

INTRODUCTION TO DATABASE MANAGEMENT SYSTEM

By

Dr. Satinder Bal Gupta

*B.Tech., (CSE), MCA, UGC-NET, Ph.D (CS)
Prof., Deptt. of Computer Sci. & App.
Vaish College of Engineering, Rohtak,
Haryana*

Aditya Mittal

*B.E. (Computer Science and Engg.),
Software Engineer (SAP-ABAP)
IBM India Pvt. Ltd.
Noida, (U.P.)*



UNIVERSITY SCIENCE PRESS

(An Imprint of Laxmi Publications Pvt. Ltd.)

An ISO 9001:2008 Company

**BENGALURU • CHENNAI • COCHIN • GUWAHATI • HYDERABAD
JALANDHAR • KOLKATA • LUCKNOW • MUMBAI • RANCHI • NEW DELHI
BOSTON (USA) • ACCRA (GHANA) • NAIROBI (KENYA)**

CONTENTS

<i>Preface</i>	(xi)
1. Introduction to Database Systems	1-43
1.1 Introduction	1
1.2 Basic Definitions and Concepts	1
1.2.1 Data	1
1.2.2 Information	2
1.2.3 Meta data	2
1.2.4 Data dictionary	3
1.2.5 Database	3
1.2.6 Components of a database	4
1.2.7 Database management system (DBMS)	4
1.2.8 Components of DBMS	5
1.3 Traditional File System Versus Database Systems	6
1.3.1 Disadvantages of traditional file system	7
1.3.2 Database systems or database system environment	8
1.3.3 Advantages of database systems (DBMS's)	9
1.3.4 Disadvantages of database systems	11
1.4 DBMS Users	11
1.4.1 End users or naive users	12
1.4.2 Online users	12
1.4.3 Application programmers	12
1.4.4 Database administrator	12
1.5 Database or DBMS Languages	13
1.5.1 Data definition language (DDL)	13
1.5.2 Storage definition language (SDL)	13
1.5.3 View definition language (VDL)	13
1.5.4 Data manipulation language (DML)	13
1.5.5 Fourth-generation language (4-GL)	14
1.6 Schemas, Subschema and Instances	14

1.6.1	Schema	14
1.6.2	Subschema	15
1.6.3	Instances	15
1.7	Three Level Architecture of Database Systems (DBMS)	16
1.7.1	Levels or views	16
1.7.2	Different mappings in three level architecture of DBMS	17
1.7.3	Advantages of three-level architecture	18
1.7.4	Data independence	18
1.8	Data Models	18
1.8.1	Types of data models	19
1.8.2	Comparison of various data models	24
1.8.3	Which data models to use?	25
1.9	Types of Database Systems	25
1.9.1	According to the number of users	26
1.9.2	According to the type of use	27
1.9.3	According to database site locations	27
1.10	Comparison between Client/Server and Distributed Database System	32
	<i>Test Your Knowledge</i>	32
	<i>Exercises</i>	41
2.	E-R and EER Models	44–91
2.1	Introduction	44
2.2	Basic Concepts	45
2.3	Types of Attributes	46
2.3.1	Simple and composite attributes	46
2.3.2	Single valued and multi-valued attributes	46
2.3.3	Derived attributes and stored attributes	46
2.4	Relationship Sets	47
2.4.1	Degree of relationship sets	47
2.4.2	Role and recursive relationship set	48
2.5	Mapping Constraints	48
2.5.1	Mapping cardinalities (cardinality ratios)	48
2.5.2	Participation constraints	50
2.6	Keys	50
2.6.1	Types of keys	50
2.7	Entity—Relationship Diagram	51
2.7.1	Advantages of E-R model	53
2.7.2	Limitation of E-R model	54
2.8	Types of Entity Sets	54
2.8.1	Strong entity sets	54

2.8.2 Weak entity sets	54
2.9 Enhanced Entity-Relationship (EER) Model	55
2.9.1 Superclass and subclass entity types	55
2.9.2 Specialization	57
2.9.3 Generalization	57
2.9.4 Attribute inheritance	57
2.9.5 Aggregation	57
2.9.6 Specialization and generalization constraints	58
2.9.7 Categorization	59
2.10 Reduction of an E-R and EER Diagram into Tables	60
<i>Test Your Knowledge</i>	78
<i>Exercises</i>	87
3. File Organization	92–130
3.1 Introduction	92
3.2 Basic Concepts of Files	92
3.2.1 Records and record types	92
3.2.2 Types of files	94
3.3 File Organization Techniques	94
3.3.1 Heap file organization	95
3.3.2 Sequential file organization	95
3.3.3 Index sequential file organization	99
3.3.4 Hashing	101
3.3.5 Direct file organization	105
3.4 Indexing	106
3.4.1 Ordered indexing	106
3.4.2 Hashed indexing	110
3.5 B-tree Index Files	110
3.6 B ⁺ -Tree Index Files	111
3.7 Comparison of Different File Organizations	113
3.8 Factors Affecting Choice of File Organization	113
<i>Test Your Knowledge</i>	117
<i>Exercises</i>	126
4. Data Models	131–161
4.1 Introduction	131
4.2 Hierarchical Model	132
4.2.1 Tree structure diagrams	132
4.2.2 Operations on hierarchical data model	135
4.2.3 Query language for hierarchical databases	135
4.2.4 Virtual records	137

4.2.5 Implementation of tree structure diagram	138
4.3 Network Model	139
4.3.1 Graph structure diagrams	139
4.3.2 Operations on network data model	143
4.3.3 Query language for network databases	143
4.3.4 Implementation of graph structure diagram	144
4.4 Relational Model	146
4.4.1 Definition of relation	146
4.4.2 Data structure of relational database	146
4.4.3 Integrity constraints	147
4.5 CODD's Rules	149
4.5.1 Operations on relational model	151
4.6 Comparison of DBMS and RDBMS	153
4.7 Comparison of Data Models	153
<i>Test Your Knowledge</i>	155
<i>Exercises</i>	160
5. Relational Algebra and Calculus	162–192
5.1 Introduction	162
5.2 Relational algebra	162
5.2.1 Operations in relational algebra	163
5.3 Relational Calculus	172
5.3.1 Tuple relational calculus	173
5.3.2 Domain relational calculus	175
5.4 Comparison of Domain Relational Calculus and Tuple Relational Calculus	177
5.5 Comparison of Relational Calculus and Relational Algebra	177
<i>Test Your Knowledge</i>	183
<i>Exercises</i>	190
6. Functional Dependency and Normalisation	193–272
6.1 Introduction	193
6.2 Informal Design Guidelines for Relation Schemas	193
6.3 Functional Dependencies	195
6.3.1 Functional dependency chart/diagram	195
6.3.2 Types of functional dependencies	196
6.4 Anomalies in Relational Database	199
6.5 Dependencies and Logical Implications	200
6.5.1 Armstrong's axioms	200
6.6 Closure of a Set of Functional Dependencies	202
6.6.1 Closure of an attribute w.r.t. the set of FD's F	203
6.7 Covers	204

6.7.1	Types of cover	205
6.7.2	Identifying redundant functional dependencies	206
6.7.3	Covers and equivalent sets of FDs	208
6.7.4	Minimal set of functional dependencies	209
6.7.5	Finding a minimal cover for a set of FDs	209
6.8	Keys and Functional Dependencies	211
6.9	Decompositions	214
6.9.1	Properties of a good decomposition	214
6.10	Normalisation	215
6.10.1	Benefits of normalisation	215
6.10.2	Various normal forms	216
6.11	Denormalisation	228
6.11.1	The need of denormalization	229
6.11.2	Issues to be considered when deciding denormalization	229
6.11.3	Advantages of denormalization	230
6.11.4	Disadvantages of denormalization	230
	<i>Test Your Knowledge</i>	250
	<i>Exercises</i>	262
7.	Query Languages	273–347
7.1	Introduction	273
7.2	Structured Query Language (SQL)	273
7.2.1	Characteristics of SQL	273
7.2.2	Advantages of SQL	274
7.2.3	Parts (Components) of SQL language	274
7.2.4	Basic data types	276
7.2.5	Data manipulation in SQL	276
7.2.6	Data definition language (DDL)	303
7.2.7	Data control language (DCL)	308
7.3	Query By Example (QBE)	311
7.3.1	QBE dictionary	311
7.3.2	Data types	312
7.3.3	Data definition functions	312
7.3.4	Update operations	314
7.3.5	Data retrieval operations	316
7.3.6	Built-in functions	320
7.3.7	Features of QBE	321
7.3.8	Advantages of QBE	321
7.3.9	Limitations of QBE	321
7.3.10	Commercial database management systems providing QBE feature	321

7.4	Comparison of SQL and QBE	321
	<i>Test Your Knowledge</i>	334
	<i>Exercises</i>	343
8.	Transactions and Concurrency Control	348–393
8.1	Introduction	348
8.2	Transaction	348
	8.2.1 ACID properties of transaction	348
	8.2.2 Transaction states	349
8.3	Some Definitions	349
8.4	Why Concurrency Control is Needed?	350
8.5	Concurrency Control Techniques	352
	8.5.1 Lock-based protocols	352
	8.5.2 Graph based protocols	358
	8.5.3 Timestamp-based protocols	360
	8.5.4 Validation based protocols	361
	8.5.5 Multiversion concurrency control protocols	363
8.6	Deadlocks	364
	8.6.1 Necessary conditions for deadlock	364
	8.6.2 Methods for handling deadlocks	364
	<i>Test Your Knowledge</i>	384
	<i>Exercises</i>	389
9.	Database Security and Authorization	394–403
9.1	Introduction	394
9.2	Security Violations	394
9.3	Authorization (Access Rights)	395
9.4	Views	396
9.5	Granting of Privileges	396
9.6	Notion of Roles (Grouping of Users)	397
9.7	Audit Trails	398
	9.7.1 Advantages of audit trails	398
	9.7.2 Audit trails and logs	399
	9.7.3 Review of audit trails	399
	<i>Test Your Knowledge</i>	400
	<i>Exercises</i>	402
10.	Database Recovery System	404–418
10.1	Introduction	404
10.2	Classification of Failures	404
10.3	Recovery Concept	405

10.3.1 Log based recovery	405
10.4 Shadow Paging	408
<i>Test Your Knowledge</i>	412
<i>Exercises</i>	416
11. Query Processing and Optimization	419–447
11.1 Introduction	419
11.2 Basics of Query Processing	419
11.2.1 General strategy for query processing	420
11.2.2 Steps in query processing	420
11.3 Query Optimization	425
11.3.1 Transformation rules for relational algebra	426
11.3.2 Heuristic query optimization	428
11.3.3 Cost based query optimization	432
<i>Test Your Knowledge</i>	439
<i>Exercises</i>	444
12. Parallel and Distributed Databases	448–479
12.1 Introduction	448
12.2 Parallel Databases	448
12.2.1 Parallel database architecture	449
12.2.2 The key elements of parallel database processing	451
12.2.3 Query parallelism	452
12.2.4 Advantages of parallel databases	455
12.2.5 Disadvantages of parallel databases	455
12.3 Distributed Databases	455
12.3.1 Basic concepts of distributed databases	456
12.3.2 Distributed database management system (DDBMS)	457
12.3.3 Advantages of distributed databases	457
12.3.4 Disadvantages of distributed databases	458
12.3.5 Data distribution	459
12.3.6 Data replication and allocation	460
12.3.7 Distributed DBMS architectures	463
12.3.8 Comparison of DBMS and DDBMS	464
12.3.9 Query processing in distributed databases	464
12.4 Comparison of Parallel and Distributed Databases (DDB's)	466
<i>Test Your Knowledge</i>	470
<i>Exercises</i>	476
13. Data Warehouses and Data Mining	480–510
13.1 Introduction	480

13.2	Data Warehouse	481
13.2.1	Distinctive characteristics of data warehouses	481
13.2.2	Difference between database and data warehouse	482
13.2.3	Data warehouse architecture	483
13.2.4	Data warehouse components	484
13.2.5	Advantages of data warehouse	485
13.2.6	Disadvantages/limitations of data warehouse	485
13.3	Data Mart	486
13.3.1	Benefits of a data mart	486
13.3.2	Types of data marts	486
13.3.3	Steps to implement a data mart	488
13.3.4	How data mart is different from a data warehouse?	489
13.4	OLTP (On-line Transaction Processing)	490
13.4.1	Limitations of OLTP	490
13.5	OLAP (On-Line Analytical Processing)	491
13.5.1	Codd's OLAP characteristics	491
13.5.2	Difference between OLTP and OLAP	492
13.5.3	OLAP operations	493
13.5.4	Types of OLAP systems	494
13.6	Data Mining	497
13.6.1	Data mining process as a part of knowledge discovery process	497
13.6.2	Goals of data mining	498
13.6.3	Elements of data mining	499
13.6.4	Types of knowledge discovered during data mining	499
13.6.5	Models of data mining	499
13.6.6	Techniques used in data mining	500
13.6.7	Data mining tools	500
13.6.8	Data mining applications	501
13.6.9	Advantages of data mining	502
13.6.10	Disadvantages of data mining	502
13.6.11	Scope of improvement in data mining	502
13.7	Comparison of Data Mining and Structured Query Language (SQL)	502
13.8	Comparison of Data Mining and Data Warehousing	503
	<i>Test Your Knowledge</i>	504
	<i>Exercises</i>	509
14.	Database Design Project	511–519
14.1	Introduction	511
14.2	Database Design Cycle	511

14.2.1 Phase 1—Definition of the problem	511
14.2.2 Phase 2—Analyses of existing system and procedure	513
14.2.3 Phase 3—Preliminary design	514
14.2.4 Phase 4—Computing system requirements	515
14.2.5 Phase 5—Final design	515
14.2.6 Phase 6—Implementation and testing	516
14.2.7 Phase 7—Operation and tuning	517
14.3 Advantages of Database System Design Cycle	517
14.4 Limitations	517
<i>Test Your Knowledge</i>	517
<i>Exercises</i>	519
15. Additional Topics	520–524
15.1 Database Tuning	520
15.1.1 Why need database tuning	520
15.1.2 Types of tuning	520
15.2 Data Migration	522
15.2.1 Basic concepts of data migration	522
15.2.2 Why we need data migration	523
15.2.3 Key factors in data migration	523
15.2.4 Stages of data migration process	524
Index	525–532