

Contents

1	Data Management	1
1.1	Information Systems and Databases	1
1.2	SQL Databases	3
1.2.1	Relational Model	3
1.2.2	Structured Query Language (SQL)	6
1.2.3	Relational Database Management System	8
1.3	Big Data	10
1.4	NoSQL Databases	12
1.4.1	Graph-based Model	12
1.4.2	Graph Query Language Cypher	14
1.4.3	NoSQL Database Management System	16
1.5	Organization of Data Management	18
1.6	Further Reading	21
	References	22
2	Data Modeling	25
2.1	From Data Analysis to Database	25
2.2	The Entity-Relationship Model	28
2.2.1	Entities and Relationships	28
2.2.2	Association Types	29
2.2.3	Generalization and Aggregation	32
2.3	Implementation in the Relational Model	35
2.3.1	Dependencies and Normal Forms	35
2.3.2	Mapping Rules for Relational Databases	46
2.3.3	Structural Integrity Constraints	54
2.4	Implementation in the Graph Model	57
2.4.1	Graph Properties	57
2.4.2	Mapping Rules for Graph Databases	68
2.4.3	Structural Integrity Constraints	75
2.5	Enterprise-Wide Data Architecture	76

2.6	Formula for Database Design	79
2.7	Further Reading	81
	References	82
3	Database Languages	85
3.1	Interacting with Databases	85
3.2	Relational Algebra	87
3.2.1	Overview of Operators	87
3.2.2	Set Operators	89
3.2.3	Relational Operators	91
3.3	Relationally Complete Languages	95
3.3.1	SQL	96
3.3.2	QBE	99
3.4	Graph-based Languages	101
3.4.1	Cypher	103
3.5	Embedded Languages	107
3.5.1	Cursor Concept	108
3.5.2	Stored Procedures and Stored Functions	108
3.5.3	JDBC	109
3.5.4	Embedding Graph-based Languages	110
3.6	Handling NULL Values	111
3.7	Integrity Constraints	113
3.8	Data Protection Issues	116
3.9	Further Reading	120
	References	121
4	Ensuring Data Consistency	123
4.1	Multi-User Operation	123
4.2	Transaction Concept	124
4.2.1	ACID	124
4.2.2	Serializability	125
4.2.3	Pessimistic Methods	128
4.2.4	Optimistic Methods	131
4.2.5	Troubleshooting	133
4.3	Consistency in Massive Distributed Data	134
4.3.1	BASE and the CAP Theorem	134
4.3.2	Nuanced Consistency Settings	136
4.3.3	Vector Clocks for the Serialization of Distributed Events	137
4.4	Comparing ACID and BASE	139
4.5	Further Reading	141
	References	141

5	System Architecture	143
5.1	Processing of Homogeneous and Heterogeneous Data	143
5.2	Storage and Access Structures	146
5.2.1	Indexes and Tree Structures	146
5.2.2	Hashing Methods	149
5.2.3	Consistent Hashing	150
5.2.4	Multidimensional Data Structures	152
5.3	Translation and Optimization of Relational Queries	155
5.3.1	Creation of Query Trees	155
5.3.2	Optimization by Algebraic Transformation	156
5.3.3	Calculation of Join Operators	158
5.4	Parallel Processing with MapReduce	161
5.5	Layered Architecture	162
5.6	Use of Different Storage Structures	164
5.7	Further Reading	166
	References	167
6	Postrelational Databases	169
6.1	The Limits of SQL—and Beyond	169
6.2	Federated Databases	170
6.3	Temporal Databases	173
6.4	Multidimensional Databases	176
6.5	Data Warehouse	180
6.6	Object-Relational Databases	183
6.7	Knowledge Databases	187
6.8	Fuzzy Databases	190
6.9	Further Reading	195
	References	197
7	NoSQL Databases	201
7.1	Development of Nonrelational Technologies	201
7.2	Key-Value Stores	202
7.3	Column-Family Stores	205
7.4	Document Stores	207
7.5	XML Databases	210
7.6	Graph Databases	215
7.7	Further Reading	217
	References	218
	Glossary	219
	References	223
	Index	225