

Database Solutions

A step-by-step guide to building databases

Second edition

Thomas M. Connolly

Carolyn E. Begg



Brief contents

Part 1 Background

1	Introduction	3
2	The relational model	21
3	SQL and QBE	37
4	The database system development lifecycle	78
5	Database administration and security	97

Part 2 Database analysis and design techniques

6	Fact-finding	115
7	Entity-Relationship modeling	146
8	Normalization	171

Part 3 Logical database design

9	Logical database design – Step 1	191
10	Logical database design – Step 2	219
11	Enhanced ER modeling techniques	246

Part 4 Physical database design

12	Physical database design – Step 3	261
13	Physical database design – Step 4	278
14	Physical database design – Steps 5 and 6	296
15	Physical database design – Step 7	305
16	Physical database design – Step 8	319

Part 5 Second worked example

17	<i>PerfectPets</i> – Logical database design	331
18	<i>PerfectPets</i> – Physical database design	350

Part 6 Current and emerging trends

19	Current and emerging trends	377
----	-----------------------------	-----

Appendices

Appendix A	Alternative data modeling notations	413
Appendix B	Summary of the database design methodology	420
Appendix C	Advanced logical database design	428
Appendix D	File organizations and indexes	442
Appendix E	Common data models	462
Glossary		503
References		513
Index		515

Contents

Preface

xvii

Part 1 Background

1	Introduction	3
1.1	Examples of the use of database systems	4
1.2	Database approach	7
1.2.1	The database	7
1.2.2	The Database Management System (DBMS)	8
1.2.3	(Database) application programs	8
1.2.4	Views	8
1.2.5	Components of the DBMS environment	10
1.2.6	DBMS architectures	10
1.3	Functions of a DBMS	12
1.4	Database design	17
1.5	Advantages and disadvantages of DBMSs	17
	Chapter summary	19
	Review questions	20
2	The relational model	21
2.1	What is a data model?	22
2.2	Terminology	22
2.2.1	Relational data structure	22
2.2.2	Properties of relational tables	26
2.2.3	Relational keys	26
2.2.4	Representing relational databases	29
2.3	Relational integrity	32
2.3.1	Nulls	32
2.3.2	Entity integrity	33
2.3.3	Referential integrity	33

2.3.4	Other business rules	34
2.4	Relational languages	34
	Chapter summary	35
	Review questions	35
3	SQL and QBE	37
3.1	Structured Query Language (SQL)	38
3.1.1	Objectives of SQL	39
3.1.2	Terminology	39
3.1.3	Writing SQL commands	40
3.2	Data manipulation	41
3.2.1	Simple queries	42
3.2.2	Row selection (WHERE clause)	45
3.2.3	Sorting results (ORDER BY clause)	50
3.2.4	Using the SQL aggregate functions	51
3.2.5	Grouping results (GROUP BY clause)	53
3.2.6	Subqueries	56
3.2.7	Multi-table queries	58
3.2.8	INSERT, UPDATE, and DELETE statements	61
3.3	Data definition	63
3.3.1	CREATE TABLE	63
3.3.2	CREATE VIEW	67
3.4	Query-by-Example (QBE)	67
	Chapter summary	73
	Review questions	75
	Exercises	75
4	The database system development lifecycle	78
4.1	The software crisis	78
4.2	The information systems lifecycle	79
4.3	The database system development lifecycle	80
4.4	Database planning	80
4.5	System definition	82
4.5.1	User views	83
4.6	Requirements collection and analysis	83
4.7	Database design	86
4.8	DBMS selection	88
4.9	Application design	89
4.9.1	Transaction design	89
4.9.2	User interface design	90
4.10	Prototyping	90
4.11	Implementation	91
4.12	Data conversion and loading	92
4.13	Testing	93
4.14	Operational maintenance	94
	Chapter summary	94
	Review questions	96

5	Database administration and security	97
5.1	Data administration and database administration	98
5.1.1	Data administration	98
5.1.2	Database administration	99
5.1.3	Comparison of data and database administration	99
5.2	Database security	100
5.2.1	Threats	102
5.2.2	Countermeasures – computer-based controls	103
	Chapter summary	111
	Review questions	112

Part 2 Database analysis and design techniques

6	Fact-finding	115
6.1	When are fact-finding techniques used?	116
6.2	What facts are collected?	116
6.3	Fact-finding techniques	118
6.3.1	Examining documentation	118
6.3.2	Interviewing	118
6.3.3	Observing the business in operation	120
6.3.4	Research	120
6.3.5	Questionnaires	121
6.4	The <i>StayHome</i> case study	123
6.4.1	The <i>StayHome</i> case study – an overview	123
6.4.2	The <i>StayHome</i> case study – database planning	126
6.4.3	The <i>StayHome</i> case study – system definition	133
6.4.4	The <i>StayHome</i> case study – requirements collection and analysis	135
6.4.5	The <i>StayHome</i> case study – database design	143
	Chapter summary	144
	Review questions	144
7	Entity-Relationship modeling	146
7.1	Entities	147
7.2	Relationships	148
7.2.1	Degree of a relationship	149
7.2.2	Recursive relationships	150
7.3	Attributes	151
7.3.1	Simple and composite attributes	151
7.3.2	Single-valued and multi-valued attributes	152
7.3.3	Derived attributes	152
7.3.4	Keys	153
7.4	Strong and weak entities	155

7.5	Multiplicity constraints on relationships	155
7.5.1	One-to-one (1:1) relationships	156
7.5.2	One-to-many (1:*) relationships	157
7.5.3	Many-to-many (*:*) relationships	159
7.5.4	Multiplicity for complex relationships	159
7.5.5	Cardinality and participation constraints	162
7.6	Attributes on relationships	162
7.7	Design problems with ER models	163
7.7.1	Fan traps	163
7.7.2	Chasm traps	165
	Chapter summary	168
	Review questions	169
	Exercises	169
8	Normalization	171
8.1	Introduction	172
8.2	Data redundancy and update anomalies	172
8.2.1	Insertion anomalies	174
8.2.2	Deletion anomalies	174
8.2.3	Modification anomalies	174
8.3	First normal form (1NF)	175
8.4	Second normal form (2NF)	177
8.5	Third normal form (3NF)	180
	Chapter summary	185
	Review questions	186
	Exercises	186

Part 3 Logical database design

9	Logical database design – Step 1	191
9.1	Introduction to the database design methodology	192
9.1.1	What is a design methodology?	192
9.1.2	Phases of database design	192
9.1.3	Critical success factors in database design	193
9.2	Overview of the database design methodology	194
9.3	Introduction to Step 1 of the logical database design methodology	196
Step 1	Create and check ER model	197
Step 1.1	Identify entities	197
Step 1.2	Identify relationships	199
Step 1.3	Identify and associate attributes with entities or relationships	204
Step 1.4	Determine attribute domains	208
Step 1.5	Determine candidate, primary, and alternate key attributes	209

Step 1.6 Specialize/Generalize entities (optional step)	211
Step 1.7 Check model for redundancy	212
Step 1.8 Check model supports user transactions	215
Step 1.9 Review model with user	217
Chapter summary	217
Review questions	218
Exercise	218
10 Logical database design – Step 2	219
Step 2 Map ER model to tables	219
Step 2.1 Create tables	220
Step 2.2 Check table structures using normalization	234
Step 2.3 Check tables support user transactions	235
Step 2.4 Check business rules	239
Step 2.5 Review logical database design with users	242
Chapter summary	243
Review questions	244
Exercise	245
11 Enhanced ER modeling techniques	246
11.1 Specialization/Generalization	247
11.1.1 Superclasses and subclasses	247
11.1.2 Superclass/Subclass relationships	247
11.1.3 Attribute inheritance	249
11.1.4 Specialization process	249
11.1.5 Generalization process	250
11.1.6 Constraints on superclass/subclass relationships	252
11.2 Creating tables to represent specialization/generalization	254
Chapter summary	256
Review questions	257
Exercises	257

Part 4 Physical database design

12 Physical database design – Step 3	261
12.1 Comparison of logical and physical database design	262
12.2 Overview of the physical database design methodology	263
Step 3 Translate logical database design for target DBMS	264
Step 3.1 Design base tables	265
Step 3.2 Design representation of derived data	270
Step 3.3 Design remaining business rules	272
Chapter summary	276
Review questions	277
Exercise	277

13	Physical database design – Step 4	278
	Step 4 Choose file organizations and indexes	279
	Step 4.1 Analyze transactions	280
	Step 4.2 Choose file organizations	286
	Step 4.3 Choose indexes	287
	13.1 File organizations and indexes for <i>StayHome</i> with Microsoft Access 2002	292
	13.1.1 Guidelines for indexes	292
	13.1.2 Indexes for <i>StayHome</i>	293
	Chapter summary	294
	Review questions	295
	Exercise	295
14	Physical database design – Steps 5 and 6	296
	Step 5 Design user views	296
	Step 6 Design security mechanisms	298
	Chapter summary	303
	Review questions	304
	Exercise	304
15	Physical database design – Step 7	305
	Step 7 Consider the introduction of controlled redundancy	305
	Chapter summary	318
	Review questions	318
	Exercise	318
16	Physical database design – Step 8	319
	Step 8 Monitor and tune the operational system	320
	Understanding system resources	321
	Summary	324
	New requirement from <i>StayHome</i>	325
	Chapter summary	326
	Review questions	327

Part 5 Second worked example

17	<i>PerfectPets</i> – Logical database design	331
	17.1 <i>PerfectPets</i>	331
	17.1.1 Data requirements	331
	17.1.2 Transaction requirements	334
	17.2 Using the logical database design methodology	335
18	<i>PerfectPets</i> – Physical database design	350
	18.1 Using the physical database design methodology	350

Part 8 Current and emerging trends

19	Current and emerging trends	377
19.1	Advanced database applications	378
19.2	Weaknesses of Relational DBMSs (RDBMSs)	381
19.3	Distributed DBMSs and replication servers	383
19.3.1	Advantages and disadvantages of DDBMSs	385
19.3.2	Replication servers	388
19.4	Object-oriented DBMSs and object-relational DBMSs	390
19.4.1	Object-oriented DBMSs (OODBMSs)	391
19.4.2	Object-relational DBMSs (ORDBMSs)	393
19.5	Data warehousing	395
19.5.1	Data marts	397
19.6	Online Analytical Processing (OLAP)	399
19.7	Data mining	401
19.8	Web-database integration and XML	402
19.8.1	Static and dynamic Web pages	403
19.8.2	Requirements for Web-DBMS integration	404
19.8.3	Approaches to integrating the Web and DBMSs	405
19.8.4	XML	405
	Chapter summary	407
	Review questions	409

Appendices

A	Alternative data modeling notations	413
A.1	ER modeling using the Chen notation	413
A.2	ER modeling using the Crow's Feet notation	413
B	Summary of the database design methodology	420
C	Advanced logical database design	428
C.1	The Business user views of <i>StayHome</i>	429
C.1.1	Users' requirements specification	429
C.1.2	Local logical data model	431
	Step 2.6 Build and check global logical data model	431
	Appendix summary	441
D	File organizations and indexes	442
D.1	Basic concepts	443
D.2	Heap files	445
D.3	Ordered files	445
D.4	Hash files	447
D.5	Indexes	448

- D.5.1 Types of indexes
- D.5.2 Secondary indexes
- D.5.3 Multilevel indexes
- D.5.4 B⁺-Trees
- D.5.5 Bitmap indexes
- D.5.6 Join indexes
- D.6 Guidelines for selecting file organizations
- D.7 Clustered and non-clustered tables
 - D.7.1 Index clusters
 - D.7.2 Hash clusters
- Appendix summary

- E Common data models
 - E.1 Customer order entry
 - E.2 Inventory control
 - E.3 Asset management
 - E.4 Project management
 - E.5 Course management
 - E.6 Human resource management
 - E.7 Payroll management
 - E.8 Vehicle rentals
 - E.9 Student accommodation
 - E.10 Client transportation
 - E.11 Publisher printing
 - E.12 County library
 - E.13 Real estate rentals
 - E.14 Travel agent
 - E.15 Student results

Glossary

References

Index