C.J.Date

# Database
# A Primer

C. J. Date

# DATABASE: A PRIMER

Addison-Wesley Publishing Company
Reading, Massachusetts • Menlo Park, California
London • Amsterdam • Don Mills, Ontario • Sydney

This book is in the
Addison-Wesley Microcomputer Books
Popular Series
Marshall Henrichs, *Cover Design*
Ellen Klempner, *Design*



**Library of Congress Cataloging in Publication Data**

Date, C J.
. Database· a primer.

Bibliography p
Includes index
1 Data base management I Title
QA76.9.D3D366 1983 001.64'2 83-3786
ISBN 0-201-11358-9



ISBN 0-201-11358-9
ABCDEFGHIJ-HA-89876543

# FOREWORD

This book is a user's introduction to database technology—with the emphasis on the *user*. It is concerned with how database systems are used, rather than with how they work. Database technology is, of course, built on an underlying foundation of computer technology; indeed, as you are probably aware, database systems represent one of the most important application areas for the modern computer. But the book is not really about computers as such at all—it is (to repeat) about how those computers can be used. It treats computers as a means, not an end.

The assumption is, then, that you need to use or understand a database system but have little or no knowledge of how computers work. If that description fits your case, then this book is for you. The fact is, you should not *need* to know much about the computer itself in order to use a database system. (What little you do need will be explained at appropriate points in the text.) It is the great achievement of modern computer technology that computers—in particular, microcomputers—are at last beginning to become tools in the service of people who do not have, and do not need to have, any knowledge of how the machine actually works internally. Fundamentally, there is no more need to know how a computer functions internally in order to use it than there is to know how the internal combustion engine works in order to drive a car. However, it is a fact that, until comparatively recently, only those with very specialized training were capable of exploiting computer systems to the full. But this situation is changing rapidly, for a variety of reasons, one of which is (laudably enough) simple user demand.

So who exactly is this book intended for? The answer is, "All of the following:"

- **Home computer owners**

  In this category I include anyone who owns, or is considering owning, a personal computer as a tool to help with some kind of home filing system (as opposed to someone who is interested in a home computer more as an electronic hobby). The scope for ordinary households in applying such a tool will grow by leaps and bounds over the next ten or fifteen years.

- **Small business owners**

  It is becoming increasingly common to find small corporations acquiring their own (micro- or mini-) computer to help in running their business. Frequently, the fact that there is a database system available on that computer is the overriding reason for using the computer at all. As with the home computer owner above, the person who actually interacts with that system (who may even be the corporation president) will not be a computer specialist, nor will he or she have any desire to be.

- **End-users in a larger corporation**

  In a medium-to-large corporation there will typically be some large central computer and some large number of "end-users," who are certainly not computer specialists themselves but who need to use that computer and are therefore allowed to access it from some remote terminal. Once again, those users will probably have no particular interest in exactly how the machine works, but they *will* be interested in what the database system can do for them and what they have to do to make use of it.

If you fit into any of these three categories, then this book should give you a good idea of what to expect. It describes the kinds of things you will have to do and the kinds of behavior you can expect on the part of the system. But there are others who should also gain from reading this book:

- **User management**

  If you are not a direct user yourself but instead are in the position of being able to request other people (the end-users mentioned above) to obtain information from the system on your behalf, then you should have some idea of what those other people have to do, so that you can communicate with them properly and can understand what is and is not reasonable to expect of them. This book will provide you with that necessary background knowledge (but, of course, you may be able to skip some of the detail).

- **College students**

  The field of database management is (as indicated above) becoming increasingly important, not only in computer science per se but also in many related disciplines such as business administration. A student in a community college or in an undergraduate course in one of those related disciplines should find that the book provides a good grounding in this new field.

- **Home students**

  Finally, the book should prove suitable for the "intelligent amateur"—by which I mean anyone who may be interested (for whatever reason) in studying database technology at home, without any kind of formal instruction (especially if that person has a home system at his or her disposal on which to try out some of the examples and exercises).

The book is *not* intended for data processing professionals such as application programmers or database software specialists—though such readers may perhaps learn something too, inasmuch as they may gain an appreciation of how end-users wish or need to use the systems that they build.

Some of you will be aware that I have already published two other books on the subject of database technology—*An Introduction to Database Systems*, Volume I (Addison-Wesley, 3rd edition, 1981) and Volume II (Addison-Wesley, 1st edition, 1983)—and may be wondering how the present book differs from those existing books. In fact, there is not really very much overlap at all. The present book does necessarily cover some of the same ground, but:

- As already stressed, it is aimed at the user, not the data processing specialist. There is far less technical detail in this book. (What detail there is is detail the user needs in order to use the system properly.) The overall slant is different.
- It assumes less. Specifically, the two existing books assume that the reader is familiar with at least one programming language and understands what is involved in "file processing" in that language. By contrast, the present book assumes simply that the reader is interested in getting the system to do something useful.

- It includes a good deal of different (more user-oriented) material. For example, it covers NOMAD (a very "user friendly" system) and dBASE (one of the most powerful microcomputer systems) in some depth; neither of those systems is touched on at all in the other books. It also devotes an entire chapter to the question of displaying results and formatting reports, and it provides some guidelines for database design (another topic that is not really addressed in the existing books).

One technical point: All of the systems described in this book are *relational* systems. Those of you who follow the trade or technical press will be aware that a debate has been going on for some time in the database world regarding the relative merits of the newer, so-called relational systems versus the older *hierarchical* and *network* systems. There is no need to air the issues in that debate here. Suffice it to say that (a) most database professionals now believe that relational technology is the way of the future—so much so that new users have no need even to be aware of the older technologies—and (b) just about every product announcement in the database field these days is either for an entirely new relational system or for relational enhancements to one of the older systems. In particular, systems for smaller machines (micros or minis) are almost invariably relational. (But if you do wish to understand exactly what it means for a system to be relational, I have included in Appendix A a more technical discussion of the subject, under the heading "The Relational Model." More information on this topic, and specifically on the differences between relational and nonrelational systems, can be found in the two books mentioned above.)

**Acknowledgments:** It is a great pleasure to acknowledge the following people, who assisted me in the preparation of this book in a variety of ways: Roger Reinsch of IBM and Tom Dwyer of the University of Pittsburgh, both of whom read the manuscript in its entirety and made many helpful suggestions; Roger Harvey of IBM, who answered numerous technical questions concerning Query By Example; Brad Whitlock of NCSS, who very kindly furnished me with a copy of the NOMAD *Reference Manual*; and George Tate of Ashton-Tate, who equally kindly supplied me with a copy of the dBASE system for my own use in trying out some of the examples. To all these people, my grateful thanks. I would also like to thank Tom Bell of Addison-Wesley for his cheerful support throughout the entire production process.

Finally, I am pleased to acknowledge the support I have received from IBM in preparing this book. Needless to say, however, the content of the book is entirely my own responsibility; the views expressed are my own and in no way represent an official statement on the part of IBM.

*Saratoga, California* C. J. D.
*June 1983*

# CONTENTS

# PART I

## *Database Access and Maintenance*

# 1 Introduction to Part I

## OBJECTIVES

Database management is one of the most important functions provided by modern computer systems. So important is it, in fact, that very often it is the principal justification for acquiring the computer in the first place. The primary objective of this book is to explain what database management means to you, the user. Thus the aim is to describe what database systems look like *externally*, rather than how they work internally—though we will delve (not very deeply) into some internal aspects from time to time. But first let me make clear what I mean by the term "user."

Broadly speaking, a user is anyone who is interested in using some automated (i.e., computerized) system for maintaining and interrogating data files, on any subject whatsoever, at any level of sophistication. More specifically, the term "user" will be taken throughout this book to mean any one of the following:

1. A person with his or her own private home computer, presumably some kind of microcomputer such as an Apple II or a TRS 80 or an IBM Personal Computer or any of the many others now available;
2. A person in a small business, such as a small accounting corporation or a local retail store, with a micro- or minicomputer to help in running that business;
3. An end-user in a larger organization, such as a bank, industrial company, hospital, or university, who has access via a computer terminal to some large, possibly remote, central computer operated by some specialist data processing department.

These different classes of people, though clearly having different overall requirements, nevertheless have a lot in common when it comes to the question of what they want the database system to do for them. In all cases the requirements are to be able to store data records in the system, to be able to retrieve them on demand, and to be able to update them as and when necessary. Externally,

therefore, systems should look pretty much the same in all three cases, even though internally their structure will probably vary considerably. Since (as already indicated) this book is concerned with externals, there is little need for us to make much distinction among the three cases.

A secondary objective of this book—or, perhaps, a different way of expressing the primary objective—is to *remove the mystery* from database systems. For some reason a considerable mystique has grown up around the topic of database management, and it is not uncommon to find people (even computer professionals) who seem to be almost scared of the subject: "Oh, that's much too complex for me, I'd never understand it, don't you need a math degree before you can use a database?" and so on. We might add that the terminology of the subject tends to foster this attitude: In a field that is generally notorious for its bewildering array of jargon—namely, computer science—database management stands out specifically as one of the worst offenders. Now it is true that database *internals* is a complex subject. It is also true that the job of designing and administering a large, sophisticated database is a complex one. As a result, the tasks of *database system implementation* and *database administration* (in a large organization) are both very specialized, and both require highly skilled people to do them. But those tasks are not performed by *users*, in our sense of the term. The whole point of all that complexity is to make life *easy* for the user. From the user's perspective, database systems are simple. (Perhaps in honesty we should add "or should be"—some of the older systems are not as easy to use as they might be! But the claim is generally becoming truer and truer as time goes on, and it is certainly true of the newer systems.) It is, however, unfortunately still the case that the terminology is sometimes obscure. We will try to lighten some of that obscurity in this book.

## STRUCTURE OF THIS BOOK

As you can see, the book is divided into two parts. Part I consists of ten chapters and Part II consists of eight. The questions addressed in Part I are the ones you should be most immediately interested in—namely, how to get data into the database in the first place and how to get it back out subsequently (and how to change it, when necessary). These aspects are referred to generically as *database access and maintenance*. The topics of Part II, by contrast, are ones you do not need to understand just to start using the system initially, but should know something about as part of your