 5–7. In addition, we present some implementation techniques in a unified manner in Chapter 8. We hope that the reader will find this organization helpful. It should enable him to grasp DBMS concepts as a unified approach and not as a fragmented set of ideas.

Implementation considerations are incorporated to provide some insight into the problems and resulting limitations of DBMS's. However, no detailed solutions to implementation problems are given. We concentrate mainly on the facilities provided by DBMS's and not on the details of their implementation. Such mechanisms are continually evolving, especially with hardware changes. New insights and techniques are emerging frequently. Any in-depth discussion of specific mechanisms would, of necessity, be incomplete and soon out of date.

We emphasize mainly the data-structuring aspect of DBMS's. Operational requirements such as security, integrity, and concurrency, which are an integral part of a successful system, are discussed briefly in Chapter 9. Although they deserve special attention, it is difficult to present general solutions concisely. Much research is still needed and is currently proceeding in these important areas.

In a book of this size it is not possible to present complete applications. However, since we feel that examples help illustrate the concepts discussed, we have included numerous examples of realistic applications. This book is not a programming manual. Some of our examples are neither complete nor particularly attractive programs. We try to deal with the applications and not with their particular representation in a programming language. Different languages are presented to

BIBLIOTHEQUE DU CERIST

illustrate features and concepts and not to promote any specific syntax for commands. The programs presented have been checked as thoroughly as possible for correctness. However, we did not ever intend to run them at any particular installation.

We strongly advise the reader to use a DBMS. It is a sobering experience to be confronted with the idiosyncrasies of a particular DBMS and to overcome them for a particular problem. Our DBMS teaching environment, described in the Appendix, is specifically designed for the purpose of exposing potential users to different DBMS approaches. However, for pedagogical use, any DBMS is adequate, provided it includes a wide range of facilities and operates within the budget of our readers.

We hope this book presents a concise overview of the DBMS area. We also hope that the material presented will not quickly become obsolete in the rapidly changing world of DBMS's. The first part of this book, which deals with basic concepts and organizations, will probably remain relevant for some time. The second part deals with particular systems which evolve continually. However, the investment of both manufacturers and users in these systems is so great that they will probably be with us for quite some time.

There is currently a debate among DBMS professionals concerning the advantages of each DBMS approach [Canning, 1972a; Ashenhurst, 1974; Bachman, 1974a; Codd and Date, 1974; Date and Codd, 1974; Sibley, 1974; Steel, 1975a]. The debate concerns not only specific, desirable features for a DBMS but also the possibility and desirability of a standardization [SPARC, 1974]. The different positions in this debate can be summarized as follows:

1. A group of proponents of network systems want the CODASYL DBTG proposal [CODASYL DBTG, 1971], with some possible modifications, to become a standard. Such standardization, they argue, will have the same beneficial effect on DBMS's as the standardization of COBOL had on EDP.

2. A group of independents strongly oppose any effort at standardization. They claim that it is too early to impose a standard in an area that is developing as rapidly as DBMS's. They would like to postpone any standardization effort. In this way, the successful commercial DBMS may become a *de facto* standard.

3. Some people prefer the relational model as a basic data model for DBMS's. They cite as advantages simplicity, data independence, etc. These people do not disregard hierarchies or networks for implementation, but simply prefer the user interface to be relational. Although they do not push for standardization, they would like DBMS development to be heavily influenced by the relational model.

4. Some people propose that a different, more flexible model should be the basis for DBMS's. Such a model is proposed as a *coexistence model.* It is a compromise between the hierarchical, network, and relational approaches, being identical to none, but able to accommodate all.

We shall try to avoid taking a particular position. We have our own prejudices which may show in this book; however, we think it would be unfortunate to base the book on a particular approach. We hope that time and maturity will eventually

BIBLIOTHEQUE DU CERIST

minimize the differences between the proponents of different approaches. At least the air will be cleared of misunderstandings, misconceptions, and differences of terminology which generate much of the controversy. The approaches are not really all that different [Stonebraker and Held, 1975]. In addition, none of them is a panacea. No approach by itself will solve all the problems associated with the design, implementation, and operation of DBMS's. The debate will probably evolve as an argument between efficiency and ease of programming. As such, it reminds one of the old controversy on the merits of different programming languages.

BIBLIOTHEQUE DU CERIST

BIBLIOTHEQUE DU CERIST

ACKNOWLEDGMENTS

We would like to thank Professor S. A. Schuster, Dr. E. A. Ozkarahan, J. Klebanoff, and all the people in the Data Base Group at the University of Toronto for many stimulating ideas and discussions.

Dr. P. A. Bernstein, Dr. H. A. Schmid, Dr. G. M. Nijssen, Professor W. R. Bezanson, B. Czarnik, M. L. Brodie, B. W. Wade, and J. R. Swenson provided us with many useful comments. Their criticism was helpful in turning a set of lecture notes into a book. We would also like to thank an Academic Press reviewer for his very helpful suggestions and comments.

The DBMS examples in Part II of this book were originally produced by R. Stark, S. Chan, D. Schafheitlin and B. Czarnik. The initial editing was done by J. Klebanoff. We would like to thank T. Cassidy and M. Kallaur for commenting on the original drafts of some of the descriptions. Their suggestions were very helpful in improving the content of the descriptions. We also acknowledge, with thanks, the help of the following organizations: IBM Canada Ltd., MRI Systems Corp., Cullinane Corp., Cincom Systems of Canada Ltd., and R. D. Nickel & Associates.

We would like to acknowledge, with thanks, many interesting discussions with persons at the IBM San Jose Data Base Group and the University of California at Berkeley Data Base Group.

Finally, we would like to thank I. Weber for typing the original draft of this book.