
INTERCONNECTING HETEROGENEOUS INFORMATION SYSTEMS

by

Athman Bouguettaya

Queensland University of Technology

Brisbane, AUSTRALIA

Boualem Benatallah

Queensland University of Technology

Brisbane, AUSTRALIA

Ahmed Elmagarmid

Purdue University

Lafayette, IN USA



KLUWER ACADEMIC PUBLISHERS

Boston / Dordrecht / London

Contents

List of Figures	xi
List of Tables	xiii
Preface	xv
Acknowledgments	xix
Foreword	xxi
1. INTRODUCTION	1
2. MODELING ARCHITECTURE	5
2.1 Terminology	6
2.2 Global Schema Integration	9
2.3 Federated Database Systems	11
2.3.1 Loosely Coupled FDBS	13
2.3.2 Tightly Coupled FDBS	14
2.4 Multidatabase Language Approach	14
3. CRITICAL ASPECTS OF INTEROPERABILITY	19
3.1 Schema and Language Translation	20
3.2 Schema Integration	25
3.2.1 Schema and Data Heterogeneity	25
Types of Discrepancies	27
Examples of Semantic Conflict Resolution	30
Tools to Aid Conflicts Resolution	32
3.2.2 Schema Integration Process	33
3.3 Query Processing	35
3.3.1 Query Processing in Centralized DBMS	36
3.3.2 Query Processing in Distributed Database Systems	36
3.3.3 Query Processing in Heterogeneous Database Systems	38
Major Issues in Multidatabase Query Optimization	38
Generic Architecture of a Multidatabase Query Processor	39
Proposed Solutions for Query Optimization	40
3.4 Transactions	44
3.4.1 Quasi-serializability	46
3.4.2 Hierarchical Concurrency Control	50
3.4.3 Value Dates	51

3.4.4	Extended Transaction Models	52
4.	RESEARCH AND COMMERCIAL PRODUCTS	55
4.1	DAPLEX	55
4.1.1	Data Definition	56
4.1.2	Data Manipulation	56
4.1.3	Higher Level Capabilities	57
4.1.4	Databases Integration	57
4.2	IRIS	58
4.2.1	Objects and Types	58
4.2.2	Functions	59
4.3	Pegasus	60
4.3.1	Architecture	60
4.3.2	Common Data Model	61
4.3.3	Heterogeneous Object SQL	62
4.3.4	Query Processing	62
4.3.5	Translation and Integration	63
4.3.6	Schema Integration	64
4.4	Comandos Integration System	64
4.4.1	System Architecture	65
4.4.2	Common Data Model	65
4.4.3	Translation and Integration	66
4.4.4	Query Processing	66
4.5	Federated Objectbase System	66
4.5.1	System Architecture	66
4.5.2	Common Data Model	67
4.5.3	Translation and Integration	67
4.6	Myriad	68
4.6.1	Architecture	68
4.6.2	Query Processing	69
4.6.3	Federated Query Managers and Agents	69
4.6.4	Query Optimization	69
4.6.5	Transaction Management	70
4.6.6	Translation and Integration	71
4.7	Multibase	72
4.7.1	System Architecture	72
4.7.2	Common Data Model, Translation, and Integration	73
4.7.3	Query Processing	73
4.8	Mermaid	73
4.8.1	System Architecture	73
4.8.2	Common Data Model and Data Translation	74
4.8.3	Query Processing	75
4.9	SIMS Project	77
4.9.1	Knowledge Representation or Modeling	78
4.9.2	Selecting the Right Databases	78
4.9.3	Access Planning	78
4.9.4	Global Query Optimization	78
4.10	InterBase	79

4.11 OMNIBASE	80
4.12 Other Systems	81
5. MIDDLEWARE	87
5.1 CORBA	88
5.1.1 The Common Object Request Broker Architecture	89
Static Invocation	90
Dynamic Invocation	90
5.1.2 ORBs Interoperability	92
5.1.3 Examples of CORBA Commercial ORBs	93
5.2 OLE/COM	94
5.2.1 Component Object Model (COM)	94
5.2.2 OLE	95
5.2.3 OLE/DB and ActiveX	96
5.3 DCE	97
5.4 OpenDoc	100
5.5 ANSAware	102
5.6 Other Relevant Middleware Technologies	103
5.6.1 Java Technology	104
5.6.2 HTML/CGI Technology	104
5.6.3 Message-Oriented Middleware and TP monitors	105
5.6.4 Distributed Object Manager System	106
5.6.5 Arjuna	106
5.7 Conclusion	106
6. RESEARCH TRENDS	109
6.1 Information Sharing on the Web	109
6.1.1 Distributed Naming Systems	111
6.1.2 Information Retrieval	112
Network-based Information Retrieval	113
Internet Information Gathering	113
Multi-Agent Information Retrieval	114
6.1.3 WWW Search Tools	114
Robots	115
Meta-Robots	116
Database-like WWW Languages	116
6.1.4 Information Brokering Techniques	117
6.1.5 Summary	121
6.2 Transactional Workflows	121
Definition	122
6.2.1 Background and Motivation	122
6.2.2 Features of Workflow Systems	123
Types of Workflow Systems	123
Characteristics of a Workflow Management Systems	124
Workflow Management Coalition (WfMC)	125
6.2.3 Modeling Techniques and Prototypes	126
Advanced Transaction Management (ATM)	126
Petri-net	127

Goal-Oriented Workflow Analysis and Infrastructure	127
Mentor	127
METEOR2	128
CBORD	128
WASA	129
ORCHESTRA	129
Other Techniques	129
6.2.4 Limitations of Existing Systems	130
6.2.5 Research Efforts	130
6.2.6 Summary	132
6.3 Advances in Data Warehousing	133
6.3.1 Background and Motivation	134
Data Mining and Knowledge Discovery	135
On-Line Analysis and Processing	136
Architectural Requirements	137
6.3.2 Basic Architectures	138
6.3.3 Warehouse Consistency and View Maintenance	139
Levels of Consistency	139
Maintaining Data Consistency	140
View Maintenance for Warehouse	140
Concurrency Control Problem	143
6.3.4 Warehouse Design Issues	143
Techniques for the View-Selection Problem	143
6.3.5 Summary	144
7. CASE STUDY 1: WEBFINDIT PROTOTYPE	145
7.1 Introduction	145
7.2 Design Overview	149
7.2.1 Core Concepts	149
Coalitions	150
Service Links	150
7.2.2 Documentation	151
7.2.3 Co-databases	153
7.3 WebTassili Language	155
7.3.1 Rationale	155
7.3.2 WebTassili Salient Features	157
7.3.3 Evolution Features	160
Rules for Joining and Leaving WebFINDIT	160
Coalition Formation and Update	161
Service Formation and Update	165
7.4 Implementing WebFINDIT	166
7.4.1 Overview	166
7.4.2 Browser	168
7.4.3 Query Processor	169
7.4.4 Co-database Server	170
7.4.5 Broker	173
7.4.6 Information Source Interface	174
7.5 Scalability Evaluation	175
7.6 Related Work	176

7.7 Conclusion	177
8. CASE STUDY 2: A TRANSACTION-ORIENTED WORKFLOW ENVIRONMENT	179
8.1 TOWE Architecture	180
8.1.1 Process Model	180
8.1.2 Scheduling Processes	183
8.1.3 The Class Library	185
8.2 Process Definition	187
8.2.1 The Scheduler Class	187
8.2.2 Implementation of Process Objects	188
8.2.3 Recovery	190
8.3 Implementation	191
8.4 Conclusion	193
References	195
Index	213