
A n I n t r o d u c t i o n

t o

D a t a b a s e S y s t e m s

Eighth Edition

by

C. J. Date

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

*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 magazines *Database Programming & Design* and *Intelligent Enterprise*. 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).

P r e f a c e t o t h e E i g h t h E d i t i o n

*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, I should note that the first few editions of this book included a tutorial on such matters. But file management is not specific to database systems; furthermore, it is a huge subject in its own right, and it has textbooks of its own__see, e.g., *File Organization for Database Design*, by Gio Wiederhold, published by McGraw-Hill in 1987. (Note that this book, despite its title, is really about files, not databases.) For such reasons, I have dropped the coverage of such material from more recent editions.*

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 1000 pages or so (indeed, most of the book's competitors are also around 1000 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 three appendixes. One is an overview of a dramatic new and radically different implementation technology called **The TransRelational™ Model**; one gives a **BNF grammar** for SQL expressions; and the third contains a **glossary** of the more important abbreviations, acronyms, and symbols introduced in the body of the text.

<<< next para OK ??? PPT slides ??? scripts ??? >>>

An online **Instructor's Manual** is also available, giving 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 answers to exercises and other supporting material. For instructions on how to access the Manual, please contact your local Addison-Wesley sales representative. To locate your local representative, please visit us on the Web at <http://hepg.awl.com/rep-locator>.

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,* but skip Chapter 13

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

- 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

The major differences between this edition and its immediate predecessor are summarized below.

- *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, 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 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 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 the old Appendixes A and C, respectively. The material from the old Appendix B has been incorporated into the body of the book.

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 prescriptive; 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).^{*} 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.

^{*} There is a website, too: <http://www.thethirdmanifesto.com>. See also <http://www.dbdebunk.com> for much related material.

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.

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 quoted above 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

<<< following to be completed >>>

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 << list to be supplied ... Stevens Institute ... San Diego State University ... Boston University ... Florida Atlantic University ... >>>. 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 editor Elisabeth Beller for another sterling job.

Healdsburg, California
2003

C. J. Date

***** End of Preface *****