

XML и реляционные СУБД



**Хранение и работа с XML
в реляционных системах
управления базами данных**

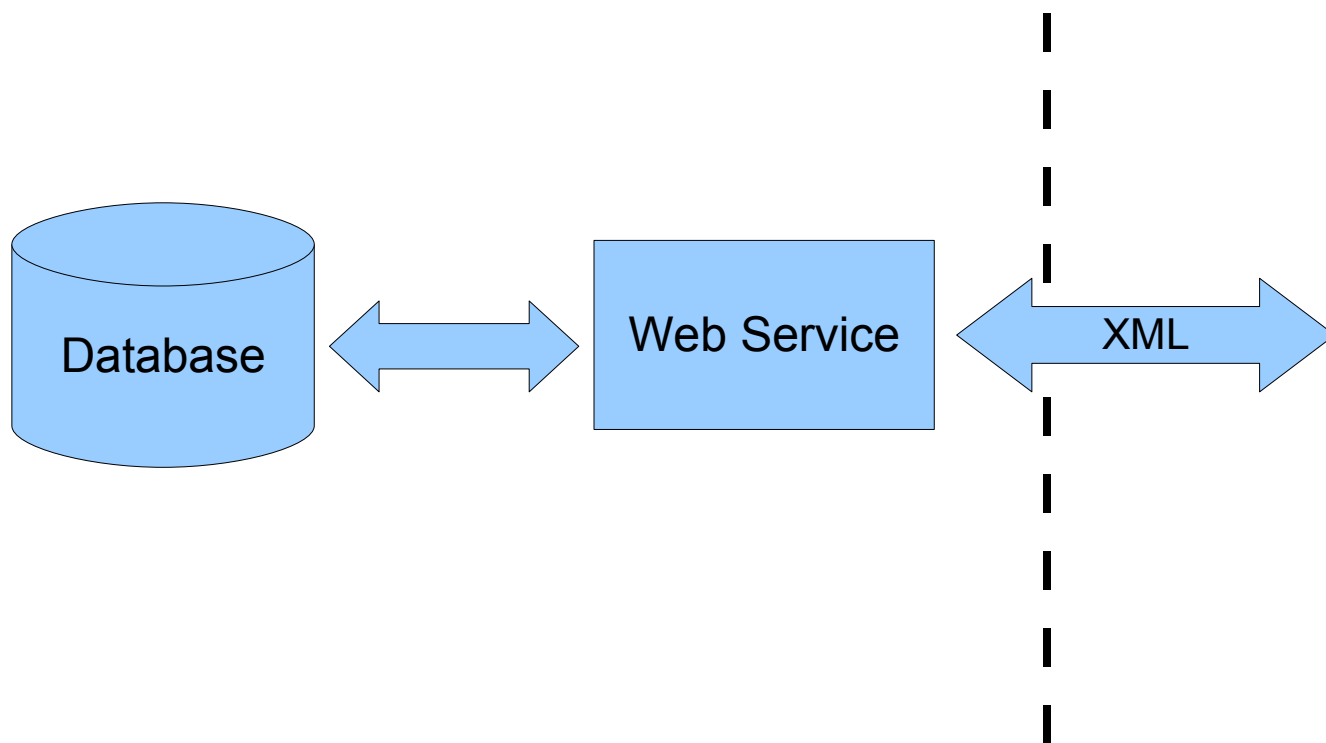
Самохвалов Николай
Компания «Постгресмен»
<http://postgresmen.ru>
rit2007@samokhvalov.com

XML и СУБД

Зачем?

XML и СУБД

- XML для общения с «внешним миром»

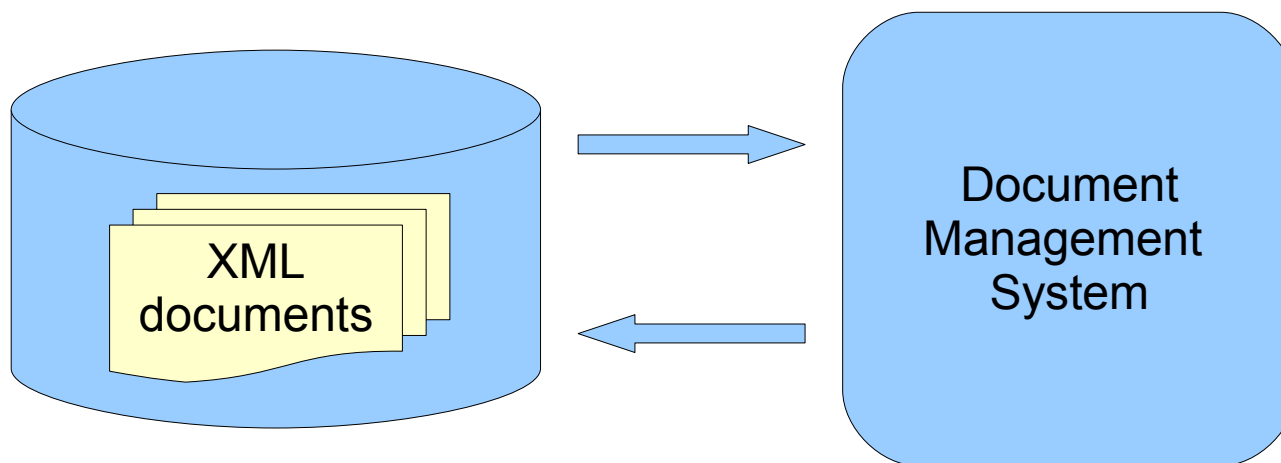


Зачем хранить XML в РСУБД?

- Системы документооборота
- Полуструктурированные данные
- Архивы документов

Пример 1: система документооборота

Цель: работа с документами в оригинальном формате



Пример 2: разнородный каталог

Цель: работа с объектами с различными наборами свойств

Авто [+ Подать объявления в этот раздел](#)

Кредиты и ссуды (22)	Комиссионное оформление и страхование (5)
Легковые автомобили (35679)	Прицепы, автодома и фургоны (146)
Новые автомобили	Автобусы (231)
Автомобили с пробегом	Спецтехника (1426)
Малый коммерческий транспорт (2320)	Мотоциклы и мопеды (1505)
Средние и тяжелые грузовики (2039)	Снегоходы (37)
Грузовые прицепы (508)	Автосервис и услуги (780)
Водный транспорт (365)	Сервис и ремонт
Автозапчасти и принадлежности (10507)	Экспертиза и оценка
Для легковых автомобилей	Юридическая помощь
Для коммерческих автомобилей	Автошколы
Для мотоциклов и мопедов	Другое
Диски, шины, колеса	
Для спецтехники и прицепов	
Авто на запчасти	
Средства для ухода	
Другое	
Другое (400)	

IRR.ru, раздел «Авто»

Пример 2: разнородный каталог

item	
obj_id	INT8
item_section_id	INT8
item_vendor_id	INT8
item_model_id	INT8
item_year	INT2
item_price	NUMERIC(30, 6)
item_prop1	INT4
item_prop2	INT4
item_prop3	INT4
item_prop4	INT4
...	
item_prop21	TEXT
item_prop22	TEXT
item_prop23	TEXT
...	
item_prop41	BOOLEAN
...	



item	
obj_id	INT8
item_section_id	INT8
item_vendor_id	INT8
item_model_id	INT8
item_year	INT2
item_price	NUMERIC(30, 6)
item_props	hstore

hstore – тип данных в PostgreSQL, множество пар «ключ – значение» (аналог хэшей Perl), с поддержкой индексов (GiST)

Пример 3: журнал действий

Цель: гибкость схемы (возможность добавить новые свойства)

action	
action_id	SERIAL
action_type_id	INT4
action_status_id	INT4
action_person_id	INT4
action_data	hstore

SQL/XML (ISO/IEC 9075-14:2006)

- тип данных XML (XQuery 1.0 and XPath 2.0 Data Model)
- функции публикации (XMLDOCUMENT, XMLELEMENT, XMLATTRIBUTES, XMLFOREST, XMLAGG, XMLCONCAT, XMLPI, XMLCOMMENT)
- поддержка XQuery (операция XMLQUERY)

Тип данных XML

doc		
PK	id	INTEGER
	data	XML
	created	TIMESTAMP

```

- <rss version="2.0">
- <rss version="2.0">
- <channel>
  <title>MoiKrug.ru - Вакансии (RSS)</title>
- <description>
  MoiKrug.ru - Последние вакансии в первой социальной сети
</description>
  <link>http://moikrug.ru/?from=rss</link>
  <language>ru-ru</language>
- <image>
  <link>http://moikrug.ru/?from=rss</link>
  <url>http://moikrug.ru/images/logo.gif</url>
- <title>
  MoiKrug.ru - найти хороших людей по рекомендациям
</title>
</image>
- <item>
  <title>Менеджер по персоналу ($1500)</title>
  <link>http://moikrug.ru/vacancies/736200561/?from=rss</link>
- <description>
  Менеджер по персоналу ($1500)<br /> HFLabs<br />
  персоналу.<br /> <br /> Обязанности:<br /> * кадровое
  поддержание корпоративной культуры компании.<br />
  года в роли HR в западной компании или компании с
  /> * английский язык на уровне чтения резюме;<br />
  зарплата 1000-1700 у.е. плюс бонусы 300-500 у.е. за
  "HR".
  </description>
  <dc:creator>Холодова Елена</dc:creator>
  <pubDate>Fri, 24 Nov 2006 18:08:36 GMT</pubDate>
</item>
</channel>
</rss>
  
```

```

CREATE TABLE doc (
  id INTEGER PRIMARY KEY,
  data XML,
  created TIMESTAMP WITH TIMEZONE
);
  
```

Производительность

- Выборка всего XML-значения по условию XPath

```
SELECT xmlcol  
FROM tbl1  
WHERE (xpath('/a/@b', xmlcol))[1] = '123';
```

```
SELECT xmlcol  
FROM tbl1  
WHERE (xpath('/a[@c="s"]/@b', xmlcol))[1] = '123';
```

- Выборка произвольной части XML-значения

```
SELECT xpath('/a/@b', xmlcol)  
FROM tbl1  
WHERE id = 111;
```

- Модификация части XML-значения

Oracle 10g

- Тип XMLType. Варианты хранения:
 - object-relational storage
 - CLOB
- XML Schemas
- SQL/XML (включая ф-ции публикации, XMLQUERY, XMLVALIDATE, XMLTABLE, ...)
- XQuery
- XSLT
- DOM
- Node-level updates
- Индексы:
 - обычные B-деревья в случае object-relational storage
 - XMLIndex для обоих типов хранения (индексация путей, значений, ...)
 - функциональные индексы
 - индексы для полнотекстового поиска

MS SQL 2005

- Тип XML. Хранение: BLOB.
- XML Schemas
- SQL/XML publishing functions
- XQuery
- XSLT
- Node-level updates
- Индексы:
 - «первичные» XML-индексы (ORDPATH)
 - «вторичные» XML-индексы
 - PATH
 - VALUE
 - PROPERTY
- Web services

IBM DB2 version 9

- Тип XML. Собственное хранилище (native storage: pureXML)
- XML Schemas (с поддержкой эволюции)
- SQL/XML (включая ф-ции публикации, XMLQUERY, XMLVALIDATE, XMLTABLE, ...)
- XQuery
- XSLT
- Node-level updates
- Индексы:
 - XML value indexes
 - XML full text indexes
- Annotated XML Schemas для декомпозиции (XML->rel)

MySQL 5.1

```
mysql> SET @xml = '<a>111<b:c>222<d>333</d><e:f>444</e:f></b:c></a>';  
Query OK, 0 rows affected (0.00 sec)
```

```
mysql> SELECT ExtractValue(@xml, '//e:f');
```

```
+-----+  
| ExtractValue(@xml, '//e:f') |  
+-----+  
| 444 |  
+-----+
```

```
1 row in set (0.00 sec)
```

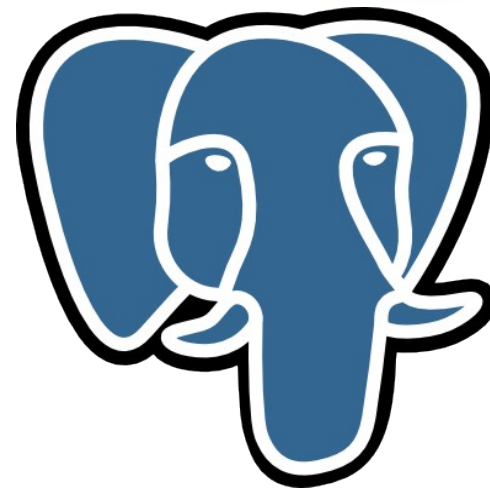
```
mysql> SELECT UpdateXML(@xml, '//b:c', '<g:h>555</g:h>');
```

```
+-----+  
| UpdateXML(@xml, '//b:c', '<g:h>555</g:h>') |  
+-----+  
| <a>111<g:h>555</g:h></a> |  
+-----+
```

```
1 row in set (0.00 sec)
```

PostgreSQL 8.3

- Тип XML (over VARCHAR), поддержка XML-фрагментов
- SQL/XML
 - SQL/XML publishing functions
 - XMLPARSE, XMLSERIALIZE
- XSLT (via contrib)
- XPath с поддержкой XML Namespaces (`xpath()`)
- Индексы:
 - функциональные индексы над `xpath()`
 - XLABEL (GiST based; в разработке)
- Альтернативные функции отображения rel->XML
 - `table_to_xml()`, `query_to_xml()`, `cursor_to_xml()`
 - `table_to_xmlschema()`, ...
 - `table_to_xml_and_xmlschema()`, ...
 - `database_to_xml()`, `database_to_xmlschema()`, ...

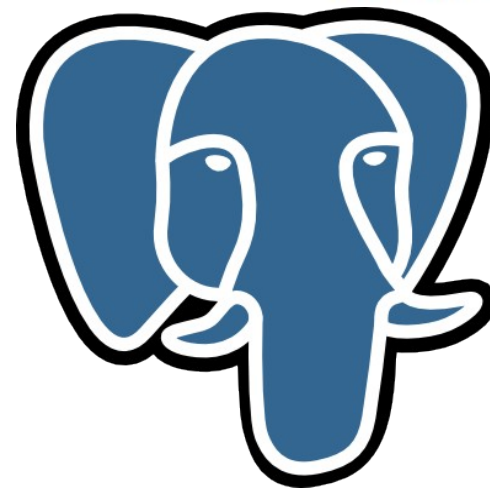


PostgreSQL 8.3

```
xmltest=# SELECT XMLELEMENT(  
           NAME "person",  
           XMLATTRIBUTES(  
               person_name || ' ' || person_lastname  
                   AS "name",  
               age(person_birthday) AGE "age"  
           )  
       )  
FROM person;
```

xml element

```
-----  
<person name="Ivan Sidorov" age="29 years 6 mons 6 days" />  
<person name="Petr Ivanov" age="22 years 8 mons 14 days" />
```



PostgreSQL 8.3

```
xmltest=# CREATE TABLE table1(  
    id INTEGER PRIMARY KEY,  
    created TIMESTAMP NOT NULL  
    DEFAULT CURRENT_TIMESTAMP,  
    xdata XML  
);  
CREATE TABLE  
  
xmltest=# INSERT INTO  
    table1(id, xdata)  
VALUES(  
    1,  
    '<dept xmlns:smpl="http://example.com" smpl:did="DPT011-IT">  
        <name>IT</name>  
        <persons>  
            <person smpl:pid="111">  
                <name>John Smith</name>  
                <age>24</age>  
            </person>  
            <person smpl:pid="112">  
                <name>Michael Black</name>  
                <age>28</age>  
            </person>  
        </persons>  
    </dept>'  
);  
INSERT 0 1
```



PostgreSQL 8.3

```
xmltest=# SELECT xpath('//person/name/text()', xdata)
FROM table1;
```

 xpath

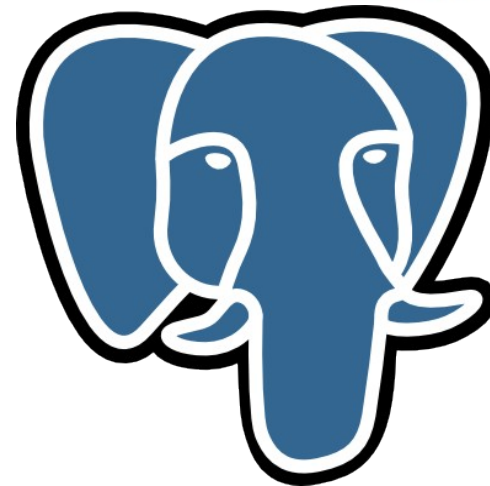
```
-----
{"John Smith","Michael Black"}
(1 row)
```

```
xmltest=# SELECT xpath('//person/@smpl:pid', xdata,
                  ARRAY[ARRAY['smpl', 'http://example.com']])
FROM table1;
```

 xpath

```
-----
{111,112}
(1 row)
```

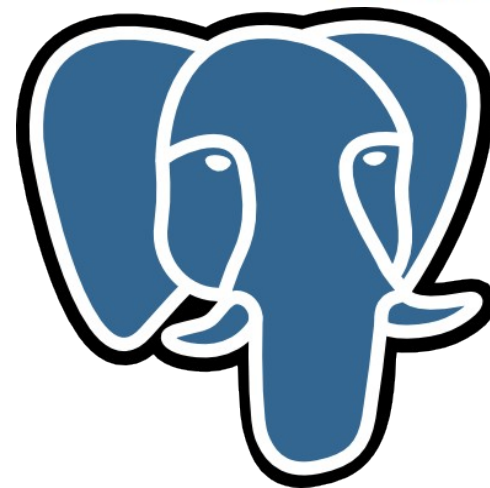
```
xmltest=# CREATE INDEX i_table1_xdata ON table1 USING btree(
          xpath('//person/@name', xdata)
);
CREATE INDEX
```



PostgreSQL 8.3

Планы по развитию:

- XLABEL
- XML Schemas catalogue
- полнотекстовый поиск по XML-данным
- decomposition with annotated XML Schemas
- XQuery (XMLQUERY)



Пример использования

Хранение метаданных в наблюдательных архивах
International Virtual Observatory

The screenshot shows a Microsoft Internet Explorer browser window displaying the ASPID-SR Search Interface. The search results table is as follows:

FITS	Proposal	Target	J2000 Coordinates	Local Date&Time	Exp.Time	Inst
	unknown	MKN 315	(23h 04m 02.64s, +22d 37m 31.6s)	2000-12-02 21:11:04	3600	M
	unknown	MKN 315	(23h 04m 03.12s, +22d 37m 35.0s)	2000-12-02 17:59:44	4800	M
	Environment of SyG	MKN315	(23h 04m 03.31s, +22d 37m 34.8s)	2003-08-24 23:22:45	13200	M

Below the browser, the VOSpec software interface is shown. It features a 'VOSpec Spectra Viewer' window displaying a plot of Flux (Jy/logarithmic) versus Wavelength (micron, logarithmic). The plot shows a spectrum with several emission lines, notably around 5.0 and 5.2 microns. The Y-axis ranges from 0.0 to 10.0 (scaled by 10⁻¹), and the X-axis ranges from 4.80 to 6.40 (scaled by 10⁻¹).

The VOSpec interface also includes a 'Spectra List' panel at the bottom, showing 'Local Data' and 'Local Spectrum' (checked). The software is identified as 'Aladin v4.0 BETA VERSION (based on v4.007)' and 'ESO DSS2-red 23:04:03.31 +22:37:34.8'.

Спасибо за внимание!

Ссылки, литература

- **SQL:200x, 2006, Part 14: XML-Related Specifications (SQL/XML)**, ISO/IEC JTC 1/SC 32, CD 9075-14:200x(E). International Standard, ISO, ANSI.
- **R. Bourret. XML Database Products.** <http://www.rpbourret.com/xml/XMLDatabaseProds.htm>
- **M. Nicola and B. van der Linden. Native XML Support in DB2 Universal Database.**
In VLDB, pages 1164 1174, Trondheim, Norway, 2005.
- **P. O Neil, E. O Neil, S. Pal, I. Cseri, G. Schaller, and N. Westbury. ORDPATHs: insert-friendly XML node labels.** // In SIGMOD 04: Proceedings of the 2004 ACM SIGMOD international conference on Management of data, pages 903 908, New York, NY, USA, 2004. ACM Press.
- **Бартунов О. С. Что такое PostgreSQL?** // Корпоративные базы данных-2005. CIT. М.: 2005. http://www.citforum.ru/database/postgres/what_is/
- **N. Samokhvalov. XML Support in PostgreSQL.** // Proceedings of the SYRCoDIS-2007, Moscow, Russia, 2007.
- **I. Zolotukhin, N. Samokhvalov, F. Bonnarel, and I. Chilingarian. Comprehensive Metadata Query Interface for Heterogeneous Data Archives Based on Open Source PostgreSQL ORDBMS.** // In *Astronomical Data Analysis Software & Systems XVI*, p. P3.23, Tucson, Arizona, USA, 2006.