

MP_0373.
**Lenguajes de marcas y sistemas de
gestión de información**

**UF5. Almacenamiento y
manipulación de la información**

**5.4. Almacenamiento
XML nativo**

Índice

☰	Objetivos	3
☰	Conceptos	4
☰	Características	5
☰	Aplicaciones	6
☰	Manejo de colecciones	8
☰	Base de datos clásica frente a base de datos nativa	9
☰	Áreas de aplicación	10
☰	Resumen	11

Objetivos

Con esta unidad perseguimos los siguientes objetivos:

- 1 Conocer los conceptos fundamentales de las bases de datos XML nativas y sus características.
- 2 Conocer cómo es el manejo de colecciones en bases de datos XML nativas.
- 3 Conocer las diferencias con otros sistemas.
- 4 Conocer las áreas de aplicación.

Conceptos

XML está provocando la aparición de una nueva generación de bases de datos.

Las bases de datos XML nativas (NXD) son muy distintas a las relacionales, ya que tienen capacidades para soportar el almacenamiento de documentos XML, aunque los datos almacenados se derivan al formato relacional usando tablas, registros y columnas, o en otras ocasiones almacenando todo el documento en formato *Binary Large Object* (BLOB).

La cualidad principal de las bases de datos no nativas XML que trabajan con documentos XML es la capacidad de obtener resultados de las consultas en formato XML; es por ello que dichas bases de datos pertenecen a la categoría de "XML-Enabled database".

Un **SGBD XML-Enabled** es un SGBD tradicional con capacidad para el tratamiento y obtención de documentos XML.

Hay que tener en cuenta que los grandes fabricantes de SGBD están aumentando las capacidades para la inserción y el tratamiento de documentos XML dentro de sus productos, incluyendo características de SGBD XML nativas, por lo que en un futuro próximo es posible que la frontera entre estos dos tipos de gestores quede diluida.

Las bases de datos más relacionadas con XML son las "Native XML Database" NXD o bases de datos XML nativas.

Características

¿En qué consisten las bases de datos nativas en XML?

1

Modelo lógico.

Las bases de datos nativas definen un modelo lógico para el documento XML, almacenan y recuperan documentos de la misma manera que los XML. Por lo tanto, el **modelo debe incluir atributos como PCDATA y documentos en orden**.

Ejemplos de estos modelos son: XPath, XML Infoset y modelos que implican DOM y SAX 1.0.

2

Unidad de información.

Tienen un documento de XML como **unidad fundamental del almacenamiento**, como si en una base de datos relacional se tiene una fila en una tabla como unidad fundamental del almacenamiento lógico.

3

Estructura física.

No se requiere tener ningún modelo de almacenamiento físico en particular. Por ejemplo, el documento puede ser creado en una base de datos relacional, jerárquica, u orientada a objetos, o utilizar un formato de almacenamiento propietario tal como archivos indexados y comprimidos.

Aplicaciones

Existen diversos productos que brindan diferentes funcionalidades para las bases de datos nativas en XML, pero generalmente tienen las siguientes características.

Procesamiento de datos

El procesamiento de datos en este tipo de bases de datos es costoso y muy poco amigable, esto es debido al formato jerárquico en el que se almacena la información.

En la mayoría de SGBD XML nativas la recuperación del documento es completa, su tratamiento se realiza a través del API de referencias SAX o DOM y posteriormente se vuelve a almacenar el documento en el repositorio. Esto ocurre porque aún no hay un lenguaje estándar que permita de forma sencilla y procedimental la actualización, inserción o eliminación de elementos de un documento XML.

Xupdate es un lenguaje que permite realizar actualizaciones en un documento XML, pero aún no es un estándar y muchos gestores de este tipo de bases de datos no lo soportan.

Almacenamiento

Una base de datos nativa en XML almacena la información en documentos XML. Este tipo de bases de datos tienen repositorios con un formato "tipo XML", como puede ser DOM o Infoset.

En este mismo "repositorio" (paquete de archivos) se almacenan los índices que se generan por cada documento XML almacenado.

Búsquedas

Este tipo de bases de datos no utiliza SQL como lenguaje de consulta. En lugar de ello utilizan **Xpath**. Algunas bases de datos permiten seleccionar los elementos que deberán tener índices, mientras que otras bases de datos indexan todo el contenido del documento.

El problema que tienen las búsquedas en este tipo de bases de datos es que no permiten realizar búsquedas muy complicadas, como por ejemplo ordenamiento y *cross join*, debido a que Xpath no fue creado realmente para búsquedas en bases de datos, sino simplemente para búsquedas en un solo documento.

Muchas bases de datos permiten realizar búsquedas utilizando la tecnología *Full-Text Search*, de esta manera se puede agilizar la búsqueda de datos en los documentos XML.

Manejo de colecciones

Se espera que en un futuro los esquemas propuestos por el W3C se impongan como estándar para las NXD.

Las NXD manejan colecciones de documentos, permitiendo que se consulten y se manejen como un conjunto. Este método es muy similar al concepto relacional de una tabla.

Las NXD se diferencian del concepto de las tablas relacionales en que no todas las bases de datos nativas de XML requieren un esquema para estar asociado a una colección. Esto significa que se puede almacenar cualquier documento de XML en la colección, sin importar el esquema.

Al hacer esto se pueden construir consultas a través de todos los documentos en la colección. Las NXD que utilizan estas funciones se llaman esquema-independientes. Tener colecciones esquema-independientes del documento da a la base de datos mucha flexibilidad y hace que el desarrollo de las aplicaciones sea más fácil.

Si se necesita una estructura de esquemas fuerte, hay que asegurarse de que se utiliza una NXD que soporte esquemas, o se deberá encontrar otra manera de almacenar los datos del XML.

Base de datos clásica frente a base de datos nativa

A continuación veremos las diferencias entre una base de datos clásica y una base de datos nativa.

- 1 Una base de datos XML nativa provee de su propio modelo de datos.
- 2 Un SGBD XML-Enabled solo puede manejar y almacenar los documentos que encajan dentro del modelo definido para ellos, mientras que un **SGMD XML nativo** debe manejar todos los tipos de documentos posibles.
- 3 Todas las bases de datos relacionales están centradas en los datos, pues lo que ellas almacenan en sus campos son datos simples, más conocidos como datos atómicos.
- 4 Una base de datos nativa en XML no posee campos, ni almacena datos atómicos, lo que almacena son documentos XML, por lo tanto a este tipo de bases de datos se les denomina bases de datos centradas en documentos.

Áreas de aplicación

En general, cualquier NXD es una excelente herramienta para almacenar documentos orientados a datos. Básicamente, si la información es representada por XML una NXD probablemente es una buena solución.

Una NXD puede almacenar cualquier tipo de información XML, pero probablemente no sería la mejor herramienta para algo como un sistema de contabilidad, donde los datos están muy bien definidos y son rígidos.

Algunas áreas potenciales de aplicación podrían ser:

- Portales de información corporativa.
- Información de catálogos.
- BD en partes de manufactura.
- Información médica.
- BD personalizadas.

Resumen

Has terminado la lección, veamos los puntos más importantes que hemos tratado.

- En esta lección hemos visto una introducción a la importancia que tendrán en un futuro las bases de datos XML nativas.
- XML está provocando la aparición de una nueva generación de bases de datos, que aunque en la actualidad se encuentran en una fase de investigación y desarrollo, en breve se convertirán en una buena alternativa a las bases de datos tradicionales.
- Las bases de datos más relacionadas con XML son las "Native XML Database", NXD o bases de datos XML nativas.



PROEDUCA