

The Definitive Guide to Berkeley DB XML



Danny Brian

Apress®



The Definitive Guide to Berkeley DB XML

Copyright © 2006 by Danny Brian

All rights reserved. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval system, without the prior written permission of the copyright owner and the publisher.

ISBN-13: 978-1-59059-666-1

ISBN-10: 1-59059-666-8

Library of Congress Cataloging-in-Publication data is available upon request.

Printed and bound in the United States of America 9 8 7 6 5 4 3 2 1

Trademarked names may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, we use the names only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

Lead Editor: Matt Wade

Technical Reviewer: George Feinberg

Editorial Board: Steve Anglin, Ewan Buckingham, Gary Cornell, Jason Gilmore, Jonathan Gennick, Jonathan Hassell, James Huddleston, Chris Mills, Matthew Moodie, Dominic Shakeshaft, Jim Sumser, Keir Thomas, Matt Wade

Project Manager: Kylie Johnston

Copy Edit Manager: Nicole LeClerc

Copy Editor: Nancy Sixsmith

Assistant Production Director: Kari Brooks-Copony

Production Editor: Kelly Winkist

Compositor: Molly Sharp

Proofreader: Linda Seifert

Indexer: John Collin

Artist: April Milne

Cover Designer: Kurt Krames

Manufacturing Director: Tom Debolski

Distributed to the book trade worldwide by Springer-Verlag New York, Inc., 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax 201-348-4505, e-mail orders-ny@springer-sbm.com, or visit <http://www.springeronline.com>.

For information on translations, please contact Apress directly at 2560 Ninth Street, Suite 219, Berkeley, CA 94710. Phone 510-549-5930, fax 510-549-5939, e-mail info@apress.com, or visit <http://www.apress.com>.

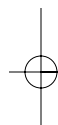
The information in this book is distributed on an “as is” basis, without warranty. Although every precaution has been taken in the preparation of this work, neither the author(s) nor Apress shall have any liability to any person or entity with respect to any loss or damage caused or alleged to be caused directly or indirectly by the information contained in this work.

The source code for this book is available to readers at <http://www.apress.com> in the Source Code section.



Contents

About the Author	xv
About the Technical Reviewer	xvi
Acknowledgments	xvii
Introduction	xix
■ CHAPTER 1 A Quick Look at Berkeley DB XML	1
A Complete Example	1
Creating and Using a Database	2
Querying a Database	3
Creating and Querying a Second Database	3
Metadata	4
XQuery	4
Conclusion	5
■ CHAPTER 2 The Power of an Embedded XML Database	7
Database Servers vs. Embedded Databases	7
Architecture Example	9
Embedded Databases You Might Know	10
SQLite	11
Wordnet	11
Embedded Databases on the Desktop	14
XML for Data Exchange	14
XML for Data Storage	16
Indexing XML	18
High-Performance XML Databases	20
BDB XML for Quality Architecture	21
Conclusion	23
■ CHAPTER 3 Installation and Configuration	25
BDB XML Packages and Layout	25
Berkeley DB	25
Xerces C++	25
Pathan	26



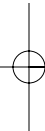
XQuery	26
Berkeley DB XML	26
Installation	26
Windows	26
Unix	28
Building and Using Individual Packages	29
Unix Variants	30
Building Bindings	31
Conclusion	33

CHAPTER 4 **Getting Started** 35

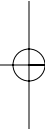
Core Concepts	35
The Shell	35
Shell Options	36
Creating Containers	36
Adding and Deleting Documents	37
Querying Containers	38
Indexing Containers	40
Using XQuery	41
Metadata	43
Transactions	44
Conclusion	45

CHAPTER 5 **Environments, Containers, and Documents** 47

Environments	47
Creating and Opening Environments	48
Additional Environment Configuration	49
Containers	50
Creating and Opening Containers	50
Container Types	51
Some Container Operations	52
Documents	53
Adding Documents	54
Retrieving a Document	54
Replacing Documents	55
Modifying Documents Programmatically	55
Deleting Documents	56
Transactions	56
Validation	56
Metadata	57
Conclusion	57



CHAPTER 6	Indexes	61
	Creating and Manipulating Indexes	61
	Index Nodes	62
	Index Types	62
	Uniqueness	63
	Path Types	63
	Node Types	64
	Key Types	64
	Syntax Types	65
	Managing Indexes	65
	Adding Indexes	65
	Listing Indexes	67
	Deleting and Replacing Indexes	67
	Default Indexes	68
	Index Strategies	68
	Query Plans	70
	Conclusion	72
 CHAPTER 7	 XQuery with BDB XML	 73
	Trying XQuery	73
	Sample Data	74
	XPath	75
	Expressions	76
	Sequences	77
	A Complete Example	78
	FLWOR Expressions	80
	for	80
	let	81
	where	81
	order by	81
	return	82
	Data Types	82
	Nodes	84
	Atomic Values	84
	Navigation	84
	Comparisons	85
	User Functions	86
	Modules	87
	Some XQuery Tricks	87
	Iteration vs. Filtering	87
	Regular Expressions	88
	Querying for Metadata	89



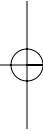
Querying Multiple Data Sources	90
Recursion	90
Reshaping Results	92
Utilizing Hierarchy	94
Ranges	96
Unions, Intersections, and Differences	98
Indexes and Queries	99
Query Plans	99
Node Names and Wildcards	101
Queries Against Results	102
Conclusion	102

CHAPTER 8 **BDB XML with C++** 103

Compiling Applications	103
Class Organization	104
Errors and Exception Handling	105
Opening Environments	107
XmlManager Class	108
Instantiating XmlManager Objects	108
Managing Containers	108
Loading Documents	110
Preparing and Executing Queries	112
Using Query Results	114
Creating Other Objects	115
Using XmlContainer	116
Using XmlDocument	119
Using XmlModify	120
Using XmlTransaction	121
BDB XML Event API	123
Conclusion	124

CHAPTER 9 **BDB XML with Python** 125

Running Applications	125
Class Organization	125
Errors and Exception Handling	126
Environments	126
XmlManager	127
Instantiating XmlManager Objects	127
Managing Containers	127
Loading Documents	128



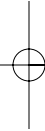
Preparing and Executing Queries	129
Using Query Results	131
Creating Other Objects	133
Using XmlContainer	133
Using XmlDocument and XmlModify	136
Transactions	138
Conclusion	139

■ CHAPTER 10 **BDB XML with Java** 141

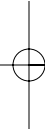
Running Applications	141
Class Organization	142
Errors and Exception Handling	142
Environments	144
XmlManager	145
Instantiating XmlManager Objects	145
Managing Containers	145
Loading Documents	147
Preparing and Executing Queries	148
Using Query Results	151
Creating Other Objects	153
Using XmlContainer	153
Using XmlDocument and XmlModify	156
Conclusion	159

■ CHAPTER 11 **BDB XML with Perl** 161

Running Applications	161
Class Organization	161
Errors and Exception Handling	162
Environments	163
XmlManager	163
Instantiating XmlManager Objects	163
Managing Containers	163
Loading Documents	165
Preparing and Executing Queries	166
Using Query Results	168
Creating Other Objects	169
Using XmlContainer	169
Using XmlDocument	172
Using XmlModify	173
Conclusion	175



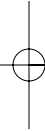
CHAPTER 12	BDB XML with PHP	177
	Running Applications	177
	Class Organization	177
	Environments	178
	XmlManager	179
	Instantiating XmlManager Objects	179
	Managing Containers	179
	Loading Documents	180
	Preparing and Executing Queries	181
	Using Query Results	183
	Creating Other Objects	184
	Using XmlContainer	185
	Using XmlDocument	187
	Using XmlModify	188
	Conclusion	189
CHAPTER 13	Managing Databases	191
	Populating Containers	191
	Dumping Containers	192
	Loading Containers	193
	Managing Logs	193
	Detecting Deadlocks	194
	Checkpointing Transactions	195
	Recovery	195
	Debugging Databases	196
	Backup and Restore	196
	Conclusion	197
APPENDIX A	XML Essentials	199
	It's About the Data	199
	XML Building Blocks	203
	Elements	203
	Attributes	203
	Well-Formedness	204
	CDATA	205
	Relationships	206
	Namespaces	206
	Validation	207
	XML Schemas	209
	XPath: the Gist	210
	Paths	211
	Nodes	211



Document Object Model (DOM)	212
XPath: the Details	214
Contexts	214
Path Operators	214
Predicates	215
Operators	217
Axes	217
Functions	219
XML DOM, Continued	221
Implementation Considerations	222
Reading and Writing XML	223
Other XML Technologies	226
XSLT	226
SAX	229
RPC-XML and SOAP	229
Conclusion	230

■ APPENDIX B **BDB XML API Reference** 231

Language Notes	231
DbEnv	232
DbXml	235
XmlContainer	238
XmlContainerConfig	260
XmlDocument	261
XmlDocumentConfig	269
XmlException	270
XmlIndexLookup	270
XmlIndexSpecification	275
XmlInputStream	282
XmlManager	284
XmlManagerConfig	310
XmlMetadataIterator	311
XmlModify	312
XmlQueryContext	317
XmlQueryExpression	324
XmlResults	327
XmlStatistics	332
XmlTransaction	334
XmlUpdateContext	336
XmlValue	337



APPENDIX C

XQuery Reference

343

Expressions

343

Functions

347

Data Types

351

INDEX

355

