# Contents


| | | |
|---|---|---|
| 1 | **Introduction** | 1 |
| 1.1 | Information and data | 1 |
| 1.2 | Databases and database management systems | 3 |
| 1.3 | Data models | 5 |
| | 1.3.1 Schemas and instances | 6 |
| | 1.3.2 Abstraction levels in DBMSs | 6 |
| | 1.3.3 Data independence | 7 |
| 1.4 | Languages and users | 8 |
| | 1.4.1 Database languages | 8 |
| | 1.4.2 Users and designers | 9 |
| 1.5 | Advantages and disadvantages of DBMSs | 10 |
| 1.6 | Bibliography | 10 |

**Part I. Relational databases** **13**

| | | |
|---|---|---|
| 2 | **The relational model** | 15 |
| 2.1 | The structure of the relational model | 15 |
| | 2.1.1 Logical models in database systems | 15 |
| | 2.1.2 Relations and tables | 16 |
| | 2.1.3 Relations with attributes | 18 |
| | 2.1.4 Relations and databases | 20 |
| | 2.1.5 Incomplete information and null values | 26 |
| 2.2 | Integrity constraints | 28 |
| | 2.2.1 Tuple constraints | 30 |
| | 2.2.2 Keys | 30 |
| | 2.2.3 Keys and null values | 33 |
| | 2.2.4 Referential constraints | 34 |
| 2.3 | Conclusions | 38 |
| 2.4 | Bibliography | 38 |
| 2.5 | Exercises | 39 |

| | | |
|---|---|---|
| 3 | **Relational algebra and calculus** | 41 |
| 3.1 | Relational algebra | 42 |
| | 3.1.1 Union, intersection, difference | 42 |
| | 3.1.2 Renaming | 43 |
| | 3.1.3 Selection | 45 |
| | 3.1.4 Projection | 47 |
| | 3.1.5 Join | 49 |
| | 3.1.6 Queries in relational algebra | 56 |
| | 3.1.7 Equivalence of algebraic expressions | 59 |
| | 3.1.8 Algebra with null values | 62 |
| | 3.1.9 Views | 65 |
| 3.2 | Relational calculus | 67 |
| | 3.2.1 Domain relational calculus | 68 |
| | 3.2.2 Qualities and drawbacks of domain calculus | 72 |
| | 3.2.3 Tuple calculus with range declarations | 74 |
| 3.3 | Datalog | 77 |
| 3.4 | Bibliography | 80 |
| 3.5 | Exercises | 81 |
| 4 | SQL | 85 |
| 4.1 | Data definition in SQL | 87 |
| | 4.1.1 Elementary domains | 87 |
| | 4.1.2 Schema definition | 90 |
| | 4.1.3 Table definition | 91 |
| | 4.1.4 User defined domains | 91 |
| | 4.1.5 Default domain values | 92 |
| | 4.1.6 Intra-relational constraints | 93 |
| | 4.1.7 Inter-relational constraints | 94 |
| | 4.1.8 Schema updates | 97 |
| | 4.1.9 Relational catalogues | 99 |
| 4.2 | SQL queries | 100 |
| | 4.2.1 The declarative nature of SQL | 101 |
| | 4.2.2 Simple queries | 101 |
| | 4.2.3 Aggregate queries | 113 |
| | 4.2.4 Group by queries | 116 |
| | 4.2.5 Set queries | 120 |
| | 4.2.6 Nested queries | 122 |
| 4.3 | Data modification in SQL | 128 |
| | 4.3.1 Insertions | 128 |
| | 4.3.2 Deletions | 129 |
| | 4.3.3 Updates | 130 |
| 4.4 | Other definitions of data in SQL | 131 |
| | 4.4.1 Generic integrity constraints | 131 |
| | 4.4.2 Assertions | 132 |
| | 4.4.3 Views | 133 |

| | | |
|---|---|---|
| | 4.4.4 Views in queries | 135 |
| 4.5 | Access control | 136 |
| | 4.5.1 Resources and privileges | 136 |
| | 4.5.2 Commands for granting and revoking privileges | 137 |
| 4.6 | Use of SQL in programming languages | 138 |
| | 4.6.1 Integration problems | 139 |
| | 4.6.2 Cursors | 140 |
| | 4.6.3 Dynamic SQL | 142 |
| | 4.6.4 Procedures | 145 |
| 4.7 | Summarizing examples | 147 |
| 4.8 | Bibliography | 150 |
| 4.9 | Exercises | 150 |

**Part II. Database design** **155**

| | | |
|---|---|---|
| 5 | **Design techniques and models** | 157 |
| 5.1 | The database design process | 158 |
| | 5.1.1 The life cycle of information systems | 158 |
| | 5.1.2 Methodologies for database design | 160 |
| 5.2 | The Entity-Relationship model | 163 |
| | 5.2.1 The basic constructs of the model | 165 |
| | 5.2.2 Other constructs of the model | 170 |
| | 5.2.3 Final overview of the E-R model | 177 |
| 5.3 | Documentation of E-R schemas | 179 |
| | 5.3.1 Business rules | 180 |
| | 5.3.2 Documentation techniques | 182 |
| 5.4 | Bibliography | 183 |
| 5.5 | Exercises | 184 |
| 6 | **Conceptual design** | 189 |
| 6.1 | Requirements collection and analysis | 189 |
| 6.2 | General criteria for data representation | 194 |
| 6.3 | Design strategies | 196 |
| | 6.3.1 Top-down strategy | 196 |
| | 6.3.2 Bottom-up strategy | 198 |
| | 6.3.3 Inside-out strategy | 201 |
| | 6.3.4 Mixed strategy | 202 |
| 6.4 | Quality of a conceptual schema | 203 |
| 6.5 | A comprehensive method for conceptual design | 204 |
| 6.6 | An example of conceptual design | 205 |
| 6.7 | CASE tools for database design | 209 |
| 6.8 | Bibliography | 211 |
| 6.9 | Exercises | 211 |
| 7 | **Logical design** | 217 |
| 7.1 | Performance analysis on E-R schemas | 218 |

7.2 Restructuring of E-R schemas 222
7.2.1 Analysis of redundancies 223
7.2.2 Removing generalizations 226
7.2.3 Partitioning and merging of entities and relationships 229
7.2.4 Selection of primary identifiers 233
7.3 Translation into the relational model 234
7.3.1 Entities and many-to-many relationships 234
7.3.2 One-to-many relationships 236
7.3.3 Entities with external identifiers 237
7.3.4 One-to-one relationships 238
7.3.5 Translation of a complex schema 239
7.3.6 Summary tables 241
7.3.7 Documentation of logical schemas 241
7.4 An example of logical design 245
7.4.1 Restructuring phase 246
7.4.2 Translation into the relational model 249
7.5 Logical design using CASE tools 250
7.6 Bibliography 251
7.7 Exercises 251

8 **Normalization** 255
8.1 Redundancies and anomalies 256
8.2 Functional dependencies 257
8.3 Boyce–Codd normal form 259
8.3.1 Definition of Boyce–Codd normal form 259
8.3.2 Decomposition into Boyce–Codd normal form 260
8.4 Decomposition properties 262
8.4.1 Lossless decomposition 262
8.4.2 Preservation of dependencies 265
8.4.3 Qualities of decompositions 266
8.5 Third normal form 267
8.5.1 Definition of third normal form 267
8.5.2 Decomposition into third normal form 268
8.5.3 Other normalization techniques 269
8.6 Database design and normalization 270
8.6.1 Verification of normalization on entities 271
8.6.2 Verification of normalization on relationships 272
8.6.3 Further decomposition of relationships 274
8.6.4 Further restructurings of conceptual schemas 275
8.7 Bibliography 276
8.8 Exercises 276

**Part III. Database technology 281**

9 **Technology of a database server** 283
9.1 Definition of transactions 284

| | | |
|---|---|---|
| | 9.1.1 ACID properties of transactions | 285 |
| | 9.1.2 Transactions and system modules | 287 |
| 9.2 | Concurrency control | 287 |
| | 9.2.1 Architecture of concurrency control | 287 |
| | 9.2.2 Anomalies of concurrent transactions | 288 |
| | 9.2.3 Concurrency control theory | 290 |
| | 9.2.4 Lock management | 301 |
| | 9.2.5 Deadlock management | 305 |
| 9.3 | Buffer management | 307 |
| | 9.3.1 Architecture of the buffer manager | 307 |
| | 9.3.2 Primitives for buffer management | 308 |
| | 9.3.3 Buffer management policies | 310 |
| | 9.3.4 Relationship between buffer manager and file system | 310 |
| 9.4 | Reliability control system | 311 |
| | 9.4.1 Architecture of the reliability control system | 312 |
| | 9.4.2 Log organization | 313 |
| | 9.4.3 Transaction management | 315 |
| | 9.4.4 Failure management | 317 |
| 9.5 | Physical access structures | 320 |
| | 9.5.1 Architecture of the access manager | 321 |
| | 9.5.2 Organization of tuples within pages | 321 |
| | 9.5.3 Sequential structures | 323 |
| | 9.5.4 Hash-based structures | 325 |
| | 9.5.5 Tree structures | 327 |
| 9.6 | Query optimization | 332 |
| | 9.6.1 Relation profiles | 333 |
| | 9.6.2 Internal representation of queries | 336 |
| | 9.6.3 Cost-based optimization | 339 |
| 9.7 | Physical database design | 341 |
| | 9.7.1 Definition of indexes in SQL | 343 |
| 9.8 | Bibliography | 343 |
| 9.9 | Exercises | 344 |
| 10 | **Distributed architectures** | 349 |
| 10.1 | Client-server architecture | 351 |
| 10.2 | Distributed databases | 353 |
| | 10.2.1 Applications of distributed databases | 354 |
| | 10.2.2 Local independence and co-operation | 355 |
| | 10.2.3 Data fragmentation and allocation | 356 |
| | 10.2.4 Transparency levels | 358 |
| | 10.2.5 Classification of transactions | 360 |
| 10.3 | Technology of distributed databases | 361 |
| | 10.3.1 Distributed query optimization | 362 |
| | 10.3.2 Concurrency control | 363 |
| | 10.3.3 Failures in distributed systems | 368 |

10.4 Two-phase commit protocol 369
10.4.1 New log records 369
10.4.2 Basic protocol 370
10.4.3 Recovery protocols 372
10.4.4 Protocol optimization 374
10.4.5 Other commit protocols 375
10.5 Interoperability 377
10.5.1 Open Database Connectivity (ODBC) 377
10.5.2 X-OPEN Distributed Transaction Processing (DTP) 378
10.6 Co-operation among pre-existing systems 381
10.7 Parallelism 383
10.7.1 Inter-query and intra-query parallelism 384
10.7.2 Parallelism and data fragmentation 385
10.7.3 Speed-up and scale-up 386
10.7.4 Transaction benchmarks 387
10.8 Replicated databases 388
10.8.1 New functions of replication managers 390
10.9 Bibliography 391
10.10 Exercises 391

**Part IV. Database evolution 395**

11 **Object databases** 397
11.1 Object-Oriented databases (OODBMSs) 398
11.1.1 Types 399
11.1.2 Classes 402
11.1.3 Methods 404
11.1.4 Generalization hierarchies 408
11.1.5 Persistence 411
11.1.6 Redefinition of methods 412
11.1.7 Refinement of properties and methods 414
11.1.8 The object-oriented database manifesto 416
11.2 The ODMG standard for object-oriented databases 417
11.2.1 Object Definition Language: ODL 417
11.2.2 Object Query Language: OQL 419
11.3 Object-Relational databases (ORDBMSs) 423
11.3.1 SQL-3 data model 423
11.3.2 SQL-3 query language 427
11.3.3 The third generation database manifesto 428
11.4 Multimedia databases 429
11.4.1 Types of multimedia data 429
11.4.2 Queries on multimedia data 430
11.4.3 Document search 431
11.4.4 Representation of spatial data 432
11.5 Technological extensions for object-oriented databases 434

| | | |
|---|---|---|
| | 11.5.1 Representation of data and identifiers | 435 |
| | 11.5.2 Complex indexes | 436 |
| | 11.5.3 Client-server architecture | 437 |
| | 11.5.4 Transactions | 438 |
| | 11.5.5 Distribution and interoperability: CORBA | 439 |
| 11.6 | Bibliography | 441 |
| 11.7 | Exercises | 442 |
| 12 | **Active databases** | 447 |
| 12.1 | Trigger behaviour in a relational system | 448 |
| 12.2 | Definition and use of triggers in Oracle | 449 |
| | 12.2.1 Trigger syntax in Oracle | 449 |
| | 12.2.2 Behaviour of triggers in Oracle | 450 |
| | 12.2.3 Example of execution | 450 |
| 12.3 | Definition and use of triggers in DB2 | 452 |
| | 12.3.1 Trigger syntax in DB2 | 452 |
| | 12.3.2 Behaviour of triggers in DB2 | 453 |
| | 12.3.3 Example of execution | 453 |
| 12.4 | Advanced features of active rules | 454 |
| 12.5 | Properties of active rules | 455 |
| 12.6 | Applications of active databases | 457 |
| | 12.6.1 Referential integrity management | 457 |
| | 12.6.2 Business rules | 460 |
| 12.7 | Bibliography | 461 |
| 12.8 | Exercises | 461 |
| 13 | **Data analysis** | 465 |
| 13.1 | Data warehouse architecture | 467 |
| 13.2 | Schemas for data warehouses | 469 |
| | 13.2.1 Star schema | 469 |
| | 13.2.2 Star schema for a supermarket chain | 471 |
| | 13.2.3 Snowflake schema | 473 |
| 13.3 | Operations for data analysis | 474 |
| | 13.3.1 Query formulation interfaces | 474 |
| | 13.3.2 Drill-down and roll-up | 475 |
| | 13.3.3 Data cube | 477 |
| 13.4 | Development of the data warehouse | 479 |
| | 13.4.1 Bitmap and join indexes | 480 |
| | 13.4.2 View materialization | 481 |
| 13.5 | Data mining | 481 |
| | 13.5.1 The data mining process | 482 |
| | 13.5.2 Data mining problems | 482 |
| | 13.5.3 Data mining perspectives | 485 |
| 13.6 | Bibliography | 486 |
| 13.7 | Exercises | 486 |

14 **Databases and the World Wide Web** 489
14.1 The Internet and the World Wide Web 490
14.1.1 The Internet 490
14.1.2 The World Wide Web 491
14.1.3 HTML 492
14.1.4 HTTP 494
14.1.5 Gateways 494
14.2 Information systems on the Web 495
14.2.1 Publication and consultation on the Web 496
14.2.2 Transactions on the Web 496
14.2.3 Electronic commerce and other new applications 497
14.3 Design of data-intensive Web sites 498
14.3.1 A logical model for data-intensive hypertexts 499
14.3.2 Levels of representation in Web hypertexts 502
14.3.3 Design principles for a data-intensive Web site 505
14.4 Techniques and tools for database access through the Web 508
14.4.1 Database access through CGI programs 508
14.4.2 Development tools 510
14.4.3 Shortcomings of the CGI protocol 511
14.4.4 Simulating long connections for transactions 511
14.4.5 Server-based alternatives to the CGI approach 512
14.4.6 Client-based alternatives to the CGI approach 514
14.5 Bibliography 516
14.6 Exercises 517

**Part V. Appendices & Bibliography** **519**

**Appendix A Microsoft Access** 521
A.1 System characteristics 522
A.2 Definition of tables 523
A.2.1 Specification of join paths 528
A.2.2 Populating the table 529
A.3 Query definition 530
A.3.1 Query By Example 530
A.3.2 The SQL interpreter 536
A.4 Forms and reports 538
A.5 The definition of macros 539

**Appendix B DB2 Universal Database** 543
B.1 DB2 overview 544
B.1.1 Versions of the system 544
B.1.2 Instances and schemas of DB2 545
B.1.3 Interaction with DB2 545
B.2 Database management with DB2 546
B.2.1 Interactive tools 546
B.2.2 Application programs 551

B.3 Advanced features of DB2 554
B.3.1 Extension of SQL for queries 554
B.3.2 Object-oriented features of DB2 558

**Appendix C Oracle PL/SQL** 565
C.1 Tools architecture of Oracle 565
C.2 Base domains 567
C.3 The object-relational extension of Oracle 569
C.4 PL/SQL language 572
C.4.1 Execution of PL/SQL in a client-server environment 573
C.4.2 Declarations of variables and cursors 574
C.4.3 Control structures 576
C.4.4 Management of exceptions 578
C.4.5 Procedures 580
C.4.6 Packages 585

**Bibliography** 587

**Index** 593