

Contents

- 1 Data Cloudification** 1
 - 1.1 Big Data, Big Problems..... 1
 - 1.2 Classic Database Solutions as an Alternative..... 4
 - 1.3 Cloud Computing 7
 - 1.4 Data Management and Cloud Computing 9
 - 1.5 Summary 11
 - References 11
- 2 Virtualization for Data Management Services**..... 13
 - 2.1 Core Concepts of Virtualization 13
 - 2.1.1 Limits of Virtualization..... 16
 - 2.1.2 Use Cases for Virtualization..... 16
 - 2.1.3 Summary 20
 - 2.2 Virtualization Stack 20
 - 2.3 Hardware Virtualization 24
 - 2.3.1 Machine Virtualization 26
 - 2.3.2 Virtual Storage 28
 - 2.4 Schema Virtualization 34
 - 2.4.1 Sharing 37
 - 2.4.2 Customization..... 40
 - 2.5 Configuring the Virtualization 51
 - 2.6 Summary 56
 - References 56
- 3 Transactional Data Management Services for the Cloud**..... 59
 - 3.1 Foundations 59
 - 3.1.1 Transaction Processing 59
 - 3.1.2 Brewer’s CAP Theorem and the BASE Principle 66
 - 3.2 Consistency Models..... 68
 - 3.3 Data Fragmentation 69
 - 3.3.1 Classification of Fragmentation Strategies 70

3.3.2	Horizontal Versus Vertical Fragmentation	70
3.3.3	Consistent Hashing	72
3.4	Data Replication	74
3.4.1	Classification of Replication Strategies	74
3.4.2	Classic Replication Protocols	76
3.4.3	Optimistic Replication	78
3.4.4	Conflict Detection and Repair	79
3.4.5	Epidemic Propagation	81
3.5	Consensus Protocols	83
3.6	Summary	88
	References	89
4	Web-Scale Analytics for BIG Data	91
4.1	Foundations	91
4.1.1	Query Processing	92
4.1.2	Parallel Data Processing	96
4.1.3	Fault Tolerance	102
4.2	The MapReduce Paradigm	103
4.2.1	The MapReduce Programming Model	104
4.2.2	The MapReduce Execution Model	108
4.2.3	Expressing Relational Operators in MapReduce	114
4.2.4	Expressing Complex Operators in MapReduce	118
4.2.5	MapReduce Shortcomings and Enhancements	121
4.2.6	Alternative Model: DAG Dataflows	123
4.3	Declarative Query Languages	124
4.3.1	Pig	124
4.3.2	JAQL	126
4.3.3	Hive	129
4.4	Summary	131
	References	133
5	Cloud-Specific Services for Data Management	137
5.1	Service Level Agreements	137
5.1.1	The Notion of QoS and SLA	138
5.1.2	QoS in Data Management	138
5.1.3	Workload Management	140
5.1.4	Resource Provisioning	142
5.2	Pricing Models for Cloud Systems	145
5.2.1	Cost Types	146
5.2.2	Subscription Types	146
5.3	Security in Outsourced Data Management	148
5.3.1	Secure Joins	149
5.3.2	Secure Aggregation and Selection	152
5.3.3	Secure Range Queries	155
5.4	Summary	158
	References	159

6 Overview and Comparison of Current Database-as-a-Service Systems	161
6.1 Infrastructure-as-a-Service	161
6.1.1 Amazon Elastic Compute Cloud	162
6.1.2 Further IaaS Solutions	165
6.2 Platform-as-a-Service	166
6.2.1 Amazon S3, DynamoDB, and RDS	166
6.2.2 Amazon Elastic MapReduce	170
6.2.3 Microsoft Windows Azure	175
6.3 Software- and Data-as-a-Service	179
6.3.1 The Open Data Trend	180
6.3.2 Data Markets and Pricing models	184
6.3.3 Beyond “Data-as-a-Service”	185
6.4 Summary	186
References	187
7 Summary and Outlook	189
7.1 Outlook, Current Trends, and Some Challenges	190
7.2 ... the End	193
References	193

Web-Scale Data Management for the Cloud

Lehner, W.; Sattler, K.-U.

2013, XV, 193 p., Hardcover

ISBN: 978-1-4614-6855-4