

DATABASE

Step-by-Step

Second Edition

DATABASE

Step-by-Step

Second Edition

Mark L. Gillenson

University of Miami



WILEY

John Wiley & Sons

New York / Chichester / Brisbane / Toronto / Singapore

Copyright ©1990 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.

Library of Congress Cataloging in Publication Data:

Gillenson, Mark L.

Database: step-by-step / Mark L. Gillenson. – 2nd ed.

p. cm.

Includes bibliographical references.

ISBN 0-471-61759-8

1. Database management.

QA76.9.D3G523 1990

005.74-dc20

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Preface

In the four short years since the publication of the first edition, the field of database has continued its rapid expansion. The major changes to the field have been exciting and numerous. Relational database management systems (DBMSs) are constantly improving their performance characteristics and increasingly have been used for serious commercial applications. The SQL language has been adopted as both a U.S. and an international standard. DBMSs are now considered "standard equipment" on personal computers and several mainframe DBMSs have been adapted for the PC. Interest is beginning to be shown in converting some of the huge set of industrial hierarchical and network database applications to relational database. Increasing interest is being shown in distributed databases and data dictionaries. The new concept of object-oriented databases is being developed.

In order to keep up with this rapidly changing field, this edition includes extensive updating of many first-edition chapters. In addition, minor improvements have been made throughout; new topics have been added; the number of chapters has been increased from 15 to 17. The major changes are as follows.

The chapters on relational and pseudo-relational database have been moved before the chapters on hierarchical and network database, to reflect the present intense interest in relational database and the projected long-term importance of the relational concept. But, the hierarchical and network chapters have not been reduced in content or relegated to appendices because these topics continue to be of major importance in the commercial world and will be for some time to come.

The most major change to the book regards the topic of relational database, which was covered in one chapter in the first edition, but now has three chapters devoted to it (plus a separate chapter on pseudo-relational database). The introductory material on relational database has been rewritten and expanded. The material on the now standard SQL language has been expanded and given its own chapter. Joining Query-by-Example (QBE), in another new chapter devoted to other relational tools, are Query Management Facility (QMF) and the popular personal computer DBMS: dBASE III + and dBASE IV. The material on the IBM System/38 in the chapter on pseudo-relational database has been updated to include the follow-on IBM AS/400 system.

Other changes include the inclusion of the Date-Wilson method of database design in the database design chapter, additional material on relational-style catalogs in the chapter on data dictionaries, and the inclusion of material on object-oriented databases and expert databases in the chapter on evolving topics.

Always remember that no matter how fast or exotic computing hardware becomes, the fundamental issues in how to efficiently and effectively handle a company's data will remain.

I would like to thank Raymond Frost and Vicar Hernandez for their help in the preparation of the second edition. I would also like to thank my Wiley editor, Diane Cerra, for her encouragement and diligent support of this project.

Mark L. Gillenson

*Coral Gables, Florida
October 1989*

Preface To The First Edition

Database, the study of data storage and management, is a burgeoning discipline in the rapidly expanding fields of computer science and data processing. As in the evolution of any discipline, it began as the study of a relatively small body of knowledge, by a limited group of practitioners in industry, and, by faculty and graduate students in universities. Today, we find that both the number and depth of database subtopics have grown dramatically. A wide variety of data processing and non-data processing personnel, across the gamut of industrial and nonindustrial organizations, are exposed to the topic. As a university subject, it is increasingly migrating to the undergraduate computer science and business administration curriculums.

An unfortunate, if not surprising, characteristic of the database field, is that to have a thorough understanding of it, to eliminate the mysteries of how so significant an accomplishment can actually be made to work in the cold silicon and metals of the machines, a large amount of interconnected information must be assimilated. It is difficult to understand why a hierarchical data structure might be used without understanding the problems of data redundancy and multiple relationships. It is difficult to understand how to gain access to the records in a network structure without understanding the fundamentals of access methods. It is difficult to understand the need for a data administration function without understanding the principles behind the database approach. Such a list of interconnected concepts could go on indefinitely in this field.

The purpose of this book is to explain the PRINCIPLES behind the database and the broad range of database subtopics IN A MANNER WHICH LEAVES

NOTHING TO THE IMAGINATION. On the other hand, it does not attempt to exhaustively cover every detail of the field. It does not require a substantial data processing background as a prerequisite, but assumes only that the reader understands elementary computer principles and elementary concepts of computer programming. It approaches the subject in a methodical progressive manner, which will insure a *confident understanding* of the subject of database.

This book is intended to be used as the text for a first course in database, or as an introduction to the field for data processing professionals or others whose work exposes them to the database environment. Included is material on basic data definitions and structures, access methods, database management systems characteristics and approaches, database design, management aspects of the environment, and other assorted topics in the field.

I would like to leave the following thought with my fellow computer scientists, lest I offend their collective sense of precision. In its attempt to cut through the maze of database detail and yet hit hard at the basic concepts while surveying the entire field, this book may generalize or oversimplify a few points. I simply suggest that the tradeoff is well worth it. The readers of this book who need only a basic understanding of database will benefit from the tradeoff, while those who expect to go on in database will, after building a solid foundation here, have plenty of opportunity to go into more detail and be more precise in further courses, work experiences, and books. This book is intended to be a "user-friendly" introduction to the (entire) field of database-period.

A number of my database colleagues reviewed chapters of this book, making invaluable comments and suggestions in the process. Chief among them were Charley Bontempo and Judy King, both of whom read large sections of the book and were always available to discuss ideas about its construction. Additionally, I would like to acknowledge Art Amman, Jake Ever, Chuck Haspel, Alice Jones, Jack Lebow, Mike McGuire, Fred Page, Dick Schlough, John Sears, and Moshe Zloof. My thanks also to those students of the IBM Systems Research Institute and of Pace University in New York City who read and commented on the manuscript and tested the exercises at the ends of the chapters.

I would also like to thank the management of the IBM Systems Research Institute, specifically Allen L. Morton, Jr., Program Director, Alfred M. Pietrasanta, Director, and Robert W. DeSio, Director of The Corporate Technical Institutes, for providing a stimulating and supportive atmosphere for the writing of this book.

To Jim Gaughan and the crew at John Wiley & Sons, my thanks for their support of this project and their professionalism in publishing.

Lastly, I am grateful to Irene Gillenson, Edith and Marty Sherr, and Marion and Jack Zack, for a very special kind of help given to me while I was writing this book.

The views expressed in this book are those of the author and do not necessarily reflect those of the IBM Corporation.

Mark L. Gillenson

New York, New York

July 1985

Contents

PREFACE	xvii
PREFACE TO THE FIRST EDITION	xix
CHAPTER 1: INTRODUCTION	1
Introduction,	1
Data and Computers,	1
<i>The Origins of Record Keeping,</i>	1
<i>Data Through the Ages,</i>	3
<i>Early Calculating Devices,</i>	4
<i>Practical Large-Scale Calculating,</i>	8
<i>Commercial Computers,</i>	13
<i>Modern Data Storage Media,</i>	16
<i>Today's Data Processing Environment,</i>	19
The Information Age,	24
This Book,	25
References,	25
Questions and Exercises,	26

CHAPTER 2: SIMPLE FILES AND STORAGE MEDIA 27

- What is Data?, 27
 - Records and Entities*, 27
 - Relationships*, 29
 - Simple Files*, 30
- Data Storage Concepts, 30
 - Data Representations*, 30
 - Primary Memory*, 32
 - Secondary Memory*, 33
- Storing Simple Files, 40
- References, 40
- Questions and Exercises, 40

CHAPTER 3: FILE ORGANIZATIONS AND ACCESS METHODS 43

- Accessing Data, 43
- File Organizations and Access Methods, 44
 - Introduction*, 44
 - Simple Linear Files*, 44
 - Indexed Files*, 46
 - Indexed-Sequential Files*, 51
 - Hashed Files*, 62
- Conclusion, 67
- References, 67
- Questions and Exercises, 67

CHAPTER 4: BENEFITS OF THE DATABASE APPROACH 70

- Introduction, 70
- Data Processing Environment Characteristics, 71
 - Data as a Manageable Resource*, 71
 - Standardization and Specialization*, 72
- Redundant Data, 72
 - Redundancy within One File*, 73
 - Redundancy among Many Files*, 76
- Multiple Relationships, 76

Data Management Control Topics,	79
Security,	79
Backup and Recovery,	80
Concurrency,	80
Auditability,	80
Independent Files and General Data Management Issues,	81
Data Independence,	81
Conclusion,	81
References,	82
Questions and Exercises,	82

CHAPTER 5: DATABASE MANAGEMENT SYSTEMS CHARACTERISTICS

83

The Nature of a Database Management System,	83
Data Definition Languages,	86
Introduction,	86
The Four Approaches to Structure,	87
Data Manipulation Languages,	92
Embedded Statements,	93
Query Languages,	93
Comparing Embedded Statements to Query Languages,	93
Security, Backup and Recovery, and Concurrency,	94
Data Communications Interface,	95
References,	95
Questions and Exercises,	96

CHAPTER 6: THE RELATIONAL APPROACH TO DATABASE

97

Introduction,	97
Review of the Database Premise,	98
The Relational Approach,	99
What is a Relation?,	99
Extracting Data from a Relation,	101
Extracting Data across Multiple Relations: Data Integration,	105
Foreign Keys and Multiple Relationships,	107
Adding, Deleting, or Updating Data: Referential Integrity,	110
Performance Characteristics,	111

References,	112
Questions and Exercises,	112

CHAPTER 7: STRUCTURED QUERY LANGUAGE (SQL) 114

SQL Concepts,	114
Introduction,	114
The SQL Data Structure,	115
SQL Performance and Query Optimizer Concepts,	115
The SQL Language,	117
Introduction,	117
SQL Simple Single Table Retrievals,	120
SQL Built-In Functions,	125
SQL Multiple Table Retrievals,	127
SQL Update, Delete, and Insert Operations,	132
SQL Maintenance Commands,	133
Embedded SQL,	135
References,	136
Questions and Exercises,	136

CHAPTER 8: OTHER RELATIONAL TOOLS 140

Introduction,	140
Query-by-Example (QBE),	141
Introduction,	141
QBE Simple Single File Retrievals,	141
QBE Built-In Functions,	146
QBE Multiple File Retrievals (Joins),	148
QBE Update, Delete, and Insert Operations,	152
Query Management Facility (QMF),	153
Introduction,	153
The QMF Report Generator,	154
dBASE III+ and dBASE IV,	154
Introduction,	156
dBASE Simple Single File Retrievals,	158
dBASE Built-In Functions,	162
dBASE Multiple File Retrievals (Joins),	163
dBASE Update, Delete, and Insert Operations,	167
Other dBASE Commands,	168
References,	168
Questions and Exercises,	168

CHAPTER 9: THE PSEUDO-RELATIONAL APPROACH TO DATABASE**172**

Introduction, 172

Pseudo-Relational Database Concepts, 173

IBM AS/400-System/38 Database, 175

*Introduction, 175**Data Retrieval and Manipulation, 176**Physical and Logical Files, 176**The Binary Index Tree, 180*

References, 186

Questions and Exercises, 186

CHAPTER 10: THE HIERARCHICAL APPROACH TO DATABASE**188**

Introduction, 188

Hierarchies, 189

IMS Physical Structure, 190

*Simple Hierarchies, 191**Logical Relationships, 198**Logical versus Physical Databases, 206**Access Methods, 209**Secondary Indexes, 210**Control Blocks, 213*

DL/I Language Interface, 214

*Introduction, 214**The Calling Mechanism, 214**Direct Retrieval, 215**Sequential Retrieval, 218**Insertion, Deletion, and Update, 220**Additional Topics, 222**Application Development Facility (ADF), 223*

References, 226

Questions and Exercises, 227

CHAPTER 11: THE NETWORK APPROACH TO DATABASE**231**

Introduction, 231

Networks, 232

CODASYL Physical Structure, 234*The Set Concept, 234**Juncture Records, 239**Schemas and Subschemas, 241**Access Methods, 242***CODASYL Language Interface, 243***Introduction, 243**Preliminary Topics, 244**The FIND Statement, 245**The GET Statement, 250**Record Maintenance Statements, 250**Set Maintenance Statements, 252***Comparison Between the Relational and Other Approaches, 253***Points of Comparison, 253**A Join Is a Join Is a Join, 253**Flexibility in Design and Modification, 255**Performance Characteristics, 256**Integrity Exposure on Insertion and Deletion, 257***References, 258****Questions and Exercises, 258****CHAPTER 12: DATABASE DESIGN****261****The Concept of Database Design, 261***Introduction, 261**Information Systems Strategic Planning, 262**Systems Analysis, 264***The Need for Database Design, 269***Intrafile and Multifile Redundancy, 269**Performance, 270**Data Independence, 270**Data Security, 271**Ease of Programming, 271***Database Design Methodologies, 271***Data Normalization and Data Structuring, 272**The Entity-Relationship Model, 287**The Date-Wilson Technique, 290***Other Tools for Database Design, 292***Automated Database Design Aids, 293**Performance Prediction and Monitoring Aids, 293*

Conclusion,	293
References,	294
Questions and Exercises,	295

CHAPTER 13: DATA DICTIONARIES

300

Introduction,	300
Definition and Characteristics,	301
Dictionary Contents,	304
<i>Entities,</i>	304
<i>Attributes,</i>	305
<i>Relationships,</i>	305
Characteristics and Features,	306
Input Forms,	307
<i>Batch Forms,</i>	307
<i>On-Line Commands,</i>	308
<i>Full Screen Panels,</i>	308
<i>Extraction,</i>	309
Output Forms,	309
Implementation,	311
<i>Benefits,</i>	311
<i>Costs,</i>	313
<i>Requirements for Success,</i>	313
Active versus Passive Dictionaries,	314
Relational DBMS Catalogs,	315
References,	317
Questions and Exercises,	317

CHAPTER 14: DATABASE CONTROL TOPICS

320

Introduction,	320
Security,	320
<i>Security Breaches,</i>	320
<i>Security Measures,</i>	322
<i>SQL Security,</i>	324
Concurrency,	325
<i>The Concurrency Problem,</i>	325
<i>Locks and Deadlock,</i>	327

Backup and Recovery, 329

Introduction, 329

Forward Recovery, 330

Backward Recovery, 331

Auditability, 332

References, 333

Questions and Exercises, 334

CHAPTER 15: DATA ADMINISTRATION

335

Introduction, 335

Advantages of the Data Administration Approach, 337

Data as a Shared Corporate Resource, 337

Efficiency in Job Specialization, 338

Data Detail Lifted From Programmers, 338

Operational Management of Data, 339

Data Administration Responsibilities, 340

Data Management Responsibilities, 340

Database Administration Responsibilities, 343

Data Administration Organizational Placement, 345

Data Administration Organizational Impact, 347

Data Administration and New Computing Technologies, 348

Information Resource Management, 349

References, 349

Questions and Exercises, 350

CHAPTER 16: EVOLVING TOPICS

352

Introduction, 352

Database Machines, 352

The Database Machine Concept, 352

Database Machine Approaches, 353

Database Machines Analysis, 355

Distributed Databases, 357

The Distributed Database Concept, 357

The Placement of the Data, 359

Integrity: Updating and Controlling Duplicate Data, 362

Multisite Joins, 365

Directory Management, 365

Object-Oriented Databases,	367
Expert Databases and Knowledge Bases,	368
References,	369
Questions and Exercises,	370
 CHAPTER 17: THE FUTURE	 372
The Growth of Database,	372
Future Application Areas,	373
References,	376
Questions and Exercises,	376
 GLOSSARY	 378
 INDEX	 385