

DB2:

Concepts, Design, and Programming



DB2:

Concepts, Design, and Programming

JAMES MARTIN

with

Kathleen Kavanagh Chapman

Joe Leben



PRENTICE HALL, Englewood Cliffs, New Jersey 07632

Library of Congress Cataloging-in-Publication Data

Martin, James

DB2, concepts, design, and programming.

"A James Martin book."

Includes index.

1. Data base management. 2. IBM Database 2
(Computer system) I. Chapman, Kathleen Kavanagh.
II. Leben, Joe. III. Title.

QA76.9.D3M363 1989 005.75'65 88-30727

ISBN 0-13-198581-7

Editorial/production supervision: *Kathryn Gollin Marshak and Karen Skrable Fortgang*

Jacket design: *Bruce Kenselaar*

Manufacturing buyer: *Mary Ann Gloriande*

© 1989 by James Martin

All rights reserved. No part of this book may be
reproduced, in any form or by any means,
without permission in writing from the publisher.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN 0-13-198581-7

PRENTICE-HALL INTERNATIONAL (UK) LIMITED, *London*

PRENTICE-HALL OF AUSTRALIA PTY. LIMITED, *Sydney*

PRENTICE-HALL CANADA INC., *Toronto*

PRENTICE-HALL HISPANOAMERICANA, S.A., *Mexico*

PRENTICE-HALL OF INDIA PRIVATE LIMITED, *New Delhi*

PRENTICE-HALL OF JAPAN, INC., *Tokyo*

SIMON & SCHUSTER ASIA PTE. LTD., *Singapore*

EDITORA PRENTICE-HALL DO BRASIL, LTDA., *Rio de Janeiro*

PREFACE

Databases are playing an increasingly important role in supporting an enterprise's information processing requirements. The development of corporate databases will continue to be an extremely important data processing activity for many years to come. Data will be regarded as a vital corporate resource that must be managed so as to maximize its value. Both the quantity of data stored and the complexity of its organization are increasing by leaps and bounds.

The pointer-structured databases of an earlier era are giving way to *relational* database technology because relational databases are more flexible, are based on formal mathematics, and support a wide diversity of user languages. SQL is now an international standard, and many different user tools produce SQL code that can execute with DB2 or other relational databases.

DB2 is a *relational* database management system (DBMS) and is IBM's primary strategic product in the database area. Much other IBM software is being designed to link to DB2, and this represents a major part of IBM's strategic thrust in software. DB2, along with associated products such as Query Management Facility (QMF) and Data Extract (DXT), provides a wide range of capabilities, suitable for both ad hoc access by end users and the development of production applications by professional programmers.

ACKNOWLEDGMENTS The authors would like to thank Peter Tierney, Vice President of Marketing for Oracle Corporation, who supplied us with a copy of the very fine Oracle relational DBMS. We were able to use Oracle, running on an IBM Personal Computer AT, to initially develop the SQL examples and to validate the content of tables in Chapter 2 and in Appendix I. Two people at IBM contributed to the writing of this book: John Gridley made it possible for us to develop the QBE examples on an IBM main-

frame in Chicago, and Bill Kellow provided the technical information needed to write the parts of the book that deal with the referential integrity features of DB2.

James Martin
Kathleen K. Chapman
Joe Leben

CONTENTS

Preface xv

PART I

INTRODUCTION

1 The DB2 Environment 3

The DB2 Product Family 3; The DB2 Operating Environment 4; The SQL Data Language 5; The DB2 Catalog 5; DB2 Utilities 6; DBMS Characteristics 6; Database Objectives 7; Data Redundancy 8; Data Independence 8; Security 9; Recovery 9; Concurrency 9

2 DB2 and the Relational Data Model 11

Tables 11; Alternative Terminology 12; Table Properties 12; Keys 13; Primary Key 14; Representing Relationships 15; Relational Operators 17; Set Operators 17; The Union Operator 18; The Intersection Operator 19; The Difference Operator 20; The Cartesian Product Operator 21; Other Relational Operators 22; The Select Operator 23; The Project Operator 23; The Join Operator 25; Inner Joins 27; Outer Joins 28; The Divide Operator 29; The SQL Language 30; Integrity Constraints 30; Entity Integrity 30; Referential Integrity 31; Codd's Relational Rules 31; Conclusion 33; References 33

3 Logical and Physical Data Structures 35

ANSI/SPARC Architecture 36; DB2 Logical Data Model 37; DB2 User Views 37; DB2 Physical Data Structures 38; Logical versus Physical Representation 42; Data Views and SQL 42; Data Definition 42; Data Retrieval and Manipulation 43; Processing Support 44; Summary of Part I 45

PART



STRUCTURED QUERY LANGUAGE (SQL)

4 Simple Queries 49

Executing SQL Statements 49; The SELECT Statement 50; Selecting All Columns, All Rows 51; Displaying Selected Columns 52; Displaying Selected Rows 53; Selection Using Arithmetic Expressions 54; Selection Using a List or Range 55; Pattern Matching 56; Eliminating Duplicate Rows 57; Multiple Conditions 58; Null Values 59; Negative Conditions 60; Using the NOT Operator 61; Generated Columns 64; Sequencing the Rows 64; Sorting on a Generated Column 65

5 Complex Queries 67

Joining Tables 68; Joining More than Two Tables 69; Duplicate Column Names 70; Built-in Functions 71; Grouping Rows 72; Summary Data in Conditions 72; Multiple Columns in a GROUP BY Clause 73; Nested Queries 74; Built-in Functions in Subqueries 74; Subqueries That Return Multiple Values 75; Using NOT with ANY and ALL 77; Multiple Levels of Nesting 78; Correlated Subqueries 79; The EXISTS Keyword 80; Using NOT with EXISTS 81; Using EXISTS for All Values 81; Combining SELECT Statements 82; Performing an Outer Join 85

6 Advanced Functions 87

Creating Tables 87; Table Names 88; Column Definitions 89; NOT NULL Specifications 90; Database Name 90; Creating Tables with Referential Constraints 91; Data Manipulation 92; Inserting Data 92; Changing Data 94; Deleting Data 95; Altering and Dropping Tables 97; Deleting a Table 98; Creating and Dropping Synonyms 98; Creating and Dropping Views 99; Authorization 100; Summary of Part II 102

PART III **QUERY MANAGEMENT FACILITY (QMF)**

7 The QMF Environment 107

Query Management Facility 107; The QMF Interface 107; Choice of Languages 108; QMF Functions 108; Queries 108; Reports 109; Procedures 109; QMF Session 109; QMF Command Entry 109; QMF Items 110; QMF Commands 110; QMF Queries 110; SQL Queries 112; QBE Queries 114; Generating a Report 115; Printing and Formatting a Report 115; HELP Facility 117; Data Entry and Update 117

8 Report Formatting and QMF Procedures 119

Displaying or Printing a Report 119; Customizing a Report 121; Report-customizing Options 121; Sample Customized Report 122; Additional Formatting Options 124; Saving a Query 124; Variables 125; QMF Procedures 126; Procedure Variables 127

9 Query-by-Example (QBE) 129

Displaying Table Structures 129; Entering a QBE Query 130; Use of IN, BETWEEN, and LIKE 133; Unique Rows 134; Multiple Conditions 135; Example Elements 136; Condition Box 136; Expressions in Conditions 137; Negatives 137; Columns of Calculated Values 138; Specifying Row Sequence 139; Joining Tables 140; Built-in Functions and GROUP BY 141; Linked Queries 143; Condition Value in a Different Table 143; Condition Value in the Same Table 144; Combining Query Results 146; Updating Tables 147; Operations Not Supported 148; Summary of Part III 148

PART IV APPLICATION PROGRAMMING

10 Embedded SQL Statements 153

DB2 Programming Environment 153; Syntax for Embedded SQL Statements 153; Host Variables 154; Processing Query Results 156; DB2 Retrieval Program 156; Using a Cursor 158; The SQL Communication Area (SQLCA) 160; Processing SQL Return Codes 162; Declaring Tables and Views 165; The DCLGEN Facility 165

11 DB2 Program Development 167

Precompilation 167; Compilation and Link Editing 169; Binding 169; Program Execution 170; Authorization 170; Using DB2I 171; DB2I Program Preparation 172; DCLGEN 175; SPUFI 176; SPUFI Input 176; SPUFI Output 177

12 Loading and Updating Tables 179

Loading Data into Tables 179; Loading Tables with INSERT 180; Inserting Multiple Rows 180; Modifying Data 181; Modifying Multiple Rows 182; Updating Using a Cursor 182; Deleting Data Using a Cursor 185; Change Commitment and Concurrency 185; Specifying Commit Points 186; Unit of Recovery 187; Processing Related Transactions 189

13 Dynamic SQL 191

Non-SELECT Statements 192; Fixed-List SELECT Statements 192; Varying-List SELECT Statements 193; Designing Dynamic SQL Applications 194; Summary of Part IV 194

PART V**DB2 SYSTEM
ADMINISTRATION****14 DB2 Storage Structures 199**

DB2 Objects 199; Defining Storage Groups 201; Defining Databases 202; Defining Table Spaces 203; Defining Tables 204; Altering and Dropping Tables 205; Defining Indexes 205; Defining Views 207; The DB2 System Catalog 208

15 DB2 Utilities 211

Utilities for Loading and Recovering Data 211; The LOAD Utility 211; DB2 Recovery 213; Incremental Copies 213; Table Space Sets 214; Dual Logging 214; The COPY Utility 215; The MERGECOPY Utility 215; The QUIESCE Utility 215; The REPORT Utility 215; The RECOVER Utility 216; The MODIFY Utility 216; The CHECK Utility 216; The REPAIR Utility 216; Utilities for Monitoring and Tuning 216; The RUNSTATS Utility 217; The STOSPACE Utility 217; The REORG Utility 217; Stand-alone Utilities 217; The Change Log Inventory Utility 217; The Print Log Map Utility 217; Other Monitoring and Tuning Tools 218

16 The LOAD Utility 219

Loading an Empty Table 219; Adding or Replacing Data 219; LOAD Utility Options 220; Loading Selected Records 220; Specifying Input Fields 220; Loading Multiple Tables 221; Loading Partitioned Tables 221; Loading Tables with Unique Indexes 222; Using the CHECK Utility to Check Indexes 222; Loading Tables with Referential Constraints 222; Using the CHECK Utility to Check Referential Constraints 223

17 Data Extract (DXT) 225

Describing DXT Source Data 226; Building and Submitting a DXT Extract Request 226; Running a DXT Extract Request 227

18 DB2 Security 229

Table Privileges 230; Application Plan Privileges 231; Database Privileges 231; System Resource Privileges 232; Administrative Privileges 232; Granting Privileges 233; GRANT Options 236; Revoking Privileges 236; Authorization Information in the DB2 Catalog 236; Other Security Mechanisms 237

19 Locking Facilities 239

Unrepeatable Reads 239; Access to Uncommitted Data 239; Lost Updates 243; DB2 Resource Locks 243; Lock Size 243; Lock Duration 244; Lock Modes 244; Controlling Locks 245; Summary of Part V 246

PART VI IBM Application Development Products

20 Application System (AS) 251

AS Applications 252; Interactive Processing 252; Using AS 252; Operating Environment 252; Accessing Files and Tables 253; Entering SQL Statements 254; The AS Command Language 254; The VIEW Command 255; Arithmetic Expressions 256; Statement Continuation 257; Report Generation 257; Viewing Selected Rows 259; Logical Expressions 259; Canceling a Selection 260; AS Output Files and Tables 260; Sorting 261; Business Graphics 262; The PLOT Command 262; The HISTOGRAM Command 264; AS Languages 265; The PROCEDURE Language 270; References 270

21 Cross System Products (CSP) 271

CSP/AD 272; Maps 272; Records 273; Processes 273; Ease of Use 274; Online Prototyping and Testing 274; Using CSP 274; CSP/AD Menus 275; Terminal PF Keys 276; Application Definition Facility 276; Data Definition 277; Map Definition 279; Application Definition 281; Process Types 282; Statement Groups 283; Test Facility 283; Application Generation 284; List Processor, Utilities, and File Maintenance 284; Application Execution 285

PART VII APPENDIXES

Appendix I Normalization 289

First Normal Form 289; Second Normal Form 291; Third Normal Form 294; Fourth Normal Form 297; Fifth Normal Form 298

Appendix II Data Types and Expressions 303

Numeric Data 303; String Data 304; Data and Time Data 304

Appendix III Referential Integrity: Hierarchies and Cycles 309

Hierarchies 309; Cycles 310; Delete Considerations 311; Load Considerations 314

Index 317

PART



INTRODUCTION

Part I of this book examines key concepts underlying DB2, IBM's large-system relational database management system. Chapter 1 examines the characteristics generally associated with a *database management system* (DBMS) and the functions and facilities provided by the DB2 program product and related products such as *Query Management Facility* (QMF) and *Data Extract* (DXT). Chapter 2 then introduces the *relational model*, examines key concepts associated with it, and shows how DB2 implements relational DBMS facilities. Chapter 3 looks at the three views of data that are used with database management systems and examines how these views are implemented by DB2.

1 THE DB2 ENVIRONMENT

THE DB2 PRODUCT FAMILY

DATABASE 2, or DB2 as it is more commonly called, is one of several database products offered by IBM. DB2 is a relational database management system designed for the large-system (MVS) environment.* IBM offers another relational database management system, called SQL/DS, for midrange systems running DOS/VS or VM. The first database management system offered by IBM was Information Management System (IMS), which also runs in the MVS environment. A product called DL/I is the counterpart to IMS for midrange systems running the DOS/VS operating system.

In addition to DB2 itself, there are two other important program products that are widely used in the DB2 environment to extend the capabilities of DB2. These two products are closely associated with DB2, and we will be making numerous references to them throughout the book:

- **Query Management Facility (QMF).** QMF is a product that serves as an easy-to-use front end to DB2. It allows end users to retrieve data interactively, to format data into reports, and to display the data graphically.
- **Data Extract (DXT).** DXT is used to extract data from IMS, SQL/DS, or DB2 databases, from VSAM files, or from physical sequential files in order to make the data available to DB2.

*Many terms used in data processing literature begin as two or more words, such as *real-time* or *offline*. It is quite common in the computer industry for a widely accepted, often-used term to evolve into a new single word. The term *database* is now widely used and generally accepted as one word. In this book, we use the term *database* except when we are referring to an IBM product or manual name that uses the older two-word form. It is interesting to note that at the time of this writing, IBM uses the one-word form in naming the DB2 product—DATABASE 2—but continues to use the two-word form in the text of the DB2 reference manuals.

BOX 1.1 IBM products that can be used with DB2.

- **DB2 Performance Monitor.** This product is used to collect and format information about DB2 system and application activity.
- **Data Base Relational Application Directory.** This product provides directory support for application development in the DB2 environment.
- **Host Data Base View.** This product can be used to extract DB2 data and make it available to an IBM Personal Computer.
- **Data Base Migration Aid Utility.** This product helps migrate data and catalog information from one DB2 system to another.
- **Data Base Edit Facility.** This product provides a simple-to-use interface for DB2 data manipulation.
- **Cross System Product (CSP).** This product is a fourth-generation language (4GL), primarily intended for professional application developers, that can be used to generate applications that access DB2 databases.
- **Application Development Facility (ADF).** This is another 4GL, also primarily intended to be used by information systems professionals, to generate IMS and DB2 applications.
- **Application System (AS).** This product is a 4GL, primarily intended for end users, that can be used to access DB2 databases.

We will not discuss the many other IBM program products that can be used with DB2. Some of these are briefly described in Box 1.1. Also, many products from other vendors can also be used to access DB2 databases.

**THE DB2
OPERATING
ENVIRONMENT**

Describing the operating system and telecommunications environment in which the DB2 software runs requires generating a little IBM "alphabet soup." If you are not well acquainted with the software that is used in the IBM large-system environment, this section can safely be skipped over and the alphabet soup ignored. Having said that, DB2 runs under any of the currently supported versions of the MVS operating systems and can be used in conjunction with either IMS, CICS, or TSO. When run with TSO, DB2 data can be accessed from a TSO terminal running in foreground or by a batch job running in background. When using DB2 with TSO, DB2 provides the database services and TSO provides data communication services. When DB2 is run in the IMS environment, an application program can access data in both DB2 and IMS databases; IMS-DC provides data communication services in the IMS en-

vironment. When DB2 is run in the CICS environment, CICS provides the data communication services. CICS also provides an interface to IMS databases so that a single CICS application program can access both DB2 and IMS data. When an application program uses the data communication services of IMS or CICS, DB2 provides the facilities needed to be able to access both DB2 and IMS databases with full data integrity and recoverability; the user need not know the physical locations of the data in either environment.

IBM has stated that DB2 is an important part of its *Systems Application Architecture* (SAA). SAA is IBM's long-range plan for integrating its diverse computing system lines and providing common user and programming interfaces for all of its large, medium, and small systems.

DB2 provides services that are used to access data either interactively using a terminal or via an application program. DB2 application programs can be written in a variety of programming languages, including COBOL, PL/I, FORTRAN, BASIC, APL2, Assembler, C, and many fourth-generation languages.

THE SQL DATA LANGUAGE

The primary language used by both application programmers and end users in accessing a DB2 database is the *Structured Query Language* (SQL). SQL (often pronounced "sequel") is the standard language interface to DB2 and supports facilities that allow it to be used as both a *data definition language* (DDL) and a *data manipulation language* (DML). Many products other than DB2, such as IBM's Application System (AS), support interfaces to SQL and thus to DB2. SQL statements can be used to define, retrieve, and update data in DB2 databases. In addition to the basic data definition and data manipulation statements, there are also SQL statements that provide for processing related to security, integrity, and recovery and for the physical administration of DB2 databases. SQL statements can be executed interactively from a terminal, or they can be embedded in an application program.

Having a single data language that encompasses all functions needed to both support and use databases helps make DB2 easy to use while still providing a powerful set of facilities. All SQL statements use the same syntax. This allows a person to perform a full range of functions, from simple retrieval of data at a terminal through application system development to complete administration of a DB2 system, without having to learn several languages.

THE DB2 CATALOG

DB2 maintains information about the data that it manages in a set of tables known as the *DB2 catalog*. Since the DB2 catalog consists of DB2 databases, authorized users can access and manipulate data in the catalog using SQL statements in the same manner as they manipulate data in application databases. Data in the catalog can be used to determine what data items exist in the DB2 data-