Lecture 0 (1 hour)

XML and DataBase

Sang Shin
Java™ Technology Evangelist
sang.shin@sun.com

(You can use this material in any way you want, but if you can drop me an email when you do, that will be greatly appreciated.)

Topics

- Motivation
- Similarities and differences between XML and Database
- XML database types
- Issues of storing XML data in RDBMS
- Products

Motivation

- Applications want to deal with data in XML form
- Need for XML data/documents management
  - Store, Retrieve, Search, Update
- Database has been used successfully for data management
- So how an we leverage them both?

Similarities with Database

- Storage: Tables vs. XML documents
- Schemas: Database schema vs. DTD, XML schema languages
- Query languages: SQL vs. XQL, XML-QL, QUILT
- Programming interface: JDBC vs. SAX, DOM

Differences from Database

- Storage mechanism
- Indexing
- Built-in security
- Transaction support
- Data integrity
- Multi-user access
- Query across multiple data sources

Data Management vs. Document Management

- Data management
  - Need database tuned for data storage
  - Relational or Object database
- Document management
  - Need content management system
  - Content management system might be built over databases

XML Database Types

- XML enabled database
  - XML is used for input and output
  - Relational tables internally
  - Middleware converts between XML and relational database tables
- Native XML database
  - XML data is stored in its native format

XML Enabled Database

- Pros
  - Proven database technology
- Cons
  - Conversion between XML and relational tables is needed
    - Conversion performance overhead
    - Complex XML hierarchy information is hard to convert to tables
    - Round-tripping to exactly same document is hard
  - Might not handle Unicode well enough

Native XML Database

- Pros
  - Preserves the XML hierarchy information
- Cons
  - Not proven yet
  - Scalability concern
- When to use
  - to integrate information from many different platforms and formats and send it to business partners or customers

Mapping XML Structure to DB Structure

- Template-driven mapping
- Model-driven mapping
  - Table model
  - Data-specific object model

Template-driven Mapping

- No predefined mapping
- Embed commands in a template
- Pocessed by middleware
- Very flexible

Example: Template-driven Mapping - Template

```
<?xml version="1.0"?>

  <FlightInfo>
    <Intro>The following flights have available seats:</Intro>

    <SelectStmt>
    SELECT Airline, FltNumber, Depart, Arrive FROM Flights
    </SelectStmt>

    <Conclude>We hope one of these meets your
   needs</Conclude>
  </FlightInfo>
```

Example: Template-driven Mapping - XML Structure

```
<?xml version="1.0"?>
  <FlightInfo>
    <Intro>The following flights have available seats:</Intro>
    <Flights>
      <Row>
        <Airline>ACME</Airline>
        <FltNumber>123</FltNumber>
        <Depart>Dec 12, 1998 13:43</Depart>
        <Arrive>Dec 13, 1998 01:21</Arrive>
      </Row>
      <!--  More Row's -->
    </Flights>
    <Conclude>We hope one of these meets you needs</Conclude>
  </FlightInfo>
```

Model-Driven Mapping

- Data model is imposed on the structure of the XML document
- Not as flexible as Template-model
  - Typically used with XSLT to compensate it
- Simple
- Two models
  - Table model
  - Data-specific object model

Table Model

- Models the XML document as a single table or set of tables
- Most popular for relational database

```
<database>
    <table>
      <row>
        <column1>...</column1>
        <column2>...</column2>
        ...
      </row>
      ...
    </table>
    ...
</database>
```

Data-specific Object Model

- Direct mapping between XML elements and data
- Popular with object or hierarchical databases

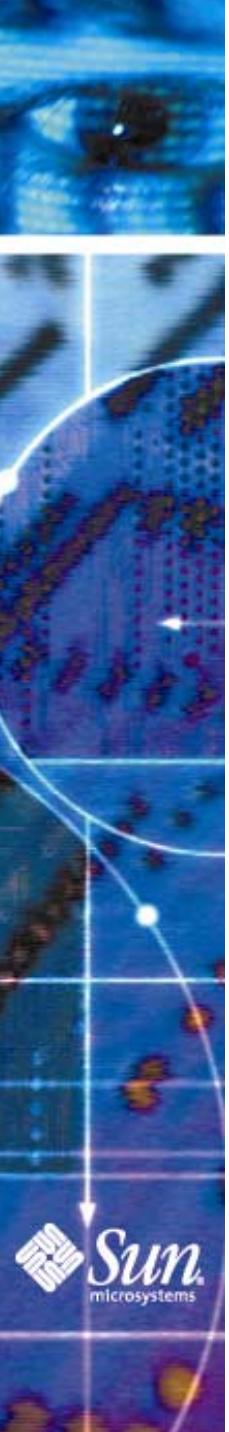
Example: Data-specific Object Model

Issues of Storing XML data into Relational Database

- Data types
- Binary data
- Character set
- Processing instruction
- Storing markup
- DTD generation from DB schema and vice versa

Data Types

- All data in XML document is text
  - Need to be translated to datatypes of database
- Issues
  - Translation is not always easy



Binary Data

- Two ways of storing binary data in XML document
  - Unparsed entities
  - Base64 encoding
- Issues
  - No XML standard notation for Base64-encoded data
  - Application specific notation is needed
    - Could be lost during translation

Character Set

- Unicode is native encoding scheme of XML document
- Issues
  - Databases might not support Unicode

Storing Markup

- Storing markup information of XML document could be useful
- Issues
  - Metadata describing markup data need to be created and maintained

Advanced Features

- Indexing when storing XML documents
  - Allows faster search

Products

- Content@XML from Xyvision
  - Content management system that stores XML documents in any relational database
- XML-DBMS
  - Translator between XML documents and RDBMS

Products: Native XML Database

- Tamino from Software AG
- dbXML
- eXcelon Data server from eXcelon
- X-Hive/DB from Rotterdam

Summary

- Motivation
- Similarities and differences between XML and Database
- XML database types
- Issues of storing XML data in RDBMS
- Products

References

- Article “XML and Databases”, written by Ronald Bourret, Nov. 2000
- Article “XML Enters The DBMS Area”, written by Edmund Dejesus, Oct. 2000