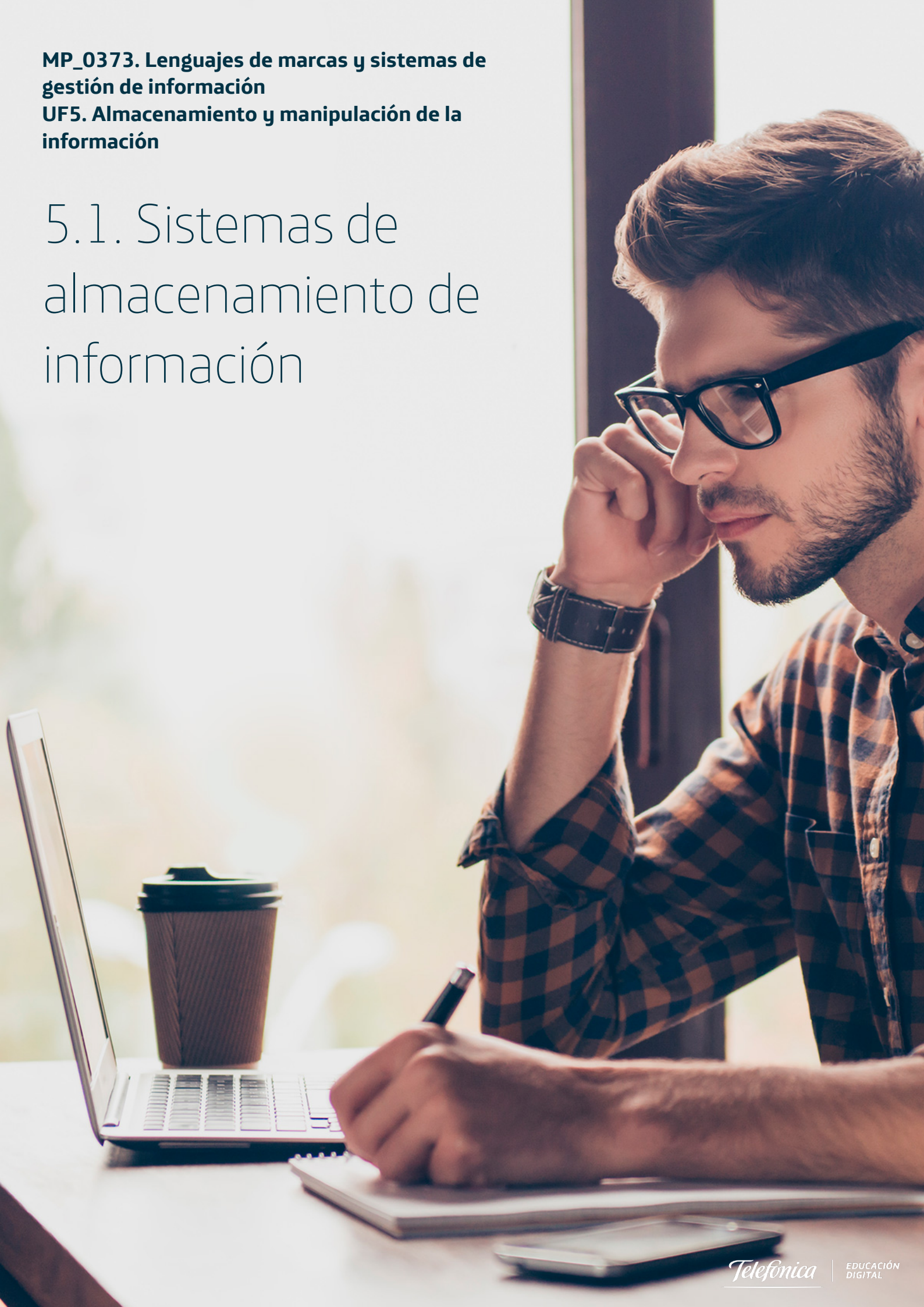


## 5.1. Sistemas de almacenamiento de información



# Índice

---

Objetivos .....	3
Conceptos básicos .....	4
Bases de datos relacionales.....	4
Bases de datos no relacionales .....	4
Ventajas y desventajas del uso de XML .....	5
Aplicación de XML .....	5
Almacenamiento en XML.....	6
Tipos de bases de datos XML .....	6
Semejanzas entre una BD y XML .....	6
Bases de datos relacionales con XML .....	7
XML y bases de datos orientadas a objetos .....	8
Despedida .....	9
Resumen.....	9

# Objetivos

En esta unidad perseguimos los siguientes objetivos:

- Saber cuáles son los **sistemas de almacenamiento y su finalidad**.
- Conocer las **características de los sistemas** que almacenan ficheros XML.
- Conocer la **aplicación de XML** en el almacenamiento de datos.
- Aprender los **diferentes formatos** de almacenamiento de XML.
- Saber cómo se adaptan las **bases de datos relacionales** a trabajar con XML.
- Conocer la relación entre **XML y las bases de datos orientadas a objetos**.

# Características de XML

## Conceptos básicos

Los sistemas de almacenamiento de datos se han adaptado a la entrada del estándar XML en la persistencia de los datos y su estructuración.

Este tipo de sistemas de almacenamiento trabajan con varios sistemas lógicos y físicos.

- En cuanto a los **sistemas físicos**, son conocidos y su arquitectura no afecta al aspecto, que es de lo que vamos a tratar en esta unidad.
- Sin embargo, sí debemos revisar los **sistemas lógicos**, porque dependiendo de cómo se organice la información y cómo se acceda a ella, comprobaremos que existen distintos tipos.

## Bases de datos relacionales

Almacenar los documentos XML en bases de datos relacionales supone una gran ventaja, ya que se usan ampliamente en las aplicaciones existentes. Podemos almacenar datos XML en dichas bases de datos de forma que se pueda acceder a ellos de una forma relacional.

- **Almacenamiento como cadena.** Una forma sencilla es almacenar cada elemento hijo del elemento de mayor nivel como una cadena en una tupla (registro) separada de la base de datos. Esta es una arquitectura en la que la búsqueda es bastante complicada, ya que se debe recorrer todo el XML para encontrar un registro concreto.
- **Representación en árbol.** Los datos XML se pueden modelar como un árbol y almacenar mediante el uso de un par de relaciones. De este modo, a cada elemento y atributo se le proporciona un identificador único.
- **Asignación a relaciones.** Según este enfoque, los elementos XML cuyo esquema es conocido se asignan a relaciones y atributos.

## Bases de datos no relacionales

Existen varias alternativas para almacenar datos XML en sistemas de almacenamientos de datos no relacionales:

- **Almacenamiento en archivos planos.** Puesto que XML es principalmente un formato de archivo, un mecanismo de almacenamiento natural es, simplemente, un archivo plano ordenado en el sistema.
- **Almacenamiento en una base de datos XML.** Las bases de datos XML usan XML como su modelo de datos básico. Las bases de datos XML antiguas implementaban el modelo de objetos documento sobre una base de datos orientada a objetos basada en C++.

## Ventajas y desventajas del uso de XML

El uso de XML para almacenar datos de forma legible presenta algunas ventajas, y también inconvenientes.

### Ventajas

- **XML es un estándar potente y de amplia aceptación para guardar y comunicar información** acerca de objetos. Su forma de estructurar la información en árbol permite añadir una dimensión más a los registros de una base de datos.
- **Permite la codificación de información separada**, tal y como se debe presentar al usuario.
- Una base de datos XML se puede ver como **una colección de documentos XML**, en la que cada documento representa un registro.
- **Cada documento XML es un archivo en el sistema de archivos y contiene una cadena válida XML.**
- La estructura de un documento XML suele seguir un mismo esquema XML, aunque no es necesario que sea así. De este modo, cada archivo se puede configurar de forma estructurada, por lo que **es independiente, pero fácilmente accesible.**
- Tener colecciones de documentos con un esquema independiente **proporciona a la base de datos flexibilidad y facilita el desarrollo de la aplicación.**
- En los últimos años se ha hecho necesaria la existencia de estándares de intercambio de información con el objetivo de que las organizaciones puedan **compartir su información de una manera más cómoda, automática y eficiente.**

### Inconvenientes

- Para encontrar un fragmento específico de información en los contenidos de un nodo o atributo XML **hay que procesar completamente el archivo XML.**
- Los documentos XML son indexados de forma compleja para su acceso rápido.
- Aunque conserven relaciones entre ellos, es **complicado realizar consultas Join sobre documentos XML.**
- Los elementos anidados y los elementos que se repiten (correspondientes a atributos con valores de conjunto) **complican el almacenamiento de los datos XML en un formato relacional.**

## Aplicación de XML

Los sistemas de almacenamiento XML deben acomodarse dependiendo del uso y ámbito de la aplicación que los va a usar.

Las **bases de datos relacionales tradicionales son mejores para tratar con requerimientos centrados en los datos**, mientras que los sistemas de **administración de contenido y de documentos suelen ser mejores para almacenar datos centrados en el documento.**

Los sistemas de bases de datos deben ser capaces de exponer los datos relacionales como XML, y almacenar el XML recibido como datos relacionales para transferir, obtener y almacenar los datos requeridos por la aplicación.

Los documentos y los requerimientos de almacenamiento de datos XML pueden ser agrupados en dos categorías generales:

### Centrados en datos

Si el documento XML tiene una estructura definida y contiene datos que pueden ser actualizados y usados de diversos modos, se dice que el documento está centrado en los datos. En este caso, el sistema responde a la organización de **los datos agrupados en documentos que tienen sentido por sí mismos**, como pueden ser documentos de facturación, órdenes de compra y documentos estructurados. También se adapta para generar documentación estructurada, como son los contenidos de periódicos, artículos y publicidad.

### Centrados en los documentos

Los documentos tienden a ser más impredecibles en tamaño y contenido que los centrados en los datos, que son altamente estructurados, con tipos de datos de tamaño limitado y reglas menos flexibles para campos opcionales y contenido.

## Almacenamiento en XML

### Tipos de bases de datos XML

Los datos almacenados en la base de datos XML se pueden consultar mediante **XQuery** serializada y exportar al formato deseado, que puede ser:

- XML enabled.
- XML nativo (NXD).

### Semejanzas entre una BD y XML

Entre una base de datos y un fichero XML con su esquema asociado, podemos establecer las siguientes semejanzas:

- La tecnología XML usa uno o más documentos para almacenar la información.
- Define esquemas sobre la información.
- Tiene lenguajes de consulta específicos para recuperar la información requerida.
- Dispone de API (SAX, DOM).



Pero hay **muchas más cosas que los diferencian**, debido a que XML no es una base de datos:

- La tecnología XML carece, entre otras cosas, de almacenamiento y actualización eficientes.
- No tiene índices.
- Carece de seguridad.
- No dispone de transacciones.
- Carece de integridad de datos.
- No facilita el acceso concurrente.
- Y, por último, carece también de disparadores, que son algunas de las características habituales en las bases de datos.

Por tanto, es imposible pensar que XML se vaya a utilizar para las tareas transaccionales de una organización para las que está más que justificado utilizar alguna base de datos de sistemas muy difundidos, como MySQL, IBM DB, Microsoft SQL, ORACLE, etc.

Por otro lado, XML **permite integrar sistemas de información hasta ahora separados**:

### Almacenamiento directo de ficheros

Es una opción pobre; las opciones que podemos hacer sobre ellos son limitadas y definidas por el sistema. No se pueden realizar operaciones sobre el contenido y deberemos limitarnos al movimiento del documento como unidad.

### Almacenamiento sobre una base de datos

En documentos centrados en datos se puede realizar un mapeo entre los distintos elementos definidos en el documento y el modelo de datos del SGBD.

Esta posibilidad está centrada en documentos debido a que una estructura regular y bien controlada es fácilmente transformable en un esquema relacional. Sin embargo, posee el inconveniente de que **sólo se almacenan los datos que nos interesa conservar y partes del documento se pierden** por el camino (por ejemplo, el formato, comentarios, instrucciones de proceso, etc.) y **a la hora de reconstruir el documento a partir de los datos almacenados obtendremos otro distinto**.

## Bases de datos relacionales con XML

Aún cuando no sea de forma nativa, existen razones para usar los tipos de bases de datos relacionales y los productos de bases de datos existentes para almacenar XML.

- Las bases de datos relacionales y orientadas a objetos son bien conocidas, mientras que **las bases de datos XML nativas son nuevas**.
- Como resultado de la familiaridad con las bases de datos relacionales y orientadas a objetos, los usuarios se inclinan por ellas, especialmente **por el rendimiento**.

- Las bases de datos relacionales se basan en las **relaciones entre tablas mediante claves primarias y foráneas** como único medio para representar los datos del mundo real.
- Su manejo, a nivel administración y programación, está asociado a **lenguajes estándares como DDL, DLM, DCL**.
- Los objetos o estructuras jerarquizadas XML han requerido del estudio de teorías complejas para acoplarse a bases de datos relacionales.
- Existen numerosos **middleware** (software que permite la comunicación entre dos aplicaciones de software independientes) para poder integrar estos sistemas.
- Un programa de **middleware** podría permitir a una base de datos, o sistema gestión de bases de datos SGDB, **acceder a los datos de otra**, que se encarga automáticamente de la transferencia de información desde ficheros o estructuras complejas XML hacia las bases de datos relacionales.

## XML y bases de datos orientadas a objetos

Las bases de datos orientadas a objetos están fabricadas para manejar, almacenar y mantener un modelo de objetos puro. Normalmente no se basan en el modelo relacional, aunque también podrían derivar de él.

Las funcionalidades de las bases de datos orientadas a objetos, su diseño y sus relaciones, están basados en el **paradigma de la programación orientada a objetos** y la usan como diseño funcional.

Como los documentos XML pueden asociarse a objetos estructurados, la relación de esta forma de guardar la información y las bases de datos orientadas a objetos pueden ser una alternativa para el almacenamiento y gestión de documentos XML.

Cuando diseñamos los documentos XML como objetos pasan a ser **gestionados directamente por el SGBD00** (Sistema Gestor de Base de Datos Orientado a Objetos). Para acceder, manipular, borrar, etc. los objetos se usa el lenguaje de consulta OQL (Object Query Language).

EL SGBD00 implementará los mecanismos de indexación, optimización, procesamiento de consultas, etc., ya que estos no son específicos para el modelo XML ni los objetos en general.



# Despedida

## Resumen

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

- En esta unidad hemos conocido los **sistemas de almacenamiento y su finalidad**, aprendiendo las características de los sistemas que almacenan ficheros XML.
- Se han comentado las **ventajas y desventajas que nos ofrece XML** para el almacenamiento de datos, así como los diferentes formatos que podemos usar.
- Hemos visto los diferentes tipos de sistemas de almacenamiento XML, las diferencias entre bases de datos relacionales y no relacionales, y los sistemas basados en datos o en documentos.
- Por último, hemos realizado una introducción a los sistemas de almacenamiento XML utilizando el paradigma orientado a objetos.