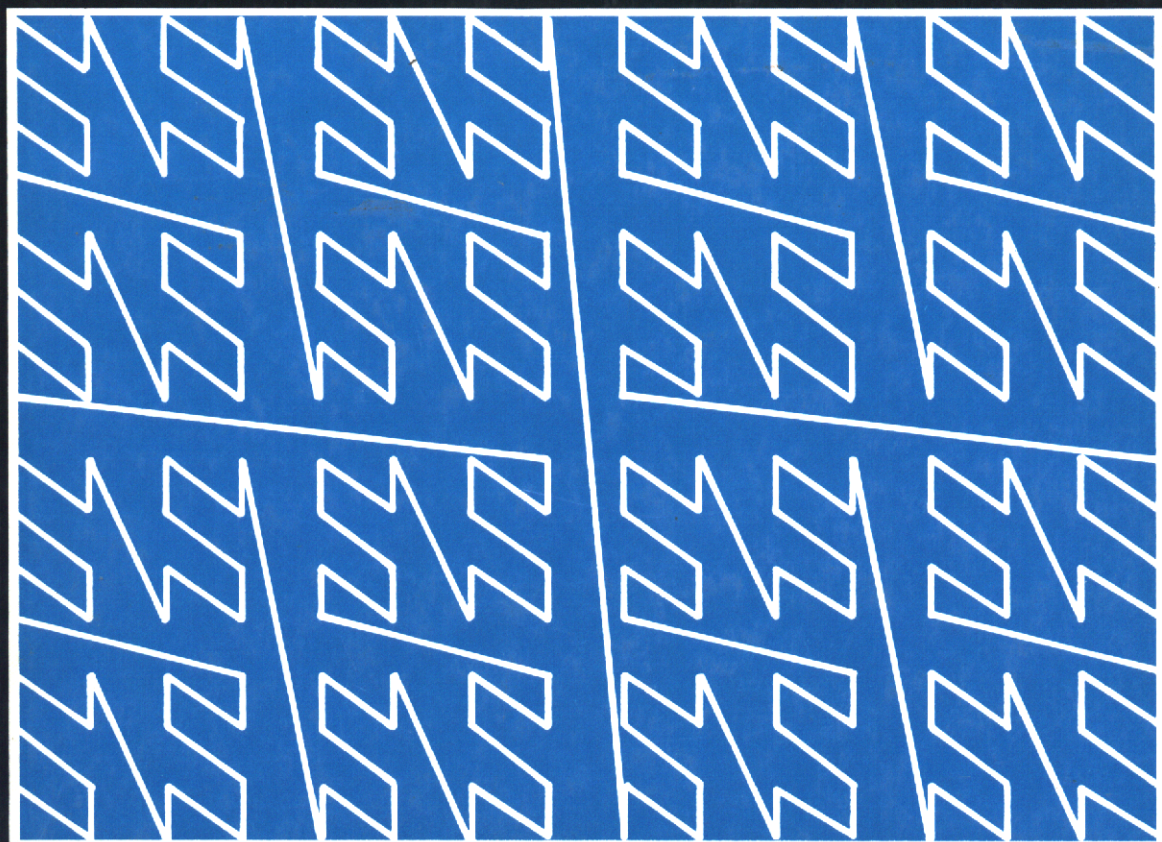
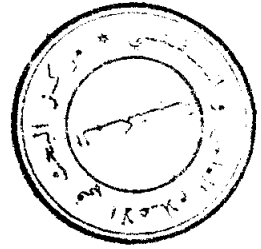


RELATIONAL INFORMATION SYSTEMS



T. H. MERRETT

Relational Information Systems



Relational Information Systems

T.H. Merrett

McGill University, Montreal



Reston Publishing Co.
A Prentice-Hall Company
Reston, Virginia

Library of Congress Cataloging in Publication Data

Merrett, Timothy Howard

Relational information systems.

Bibliography: p.

1. Data base management. 2. Information storage
and retrieval systems. I. Title.

QA76.9.D3M45 1983 001.64 82-25036

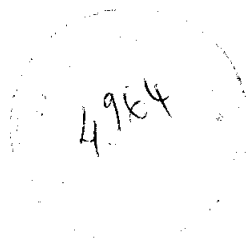
ISBN 0-8359-6642-9

© 1984 by Reston Publishing Company, Inc.
A Prentice-Hall Company
Reston, Virginia 22090

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

10 9 8 7 6 5 4 3

Printed in the United States of America



To
Hazel Howard
and
John Campbell Merrett

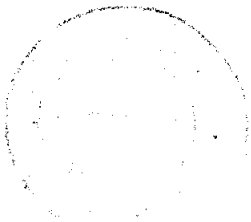


TABLE OF CONTENTS

PREFACE	xiii
----------------------	-------------

PART 1 RELATIONS	1
------------------------------------	----------

Chapter 1.1 BASICS: About Data	3
--	----------

 1.1-1 Instances of Relations, 3

 1.1-2 Changing Data, 12

 Review Questions and Exercises, Chapter 1.1, 16

Chapter 1.2 IMPLEMENTATION: Data Structures ..	19
--	-----------

 1.2-1 Secondary Storage, 20

 1.2-2 Files, 25

 1.2-3 Direct Files, 32

 1.2-3a Hashing, 32

 Volatility, 34

 Virtual Hashing, 35

 Activity and Symmetry, 37

 1.2-3b Tidying, 38

 Multipaging, 42

 Dynamic Multipaging, 51

 1.2-3c Summary of Direct Access, 57

 1.2-4 Logarithmic Files, 58

 B-Trees as Secondary Indexes, 59

 Volatility, 61

 Activity, 64

 Symmetry, 66

 Tries, 72

 Summary of Logarithmic Files, 78

 1.2-5 Sequential Files, 79

 Sorting, 80

 1.2-6 Auxiliary Methods, 83

 Differential Files, 83

 Sorted Batches, 86

 Indexing, 87

 Review Questions and Exercises, Chapter 1.2, 88

Chapter 1.3	COST ANALYSIS	91
1.3-1	Computational Complexity, 91	
1.3-2	Practical Cost Analysis, 93	
1.3-3	Direct Files, 94	
	Hashing, 95	
	Virtual Hashing—Applying the Analysis, 102	
	Activity—Usage Distributions, 104	
1.3-4	Logarithmic Files, 107	
1.3-5	Sequential Files, 108	
	Sorting, 108	
	Request Distributions, 111	
	Sequential Access Time, 113	
1.3-6	Summary of Cost Analysis, 114	
Chapter 1.4	ADVANCED TOPICS	117
1.4-1	Relations, 117	
1.4-2	Functional Dependence, 119	
1.4-3	Integrity Constraints and Data Semantics, 122	
1.4-4	File Structures, 126	
	Direct Access Files, 126	
	Multidimensional Storage Structures, 127	
	Logarithmic Files, 127	
1.4-5	Cost Analysis, and Modeling, 128	

PART 2 OPERATIONS ON RELATIONS ... 131

Chapter 2.1	BASICS: Relational Algebra	133
2.1-1	Assignments, 133	
2.1-2	Taking Relations Apart—Unary Operations, 134	
	The Relational Editor, 135	
2.1-3	Putting Relations Together—Binary Operations, 136	
	Natural Join, 136	
	Generalized Joins, 141	
	Set Selectors, 143	
	Range Joins, 146	
2.1-4	Null Values, 147	
2.1-5	Simplicity, Normalization and Hierarchies, 150	
	Review Questions and Exercises, Chapter 2.1, 156	
Chapter 2.2	IMPLEMENTATION	159
2.2-1	Sort-Merge Techniques, 159	
	Projection, 159	

	μ -Join, 160,	
	σ -Join, 161	
	Range Joins, 163	
2.2-2	Multipaging Techniques, 164	
	Select, 166	
	Project, 167	
	Sigma-Join, 167	
	Mu-Join, 168	
2.2-3	Techniques for Low Activity, 168	
2.2-4	Techniques for Algebraic Expressions, 170	
	Review Questions and Exercises, Chapter 2.2, 176	
Chapter 2.3	COST ANALYSIS	177
2.3-1	Sort-Merge is Best for Natural Join, 177	
2.3-2	Tuple-Density Models of Relations, 181	
	The Tuple-Density Model, 181	
	Projection, 184	
	Natural Join, 186	
	μ -Join, 189	
	σ -Join, 190	
2.3-3	Costs of Multipaging, 192	
2.3-4	Summary of Costs, 195	
Chapter 2.4	ADVANCED TOPICS	197
2.4-1	Operations on Relations, 197	
2.4-2	Decomposition, 199	
2.4-3	Null Values, 201	
2.4-4	Implementing Operations and Expressions, 202	
2.4-5	Systems and Languages, 204	
2.4-6	Distributions, Correlations and Selectivities, 210	

PART 3 OPERATIONS ON ATTRIBUTES . 215

Chapter 3.1	BASICS: Domain Algebra	217
3.1-1	Scalar Operations, 218	
3.1-2	Reduction, 218	
3.1-3	Functional Mapping, 221	
	Review Questions and Exercises, Chapter 3.1, 224	
Chapter 3.2	IMPLEMENTATION: Actualization	225
3.2-1	Sequential Techniques, 225	

3.2-2	Multipaging Techniques, 227	
3.2-3	Pipelining, 228	
Chapter 3.4	ADVANCED TOPICS	233
3.4-1	The Domain Algebra, 233	

PART 4 QUERIES AND QUANTIFIERS ... 235

Chapter 4.1	BASICS: QT-Expressions	237
4.1-1	QT-Selectors, 238	
4.1-2	Updates, 242	
4.1-3	Integrity, Concurrency, Reliability, 245	
	Review Questions and Exercises, Chapter 4.1, 248	
Chapter 4.2	IMPLEMENTATION	253
4.2-1	Using Projection and Equivalence Reduction, 253	
4.2-2	Using a Single Pass, 254	
Chapter 4.4	ADVANCED TOPICS: Queries and Updates	257
4.4-1	Query Languages, 257	
4.4-2	Updating Views, 263	
4.4-3	Enforcing Integrity Constraints, 267	
4.4-4	Concurrency, 273	
4.4-5	Reliability, 281	

PART 5 A MANUFACTURING PROFIT APPLICATION 285

Chapter 5.1	BASICS: A Relational Model of a Manufacturing Firm	287
5.1-1	Bill-of-Materials and Routing, 287	
5.1-2	Production Costs, 294	
5.1-3	Finished Goods Cost, 297	
5.1-4	Finished Goods Profit, 299	
5.1-5	System Overview, 301	
Chapter 5.2	IMPLEMENTATION: Relational Closure	303
5.2-1	Closure Using Relational Algebra, 304	
5.2-2	Topological Sorting of Large Files, 306	
5.2-3	Closure by Direct Methods, 309	
5.2-4	Closure Joins, 313	

Chapter 5.3	COST ANALYSIS: Mainly Topological Sorting	321
5.3-1	The Cost of Hashing, 321	
5.3-2	Extrapolating Graphs, 323	
5.3-3	Cost of Hardy Topological Sort, 324	
5.3-4	Closure and Closure Joins, 327	

PART 6 A FINANCIAL ACCOUNTING SYSTEM 333

Chapter 6.1	BASICS: An Information System Based on Funds Flow	335
6.1-1	Accounting and Information Systems, 335	
6.1-2	Spread Sheets and Financial Statements, 338	
6.1-3	Operational Data and Spread Sheets, 345	
Chapter 6.2	IMPLEMENTATION: Chronological Relations	351
6.2-1	Multipaging the Spread Sheet, 352	
6.2-2	The Differential Files, 356	
	The Main File, 357	
	The Differential Post-file, 358	
	The Differential Ante-file, 360	
6.2-3	The Monthly Reorganization, 361	
	Downdating, 361	
6.2-4	Advantages of Chronological Relations, 362	
Chapter 6.3	COST ANALYSIS	365
6.3-1	Multipaged Files, 365	
6.3-2	Indexed Files, 369	
6.3-3	Via Relational Algebra, 370	

APPENDIX 371

Graphs and Networks, 371
Trees and Forests, 375

ANSWERS TO SELECTED EXERCISES	379
BIBLIOGRAPHY	455
GLOSSARY OF SYMBOLS	475
INDEX	489