

DATABASE DESIGN AND DEVELOPMENT

DATABASE DESIGN AND DEVELOPMENT

An Essential Guide for IT Professionals

PAULRAJ PONNIAH



IEEE PRESS



A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2003 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

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, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400, fax 978-750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, e-mail: permreq@wiley.com.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993 or fax 317-572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print, however, may not be available in electronic format.

Library of Congress Cataloging-in-Publication Data:

Ponniah, Paulraj.

Database design and development: an essential guide for IT professionals/Paulraj Ponniah.

p. cm.

“A Wiley-Interscience publication.”

Includes bibliographical references and index.

ISBN 0-471-21877-4 (cloth)

1. Database design. 2. Database management. I. Title.

QA76.9.D26P58 2003

005.74—dc10

2002192402

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

In loving memory
of
my dear parents.

CONTENTS

Preface	xxv
PART I BASIC DATABASE CONCEPTS	1
1 THE DATABASE APPROACH	3
Chapter Objectives / 3	
Evolution of Data Systems / 4	
Technology Explosion / 5	
Demand for Information / 8	
Waves of Evolution / 9	
File-Oriented Data Systems / 10	
Why Database Systems? / 11	
The Driving Forces / 11	
Inadequacy of Earlier Data Systems / 12	
Database Systems Meet the Challenges / 16	
The Database Approach / 19	
Database: A Formal Definition / 19	
Data-Driven, Not Process-Driven / 20	
Basic Concepts / 21	
How Databases Are Used / 23	
Overview of Data Models / 23	
Hierarchical / 24	
Network / 25	

Relational / 27
Object-Relational / 28
Types of Databases / 29
Centralized / 29
Distributed / 30
Survey of the Database Industry / 31
Brief History / 31
Leading Commercial Databases / 33
Peek into the Future / 33
Chapter Summary / 34
Review Questions / 34
Exercises / 35

2 OVERVIEW OF MAJOR COMPONENTS

36

Chapter Objectives / 36
What Makes Up a Database Environment? / 37
Overall Architecture / 38
Hardware Infrastructure / 40
Supporting Software / 40
People and Procedures / 41
Database and DBMS / 42
DB and DBMS – Not Synonymous / 42
Why Use a DBMS? / 42
DBMS Classifications / 45
Languages and Interfaces / 46
Functions and Features / 48
Hardware / 48
Storage of Data / 50
Operating System Software / 50
Database Software / 52
Users / 52
Practitioners / 53
Methods and Procedures / 54
How Databases Are Used / 55
Inside a DBMS / 56
Database Engine / 56
Data Dictionary / 58
Query Processor / 59
Forms Generator / 61
Report Writer / 62
Applications Developer / 63

Communications Interface / 64	
Utilities / 64	
Chapter Summary / 64	
Review Questions / 65	
Exercises / 66	

PART II DATABASE DEVELOPMENT PROCESS 67

3 SIGNIFICANCE OF THE DATABASE ENVIRONMENT 69

Chapter Objectives / 69	
Organizational Context / 70	
Core Business / 70	
Primary Processes / 71	
Information as a Major Asset / 72	
DB System in the Organization / 74	
Information Requirements / 75	
At User Levels / 76	
In Functional Divisions / 77	
At Geographic Regions / 78	
Providing Information / 79	
Information Sharing / 80	
Why Share Information? / 80	
Major Benefits / 82	
Information Sharing Schemes / 84	
DB System as Enabler / 86	
Pivotal Role of the Database System / 87	
Data Storage / 87	
Information Delivery / 88	
Foundation for Applications / 89	
Indispensable for Modern Enterprise / 90	
Chapter Summary / 91	
Review Questions / 91	
Exercises / 92	

4 DATABASE DEVELOPMENT LIFE CYCLE 93

Chapter Objectives / 93	
Major Development Steps / 94	
Starting the Process / 94	
Design and Development / 96	
Development and Implementation / 98	

Steps and Tasks /	98
Roles and Responsibilities /	99
Management and Control /	101
Planning for the DB System /	103
Scope of Overall Planning /	104
Who Does the Planning? /	105
Impact of the Business Plan /	105
The Database Plan /	107
Critical Planning Issues /	108
Feasibility Study /	109
Purpose and Extent /	109
Technological Infrastructure /	111
Skills Review /	112
Estimation of Costs /	113
Assessment of Benefits /	115
Weighing the Options /	117
Requirements Definition /	118
Requirements Gathering Methods /	119
Conducting User Interviews /	120
Observation of Processes /	124
Review of Applications /	125
Study of Documents /	126
Consolidating Information Requirements /	126
Requirements Definition Document /	127
The Design Phase /	128
Design Objectives /	129
Logical Versus Physical Design /	129
The External Schema /	130
The Conceptual Schema /	130
The Internal Schema /	131
Implementation and Deployment /	132
Conceptual to Internal Schema /	133
DBMS Installation /	134
Building the Data Dictionary /	134
Populating the Databases /	135
Developing Application Interfaces /	135
Maintenance and Growth /	135
Administration Tools /	136
Ongoing Monitoring /	137
Performance Tuning /	137
Managing Growth /	138
Chapter Summary /	139

Review Questions / 139

Exercises / 140

PART III CONCEPTUAL DATA MODELING 141

5 DATA MODELING BASICS 143

Chapter Objectives / 143

What is a Data Model? / 144

Why Create a Data Model? / 144

Real-World Information Requirements / 145

Data Model as the Replica / 146

Data Model Components / 147

Data Modeling Concepts / 148

Representation of Information Requirements / 149

Filtering Out Irrelevant Data / 150

Mapping of Components / 151

Data Model to Data Design / 151

Patterns of Data Modeling / 153

High-Level Data Model / 154

Object-Based Modeling Technique / 154

Entity-Relationship Modeling Technique / 155

Data Modeling Aids / 156

Data Views / 156

What are Data Views? / 157

Collection of Information Requirements / 157

Windows into the Data System / 158

Data Views: Two Perceptions / 159

View Integration / 160

Merging Individual User Views / 160

Integrating Partial Data Models / 162

Enhancement and Validation / 162

Consolidated Data Model / 163

Chapter Summary / 164

Review Questions / 165

Exercises / 165

6 OBJECT-BASED DATA MODEL: PRINCIPLES AND COMPONENTS 167

Chapter Objectives / 167

Overview of Object-Based Modeling / 168

A Generic Data Model / 168	
Benefits of Object-Based Data Model / 169	
Introduction to Components / 169	
Mapping of Real-World Information / 171	
Example of a Model Diagram / 174	
Business Objects / 177	
Object Sets and Instances / 177	
Types of Objects / 178	
Recognizing Object Sets / 178	
Attributes / 179	
Identifiers for Instances / 182	
Relationships between Objects / 183	
Role of Relationships / 183	
Cardinality in Relationships / 184	
Aggregate Objects / 189	
Degrees of Relationships / 192	
Generalization and Specialization / 192	
Supersets and Subsets / 194	
Generalization Hierarchy / 194	
Inheritance of Attributes / 196	
Inheritance of Relationships / 196	
Special Cases / 197	
Special Object Types and Relationships / 199	
Conceptual and Physical Objects / 200	
Recursive Relationships / 201	
Assembly Structures / 202	
Review of Object-Based Data Model / 203	
Summary of Components / 203	
Comprehensive Data Model Diagram / 203	
Chapter Summary / 203	
Review Questions / 204	
Exercises / 205	

7 ENTITY-RELATIONSHIP DATA MODEL

206

Chapter Objectives / 206	
Introduction to E-R Model / 207	
Basic Concepts / 207	
Data Modeling Process / 208	
Major Components / 209	
Entities / 210	
Entity Definition / 211	

Entity Types /	211
Entity Sets /	212
Weak Entity Types /	212
Attributes /	214
Attribute Specification /	214
Values and Domains /	215
Attribute Types /	215
Candidate and Primary Keys /	218
Relationships /	219
Association Between Entities /	219
Degree of a Relationship /	221
Cardinality in Relationships /	221
Optional and Mandatory Conditions /	221
Special Cases /	226
Modeling Time-Dependent Components /	226
Identifying and Nonidentifying Relationships /	226
Attributes of Relationship Types /	228
When to Use a Gerund /	228
Generalization and Specialization /	229
The Entity-Relationship Diagram (ERD) /	230
Review of Components and Notations /	230
Sample ERD /	231
Highlights of Sample ERD /	232
Chapter Summary /	233
Review Questions /	234
Exercises /	234

PART IV THE RELATIONAL DATA MODEL 237

8 RELATIONAL DATA MODEL FUNDAMENTALS 239

Chapter Objectives /	239
Structure and Components /	240
Strength of the Relational Model /	240
Relation: The Single Modeling Concept /	241
Columns as Attributes /	242
Rows as Instances /	242
Primary Key /	243
Relationship Through Foreign Keys /	244
Relational Model Notation /	245
Data Integrity Constraints /	246

Why Data Integrity? /	247
Basic Relational Properties /	247
Entity Integrity /	248
Referential Integrity /	249
Functional Dependencies /	250
Data Manipulation /	251
Role of Data Languages /	252
Data Manipulation Languages /	252
Relational Algebra /	253
Relational Calculus /	262
Comparison of Generic Languages /	264
Relational Model Design /	266
Requirements to Data Model /	267
Design Approaches /	267
Semantic to Relational Model /	267
Traditional Method /	268
Evaluation of the Two Methods /	269
Chapter Summary /	270
Review Questions /	271
Exercises /	271

9 SEMANTIC DATA MODEL TO RELATIONAL DATA MODEL

273

Chapter Objectives /	273
Model Transformation Approach /	274
Merits /	274
When to Use This Method /	276
Steps and Tasks /	277
Critical Issues /	277
Mapping of Components /	278
Mapping and Transformation /	279
Object Sets to Relations /	279
Attributes /	280
Instance Identifiers /	281
Transformation of Relationships /	282
One-to-One Relationships /	282
One-to-Many Relationships /	284
Many-to-Many Relationships /	286
Mandatory and Optional Conditions /	291
Aggregate Objects as Relationships /	296
Identifying Relationship /	296
Supersets and Subsets /	297

Logical Structure Versus Physical Structure /	334
Logical Design Phase in DDLC /	335
Why This Phase Is Necessary /	335
Input to Physical Design /	337
Ensuring Design Completeness /	337
Data Modeling in Logical Design /	339
Steps for Completing Logical Design /	339
Representing Entities /	340
Representing Attributes /	340
Representing Relationships /	340
Rules and Constraints /	342
Design for the Relational Data Model /	342
Relation as the Single Design Concept /	343
Logical Design Components /	344
Logical Schema /	344
Special Considerations /	345
Documentation of Logical Design /	346
Logical Design Outputs /	346
Usage of Outputs /	347
Use of CASE Tools /	347
Documentation Outline /	348
Chapter Summary /	349
Review Questions /	349
Exercises /	350

12 THE PHYSICAL DESIGN PROCESS

351

Chapter Objectives /	351
Introduction to Physical Design /	352
Logical to Physical Design /	352
Goals and Design Decisions /	354
Physical Design Components /	355
Physical Design Tasks /	356
Use of Data Dictionary /	356
Data Storage and Access /	358
Storage Management /	359
Access of Physical Data /	360
Files, Blocks, and Records /	360
File Organization /	363
Linking of Related Data Elements /	366
RAID Technology Basics /	367

Indexing Techniques /	371
Primary Indexes /	372
Binary Search /	373
B-Tree Index /	374
Secondary Indexes /	375
Bitmapped Indexing /	376
Other Performance Considerations /	377
Clustering /	378
Denormalization /	379
Fragmentation /	380
Memory Buffer Management /	380
Preconstructed Joins /	381
Chapter Summary /	382
Review Questions /	383
Exercises /	383

13 SPECIAL IMPLEMENTATION CONSIDERATIONS

385

Chapter Objectives /	385
Implementation Languages /	386
Meaning of Model Implementation /	386
Role of Languages /	387
Languages for the Relational Model /	389
SQL: The Relational Standard /	391
Overall Features /	392
Brief History and Evolution /	393
Data Definition in SQL /	394
Data Retrieval in SQL /	397
Data Maintenance in SQL /	399
Data Control in SQL /	401
Queries /	403
Summary of SQL Query Components /	406
Database Access from Application Program /	406
Query Processing /	407
Query Processing Steps /	409
The Query Optimizer /	410
Heuristic Approach /	411
Cost-Based Optimization /	412
Database System Deployment /	415
Deployment Tasks /	416
Implementation in Centralized Architecture /	417
Implementation in Client/Server Architecture /	418

Chapter Summary /	420
Review Questions /	420
Exercises /	421

PART VI DATABASE ADMINISTRATION AND MAINTENANCE **423**

14 OVERVIEW OF ADMINISTRATION FUNCTIONS **425**

Chapter Objectives /	425
Significance of Administration /	426
Essential Need for Administration /	427
Administration Objectives /	430
Data Administration /	432
Database Administration /	432
Skills and Responsibilities /	433
Administrative Roles /	433
Areas of Responsibility /	434
Skills, Knowledge, and Experience /	436
Interaction with Users and Practitioners /	437
During Development /	437
Planning and Feasibility Study /	438
Requirements Definition /	439
Logical Design /	439
Physical Design /	440
Implementation and Deployment /	440
After Deployment /	441
Ongoing Functions /	441
Maintenance of Service Levels /	442
Enhancements to Database System /	442
Growth and Change /	442
Chapter Summary /	443
Review Questions /	444
Exercises /	444

15 DATA INTEGRITY **446**

Chapter Objectives /	446
Transaction Processing /	447
Transaction Concepts /	447
Properties of Transactions /	449
Transaction States /	451

Processing of Transactions /	452
Integrity Considerations /	454
Concurrent Transactions /	456
Motivation for Concurrency /	458
Concurrency Problems /	458
Transactions and Schedules /	460
Serializability /	462
Recoverability /	465
Concurrency Control /	466
Lock-Based Resolution /	467
Application of Lock-Based Techniques /	470
Deadlock: Prevention and Detection /	473
Timestamp-Based Resolution /	476
Optimistic Techniques /	479
Database Failures and Recovery /	481
Classification of Failures /	481
Recovery Concepts /	482
Logging /	483
Checkpoint /	485
Log-Based Recovery Techniques /	486
Shadow Paging /	489
A Recovery Example /	491
Chapter Summary /	491
Review Questions /	493
Exercises /	493

16 DATABASE SECURITY

495

Chapter Objectives /	495
Security Issues /	496
Goals and Objectives /	496
Security Problems /	497
Solution Options /	499
Privacy Issues /	500
Web Security /	501
Access Control /	501
Levels and Types of Data Access /	502
Discretionary Control /	503
Use of Views /	506
SQL Examples /	507
Mandatory Control /	508
Special Security Considerations /	510

Authorization /	510
Authentication /	512
Role of the DBA /	513
Statistical Databases /	513
Encryption /	516
What is Encryption? /	516
Encryption Methods /	518
Data Encryption Standard /	519
Public Key Encryption /	520
Chapter Summary /	523
Review Questions /	523
Exercises /	524

17 ONGOING MAINTENANCE AND GROWTH

525

Chapter Objectives /	525
Routine Maintenance /	527
Backup and Recovery /	527
Security Maintenance /	529
Space Management /	531
Concurrency Control /	532
Problem Resolution /	532
Monitoring and Review /	533
Purpose of Monitoring /	534
The Monitoring Process /	534
Gathering of Statistics /	536
Review of Operations /	536
Use of Benchmarks /	537
Growth and Enhancements /	538
Ongoing Growth and Enhancements /	538
Application Enhancements /	539
Schema Revisions /	541
DBMS Version Upgrades /	542
Tuning for Performance /	543
Goals and Solution Options /	544
Tuning Indexes /	545
Tuning Queries and Transactions /	546
Tuning the Schema /	547
Chapter Summary /	547
Review Questions /	548
Exercises /	549

PART VII ADVANCED DATABASE TOPICS **551****18 DISTRIBUTED DATABASE SYSTEMS** **553**

Chapter Objectives /	553
Fundamental Principles /	554
What is a Distributed Database? /	555
Basic Concepts /	556
Motivation and Goals /	557
Advantages and Disadvantages /	557
Distributed Databases /	559
Types and Configurations /	559
DDBMS /	561
Network Component /	566
Data Distribution /	568
Architectural Options /	572
Design and Implementation Issues /	574
Transparencies /	576
Transparency: Key Ideal /	576
Fragmentation Transparency /	577
Replication Transparency /	578
Location Transparency /	578
Network Transparency /	578
Naming Transparency /	579
Failure Transparency /	579
Distributed Processing /	579
Query Processing /	580
Transaction Processing /	585
Concurrency Control /	588
Distributed Recovery /	592
Chapter Summary /	595
Review Questions /	596
Exercises /	596

19 DATABASE SYSTEMS AND THE WEB **598**

Chapter Objectives /	598
Web Technology: A Refresher /	599
The Internet and the Web /	599
HyperText Transfer Protocol (HTTP) /	601
Uniform Resource Locator (URL) /	602

HyperText Markup Language (HTML) /	603
Beyond HTML /	605
Intranets and Extranets /	606
Web-Database Integration /	607
Motivation for Integration /	608
Requisites for Integration /	609
Architecture for Integration /	609
Advantages and Disadvantages /	611
Integration Approaches /	613
Common Gateway Interface (CGI) /	613
Application Programming Interface (API) /	616
Server-Side Includes (SSI) /	618
Cookies /	619
Use of Java Applications /	620
Use of Scripting Languages /	622
Database Tools /	623
Security Options /	625
Significance of Protection /	625
Firewalls, Wrappers, and Proxies /	626
Digital Signatures and Digital Certificates /	628
SET and SST /	629
SSL and S-HTTP /	630
Java Security /	630
Chapter Summary /	633
Review Questions /	634
Exercises /	634

20 TRENDS IN DATABASE TECHNOLOGY

636

Chapter Objectives /	636
Object-Oriented Databases /	637
Basic Concepts /	638
Objects and Classes /	640
Methods and Messages /	641
Inheritance /	643
Polymorphism /	643
Object-Oriented Data Model /	644
Object-Relational Databases /	646
The Driving Forces /	647
What is an ORDBMS? /	648
Feature Highlights /	648
SQL-3: Object-Relational Support /	649

Databases for Decision Support /	649
Data Warehousing /	650
Online Analytical Processing (OLAP) /	661
Data Mining /	667
Leading Trends: Basic Overview /	671
Parallel Databases /	672
Active Databases /	674
Intelligent Databases /	675
Deductive Databases /	676
Multimedia Databases /	676
Mobile Databases /	678
Geographic Databases /	681
Chapter Summary /	683
Review Questions /	683
Exercises /	684

APPENDICES	685
A Legacy System Models: Hierarchical and Network	687
B Codd's Relational Rules	692
C Diagramming Conventions and Symbols	694
D Use of CASE tools	700
E Review of major commercial DBMSs	703
F Database Design and Development Summary	712
 References	 716
Glossary	718
Index	729

PREFACE

Are you a programmer, systems analyst, network specialist, project leader, or any other type of information technology professional? Alternatively, are you a student aspiring for a career in information technology? Then you definitely need to know how database systems are designed and developed. You have to understand the fundamentals of database systems clearly.

As you know, in today's business environment, companies depend on their databases to provide crucial information essential for running their businesses. Gone are the days of file-oriented data systems. Now database systems form the centerpiece of the growing and maturing electronic commerce. Database and Web technologies have merged. Over the years, commercial database products have become sophisticated and robust.

In this transformed computing environment, knowledge of database systems can no longer be confined only to specialists such as data analysts and database administrators. All IT professionals need basic knowledge of database technology and its applications. This book comes to you as an essential guide on database design and development, covering all the necessary topics in proper measure, written especially for IT professionals—present and future.

THE SCENARIO

In every industry across the board, from retail chain stores to financial institutions, from manufacturing organizations to government departments, and from airline companies to utility businesses, database systems have become the norm for information storage and retrieval. Whether it is a Web-based application driving electronic commerce or an inventory control application managing just-in-time inventory or a data warehouse system supporting strategic decision making, you need an effective and successful technology to store and retrieve data. It is no

wonder that companies have adopted database technology without reservations. The modern relational database system, proven to be eminently suitable for data management, has become more and more pervasive.

Over the recent years, vendors of all leading database products have released more sophisticated and powerful software versions. Database management systems such as DB2, Informix, Oracle, SQL Server, and Sybase have all expanded with several useful features. Database management systems have become the centerpiece of e-business. Numerous books feature commercial database products.

THE ROLE OF IT

In this scenario, the information technology department of every organization has a primary responsibility. The department has to support and keep the database systems running. Without the database system, the day-to-day business of the organization will come to a grinding halt.

Therefore, all IT staff must understand the workings of database systems. It is not enough to have just a handful of specialists knowledgeable in database technology. All applications in the enterprise now work with databases. Every IT professional, therefore, must know the basics of the technology. Everyone must learn how database systems are designed and developed. Every IT professional must understand the fundamental principles.

WHAT THIS BOOK CAN DO FOR YOU

This book provides you with necessary information on the basics of database technology. It covers all the essential topics carefully with proper emphasis as required by each topic. If you are new to the fundamentals of database technology, this book is an essential pre-requisite before you determine the next steps toward specialization in the database field. If you are already familiar with the technology, this book is a suitable refresher to reinforce your grasp of the subject.

More specifically, here is a summary of what this book can do for you:

Specially designed for IT professionals

Specifically intended for IT professionals like you, this book builds on what you already know. The book takes into account the background, knowledge, and terminology of IT professionals; it presents the topics in a suitable direct style.

Comprehensive with just the necessary details

The book deals with every significant topic needed by IT professionals looking for the fundamentals. It encompasses database concepts, terminology, planning, implementation, and administration; the book also includes significant technology trends.

Suitably organized

The book follows an organization most apt and logical for IT professionals concentrating on the fundamentals. It is the type of organization these professionals is most familiar in their day-to-day work experience. Beginning with an overview of basic concepts, the book moves on to an overview of the database system development process, then to the important topic of data modeling, on to design, then to implementation, and concludes with ongoing maintenance and growth.

Feature highlights

Every chapter opens up with chapter objectives and concludes with a chapter summary. At the end of each chapter, you find a set of review questions and exercises. These features make the book eminently suitable for self-study or for use as a textbook in a college course.

Exposure to real-world situations

Throughout the book, each concept or technique is illustrated with real-world examples. An appendix is devoted to the review of leading commercial database management systems.

Preparation for database specialists

Although intended as a first course on the fundamentals, the book provides sufficient coverage of each topic so that you may easily proceed to the next phase of specialization for specific roles such as data modeler, database designer, data analyst, data administrator, or database administrator.

ACKNOWLEDGMENTS

The authors listed in the reference section at the end of the book greatly enhanced my understanding and appreciation for database technology. I am deeply indebted to the authors for their insights and observations; I wish to express my sincere thanks to them.

I must also record my appreciation for the several professional colleagues who had worked with me on various database projects during my 25-year consulting career. Also, thanks are due to the many students in my database classes over the years. Interactions with my students and colleagues have enabled me to shape this book according to the specific needs of IT professionals.

PAULRAJ PONNIAH, Ph.D.

*Milltown, New Jersey
November 2002*