Chapter 10: XML and Databases

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

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?

Comparing XML-Based Data and Relational Data

- Relational
  - Structured
  - No hierarchy
  - Table: rows and columns
  - Order of rows in table not important
  - Data in tables associated using keys
- XML-Based
  - Semi-structured as well as structured
  - Heirarchical,
    - XPath Model
    - Document Object Model
    - XML Infoset
  - Elements and attributes
  - Order of elements is important

Approaches to Storing XML

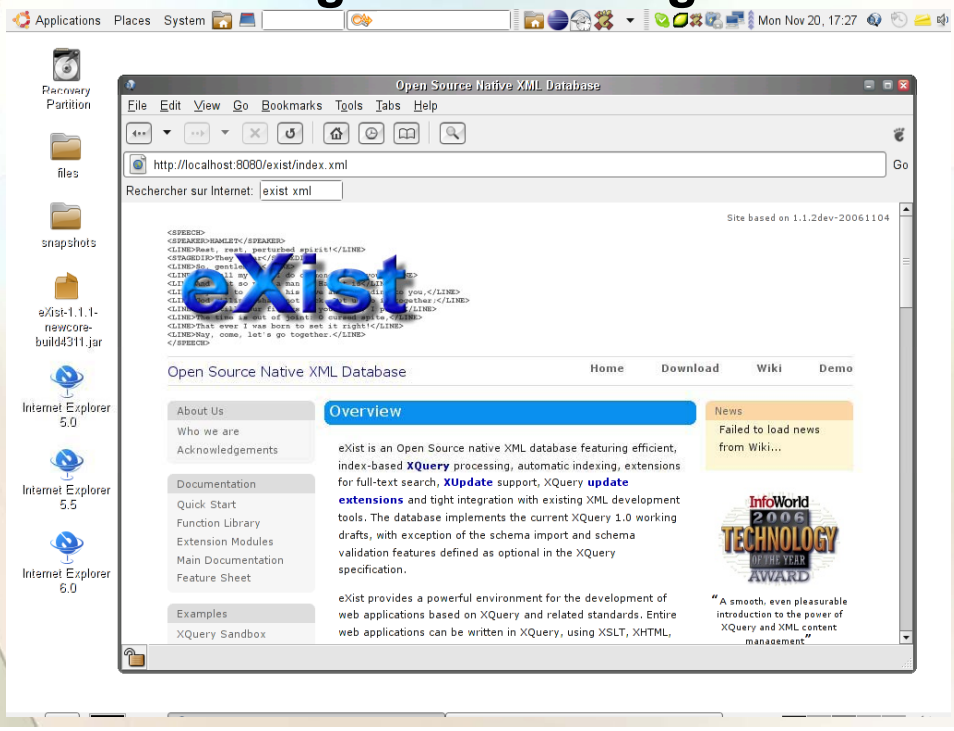
- Storing XML on File Systems
  - Document Size
  - Updates
  - Indexes
  - Building Your Own
- Using XML With Conventional Databases
  - Producing XML from Relational Databases
  - Moving XML to Relational Databases
    - Shredding
  - Data Binding
    - Representation mapping between XML and RDBs
-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

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

Using Native XML Databases

- Obtaining and Installing eXist



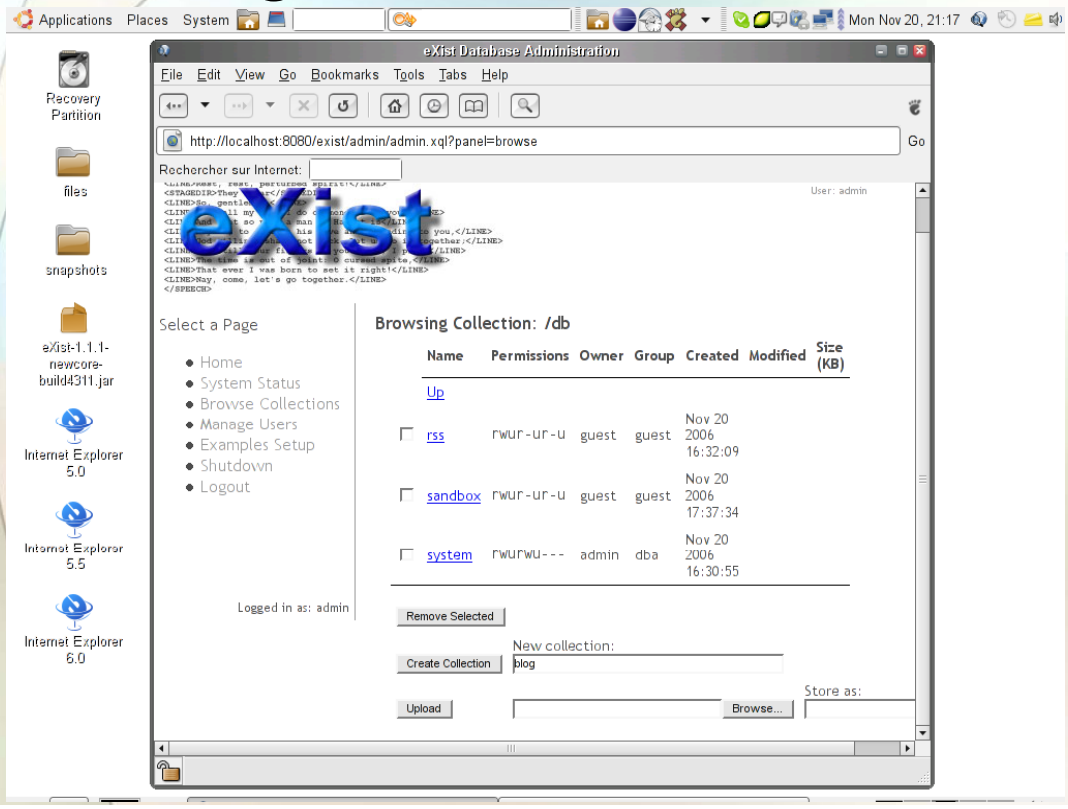
Using eXist

- 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

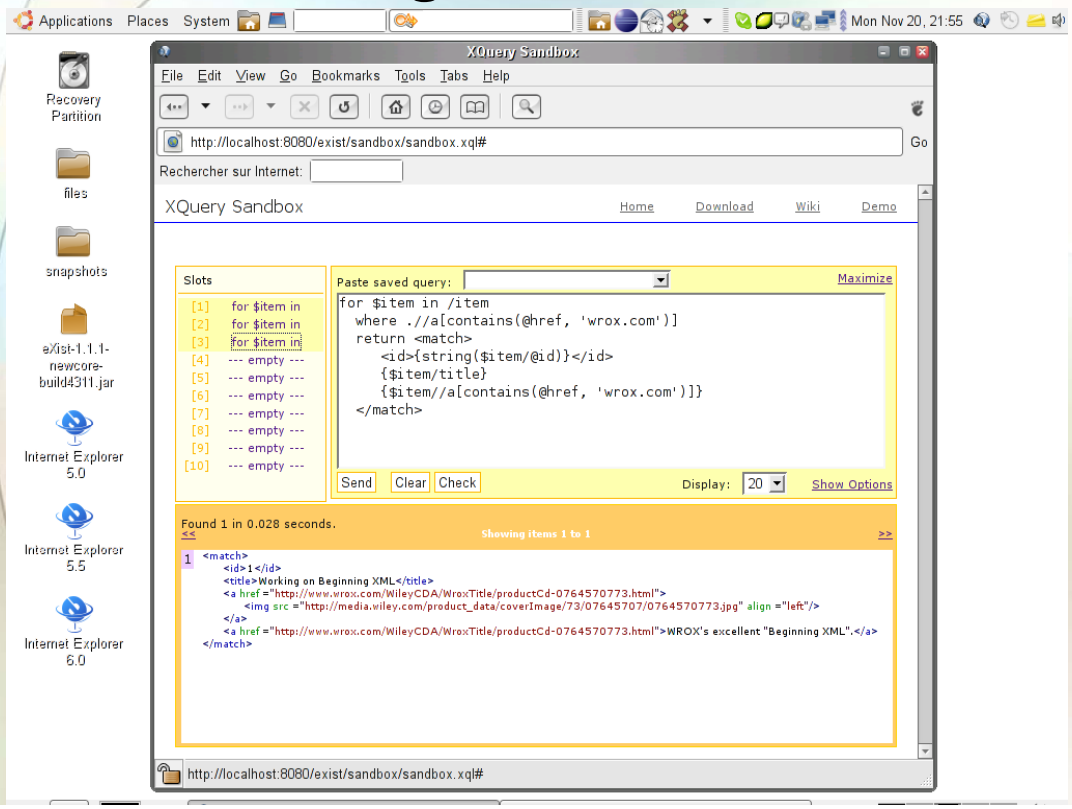
Accessing eXist

- 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)

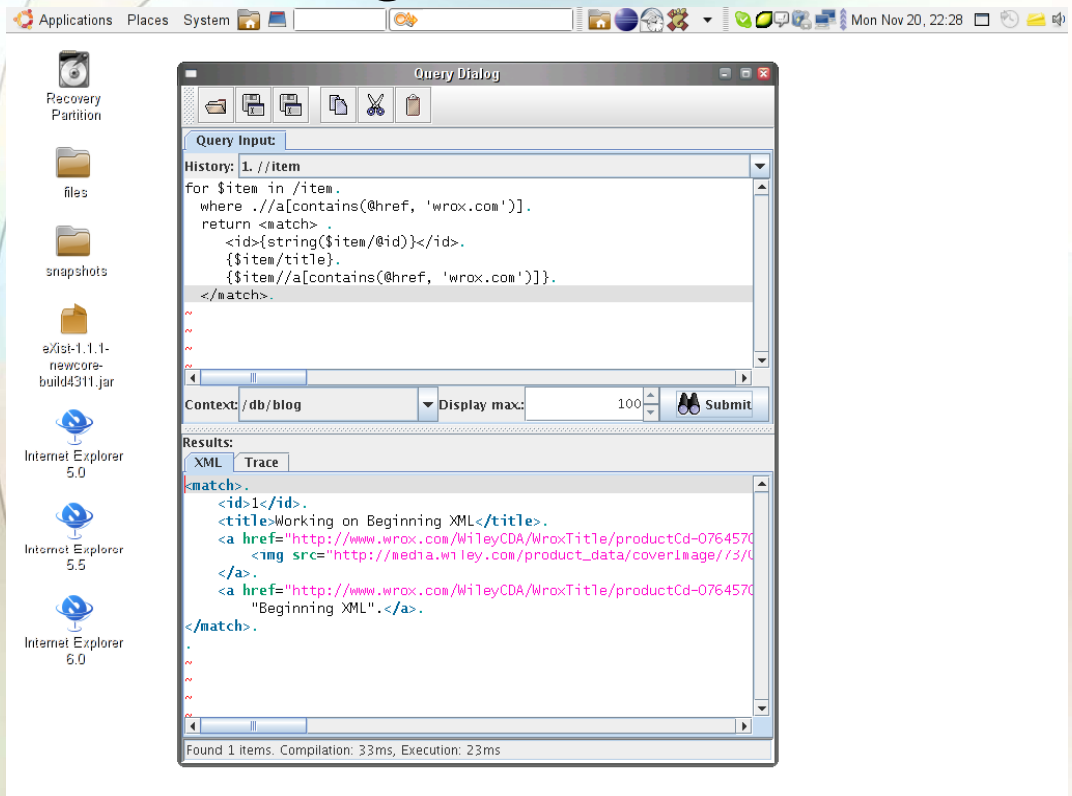
Using the eXist Web Interface



Using the Sandbox



Using the eXist Client



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

Looking Ahead

- Time to give up Relational Databases?

- No standard yet
- Waiting on XQuery
- W3C is involved