

Análisis comparativo de dos Sistemas de Bases de Datos XML

Silvina Migani, Cristina Vera

Departamento de Informática / FCFN/ UNSJ
Av. Ignacio de la Roza 590 (O)
Teléfonos: 4260353 - 4260355
silvina.migani@gmail.com; civera2@yahoo.com.ar

Resumen

El mundo de la informática se ha visto profundamente influenciado por el advenimiento de la World Wide Web y por una tecnología relacionada, el lenguaje XML. Es un lenguaje de marcado de carácter general que brinda una forma simple para representar datos. Sus propiedades de ser universal y extensible le abrieron paso a un extenso y variado uso en las aplicaciones informáticas, generando la necesidad de almacenar y manipular este nuevo tipo de datos [1] [3] [9].

Consecuentemente los fabricantes de motores de bases de datos debieron asumir esa nueva realidad y brindar las funcionalidades requeridas. Por un lado, los SGBDs tradicionales extendieron sus capacidades y, por otro, surgieron SGBDs XML específicos, llamados comúnmente XML nativos [2].

Este trabajo muestra un estudio comparativo de los motores de bases de datos eXist 2.1 y Oracle 11g R2, representantes de los SGBDs XML Nativos, y SGBDs Tradicionales con soporte XML, respectivamente. El mismo se realizó aplicando el benchmark XMark [12][19].

Palabras clave: Bases de Datos - XML

Contexto

El trabajo se enmarca dentro del Proyecto de Investigación (CICITCA) “XML: TÉCNICAS DE GESTIÓN E INTERCAMBIO DE DATOS” - 21/E915, en concordancia con uno de los objetivos perseguidos: “Estudiar la tecnología XML en distintos tipos de motores de bases de datos”.

Introducción

El impacto de la utilización masiva de los datos XML provocó hechos salientes en el área de las bases de datos:

- El surgimiento del lenguaje SQL/XML: El lenguaje estándar utilizado por los motores de bases de datos relacionales, SQL, ha sido ampliado con el objetivo de manejar estructuras XML, dando lugar a la aparición de SQL/XML (definido por la ISO/IEC 9075-14) [7].
- La extensión de los SGBDs Tradicionales: Los motores de bases de datos comerciales tradicionales implementaron las extensiones propuestas al lenguaje SQL para poder almacenar y manipular datos XML [2].

- La aparición de los SGBDs XML Nativos: Estos sistemas brindan un soporte nativo a los datos XML, esto es, XML es la estructura subyacente, además dan soporte a los lenguajes XML específicos XPath y XQuery [1] [2][3][9].

Por lo tanto, hoy en día existe una gran cantidad de Sistemas de Bases de Datos disponibles en el mercado que manipulan datos XML, cada uno con características de funcionalidad y rendimiento diferentes.

Los benchmarks constituyen un marco de referencia que posibilitan la evaluación de diferentes tipos de sistemas en general. Específicamente para el caso de motores de bases de datos XML, se han propuesto varios, y XMark es uno de ellos [12][19]. El mismo toma como escenario de aplicación un sitio de subastas en internet, e incluye veinte consultas XQuery que intentan cubrir los usos más representativos.

Este artículo describe, en líneas generales, la utilización del benchmark XMark como marco de comparación de los motores de bases de datos eXist 2.1 (SGBD XML Nativo) y Oracle 11g R2 (SGBD Tradicional con soporte XML).

Los SGBDs eXist 2.1 y Oracle 11g R2

eXist es un SGBD XML nativo open source [11][18] que brinda soporte a los estándares XPath, XQuery, XInclude y XSLT [6][8] [13] [14] [15][16] [17]. Asimismo integra algunas técnicas avanzadas como búsquedas de términos, por proximidad de términos, y basadas en expresiones regulares. En cuanto al modelo de almacenamiento, los documentos se almacenan en colecciones, las cuales pueden estar anidadas; desde un punto de vista práctico el almacén de datos funciona como un sistema de ficheros. Cada

documento está en una colección. Los documentos no necesitan tener una DTD o esquema asociado [1][4], y dentro de una colección pueden almacenarse documentos de cualquier tipo [11].

El Motor de Base de Datos Oracle [10] [5] desde la versión 9iR2 proporciona un tipo de dato nativo llamado XMLTYPE; el cual permite guardar información en formato XML con soporte a XSD, XSLT, XPath, XQuery, indexación y particionamiento de documentos XML. En las versiones 9i y 10g los documentos XML se almacenan internamente como CLOB, pero a partir de la versión Oracle Database 11g se guardan en formato binario. Este nuevo modo de almacenar es más eficiente en el consumo de espacio y en el tiempo de respuesta para acceder a los datos y es el modo de almacenamiento por defecto a partir de la versión 11.2.0.2.

Evaluación de los SGBDs según XMark

El benchmark XMark provee veinte consultas XQuery y tres documentos XML sobre los que se deben ejecutar las consultas, cada uno correspondientes a los tamaños Tiny, Small y Standard [12]. Las consultas pueden agruparse según el tipo de tarea que realizan en:

- Grupo I: Consultas relacionales simples con comparaciones sobre varios tipos de valores de datos.
- Grupo II: Consultas de documentos que preservan el orden de los elementos.
- Grupo III: Consultas de navegación.
- Grupo IV: Consultas con operaciones de agregación y ordenamiento.

Cabe aclarar que no pudieron ser ejecutadas todas las consultas para los tres tipos de documentos de entrada en

ambos motores, ya que en algunos casos dieron error.

A continuación, a modo de ejemplo, se muestra la ejecución de la consulta Q1 en ambos motores sobre el tamaño de documento Tiny (Figura 1 y Figura 2).

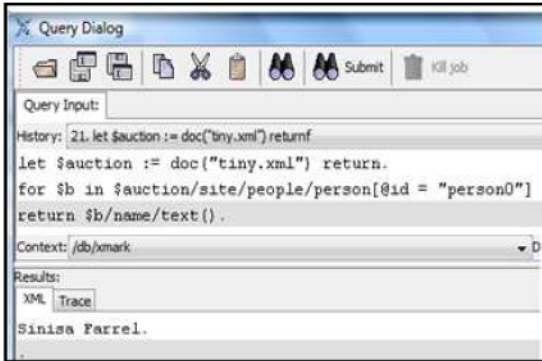


Figura 1: Ejecución de la consulta Q1 en eXist

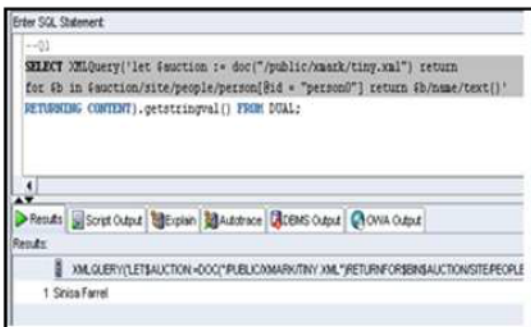


Figura 2: Ejecución de la consulta Q1 en Oracle

La tabla que se detalla a continuación resume los tiempos de respuesta obtenidos luego de la ejecución de las consultas en ambos motores.

Consultas	eXist		Oracle	
	Tiny	Small	Tiny	Small
Q1	0,191	0,22	0,0009	7,4858
Q2	0,311	1,362	5,1407	12,1018
Q3	1,011	0,611	5,6475	26,5018
Q4	0,761	1,973	4,8153	7,4162
Q5	0,641	4,606	14,4587	19,4509
Q6	0,151	0,07	4,6337	7,8848
Q7	0,841	0,6	5,0249	8,6359
Q8	1,352	44,535	65,9427	-
Q9	4,296	59,966	93,1457	-
Q10	1,762	7,792	11,6325	-
Q11	1,653	65,164	15,861	132,3187
Q12	1,883	66,265	26,181	98,1834
Q13	0,49	0,661	4,5853	8,0273
Q14	0,391	1,262	-	-
Q15	1,202	1,833	5,2759	7,5954
Q16	1,222	2,564	5,5686	6,7993
Q17	0,18	0,43	5,8153	8,722
Q18	0,411	1,663	-	-
Q19	0,301	1,682	5,3592	14,2544
Q20	0,22	0,58	1,194	22,7338

Tabla 1: Tiempos de Respuestas en eXist y Oracle

Luego de la ejecución de las consultas en ambos motores se procedió a la comparación, análisis y evaluación de los mismos, según los criterios que propone XMark. Para facilitar el análisis, se confeccionaron gráficos de barra con los valores obtenidos.

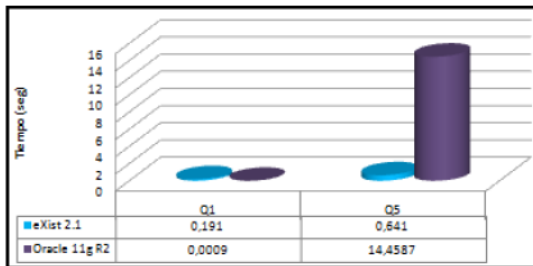


Figura 3: Consultas Grupo I

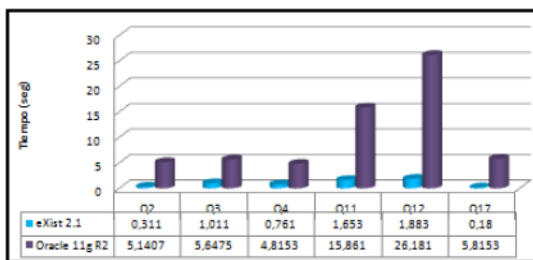


Figura 4: Consultas Grupo II

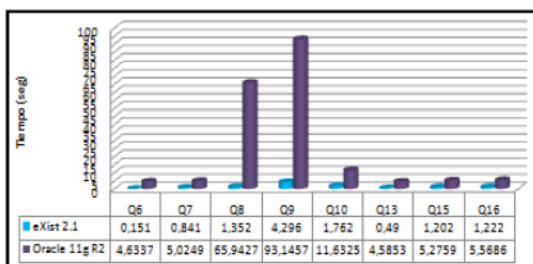


Figura 5: Consultas Grupo III

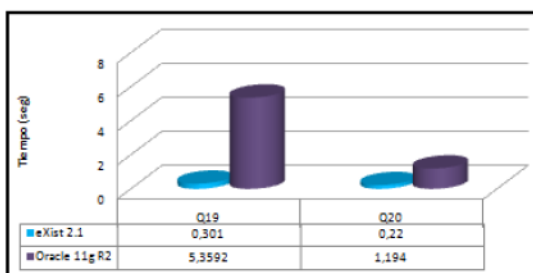


Figura 6: Consultas Grupo IV

Los tiempos de respuesta obtenidos en las diferentes consultas analizadas en ambos SGBDs, demuestran claramente una superioridad por parte de eXist sobre Oracle. Y concretamente, donde eXist alcanza la diferencia más significativa, es en las consultas de navegación. En el otro extremo, quedan las consultas que

requieren agregación y ordenamiento para la obtención de sus resultados.

Por lo tanto, se desprende que la mayor fortaleza de eXist sobre Oracle, se pone de manifiesto en operaciones típicas de aplicaciones centradas en los documentos. Mientras que su comportamiento tiende a descender, en operaciones que son más representativas de las aplicaciones centradas en los datos.

Líneas de Investigación, Desarrollo e Innovación

Además del estudio y profundización de los conceptos inherentes al soporte de XML en bases de datos, se investigaron y experimentaron:

- Oracle 11 g con licencia OTN. Sitio de descarga <http://www.oracle.com.ar>
- eXist 2.1, con licencia GPL. Sitio de descarga <http://www.exist-db.org/download.html>

Resultados y Objetivos

Este trabajo se alinea fundamentalmente a los siguientes objetivos planteados inicialmente:

- Profundizar en el estudio de la tecnología XML y de los motores de bases de datos que gestionan este nuevo tipo de datos.
- Efectuar un análisis comparativo de la gestión de datos XML abordada con un motor de base de datos nativo y con un motor de base de datos relacional extendido.

Formación de Recursos Humanos

- Se concluyó la Beca Interna de Investigación en la CATEGORÍA INICIACIÓN, convocatoria 2012, en el marco de la Ordenanza Nro. 10/05-CS: Licenciada Cristina Vera. El

informe final fue presentado en Setiembre de 2014.

- Se iniciaron dos trabajos de Tesis de Maestría dentro de esta línea de investigación.

Referencias

- [1]. "Introducción a XPath 2.0." Disponible en <http://www.mulberrytech.com/papers/XPath-2-0-User-Grp-HTML/contents.html>.
- [2]. "W3c XPath". Disponible en <http://www.w3.org/TR/xpath/#section-Introduction>.
- [3]. "W3C XSLT 2.0." Disponible en http://www.w3c.es/Prensa/2005/nota051103_xslt-xquery-xpath.html.
- [4]. "W3C XSLT 3.0." Disponible en <http://www.w3.org/TR/xslt-30/>.
- [5]. "XPath W3C". Disponible en <http://www.w3.org/TR/xpath-30>.
- [6]. "XQuery 3.0" W3C. Disponible en <http://www.w3.org/TR/xquery-30>.
- [7]. Benz, Durant. "XML Programming Bible". Wiley Publishing Inc.
- [8]. Chaudhri, Rashid, Zicari. "XML Data Management: Native XML and XML-Enabled Database Systems". Addison Wesley 2003.
- [9]. Elmasri, Navathe. "FUNDAMENTALS OF DATABASE SYSTEMS". Cuarta Edición. Addison Wesley.
- [10]. Evjenet. "Professional XML". Wrox Press 2007.
- [11]. Francisco Riccio. "Manejando XMLType en Oracle Database 11gR2 a través del componente XDB". Publicado en abril 2013. Disponible en <http://www.oracle.com/technetwork/es/articles/sql/xmltype-en-database11g-a-traves-xdb-1931103-esa.html>.
- [12]. <http://www.xml-benchmark.org/> "XMark — An XML Benchmark Project".
- [13]. ISO/IEC 9075-14:2003 Information technology – Database languages – SQL – Part 14: "XML-Related Specifications, (SQL/XML)"
- [14]. Jim Melton, Stephen Buxton. "Querying Xml: XQuery, XPath, and SQL/XML in Context". Morgan Kaufmann 2006.
- [15]. Molina, Ullman, Widom. "DATABASES SYSTEMS: THE COMPLETE BOOK." Prentice Hall.
- [16]. Oracle® "XML DB Developer's Guide 11g Release 1 (11.1)." Mayo de 2008.
- [17]. Sitio oficial de eXist. Disponible en <http://exist-db.org>.
- [18]. Yin-Fu Huang, Shing-Hang Wang. "eXist: An Open Source Native XML database". 30 de Julio de 2009.
- [19]. Yu Yang. "Benchmarking of Native XML Database System". Universidad de Wollongong. 2005.