
Contents

1	Overview of Database Management System	1
1.1	Introduction	1
1.2	Data and Information	2
1.3	Database	2
1.4	Database Management System	3
1.4.1	Structure of DBMS	3
1.5	Objectives of DBMS	4
1.5.1	Data Availability	4
1.5.2	Data Integrity	4
1.5.3	Data Security	4
1.5.4	Data Independence	5
1.6	Evolution of Database Management Systems	5
1.7	Classification of Database Management System	6
1.8	File-Based System	7
1.9	Drawbacks of File-Based System	8
1.9.1	Duplication of Data	8
1.9.2	Data Dependence	8
1.9.3	Incompatible File Formats	8
1.9.4	Separation and Isolation of Data	9
1.10	DBMS Approach	9
1.11	Advantages of DBMS	10
1.11.1	Centralized Data Management	10
1.11.2	Data Independence	10
1.11.3	Data Inconsistency	10
1.12	Ansi/Spark Data Model	11
1.12.1	Need for Abstraction	11
1.12.2	Data Independence	12
1.13	Data Models	13
1.13.1	Early Data Models	14
1.14	Components and Interfaces of Database Management System	14

1.14.1	Hardware	14
1.14.2	Software	15
1.14.3	Data	16
1.14.4	Procedure	16
1.14.5	People Interacting with Database	16
1.14.6	Data Dictionary	20
1.14.7	Functional Components of Database System Structure	21
1.15	Database Architecture	22
1.15.1	Two-Tier Architecture	22
1.15.2	Three-tier Architecture	24
1.15.3	Multitier Architecture	24
1.16	Situations where DBMS is not Necessary	26
1.17	DBMS Vendors and their Products	26
2	Entity–Relationship Model	31
2.1	Introduction	31
2.2	The Building Blocks of an Entity–Relationship Diagram	32
2.2.1	Entity	32
2.2.2	Entity Type	32
2.2.3	Relationship	32
2.2.4	Attributes	32
2.2.5	ER Diagram	33
2.3	Classification of Entity Sets	34
2.3.1	Strong Entity	34
2.3.2	Weak Entity	34
2.4	Attribute Classification	35
2.4.1	Symbols Used in ER Diagram	35
2.5	Relationship Degree	39
2.5.1	Unary Relationship	39
2.5.2	Binary Relationship	40
2.5.3	Ternary Relationship	40
2.5.4	Quaternary Relationships	40
2.6	Relationship Classification	41
2.6.1	One-to-Many Relationship Type	41
2.6.2	One-to-One Relationship Type	41
2.6.3	Many-to-Many Relationship Type	41
2.6.4	Many-to-One Relationship Type	42
2.7	Reducing ER Diagram to Tables	42
2.7.1	Mapping Algorithm	42
2.7.2	Mapping Regular Entities	43
2.7.3	Converting Composite Attribute in an ER Diagram to Tables	44
2.7.4	Mapping Multivalued Attributes in ER Diagram to Tables	45

2.7.5	Converting “Weak Entities” in ER Diagram to Tables	45
2.7.6	Converting Binary Relationship to Table	46
2.7.7	Mapping Associative Entity to Tables	47
2.7.8	Converting Unary Relationship to Tables	49
2.7.9	Converting Ternary Relationship to Tables	50
2.8	Enhanced Entity–Relationship Model (EER Model)	51
2.8.1	Supertype or Superclass	51
2.8.2	Subtype or Subclass	52
2.9	Generalization and Specialization	52
2.10	ISA Relationship and Attribute Inheritance	53
2.11	Multiple Inheritance	53
2.12	Constraints on Specialization and Generalization	54
2.12.1	Overlap Constraint	54
2.12.2	Disjoint Constraint	55
2.12.3	Total Specialization	55
2.12.4	Partial Specialization	56
2.13	Aggregation and Composition	56
2.14	Entity Clusters	57
2.15	Connection Traps	58
2.15.1	Fan Trap	59
2.15.2	Chasm Trap	59
2.16	Advantages of ER Modeling	60
3	Relational Model	65
3.1	Introduction	65
3.2	CODD’S Rules	65
3.3	Relational Data Model	67
3.3.1	Structural Part	67
3.3.2	Integrity Part	67
3.3.3	Manipulative Part	68
3.3.4	Table and Relation	69
3.4	Concept of Key	69
3.4.1	Superkey	69
3.4.2	Candidate Key	70
3.4.3	Foreign Key	70
3.5	Relational Integrity	70
3.5.1	Entity Integrity	70
3.5.2	Null Integrity	71
3.5.3	Domain Integrity Constraint	71
3.5.4	Referential Integrity	71
3.6	Relational Algebra	72
3.6.1	Role of Relational Algebra in DBMS	72
3.7	Relational Algebra Operations	72
3.7.1	Unary and Binary Operations	72

3.7.2	Rename operation (ρ)	76
3.7.3	Union Operation	77
3.7.4	Intersection Operation	78
3.7.5	Difference Operation	79
3.7.6	Division Operation	80
3.7.7	Cartesian Product Operation	82
3.7.8	Join Operations	83
3.8	Advantages of Relational Algebra	89
3.9	Limitations of Relational Algebra	89
3.10	Relational Calculus	90
3.10.1	Tuple Relational Calculus	90
3.10.2	Set Operators in Relational Calculus	92
3.11	Domain Relational Calculus (DRC)	97
3.11.1	Queries in Domain Relational Calculus:	98
3.11.2	Queries and Domain Relational Calculus Expressions	98
3.12	QBE	102
4	Structured Query Language	111
4.1	Introduction	111
4.2	History of SQL Standard	112
4.2.1	Benefits of Standardized Relational Language	113
4.3	Commands in SQL	113
4.4	Datatypes in SQL	114
4.5	Data Definition Language (DDL)	117
4.6	Selection Operation	121
4.7	Projection Operation	122
4.8	Aggregate Functions	124
4.8.1	COUNT Function	124
4.8.2	MAX, MIN, and AVG Aggregate Function	127
4.9	Data Manipulation Language	135
4.9.1	Adding a New Row to the Table	136
4.9.2	Updating the Data in the Table	137
4.9.3	Deleting Row from the Table	138
4.10	Table Modification Commands	138
4.10.1	Adding a Column to the Table	139
4.10.2	Modifying the Column of the Table	141
4.10.3	Deleting the Column of the Table	142
4.11	Table Truncation	143
4.11.1	Dropping a Table	145
4.12	Imposition of Constraints	146
4.12.1	NOT NULL Constraint	147
4.12.2	UNIQUE Constraint	149
4.12.3	Primary Key Constraint	151
4.12.4	CHECK Constraint	154

4.12.5	Referential Integrity Constraint	155
4.12.6	ON DELETE CASCADE	159
4.12.7	ON DELETE SET NULL	161
4.13	Join Operation	163
4.13.1	Equijoin	165
4.14	Set Operations	166
4.14.1	UNION Operation	166
4.14.2	INTERSECTION Operation	168
4.14.3	MINUS Operation	169
4.15	View	169
4.15.1	Nonupdatable View	172
4.15.2	Views from Multiple Tables	176
4.15.3	View From View	179
4.15.4	VIEW with CHECK Constraint	186
4.15.5	Views with Read-only Option	187
4.15.6	Materialized Views	191
4.16	Subquery	192
4.16.1	Correlated Subquery	194
4.17	Embedded SQL	201
5	PL/SQL	213
5.1	Introduction	213
5.2	Shortcomings in SQL	213
5.3	Structure of PL/SQL	214
5.4	PL/SQL Language Elements	215
5.5	Data Types	222
5.6	Operators Precedence	223
5.7	Control Structure	224
5.8	Steps to Create a PL/SQL Program	226
5.9	Iterative Control	228
5.10	Cursors	231
5.10.1	Implicit Cursors	232
5.10.2	Explicit Cursor	234
5.11	Steps to Create a Cursor	235
5.11.1	Declare the Cursor	235
5.11.2	Open the Cursor	236
5.11.3	Passing Parameters to Cursor	237
5.11.4	Fetch Data from the Cursor	237
5.11.5	Close the Cursor	237
5.12	Procedure	243
5.13	Function	247
5.14	Packages	252
5.15	Exceptions Handling	255
5.16	Database Triggers	264
5.17	Types of Triggers	267

6	Database Design	283
6.1	Introduction	283
6.2	Objectives of Database Design	285
6.3	Database Design Tools	286
6.3.1	Need for Database Design Tool	286
6.3.2	Desired Features of Database Design Tools	286
6.3.3	Advantages of Database Design Tools	287
6.3.4	Disadvantages of Database Design Tools	287
6.3.5	Commercial Database Design Tools	287
6.4	Redundancy and Data Anomaly	288
6.4.1	Problems of Redundancy	288
6.4.2	Insertion, Deletion, and Updation Anomaly	288
6.5	Functional Dependency	289
6.6	Functional Dependency Inference Rules (Armstrong's Axioms)	292
6.7	Closure of Set of Functional Dependencies	294
6.7.1	Closure of a Set of Attributes	294
6.7.2	Minimal Cover	295
6.8	Normalization	296
6.8.1	Purpose of Normalization	296
6.9	Steps in Normalization	296
6.10	Unnormal Form to First Normal Form	298
6.11	First Normal Form to Second Normal Form	300
6.12	Second Normal Form to Third Normal Form	301
6.13	Boyce–Codd Normal Form (BCNF)	304
6.14	Fourth and Fifth Normal Forms	307
6.14.1	Fourth Normal Form	307
6.14.2	Fifth Normal Form	311
6.15	Denormalization	311
6.15.1	Basic Types of Denormalization	311
6.15.2	Table Denormalization Algorithm	312
7	Transaction Processing and Query Optimization	319
7.1	Transaction Processing	319
7.1.1	Introduction	319
7.1.2	Key Notations in Transaction Management	320
7.1.3	Concept of Transaction Management	320
7.1.4	Lock-Based Concurrency Control	326
7.2	Query Optimization	332
7.2.1	Query Processing	333
7.2.2	Need for Query Optimization	333
7.2.3	Basic Steps in Query Optimization	334
7.2.4	Query Optimizer Architecture	335
7.2.5	Basic Algorithms for Executing Query Operations	341

7.2.6	Query Evaluation Plans	344
7.2.7	Optimization by Genetic Algorithms	346
8	Database Security and Recovery	353
8.1	Database Security	353
8.1.1	Introduction	353
8.1.2	Need for Database Security	354
8.1.3	General Considerations	354
8.1.4	Database Security System	356
8.1.5	Database Security Goals and Threats	356
8.1.6	Classification of Database Security	357
8.2	Database Recovery	368
8.2.1	Different Types of Database Failures	368
8.2.2	Recovery Facilities	368
8.2.3	Main Recovery Techniques	370
8.2.4	Crash Recovery	370
8.2.5	ARIES Algorithm	371
9	Physical Database Design	381
9.1	Introduction	381
9.2	Goals of Physical Database Design	382
9.2.1	Physical Design Steps	382
9.2.2	Implementation of Physical Model	383
9.3	File Organization	384
9.3.1	Factors to be Considered in File Organization	384
9.3.2	File Organization Classification	384
9.4	Heap File Organization	385
9.4.1	Uses of Heap File Organization	385
9.4.2	Drawback of Heap File Organization	385
9.4.3	Example of Heap File Organization	386
9.5	Sequential File Organization	386
9.5.1	Sequential Processing of File	387
9.5.2	Draw Back	387
9.6	Hash File Organization	387
9.6.1	Hashing Function	387
9.6.2	Bucket	388
9.6.3	Choice of Bucket	389
9.6.4	Extendible Hashing	391
9.7	Index File Organization	392
9.7.1	Advantage of Indexing	392
9.7.2	Classification of Index	392
9.7.3	Search Key	393
9.8	Tree-Structured Indexes	394
9.8.1	ISAM	394

XVIII Contents

9.8.2	B-Tree	394
9.8.3	Building a B ⁺ Tree	394
9.8.4	Bitmap Index	396
9.9	Data Storage Devices	397
9.9.1	Factors to be Considered in Selecting Data Storage Devices	397
9.9.2	Magnetic Technology	397
9.9.3	Fixed Magnetic Disk	398
9.9.4	Removable Magnetic Disk	398
9.9.5	Floppy Disk	398
9.9.6	Magnetic Tape	398
9.10	Redundant Array of Inexpensive Disk	398
9.10.1	RAID Level 0 + 1	399
9.10.2	RAID Level 0	400
9.10.3	RAID Level 1	401
9.10.4	RAID Level 2	401
9.10.5	RAID Level 3	402
9.10.6	RAID Level 4	403
9.10.7	RAID Level 5	404
9.10.8	RAID Level 6	405
9.10.9	RAID Level 10	406
9.11	Software-Based RAID	406
9.12	Hardware-Based RAID	407
9.12.1	RAID Controller	407
9.12.2	Types of Hardware RAID	408
9.13	Optical Technology	409
9.13.1	Advantages of Optical Disks	409
9.13.2	Disadvantages of Optical Disks	409
10	Data Mining and Data Warehousing	415
10.1	Data Mining	415
10.1.1	Introduction	415
10.1.2	Architecture of Data Mining Systems	416
10.1.3	Data Mining Functionalities	417
10.1.4	Classification of Data Mining Systems	417
10.1.5	Major Issues in Data Mining	418
10.1.6	Performance Issues	419
10.1.7	Data Preprocessing	420
10.1.8	Data Mining Task	423
10.1.9	Data Mining Query Language	425
10.1.10	Architecture Issues in Data Mining System	426
10.1.11	Mining Association Rules in Large Databases	427
10.1.12	Mining Multilevel Association From Transaction Databases	430

10.1.13	Rule Constraints	433
10.1.14	Classification and Prediction	434
10.1.15	Comparison of Classification Methods	436
10.1.16	Prediction	441
10.1.17	Cluster Analysis	442
10.1.18	Mining Complex Types of Data	449
10.1.19	Applications and Trends in Data Mining	453
10.1.20	How to Choose a Data Mining System	456
10.1.21	Theoretical Foundations of Data Mining	458
10.2	Data Warehousing	461
10.2.1	Goals of Data Warehousing	461
10.2.2	Characteristics of Data in Data Warehouse	462
10.2.3	Data Warehouse Architectures	462
10.2.4	Data Warehouse Design	465
10.2.5	Classification of Data Warehouse Design	467
10.2.6	The User Interface	471
11	Objected-Oriented and Object Relational DBMS	477
11.1	Objected oriented DBMS	477
11.1.1	Introduction	477
11.1.2	Object-Oriented Programming Languages (OOPLs)	479
11.1.3	Availability of OO Technology and Applications	481
11.1.4	Overview of OODBMS Technology	482
11.1.5	Applications of an OODBMS	487
11.1.6	Evaluation Criteria	491
11.1.7	Evaluation Targets	519
11.1.8	Object Relational DBMS	525
11.1.9	Object-Relational Model	526
11.1.10	Aggregation and Composition in UML	529
11.1.11	Object-Relational Database Design	530
11.1.12	Comparison of OODBMS and ORDBMS	537
12	Distributed and Parallel Database Management Systems ..	559
12.1	Distributed Database	559
12.1.1	Features of Distributed vs. Centralized Databases	561
12.2	Distributed DBMS Architecture	562
12.2.1	DBMS Standardization	562
12.2.2	Architectural Models for Distributed DBMS	563
12.2.3	Types of Distributed DBMS Architecture	564
12.3	Distributed Database Design	565
12.3.1	Framework for Distributed Database Design	566
12.3.2	Objectives of the Design of Data Distribution	567
12.3.3	Top-Down and Bottom-Up Approaches to the Design of Data Distribution	568
12.3.4	Design of Database Fragmentation	568

12.4	Semantic Data Control	572
12.4.1	View Management	572
12.4.2	Views in Centralized DBMSs	573
12.4.3	Update Through Views	573
12.4.4	Views in Distributed DBMS	574
12.4.5	Data Security	574
12.4.6	Centralized Authorization Control	575
12.4.7	Distributed Authorization Control	575
12.4.8	Semantic Integrity Control	576
12.4.9	Distributed Semantic Integrity Control	577
12.5	Distributed Concurrency Control	578
12.5.1	Serializability Theory	578
12.5.2	Taxonomy of Concurrency Control Mechanism	578
12.5.3	Locking-Based Concurrency Control	580
12.5.4	Timestamp-Based Concurrency Control Algorithms	582
12.5.5	Optimistic Concurrency Control Algorithms	583
12.5.6	Deadlock Management	583
12.6	Distributed DBMS Reliability	586
12.6.1	Reliability Concepts and Measures	586
12.6.2	Failures in Distributed DBMS	588
12.6.3	Basic Fault Tolerance Approaches and Techniques	590
12.6.4	Distributed Reliability Protocols	590
12.7	Parallel Database	592
12.7.1	Database Server and Distributed Databases	593
12.7.2	Main Components of Parallel Processing	595
12.7.3	Functional Aspects	597
12.7.4	Various Parallel System Architectures	599
12.7.5	Parallel DBMS Techniques	602
13	Recent Challenges in DBMS	611
13.1	Genome Databases	612
13.1.1	Introduction	612
13.1.2	Basic Idea of Genome	612
13.1.3	Building Block of DNA	612
13.1.4	Genetic Code	614
13.1.5	GDS (Genome Directory System) Project	614
13.1.6	Conclusion	619
13.2	Mobile Database	619
13.2.1	Concept of Mobile Database	619
13.2.2	General Block Diagram of Mobile Database Center	620
13.2.3	Mobile Database Architecture	620
13.2.4	Modes of Operations of Mobile Database	622
13.2.5	Mobile Database Management	622
13.2.6	Mobile Transaction Processing	623
13.2.7	Distributed Database for Mobile	624

13.3	Spatial Database	626
13.3.1	Spatial Data Types	627
13.3.2	Spatial Database Modeling	628
13.3.3	Discrete Geometric Spaces	628
13.3.4	Querying	629
13.3.5	Integrating Geometry into a Query Language	630
13.3.6	Spatial DBMS Implementation	631
13.4	Multimedia Database Management System	632
13.4.1	Introduction	632
13.4.2	Multimedia Data	632
13.4.3	Multimedia Data Model	633
13.4.4	Architecture of Multimedia System	635
13.4.5	Multimedia Database Management System Development	636
13.4.6	Issues in Multimedia DBMS	636
13.5	XML	637
13.5.1	Introduction	637
13.5.2	Origin of XML	637
13.5.3	Goals of XML	638
13.5.4	XML Family	638
13.5.5	XML and HTML	638
13.5.6	XML Document	639
13.5.7	Document Type Definitions (DTD)	640
13.5.8	Extensible Style Sheet Language (XSL)	640
13.5.9	XML Namespaces	641
13.5.10	XML and Database Applications	643
14	Projects in DBMS	645
14.1	List of Projects	645
14.2	Overview of the Projects	645
14.2.1	Front-End: Microsoft Visual Basic	645
14.2.2	Back-End: Oracle 9i	646
14.2.3	Interface: ODBC	646
14.3	First Project: Bus Transport Management System	647
14.3.1	Description	647
14.3.2	Features of the Project	647
14.3.3	Source Code	649
14.4	Second Project: Course Administration System	656
14.4.1	Description	656
14.4.2	Source Code	656
14.5	Third Project: Election Voting System	666
14.5.1	Description	666
14.5.2	Source Code	666
14.6	Fourth Project: Hospital Management System	673
14.6.1	Description	673
14.6.2	Source Code	674

XXII Contents

14.7 Fifth Project: Library Management System	680
14.7.1 Description	680
14.7.2 Source Code	680
14.8 Sixth Project: Railway Management System	690
14.8.1 Description	690
14.8.2 Source Code	690
14.9 Some Hints to Do Successful Projects in DBMS	696
A Dictionary of DBMS Terms	699
B Overview of Commands in SQL	721
C Pioneers in DBMS	727
C.1 About Dr. Edgar F. Codd	728
C.2 Ronald Fagin	736
C.2.1 Abstract of Ronald Fagin's Article	737
D Popular Commercial DBMS	739
D.1 System R	739
D.1.1 Introduction to System R	739
D.1.2 Keywords Used	739
D.1.3 Architecture and System Structure	740
D.1.4 Relational Data Interface	742
D.1.5 Data Manipulation Facilities in SEQUEL	743
D.1.6 Data Definition Facilities	745
D.1.7 Data Control Facilities	746
D.2 Relational Data System	749
D.3 DB2	752
D.3.1 Introduction to DB2	752
D.3.2 Definition of DB2 Data Structures	753
D.3.3 DB2 Stored Procedure	753
D.3.4 DB2 Processing Environment	755
D.3.5 DB2 Commands	757
D.3.6 Data Sharing in DB2	759
D.3.7 Conclusion	760
D.4 Informix	760
D.4.1 Introduction to Informix	760
D.4.2 Informix SQL and ANSI SQL	761
D.4.3 Software Dependencies	762
D.4.4 New Features in Version 7.3	763
D.4.5 Conclusion	766
Bibliography	767



<http://www.springer.com/978-3-540-48397-7>

Fundamentals of Relational Database Management
Systems

Sumathi, S.; Esakkirajan, S.

2007, XXVII, 776 p., Hardcover

ISBN: 978-3-540-48397-7