

Chapter 10: XML and Databases

1

Chapter 10 Objectives

- **XML vs. Relational Databases**
- **Approaches to Storing XML**
- **How to perform foundational tasks using eXist, an Open Source native XML database.**
- **Aspects of using Native XML Databases**
- **Looking ahead**

2

The Need for Efficient XML Data Stores

- **Increase in Amount of XML**
 - Diverse Data (sources etc.)
 - Structured Data
 - Semi-Structured Data
 - Loosely Structured Data
 - Document-centric XML
 - Flexible-structured

XML - Document or Data?

3

Comparing XML-Based Data and Relational Data

- | | |
|--|---|
| <ul style="list-style-type: none">• Relational<ul style="list-style-type: none">• Structured• No hierarchy• Table: rows and columns• Order of rows in table not important• Data in tables associated using keys | <ul style="list-style-type: none">• XML-Based<ul style="list-style-type: none">• Semi-structured as well as structured• Heirarchical,<ul style="list-style-type: none">• XPath Model• Document Object Model• XML Infoset• Elements and attributes• Order of elements is important |
|--|---|

4

Approaches to Storing XML

- **Storing XML on File Systems**
 - Document Size
 - Updates
 - Indexes
 - Building Your Own
- **Using XML With Conventional Databases**
 - Producing XML from Relational Databases
 - Moving XML to Relational Databases
 - Shredding
 - Data Binding
 - Representation mapping between XML and RDBs
- **Native XML Databases**
 - eXist – most popular open-source XML DB

5

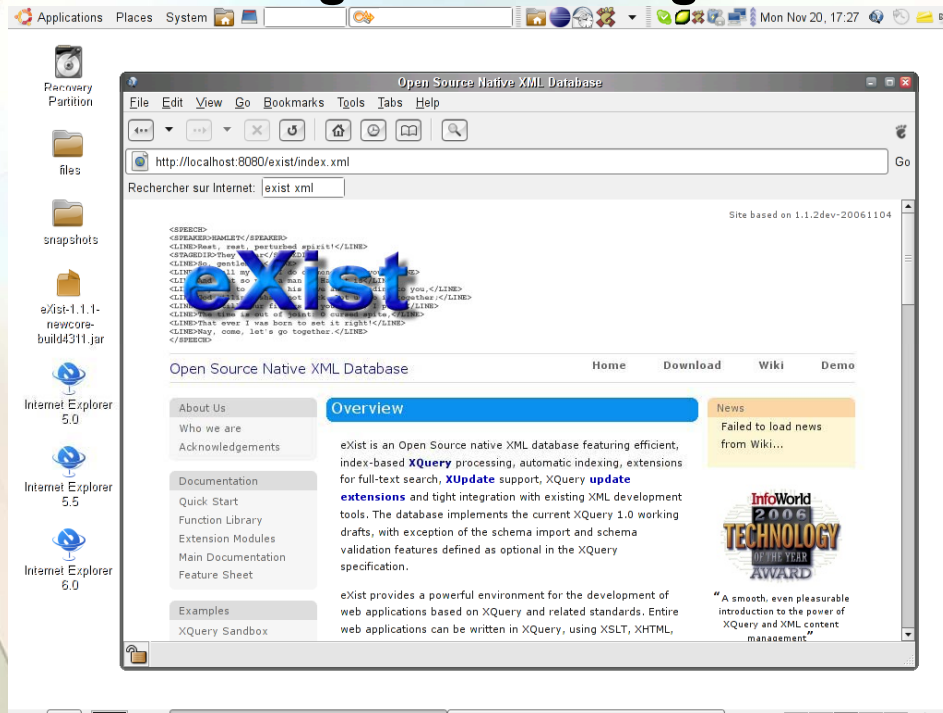
Aspects of Using Native XML Databases

- **Storage**
 - Large string
 - Binary object (BLOB or CLOB)
 - “XML” data type
- **Concurrency, locking, and transactional control**
- **Reading native XML DBs**
 - XQuery, etc.
- **Changing content in native XML DBs**
 - XQuery 1.0 can't do it
 - XQuery 1.0 Update facility
 - XQuery 1.1

6

Using Native XML Databases

- **Obtaining and Installing eXist**



7

Using eXist

- **As a Java library to embed database server in Java application**
- **As a standalone database server**
- **Embedded in Web server, both**
 - **Accessible via Web interface**
 - **Functioning as standalone database**

8

Accessing eXist

- eXist Web Interface *
- eXist Client *
- XML:DB API
 - Like JDBC for relational DBs
- XML RPC (Remote Procedure Call)
- Web Services
 - SOAP (originally Simple Object Access Protocol)
 - REST * (Representational State Transfer)
- Atom Publishing Protocol (APP)

**Which Interface?
It's up to you.**

9

Using the eXist Web Interface

Applications Places System

Recovery Partition

files

snapshots

exist-1.1.1-newcore-build4311.jar

Internet Explorer 5.0

Internet Explorer 5.5

Internet Explorer 6.0

eXist Database Administration

File Edit View Go Bookmarks Tools Tabs Help

http://localhost:8080/exist/admin/admin.xql?panel=browse

Go

Rechercher sur Internet:

User: admin

Select a Page

- Home
- System Status
- Browse Collections
- Manage Users
- Examples Setup
- Shutdown
- Logout

Logged in as: admin

Browsing Collection: /db

	Name	Permissions	Owner	Group	Created	Modified	Size (KB)
	Up						
<input type="checkbox"/>	rss	rwur-ur-u	guest	guest	Nov 20 2006 16:32:09		
<input type="checkbox"/>	sandbox	rwur-ur-u	guest	guest	Nov 20 2006 17:37:34		
<input type="checkbox"/>	system	rwurwu---	admin	dba	Nov 20 2006 16:30:55		

Remove Selected

Create Collection

Upload

New collection:

Store as:

Browse...

10

Using the Sandbox

The screenshot shows a web browser window titled "XQuery Sandbox" at the URL `http://localhost:8080/exist/sandbox/sandbox.xql#`. The interface includes a menu bar (File, Edit, View, Go, Bookmarks, Tools, Tabs, Help), a search bar, and a "Rechercher sur Internet:" field. Below the search bar are links for Home, Download, Wiki, and Demo. The main area is divided into two panes. The left pane, titled "Slots", contains a list of 10 slots, with the first three containing XQuery snippets. The right pane, titled "Paste saved query:", contains a text area with an XQuery snippet. Below the text area are buttons for Send, Clear, and Check, and a "Display:" dropdown set to 20. The bottom section shows the results of the query, indicating "Found 1 in 0.028 seconds" and "Showing items 1 to 1". The result is an XML snippet.

```
for $item in /item
where ../a[contains(@href, 'wrox.com')]
return <match>
  <id>{string($item/@id)}</id>
  { $item/title }
  { $item/..a[contains(@href, 'wrox.com')] }
</match>
```

Found 1 in 0.028 seconds. Showing items 1 to 1

```
<match>
  <id>1</id>
  <title>Working on Beginning XML</title>
  <a href="http://www.wrox.com/WileyCDA/WroxTitle/productCd-0764570773.html">
    
  </a>
  <a href="http://www.wrox.com/WileyCDA/WroxTitle/productCd-0764570773.html">WROX's excellent "Beginning XML".</a>
</match>
```

11

Using the eXist Client

The screenshot shows the "Query Dialog" window in the eXist Client. The "Query Input" tab is active, displaying an XQuery snippet. The "History" list shows the current query. The "Context" is set to `/db/blog` and "Display max:" is set to 100. The "Submit" button is visible. The "Results" section shows the XML output of the query, which is an XML snippet.

```
for $item in /item.
where ../a[contains(@href, 'wrox.com')].
return <match> .
  <id>{string($item/@id)}</id>.
  { $item/title }.
  { $item/..a[contains(@href, 'wrox.com')] }.
</match>.
```

Context: /db/blog Display max: 100 Submit

Results:

```
<match> .
  <id>1</id> .
  <title>Working on Beginning XML</title> .
  <a href="http://www.wrox.com/WileyCDA/WroxTitle/productCd-0764570773.html">
    
  </a> .
  <a href="http://www.wrox.com/WileyCDA/WroxTitle/productCd-0764570773.html">WROX's excellent "Beginning XML".</a> .
</match>.
```

Found 1 items. Compilation: 33ms, Execution: 23ms

12

XML in Commercial/Open Source RDBMS

- Huge volumes in RDBMSs
- XML-enabled databases
- Retrieved into XML based files
- Varying vendor solutions
- RDBMS enabled with XML
 - IBM DB2
 - Oracle
 - Sybase
 - MySQL
 - Microsoft SQL Server

13

Looking Ahead

- Time to give up Relational Databases?
- No standard yet
- Waiting on XQuery
- W3C is involved

Stayed tuned and informed.

14