DATA BASE

Structured Techniques for Design, Performance, and Management 2nd Edition

Shaku Atre

Data Base: Structured Techniques for Design, Performance, and Management

Second Edition

Shaku Atre
Atre International Consultants, Inc.
Rye, New York

JOHN WILEY & SONS
New York • Chichester • Brisbane • Toronto • Singapore

Copyright © 1988 by Shaku Atre
Published by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by Section 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional service. If legal advice or other expert assistance is required, the services of a competent professional person should be sought. *From a Declaration of Principles jointly adopted by a Committee of the American Bar Association and a Committee of Publishers.*

ISBN 0-471-85251-1

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Preface

We are reaching the stage where the problems that we must solve are going to become insoluble without computers. I do not fear computers. I fear the lack of them.

Isaac Asimov

The computer revolution promises to simplify, ease, and otherwise enhance life in ways undreamed of even by the Utopians. In organizations of all kinds, the computer stores and retrieves vast amounts of information. The volume of data processed in all types of industries, medical services, financial institutions, and governmental agencies is expanding geometrically, and it is very difficult to imagine how these organizations would function without computers. Organizations of various natures are called "enterprises."

The repository of information, called a "data base," is the foundation of the entire computer system for an enterprise. Thus the design and the performance of the data base are perhaps two of the most important areas in information processing.

Although several articles and books have been published that give scholarly treatment to this area, or give survey information regarding existing data base management systems and data models, few offer specific answers to these questions:

1. What are the responsibilities of the people involved?
2. How can a data base that will satisfy performance criteria be designed?
3. What problems may arise in different phases?

This book is intended to fill the need for this and related information. It demonstrates the principles for designing a data base that will achieve the functions and performance needed to satisfy the requirements of the information processing environment. It addresses the roles of the data

base administrators, systems programmers, systems analysts, application programmers, information processing managers, data processing managers, and data base users. The material is simple without becoming simplistic and shows a step-by-step approach to data base design. The chapters are organized so that data base designers can follow their sequence during the process of designing a data base. The methodology is independent of any software package; it does not concentrate on any one manufacturer's data base management system.

This book can be used effectively by practitioners in business as well as in government, and also as a single-semester course in a business or technical school where students are interested in the "how to" as well as the "why" of information processing approaches.

A case study approach has been adopted. A case study for designing a data base for a banking environment is carried through from the very beginning to the end in Appendix A. In Appendix B a case study of a university environment is presented in its entirety. The step-by-step approach to designing a data base and the application of this approach to the two case studies are major features of the book.

The material presented was developed in the courses that I have taught for a number of years at IBM's Systems Research Institute. A version of the manuscript was used as a text in a course with the same title, "Data Base: Structured Techniques for Design, Performance, and Management." Portions were presented at various professional seminars, including local chapter meetings of the Association for Computing Machinery (ACM) and the Data Processing Management Association (DPMA). Questions and suggestions from students and colleagues helped to make the material more timely and digestible, while the overall enthusiastic response at these lectures led to the decision to present the material as a book.

The book is divided into three parts, each consisting of a number of chapters. Part 1, "Data Base Administration," deals with installing and maintaining a data base. The first chapter sets the stage by discussing the issues that will be treated later. Chapter 2 describes the function of the data base administrator (DBA) and the roles that the DBA and his or her staff play in the information processing system. The third chapter, titled "Data Dictionary," discusses a valuable tool for collecting information about the data base environment.

Part 2, "Data Base Design (Conceptual Model and Logical Model)," demonstrates the conceptual and logical data base design process. The fourth chapter creates the basis of any data base management system—the data models. A brief treatment is given of three popular approaches—the relational, hierarchical, and network data models. Chapter 5 describes the design process for developing a conceptual model. Chapter 6 gives the design process for a logical model. Both chapters demonstrate the techniques in detail, with specific examples.

此为试读,需要完整PDF请访问: www.ertongbook.com

Part 3, "Data Base Performance," describes the equally important aspects of the performance of information processing with the data base. The main emphasis is on how to design for good performance, that is, how to optimize the design. Chapter 7 reviews data access techniques, which are key issues of performance. The eighth chapter presents some implementations of these access methods. Chapter 9 discusses the data base design of a physical model with the estimation of space and time. Chapter 10 treats specific performance problems in the design phase, implementation phase, and operations phase.

Appendix A demonstrates a case study of a banking environment. Appendix B contains a case study of a university environment. Finally, Appendix C gives the basic concepts of probability theory and mathematical statistics used in Chapter 9.

And now only the pleasant task of acknowledging my students, friends, and colleagues remains. I am grateful to the numerous students who patiently read the manuscript and made many constructive comments. I also want to thank Charles Bontempo, Robert Bower, Gary D. Bowers, Robert G. Brown, George T. Fadok, Joan Foster, Ron Gale, George Hubbard, Judy King, Barry Kingsbury, Frank Post, Jay Schmoll, Cora Tangney, and everyone else who helped along the way.

S. ATRE

New York, New York
July 1980

Preface to the Second Edition

The urgent need for a coordinated and complete data base design methodology, surfacing in all sectors of the data processing community, has resulted in the continuing success of this book. Moreover, seminars emphasizing the practical step-by-step nature of the methodology have been enthusiastically received and applied in commercial and academic as well as many government installations throughout the world. The book, in fact, has been translated into several foreign languages. These events have prompted revision to include new developments and insights gained, while preserving all the strengths of the first edition.

Part 1, "Data Administration," has undergone minor alterations. The introductory Chapter 1 remains unchanged. Chapter 2 has been revised to reflect the evolution of the previous DBA function into two functions: Data Administration (DA), responsible for planning and setting policies; and Data Base Administration (DBA), responsible for day-to-day technical implementation of policies. Chapter 3, "Data Dictionary," has been slightly revised to sharpen distinctions between active and passive and between integrated and freestanding data dictionaries.

Part 2, "Data Base Design (Conceptual Model and Logical Model)," has been expanded to include the inverted file data model. Several prominent and widely used data base management systems are based predominantly upon file inversion. The needs of data base designers for these systems have to be addressed in a methodology that claims to be comprehensive. Accordingly, Chapter 4 now includes the inverted file data model as a separate and distinct data model. Chapter 5, "Data Base Design (Conceptual Model)," which is independent of the DBMS used, remains unchanged. Chapter 6 treats mapping the conceptual model to a logical model based on file inversion, in addition to the other popular logical models.

The biggest changes in the new edition are in Part 3, "Data Base Performance," although Chapter 7, "Data Storage and Access Methods," remains the same. Chapter 8 previously discussed Implementation of the Access Methods for several data base management systems. Chapter 9 did the same for Designing and Evaluating the Physical Model. The new Chapters 8, 9, and 10 treat IMS/VS, IDMS/R, and ADABAS, respectively.

Each has three sections: (1) Implementations of the Access Methods; (2) Design a Physical Model of a Data Base; (3) Evaluate the Physical Model of a Data Base. Chapter 8 unifies material previously found in the old Chapters 8 and 9. The new Chapters 9 and 10 contain completely new material and are much more detailed and specific than anything in the first edition. Recasting the treatments of the three DBMSs—IMS, IDMS/R, and ADABAS—in a unified and balanced form will make them more readily accessible to a reader interested in just one of them. Contrasting and comparing the three systems is also facilitated. Finally, the new Chapter 11, "Performance Issues," is the old Chapter 10 unchanged.

The even and comprehensive handling of three different data base management systems has also been carried over into Appendices A and B, the Banking and University Case Studies. These have been improved for the conceptual and IMS models. In addition, logical and physical models have been created and evaluated for IDMS/R and for ADABAS.

Finally, to enhance the instructional value of the book and to respond to requests from teachers in numerous colleges and universities, sets of questions have been included for all chapters.

And now I would like to thank two individuals, in particular, who made this second edition possible: Franz Schneider, Principal Consultant, and Nancy Carafa, Office Manager, who have worked with me at Atre International Consultants, Inc. for the past several years.

SHAKU ATRE

Rye, New York
November 1987

Works by the Author

PAPERS & ARTICLES BY SHAKU ATRE

"End-User Designed/Driven Data Bases," S. Atre, *Computerworld/Extra*, September 1, 1982.

"Spreading the Wealth of DBMS," S. Atre, *Computerworld Buyer's Guide: Software*, December 14, 1983.

"Information Center & Data Base: Synergy for Success!" S. Atre, *Technology Transfer Institute Newsletter*, Vol. 2, No. 1, 1983.

"Nothing Succeeds Like Access," S. Atre, *Data Training*, January 1984.

"Information Center Implementation," interview with S. Atre, *Data Base Newsletter*, January/February 1984.

"Info Centers and P. C.s: Should They Be Synonymous?" S. Atre, *P. C. Week*, July 3, 1984.

"Information Center Software," S. Atre, *Computerworld: Office Automation*, August 15, 1984.

"Create That Opportunity," interview with S. Atre, *InfoSystems*, December 1984.

"Will The Real Fourth-Generation Languages Please Stand Up?" S. Atre, *Softline*, *Computerworld*, June 17, 1985.

"Technology, Intelligent Software Guide Market Growth," S. Atre, *Computerworld Special Report*, June 24, 1985.

"The Information Center and Productivity Tools: Working in Harmony," S. Atre, *Computerworld Focus*, September 18, 1985.

"Information Center Strategy," interview with S. Atre, by Elisabeth Horwitt, *Business Computer Systems*, September 1985.

"Mini and Mainframe DBMS, The Seven-Step Solution," S. Atre, Product Spotlight, *Computerworld*, March 31, 1986.

"IC Directors' Problems: Management, Understaffing," interview with Shaku Atre, by Bob Rhein, *MIS Week*, August 4, 1986.

"Integrating tools speeds and simplifies application building process," Shaku Atre, Product Spotlight/Application Generators, *Computerworld*, Vol. XX, No. 31, August 4, 1986.

"The Information Center Holds," S. Atre, *PC World*, August 1986.

"Storing Data About Data," S. Atre, *Information Center*, Weingarten Publications, Inc., July 1986.

BOOKS BY SHAKU ATRE

Data Base: Structured Techniques for Design, Performance, and Management, S. Atre, New York: John Wiley, 1980.

Data Base Management Systems for the Eighties, S. Atre, Wellesley, Massachusetts: Q.E.D. Information Sciences, 1984.

Information Center: Strategies and Case Studies, S. Atre, Rye, New York: Atre International Consultants, Inc., 1986.

CONSULTING EDITOR, SHAKU ATRE

The Macmillan Database/Data Communications Series:

AVAILABLE

Cave-Maymon *Software Lifecycle Management*
Fadok *Effective Design of CODASYL Data Base*
Potter *Local Area Networks: Applications and Design*
Ranade/Ranade *VSAM: Multiprogramming with Virtual Storage*
St. Amand *A Guide to Packet-Switched Value-Added Networks*
Towner *The ADS/On-Line Cookbook*

Contents

1. DATA 1

1.1. Sources of Data, 2
1.1.1. *Enterprise*, 2
1.1.2. *Entity*, 3
1.1.3. *Attributes (or Data Elements)*, 3
1.1.4. *Data Value*, 3
1.1.5. *Key Data Element*, 4
1.1.6. *Data Record*, 5
1.1.7. *Data File*, 5
1.2. Data File Environment, 5
1.2.1. *Advantages of Generalized Routines*, 7
1.2.2. *Drawbacks of the Conventional Data File Environment*, 9
1.3. Data Base Environment, 11
1.3.1. *What Is a Data Base?* 11
1.3.2. *Data Base Management System (DBMS)*, 11
1.3.3. *Drawbacks of Data Integration*, 12
1.4. Data Administrator, 13
1.5. Data Independence, 15
1.5.1. *What Is Data Independence?* 15
1.5.2. *Two Stages of Data Independence*, 16
1.5.3. *How to Achieve Data Independence*, 18
1.6. Data Dictionary, 19
1.7. Data Base Design and Performance—Preview, 21

PART 1 DATA ADMINISTRATION

2. DATA ADMINISTRATION 25

2.1. Data Administration and Data Base Administration Functions, 25
2.1.1. *DA's Responsibilities*, 27

此为试读,需要完整PDF请访问: www.ertongbook.com

2.1.2. *DA and Management*, 28
2.1.3. *DA and User Community*, 30
2.1.4. *DA and Application Development Staff*, 32
2.1.5. *DA and Systems Staff*, 33
2.1.6. *DA and Operations Staff*, 33
2.1.7. *DA and Software Vendors*, 34
2.1.8. *DA and Hardware Vendors*, 35
2.2. Data Base System's Life Cycle, 36
2.2.1. *Design of the Data Base*, 39
2.2.2. *Physical Creation of the Data Base*, 40
2.2.3. *Conversion of Existing Data Sets and Applications to Match the Newly Created Data Base*, 43
2.2.4. *Integration of Converted Applications and New Applications into the New Data Base*, 43
2.2.5. *Operations Phase*, 44
2.2.6. *Growth, Change, and Maintenance Phase*, 44
2.3. DA, DA's Staff, and Their Responsibilities, 44
2.3.1. *DA's Placement in Organization*, 45
2.3.2. *DBA's Staff*, 47
2.3.3. *Responsibilities of the Individual Staff Members*, 49

3. DATA DICTIONARY **58**

3.1. What Is a Data Dictionary? 58
3.1.1. *Basic Objectives of a Data Dictionary*, 59
3.1.2. *Data Base Management System and Data Dictionary*, 61
3.1.3. *Data Dictionary and Its Interfaces*, 63
3.1.4. *An Ideal Data Dictionary: Its Requirements and Organization*, 65
3.2. Implementation Strategies for a Data Dictionary, 68
3.2.1. *Cost/Benefit Considerations*, 68
3.2.2. *When to Implement a Data Dictionary*, 71
3.2.3. *Data Definition Guidelines*, 73

PART 2 DATA BASE DESIGN (CONCEPTUAL MODEL AND LOGICAL MODEL)

4. DATA MODELS **79**

4.1. What Is a Data Model? 80
4.2. Relationships within a Data Model, 81
4.2.1. *One-to-One Relationship (between Two Entities)*, 81

4.2.2. *One-to-Many Relationship (between Two Entities)*, 81
4.2.3. *Many-to-Many Relationship (between Two Entities)*, 81
4.2.4. *One-to-One Relationship (between Two Attributes)*, 82
4.2.5. *One-to-Many Relationship (between Two Attributes)*, 83
4.2.6. *Many-to-Many Relationship (between Two Attributes)*, 83
4.2.7. *Overview of the Data Models*, 84
4.3. The Relational Data Model, 85
4.3.1. *Advantages of a Relational Data Model*, 90
4.3.2. *Disadvantages of a Relational Data Model*, 90
4.4. The Hierarchical Data Model, 91
4.4.1. *A Hierarchical Tree Structure*, 91
4.4.2. *Storage Operations with a Hierarchical Data Model*, 98
4.4.3. *Advantages of a Hierarchical Data Model*, 101
4.4.4. *Disadvantages of a Hierarchical Data Model*, 101
4.5. The Network Data Model, 101
4.5.1. *Representation of One-to-Many Relationship*, 106
4.5.2. *Three More Classes of Set Types*, 108
4.5.3. *Storage Operations with a Network Data Model*, 113
4.5.4. *Advantages of a Network Data Model*, 113
4.5.5. *Disadvantages of a Network Data Model*, 114
4.6. The Inverted File Data Model, 114
4.6.1. *Advantages of an Inverted File Data Model*, 116
4.6.2. *Disadvantages of an Inverted File Data Model*, 116

5. DATA BASE DESIGN (CONCEPTUAL MODEL) 120

Design a Conceptual Model of a Data Base, 120
5.1. Data Analysis, 121
5.1.1. *Collecting Information about Data for Existing Applications*, 121
5.1.2. *Collecting Information about Data for Future Applications*, 124
5.2. Relational Concepts for the Design Process, 125
5.3. Graphical Representation, 142
5.4. Design Process, 144

6. DATA BASE DESIGN (LOGICAL MODEL) 152

Design a Logical Model of a Data Base, 152

6.1. Mapping to a Relational Data Model, 153
6.2. Mapping to a Hierarchical Data Model, 157
6.3. Mapping to a Network Data Model, 172
6.4. Mapping to an Inverted File Data Model, 178

PART 3 DATA BASE PERFORMANCE 189

7. DATA STORAGE AND ACCESS METHODS 191

7.1. Interfaces between User and Data Base(s), 191
7.2. Internal Model (Physical Model) Access Methods, 193
7.2.1. *Physical Sequential*, 194
7.2.2. *Indexed Sequential*, 196
7.2.3. *Indexed Random*, 199
7.2.4. *Inverted*, 202
7.2.5. *Direct*, 203
7.2.6. *Hashing*, 204
7.3. External Model (User View) Access Methods, 206
7.3.1. *Relationships between Two Records*, 207
7.3.2. *External Model Access Methods*, 207

8. HIERARCHICAL DATA BASE MANAGEMENT SYSTEMS 217

8.1. Implementations of the Access Methods (IMS: Information Management System), 217
8.2. Design a Physical Model of a Data Base, 235
8.3. Evaluate the Physical Model of a Data Base, 240
8.3.1. *Space Estimates*, 241
8.3.2. *Time Estimates*, 247
8.3.3. *Data Base Applications*, 252

9. NETWORK DATA BASE MANAGEMENT SYSTEMS 256

9.1. Implementations of the Access Methods (IDMS/R: Integrated Database Management System/Relational), 257
9.1.1. *Areas and Pages*, 257
9.1.2. *Database Keys*, 257
9.1.3. *Internal Model Access Methods*, 258
9.1.4. *Location Mode CALC*, 258
9.1.5. *Location Mode VIA SET*, 259
9.1.6. *Location Mode DIRECT*, 259
9.1.7. *Set Order*, 260
9.1.8. *Set Mode*, 261
9.1.9. *File Inversion*, 263

9.1.10. *External Model Access Methods*, 264
9.1.11. *Currency*, 264
9.1.12. *DML Commands*, 265
9.1.13. *Relational External Models*, 266
9.2. Design a Physical Model of a Data Base, 267
9.2.1. *Set Options*, 267
9.2.2. *Record Options*, 270
9.2.3. *Areas and Files*, 271
9.2.4. *Dos and Don'ts of Data Structuring*, 271
9.2.5. *Bachman Diagram*, 275
9.3. Evaluate the Physical Model of a Data Base, 277
9.3.1. *Space Estimates*, 277
9.3.2. *Time Estimates*, 279

10. INVERTED FILE DATA BASE MANAGEMENT SYSTEMS **280**

10.1. Implementations of the Access Method (ADABAS: Adaptable Data Base System), 281
10.1.1. *Descriptors*, 281
10.1.2. *Internal Sequence Numbers*, 281
10.1.3. *Examples*, 281
10.1.4. *Implementation of File Inversion in ADABAS*, 284
10.1.5. *ADABAS Components*, 286
10.1.6. *Storage Methods (Internal Model)*, 288
10.1.7. *Retrieval Methods (External Model)*, 289
10.1.8. *Accessing Related Records*, 290
10.2. Design a Physical Model of a Data Base, 292
10.2.1. *Field Definition Options*, 292
10.2.2. *Compression*, 293
10.2.3. *Multiple Value Fields and Periodic Groups*, 293
10.2.4. *Descriptors*, 294
10.2.5. *Field and Component Placement*, 295
10.2.6. *Free Space*, 297
10.2.7. *Buffer Pool*, 297
10.2.8. *Number of Files*, 298
10.3. Evaluate the Physical Model of a Data Base, 299
10.3.1. *Space Estimates*, 299
10.3.2. *Time Estimates*, 303

11. PERFORMANCE ISSUES **305**

11.1. Implementation, 306
11.1.1. *Physical Creation*, 306

11.1.2. *Conversion and Integration*, 308
11.2. Operations, 309
11.2.1. *Backup and Recovery*, 309
11.2.2. *Reorganization*, 312
11.2.3. *Restructuring*, 313
11.2.4. *Performance Monitoring and Tuning*, 315
11.2.5. *Security in the Data Base Environment*, 321
11.2.6. *Privacy in the Data Base Environment*, 325

APPENDIX A: DESIGNING A DATA BASE FOR A BANKING ENVIRONMENT "POPULAR BANK"—CASE STUDY 1 **333**

APPENDIX B: DESIGNING A DATA BASE FOR A UNIVERSITY ENVIRONMENT "THE UNIVERSITY"—CASE STUDY 2 **376**

APPENDIX C: BASIC CONCEPTS: PROBABILITY THEORY AND MATHEMATICAL STATISTICS **419**

Index **423**

CHAPTER 1

Data

1.1 Sources of Data 2
- 1.1.1 Enterprise, 2
- 1.1.2 Entity, 3
- 1.1.3 Attributes (or Data Elements), 3
- 1.1.4 Data Value, 3
- 1.1.5 Key Data Element, 4
- 1.1.6 Data Record, 5
- 1.1.7 Data File, 5

1.2 Data File Environment 5
- 1.2.1 Advantages of Generalized Routines, 7
- 1.2.2 Drawbacks of the Conventional Data File Environment, 9

1.3 Data Base Environment 11
- 1.3.1 What Is a Data Base? 11
- 1.3.2 Data Base Management System (DBMS), 11
- 1.3.3 Drawbacks of Data Integration, 12

1.4 Data Administrator 13

1.5 Data Independence 15
- 1.5.1 What Is Data Independence? 15
- 1.5.2 Two Stages of Data Independence, 16
- 1.5.3 How to Achieve Data Independence, 18

1.6 Data Dictionary 19

1.7 Data Base Design and Performance—Preview 21

One of the foundations of an enterprise is the information needed for its survival and prosperity. This chapter indicates where this information comes from and how data processing requirements for using the information caused data base systems to evolve. One of the desirable characteristics of a successful data base is flexibility. Flexibility makes a data base less sensitive to changes in information processing requirements. This chapter discusses the issues regarding flexibility and gives a preview of the steps involved in designing a data base to provide adequate performance.

此为试读,需要完整PDF请访问: www.ertongbook.com