



Podpora XML v databázích

J. Zendulka



Obsah

- Typy XML dokumentů
- XML databázové produkty
- Standard SQL a podpora XML
- Podpora XML u serveru Oracle11g (Oracle XML DB)
- Podpora formátu JSON



Typy XML dokumentů (1)

- *Datové XML dokumenty* (data-centric XML documents)
 - XML použit pro zápis dat (typicky k přenosu)
 - Typický životní cyklus: DB → XML → DB
 - Pravidelná struktura, jemná granularita, nezáleží na pořadí dat (rozhodující název elementu/atributu)
 - Příklady: faktury, objednávky; dokumenty vytvořené podle šablony; export dat z DB
 - Odpovídající aplikace – **datově zaměřené**.

Typy XML dokumentů (2)

■ Př)

```
<?xml version="1.0" encoding="iso-8859-2"?>
<FAKTURA cislo_faktury="10022006/159">
  <DATUM_VYSTAVENI>1.4.2006</DATUM_VYSTAVENI>
  <DATUM_SPLATNOSTI>1.5.2006</DATUM_SPLATNOSTI>
  <DODAVATEL>
    <JMENO>Jan Novák</JMENO>
    <ULICE>Božetěchova 2</ULICE>
    <MESTO>Brno</MESTO>
    <PSC>61200</PSC>
    <ICO>12345678</ICO>
  </DODAVATEL>
  ...
</FAKTURA>
```



Typy XML dokumentů (3)

- *Dokumentově zaměřené XML dokumenty*
(document-centric XML documents)
 - Často určen pro čtení či zpracování lidmi
 - Méně pravidelná struktura, pořadí elementů zásadní.
 - Data zpravidla nepochází z databází
 - Příklady: Knihy, e-maily, WWW stránky v XHTML
 - Odpovídající aplikace – *dokumentově zaměřené*.

Typy XML dokumentů (4)

■ Př)

```
<?xml version="1.0" encoding="iso-8859-2"?>
```

```
<PRIRUCKA>
```

```
<NADPIS1>Uživatelská příručka k systému</NADPIS1>
```

```
<ODSTAVEC>Pokud nejste zaregistrován do systému a rád  
byste se zapsal na nějaký kurz, tak se musíte nejdříve
```

```
<ODKAZ url="/registrace">zaregistrovat</ODKAZ>.
```

```
Obdržíte tak své klientské číslo a heslo, pod kterými se  
budete přihlašovat.
```

```
</ODSTAVEC>
```

```
</PRIRUCKA>
```



Obsah

- Typy XML dokumentů
- XML databázové produkty
- Standard SQL a podpora XML
- Podpora XML u serveru Oracle11g (Oracle XML DB)
- Podpora formátu JSON



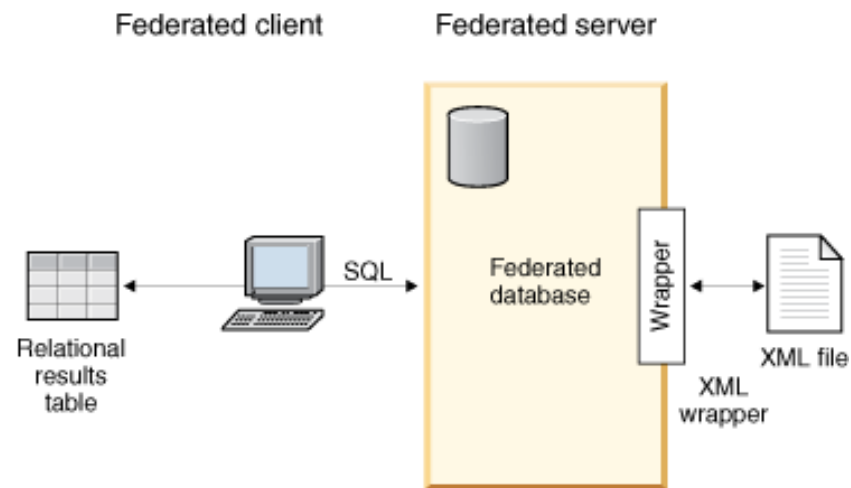
XML databázové produkty (1)

- Databázové systémy s podporou XML (XML - enabled)
 - DB → XML, XML → DB
 - většina komerčních relačních DB
 - pro datově zaměřené aplikace.
- Nativní XML databázové systémy
 - datový model vychází z modelu XML dokumentu (zpravidla varianta DOM mapovaná na příslušné úložiště) → rychlost
 - například Tamino (Software AG), eXist (open source <http://exist-db.org/exist/apps/demo/index.html>)
 - pro datově i dokumentově zaměřené aplikace

XML databázové produkty (2)

■ Wrappery

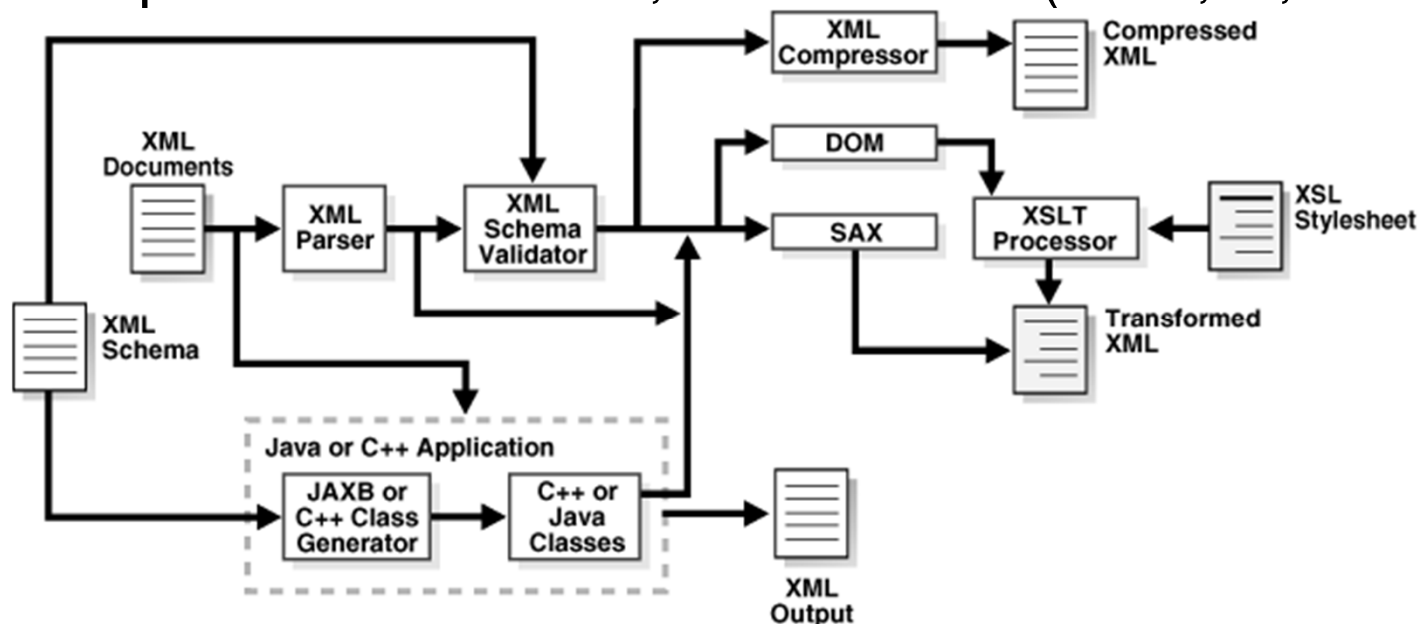
- programy umožňující relační pohled nad XML dokumentem nebo dokumenty → dotazování v SQL
- pro datově zaměřené aplikace
- např. wrapper systému IBM DB2 pro přístup k XML zdroji federované DB



XML databázové produkty (3)

■ Middleware

- prostředník mezi aplikacemi
DB → XML, XML → DB, podpora pro XML ne v DB
- pro datově zaměřené aplikace
- např. Microsoft ADO, Oracle XDK (Java, C, C++)

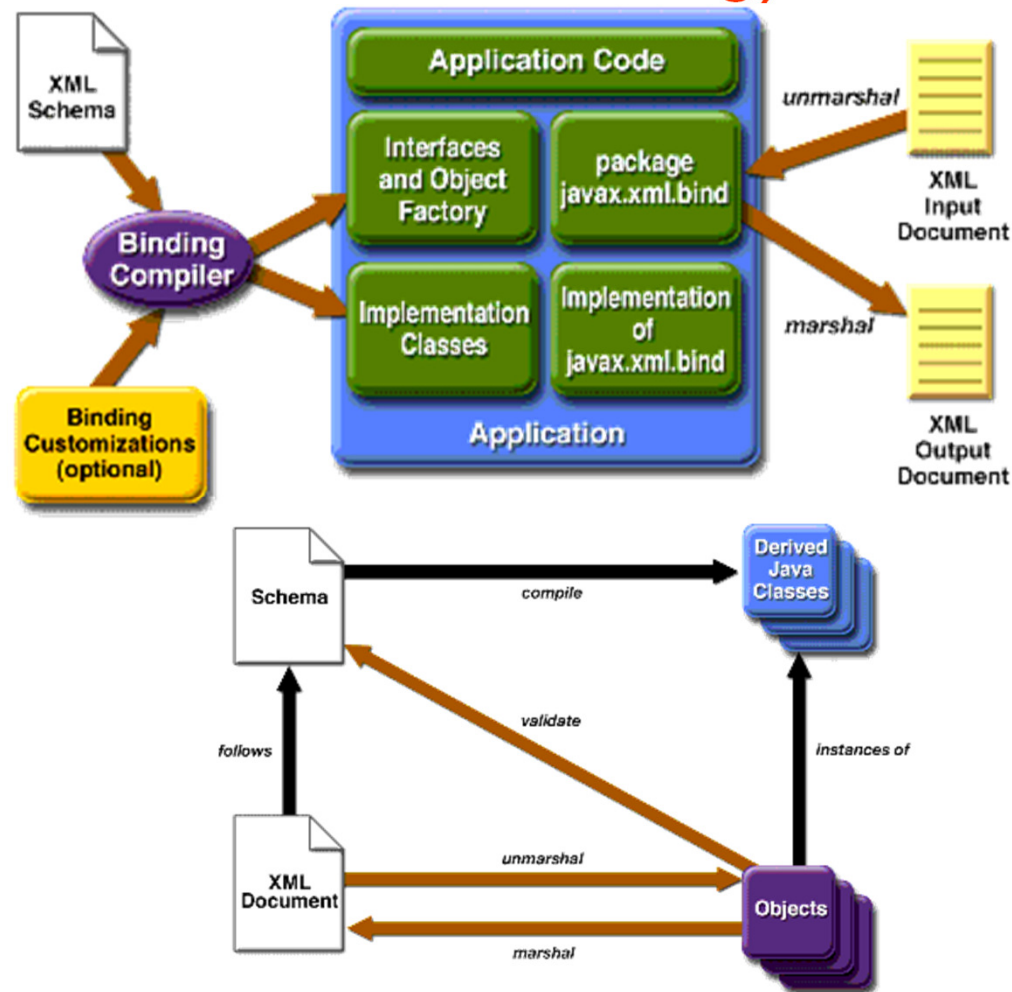




XML databázové produkty (4)

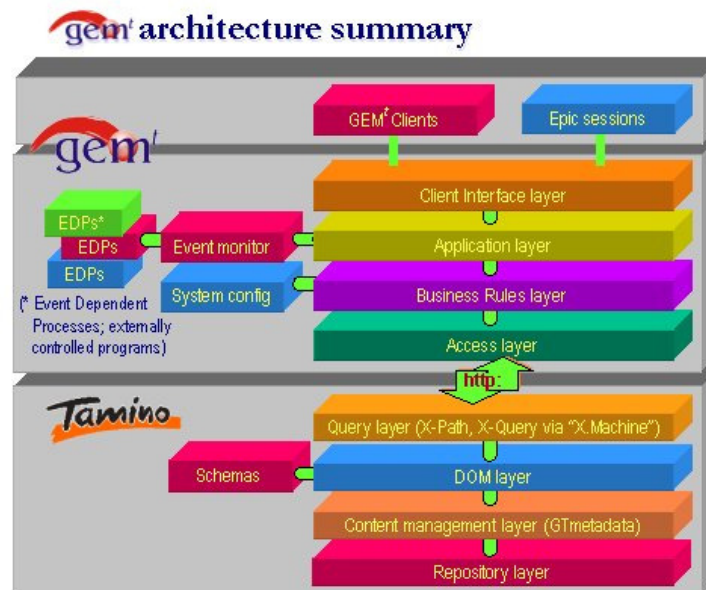
- Nástroje pro navázání XML dat (XML data binding)
 - z XML dokumentu vytvoří objekty zapouzdřující data, resp. z XML schématu třídy pro práci s dokumentem → bližší realitě než dle DOM
 - pro datově zaměřené aplikace
 - např. JAXB (Java Architecture for XML Binding)

Podstata JAXB (Java Architecture for XML Binding)



XML databázové produkty (5)

- Systémy pro správu obsahu (Content Management Systems)
 - nad nativní XML DB, souborovým systémem nebo RDB, prostředky pro kontroly, editaci, verzování atd.
 - pro dokumentově zaměřené aplikace
 - např. GEM^t (X.Systems)





Obsah

- Typy XML dokumentů
- XML databázové produkty
- **Standard SQL a podpora XML**
- Podpora XML u serveru Oracle10g (Oracle XML DB)
- Podpora formátu JSON



Standard SQL a podpora XML

- podpora XML zahrnuta do SQL:2003, tzv. SQL/XML
 - datový typ XML - standardní typ, hodnotami jsou 'XML hodnoty'
 - funkce pro publikování v XML (konstruktory hodnot typu XML) - lze použít v příkazech SQL
 - xmlelement(), xmlattributes(), xmlroot(), xmlparse(), xmlforest(), ...
 - pravidla mapování - mapování SQL a XML datových typů, mapování SQL a XML metadat
 - často nekompatibilita SQL/XML řešení u výrobců



Obsah

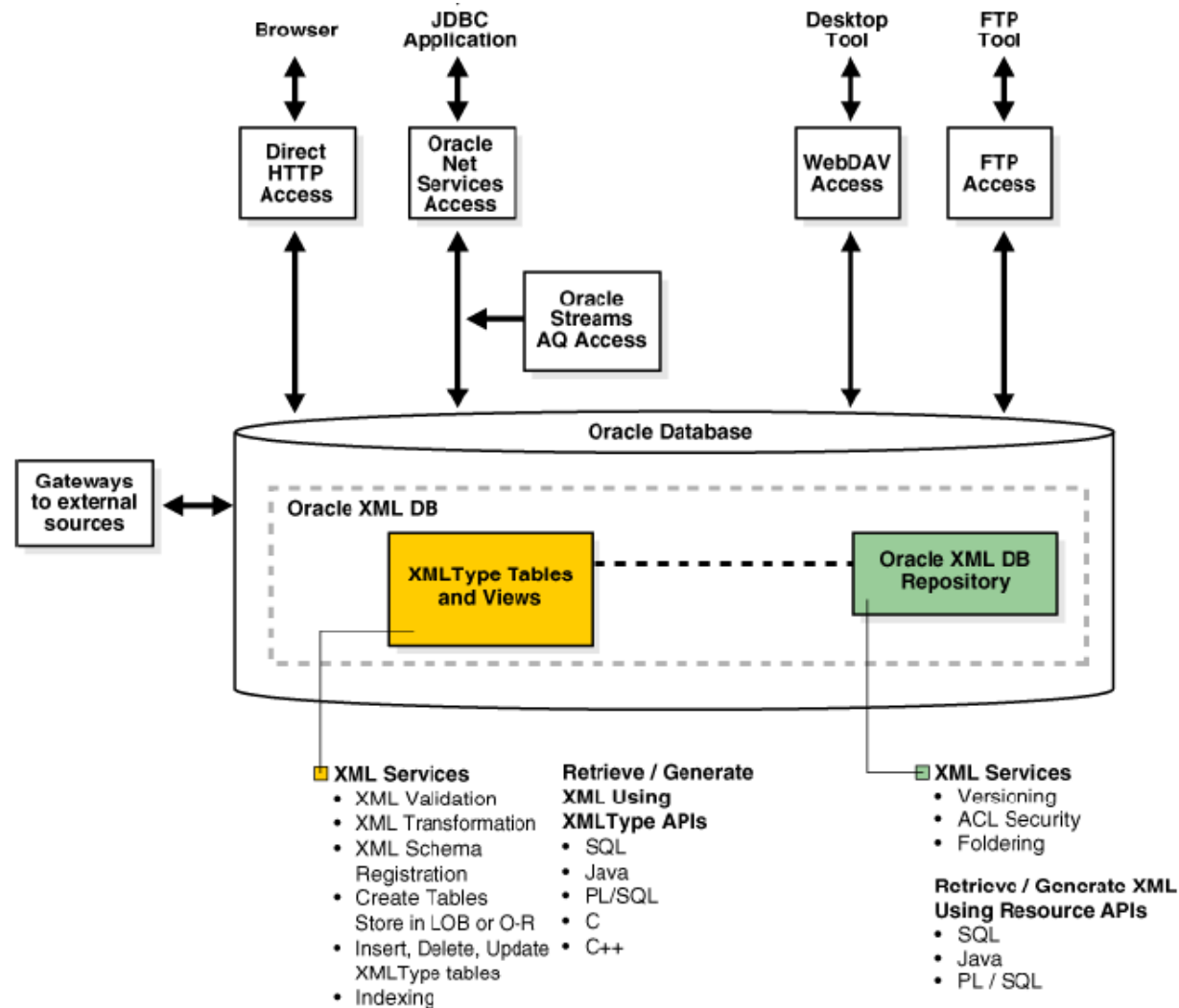
- Typy XML dokumentů
- XML databázové produkty
- Standard SQL a podpora XML
- Podpora XML u serveru Oracle11g (Oracle XML DB)
- Podpora formátu JSON



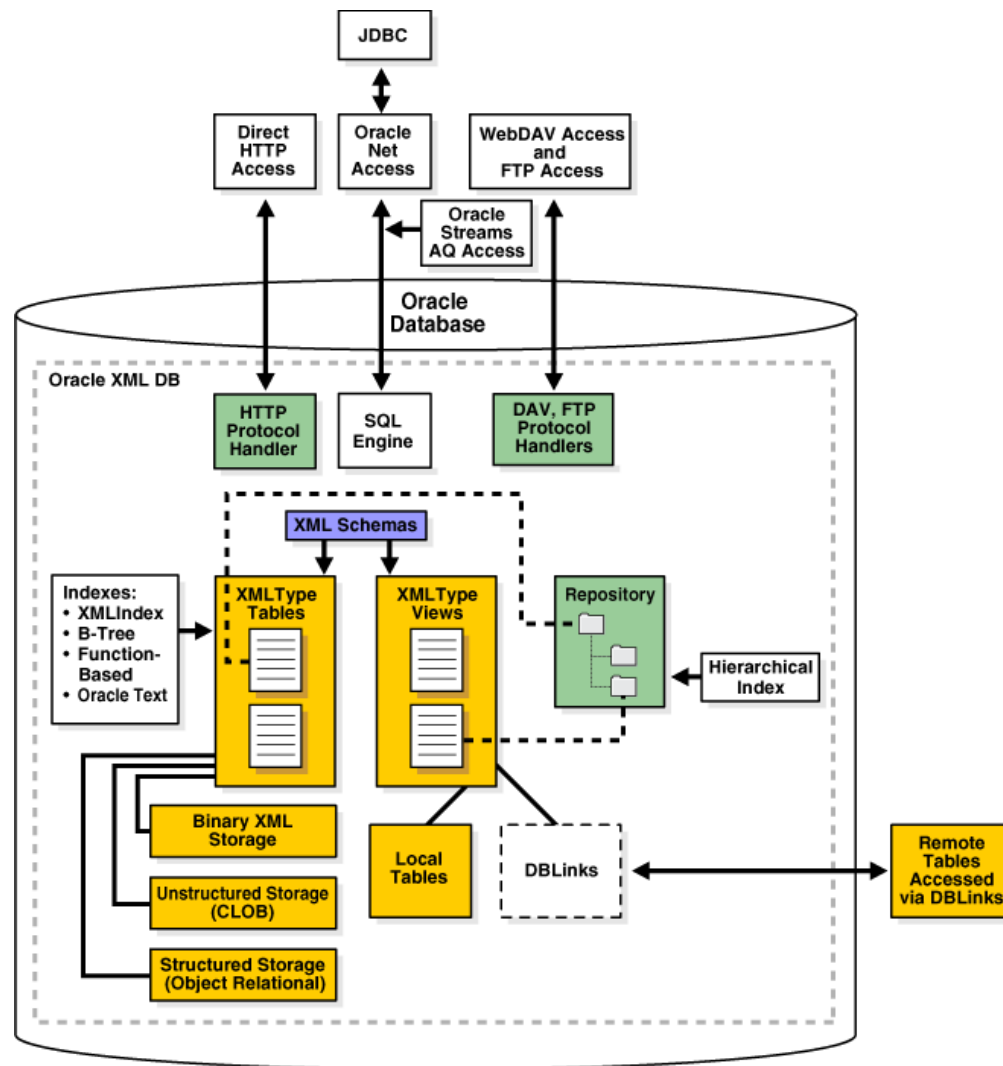
Podpora XML u serveru Oracle11g

- Od verze 9.2 (2002)
- Typ XMLType - odpovídá standardnímu typu XML
- SQL funkce a operátory pro práci s XML
- Podpora XML Schema
- Dualita XML a SQL - lze používat SQL operace nad XML daty a naopak
- Podpora dotazování SQL/XML, XPath, XQuery
- Hierarchická organizace obsahu v XML DB Repository v podobě adresářů a souborů.

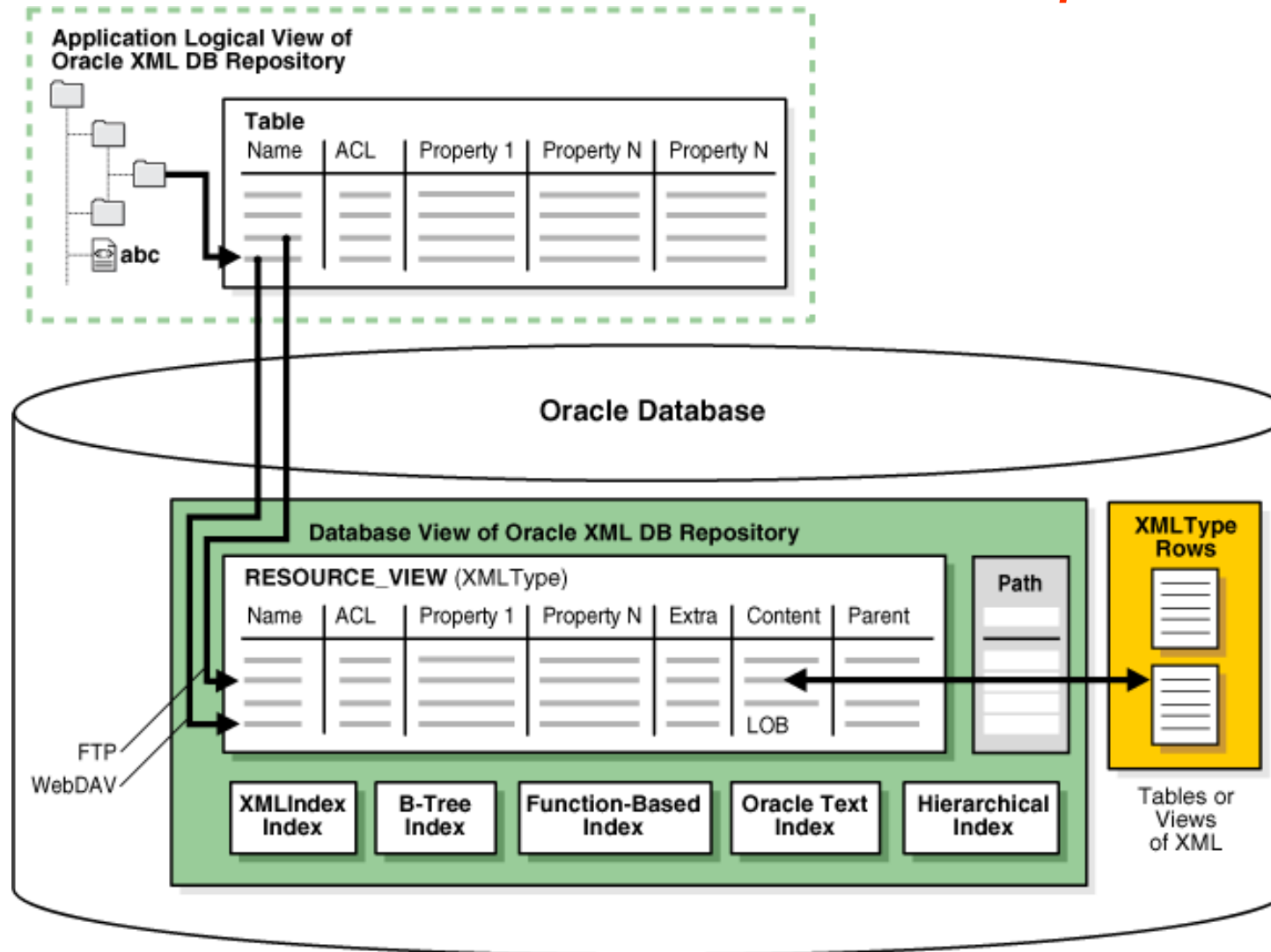
Architektura Oracle XML DB



Architektura úložiště hodnot XMLtypu



Architektura Oracle XML DB Repository







Ukládání XML dokumentů do databáze(1)

- Uložení do relační tabulky
 - používaný dříve, ztrácí výhody XML
 - rozklad/sestavení XML dokumentu na aplikační úrovni ↔ standardní uložení/načtení do/z DB
- Uložení do CLOB
 - uložení celého dokumentu do CLOB
 - flexibilní s ohledem na změnu schématu
 - rychlé získání dokumentu, operace nad částmi pomalé a paměťově náročné
 - pomalé aktualizace



Ukládání XML dokumentů do databáze (2)

■ Objektově-relační uložení

- ☐ použití typu XMLType pro vytvoření sloupcových objektů a objektových tabulek
- ☐ mapování a validace na základě XML schématu
- ☐ efektivní práce s částí dokumentu pomocí XPath
- ☐ možnost využití podpory pro omezení v SQL (např. PRIMARY KEY)



Dotazování nad XML

- Použití XPath a XSLT
- Použití XQuery
- Použití XSQL
- Použití SQL/XML

Příklad

```
<?xml version="1.0" encoding="windows-1250" ?>
<ZAKAZNICI>
  <ZAKAZNIK id="1">
    <LOGIN>xmalyj02</LOGIN>
    <JMENO>Jan</JMENO>
    <PRIJMENI>Malý</PRIJMENI>
    <ADRESA>
      <CISLO>20</CISLO>
      <ULICE>Žádná</ULICE>
      <MESTO>Brníčko</MESTO>
      <PSC>060200</PSC>
    </ADRESA>
    <EMAIL>jan@seznam.cz</EMAIL>
    <TELEFON telefon2="987654321">123456789</TELEFON>
  </ZAKAZNIK>
  ...
</ZAKAZNICI>
```

Použití XPath

■ Příklady dotazů

- **/ZAKAZNICI** - výraz vybere všechny elementy ZAKAZNICI
- **ZAKAZNICI/*** - výraz vybere všechny synovské elementy (element ZAKAZNIK)
- **/ZAKAZNICI/ZAKAZNIK/@id** - výraz vybere všechny atributy id
- **/ZAKAZNICI/ZAKAZNIK/JMENO[1]** - vybere první jméno dokumentu
- **/ZAKAZNICI/ZAKAZNIK/JMENO[last()]** - vybere poslední jméno dokumentu
- **//ZAKAZNIK[ADRESA/PSC='61200']/PRIJMENI** - vybere příjmení zákazníka, jehož PSČ je 61200

Použití XQuery (1) – v XMLQuery

■ Příklad - „Najdi zákazníky z Brna“

xquery version "1.0";

{--toto je komentar--}

for \$p in doc('zakaznici.xml')/ZAKAZNICI/ZAKAZNIK

where \$p/ADRESA/MESTO = 'Brno'

order by \$p/@id

return

<BRNAK id="{ \$p/@id }">

<JMENO>{ string(\$p/JMENO) }</JMENO>

<PRIJMENI>{ string(\$p/PRIJMENI) }</PRIJMENI>

</BRNAK>



Použití XQuery (2)

■ Příklad - výsledek

<BRNAK id="2">

 <JMENO>Petr</JMENO>

 <PRIJMENI>Novák</PRIJMENI>

</BRNAK>



Použití SQL/XML

- Dostupné jsou např. funkce:
 - *XMLElement()* – vytvoří XML element se zadaným jménem a do obsahu uloží hodnotu.
 - *XMLForest()* – vytvoří skupinu XML elementů z názvů sloupců relační tabulky.
 - *XMLConcat()* – funkce sloučí dvě a více XML hodnot do jedné XML hodnoty.
 - *XMLAgg()* – funkce vytvoří les XML elementů ze seznamu elementů.



Balíčky a knihovny pro práci s XML

- Existuje řada balíčků pro PL/SQL, např.
 - DBMS_XMLDOM - pro práci s modelem DOM dokumentu
 - DBMS_XMLPARSER - pro procházení dokumentu a získávání obsahu a metadat
 - DBMS_XSLPROCESSOR - procesor XSLT
 - DBMS_XMLSTORE - pro vkládání, aktualizaci a rušení XML dokumentů uložených v objektově-relační databázi
- Pro Javu existuje DOM API



Práce s XML na Oracle

- Ukážeme si
 - ☐ Registraci schématu
 - ☐ Vytvoření a manipulace s tabulkou se sloupcem typu XMLType
 - ☐ Generování XML z dat v relačních tabulkách
 - ☐ Vytvoření a manipulace s objektovou tabulkou typu XMLType
 - ☐ Indexování XML dat



Obsah

- Typy XML dokumentů
- XML databázové produkty
- Standard SQL a podpora XML
- Podpora XML u serveru Oracle11g (Oracle XML DB)
- Podpora formátu JSON

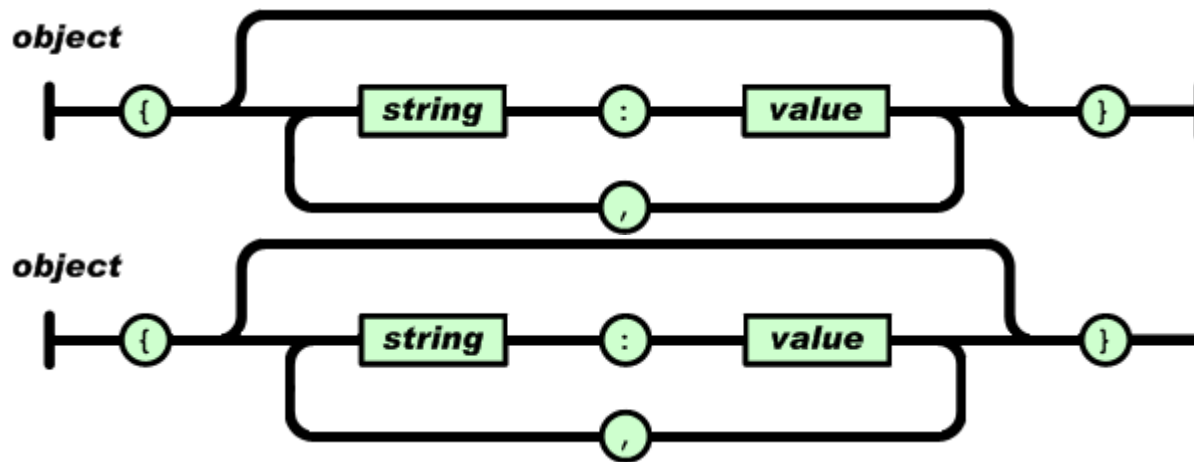


Formát JSON (1)

- JSON (JavaScript Object Notation)
 - Notace pro popis dat (pro ukládání a výměnu)
 - Výhody oproti XML
 - Úspornější, čitelnější člověkem, méně přísný → snadněji zpracovatelný v různých programovacích jazycích, mnohem snadnější práce s poli
 - Výhody XML
 - Přísnější (jmenné prostory → sdílení standardních struktur), možnost poznámek, koncept schémat
 - Oba přístupy lze kombinovat

Formát JSON (2)

- Základní struktury – objekt, pole



- Datové typy, resp. *hodnoty*
 - String, number, *true*, *false*, *null*, object



Příklad

```
{
  "father": {
    "firstname" : "David",
    "lastname"  : "Njoku",
    "occupation" : "Software developer",
    "gender"     : "Male",
    "isAdult"    : true
  },
  "mother": {
    "firstname" : "Sola",
    "lastname"  : "Njoku",
    "occupation" : "Radio producer",
    "gender"     : "Female",
    "isAdult"    : true
  },
}
```



Podpora pro práci s JSON

- Kde ukládat JSON data?
 - Ideálně NoSQL (dokumentové) DB – bez/volné schéma, dvojice klíč – hodnota.
 - Existuje podpora u RDBMS?
 - Existuje (např. Oracle JSON Store)
 - Aktuálně připravována standardizace v SQL



Podpora pro práci s JSON v Oracle

- Ukládání
 - VARCHAR2
 - CLOB
- Kontrola správnosti (CHECK)
- Dotazování – podobné zadávání cesty jako v Xpath
- Podpora generování – podobné SQL funkce jako u XML pro vytvoření JSON objektu a pole, podpora v PL/SQL

Příklad (1)

```
-- Vytvoření tabulky se sloupcem ve formátu JSON
CREATE TABLE json_test
(ID          RAW(16) NOT NULL,
 json_data   CLOB    NOT NULL,
 created_date DATE    DEFAULT SYSDATE
);
-- Přidání omezení pro kontrolu formátu
ALTER TABLE json_test
ADD CONSTRAINT check_is_json
CHECK (json_data IS JSON);
-- Vložení řádku
INSERT INTO json_test (id, json_data)
VALUES (sys_guid(),
'{'
  "father":  { ...
  }`);
```



Příklad (2)

```
-- Vypíše vše
SELECT jt.json_data
FROM json_test jt;

-- Použití specifikace cesty jako u XPath
SELECT JSON_VALUE(json_data, '$.father.firstname' ERROR ON ERROR)
name_of_the_father
FROM json_test;
```



Literatura (1)

- Bourret, R.: XML and Databases. Dostupné na <http://www.rpbourret.com/xml/XMLAndDatabases>.
- Bourret, R.: Database Products. Dostupné na <http://www.rpbourret.com/xml/XMLDatabaseProds.htm>.
- XML Wrapper. Dostupné na <http://publib.boulder.ibm.com/infocenter/db2luw/v9/index.jsp?topic=/com.ibm.websphere.ii.foundation.conn.fw.xml.doc/configuring/c0007799.html>
- Introduction to Oracle XML Developer's Kit. Dostupné na http://download.oracle.com/docs/cd/E11882_01/appdev.112/e10708/adx_overview.htm#i1013836.
- JAXB Architecture. Dostupné na http://download.oracle.com/docs/cd/E17802_01/webservices/webse/vices/docs/1.5/tutorial/doc/JAXBWorks2.html.
- GEMt Content Management System. Dostupné na <http://02a0e19.netsolhost.com/gemcontentsystem.htm>.



Literatura (2)

- Oracle XML DB Developer's Guide, 10g Release 2, 1, Oracle Corporation, 2005, s. 852, Dostupné na http://download.oracle.com/docs/cd/B19306_01/appdev.102/b14259/toc.htm.
- New Features in Oracle XML DB for Oracle 11g Release 1. Dostupné na <http://www.oracle.com/technology/products/database/oracle11g/pdf/xml-db-11g-whitepaper.pdf>
- DP a BP:
 - Sobotka, J.: Podpora pro práci s XML u databázového serveru Oracle. Diplomová práce FIT VUT. 2006.
 - Dvorštiak, M.: Demonštrácia podpory pre prácu s XML v moderných databázových systémoch. Bakalárska práce FIT VUT. 2009.
 - Kaňovský, J.: Podpora pro práci s XML v moderních databázových systémech. Bakalárska práce FIT VUT. 2009.



Literatura (3)

- Njoku,D.: JSON For Absolute Beginners. Part 1 - 6. Part 1 on <https://www.red-gate.com/simple-talk/sql/oracle/json-for-absolute-beginners-part-1-introduction/>.