

73

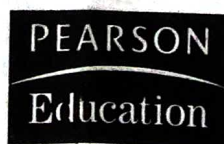
Principles of Distributed Database Systems

M. Tamer Özsu

University of Alberta
Edmonton, Canada

Patrick Valduriez

INRIA
Paris, France



CONTENTS

PREFACE

xiii

1 INTRODUCTION

1

1.1 DISTRIBUTED DATA PROCESSING

2

1.2 WHAT IS A DISTRIBUTED DATABASE SYSTEM?

3

1.3 PROMISES OF DDBSs

6

1.3.1 Transparent Management of Distributed and Replicated Data

7

1.3.2 Reliability through Distributed Transactions

13

1.3.3 Improved Performance

14

1.4 COMPLICATING FACTORS

16

1.5 PROBLEM AREAS

18

1.5.1 Distributed Database Design

18

1.5.2 Distributed Query Processing

18

1.5.3 Distributed Directory Management

18

1.5.4 Distributed Concurrency Control

19

1.5.5 Distributed Deadlock Management

19

1.5.6 Reliability of Distributed DBMS

19

1.5.7 Operating System Support

19

1.5.8 Heterogeneous Databases

20

1.5.9 Relationship among Problems

20

2 OVERVIEW OF RELATIONAL DBMS

23

2.1 RELATIONAL DATABASE CONCEPTS

24

2.2 NORMALIZATION

25

2.2.1 Dependency Structures

27

2.2.2 Normal Forms

29

iii

2.3	INTEGRITY RULES	31
2.4	RELATIONAL DATA LANGUAGES	32
2.4.1	Relational Algebra	32
2.4.2	Relational Calculus	40
2.5	RELATIONAL DBMS	45
3	REVIEW OF COMPUTER NETWORKS	48
3.1	DATA COMMUNICATION CONCEPTS	49
3.2	TYPES OF NETWORKS	50
3.2.1	Topology	50
3.2.2	Communication Schemes	53
3.3	PROTOCOL STANDARDS	54
3.4	BROADBAND NETWORKS	56
3.5	WIRELESS NETWORKS	58
3.6	INTERNET	59
4	DISTRIBUTED DBMS ARCHITECTURE	61
4.1	DBMS STANDARDIZATION	62
4.2	ARCHITECTURAL MODELS FOR DISTRIBUTED DBMSs	66
4.2.1	Autonomy	67
4.2.2	Distribution	69
4.2.3	Heterogeneity	70
4.2.4	Architectural Alternatives	70
4.3	DISTRIBUTED DBMS ARCHITECTURE	73
4.3.1	Client/Server Systems	73
4.3.2	Peer-to-Peer Distributed Systems	75
4.3.3	MDBS Architecture	79
5	DISTRIBUTED DATABASE DESIGN	84
5.1	ALTERNATIVE DESIGN STRATEGIES	86
5.1.1	Top-Down Design Process	86
5.1.2	Bottom-Up Design Process	88
5.2	DISTRIBUTION DESIGN ISSUES	88
5.2.1	Reasons for Fragmentation	88
5.2.2	Fragmentation Alternatives	89
5.2.3	Degree of Fragmentation	90
5.2.4	Correctness Rules of Fragmentation	91
5.2.5	Allocation Alternatives	92
5.2.6	Information Requirements	93

5.3	FRAGMENTATION	93
5.3.1	Horizontal Fragmentation	94
5.3.2	Vertical Fragmentation	111
5.3.3	Hybrid Fragmentation	124
5.4	ALLOCATION	126
5.4.1	Allocation Problem	126
5.4.2	Information Requirements	128
5.4.3	Allocation Model	130
5.4.4	Solution Methods	133
6	SEMANTIC DATA CONTROL	134
6.1	VIEW MANAGEMENT	135
6.1.1	Views in Centralized DBMSs	135
6.1.2	Updates through Views	137
6.1.3	Views in Distributed DBMSs	138
6.2	DATA SECURITY	139
6.2.1	Centralized Authorization Control	140
6.2.2	Distributed Authorization Control	142
6.3	SEMANTIC INTEGRITY CONTROL	143
6.3.1	Centralized Semantic Integrity Control	145
6.3.2	Distributed Semantic Integrity Control	150
7	OVERVIEW OF QUERY PROCESSING	156
7.1	QUERY PROCESSING PROBLEM	157
7.2	OBJECTIVES OF QUERY PROCESSING	160
7.3	COMPLEXITY OF RELATIONAL ALGEBRA OPERATIONS	161
7.4	CHARACTERIZATION OF QUERY PROCESSORS	162
7.4.1	Languages	163
7.4.2	Types of Optimization	163
7.4.3	Optimization Timing	164
7.4.4	Statistics	164
7.4.5	Decision Sites	164
7.4.6	Exploitation of the Network Topology	165
7.4.7	Exploitation of Replicated Fragments	165
7.4.8	Use of Semijoins	165
7.5	LAYERS OF QUERY PROCESSING	166
7.5.1	Query Decomposition	167
7.5.2	Data Localization	168

	168
7.5.3 Global Query Optimization	169
7.5.4 Local Query Optimization	
8 QUERY DECOMPOSITION AND DATA LOCALIZATION	170
8.1 QUERY DECOMPOSITION	170
8.1.1 Normalization	172
8.1.2 Analysis	175
8.1.3 Elimination of Redundancy	176
8.1.4 Rewriting	180
8.2 LOCALIZATION OF DISTRIBUTED DATA	181
8.2.1 Reduction for Primary Horizontal Fragmentation	184
8.2.2 Reduction for Vertical Fragmentation	185
8.2.3 Reduction for Derived Fragmentation	187
8.2.4 Reduction for Hybrid Fragmentation	
9 OPTIMIZATION OF DISTRIBUTED QUERIES	190
9.1 QUERY OPTIMIZATION	190
9.1.1 Search Space	191
9.1.2 Search Strategy	193
9.1.3 Distributed Cost Model	194
9.2 CENTRALIZED QUERY OPTIMIZATION	199
9.2.1 INGRES Algorithm	200
9.2.2 System R Algorithm	203
9.3 JOIN ORDERING IN FRAGMENT QUERIES	206
9.3.1 Join Ordering	206
9.3.2 Semijoin-Based Algorithms	208
9.3.3 Join versus Semijoin	211
9.4 DISTRIBUTED QUERY OPTIMIZATION ALGORITHMS	212
9.4.1 Distributed INGRES Algorithm	213
9.4.2 R* Algorithm	218
9.4.3 SDD-1 Algorithm	222
10 INTRODUCTION TO TRANSACTION MANAGEMENT	228
10.1 DEFINITION OF A TRANSACTION	229
10.1.1 Termination Conditions of Transactions	232
10.1.2 Characterization of Transactions	233
10.1.3 Formalization of the Transaction Concept	234
10.2 PROPERTIES OF TRANSACTIONS	237
10.2.1 Atomicity	237

10.2.2	Consistency	238
10.2.3	Isolation	239
10.2.4	Durability	242
10.3	TYPES OF TRANSACTIONS	242
10.3.1	Flat Transactions	244
10.3.2	Nested Transactions	244
10.3.3	Workflows	246
11	DISTRIBUTED CONCURRENCY CONTROL	250
11.1	SERIALIZABILITY THEORY	250
11.2	TAXONOMY OF CONCURRENCY CONTROL MECHANISMS	256
11.3	LOCKING-BASED CONCURRENCY CONTROL ALGORITHMS	258
11.3.1	Centralized 2PL	267
11.3.2	Primary Copy 2PL	271
11.3.3	Distributed 2PL	272
11.4	TIMESTAMP-BASED CONCURRENCY CONTROL ALGORITHMS	273
11.4.1	Basic TO Algorithm	274
11.4.2	Conservative TO Algorithm	278
11.4.3	Multiversion TO Algorithm	280
11.5	OPTIMISTIC CONCURRENCY CONTROL ALGORITHMS	281
11.6	DEADLOCK MANAGEMENT	284
11.6.1	Deadlock Prevention	285
11.6.2	Deadlock Avoidance	286
11.6.3	Deadlock Detection and Resolution	288
11.7	"RELAXED" CONCURRENCY CONTROL	291
11.7.1	Non-Serializable Schedules	291
11.7.2	Nested Distributed Transactions	292
12	DISTRIBUTED DBMS RELIABILITY	295
12.1	RELIABILITY CONCEPTS AND MEASURES	295
12.1.1	System, State, and Failure	295
12.1.2	Reliability and Availability	298
12.1.3	Mean Time between Failures/Mean Time to Repair	300
12.2	FAILURES AND FAULT TOLERANCE IN DISTRIBUTED SYSTEMS	300
12.2.1	Reasons for Failures	300
12.2.2	Basic Fault Tolerance Approaches and Techniques	303
12.3	FAILURES IN DISTRIBUTED DBMS	304
12.3.1	Transaction Failures	305

12.3.2 Site (System) Failures	305
12.3.3 Media Failures	306
12.3.4 Communication Failures	307
12.4 LOCAL RELIABILITY PROTOCOLS	307
12.4.1 Architectural Considerations	310
12.4.2 Recovery Information	315
12.4.3 Execution of LRM Commands	319
12.4.4 Checkpointing	320
12.5 DISTRIBUTED RELIABILITY PROTOCOLS	321
12.5.1 Components of Distributed Reliability Protocols	322
12.5.2 Two-Phase Commit Protocol	329
12.5.3 Variations of 2PC	330
12.6 DEALING WITH SITE FAILURES	331
12.6.1 Termination and Recovery Protocols for 2PC	337
12.6.2 Three-Phase Commit Protocol	342
12.7 NETWORK PARTITIONING	343
12.7.1 Centralized Protocols	344
12.7.2 Voting-based Protocols	346
12.7.3 Replication and Replica Control Protocols	347
12.7.4 Strict Replica Control Protocols	351
13 PARALLEL DATABASE SYSTEMS	352
13.1 DATABASE SERVERS	353
13.1.1 Database Server Approach	355
13.1.2 Database Servers and Distributed Databases	356
13.2 PARALLEL ARCHITECTURES	359
13.2.1 Objectives	364
13.2.2 Functional Aspects	368
13.2.3 Parallel System Architectures	370
13.3 PARALLEL DBMS TECHNIQUES	378
13.3.1 Data Placement	382
13.3.2 Query Parallelism	382
13.3.3 Parallel Data Processing	383
13.3.4 Parallel Query Optimization	383
13.4 PARALLEL EXECUTION PROBLEMS	
13.4.1 Initialization	
13.4.2 Interferences and Convoy Effect	
13.4.3 Load Balancing	

13.5 PARALLEL EXECUTION FOR HIERARCHICAL ARCHITECTURE	386
13.5.1 Problem Formulation	386
13.5.2 Basic Concepts	388
13.5.3 Load Balancing Strategy	390
13.5.4 Performance Evaluation	391
14 DISTRIBUTED OBJECT DATABASE MANAGEMENT SYSTEMS	392
14.1 FUNDAMENTAL OBJECT CONCEPTS AND OBJECT MODELS	393
14.1.1 Object	393
14.1.2 Abstract Data Types	396
14.1.3 Composition (Aggregation)	397
14.1.4 Class	398
14.1.5 Collection	398
14.1.6 Subtyping and Inheritance	399
14.2 OBJECT DISTRIBUTION DESIGN	400
14.2.1 Horizontal Class Partitioning	401
14.2.2 Vertical Class Partitioning	404
14.2.3 Path Partitioning	404
14.2.4 Class Partitioning Algorithms	404
14.2.5 Allocation	405
14.2.6 Replication	406
14.3 ARCHITECTURAL ISSUES	406
14.3.1 Alternative Client/Server Architectures	407
14.3.2 Cache Consistency	412
14.4 OBJECT MANAGEMENT	414
14.4.1 Object Identifier Management	414
14.4.2 Pointer Swizzling	416
14.4.3 Object Migration	417
14.5 DISTRIBUTED OBJECT STORAGE	418
14.6 OBJECT QUERY PROCESSING	421
14.6.1 Object Query Processor Architectures	423
14.6.2 Query Processing Issues	424
14.6.3 Query Execution	429
14.7 TRANSACTION MANAGEMENT	433
14.7.1 Correctness Criteria	435
14.7.2 Transaction Models and Object Structures	439
14.7.3 Transactions Management in Object DBMSs	441
14.7.4 Transactions as Objects	450

15 DATABASE INTEROPERABILITY	451
15.1 DATABASE INTEGRATION	451
15.1.1 Schema Translation	453
15.1.2 Schema Integration	456
15.2 QUERY PROCESSING	463
15.2.1 Query Processing Layers in Distributed Multi-DBMSs	464
15.2.2 Query Optimization Issues	466
15.3 TRANSACTION MANAGEMENT	470
15.3.1 Transaction and Computation Model	470
15.3.2 Multidatabase Concurrency Control	473
15.3.3 Multidatabase Recovery	477
15.4 OBJECT ORIENTATION AND INTEROPERABILITY	477
15.4.1 Object Management Architecture	479
15.4.2 CORBA and Database Interoperability	483
15.4.3 Distributed Component Object Model	486
15.4.4 COM/OLE and Database Interoperability	488
16 CURRENT ISSUES	490
16.1 DATA DELIVERY ALTERNATIVES	492
16.2 DATA WAREHOUSING	494
16.2.1 Architectures	495
16.2.2 OLAP Data Model	497
16.2.3 OLAP Servers	498
16.2.4 Research Issues	499
16.3 WORLD WIDE WEB	500
16.3.1 Architecture and Protocols	501
16.3.2 Database Access	501
16.3.3 Semistructured Data	503
16.3.4 Architectures for Information Integration	504
16.3.5 Research Projects and Open Issues	506
16.4 PUSH-BASED TECHNOLOGIES	508
16.4.1 Delivery Schedule Generation	509
16.4.2 Client Cache Management	512
16.4.3 Propagating Updates	514
16.5 MOBILE DATABASES	515
16.5.1 Directory Management	517
16.5.2 Caching	518
16.5.3 Broadcast Data	519

CONTENTS

xi

16.5.4 Query Processing and Optimization	520
16.5.5 Transaction Management	521

BIBLIOGRAPHY	523
--------------	-----

SUBJECT INDEX	587
---------------	-----