

# DESIGN OF DATABASE STRUCTURES

TOBY J. TEOREY  
JAMES P. FRY

*The University of Michigan*

Technische Hochschule Darmstadt  
FACHBEREICH INFORMATIK

BIBLIOTHEK

Inventar-Nr.: 5263

Sachgebiete: \_\_\_\_\_

Standort: \_\_\_\_\_

PRENTICE-HALL, INC., Englewood Cliffs, N.J. 07632

# CONTENTS

## Preface

xiii

## I

## INTRODUCTION

<b>1</b>	<b>Database Systems</b>	<b>3</b>
1-1	Data and Database Management	3
1-2	Levels of Data Representation	5
1-3	The Database Design Problem	12
1-4	Database System Life Cycle	14
1-4-1	<i>Analysis and design phase</i>	15
1-4-2	<i>Database implementation and operation phase</i>	16
1-5	Summary	17
<b>2</b>	<b>The Database Design Process</b>	<b>18</b>
2-1	The Concept of a Design Methodology	18
2-2	Overview of Database Design: the Basic Steps	25
2-3	Design Issues	29
2-3-1	<i>Integrity, consistency, and recovery</i>	31
2-3-2	<i>Security</i>	32
2-3-3	<i>Efficiency</i>	33
<b>3</b>	<b>Requirements Formulation and Analysis</b>	<b>35</b>
3-1	Introduction to Requirements Analysis	36
3-1-1	<i>Need for corporate requirements analysis</i>	37
3-1-2	<i>Requirements formulation and analysis steps</i>	37
3-2	Defining Scope	38

- 3-3 Collecting Information About Data Usage 39
  - 3-3-1 *Operational functions of the business* 41
  - 3-3-2 *Conducting interviews* 42
  - 3-3-3 *Management interviews for control and planning functions* 44
- 3-4 Requirements Information Transformation 45
  - 3-4-1 *Identifying data elements* 45
  - 3-4-2 *Identifying operational tasks* 46
  - 3-4-3 *Identifying control and planning tasks* 50
  - 3-4-4 *Identifying current and future operating policies* 50
- 3-5 Tools for Formulating and Documenting Requirements 51

## II

### CONCEPTUAL DESIGN

- 4 Conceptual Data Modeling 57**
  - 4-1 Framework for Conceptual Design 57
  - 4-2 Object Representation 61
  - 4-3 Entity Modeling 66
  - 4-4 Methodologies for Conceptual Design 72
    - 4-4-1 *Entity analysis* 72
    - 4-4-2 *Attribute synthesis* 74
- 5 Entity Formulation and Analysis 76**
  - 5-1 Introduction 76
    - 5-1-1 *Design objectives and scope* 76
    - 5-1-2 *Definition of design perspectives* 77
  - 5-2 Modeling of Design Perspectives 78
    - 5-2-1 *Model constructs* 78
    - 5-2-2 *Formulation of design perspectives* 79
  - 5-3 Consolidation of User Views 82
    - 5-3-1 *Concepts and principles for consolidation* 83
    - 5-3-2 *Consolidation types* 85
    - 5-3-3 *Consolidation process* 93
- 6 Attribute Synthesis: An Example of Conceptual Design 97**
  - 6-1 The Basic Model 97
  - 6-2 Analyzing Information to Identify the Components of the Information (Conceptual) Structure 99
    - 6-2-1 *Identifying entities and attributes* 99
    - 6-2-2 *Identifying relationships* 112
  - 6-3 Expressing Entities, Attributes, and Relationships Through the Graphical Notations of the Entity-Relationship Information Structure 119

- 6-4 Interpreting the Information So That It May Be Verified by All Users 123

### III

## IMPLEMENTATION DESIGN

### 7 Implementation Design Concepts 133

- 7-1 Implementation Design Components 133
- 7-2 Implementation Design Steps 135
- 7-3 Logical Database Structure Performance 138
  - 7-3-1 Logical record access approach 139
  - 7-3-2 Relationship between logical and physical performance 143
- 7-4 Alternative Implementation Design Tools 146

### 8 An Example Schema Design Problem 148

- 8-1 Requirements Specifications 148
- 8-2 Conceptual Design 150
- 8-3 Implementation Design 151
- 8-4 Summary of the Major Design Issues 160

### IV

## PHYSICAL DESIGN

### 9 Physical Database Design Principles: Basic Concepts 167

- 9-1 Introduction to Physical Design 167
  - 9-1-1 Physical design steps 168
  - 9-1-2 Physical design environment 170
  - 9-1-3 Performance measures 171
- 9-2 Performance Computation 177
  - 9-2-1 I/O service time: the basic model 178
  - 9-2-2 I/O service time for disk storage 180
  - 9-2-3 Dedicated and shared computing environments 181
- 9-3 Secondary Storage Space 186
  - 9-3-1 File storage 186
  - 9-3-2 Database storage: multiple-record types 190
- 9-4 Summary 193

### 10 Record Structure Design 194

- 10-1 Data Item Encoding and Compression 194
  - 10-1-1 Data item representation 195
  - 10-1-2 Compression techniques 196

10-2	Record partitioning	201
10-2-1	<i>Record Segmentation Algorithms</i>	201
10-2-2	<i>The bond energy algorithm</i>	205
10-2-3	<i>A heuristic partitioning model</i>	211
<b>11</b>	<b>Record Clustering</b>	<b>212</b>
11-1	Clustering in Hierarchical Databases	213
11-2	Clustering in Network Database Structures	219
<b>12</b>	<b>Primary Access Methods: Sequential Processing</b>	<b>225</b>
12-1	Introduction	225
12-1-1	<i>Access method terminology</i>	226
12-1-2	<i>Classification of access methods</i>	226
12-2	Physical Sequential Processing: GET MANY, GET ALL	228
12-2-1	<i>Physical sequential retrieval</i>	228
12-2-2	<i>Physical sequential update</i>	231
12-2-3	<i>Trade-offs in physical sequential organizations</i>	233
12-3	Linked Sequential Processing	237
12-3-1	<i>Linked sequential retrieval</i>	237
12-3-2	<i>Linked sequential update</i>	238
12-3-3	<i>Trade-offs in linked sequential organizations</i>	240
12-3-4	<i>Storage space for sequential organizations</i>	243
12-4	Report Generation: Total Cost	243
<b>13</b>	<b>Primary Access Methods: Random Processing</b>	<b>248</b>
13-1	Direct Access	248
13-2	Identifier Hashing (Random Access)	251
13-2-1	<i>Hashing functions</i>	253
13-2-2	<i>Overflow techniques</i>	255
13-2-3	<i>Performance characteristics</i>	267
13-2-4	<i>Scatter tables</i>	269
13-3	Full Index (Indexed Random)	271
13-4	Indexed Sequential	274
13-4-1	<i>Retrieval from an indexed sequential file</i>	276
13-4-2	<i>Update to an indexed sequential file</i>	287
13-4-3	<i>Storage space for indexed sequential organizations</i>	289
13-4-4	<i>Trade-off analysis in indexed sequential organizations</i>	289
<b>14</b>	<b>Primary Access Methods: Search Trees and Random Processing</b>	<b>293</b>
14-1	Binary Search Trees	293
14-1-1	<i>Retrieval performance</i>	296
14-1-2	<i>Update performance</i>	298
14-1-3	<i>Storage space</i>	304

- 14-2 B-Trees 305
  - 14-2-1 Retrieval performance 309
  - 14-2-2 Update performance 313
  - 14-2-3 Storage space 321
  - 14-2-4 B\*-trees 321
  - 14-2-5 Prefix B\*-trees 326
- 14-3 TRIE Structures 327
  - 14-3-1 Retrieval performance 330
  - 14-3-2 Update performance 330
  - 14-3-3 Storage space 331
  - 14-3-4 TRIE/B-tree trade-offs 332

## 15 Secondary Access Methods

334

- 15-1 Introduction 334
- 15-2 Multilist File 335
  - 15-2-1 Retrieval performance for query applications 338
  - 15-2-2 Generalized update performance 340
  - 15-2-3 Storage space 341
  - 15-2-4 Cellular multilist 342
- 15-3 Inverted File 344
  - 15-3-1 Retrieval performance for query applications 346
  - 15-3-2 Generalized update performance 348
  - 15-3-3 Storage space 349
  - 15-3-4 Trade-offs in secondary storage access methods 350
  - 15-3-5 Cellular inverted file 352
- 15-4 Doubly Chained Tree 355
  - 15-4-1 Retrieval performance for query applications 359
  - 15-4-2 Generalized update performance 361
  - 15-4-3 Storage space 362
  - 15-4-4 Comparison with inverted and multilist files 363
- 15-5 Inverted Index Maintenance Alternatives 366

## 16 Secondary Index Selection

370

- 16-1 Multiple-Key (Combined) Indexing 370
- 16-2 Categorization of Secondary Access Methods 373
- 16-3 The Index Selection Problem 377
- 16-4 Optimal Secondary Index Selection 380

# V

## SPECIAL DESIGN ISSUES

## 17 Reorganization

387

- 17-1 Introduction 387
- 17-2 Reorganization Strategies 389

- 17-3 The Database Administrator's Role 391
- 17-4 When to Reorganize: a Heuristic Rule 392
- 17-5 Restructuring 397
  - 17-5-1 Network restructuring 399
  - 17-5-2 Technical approaches to restructuring 401

## **18 Distributed Database Design: An Overview 408**

- 18-1 Introduction 408
- 18-2 Distributed DBMS Architecture 409
- 18-3 Design Issues in a Distributed Database Environment 414
  - 18-3-1 Data distribution strategies 415
  - 18-3-2 Network data directory distributions 419
  - 18-3-3 Homogeneous versus heterogeneous environments 420
- 18-4 A Framework for Distributed Database Design 422
  - 18-4-1 Database partitioning 422
  - 18-4-2 Database allocation 425
- 18-5 Differential Files 427

## **APPENDICES**

<b>A</b>	<b>Exercises in Conceptual and Implementation Schema Design</b>	<b>437</b>
<b>B</b>	<b>Exercises in Physical Database Design</b>	<b>445</b>
<b>C</b>	<b>List of Variables</b>	<b>455</b>
	<b>Glossary</b>	<b>460</b>
	<b>References</b>	<b>470</b>
	<b>Index</b>	<b>485</b>