

УДК 004.657(045)

Анохін А.С.

## ВИКОНАННЯ XQUERY ЗАПИТІВ В РОЗПОДІЛЬЧІЙ СИСТЕМІ НАПІВСТРУКТУРОВАНИХ ДАНИХ

Інститут комп'ютерних технологій  
Національного авіаційного університету

*Проведені експерименти з напівструктурованими даними за новими стандартами по виявленню здатності виконання XQUERY запитів в розподільчій системі із застосуванням Web servisu*

### **Вступ**

З приходом ери інформаційних технологій у суспільстві з'явилося багато способів зберігання та передачі інформації. Одним з таких видів є Розширювана Мова Розмітки (XML) [1]. XML – текстовий формат, розроблений для зберігання структурованих даних. Тобто цей вид даних може достатньо конкурувати з базами даних (БД). На сьогоднішній день комерційні системи управління базами дани (СУБД) *Microsoft, Oracle* застосовують для зберігання інформації реляційну модель даних [2].

Такі СУБД можна вважати універсальними, тобто їх можна налаштувати на будь-яку предметну область. Вони дають необмежені можливості для розвитку та використання. Але іншій класифікації БД підлягають спеціальні, тобто ті, які орієнтовані на певний клас. Тому на сьогоднішній день для зберігання та використання структурованих типів даних з'явилися XML-СУБД.

Безпосередньо до переваг XML можна віднести сам XML-документ. Тобто інформація представлена в документі зберігає структуру даних за допомогою тегів, і цей документ зберігає ієрархію даних. Достатньо невеликі розміри самих документів дозволяють зберігати сам документ в базі даних та проводити операції над ними.

Ієрархічна модель XML являє собою дуже просту модель зберігання даних, але і застарілою, так званою типізованою. На сьогодні реляційна модель вважається більш сприятливою при обробці даних. Але слід зауважити, що для розширення

можливостей XML документу було додано описуючу схему. За допомогою описуючої схеми XML-даних БД буде перевіряти документ по цих схемах. Для здійснення запитів по схемах може застосовуватися мова запитів XQUERY [3].

Застосування XML-даних с XML-СУБД дає змогу повністю замінити реляційні СУБД по обробці напівструктурованих даних на обробку даних в XML форматі.

Але невирішеною залишається задача по використанню мова запитів XQUERY в розподільчій системі. Тобто система із багатьма користувачами оперує з XML-документом та виконує операції над ним. Тому в області проектування, розробки залишається актуальною задача розробки системи керування довготривалими транзакціями при оперуванні з XML-документом в розподільній системі.

### **Постановка завдання**

Аналіз комерційних реляційних СУБД (PCСУБД) таких як *MS SQL Server 2000/2005/2008, Oracle 11g* показує, що вони підтримують встроєні стандарти XML. Але той факт, що ці СУБД були розроблені для обробки та зберігання різних типів даних не надають перевагу перед XML-СУБД для обробки XML файлів, головною перевагою таких СУБД є спорідненість системи.

XML-СУБД реалізують ідентичні властивості транзакцій: атомарність, довго тривалість, паралелізм, ізольованість.

Але одною з найголовніших проблем постає виконання запитів в СУБД. При виконанні запитів в XML-СУБД застосовується мова запитів XQUERY.

Під час виконання *XQUERY* запитів в *XML*-СУБД можуть вирішуватися наступні задачі: перевірка всіх існуючих методів керування транзакціями, застосування більш нових технологій по описанню *XML*-документу та застосування *Web* сервісу для здійснення запитів в розподільчій системі.

### Основна частина

На сьогоднішній день використання *XML* спровокувало розробку *XML*-СУБД різних виробників. Для перевірки працездатності виконання *XQUERY* запитів буде використано *XML*-СУБД "Sedna". Основною перевагою даної СУБД є використання версійного протоколу *4VXDGL* [5]. Тобто даний протокол враховує семантику *XML*-моделі даних, так і фізичних властивостей організації структур зовнішньої і оперативної пам'яті в рідних *XML*-СУБД. Протокол *4VXDGL* також враховує специфіку використання в рідних *XML*-СУБД, що забезпечує безконфліктну реалізацію читаючих та змінюючих транзакцій. Що в свою чергу дає змогу використовувати даний версійний протокол для застосування цієї бази даних в розподільчій системі, застосовуючи зв'язок клієнт-сервер.

В ході аналізу вже існуючих розробок по обробці, збереженню інформації та виконання запитів в реляційних СУБД, а також *XML*-СУБД постає необхідність у порівнянні виконання пропускну здатності зчитуючих транзакцій та змінюючих транзакцій. Тому для першого експерименту була обрана реляційна СУБД *MSSQL 2005* та *XML*-СУБД "Sedna".

Виконання транзакцій з читання та змінюючих транзакцій виконується не з самим документом *XML*, а за допомогою описуючої схеми. Для виконання першого експерименту застосовувалася *DTD* схема, але слід зауважити, що недоліком цієї описуючої схеми є те, що вона не спорідненим з *XML* форматом стандартом. Тому

для досягнення основної мети – споріднення *XML* системи, буде застосовуватися *XML schema*. Найголовнішою перевагою цього описуючого формату є те, що *XML schema* написана на мові *XML*.

На останньому етапі вирішення проблеми по виконанню *XQUERY* запитів буде задіяно *Web* сервіс із простим інтерфейсом доступу до *XML*-СУБД. За допомогою цього найпростішого способу буде визначено здатність виконання *XQUERY* запитів в розподільчій системі, як структури "клієнт-сервер".

При розгляді задачі ефективності виконання *XQUERY* запитів для *XML*-СУБД було проведено експерименти на вимірювання продуктивності з використанням *4VXDGL* протоколу та двухфазного протоколу синхронізації із гранульованістю захватів на рівні блоків. Тобто буде розглянути наступні характеристики:

- Пропускна здатність *XML*-СУБД для транзакцій зчитування *XML*-документів при паралельних запитах.
- Пропускна здатність *XML*-СУБД для записуючих транзакцій при паралельних запитах.

Перша група експериментів спрямована на вимірювання пропускну спроможності *XML*-СУБД за наявності конкурентних потоків транзакцій на читання та редагування *XML*-документів.

При цьому ми використовуємо 8 потоків транзакцій. Потік транзакцій складається з тих, хто читає або змінюють транзакцій. Загальна кількість змінюючих транзакцій становить приблизно 75%. Причому транзакції зі зміни є невеликими і виробляють зміни від 5 до 10 елементів. Розмір транзакцій з читання є оптимальним, і змінюється в бік збільшення у відсотках від загального розміру документа. Розмір *XML*-документа становить приблизно 9 Мб.

На рис. 1. зображені результати вимірів для змінюючих транзакцій.

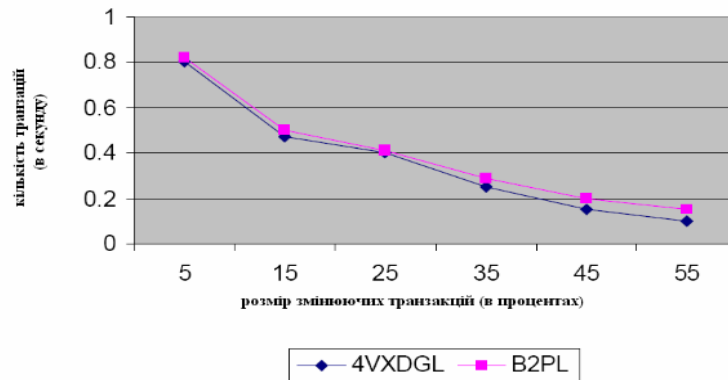


Рис. 1. Пропускна здатність XML-СУБД для змінюючих транзакцій

Криві помічені ромбом і квадратом, зображують середню кількість виконаних змінюючих транзакцій в секунду при використанні протоколів 4VXDGL і B2PL (двофазний протокол з рівнем гранулювання захоплень на рівні блоків) відповідно.

На графіку видно, що зі зростанням розміру читаючих транзакцій середня кількість виконаних змінюючих транзакцій сильно зменшується для протоколу B2PL, і залишається приблизно постійним для

4VXDGL. На рис. 2. зображена пропускна здатність XML-СУБД для транзакцій з читання. У даному випадку ми бачимо, що зі зростанням розміру зчитуюючих транзакцій середня кількість виконаних транзакцій зменшується як для протоколу 4VXDGL, так і для протоколу B2PL. Причому ми бачимо, що для тих, хто зчитує транзакцій "великого" розміру протокол B2PL трохи перевершує протокол 4VXDGL.

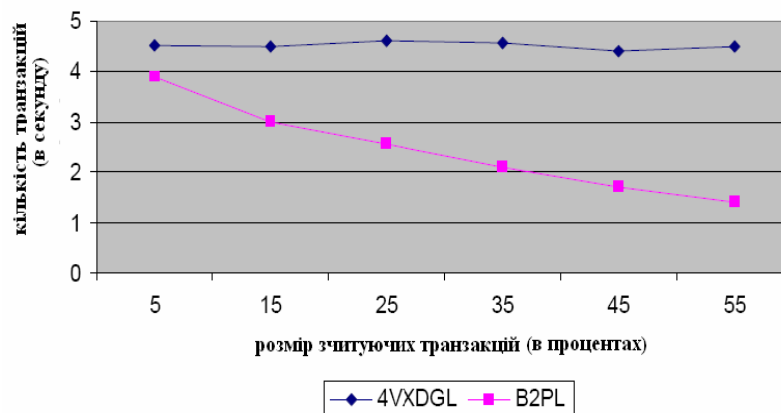


Рис. 2. Пропускна здатність XML-СУБД для транзакцій з читання

Важливою частиною специфікації XML є мова опису схеми XML-документів – DTD (Document Type Definition). Фактично, за допомогою DTD можна накладати обмеження на структурний вид XML-документа. Для накладення більш складних обмежень існує більш потужна мова опису схем XML Schema [6].

Виходячи з цього експерименту слід звернути увагу на те, що описуюча схема цього документу написана за допомогою описуючої схеми DTD. Тобто наша сис-

тема недостатньо рідна з XML-документом. Тому для подальших експериментів буде застосовуватися XML-schema (XSD). Переваги описуючої схеми:

- XSD підтримує типи даних;
- XSD підтримує простір імен;
- XSD написана на XML.

Тобто, застосовуючи описуючу XML schema можна вважати нашу систему спорідненим середовищем XML. Що дає нам змогу застосувати в нашій XML-СУБД всі документи в XML форматі.

Для перевірки здатності XML-СУБД оперувати з XML schema створимо документ XML та застосуємо до нього опи-

суючі схеми DTD та XML schema відповідно.

XML	DTD	XML schema
<pre>&lt;?xml version="1.0" ?&gt; &lt;doc&gt; &lt;person age = '55'&gt;&lt;name&gt;Peter&lt;/name&gt; &lt;addr&gt;Old Street, 25&lt;/addr&gt; &lt;child&gt; &lt;person&gt; &lt;name&gt;John&lt;/name&gt; &lt;addr&gt;UStreet, 16&lt;/addr&gt; &lt;hobby&gt;swimming&lt;/hobby&gt; &lt;hobby&gt;cycling&lt;/hobby&gt; &lt;/person&gt; &lt;/child&gt; &lt;child&gt; &lt;person&gt; &lt;name&gt;Robert&lt;/name&gt; &lt;addr&gt;Old Street, 25&lt;/addr&gt; &lt;/person&gt; &lt;/child&gt; &lt;person age='20'&gt; &lt;name&gt;Mary&lt;/name&gt; &lt;addr&gt;Queensway, 34&lt;/addr&gt; &lt;hobby&gt;painting&lt;/hobby&gt; &lt;/person&gt; &lt;/doc&gt;</pre>	<pre>&lt;!ELEMENT doc (person)*&gt; &lt;!ELEMENT person (name, addr, (hobby)*, (child)*)&gt; &lt;!ATTLIST person age CDATA #IMPLIED&gt; &lt;!ELEMENT child person&gt; &lt;!ELEMENT name #PCDATA&gt; &lt;!ELEMENT addr #PCDATA&gt; &lt;!ELEMENT hobby #PCDATA&gt;</pre>	<pre>&lt;?xml schema xmlns:xs="http://www.w3.org/2001/XMLSchema" hema"&gt; &lt;xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" &lt;xs:element name="addr"&gt; &lt;xs:complexType mixed="true" /&gt; &lt;/xs:element&gt; &lt;xs:element name="child"&gt; &lt;xs:complexType&gt; &lt;xs:sequence&gt; &lt;xs:element ref="person" /&gt; &lt;/xs:sequence&gt; &lt;/xs:complexType&gt; &lt;/xs:element&gt; &lt;xs:element name="doc"&gt; &lt;xs:complexType&gt; &lt;xs:sequence&gt; &lt;xs:element ref="person" maxOccurs="unbounded" /&gt; &lt;/xs:sequence&gt; &lt;/xs:complexType&gt; &lt;/xs:element&gt; &lt;xs:element name="hobby"&gt; &lt;xs:complexType mixed="true" /&gt; &lt;/xs:element&gt; &lt;xs:element name="name"&gt; &lt;xs:complexType mixed="true" /&gt; &lt;/xs:element&gt; &lt;xs:element name="person"&gt; &lt;xs:complexType&gt; &lt;xs:choice&gt; &lt;xs:element ref="addr" /&gt; &lt;xs:element ref="child" /&gt;</pre>

Рис. 3. XML документ та дві описуючі схеми відповідно

Виходячи з отриманих результатів можна спрогнозувати, що XML-СУБД можна застосувати для виконання XQUERY запитів в розподільчій системі. Web-сервіси (WS) позиціонуються в даний час як універсальна технологія зв'язування суттєво різномірних систем. Більш продуктивним засобом реалізувати Web сервіси є використання WebMethods оболонки Microsoft ASP.NET. З ASP.NET поставляється спеціальний клас для IHttpHandler. Asmx (званих WebServiceHandler), що забезпечує набір необхідних вам функціональних можливостей XML, XSD, SOAP і WSDL.

Для здійснення системи доступу до СУБД будемо використовувати ідеологію трьохрівневого додатку. Трьохрівневий додаток включає до себе наступні рівні:

1. Рівень представлення.
2. Рівень бізнес-логіки.
3. Рівень доступу до даних.

Рівень представлення звичайно являє собою будь-який різновид “тонкого” клієнта, який тільки викликає сервер, відображає дані і приймає введення від користувача.

Рівень бізнес-логіки виконує всю основну обробку даних. Даний рівень являє собою проміжну ланку між БД і клієнтським додатком і може працювати як, наприклад, монітор обробки транзакцій або брокер об'єктних запитів.

Рівень доступу до даних зазвичай є власне БД.

Трьохрівнева модель є доволі гнучкою, тому при зміні алгоритмів обробки даних часто достатньо змінити ПЗ на машині проміжної ланки, що може обслуговувати безліч клієнтських машин. Крім того, потужні машини проміжної ланки можуть витримувати значно більші навантаження, ніж клієнтські машини. Введення проміжної ланки також полегшує синхронізацію доступу до БД.

У нашому випадку клієнтський додаток буде представлено у вигляді ехе-файлу, а рівні бізнес-логіки і доступу до даних – у вигляді dll-бібліотек. Усі три проекти помістимо в одну групу проектів:

– Рівень доступу к даним – база даних XML.

– Рівень бізнес-логіки – бібліотеки DataAccess.dll і BusinessLogic.dll.

– Рівень представлення – клієнтський додаток.

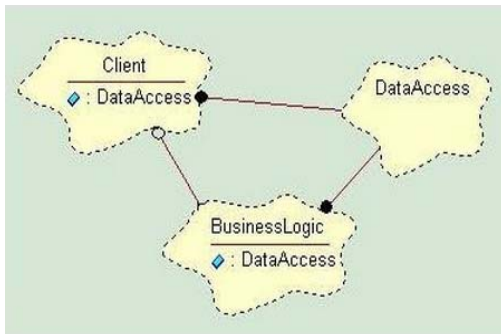


Рис. 4. Схема взаємодії між класами програми

Тому для реалізації будемо використовувати *VisualStudio2005* для написання *Web* сервісу. *Web* сервіс можна представити лише одним файлом з розширенням *ASMX* із наступним заголовком

```
<%@ WebService Language="c#"
    Class="WebService1.TWebService1" %>
```

Далі може слідувати код реалізуючий функціональність *Web* сервісу, котрий написаний на мові *C#*. Автоматично сгенерований файли *<Ім'я сервісу>.asmx* у наступному вигляді:

```
<%@ WebService Language="c#"
    Debug="true"
    Codebehind="WebService1.pas"
    Class="WebService1.TWebService1" %>
```

За для перевірки сервісу достатньо в браузері ввести строку наступного вигляду: *http://localhost/<шлях до сервісу>/<ім'я сервісу>.asmx*.

За для реалізації обміну даними між сервером та клієнтом слід застосувати *WSDL*. *WSDL* – це *Web Service Document Language*, мова описання *XML* документів, котрі використовуються для віддаленого вивозу процедур із клієнта на сервер по протоколу *SOAP*. Тому для отримання нашого *Web*-сервісу на мові *WSDL* треба ввести в браузері наступне:

*http://localhost/<шлях до сервісу>/<ім'я сервісу>.asmx?wsdl*.

Для використання с *Web*-сервісу створимо *VCL Forms* додаток.

Застосувавши параметри аутентифікації можна дістати до нашої *XML*-СУБД.

Слід ввести параметри нашого *XQUERY* запити.

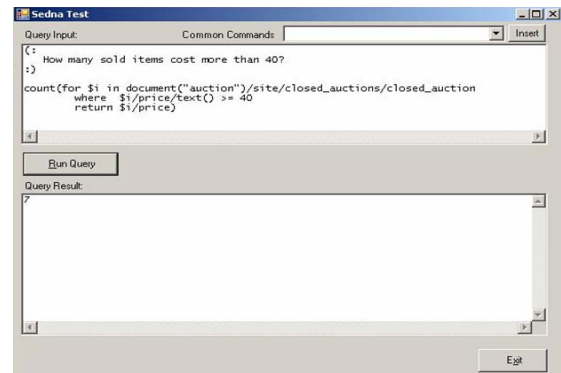


Рис. 5. Отримання результату *XQUERY* запити

### Висновки

Отримані результати доводять переваги використання *XML*-СУБД для оперування з даними формату *XML* та змогу виконання *XQUERY* запитів в розподільчій системі. Таким чином *XML*-СУБД можна розглядати як дуже гарну альтернативу при оперуванні з *XML* документами у порівнянні із комерційними СУБД.

### Список літератури

1. *The World Wide Web Consortium (W3C)*. Extensible Markup Language (XML) 1.0 (Third Edition) W3C Recommendation. Jean Paoli, C. M. Sperberg-McQueen et al. 04 Feb. 2004.

2. *Кузнецов Сергей Дмитриевич* Основы баз данных. – 1-е изд. – М.: «Интернет-университет информационных технологий – ИНТУИТ.ру», 2005. – С. 488.

3. *Чемберлин Д.* W3C XML: XQuery от экспертов. Руководство по языку запросов – М.: Кудиц-Образ, 2005. – 480 с.

4. Программирование баз данных Oracle для профессионалов. – М.: Диалектика, 2007. – С. 375–384.

5. *P. Pleshachkov, L. Novak*. Transaction Isolation In the Sedna Native XML DBMS. Proc. SYRCoDIS 2004, Saint-Petersburg, Russia.

6. *The World Wide Web Consortium (W3C)*. XML Schema Part 1: Structures :W3C recommendation. Biron P.V.,Malhotra A. (eds).–2nd edition. 2004, 28 Oct.

Подано до редакції 12.02.10