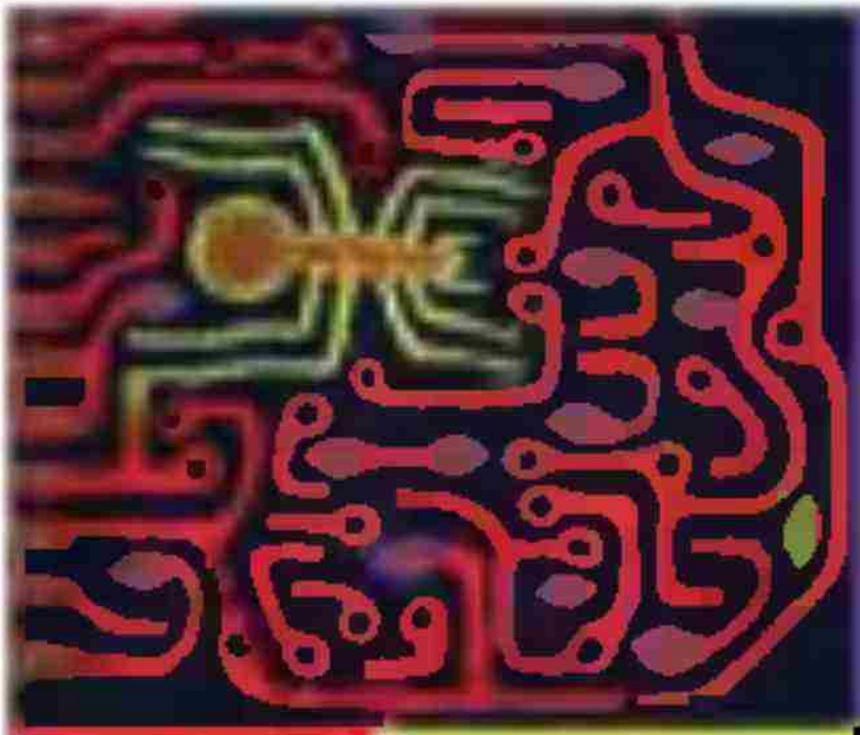

[redacted]

An Introduction to

Database Systems

C.J. Date

EIGHTH EDITION

An Introduction to

Database Systems

The material covered in *An Introduction to Database Systems* is organized into six major parts:

- **Part I** (four chapters) provides a broad introduction to the concepts of database systems in general and relational systems in particular. It also introduces the standard database language, SQL.
- **Part II** (six chapters) consists of a detailed and very careful description of the relational model, which is not only the theoretical foundation underlying relational systems but is, in fact, the theoretical foundation for the entire database field.
- **Part III** (four chapters) discusses the general question of database design. Three chapters are devoted to design theory, and the fourth considers semantic modeling and the entity/relationship model.
- **Part IV** (two chapters) is concerned with transaction management (i.e., recovery and concurrency controls).
- **Part V** (eight chapters) shows how relational concepts are relevant to a variety of further aspects of database technology—security, distributed databases, temporal data, decision support, and so on.
- **Part VI** (three chapters) describes the impact of object technology on database systems. Chapter 25 describes object systems specifically; Chapter 26 considers the possibility of a *rapprochement* between object and relational technologies and discusses object/relational systems; and Chapter 27 addresses the relevance to databases of XML.

About the Author

C. J. DATE is an author, lecturer, researcher, and independent consultant specializing in relational database systems. An active member of the database community for nearly 35 years, C. J. Date devotes the major part of his career to exploring, expanding, and expounding the theory and practice of relational technology. He enjoys a reputation second to none for his ability to explain complex technical material in a clear and understandable fashion.

"[C. J. Date's] book is the flag bearer of relational theory and mathematical treatment in general...as well as the runaway leader in discussing the SQL standards. It exercises much more respect for careful language and the importance of concepts and principles in gaining mastery of the field."

—CARL ECKBERG, *San Diego State University*

"[The] 8th Edition is an excellent and comprehensive presentation of the contemporary database field. In particular, Date's chapters on types, relations, object databases, and object-relational databases together provide an exceptionally clear, self-contained exposition of the object-relational approach to databases."

—MARTIN K. SOLOMON, *Florida Atlantic University*

"Chris Date is the computer industry's most respected expert and thinker on database technology, and his book *An Introduction to Database Systems* continues to be the definitive work for those wanting a comprehensive and current guide to database systems."

—COLIN J. WHITE, *President, Intelligent Business Strategies*

"This is the best explanation of concurrency that I have seen in literature, and it covers the ground quite thoroughly."

—BRUCE O. LARSEN, *Stevens Institute of Technology*

"...both an indispensable read and an indispensable reference. No serious information systems or database practitioner should be without this book."

—DECLAN BRADY, MICS, *Systems Architect and Database Specialist, Fujitsu*

"The author's deep insights into the area, informal treatment of profound topics, open-ended discussions of critical issues, comprehensive and up-to-date contents, as well as rich annotations on bibliography have made the book most popular in the database area for more than two decades."

—QIANG ZHU, *The University of Michigan, Dearborn*

"[The book's] appeal is its comprehensiveness and the fact that it is very up-to-date with research developments. The latter factor is due mainly to [Date's] involvement with these developments, which gives him a unique opportunity to write about them."

—DAVID LIVINGSTONE, *University of Northumbria at Newcastle*

Senior Acquisitions Editor: Maite Suarez-Rivas
Project Editor: Katherine Harutunian
Marketing Manager: Nathan Schultz
Production Supervisor: Marilyn Lloyd
Project Management: Elisabeth Beller
Composition: Nancy Logan
Technical Art: Dartmouth Publishing, Inc.
Copyeditor: Daril Bentley
Proofreader: Jennifer McClain
Design Manager: Joyce Cosentino Wells
Cover Design: Night & Day Design
Cover Image: Lindy Date
Prepress and Manufacturing: Caroline Fell

Access the latest information about Addison-Wesley titles from our World Wide Web site:
http://www.aw.com/cs

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and Addison-Wesley was aware of a trademark claim, the designations have been printed in initial caps or all caps.

The programs and applications presented in this book have been included for their instructional value. They have been tested with care but are not guaranteed for any purpose. The publisher does not offer any warranties or representations, nor does it accept any liabilities with respect to the programs or applications.

If you purchased this book within the United States or Canada you should be aware that it has been wrongfully imported without the approval of the Publisher or the Author.

Copyright © 2004 by Pearson Education, Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher. Printed in the United States of America.

ISBN 0-321-18956-6

2 3 4 5 6 7 8 9 10-HAM-06050403

This book is dedicated to my wife Lindy
and to the memory of my mother Rene—

also to the memory of Ted Codd, who, sadly,
passed away as this book was going to press

Those who cannot remember the past
are condemned to repeat it

Usually quoted in the form:

Those who don't know history are
doomed to repeat it

—George Santayana

I would like to see computer science
teaching set deliberately
in a historical framework. . .
Students need to understand
how the present situation has come
about, what was tried,
what worked and what did not, and
how improvements in hardware
made progress possible. The absence
of this element in
their training causes people to
approach every problem from
first principles. They are apt to
propose solutions that
have been found wanting in the past.
Instead of standing
on the shoulders of their precursors,
they try to go it alone.

—Maurice V. Wilkes

About the Author

C. J. Date is an independent author, lecturer, researcher, and consultant, specializing in relational database technology. He is based in Healdsburg, California.

In 1967, following several years as a mathematical programmer and programming instructor for Leo Computers Ltd. (London, England), Mr. Date moved to the IBM (UK) Development Laboratories, where he worked on the integration of database functionality into PL/I. In 1974 he transferred to the IBM Systems Development Center in California, where he was responsible for the design of a database language known as the Unified Database Language, UDL, and worked on technical planning and externals design for the IBM products SQL/DS and DB2. He left IBM in May, 1983.

Mr. Date has been active in the database field for well over 30 years. He was one of the first people anywhere to recognize the significance of Codd's pioneering work on the relational model. He has lectured widely on technical subjects—principally on database topics, and especially on relational database—throughout North America and also in Europe, Australia, Latin America, and the Far East. In addition to the present book, he is author or coauthor of a number of other database texts, including, from Morgan Kaufmann, *Temporal Data and the Relational Model* (2003) and, from Addison-Wesley, *Foundation for Future Database Systems: The Third Manifesto* (2nd edition, 2000), a detailed proposal for the future direction of the field; *Database: A Primer* (1983), which treats database systems from the nonspecialist's point of view; a series of *Relational Database Writings* books (1986, 1990, 1992, 1995, 1998), which deal with various aspects of relational technology in depth; and another series of books on specific systems and languages—*A Guide to DB2* (4th edition, 1993), *A Guide to SYBASE and SQL Server* (1992), *A Guide to SQL/DS* (1988), *A Guide to INGRES* (1987), and *A Guide to the SQL Standard* (4th edition, 1997). His books have been translated into several languages, including Braille, Chinese, Dutch, French, German, Greek, Italian, Japanese, Korean, Polish, Portuguese, Russian, and Spanish.

Mr. Date has also produced over 300 technical articles and research papers and has made a variety of original contributions to database theory. For several years, he was a regular columnist for the magazine *Database Programming & Design*. He also contributes regularly to the website *http://dbdebunk.com*. His professional seminars on database technology, offered both in North America and overseas, are widely considered to be second to none for the quality of the subject matter and the clarity of the exposition.

Mr. Date holds an Honours Degree in Mathematics from Cambridge University, England (BA 1962, MA 1966) and the honorary degree of Doctor of Technology from De Montfort University, England (1994).

Contents

Preface to the Eighth Edition xxi

PART I PRELIMINARIES 1

Chapter 1 An Overview of Database Management 3

1.1 Introduction 3
1.2 What Is a Database System? 6
1.3 What Is a Database? 11
1.4 Why Database? 16
1.5 Data Independence 20
1.6 Relational Systems and Others 26
1.7 Summary 28
Exercises 29
References and Bibliography 31

Chapter 2 Database System Architecture 33

2.1 Introduction 33
2.2 The Three Levels of the Architecture 34
2.3 The External Level 37
2.4 The Conceptual Level 39
2.5 The Internal Level 40
2.6 Mappings 41
2.7 The Database Administrator 42
2.8 The Database Management System 44
2.9 Data Communications 48
2.10 Client/Server Architecture 49
2.11 Utilities 51
2.12 Distributed Processing 51
2.13 Summary 55
Exercises 56
References and Bibliography 56

Chapter 3 An Introduction to Relational Databases 59

3.1 Introduction 59
3.2 An Informal Look at the Relational Model 60
3.3 Relations and Relvars 64
3.4 What Relations Mean 66
3.5 Optimization 69
3.6 The Catalog 71
3.7 Base Relvars and Views 72
3.8 Transactions 76
3.9 The Suppliers-and-Parts Database 77
3.10 Summary 79
Exercises 81
References and Bibliography 81

Chapter 4 An Introduction to SQL 85

4.1 Introduction 85
4.2 Overview 86
4.3 The Catalog 89
4.4 Views 90
4.5 Transactions 91
4.6 Embedded SQL 91
4.7 Dynamic SQL and SQL/CLI 97
4.8 SQL Is Not Perfect 100
4.9 Summary 101
Exercises 102
References and Bibliography 104

PART II THE RELATIONAL MODEL 109

Chapter 5 TYPES 111

5.1 Introduction 111
5.2 Values *vs.* Variables 112
5.3 Types *vs.* Representations 115
5.4 Type Definition 119
5.5 Operators 122
5.6 Type Generators 127
5.7 SQL Facilities 128
5.8 Summary 136
Exercises 137
References and Bibliography 139

Chapter 6 Relations 141

6.1 Introduction 141
6.2 Tuples 141
6.3 Relation Types 146
6.4 Relation Values 148
6.5 Relation Variables 156
6.6 SQL Facilities 161
6.7 Summary 167
Exercises 168
References and Bibliography 170

Chapter 7 Relational Algebra 173

7.1 Introduction 173
7.2 Closure Revisited 175
7.3 The Original Algebra: Syntax 177
7.4 The Original Algebra: Semantics 180
7.5 Examples 190
7.6 What Is the Algebra For? 192
7.7 Further Points 194
7.8 Additional Operators 195
7.9 Grouping and Ungrouping 203
7.10 Summary 206
Exercises 207
References and Bibliography 209

Chapter 8 Relational Calculus 213

8.1 Introduction 213
8.2 Tuple Calculus 215
8.3 Examples 223
8.4 Calculus *vs.* Algebra 225
8.5 Computational Capabilities 230
8.6 SQL Facilities 231
8.7 Domain Calculus 240
8.8 Query-By-Example 242
8.9 Summary 247
Exercises 248
References and Bibliography 250

Chapter 9 Integrity 253

9.1 Introduction 253
9.2 A Closer Look 255
9.3 Predicates and Propositions 258
9.4 Relvar Predicates and Database Predicates 259
9.5 Checking the Constraints 260
9.6 Internal *vs.* External Predicates 261
9.7 Correctness *vs.* Consistency 263
9.8 Integrity and Views 265
9.9 A Constraint Classification Scheme 266
9.10 Keys 268
9.11 Triggers (a Digression) 277
9.12 SQL Facilities 279
9.13 Summary 284
Exercises 285
References and Bibliography 288

Chapter 10 Views 295

10.1 Introduction 295
10.2 What Are Views For? 298
10.3 View Retrievals 302
10.4 View Updates 303
10.5 Snapshots (a Digression) 318
10.6 SQL Facilities 320
10.7 Summary 323
Exercises 324
References and Bibliography 325

PART III DATABASE DESIGN 329

Chapter 11 Functional Dependencies 333

11.1 Introduction 333
11.2 Basic Definitions 334
11.3 Trivial and Nontrivial Dependencies 337
11.4 Closure of a Set of Dependencies 338
11.5 Closure of a Set of Attributes 339

11.6 Irreducible Sets of Dependencies 341
11.7 Summary 343
Exercises 344
References and Bibliography 345

Chapter 12 Further Normalization I: 1NF, 2NF, 3NF, BCNF 349

12.1 Introduction 349
12.2 Nonloss Decomposition and Functional Dependencies 353
12.3 First, Second, and Third Normal Forms 357
12.4 Dependency Preservation 364
12.5 Boyce/Codd Normal Form 367
12.6 A Note on Relation-Valued Attributes 373
12.7 Summary 375
Exercises 376
References and Bibliography 378

Chapter 13 Further Normalization II: Higher Normal Forms 381

13.1 Introduction 381
13.2 Multi-valued Dependencies and Fourth Normal Form 382
13.3 Join Dependencies and Fifth Normal Form 386
13.4 The Normalization Procedure Summarized 391
13.5 A Note on Denormalization 393
13.6 Orthogonal Design (a Digression) 395
13.7 Other Normal Forms 398
13.8 Summary 400
Exercises 401
References and Bibliography 402

Chapter 14 Semantic Modeling 409

14.1 Introduction 409
14.2 The Overall Approach 411
14.3 The E/R Model 414
14.4 E/R Diagrams 418
14.5 Database Design with the E/R Model 420
14.6 A Brief Analysis 424
14.7 Summary 428
Exercises 429
References and Bibliography 430

PART IV TRANSACTION MANAGEMENT 443

Chapter 15 Recovery 445

15.1 Introduction 445
15.2 Transactions 446
15.3 Transaction Recovery 450
15.4 System Recovery 453
15.5 Media Recovery 455
15.6 Two-Phase Commit 456
15.7 Savepoints (a Digression) 457
15.8 SQL Facilities 458
15.9 Summary 459
Exercises 460
References and Bibliography 460

Chapter 16 Concurrency 465

16.1 Introduction 465
16.2 Three Concurrency Problems 466
16.3 Locking 470
16.4 The Three Concurrency Problems Revisited 472
16.5 Deadlock 474
16.6 Serializability 476
16.7 Recovery Revisited 478
16.8 Isolation Levels 480
16.9 Intent Locking 483
16.10 Dropping ACID 485
16.11 SQL Facilities 490
16.12 Summary 491
Exercises 492
References and Bibliography 494

PART V FURTHER TOPICS 501

Chapter 17 Security 503

17.1 Introduction 503
17.2 Discretionary Access Control 506

17.3 Mandatory Access Control 511
17.4 Statistical Databases 513
17.5 Data Encryption 519
17.6 SQL Facilities 523
17.7 Summary 527
Exercises 528
References and Bibliography 529

Chapter 18 Optimization 531

18.1 Introduction 531
18.2 A Motivating Example 533
18.3 An Overview of Query Processing 534
18.4 Expression Transformation 539
18.5 Database Statistics 544
18.6 A Divide-and-Conquer Strategy 545
18.7 Implementing the Relational Operators 548
18.8 Summary 553
Exercises 554
References and Bibliography 557

Chapter 19 Missing Information 575

19.1 Introduction 575
19.2 An Overview of the 3VL Approach 577
19.3 Some Consequences of the Foregoing Scheme 582
19.4 Nulls and Keys 586
19.5 Outer Join (a Digression) 589
19.6 Special Values 591
19.7 SQL Facilities 592
19.8 Summary 597
Exercises 598
References and Bibliography 600

Chapter 20 Type Inheritance 605

20.1 Introduction 605
20.2 Type Hierarchies 610
20.3 Polymorphism and Substitutability 613
20.4 Variables and Assignments 617
20.5 Specialization by Constraint 621
20.6 Comparisons 623

20.7 Operators, Versions, and Signatures 626
20.8 Is a Circle an Ellipse? 630
20.9 Specialization by Constraint Revisited 634
20.10 SQL Facilities 636
20.11 Summary 641
Exercises 642
References and Bibliography 644

Chapter 21 Distributed Databases 647

21.1 Introduction 647
21.2 Some Preliminaries 648
21.3 The Twelve Objectives 652
21.4 Problems of Distributed Systems 660
21.5 Client/Server Systems 671
21.6 DBMS Independence 674
21.7 SQL Facilities 679
21.8 Summary 680
Exercises 681
References and Bibliography 682

Chapter 22 Decision Support 689

22.1 Introduction 689
22.2 Aspects of Decision Support 691
22.3 Database Design for Decision Support 693
22.4 Data Preparation 701
22.5 Data Warehouses and Data Marts 704
22.6 Online Analytical Processing 709
22.7 Data Mining 717
22.8 SQL Facilities 719
22.9 Summary 720
Exercises 721
References and Bibliography 722

Chapter 23 Temporal Databases 727

23.1 Introduction 727
23.2 What Is the Problem? 732
23.3 Intervals 737
23.4 Packing and Unpacking Relations 743
23.5 Generalizing the Relational Operators 754
23.6 Database Design 758

23.7 Integrity Constraints 764
23.8 Summary 770
Exercises 771
References and Bibliography 772

Chapter 24 Logic-Based Databases 775

24.1 Introduction 775
24.2 Overview 776
24.3 Propositional Calculus 778
24.4 Predicate Calculus 783
24.5 A Proof-Theoretic View of Databases 789
24.6 Deductive Database Systems 793
24.7 Recursive Query Processing 798
24.8 Summary 803
Exercises 806
References and Bibliography 807

PART VI OBJECTS, RELATIONS, AND XML 811

Chapter 25 Object Databases 813

25.1 Introduction 813
25.2 Objects, Classes, Methods, and Messages 817
25.3 A Closer Look 822
25.4 A Cradle-to-Grave Example 830
25.5 Miscellaneous Issues 840
25.6 Summary 847
Exercises 850
References and Bibliography 851

Chapter 26 Object/Relational Databases 859

26.1 Introduction 859
26.2 The First Great Blunder 862
26.3 The Second Great Blunder 870
26.4 Implementation Issues 874
26.5 Benefits of True *Rapprochement* 876
26.6 SQL Facilities 878
26.7 Summary 885
Exercises 885
References and Bibliography 886

Chapter 27 The World Wide Web and XML 895

27.1 Introduction 895
27.2 The Web and the Internet 896
27.3 An Overview of XML 897
27.4 XML Data Definition 908
27.5 XML Data Manipulation 917
27.6 XML and Databases 925
27.7 SQL Facilities 928
27.8 Summary 932
Exercises 934
References and Bibliography 935

APPENDIXES 939

Appendix A The TransRelational™ Model 941

A.1 Introduction 941
A.2 Three Levels of Abstraction 943
A.3 The Basic Idea 946
A.4 Condensed Columns 952
A.5 Merged Columns 956
A.6 Implementing the Relational Operators 960
A.7 Summary 966
References and Bibliography 966

Appendix B SQL Expressions 967

B.1 Introduction 967
B.2 Table Expressions 968
B.3 Boolean Expressions 973

Appendix C Abbreviations, Acronyms, and Symbols 977

Appendix D Storage Structures and Access Methods (online at *http://www.aw.com/cssupport*)

D.1 Introduction
D.2 Database Access: An Overview
D.3 Page Sets and Files
D.4 Indexing

D.5 Hashing
D.6 Pointer Chains
D.7 Compression Techniques
D.8 Summary
Exercises
References and Bibliography

Index I-1

Preface to the Eighth Edition

This book is a comprehensive introduction to the now very large field of database systems. It provides a solid grounding in the foundations of database technology and gives some idea as to how the field is likely to develop in the future. The book is meant primarily as a textbook, not a work of reference (though I hope and believe it can be useful as a reference also, to some extent). The emphasis throughout is on insight and understanding, not just on formalisms.

PREREQUISITES

The book as a whole is meant for anyone professionally interested in computing in some way who wants to gain an understanding of what database systems are all about. I assume you have at least a basic knowledge of both:

- The storage and file management capabilities (indexing, etc.) of a modern computer system
- The features of at least one high-level programming language (Java, Pascal, PL/I, etc.)

Regarding the first of these prerequisites, however, please note that a detailed tutorial on such matters can be found in the online Appendix D, "Storage Structures and Access Methods."

STRUCTURE

I have to say that I am a little embarrassed at the size of this book. The fact is, however, that database technology has become a very large field, and it is not possible to do it justice in fewer than 1,000 pages or so (indeed, most of the book's competitors are also about 1,000 pages). Be that as it may, the book overall is divided into six major parts:

I. Basic Concepts
II. The Relational Model
III. Database Design
IV. Transaction Management

V. Further Topics

VI. Objects, Relations, and XML

Each part in turn is divided into several chapters:

- Part I (four chapters) provides a broad introduction to the concepts of database systems in general and relational systems in particular. It also introduces the standard database language SQL.
- Part II (six chapters) consists of a detailed and very careful description of the relational model, which is not only the theoretical foundation underlying relational systems, but is in fact the theoretical foundation for the entire database field.
- Part III (four chapters) discusses the general question of database design; three chapters are devoted to design theory, the fourth considers semantic modeling and the entity/relationship model.
- Part IV (two chapters) is concerned with transaction management (i.e., recovery and concurrency controls).
- Part V (eight chapters) is a little bit of a *potpourri*. In general, however, it shows how relational concepts are relevant to a variety of further aspects of database technology—security, distributed databases, temporal data, decision support, and so on.
- Finally, Part VI (three chapters) describes the impact of object technology on database systems. Chapter 25 describes object systems specifically; Chapter 26 considers the possibility of a *rapprochement* between object and relational technologies and discusses object/relational systems; and Chapter 27 addresses the relevance to databases of XML.

There are also four appendixes. The first is an overview of a dramatic new and radically different implementation technology called The TransRelational™ Model; the next gives a BNF grammar for SQL expressions; the third contains a glossary of the more important abbreviations, acronyms, and symbols introduced in the body of the text; and the last is, as already explained, an online tutorial on storage structures and access methods.

ONLINE MATERIALS

The following instructor supplements are available only to qualified instructors. For information on accessing them, please contact your local Addison-Wesley Sales Representative, or send e-mail to *aw.cse@aw.com*.

- An Instructor's Manual provides guidance on how to use the book as a basis for teaching a database course. It consists of a series of notes, hints, and suggestions on each part, chapter, and appendix, as well as other supporting material.
- Answers to exercises (included in Instructor's Manual)
- Lecture slides in PowerPoint format

The following supplements are available to all readers of this book at *http://www.aw.com/cssupport*.

- Appendix D on storage structures and access methods (as already mentioned)
- Answers to a selected subset of the exercises

HOW TO READ THIS BOOK

The book overall is meant to be read in sequence more or less as written, but you can skip later chapters, and later sections within chapters, if you choose. A suggested plan for a first reading would be:

- Read Chapters 1 and 2 "once over lightly."
- Read Chapters 3 and 4 very carefully (except perhaps for Sections 4.6 and 4.7).
- Read Chapter 5 "once over lightly."
- Read Chapters 6, 7, 9, and 10 carefully, but skip Chapter 8 (except perhaps for Section 8.6 on SQL).
- Read Chapter 11 "once over lightly."
- Read Chapters 12 and 14 carefully,[1] but skip Chapter 13.
- Read Chapters 15 and 16 carefully (except perhaps for Section 15.6 on two-phase commit).
- Read subsequent chapters selectively (but in sequence), according to taste and interest.

Each chapter opens with an introduction and closes with a summary: in addition, most chapters include exercises, and the online answers often give additional information about the topic of the exercise. Most chapters also include a set of references, many of which are annotated. This structure allows the subject matter to be treated in a layered fashion, with the most important concepts and results being presented "in line" in the main body of the text and various subsidiary issues and more complex aspects being deferred to the exercises or answers or references (as appropriate). *Note:* References are identified by two-part numbers in square brackets. For example, the reference "[3.1]" refers to the first item in the list of references at the end of Chapter 3: namely, a paper by E. F. Codd published in *CACM 25*, No. 2, in February, 1982. (For an explanation of abbreviations used in references—e.g., "CACM"—see Appendix C.)

COMPARISON WITH EARLIER EDITIONS

We can summarize the major differences between this edition and its immediate predecessor as follows:

- *Part I:* Chapters 1–4 cover roughly the same ground as Chapters 1–4 in the seventh edition, but they have been significantly revised at the detail level. In particular,

[1] You could also read Chapter 14 earlier if you like, possibly right after Chapter 4.

Chapter 4, the introduction to SQL, has been upgraded to the level of the current standard SQL:1999, as indeed has SQL coverage throughout the entire book. (This fact all by itself caused major revisions to more than half the chapters from the seventh edition.) *Note:* Facilities likely to be included in the next version of the standard—which will probably be ratified in late 2003—are also mentioned where appropriate.

- *Part II:* Chapters 5–10, on the relational model, are a totally rewritten, considerably expanded, and very much improved version of Chapters 5–9 from the seventh edition. In particular, the material on types—also known as domains—has been expanded into a chapter of its own (Chapter 5), and the material on integrity (Chapter 9) has been completely restructured and rethought. I hasten to add that the changes in these chapters do not represent changes in the underlying concepts but, rather, changes in how I have chosen to present them, based on my practical experience in teaching this material in live presentations.

  Note: Some further words of explanation are in order here. Earlier editions of the book used SQL as a basis for teaching relational concepts, in the belief that it was easier on the student to show the concrete before the abstract. Unfortunately, however, the gulf between SQL and the relational model grew and continued to grow, ultimately reaching a point where I felt it would be actively misleading to use SQL for such a purpose any longer. The sad truth is that SQL is now so far from being a true embodiment of relational principles—it suffers from so many sins of both omission and commission—that I would frankly prefer not to discuss it at all! However, SQL is obviously important from a commercial point of view; thus, every database professional needs to have some familiarity with it, and it would just not be appropriate to ignore it in a book of this nature. I therefore settled on the strategy of including (a) a chapter on SQL basics in Part I of the book and (b) individual sections in other chapters, as applicable, describing those aspects of SQL that are specific to the subject of the chapter in question. In this way the book still provides comprehensive—indeed, extensive—coverage of SQL material, but puts that coverage into what I feel is the proper context.

- *Part III:* Chapters 10–13 are a mostly cosmetic revision of Chapters 9–12 from the seventh edition. There are detail-level improvements throughout, however.

  Note: Again some further explanation is in order. Some reviewers of earlier editions complained that database design issues were treated too late. But it is my feeling that students are not ready to design databases properly or to appreciate design issues fully until they have some understanding of what databases are and how they are used; in other words, I believe it is important to spend some time on the relational model and related matters before exposing the student to design questions. Thus, I still believe Part III is in the right place. (That said, I do recognize that many instructors prefer to treat the entity/relationship material much earlier. To that end, I have tried to make Chapter 14 more or less self-contained, so that they can bring it in immediately after, say, Chapter 4.)

- *Part IV:* The two chapters of this part, Chapters 15 and 16, are completely rewritten, extended, and improved versions of Chapters 14 and 15 from the seventh edition. In

particular, Chapter 16 now includes a careful analysis of—and some unorthodox conclusions regarding—the so-called ACID properties of transactions.

- *Part V:* Chapter 20 on type inheritance and Chapter 23 on temporal databases have been totally rewritten to reflect recent research and developments in those areas. Revisions to other chapters are mostly cosmetic, though there are improvements in explanations and examples throughout and new material here and there.
- *Part VI:* Chapters 25 and 26 are improved and expanded versions of Chapters 24 and 25 from the seventh edition. Chapter 27 is new.

Finally, Appendix A is also new, while Appendixes B and C are revised versions of Appendixes A and C, respectively, from the seventh edition (the material from Appendix B in that edition has been incorporated into the body of the book). Appendix D is a significantly revised version of Appendix A from the *sixth* edition.

WHAT MAKES THIS BOOK DIFFERENT?

Every database book on the market has its own individual strengths and weaknesses, and every writer has his or her own particular ax to grind. One concentrates on transaction management issues; another stresses entity/relationship modeling; another looks at everything through an SQL lens; yet another takes a pure "object" point of view; still another views the field exclusively in terms of some commercial product; and so on. And, of course, I am no exception to this rule—I too have an ax to grind: what might be called the foundation ax. I believe very firmly that we must get the foundation right, and understand it properly, before we try to build on that foundation. This belief on my part explains the heavy emphasis in this book on the relational model; in particular, it explains the length of Part II—the most important part of the book—where I present my own understanding of the relational model as carefully as I can. I am interested in foundations, not fads and fashions. Products change all the time, but principles endure.

In this regard, I would like to draw your attention to the fact that there are several important ("foundation") topics for which this book, virtually alone among the competition, includes an entire in-depth chapter (or an appendix, in one case). The topics in question include:

- Types
- Integrity
- Views
- Missing information
- Inheritance
- Temporal databases
- The TransRelational™ Model

In connection with that same point (the importance of foundations), I have to admit that the overall tone of the book has changed over the years. The first few editions were mostly descriptive in nature; they described the field as it actually was in practice, "warts and all." Later editions, by contrast, were much more *pre*scriptive; they talked about the way the field *ought* to be and the way it ought to develop in the future, if we did things right. And the present edition is certainly prescriptive in this sense (so it is a text with an attitude!). Since the first part of that "doing things right" is surely educating oneself as to what those right things are, I hope this new edition can help in that endeavor.

Yet another related point: As you might know, I recently published, along with my colleague Hugh Darwen, another "prescriptive" book, *Foundation for Future Database Systems: The Third Manifesto* (reference [3.3] in the present book).[2] That book, which we call *The Third Manifesto* or just the *Manifesto* for short, builds on the relational model to offer a detailed technical proposal for future database systems; it is the result of many years of teaching and thinking about such matters on the part of both Hugh and myself. And, not surprisingly, the ideas of the *Manifesto* permeate the present book. Which is not to say the *Manifesto* is a prerequisite to the present book—it is not; but it *is* directly relevant to much that is in the present book, and further related information is often to be found therein.

Note: Reference [3.3] uses a language called Tutorial D for illustrative purposes, and the present book does the same. Tutorial D syntax and semantics are intended to be more or less self-explanatory (the language might be characterized, loosely, as "Pascal-like"), but individual features are explained when they are first used if such explanation seems necessary.

A CLOSING REMARK

I would like to close these prefatory notes with the following lightly edited extract from another preface—Bertrand Russell's own preface to *The Bertrand Russell Dictionary of Mind, Matter and Morals* (ed., Lester E. Denonn), Citadel Press, 1993, reprinted here by permission:

> *I have been accused of a habit of changing my opinions . . . I am not myself in any degree ashamed of [that habit]. What physicist who was already active in 1900 would dream of boasting that his opinions had not changed during the last half century? . . . [The] kind of philosophy that I value and have endeavoured to pursue is scientific, in the sense that there is some definite knowledge to be obtained and that new discoveries can make the admission of former error inevitable to any candid mind. For what I have said, whether early or late, I do not claim the kind of truth which theologians claim for their creeds. I claim only, at best, that the opinion expressed was a sensible one to hold at the time . . . I should be much surprised if subsequent research did not show that it needed to be modified. [Such opinions were not] intended as pontifical pronouncements, but only as the best I could do at the time towards the promotion of clear and accurate thinking. Clarity, above all, has been my aim.*

[2] There is a website, too: *http://www.thethirdmanifesto.com*. See also *http://www.dbdebunk.com* for much related material.

If you compare earlier editions of this book with this eighth edition, you will find that I too have changed my opinions on many matters (and no doubt will continue to do so). I hope you will accept the remarks just quoted as adequate justification for this state of affairs. I share Bertrand Russell's perception of what the field of scientific inquiry is all about, but he expresses that perception far more eloquently than I could.

ACKNOWLEDGMENTS

Once again it is a pleasure to acknowledge my debt to the many people involved, directly or indirectly, in the production of this book:

- First of all, I must thank my friends David McGoveran and Nick Tindall for their major involvement in this edition; David contributed the first draft of Chapter 22 on decision support, and Nick contributed the first draft of Chapter 27 on XML. I must also thank my friend and colleague Hugh Darwen for major help (in a variety of forms) with all SQL portions of the manuscript. Nagraj Alur and Fabian Pascal also provided me with a variety of technical background material. A special vote of thanks goes to Steve Tarin for inventing the technology described in Appendix A and for his help in getting me to understand it fully.
- Second, the text has benefited from the comments of students on the seminars I have been teaching over the past several years. It has also benefited enormously from the comments of, and discussions with, numerous friends and reviewers, including Hugh Darwen, IBM; Guy de Tré, Ghent University; Carl Eckberg, San Diego State University; Cheng Hsu, Rensselaer Polytechnic Institute; Abdul-Rahman Itani, The University of Michigan-Dearborn; Vijay Kanabar, Boston University; Bruce O. Larsen, Stevens Institute of Technology; David Livingstone, University of Northumbria at Newcastle; David McGoveran, Alternative Technologies; Steve Miller, IBM; Fabian Pascal, independent consultant; Martin K. Solomon, Florida Atlantic University; Steve Tarin, Required Technologies; and Nick Tindall, IBM. Each of these people reviewed at least some part of the manuscript or made technical material available or otherwise helped me find answers to my many technical questions, and I am very grateful to all of them.
- I would also like to thank my wife Lindy for contributing the cover art once again and for her support throughout this and all my other database-related projects over the years.
- Finally, I am grateful (as always) to everyone at Addison-Wesley—especially Maite Suarez-Rivas and Katherine Harutunian—for all of their encouragement and support throughout this project, and to my packager Elisabeth Beller for another sterling job.

Healdsburg, California C. J. Date
2003