Table of Contents

Author Biography iii

Introduction v

Chapter 1: Introduction to DBMS 1

1.1 Introduction 1

1.2 Data versus Information 2

1.3 Record 3

1.4 The File System 3

Limitations of File-Based System 4

1.5 The Evolution of Database 5

1.6 Defining a Database 7

1.7 Entities and Relationships 8

1.8 Data Models 9

1.9 Database Design 9

1.10 Types of Databases 10

Distributed Database 10

Object-Oriented Database 11

Temporal Database 12

Multimedia Database 12

Deductive Database 13

Semantic Database 13

Mobile Database 13

1.11 Transaction Management 14

Database Recovery and Security 15

1.12 Introduction to SQL 16

Defining SQL 16

Data Definition Language 17

Data Manipulation Language 19

Summary 27

Exercise 28

Multiple Choice Questions 28

Subjective Questions 30

Chapter 2: Advanced SQL 31

2.1 Introduction 31
2.2 Constraints 32
SQL NOT NULL Constraint 32
SQL UNIQUE Constraint 33
SQL PRIMARY KEY Constraint 33
SQL Foreign Key Constraint 33
SQL CHECK Constraint 34
SQL DEFAULT Constraint 34
2.3 SQL CREATE INDEX 35
Creation of Indexes 35
Dropping an Index 35
2.4 SQL Functions 36
SQL Aggregate Functions 36
2.5 The GROUP BY Statement 41
2.6 The HAVING Clause 43
2.7 SQL Special Functions 44
2.8 SQL Alias 51
2.9 SQL Join 51
SQL Inner Join 54
SQL Natural Join 56
SQL Equi Join 57
SQL Cross Join 58
SQL Left Outer Join 59
SQL Right Outer Join 60
SQL Full Outer Join 62
The SQL UNION Operator 63
2.10 Sub-Queries 65
2.11 Recursive Queries 66
2.12 Data Control Language 66
SQL GRANT Command 67
SQL REVOKE Command 67
Privileges and Roles 68
Creating and Dropping Roles 69
2.13 Views and Assertion 70
Views 70
Assertion 71

2.14 PL/SQL – A Basic Introduction ... 72
Need for PL/SQL ... 72
PL/SQL Code Structure ... 72
2.15 Triggers ... 74
Types of PL/SQL Triggers ... 76
Delete Trigger ... 77
2.16 Event Condition Action (ECA) Model ... 77
2.17 Functions and Procedures ... 77
Statements in Procedures ... 78
Procedures ... 78
Executing a Stored Procedure ... 79
PL/SQL Compilation Errors ... 80
PL/SQL Compilation Errors Executing SHOW ERRORS ... 80
Dependencies and Procedures ... 81
Uses of Procedures ... 81
Functions ... 81
2.18 Embedded SQL and Dynamic SQL ... 82
2.19 The JAVA Way to Access RDBMS: JDBC ... 85
JDBC Drivers and Connections ... 85
Creating JDBC Statements ... 86
Creating JDBC PreparedStatement ... 87
JDBC Query and Retrieving Answers ... 87
JDBC Update ... 88
Transaction and Connections ... 88
2.20 SQLJ ... 89
SQLJ and Java ... 89
Embedded SQLJ ... 89
SQLJ Translator ... 89
SQLJ Runtime Environment ... 90
SQLJ Profiles ... 90
Developing Applications Using SQLJ ... 90
Benefits of SQLJ ... 90
Codes ... 91
Summary ... 94
Exercise ... 94
Multiple-Choice Questions ... 94
Subjective Questions ... 96

Chapter 3: Advanced Transaction Processing and Recovery 97

3.1 Introduction 97
3.2 Defining a Transaction in DBMS 98
ACID Property 99
State of Transaction 100
3.3 Defining a Concurrent Transaction in DBMS 100
Schedule in Transaction 101
Problems in Concurrent Transaction 103
3.4 Serializability and Recoverability 106
Defining Serializability 106
Conflict Serializability 108
View Serializability 110
Precedence Graph and Serializability 110
Defining Recoverability 112
Cascading Rollback 113
3.5 Enhanced Lock-Based and Timestamp-Based Protocols 114
Lock-Based Protocol 114
Two-Phase Locking Protocol 117
Timestamp-Based Protocol 120
Defining Timestamp 120
3.6 Multiple Granularity 122
Multiple-Granularity Locking Protocol 123
3.7 Multi-Version Schemes 126
Multi-Version Timestamp-Based Protocol 126
Multi-Version 2PL Protocol 127
3.8 Optimistic Concurrency Control Techniques 128
3.9 Deadlock Handling 130
Causes of Deadlock 131
Deadlock Prevention 132
Deadlock Detection 133
3.10 Recovery in DBMS 135
Recoverability 135
Types of Failure 136
Methods to Control Failures 136
Different Techniques of Recoverability 137
3.11 Write-Ahead Logging Protocol 142

3.12 Advanced Recovery Techniques 142
Shadow Paging 142
Fuzzy Checkpoints 143
ARIES 144
3.13 Use of SQL in Recovery 145
3.14 RAID 147
RAID 0 147
RAID 1 148
RAID 3 149
RAID 4 150
RAID 5 150
RAID 6 151
Summary 152
Exercise 152
Multiple Choice Questions 152
Subjective Questions 154

Chapter 4: Data Security 155

4.1 Introduction 155
4.2 Introduction to Data Security Issues 155
Excessive Privilege Abuse 156
Legitimate Privilege Abuse 157
Privilege Elevation 157
Database Platform Vulnerabilities 157
SQL Injection 157
Weak Audit Trail 157
Denial of Service 157
Database Communication Protocol Vulnerabilities 158
Weak Authentication 158
Backup Data Exposure 158
4.3 Discretionary Access Control 158
Request Modification 160
4.4 Mandatory Access Control 160
Bell-La Padula Model 161
Multilevel Secure Database Management Systems 162
4.5 Role-Based Access Control 162
4.6 SQL Injection 163

SQL Manipulation 163
Code Injection 164
Function Call Injection 164
4.7 Statistical Databases 165
Security in Statistical Databases 165
4.8 Introduction to Flow Control 165
Summary 166
Exercise 166
Multiple Choice Questions 166
Subjective Questions 168

Chapter 5: Storage and Indexing 169
5.1 Introduction 169
5.2 Basics of Database Storage 169
Records and Record Types 170
File Organization 170
5.3 Operations on Files 170
5.4 Hashing Techniques 172
Internal Hashing 173
External Hashing 174
Hashing for Dynamic File Expansion 174
5.5 File Indexing 175
5.6 Types of Single-Level Ordered Indexes 176
Primary Indexes 176
Clustering Indexes 178
Secondary Indexes 179
5.7 Multi-Level Indexes 180
5.8 Dynamic Multi-Level Indexes 181
Difference between B-tree and B+-tree 181
B-Tree Structure 181
B+-Tree Structure 182
Insertion in a B+-Tree 183
Deletion from a B+-Tree 183
5.9 Indexing on Multiple Keys 185
Summary 186

Exercise 186
Multiple Choice Questions 186
Subjective Questions 188

Chapter 6: Distributed Databases 189

6.1 Introduction 189
6.2 Differences between Centralized DBMS and Distributed DBMS 190
6.3 Functionality of DDBMS 191
6.4 Advantages and Disadvantages of DDBMS 191
6.5 Centralized versus Non-Centralized Databases 192
6.6 Homogeneous and Heterogeneous DDBMS and Their Comparison 193
6.7 Architecture of DDBMS 194
Client-Server DDBMS Architecture 194
Peer-to-Peer Architecture 196
6.8 Distributed Data Storage 197
Replication of Data 197
Fragmentation of Data 197
Hybridization (Combination of Both Replication and Fragmentation) of Data 201
6.9 Transparency Issue in DDBMS 201
Network Transparency 202
Replication Transparency 202
Fragmentation Transparency 202
6.10 Distributed Database Design and Query Processing in DDBMS 203
6.11 Distributed Transaction Management 205
6.12 Distributed Commit Protocols: 2PC and 3PC 206
Two-Phase Commit Protocol 207
Three-Phase Commit Protocol 208
Difference between 2PC and 3PC 209
6.13 Distributed Concurrency Control 210
Single-Lock Manager Approach 210
Distributed Lock Manager Approach 211
6.14 Distributed Deadlock Management 212
Distributed Deadlock Detection 213
Distributed Deadlock Prevention 215
6.15 Concepts of Replication Servers 215

Summary 216
Exercise 216
Multiple Choice Questions 216
Subjective Questions 218

Chapter 7: Object Database 219
7.1 Introduction 219
7.2 Overview of the Object-Oriented Database 220
7.3 Limitations of Relational Database 223
7.4 Need for Complex Data Type 226
Collection and Large Object Types 226
7.5 Collection Types and Structured Types 228
7.6 Data Definition 229
Objects 229
Object Classes 229
Abstraction 229
Inheritance 230
7.7 Object Identity and Reference Types 234
Object Identifiers 234
Object Containment 234
Defining a Schema 235
7.8 Persistent Programming Languages 236
Persistence of Objects 237
7.9 Object – Relational Database 237
Performance Constraints 238
ORDBMS Benefits 238
Object-Relational DBMS Applications 240
ORDBMS Concepts and Terminology 241
7.10 Object Database Extensions to SQL 241
7.11 Object Definition Language 244
Abstraction 245
Encapsulation 245
Inheritance 245
7.12 Object Query Language 245
Query Input and Result 246

Summary 246
Exercise 246
Multiple Choice Questions 246
Subjective Questions 248

Chapter 8: Introduction to Data Warehousing 249
8.1 Introduction 249
8.2 The Need for Data Warehousing 250
8.3 Data Warehousing: An Information Environment 250
Benefits of Data Warehousing 251
Features of Data Warehouse 251
8.4 Increased Demand for Strategic Information 252
8.5 Inability of Past Decision Support System 253
8.6 Operational vs. Decisional Support System 254
8.7 Information Flow Mechanism 255
Key Components 256
8.8 Classification of Metadata 256
8.9 Data Warehouse and Data Marts 256
8.10 Data Warehouse Architecture 258
8.11 Different Types of Architecture 259
Data Consistency Architecture 259
Reporting Data Store and Staging Data Store Architecture 260
Data Modeling Architecture 260
Tool Architecture 260
Processing Tiers Architecture 260
Security Architecture 260
8.12 Data Warehouse Design Techniques 260
Bottom-up Design 260
Top-down Design 261
Summary 261
Exercise 262
Multiple Choice Questions 262
Subjective Questions 264

Chapter 9: Introduction to Dimensional Modeling 265
9.1 Introduction 265
9.2 Data Warehouse Modeling vs. Operational Database Modeling 266
9.3 Dimensional Model vs. the ER Model 267
9.4 Features of a Good Dimensional Model 268
9.5 Introducing the Fact Table 269
9.6 Introducing the Dimension Table 269
9.7 The Factless Face Table 270
Factless Fact Tables for Events 271
Factless Fact Tables for Conditions 271
9.8 The Star Schema 271
Query Execution in Star Schema 272
9.9 Snowflake Schema 273
9.10 Updates to Dimension Tables 274
9.11 Slowly Changing Dimensions 275
Slowly Changing Dimension Type 1 275
Slowly Changing Dimension Type 2 275
Slowly Changing Dimension Type 3 275
9.12 Large Dimension Tables 276
9.13 Rapidly Changing or Large Slowly Changing Dimensions 277
9.14 Junk Dimensions 278
9.15 Keys in the Data Warehouse Schema 278
Primary Key 279
Surrogate Key 279
Foreign Keys 281
9.16 Aggregate Table 281
9.17 Fact Constellation Schema or Family of Stars 282
Summary 282
Exercise 283
Multiple Choice Questions 283
Subjective Questions 284

Chapter 10: ETL Process 285
10.1 Introduction 285
10.2 Overview of the ETL Process 286

10.3 Need for the ETL Process 286
10.4 Data Extraction 287
10.5 Identification of Data Source 287
10.6 Extraction Methods in Data Warehouse 288
Logical Extraction Methods 289
Physical Extraction Methods 291
10.7 Change Data Capture 292
10.8 Transportation in Data Warehouses 294
Transportation Mechanisms in Data Warehouses 294
10.9 Transformation 295
Transformation Flow 295
Transformation Mechanisms 297
10.10 Staging 300
10.11 Loading 301
Loading a Data Warehouse with SQL*Loader 301
Loading a Data Warehouse with External Tables 302
Loading Data Warehouse with OCI and Direct-Path APIs 303
Loading Data Warehouse with Export/Import 303
10.12 Fact Tables 303
Aggregation in Fact Tables 303
10.13 Dimension Tables 304
Surrogate Keys 305
Referential Integrity 305
Shared Dimensions 306
10.14 Cleaning 306
Need for Data Cleaning and Transformation 306
10.15 Tools for ETL 307
ETL Tools for Data Warehouses 307
Daily Operations in Data Warehouses 308
Evolution of the Data Warehouse 308
Summary 308
Exercise 308
Multiple Choice Questions 308
Subjective Questions 310

Chapter 11: Online Analytical Processing (OLAP) 311
11.1 Introduction 311
11.2 Introduction to OLAP 311
11.3 Need for OLAP 312
11.4 OLAP Models 312
11.5 OLTP Vs OLAP 313
11.6 Multidimentional Analysis or Data Cube 314
11.7 On-Line Analytical Processing (OLAP) 317
Summary 318
Exercises 318
Multiple Choice Questions 318
Subjective Questions 320
ADBMS Lab Implementation 321
Index 331