

**DB2** Data Management Software



 e-business software

## XML and DB2

Tiina Pihlasviita

[tiina.pihlasviita@fi.ibm.com](mailto:tiina.pihlasviita@fi.ibm.com)

IBM Software Group

# Putting DB2's Power to Support XML

---

- DB2 provides stability, scalability and security
- Your mission-critical business data is currently stored in DB2
- DB2's XML support adds its power to enable your business data for XML
- With DB2's XML support, you can create applications to:
  - Store XML documents for fast search
  - compose or decompose XML documents from or into relational tables
  - Query XML data based on XPath expressions
- You can build B2B and B2C applications on top of DB2 using XML Extender

**DB2** Data Management Software

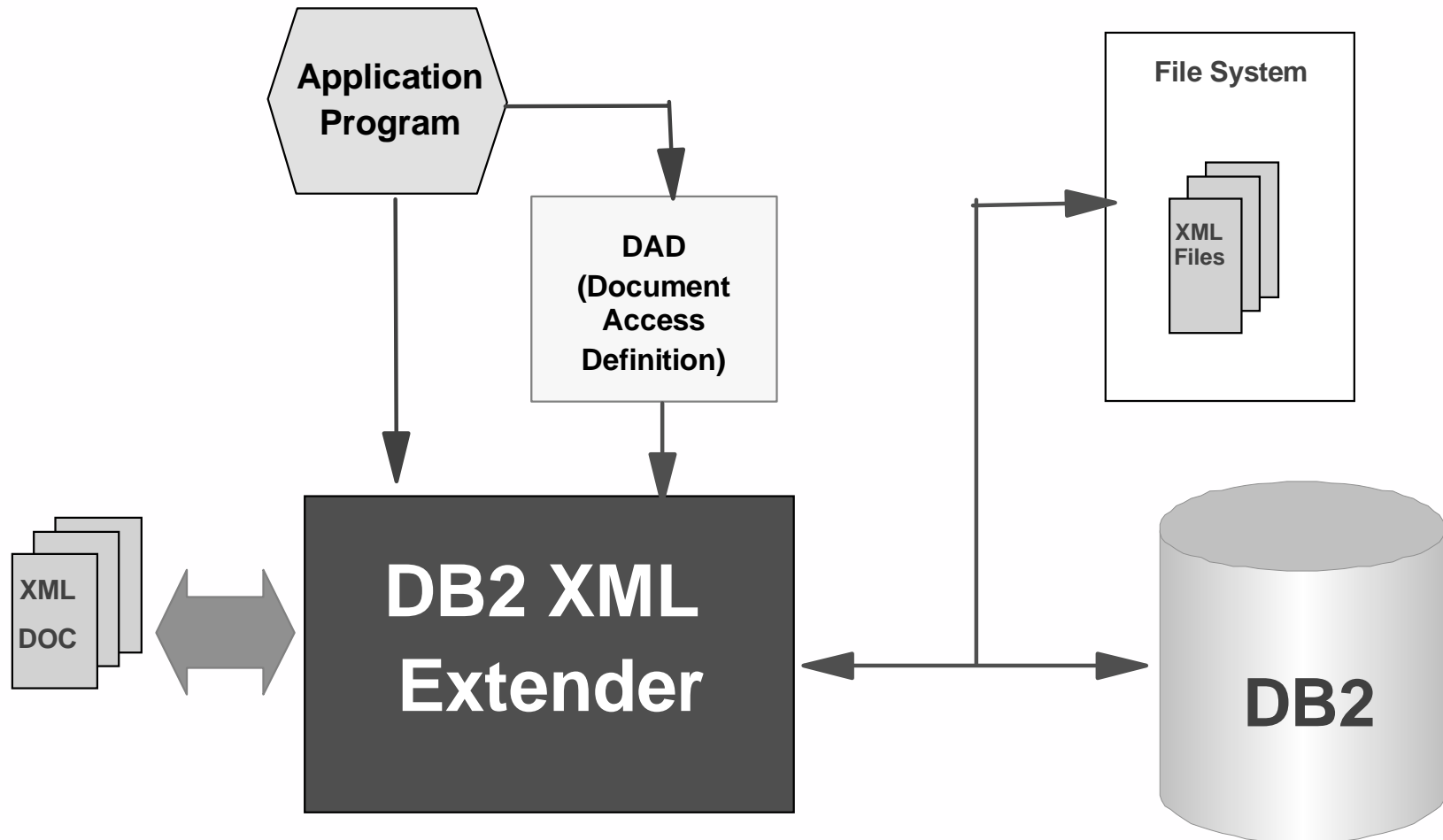


# XML Extender Architecture

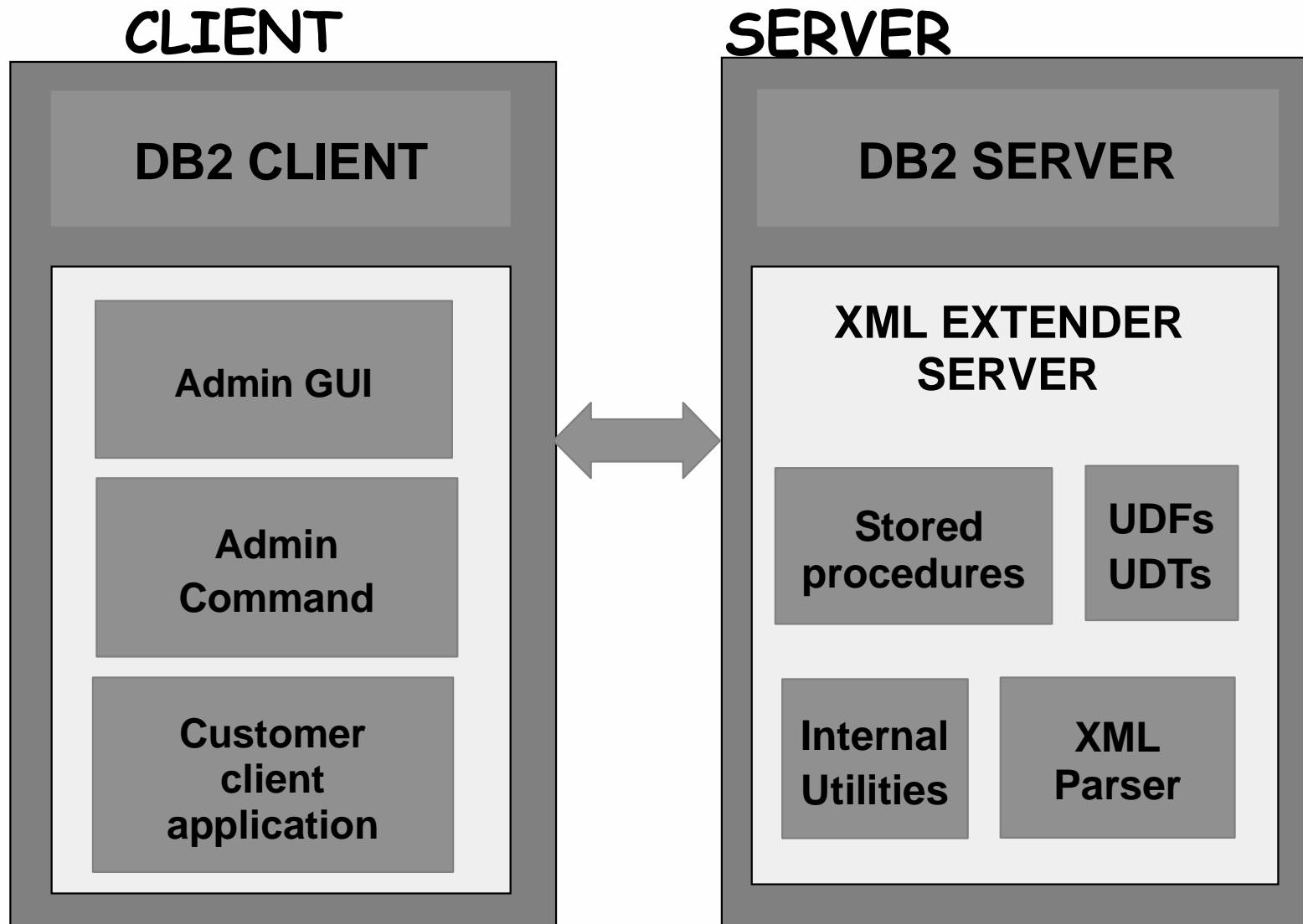
IBM Software Group

# XML Extender Overview

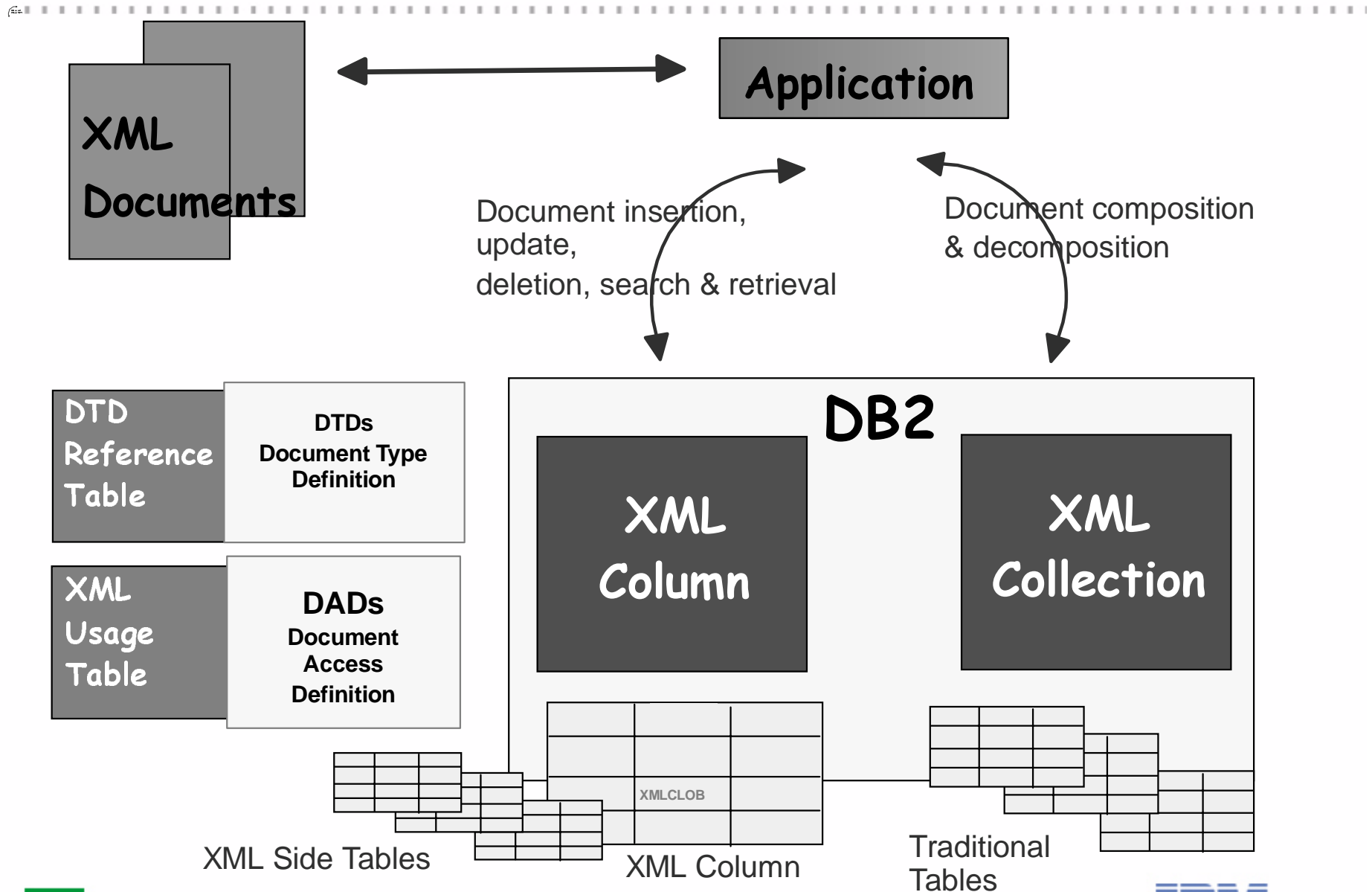
---



# XML Extender Architecture



# DB2: Two Access and Storage Methods



# DAD: Document Access Definition

- XML document itself
- Define the mapping between XML document and relational tables
- For XML Column, it largely defines how documents are indexed by defining which elements are extracted into side tables
- For XML Collections, it maps the structure of the XML document to DB2 tables and columns.
- Wizard available to assist DAD creation

```
<DAD>
<dtdid>c:\dxx\samples\dtd\getstart.dtd</dtdid>
<validation>YES</validation>
<Xcolumn>
  <table name="order_side_tab">
    <column name="order_key"
      type="integer"
      path="/Order/@key"
      multi_occurrence="NO"/>
    <column name="customer"
      type="varchar(50)"
      path="/Order/Customer/Name"
      multi_occurrence="NO"/>
  </table>
  <table name="part_side_tab">
    <column name="price"
      type="decimal(10,2)"
      path="/Order/Part/ExtendedPrice"
      multi_occurrence="YES"/>
  </table>
</Xcolumn>
</DAD>
```

# XML Column

## ■ XML Column:

XML documents stored intact as DB2 column data

with optional validation against DTDs

Supplied User Defined Types

XMLVarchar, XMLCLOB, XMLFile

Supplied User Defined Functions:

To store or retrieve entire XML documents into/from DB2

To extract or update XML elements or attribute values using XPath notation

Document Access Definition: To define special indexes {side tables}

for speedy structured search of XML document content

for use in conjunction with regular SQL data

Can also use text extender text search facilities

```
<?xml version="1.0" encoding="UTF-8"
standalone>
<!DOCTYPE personnelRec SYSTEM "prml.dtd">
<!-- This is a comment -->
<personnelRec>
  <person salary="26350.00" band="D">
    <name>
      <family>Wallace</family>
      <given>Bob</given>
    </name>
    <email>bwallace@megacorp.com</email>
    <dept>&d1</dept>
  </person>
</personnelRec>
```



XMLColumn	
	<pre>&lt;?xml version="1.0" encoding="UTF-8" standalone&gt; &lt;!DOCTYPE personnelRec SYSTEM "prml.dtd"&gt; &lt;!-- This is a comment --&gt; &lt;personnelRec&gt;   &lt;person salary="26350.00" band="D"&gt;     &lt;name&gt;       &lt;family&gt;Wallace&lt;/family&gt;       &lt;given&gt;Bob&lt;/given&gt;     &lt;/name&gt;     &lt;email&gt;bwallace@megacorp.com&lt;/email&gt;     &lt;dept&gt;&amp;d1&lt;/dept&gt;   &lt;/person&gt; &lt;/personnelRec&gt;</pre>

# XML Columns - Side Tables

---

- Increases performance of searches
- Creates DB2 tables to store indexes and frequently searched data
- The tables are synchronized with the XML document with DB2 triggers
- Requires a Document Access Definition (DAD) file to define the mapping between the XML document and the database table
- Cannot enable a XML Column without a side table

# Using Side Tables for Fast Searches in XML Documents

## XMLColumn

```
<order key='99'>
  <customer>Thompson</customer>
  <part key='82' > .... </part>
  <part key='83' > .... </part>
</order>
```

DAD  
(Document  
Access  
Definition)

## Sales\_tab

...	Order (XMLCLOB)	...
...		...

## side tables

### order\_side\_tab

order_key	customer
99	Thompson

### part\_side\_tab

part_key	....
82	....
83	....

Index

Index

Index

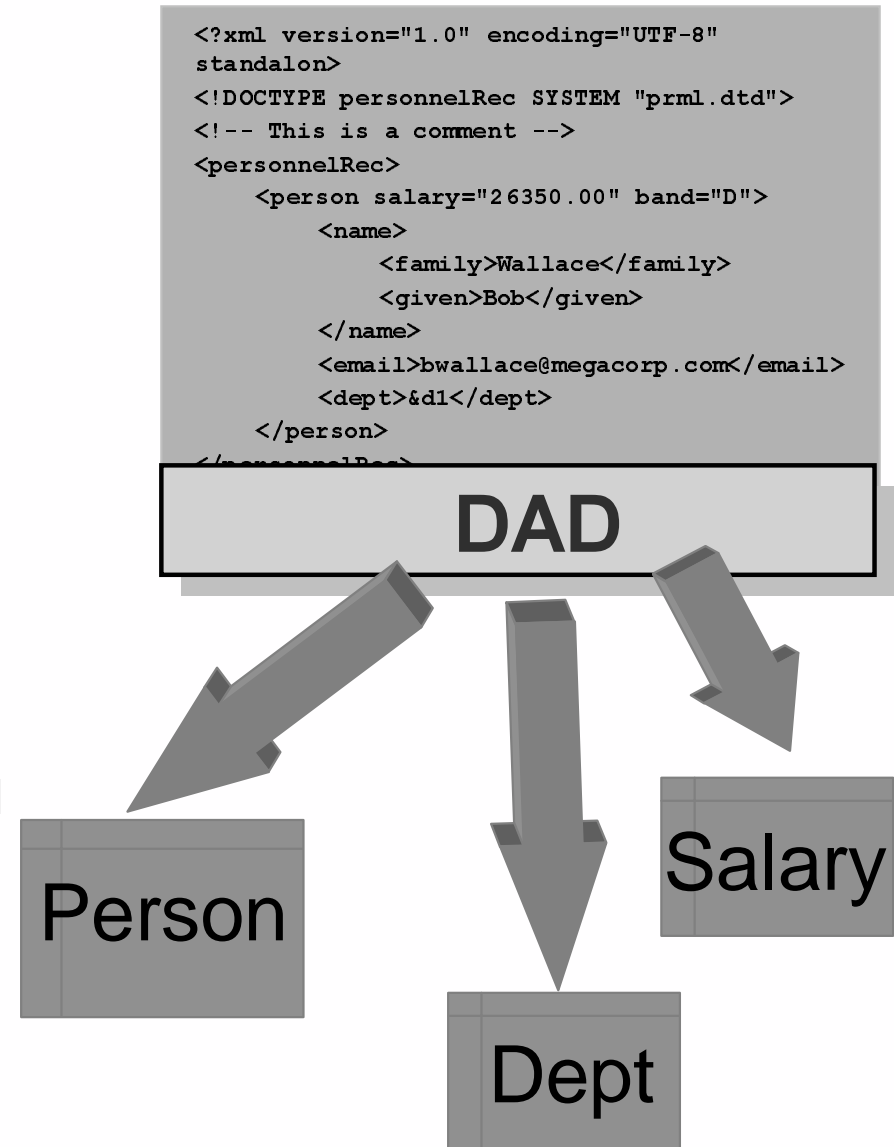
# XML Collection

## ■ XML Collection:

XML documents decomposed  
XML elements and attributes  
stored as SQL data types

XML documents composed or  
reconstructed  
from regular SQL content held in  
traditional tables  
with optional validation against  
DTDs

Document Access Definition: To  
define mappings between relational  
and XML data



# XML Collection - Two Mapping Schemes

## ■ SQL mapping (composition only)

SQL syntax

specify the SQL select statement

specify the mapping between column and XML data in attribute\_node and text\_node

certain restrictions apply

## ■ RDB\_node mapping (comp. & decomp.)

XML syntax

specify tables and relationship among tables in the RDB\_node of the root element\_node

specify RDB\_node with table name, column name and optional condition for attribute\_node and text\_node

```
<DAD>
<dtdid>c:\dxx\samples\dtd\getstart.dtd</dtdid>
<validation>YES</validation>
<Xcollection>
  <SQL_stmt>
    SELECT .....
  </SQL_stmt>
  .....
</Xcollection>
</DAD>
```

```
<DAD>
.....
<Xcollection>
.....
<element_node name="Order">
  <RDB_node>
    <table name="order_tab"/>
    <table name="part_tab"/>
    <condition>
      order_tab.order_key =
part_tab.part_key
    </condition>
  </RDB_node>
```

# Use Text Extender on Structural Text Search

---

- XML column or decomposed column in sub-tables can be enabled with Text Extender

perform structural and full text searches

- Use Text Extender section support of structural text search in query

allows matching within a specific document context

- Example:

```
SELECT order FROM sales_table
WHERE CONTAINS(orderHandle,
'MODEL order SECTION(/Order/Customer)
"John Doe" ') = 1
```

- No dependency on Text Extender if searching on SQL data types of XML element/attribute values is sufficient

# XML Columns vs XML Collections

---

## ■ Reasons for choosing XML Column:

XML documents already exist

Wish to store intact XML documents

Know which elements or attributes will be frequently searched

XML documents are frequently read but infrequently updated

Wish to keep XML documents external to DB2 on a file system

## ■ Reasons for Choosing XML Collection:

You need XML documents to be generated from your DB2 tables

You do not want to store intact XML documents, but only un-tagged data

You know what column data in database to be used to generate XML documents

You need to update small parts of XML documents often

Performance of update is critical

# Dynamic Mapping

---

- Ability to take DAD to generate or shred XML documents, where DAD can be created by applications on the fly.
- Change query criteria via overriding query condition for document generation

## SQL\_OVERRIDE:

replace the SQL query  
change conditions in the WHERE clause  
designed for DBMS experts

## XML\_OVERRIDE:

Use XPath syntax to define element or attribute,  
Specify the constraints on XML elements or attributes on the documents to be generated

An example:

```
"/department/@id='E01' AND /department/name='John Duo' "
```

**DB2** Data Management Software



Other Features

IBM Software Group



# Other Features

---

## ■ DTD Repository

Store your DTDs in DB2,

One DTD can be referenced by many XML documents and more than one XML column and XML collection.

Use DTDID to validate input XML documents at the insertion time

## ■ Validating Input or Generated XML Documents

Validation can be specified in your DAD,

Supported in both XML Column and Collection features

## ■ GUI based Administration Tool

Help to enable your database, tables, column, collection for XML

Help to create DAD

# Document Type Definition - DTD

---

- **DTD specifies the structure of an XML document**

  - XML parsers can then understand and interpret the contents

  - contains a list of allowed tags with their types and attributes

  - can be stored in a separate file, or within the XML document itself

- **It defines how elements relate to one another within the documents tree structure**

- **DTD's are used to determine the validity of an XML document**

- **DTD's are used to perform structured searches on the XML data**

- **DTD's can be shared across organisations**

# Enterprise Integration

## ■ Challenge

- ▶ Integrate islands of information to improve business effectiveness.
  - home grown, legacy applications
  - packaged apps
  - apps acquired through acquisitions

## ■ Solution

- ▶ Leverage XML as the "lingua franca" to integrate and exchange information across the enterprise.
- ▶ Build enterprise models around XML to be used by new applications.
- ▶ Provide portals with a service-oriented architecture.

## ■ Value

- ▶ Extract competitive value from existing information.
- ▶ Increase profitability through operational efficiencies.
- ▶ Shorten business cycle (decision making process)

## ■ Technical Requirements

- ▶ Integration of structured and unstructured data (including appl. data, e.g., SAP)
- ▶ Application and process integration (messaging, workflow)
- ▶ XML repository (XQuery, style sheet, XSchema, ...)
- ▶ Mining of text data (categorization, search, ...)

# DB2 Integrator Extends DM

---

- DB2 UDB is optimized for high performance transaction and warehousing support of structured information
  - ▶ supports replication, ETML, analysis and web services for structured information
  - ▶ provides XML based interfaces (e.g. SOAP, XQuery, SQLX) to support information interchange
- Content Manager is optimized for management of unstructured content
  - ▶ supports text search, workflow, versioning for unstructured information
  - ▶ provides XML based interfaces (e.g. SOAP, XQuery) for information interchange
- DB2 Integrator is optimized to integrate structured, unstructured and XML formatted information
  - ▶ supports federation, analysis, ETML, crawling & workflow/messaging integration and web services across structured & unstructured & XML information
  - ▶ provides scalable, high-performance (native) XML storage for information integration



# Going Forward

---

- What is coming out in 2003
  - ▶ Broader capabilities in handling, integrating structured and unstructured data and XML
  - ▶ Access to industry-specific data sources
  - ▶ Expanded outreach to ISVs, SIs
- Strategic direction
  - ▶ XML data store for high volume applications

# More information

## ■ DB2 UDB XML Extender Web site:

<http://www.software.ibm.com/data/db2/extenders/xmlext/index.html>

<http://www.software.ibm.com/data/db2/extenders/xmlext/library.html>

## ■ Redbooks ([www.redbooks.ibm.com](http://www.redbooks.ibm.com)):

Integrating XML with DB2 XML Extender and DB2 Text Extender  
(SG24-6130-00)

## ■ Tools:

<http://xml.apache.org/index.html>

<http://www.alphaworks.ibm.com/tech> => XML

## ■ IBM Data Management:

<http://www.fi.ibm.com/software/data/index.html>

## ■ Xperanto

<http://xperanto.dfw.ibm.com/demo/>