Table of Contents

Introduction xi

Chapter 1: Introduction to Databases 1

1.1 Database Concepts 1

Data versus Information 2

Record 3

File System 3

Limitations of File-Based System 4

Evolution of Database 5

Defining a Database 7

Entities and Relationships 7

1.2 Database Models and Types 8

Object-Based Model 9

Record-Based Logical Model 13

1.3 Concept of Normalization 16

Objectives of Functional Dependency and Normalization 16

Designing a Database Schema 17

Functional Dependencies 19

Normal Forms 20

1.4 Introduction to SQL Using MySQL 28

Defining SQL 29

Data Definition Language 30

Data Manipulation Language 36

The HAVING Clause 64

Data Control Language 88

Working with Views 91

Summary 94

Exercise 94

Multiple Choice Questions 94

Subjective Type Questions 96

Chapter 2: Advanced Database Techniques 97

2.1 Structured versus Unstructured Data 97

2.2 NoSQL Database Concepts 98

2.3 Types of NoSQL Databases 99

2.4 NoSQL Data Modeling 99

2.5 Benefits of NoSQL 101

2.6 Comparative Study of SQL and NoSQL Database Systems 103

2.7 NoSQL Using MongoDB 104

Introduction to MongoDB Shell 104

Running the MongoDB Shell 105

MongoDB Client 106

Basic Operations with the MongoDB Shell 106

Data Types 109

Basic Data Types 109

Dates 111

Arrays 112

Embedded Documents 112

_id and ObjectIds 113

2.8 Querying 114

Introduction to the find() Function 114

Specifying which Keys to Return 115

Query Criteria 116

OR Queries 117

Type-Specific Querying 118

2.9 Index Introduction 120

Types of Indexes in MongoDB 122

Indexing Properties 125

2.10 Aggregation Introduction 125

Aggregation Pipeline 125

Aggregation Using MapReduce 126

Single-Purpose Aggregation 127

Summary 128
Exercise 128
Multiple Choice Questions 128
Subjective Type Questions 130

Chapter 3: Database Transactions 131

3.1 Transaction Management in Database Systems 132
Atomicity and Durability of Transactions 135
Isolation of Transactions 137
Transaction Isolation and Atomicity 140
3.2 Concurrency Control in Database Systems 142
Lock-Based Concurrency Control Protocols 142
Snapshot Isolation 147
3.3 System Performance Tuning and Query Optimization in SQL Databases 149
Tuning SQL Database System Performance 149
Optimizing SQL Database System Queries 156
3.4 System Performance Tuning and Query Optimization in NoSQL Databases ... 161
Summary 162
Exercise 162
Multiple Choice Questions 162
Subjective Type Questions 164

Chapter 4: Database Architecture 165

4.1 Introduction to Client-Server Database Model 166
Two-Tier Client-Server Model 166
Three-Tier Client-Server Model 167
4.2 Introduction to Parallel Databases 169
Parallel Database System Architecture 170
Types of Parallelism 173
Parallel Database Implementation 173
4.3 Introduction to Distributed Databases 174
Benefits of Distributed Database Systems 175

Issues with Distributed Database Systems 178
4.4 Database Connectivity and Web Technologies 179
Two-Tier Architecture for Web Connectivity 180
Three-Tier Architecture for Web Connectivity 180
4.5 Database Administration and Management 181
Need for Database Administration and Management 181
Database Technology 181
Database Administration and Management Activities 182
4.6 Connectivity Using MongoDB and Cassandra 183
Connectivity Using MongoDB 183
Connectivity Using Cassandra 184
Summary 184
Exercise 185
Multiple Choice Questions 185
Subjective Type Questions 186

Chapter 5: Big Data Management and Programming 187
5.1 XML 188
XML Introduction 188
XML Entity References 193
Namespaces 193
XML DTDs 196
Domain-Specific DTDs 200
Querying XML Data 201
5.2 JSON 210
Data Types 210
Object Literals 211
Array Literals 212
JSON Parser 213
JSON Syntax 213
5.3 Big Data 218
5.4 Introduction to Hadoop 219

Building Blocks of Hadoop 220
Components of Hadoop....221
5.5 HBase....222
5.6 HIVE....223
5.7 Solid-State Drive....223
5.8 Cloudera 224
5.9 Oracle Cloud....225
5.10 Oracle Berkley Database (Oracle BDB)....226
5.11 MongoDB....227
5.12 Introduction to R Programming 229
Summary 229
Exercise....230
Multiple Choice Questions 230
Subjective Type Questions....231

Chapter 6: Advances in Databases Data 233

6.1 Introduction to Data Warehouse 234
Characteristics of Data Warehousing....234
Functionality of Data Warehousing 235
Architecture of Data Warehousing 235
Types of Data Warehouse Architecture 236
Data Marts 237
Data Warehousing Life Cycle....238
Data Modeling (Multidimensional Database) for DataWarehousing....239
Building of Data Warehouse 241
Data Warehouse Development....241
OLAP versus OLTP [Data Warehouses versus Operational Databases] 244
6.2 Introduction to Data Mining Techniques....245
Data Mining Technology and Its Relation to DataWarehousing....245
Association Rules....246
Classification and Clustering....247
Addressing Data Mining Problems and TheirApplications....249

Applications of Data Mining....251
Commercial Tools Used for Data Mining....252
6.3 Introduction to Business Intelligence....253
Features of BI....254
BI Component Framework....256
Types of Machine Learning....261
Approaches for Machine Learning....262
Summary....262
Exercise....263
Multiple Choice Questions....263
Subjective Type Questions....264

Index....265