Principles of Database Management
The Practical Guide to Storing, Managing and Analyzing Big and Small Data

Wilfried Lemahieu
KU Leuven, Belgium

Seppe vanden Broucke
KU Leuven, Belgium

Bart Baesens
KU Leuven, Belgium; University of Southampton, United Kingdom
# CONTENTS

About the Authors  \hspace{1cm} page xvii
Preface  \hspace{1cm} xix
Sober: 1000% Driven by Technology  \hspace{1cm} xxiv

## Part I  Databases and Database Design

### 1  Fundamental Concepts of Database Management

1.1 Applications of Database Technology  \hspace{1cm} 3
1.2 Key Definitions  \hspace{1cm} 4
1.3 File versus Database Approach to Data Management
  1.3.1 The File-Based Approach  \hspace{1cm} 5
  1.3.2 The Database Approach  \hspace{1cm} 6
1.4 Elements of a Database System
  1.4.1 Database Model versus Instances  \hspace{1cm} 8
  1.4.2 Data Model  \hspace{1cm} 9
  1.4.3 The Three-Layer Architecture  \hspace{1cm} 10
  1.4.4 Catalog  \hspace{1cm} 10
  1.4.5 Database Users  \hspace{1cm} 11
  1.4.6 Database Languages  \hspace{1cm} 12
1.5 Advantages of Database Systems and Database Management
  1.5.1 Data Independence  \hspace{1cm} 12
  1.5.2 Database Modeling  \hspace{1cm} 13
  1.5.3 Managing Structured, Semi-Structured, and Unstructured Data  \hspace{1cm} 13
  1.5.4 Managing Data Redundancy  \hspace{1cm} 14
  1.5.5 Specifying Integrity Rules  \hspace{1cm} 14
  1.5.6 Concurrency Control  \hspace{1cm} 14
  1.5.7 Backup and Recovery Facilities  \hspace{1cm} 15
  1.5.8 Data Security  \hspace{1cm} 15
  1.5.9 Performance Utilities  \hspace{1cm} 16
Summary  \hspace{1cm} 16
Key Terms List  \hspace{1cm} 16
Review Questions  \hspace{1cm} 17
Problems and Exercises  \hspace{1cm} 19

## 2  Architecture and Categorization of DBMSs

2.1 Architecture of a DBMS  \hspace{1cm} 20
  2.1.1 Connection and Security Manager  \hspace{1cm} 21
  2.1.2 DDL Compiler  \hspace{1cm} 22
  2.1.3 Query Processor  \hspace{1cm} 22
    2.1.3.1 DML Compiler  \hspace{1cm} 22
    2.1.3.2 Query Parser and Query Rewriter  \hspace{1cm} 25
    2.1.3.3 Query Optimizer  \hspace{1cm} 25
    2.1.3.4 Query Executor  \hspace{1cm} 25
  2.1.4 Storage Manager  \hspace{1cm} 25
    2.1.4.1 Transaction Manager  \hspace{1cm} 25
    2.1.4.2 Buffer Manager  \hspace{1cm} 26
    2.1.4.3 Lock Manager  \hspace{1cm} 26
    2.1.4.4 Recovery Manager  \hspace{1cm} 26
  2.1.5 DBMS Utilities  \hspace{1cm} 26
  2.1.6 DBMS Interfaces  \hspace{1cm} 27
2.2 Categorization of DBMSs  \hspace{1cm} 27
  2.2.1 Categorization Based on Data Model
    2.2.1.1 Hierarchical DBMSs  \hspace{1cm} 28
    2.2.1.2 Network DBMSs  \hspace{1cm} 28
    2.2.1.3 Relational DBMSs  \hspace{1cm} 28
    2.2.1.4 Object-Oriented DBMSs  \hspace{1cm} 28
    2.2.1.5 Object-Relational/Extended Relational DBMSs  \hspace{1cm} 29
    2.2.1.6 XML DBMSs  \hspace{1cm} 29
    2.2.1.7 NoSQL DBMSs  \hspace{1cm} 30
  2.2.2 Categorization Based on Degree of Simultaneous Access
  2.2.3 Categorization Based on Architecture
  2.2.4 Categorization Based on Usage
Summary  \hspace{1cm} 33
Key Terms List  \hspace{1cm} 33
Review Questions  \hspace{1cm} 34
Problems and Exercises  \hspace{1cm} 37
3 Conceptual Data Modeling Using the (E)ER Model and UML Class Diagram 38
3.1 Phases of Database Design 38
3.2 The Entity Relationship Model 40
3.2.1 Entity Types 40
3.2.2 Attribute Types 40
3.2.3.1 Domains 41
3.2.3.2 Key Attribute Types 42
3.2.3.3 Simple versus Composite Attribute Types 42
3.2.3.4 Single-Valued versus Multi-Valued Attribute Types 43
3.2.3.5 Derived Attribute Type 43
3.2.4 Relationship Types 43
3.2.4.1 Degree and Roles 44
3.2.4.2 Cardinalities 45
3.2.4.3 Relationship Attribute Types 46
3.2.5 Weak Entity Types 46
3.2.6 Ternary Relationship Types 48
3.2.7 Examples of the ER Model 50
3.2.8 Limitations of the ER Model 51
3.3 The Enhanced Entity Relationship (EER) Model 52
3.3.1 Specialization/Generalization 52
3.3.2 Categorization 54
3.3.3 Aggregation 55
3.3.4 Examples of the EER Model 55
3.3.5 Designing an EER Model 56
3.4 The UML Class Diagram 57
3.4.1 Recap of Object Orientation 57
3.4.2 Classes 58
3.4.3 Variables 58
3.4.4 Access Modifiers 59
3.4.5 Associations 59
3.4.5.1 Association Class 60
3.4.5.2 Unidirectional versus Bidirectional Association 60
3.4.5.3 Qualified Association 61
3.4.6 Specialization/Generalization 62
3.4.7 Aggregation 62
3.4.8 UML Example 63
3.4.9 Advanced UML Modeling Concepts 64
3.4.9.1 Changeability Property 64
3.4.9.2 Object Constraint Language (OCL) 64
3.4.9.3 Dependency Relationship 66
3.4.10 UML versus EER 66

4 Organizational Aspects of Data Management 79
4.1 Data Management 79
4.1.1 Catalogs and the Role of Metadata 80
4.1.2 Metadata Modeling 80
4.1.3 Data Quality 81
4.1.3.1 Data Quality Dimensions 82
4.1.3.2 Data Quality Problems 84
4.1.4 Data Governance 85
4.2 Roles in Data Management 86
4.2.1 Information Architect 86
4.2.2 Database Designer 87
4.2.3 Data Owner 87
4.2.4 Data Steward 87
4.2.5 Database Administrator 87
4.2.6 Data Scientist 88

Summary 88
Key Terms List 89
Review Questions 89
Problems and Exercises 90

Part II Types of Database Systems 91
5 Legacy Databases 93
5.1 The Hierarchical Model 93
5.2 The CODASYL Model 97
Summary 102
Key Terms List 102
Review Questions 102
Problems and Exercises 103

6 Relational Databases: The Relational Model 104
6.1 The Relational Model 105
6.1.1 Basic Concepts 105
6.1.2 Formal Definitions 106
6.1.3 Types of Keys 108
6.1.3.1 Superkeys and Keys 108
6.1.3.2 Candidate Keys, Primary Keys, and Alternative Keys 108
6.1.3.3 Foreign Keys 109
6.1.4 Relational Constraints 111
6.1.5 Example Relational Data Model 111
### 6.2 Normalization

- **6.2.1 Insertion, Deletion, and Update Anomalies in an Unnormalized Relational Model**
- **6.2.2 Informal Normalization Guidelines**
- **6.2.3 Functional Dependencies and Prime Attribute Type**
- **6.2.4 Normalization Forms**
  - **6.2.4.1 First Normal Form (1 NF)**
  - **6.2.4.2 Second Normal Form (2 NF)**
  - **6.2.4.3 Third Normal Form (3 NF)**
  - **6.2.4.4 Boyce–Codd Normal Form (BCNF)**
  - **6.2.4.5 Fourth Normal Form (4 NF)**

### 6.3 Mapping a Conceptual ER Model to a Relational Model

- **6.3.1 Mapping Entity Types**
- **6.3.2 Mapping Relationship Types**
  - **6.3.2.1 Mapping a Binary 1:1 Relationship Type**
  - **6.3.2.2 Mapping a Binary 1:N Relationship Type**
  - **6.3.2.3 Mapping a Binary M:N Relationship Type**
  - **6.3.2.4 Mapping Unary Relationship Types**
  - **6.3.2.5 Mapping n-ary Relationship Types**
- **6.3.3 Mapping Multi-Valued Attribute Types**
- **6.3.4 Mapping Weak Entity Types**
- **6.3.5 Putting it All Together**

### 6.4 Mapping a Conceptual EER Model to a Relational Model

- **6.4.1 Mapping an EER Specialization**
- **6.4.2 Mapping an EER Categorization**
- **6.4.3 Mapping an EER Aggregation**

### 7 Relational Databases: Structured Query Language (SQL)

- **7.1 Relational Database Management Systems and SQL**
  - **7.1.1 Key Characteristics of SQL**
  - **7.1.2 Three-Layer Database Architecture**

### 7.2 SQL Data Definition Language

- **7.2.1 Key DDL Concepts**
- **7.2.2 DDL Example**
- **7.2.3 Referential Integrity Constraints**
- **7.2.4 DROP and ALTER Command**

### 7.3 SQL Data Manipulation Language

- **7.3.1 SQL SELECT Statement**
  - **7.3.1.1 Simple Queries**
  - **7.3.1.2 Queries with Aggregate Functions**
  - **7.3.1.3 Queries with GROUP BY/HAVING**
  - **7.3.1.4 Queries with ORDER BY**
  - **7.3.1.5 Join Queries**
  - **7.3.1.6 Nested Queries**
  - **7.3.1.7 Correlated Queries**
  - **7.3.1.8 Queries with ALL/ANY**
  - **7.3.1.9 Queries with EXISTS**
  - **7.3.1.10 Queries with Subqueries in SELECT/FROM**
  - **7.3.1.11 Queries with Set Operations**
- **7.3.2 SQL INSERT Statement**
- **7.3.3 SQL DELETE Statement**
- **7.3.4 SQL UPDATE Statement**

### 7.4 SQL Views

### 7.5 SQL Indexes

### 7.6 SQL Privileges

### 7.7 SQL for Metadata Management

### 8 Object-Oriented Databases and Object Persistence

- **8.1 Recap: Basic Concepts of OO**
- **8.2 Advanced Concepts of OO**
  - **8.2.1 Method Overloading**
  - **8.2.2 Inheritance**
  - **8.2.3 Method Overriding**
  - **8.2.4 Polymorphism and Dynamic Binding**
- **8.3 Basic Principles of Object Persistence**
  - **8.3.1 Serialization**
11.1.2 The Emergence of the NoSQL Movement 302
11.2 Key–Value Stores 304
11.2.1 From Keys to Hashes 304
11.2.2 Horizontal Scaling 305
11.2.3 An Example: Memcached 306
11.2.4 Request Coordination 308
11.2.5 Consistent Hashing 309
11.2.6 Replication and Redundancy 311
11.2.7 Eventual Consistency 312
11.2.8 Stabilization 314
11.2.9 Integrity Constraints and Querying 314
11.3 Tuple and Document Stores 315
11.3.1 Items with Keys 316
11.3.2 Filters and Queries 316
11.3.3 Complex Queries and Aggregation with MapReduce 320
11.3.4 SQL After All... 330
11.4 Column-Oriented Databases 331
11.5 Graph-Based Databases 333
11.5.1 Cypher Overview 335
11.5.2 Exploring a Social Graph 336
11.6 Other NoSQL Categories 341
Summary 342
Key Terms 344
Review Questions 345
Problems and Exercises 347

Part III Physical Data Storage, Transaction Management, and Database Access 349

12 Physical File Organization and Indexing 351
12.1 Storage Hardware and Physical Database Design 351
12.1.1 The Storage Hierarchy 352
12.1.2 Internals of Hard Disk Drives 353
12.1.3 From Logical Concepts to Physical Constructs 356
12.2 Record Organization 359
12.3 File Organization 361
12.3.1 Introductory Concepts: Search Keys, Primary, and Secondary File Organization 362
12.3.2 Heap File Organization 363
12.3.3 Sequential File Organization 363
12.3.4 Random File Organization (Hashing) 365
12.3.4.1 Key-to-Address Transformation 365
12.3.4.2 Factors that Determine the Efficiency of Random File Organization 368
12.3.5 Indexed Sequential File Organization 370
12.3.5.1 Basic Terminology of Indexes 370
12.3.5.2 Primary Indexes 371
12.3.5.3 Clustered Indexes 373
12.3.5.4 Multilevel Indexes 374
12.3.6 List Data Organization (Linear and Nonlinear Lists) 375
12.3.6.1 Linear Lists 375
12.3.6.2 Tree Data Structures 377
12.3.7 Secondary Indexes and Inverted Files 379
12.3.7.1 Characteristics of Secondary Indexes 380
12.3.7.2 Inverted Files 381
12.3.7.3 Multicolumn Indexes 382
12.3.7.4 Other Index Types 383
12.3.8 B-Trees and B+-Trees 384
12.3.8.1 Multilevel Indexes Revisited 384
12.3.8.2 Binary Search Trees 385
12.3.8.3 B-Trees 386
12.3.8.4 B+-Trees 388
Summary 390
Key Terms List 391
Review Questions 392
Problems and Exercises 393

13 Physical Database Organization 395
13.1 Physical Database Organization and Database Access Methods 396
13.1.1 From Database to Tablespace 396
13.1.2 Index Design 398
13.1.3 Database Access Methods 400
13.1.3.1 Functioning of the Query Optimizer 400
13.1.3.2 Index Search (with Atomic Search Key) 402
13.1.3.3 Multiple Index and Multicolumn Index Search 403
13.1.3.4 Index-Only Access 407
13.1.3.5 Full Table Scan 408
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.3 Universal Database APIs</td>
<td>466</td>
</tr>
<tr>
<td>15.3.1 ODBC</td>
<td>466</td>
</tr>
<tr>
<td>15.3.2 OLE DB and ADO</td>
<td>467</td>
</tr>
<tr>
<td>15.3.3 ADO.NET</td>
<td>468</td>
</tr>
<tr>
<td>15.3.4 Java Database Connectivity (JDBC)</td>
<td>471</td>
</tr>
<tr>
<td>15.3.5 Intermezzo: SQL Injection and Access Security</td>
<td>477</td>
</tr>
<tr>
<td>15.3.6 SQLJ</td>
<td>479</td>
</tr>
<tr>
<td>15.3.7 Intermezzo: Embedded APIs versus Embedded DBMSs</td>
<td>480</td>
</tr>
<tr>
<td>15.3.8 Language-Integrated Querying</td>
<td>482</td>
</tr>
<tr>
<td>15.4 Object Persistence and Object-Relational Mapping APIs</td>
<td>483</td>
</tr>
<tr>
<td>15.4.1 Object Persistence with Enterprise JavaBeans</td>
<td>484</td>
</tr>
<tr>
<td>15.4.2 Object Persistence with the Java Persistence API</td>
<td>488</td>
</tr>
<tr>
<td>15.4.3 Object Persistence with Java Data Objects</td>
<td>495</td>
</tr>
<tr>
<td>15.4.4 Object Persistence in Other Host Languages</td>
<td>498</td>
</tr>
<tr>
<td>15.5 Database API Summary</td>
<td>502</td>
</tr>
<tr>
<td>15.6 Database Access in the World Wide Web</td>
<td>504</td>
</tr>
<tr>
<td>15.6.1 Introduction: the Original Web Server</td>
<td>504</td>
</tr>
<tr>
<td>15.6.2 The Common Gateway Interface: Toward Dynamic Web Pages</td>
<td>504</td>
</tr>
<tr>
<td>15.6.3 Client-Side Scripting: The Desire for a Richer Web</td>
<td>507</td>
</tr>
<tr>
<td>15.6.4 JavaScript as a Platform</td>
<td>508</td>
</tr>
<tr>
<td>15.6.5 DBMSs Adapt: REST, Other Web Services, and a Look Ahead</td>
<td>509</td>
</tr>
<tr>
<td>16 Data Distribution and Distributed Transaction Management</td>
<td>516</td>
</tr>
<tr>
<td>16.1 Distributed Systems and Distributed Databases</td>
<td>517</td>
</tr>
<tr>
<td>16.2 Architectural Implications of Distributed Databases</td>
<td>518</td>
</tr>
<tr>
<td>16.3 Fragmentation, Allocation, and Replication</td>
<td>519</td>
</tr>
<tr>
<td>16.3.1 Vertical Fragmentation</td>
<td>520</td>
</tr>
<tr>
<td>16.3.2 Horizontal Fragmentation (Sharding)</td>
<td>521</td>
</tr>
<tr>
<td>16.3.3 Mixed Fragmentation</td>
<td>521</td>
</tr>
<tr>
<td>16.3.4 Replication</td>
<td>523</td>
</tr>
<tr>
<td>16.3.5 Distribution and Replication of Metadata</td>
<td>524</td>
</tr>
<tr>
<td>16.4 Transparency</td>
<td>524</td>
</tr>
<tr>
<td>16.5 Distributed Query Processing</td>
<td>525</td>
</tr>
<tr>
<td>16.6 Distributed Transaction Management and Concurrency Control</td>
<td>528</td>
</tr>
<tr>
<td>16.6.1 Primary Site and Primary Copy 2PL</td>
<td>529</td>
</tr>
<tr>
<td>16.6.2 Distributed 2PL</td>
<td>529</td>
</tr>
<tr>
<td>16.6.3 The Two-Phase Commit Protocol (2PC)</td>
<td>530</td>
</tr>
<tr>
<td>16.6.4 Optimistic Concurrency and Loosely Coupled Systems</td>
<td>532</td>
</tr>
<tr>
<td>16.6.5 Compensation-Based Transaction Models</td>
<td>534</td>
</tr>
<tr>
<td>16.7 Eventual Consistency and BASE Transactions</td>
<td>538</td>
</tr>
<tr>
<td>16.7.1 Horizontal Fragmentation and Consistent Hashing</td>
<td>538</td>
</tr>
<tr>
<td>16.7.2 The CAP Theorem</td>
<td>539</td>
</tr>
<tr>
<td>16.7.3 BASE Transactions</td>
<td>540</td>
</tr>
<tr>
<td>16.7.4 Multi-Version Concurrency Control and Vector Clocks</td>
<td>541</td>
</tr>
<tr>
<td>16.7.5 Quorum-Based Consistency</td>
<td>542</td>
</tr>
<tr>
<td>17 Data Warehousing and Business Intelligence</td>
<td>551</td>
</tr>
<tr>
<td>17.1 Operational versus Tactical/Strategic Decision-Making</td>
<td>552</td>
</tr>
<tr>
<td>17.2 Data Warehouse Definition</td>
<td>553</td>
</tr>
<tr>
<td>17.3 Data Warehouse Schemas</td>
<td>554</td>
</tr>
<tr>
<td>17.3.1 Star Schema</td>
<td>555</td>
</tr>
<tr>
<td>16 Data Distribution and Distributed Transaction Management</td>
<td>516</td>
</tr>
<tr>
<td>Part IV Data Warehousing, Data Governance, and (Big) Data Analytics</td>
<td>549</td>
</tr>
<tr>
<td>Summary</td>
<td>512</td>
</tr>
<tr>
<td>Key Terms</td>
<td>513</td>
</tr>
<tr>
<td>Review Questions</td>
<td>513</td>
</tr>
<tr>
<td>Problems and Exercises</td>
<td>515</td>
</tr>
<tr>
<td>17 Data Warehousing and Business Intelligence</td>
<td>551</td>
</tr>
<tr>
<td>17.1 Operational versus Tactical/Strategic Decision-Making</td>
<td>552</td>
</tr>
<tr>
<td>17.2 Data Warehouse Definition</td>
<td>553</td>
</tr>
<tr>
<td>17.3 Data Warehouse Schemas</td>
<td>554</td>
</tr>
<tr>
<td>17.3.1 Star Schema</td>
<td>555</td>
</tr>
</tbody>
</table>
17.3.2 Snowflake Schema
17.3.3 Fact Constellation
17.3.4 Specific Schema Issues
  17.3.4.1 Surrogate keys
  17.3.4.2 Granularity of the Fact Table
  17.3.4.3 Factless Fact Tables
  17.3.4.4 Optimizing the Dimension Tables
  17.3.4.5 Defining Junk Dimensions
  17.3.4.6 Defining Outrigger Tables
  17.3.4.7 Slowly Changing Dimensions
  17.3.4.8 Rapidly Changing Dimensions
17.4 The Extraction, Transformation, and Loading (ETL) Process
17.5 Data Marts
17.6 Virtual Data Warehouses and Virtual Data Marts
17.7 Operational Data Store
17.8 Data Warehouses versus Data Lakes
17.9 Business Intelligence
  17.9.1 Query and Reporting
  17.9.2 Pivot Tables
  17.9.3 On-Line Analytical Processing (OLAP)
    17.9.3.1 MOLAP
    17.9.3.2 ROLAP
    17.9.3.3 HOLAP
    17.9.3.4 OLAP Operators
    17.9.3.5 OLAP Queries in SQL
17.10 Summary
17.11 Key Terms List
17.12 Review Questions
17.13 Problems and Exercises

18 Data Integration, Data Quality, and Data Governance
18.1 Data and Process Integration
  18.1.1 Convergence of Analytical and Operational Data Needs
  18.1.2 Data Integration and Data Integration Patterns
18.1.3 Data Services and Data Flows in the Context of Data and Process Integration
  18.1.3.1 Business Process Integration
  18.1.3.2 Patterns for Managing Sequence Dependencies and Data Dependencies in Processes
  18.1.3.3 A Unified View on Data and Process Integration
18.2 Searching Unstructured Data and Enterprise Search
  18.2.1 Principles of Full-Text Search
  18.2.2 Indexing Full-Text Documents
  18.2.3 Web Search Engines
  18.2.4 Enterprise Search
18.3 Data Quality and Master Data Management
18.4 Data Governance
Contents

18.4.4 Control Objectives for Information and Related Technology (COBIT) 620

18.4.5 Information Technology Infrastructure Library 621

18.5 Outlook 621

18.6 Conclusion 622

Key Terms List 622

Review Questions 623

Problems and Exercises 625

19 Big Data 626

19.1 The 5 Vs of Big Data 627

19.2 Hadoop 630

19.2.1 History of Hadoop 630

19.2.2 The Hadoop Stack 631

19.2.2.1 The Hadoop Distributed File System 631

19.2.2.2 MapReduce 635

19.2.2.3 Yet Another Resource Negotiator 641

19.3 SQL on Hadoop 643

19.3.1 HBase: The First Database on Hadoop 644

19.3.2 Pig 648

19.3.3 Hive 649

19.4 Apache Spark 652

19.4.1 Spark Core 653

19.4.2 Spark SQL 654

19.4.3 MLlib, Spark Streaming, and GraphX 656

19.5 Conclusion 659

Key Terms List 660

Review Questions 660

Problems and Exercises 662

20 Analytics 664

20.1 The Analytics Process Model 665

20.2 Example Analytics Applications 667

20.3 Data Scientist Job Profile 668

20.4 Data Pre-Processing

20.4.1 Denormalizing Data for Analysis 669

20.4.2 Sampling 670

20.4.3 Exploratory Analysis 671

20.4.4 Missing Values 671

20.4.5 Outlier Detection and Handling 672

20.5 Types of Analytics 673

20.5.1 Predictive Analytics 673

20.5.1.1 Linear Regression 673

20.5.1.2 Logistic Regression 675

20.5.1.3 Decision Trees 677

20.5.1.4 Other Predictive Analytics Techniques 681

20.5.2 Evaluating Predictive Models 682

20.5.2.1 Splitting Up the Dataset 682

20.5.2.2 Performance Measures for Classification Models 684

20.5.2.3 Performance Measures for Regression Models 687

20.5.2.4 Other Performance Measures for Predictive Analytical Models 688

20.5.3 Descriptive Analytics 689

20.5.3.1 Association Rules 689

20.5.3.2 Sequence Rules 691

20.5.3.3 Clustering 692

20.5.4 Social Network Analytics 695

20.5.4.1 Social Network Definitions 696

20.5.4.2 Social Network Metrics 696

20.5.4.3 Social Network Learning 699

20.6 Post-Processing of Analytical Models 700

20.7 Critical Success Factors for Analytical Models 701

20.8 Economic Perspective on Analytics 702

20.8.1 Total Cost of Ownership (TCO) 702

20.8.2 Return on Investment 702

20.8.3 In- versus Outsourcing 704

20.8.4 On-Premises versus Cloud Solutions 705

20.8.5 Open-Source versus Commercial Software 706
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.9 Improving the ROI of Analytics</td>
<td>708</td>
</tr>
<tr>
<td>20.9.1 New Sources of Data</td>
<td>708</td>
</tr>
<tr>
<td>20.9.2 Data Quality</td>
<td>711</td>
</tr>
<tr>
<td>20.9.3 Management Support</td>
<td>712</td>
</tr>
<tr>
<td>20.9.4 Organizational Aspects</td>
<td>712</td>
</tr>
<tr>
<td>20.9.5 Cross-Fertilization</td>
<td>713</td>
</tr>
<tr>
<td>20.10 Privacy and Security</td>
<td>714</td>
</tr>
<tr>
<td>20.10.1 Overall Considerations Regarding Privacy and Security</td>
<td>714</td>
</tr>
<tr>
<td>20.10.2 The RACI Matrix</td>
<td>715</td>
</tr>
<tr>
<td>20.10.3 Accessing Internal Data</td>
<td>716</td>
</tr>
<tr>
<td>20.10.3.1 Anonymization</td>
<td>717</td>
</tr>
<tr>
<td>20.10.3.2 SQL Views</td>
<td>719</td>
</tr>
<tr>
<td>20.10.3.3 Label-Based Access Control</td>
<td>719</td>
</tr>
<tr>
<td>20.10.4 Privacy Regulation</td>
<td>721</td>
</tr>
<tr>
<td>20.11 Conclusion</td>
<td>723</td>
</tr>
<tr>
<td>Key Terms List</td>
<td>724</td>
</tr>
<tr>
<td>Review Questions</td>
<td>725</td>
</tr>
<tr>
<td>Problems and Exercises</td>
<td>729</td>
</tr>
<tr>
<td>Appendix Using the Online Environment</td>
<td>731</td>
</tr>
<tr>
<td>Glossary</td>
<td>741</td>
</tr>
<tr>
<td>Index</td>
<td>770</td>
</tr>
</tbody>
</table>