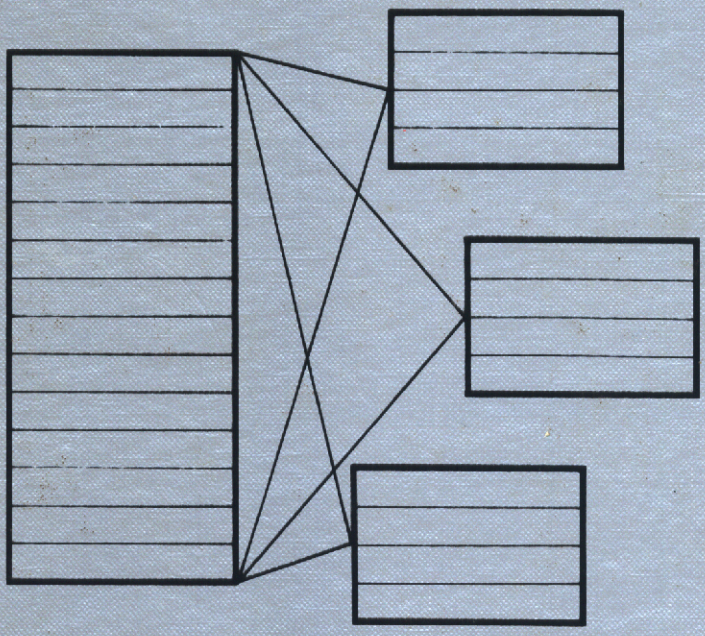
TEXTS AND MONOGRAPHS IN COMPUTER SCIENCE

RELATIONAL DATABASE TECHNOLOGY

Suad Alagić



BIBLIOTHEEK

Springer-Verlag

BIBLIOTHEQUE DU CERIST

Texts and Monographs in Computer Science

Editor
David Gries

Advisory Board
F. L. Bauer
J. J. Horning
R. Reddy
D. C. Tsichritzis
W. M. Waite

Texts and Monographs in Computer Science

Suad Alagić
Relational Database Technology

Suad Alagić and Michael A. Arbib
The Design of Well-Structured and Correct Programs

S. Thomas Alexander
Adaptive Signal Processing: Theory and Applications

Michael A. Arbib, A. J. Kfoury, and Robert N. Moll
A Basis for Theoretical Computer Science

Michael A. Arbib and Ernest G. Manes
Algebraic Approaches to Program Semantics

F. L. Bauer and H. Wössner
Algorithmic Language and Program Development

Kaare Christian
The Guide to Modula-2

Edsger W. Dijkstra
Selected Writings on Computing: A Personal Perspective

Nissim Francez
Fairness

Peter W. Frey, Ed.
Chess Skill in Man and Machine, 2nd Edition

R. T. Gregory and E. V. Krishnamurthy
Methods and Applications of Error-Free Computation

David Gries, Ed.
Programming Methodology: A Collection of Articles by Members of IFIP WG2.3

David Gries
The Science of Programming

A. J. Kfoury, Robert N. Moll, and Michael A. Arbib
A Programming Approach to Computability

E. V. Krishnamurthy
Error-Free Polynomial Matrix Computations

Franco P. Preparata and Michael Ian Shamos
Computational Geometry: An Introduction

Brian Randell, Ed.
The Origins of Digital Computers: Selected Papers

Arto Salomaa and Matti Soittola
Automata-Theoretic Aspects of Formal Power Series

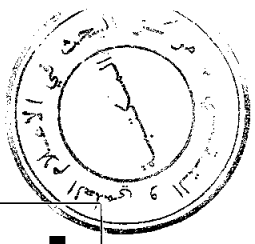
Jeffrey R. Sampson
Adaptive Information Processing: An Introductory Survey

William M. Waite and Gerhard Goos
Compiler Construction

Niklaus Wirth
Programming in Modula-2

BIBLIOTHEQUE DU CERIST

C
1877



Relational Database Technology

Suad Alagić

With 114 Illustrations

Springer-Verlag
New York Berlin Heidelberg Tokyo

BIBLIOTHEQUE DU CERIST

Suad Alagić
Department of Informatics
Faculty of Electrical Engineering
University of Sarajevo
71000 Sarajevo, Lukavica
Yugoslavia

Series Editor

David Gries
Department of Computer Science
Cornell University
Upson Hall
Ithaca, NY 14853
U.S.A.

BIBLIOTHEQUE DU CERIST

Library of Congress Cataloging in Publication Data
Alagić, Suad, 1946–
Relational database technology.
(Texts and monographs in computer science)
Bibliography: p.
Includes index.
1. Data base management. I. Title. II. Series.
QA76.9.D3A48 1986 005.75'6 85-32108

© 1986 by Springer-Verlag New York Inc.
All rights reserved. No part of this book may be translated or reproduced in any form without written permission from Springer-Verlag, 175 Fifth Avenue, New York, New York 10010, U.S.A.

4598

Typeset by Asco Trade Typesetting Ltd., Hong Kong.
Printed and bound by R.R. Donnelley & Sons, Harrisonburg, Virginia.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-96276-X Springer-Verlag New York Berlin Heidelberg Tokyo
ISBN 3-540-96276-X Springer-Verlag Berlin Heidelberg New York Tokyo

To Mara, Irena, and Gorjan

BIBLIOTHEQUE DU CERIST

BIBLIOTHEQUE DU CERIST

Preface

BIBLIOTHEQUE DU CERIST

This book presents a unified collection of concepts, tools, and techniques that constitute the most important technology available today for the design and implementation of information systems. The framework adopted for this integration goal is the one offered by the relational model of data, its applications, and implementations in multiuser and distributed environments. The topics presented in the book include conceptual modeling of application environments using the relational model, formal properties of that model, and tools such as relational languages which go with it, techniques for the logical and physical design of relational database systems and their implementations. The book attempts to develop an integrated methodology for addressing all these issues on the basis of the relational approach and various research and practical developments related to that approach.

This book is the only one available today that presents such an integration. The diversity of approaches to data models, to logical and physical database design, to database application programming, and to use and implementation of database systems calls for a common framework for all of them. It has become difficult to study modern database technology without such a unified approach to a diversity of results developed during the vigorous growth of the database area in recent years, let alone to teach a course on the subject.

It is quite clear that applications of the relational database technology are much more important to the majority of readers interested in that technology. However, knowledge about the underlying structures and algorithms, concurrency control protocols, and distributed processing techniques are becoming a common professional knowledge and are of interest not only to the implementors of such systems but also to all those readers—database

application programmers in particular—who would like to know how relational database systems actually work. On the other hand, implementors of such systems will benefit from the exposition of conceptual and relational modeling and end-user requirements and tools. Indeed, the most important implementation techniques for relational languages, for multiuser and distributed systems are presented within a framework that is an extension of the formal approach offered by various results developed around the relational model of data.

Although the book is based mainly on a collection of research papers, it is written as a textbook. This refers to the style of exposition, numerous examples which are used in the book, and an extensive set of exercises. Very little is assumed in terms of the required background. Some familiarity with the very basic concepts of modern mathematics, computer systems, and their common applications is, of course, necessary. The presentation of the basic theory of the relational model of data, its underlying relational structures and algorithms, transaction processing models, and distributed systems is sometimes formal and very technical, but it is always illustrated by very pragmatic examples.

In addition to the described integration goal the book is different from existing ones in that it applies recent object–action oriented techniques for conceptual modeling within the relational framework, and it provides a thorough exposition of the most important issues—particularly those related to data integrity—involved in the design and implementation of multiuser and distributed database systems.

The author acknowledges research support from SIZ nauke SR Bosne i Hercegovine, and valuable observations made by Mario Gaon.

Tsukuba Science City, Japan
March 1985

SUAD ALAGIĆ

BIBLIOTHEQUE DU CERIST

Contents

BIBLIOTHEQUE DU CERIST

Introduction 1

1 Modeling Application Environments 1
2 Relational Model 5
3 Relational Database System 11
4 Relational Technology 17

CHAPTER 1
Data Model 20

1 Relational Model of Data 20
1.1 Relational Representation of Entities 20
1.2 Relational Algebra 24
1.3 Relational Query Languages 29
2 Logical Dependencies 36
2.1 Functional Dependencies 36
2.2 Multivalued Dependencies 41
2.3 Join Dependencies 47
3 Hierarchical and Network Models 48
3.1 Unnormalized Relational Schemas 48
3.2 Network Model 51
Exercises: GROUP BY Clause; SET Function; Entity Relationship Diagrams; Generalized Joins; $M : N$ Relationships; Lossless Joins 56
Bibliographical Notes 63

CHAPTER 2
Logical Design 64

1 Normal Forms 64
1.1 Second Normal Form (2NF) 65

1.2 Third Normal Form (3NF) 68
1.3 Boyce–Codd Normal Form (BCNF) 72
1.4 Fourth Normal Form (4NF) 75
1.5 Projection/Join Normal Form (PJNF) 79
2 Abstractions 82
2.1 Unnormalized Relational Model 82
2.2 Aggregation 84
2.3 Generalization 88
3 Design Methodology 92
3.1 Extended Relational Model 92
3.2 Relational Database Programming Environment 95
3.3 Conceptual Modeling 107
Exercises: Views; Types of Database Entities; Generalization; Associations; Aggregation; Characterization; Cover Aggregation; Taxonomic Design Methodology; Exception Modeling; Exception Handling 118
Bibliographical Notes 130

CHAPTER 3
Structural Design 131

1 Relational Images 131
2 Decomposition of Unary Queries 135
3 Decomposition of Binary Queries 141
4 Optimization of Binary Queries 146
5 Decomposition of Queries with Set Operators 149
6 Relational Representation of Relations and Their Images 152
7 Decomposition of Data Manipulation Statements 153
8 Structure of Images 156
Exercises: Links; Network Structures; Decomposition of *n*-ary Queries; Optimization of Query Expressions; Properties of the Relational Operators; B*-Trees 162
Bibliographical Notes 172

CHAPTER 4
Data Integrity 173

1 Transactions and Integrity of Data 173
2 Concurrent Executions of Transactions 175
3 Locking Protocols 179
4 Logical Locks 188
5 Restoring a Consistent Database State 197
Exercises: Assertions; Transactions; Triggers; Tree Protocol; Hierarchical Locking Protocol 205
Bibliographical Notes 210

CHAPTER 5
Distributed Technology 212

1 Architecture of Database Systems 212
2 Distributed Executions and Integrity 216

BIBLIOTHEQUE DU CERIST

3 Distributed Query Processing 222
4 Distributed Updating 231
Exercises: Fragmentation; Transaction Structure; Integrity Constraints and Data Distribution; Generalization and Fragmentation; Multidatabase Systems; Catalog Management; Object Naming 237
Bibliographical Notes 244

References 245

Index 249

BIBLIOTHEQUE DU CERIST