

Designing XML Databases

MARK GRAVES



Prentice Hall PTR, Upper Saddle River, NJ 07458
www.phptr.com

Contents

Preface *xvii*

1

Introduction *2*

- 1.1 XML *4*
 - 1.1.1 What Is XML? *4*
 - 1.1.2 Where Does XML Come from? *8*
 - 1.1.3 Why Use XML? *9*
- 1.2 Database Systems *12*
 - 1.2.1 What Is a Database? *13*
 - 1.2.2 What Is an XML Database? *14*
 - 1.2.3 Why Use XML Databases? *16*
- 1.3 Web Databases *17*
 - 1.3.1 Flat File Database *18*
 - 1.3.2 Relational DBMS *22*
 - 1.3.3 XML DBMS *23*

1.4 Applications	24
1.5 Resources	25
1.5.1 Magazines	25
1.5.2 General Sites	26
1.5.3 XML Portals	26
1.5.4 XML Tools	27
1.5.5 XSL	27
1.5.6 W3C Documents	27
1.5.7 Example Domain-specific XML Specifications	28
1.5.8 More XML Information	28

2

Schema Design 30

2.1 Database Design	32
2.2 Conceptual Modeling	35
2.2.1 Graph Conceptual Model	35
2.2.2 Graph Conceptual Modeling Process	42
2.2.3 Conceptual Modeling Issues	47
2.2.4 XML Conceptual Model	55
2.3 Logical Modeling	57
2.3.1 Entity-relation Diagram	58
2.3.2 Relational Schema	58
2.3.3 Object Model	60
2.3.4 XML Logical Schema	66
2.4 Physical Modeling	70
2.4.1 XML Physical Schema	71
2.4.2 Data-processing-oriented View versus Document-processing-oriented View	74
2.4.3 Data Transport	77
2.4.4 Attribute versus Embedded Element Tradeoff	79
2.5 Bibliographic Remarks	82
2.5.1 References	83

3

Theoretical Foundations 84

- 3.1 Data Types 85
 - 3.1.1 XML Schema 87
 - 3.1.2 Emerging Structural Data Types 88
 - 3.1.3 Schema-driven applications 89
- 3.2 DBMSs 94
- 3.3 XML Standards 98
 - 3.3.1 XML Schema (XSDL) 99
 - 3.3.2 XSL 100
 - 3.3.3 XML Linking, Pointer, and Path 100
 - 3.3.4 XML Query 101
 - 3.3.5 XML Namespaces 102
 - 3.3.6 DOM 102
- 3.4 XML DBMS 102
 - 3.4.1 Conceptual Schema 104
 - 3.4.2 Tasks 105
 - 3.4.3 Operations 106
- 3.5 Data Modeling 108
 - 3.5.1 Existing Data Models 109
 - 3.5.2 Simple XML Data Model 114
 - 3.5.3 W3C Specification-oriented XML Data Model 119
 - 3.5.4 Relation-oriented XML Data Model 121
 - 3.5.5 Node-centric XML Data Model 131
 - 3.5.6 Edge-centric XML Data Model 136
 - 3.5.7 Generic XML Data Model 141
- 3.6 Bibliographic Remarks 149
 - 3.6.1 References 149

4

Data Storage 150

- 4.1 Storage Facilities 151
 - 4.1.1 Flat File Database 152
 - 4.1.2 Object Database 155
 - 4.1.3 Relational Database 163
- 4.2 Fine-grained Relational Schema 163
 - 4.2.1 Logical Design 164
 - 4.2.2 Physical Design 168
 - 4.2.3 Examples 173
 - 4.2.4 Implementation 179
- 4.3 Coarse-grained Relational Schema 207
- 4.4 Medium-grained Relational Schema 209
 - 4.4.1 Slice Points 210
 - 4.4.2 Database Design 212
 - 4.4.3 Implementation 216
- 4.5 Practical Considerations 227

5

Database System Architecture 230

- 5.1 System Architecture 231
 - 5.1.1 Client/Server 234
 - 5.1.2 Three-tier 237
- 5.2 XML Web Server 238
 - 5.2.1 Implementation Options 238
 - 5.2.2 Client Access 241
 - 5.2.3 Data Loading 242
 - 5.2.4 XML Generation 259
- 5.3 Relational Data Server 260
 - 5.3.1 Specifying URL Requests 262

5.3.2	Creating an SQL Query	264
5.3.3	Formatting a Report as XML	265
5.3.4	Extracting Dictionary Data	266
5.3.5	Implementation	271
5.4	XML Data Server	299
5.4.1	Implementation	303
5.5	Hybrid Relational/XML Server	325
5.5.1	Implementation	326

6

Commercial Systems 334

6.1	Overview	336
6.2	Database Adaptors	337
6.2.1	Middleware Tools	338
6.2.2	Commercial Relational Databases	338
6.2.3	Query Tools	339
6.3	DBMSs	339
6.4	XML Data Servers	340
6.4.1	dbXML	340
6.4.2	eXcelon	340
6.4.3	Tamino	340
6.5	XML Document Servers	341
6.6	Resources & Sites	342

7

User Interface 344

7.1	Overview	345
7.2	XSL-based User Interfaces	346
7.2.1	XSL Stylesheets	347
7.2.2	Rendering XML as a Table	349
7.2.3	Rendering XML Fragment as a Record	356

7.2.4	Rendering Identifiers and Proxies as Hypertext Links	358
7.2.5	Varying Presentation Based on Content	363
7.3	Java-based Visualizations	367
7.3.1	Client Architecture	368
7.3.2	Tree Example	371
7.4	Instant Applications	379

8

Querying 388

8.1	Query Classifications	389
8.2	Representation	393
8.2.1	Organization-centered Documents versus Relationship-centered Data	393
8.2.2	Node-centric versus Edge-centric Representations	395
8.2.3	Representing Links	398
8.2.4	XML Links Presentations of Edges	401
8.2.5	Storing Links	402
8.3	Query Engines	404
8.3.1	Path Querying	404
8.3.2	Tree Querying	409
8.4	Graph Querying	410
8.4.1	Graph Data Model	412
8.4.2	Graph Patterns	413
8.4.3	Visualization Tools	415
8.4.4	SQL Implementation	417
8.4.5	Graph Query Algorithm	445
8.5	Query Report Tools	454
8.5.1	Path Querying with XSL	454
8.5.2	Graph Querying	455

9

Indexing 458

- 9.1 Overview 459
- 9.2 Element Data Structures 460
- 9.3 Indexing Strategies 461
 - 9.3.1 No Indexing 461
 - 9.3.2 Complete Indexing 462
 - 9.3.3 Partial Indexing 467
 - 9.3.4 Cross-document Indexing 470
- 9.4 Document Identification 474
- 9.5 Search Technologies 477

10

Implementation 480

- 10.1 Notebook System 481
- 10.2 Biological Motivation 483
- 10.3 User Requirements 484
- 10.4 Conceptual Model 485
- 10.5 Application Description 489
 - 10.5.1 Client 489
 - 10.5.2 Middle-tier 499
- 10.6 Limitations and Extensions 533
- 10.7 Practical Considerations 534
- 10.8 Scaling Up 534
 - 10.8.1 Transaction Management 534
 - 10.8.2 Security 535
 - 10.8.3 Recovery 536
 - 10.8.4 Optimization 536