



relational database design | *clearly explained*

SECOND
EDITION

JAN L. HARRINGTON

***Relational
Database Design
Clearly
Explained
Second Edition***

This Page Intentionally Left Blank

Relational Database Design Clearly Explained

Second Edition

Jan L. Harrington



Morgan Kaufmann Publishers
An Imprint of Elsevier

Amsterdam London New York Oxford Paris Tokyo
Boston San Diego San Francisco Singapore Sydney

This book is printed on acid-free paper. (∞)

Copyright © 2002, 1999, Elsevier.

All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Permissions may be sought directly from Elsevier's Science and Technology Rights Department in Oxford, UK. Phone: (44) 1865 843830, Fax: (44) 1865 853333, e-mail: permissions@elsevier.co.uk. You may also complete your request on-line via the Elsevier homepage: <http://www.elsevier.com> by selecting "Customer Support" and then "Obtaining Permissions".

All brand names and product names mentioned in this book are trademarks or registered trademarks of their respective companies.

Academic Press

An Imprint of Elsevier

525 B Street, Suite 1900, San Diego, California 92101-4495, USA

<http://www.academicpress.com>

Academic Press

32 Jamestown Road, London NW1 7BY, UK

<http://www.academicpress.com>

Morgan Kaufmann

An Imprint of Elsevier

340 Pine Street, Sixth Floor, San Francisco, California 94104-3205, USA

<http://www.mkp.com>

Library of Congress Catalog Card Number: 2002101250

ISBN-10: 1-55860-820-6

ISBN-13: 978-1-55860-820-7

Printed in the United States of America

05 06 07 MV 9 8 7 6 5 4

Contents

Preface to the Second Edition xiii

What You Need to Know xv

Acknowledgments xv

Part One: Theory

Chapter 1: Introduction 3

Effects of Poor Database Design 4

 Unnecessary Duplicated Data and Data Consistency 6

 Data Insertion Problems 7

 Data Deletion Problems 8

 Meaningful Identifiers 9

What You Will Find in This Book 9

Chapter 2: Entities and Data Relationships 11

- Entities and Their Attributes 12
 - Entity Identifiers 13
 - Single-Valued versus Multivalued Attributes 15
 - Avoiding Collections of Entities 17
 - Documenting Logical Data Relationships 18
 - Entities and Attributes for Lasers Only 19
- Domains 20
 - Documenting Domains 21
 - Practical Domain Choices 21
- Basic Data Relationships 23
 - One-to-One Relationships 24
 - One-to-Many Relationships 26
 - Many-to-Many Relationships 27
 - Weak Entities and Mandatory Relationships 27
 - Documenting Relationships 28
 - Basic Relationships for Lasers Only 32
- Dealing with Many-to-Many Relationships 33
 - Composite Entities 34
 - Documenting Composite Entities 36
 - Resolving Lasers Only's Many-to-Many Relationships 36
- Relationships and Business Rules 37
- Data Modeling versus Data Flow 39
- Schemas 43
- For Further Reading 44

Chapter 3: Historical Antecedents 47

- File Processing Systems 48
 - Early File Processing 48
 - ISAM Files 50
 - Limitations of File Processing 52
 - File Processing on the Desktop 53
- The Hierarchical Data Model 54
 - Characteristics of the Hierarchical Data Model 54
 - IMS 57
- The Simple Network Data Model 59
 - Characteristics of a Simple Network 59
 - CODASYL 62
- The Complex Network Data Model 65

Chapter 4: The Relational Data Model 73

- Understanding Relations 74
 - Columns and Column Characteristics 75
 - Rows and Row Characteristics 76
 - Types of Tables 76
 - A Notation for Relations 77
- Primary Keys 77
 - Primary Keys to Identify People 78
 - Avoiding Meaningful Primary Keys 80
 - Concatenated Primary Keys 81
 - All-Key Relations 82
- Representing Data Relationships 82
 - Referential Integrity 85
 - Foreign Keys and Primary Keys in the Same Table 86
- Views 86
 - The View Mechanism 86
 - Why Use Views? 87
- The Data Dictionary 88
 - Sample Data Dictionary Tables 89
- A Bit of History 90
- For Further Reading 92

Chapter 5: Normalization 93

- Translating an ER Diagram into Relations 94
- Normal Forms 95
- First Normal Form 96
 - Understanding Repeating Groups 97
 - Handling Repeating Groups 98
 - Problems with First Normal Form 100
- Second Normal Form 102
 - Understanding Functional Dependencies 103
 - Using Functional Dependencies to Reach 2NF 104
 - Problems with 2NF Relations 105
- Third Normal Form 106
 - Transitive Dependencies 106
- Boyce–Codd Normal Form 108
- Fourth Normal Form 110
 - Multivalued Dependencies 111

Normalized Relations and Database Performance	112
Equi-Joins	113
What's Really Going On: PRODUCT and RESTRICT	116
The Bottom Line	119
For Further Reading	119

Chapter 6: Database Structure and Performance Tuning 121

Indexing	122
Deciding Which Indexes to Create	124
Clustering	124
Partitioning	126
Horizontal Partitioning	126
Vertical Partitioning	127
For Further Reference	128

Chapter 7: Codd's Rules 129

Rule 1: The Information Rule	130
Rule 2: The Guaranteed Access Rule	132
Rule 3: Systematic Treatment of Null Values	133
Rule 4: Dynamic Online Catalog Based on the Relational Model	134
Rule 5: The Comprehensive Data Sublanguage Rule	135
Rule 6: The View Updating Rule	136
Rule 7: High-Level Insert, Update, and Delete	137
Rule 8: Physical Data Independence	138
Rule 9: Logical Data Independence	139
Rule 10: Integrity Independence	139
Rule 11: Distribution Independence	141
Rule 12: Nonsubversion Rule	142

Chapter 8: Integrating Objects 143

An Introduction to Object-Oriented Concepts	144
Writing Instructions	145
Objects	146
Classes	151
Class Relationships	157
Benefits of Object Orientation	165

Integrating Objects into a Relational Database	166
ER Diagrams for Object-Relational Designs	167
For Further Reading	172

Part Two: Practice

Chapter 9: Using SQL to Implement a Relational Design 177

Database Object Hierarchy	178
Naming and Identifying Objects	179
Schemas	181
Creating a Schema	181
Identifying the Schema You Want to Use	182
Domains	183
Tables	184
Column Data Types	185
Default Values	190
NOT NULL Constraints	190
Primary Keys	191
Foreign Keys	191
Additional Column Constraints	194
Views	198
Deciding Which Views to Create	198
View Updatability Issues	199
Creating Views	199
Temporary Tables	201
Creating Temporary Tables	202
Loading Temporary Tables with Data	203
Disposition of Temporary Table Rows	203
Creating Indexes	204
Modifying Database Elements	205
Adding New Columns	205
Adding Table Constraints	206
Modifying Columns	206
Deleting Elements	207
Renaming Elements	208
Deleting Database Elements	209

Granting and Revoking Access Rights	209
Types of Access Rights	210
Storing Access Rights	210
Granting Rights	211
Revoking Rights	212
Object-Relational Extensions	213

Chapter 10: Using CASE Tools for Database Design 215

CASE Capabilities	216
ER Diagram Reports	218
Data Flow Diagrams	220
The Data Dictionary	222
Code Generation	225
Sample Input and Output Designs	228
The Drawing Environment	229
For Further Reading	230

Chapter 11: Database Design Case Study #1: Mighty-Mite Motors 231

Corporate Overview	232
Product Development Division	232
Manufacturing Division	238
Marketing & Sales Division	239
Current Information Systems	239
Reengineering Project	240
New Information Systems Division	241
Basic System Goals	241
Current Business Processes	242
Designing the Database	251
Examining the Data Flows	253
Creating the ER Diagram	256
Creating the Tables	261
Generating the SQL	262
The Object-Relational Design	268
Creating the Classes	273
Using the Classes in the Schema	275

Chapter 12: Database Design Case Study #2: East Coast Aquarium 281

- Organizational Overview 282
 - Animal Tracking Needs 284
 - The Volunteer Organization 288
- The Volunteers Database 290
 - Creating the Application Prototype 290
 - Creating the ER Diagram 300
 - Designing the Tables 300
 - Generating the SQL 301
- The Animal Tracking Database 301
 - Highlights of the Application Prototype 303
 - Creating the ER Diagram 308
 - Creating the Tables 311
 - Generating the SQL 312
- The Object-Relational Design 316
 - The Volunteers Database 316
 - The Animal Tracking Database 317

Chapter 13: Database Design Case Study #3: Independent Intelligence Agency 321

- Organizational Overview 322
 - Current Information Systems 324
 - Summary of IS Needs 327
 - System Specifications 328
- Designing the Database 333
 - ER Diagram for People 334
 - ER Diagram for Intelligence Gathering and Sales 337
 - ER Diagram for System Security 340
 - ER Diagram for Equipment Tracking 341
 - Designing the Tables 343
 - Generating the SQL 344
- The Object-Relational Design 355
 - Creating the Classes 362
 - Writing the Schema 365

Glossary 375

Index 387

This Page Intentionally Left Blank

Preface to the Second Edition

My favorite opening line for the database courses I teach is: “Probably the most misunderstood term in all of business computing is *database*, followed closely by the word *relational*.” At that point, the students often snicker, because they are absolutely, positively sure that they know what a database is and that they also know what it means for a database to be relational. Unfortunately, the popular press, with the help of some software developers, long ago distorted the meaning of both those terms, which led many businesses to think that designing a database is a task that could be left to a clerical worker who had taken a few days training in using database software. As you will see throughout this book, nothing could be further from the truth.

Note: By the same token, I have received applications for adjunct faculty positions from people who say they know how to use a specific piece of database software and therefore believe

they are qualified to teach a database theory course. From where I sit, that is pretty scary.

Before preparing the first edition of this book, I had wanted to write a book like this for a long time. We teach relational database design theory to college students, but it is a rare pleasure to get the chance to share that theory—and the practice of that theory—with the business community, with people who are actually doing such designs in their jobs. It's just as great a pleasure to be able to prepare a second edition, giving me a chance to correct some nasty typos and add material that reviewers have indicated was lacking.

This book is intended for anyone who has been given the responsibility of designing or maintaining a relational database. The first part will teach you how to look at the environment your database serves and to tailor the design of that database to the environment. It will also teach you ways of designing the database so that it provides accurate and consistent data, avoiding the problems that are common to poorly designed databases. In addition, you will read about design compromises that you might choose to make in the interest of database application performance and the consequences of making such choices.

The second edition introduces a chapter on the history of data management. This chapter will show you just how far the IT industry has come in the past 40 years or so in terms of handling business data. It may also help you appreciate why so many people prefer relational databases to any other alternative. The placement of this new chapter is somewhat problematic. Ideally, it should be Chapter 0, but it requires concepts that are taught in Chapter 2. Therefore, although it is slightly out of place, it appears as Chapter 3.

Part II looks at the more practical aspects of performing database design, covering the implementation of a design using SQL and using a CASE tool to help document and support the design process. In addition, Part II contains three large database design case studies, each of which presents one or more design challenges that you may encounter when designing relational databases.

Probably the biggest trend in relational database design since the first edition of this book appeared is the integration of objects into relational databases. The major relational DBMSs on the market have embraced this hybrid object-relational approach. You will therefore find an entire chapter on the object-oriented paradigm and how it has been merged with a relational database. Each of the case studies in Part II also concludes with an example of an object-relational solution.

What You Need to Know

Because this book deals primarily with database design, you do not need any special computing background to read it. You should, however, have some basic computer literacy. If you know how to get around your computer's operating system and how to run programs someone has written for you, then you know enough to understand the material in Part I and most of Part II of this book. To get the most out of Chapter 9 (using SQL to implement a relational design), you should be familiar with some type of database environment that provides an interface for sending SQL commands to a database.

Acknowledgments

Writing a book for the folks at Morgan Kaufman is always a joy. I would therefore like to thank the following individuals who helped make this one possible:

- ◆ Diane Cerra, editor at Morgan Kaufman
- ◆ Mona Buelher, editorial assistant at Morgan Kaufman
- ◆ Debbie Liehs, project manager
- ◆ Edward Wade, assistant publishing services manager
- ◆ Mei Levenson, production coordinator
- ◆ Adrienne Rubello, copy editor
- ◆ Tara Masih, proofreader

- ◆ The three reviewers:
 - Sheldon Barry, Memorial University of Newfoundland
 - Karen Watterson, industry consultant
 - Russell Belfer, independent software developer/consultant

And above all, to my very active two-year-old son, Sean, who slept long enough for me to finish this book.

JLH



This Page Intentionally Left Blank

Part One

Theory

The first part of this book considers the theoretical aspects of relational database design. You will read about identifying data relationships in your database environment, the details of the relational data model, and how to translate data relationships into a well-designed relational database that avoids most of the problems associated with bad designs.

This Page Intentionally Left Blank

1

Introduction

Many of today's businesses rely on their database systems for accurate, up-to-date information. Without those repositories of mission-critical data, most businesses are unable to perform their normal daily transactions, much less create summary reports that help management make strategic corporate decisions. To be useful, the data in a database must be accurate, complete, and organized in such a way that data can be retrieved when needed and in the format required.

Well-written database application programs—whether they execute locally, run over a local area network, or feed information to a Web site—are fundamental to timely and accurate data retrieval. However, without a good underlying database design, even the best program cannot avoid problems with inaccurate and inconsistent data. That is what this book is all about: to help you learn to