

TRATAMIENTO DE LA INFORMACIÓN Y COMPETENCIA DIGITAL

Tema 3. Parte 5. Introducción a las bases de datos

Departament d'informàtica.

Autor: Francisco Aldarias Raya

Diciembre-2023

**Preparació
Proves
d'Accés**

ÍNDEX

1 ¿POR QUÉ GESTIONAR LA INFORMACIÓN?	2
2 BASES DE DATOS	4
2.1 Conceptos	4
2.2 Estructura de una base de datos	6
2.3 Usos de las bases de datos	6
3 SISTEMAS GESTORES DE BASE DE DATOS	6
3.1 Concepto de Sistema Gestor de Base de Datos	6
3.2 Funciones de un SGBD	7
3.3 El lenguaje SQL	8
3.4 Tipos de SGBD	8
3.5 La inteligencia artificial en las base de datos	9
4 BIBLIOGRAFÍA	9

1 ¿POR QUÉ GESTIONAR LA INFORMACIÓN?

En el mundo actual existe cada vez una mayor demanda de datos. Esta demanda siempre ha sido patente en empresas y sociedades, pero en estos últimos años todavía se ha disparado más debido al acceso multitudinario a Internet.

El propio nombre Informática (información + automática) hace referencia al hecho de ser una ciencia que tiene como objetivo automatizar el trabajo con información.

En informática se conoce como dato a cualquier elemento informativo que tenga relevancia para el sistema.

Para poder almacenar datos y cada vez más datos, el ser humano ideó nuevas herramientas como archivos, cajones, carpetas y fichas en las que almacenarlos.

Antes de la aparición del ordenador, el tiempo requerido para manipular esos datos era enorme. Sin embargo el proceso de aprendizaje era relativamente sencillo ya que se usaban elementos que el usuario reconocía perfectamente.

Por esa razón, la informática adaptó sus herramientas para que los elementos que el usuario maneje en el ordenador se parezcan a los que utilizaba manualmente. Así en informática se sigue hablado de ficheros, formularios, carpetas, directorios,....etc.

En los años 70-80 se utilizaban ficheros de texto donde se guardaba la información:

```

alumnos.txt
DNI          NOMBRE          DIRECCIÓN          FECHA NTO
-----
2894512X     José Jiménez Perez    C/ Corredera,34    21-10-90
28924896D    Alejandra Gómez Marín C/ Picaso, 23      11-02-91

asignaturas.txt
DNI          NOMBRE          ASIGNATURA  NOTA
-----
-----
2894512X José Jiménez Perez    Matemáticas 5
2894512X José Jiménez Perez    Lengua      8
.....
28924896D Alejandra Gómez Marín Matemáticas 7
28924896D Alejandra Gómez Marín Inglés      3

```

A partir de los años 90 empiezan a utilizarse los Sistemas Gestores de Bases de Datos (SGBD) donde se comienza a almacenar la información en forma de tablas:

```

Alumnos (DNI, Nombre, Dirección, Fecha nacimiento)
2894512X José Jiménez Perez C/ Corredera,34 21-10-90
28924896D Alejandra Gómez Marín C/ Picaso, 23 11-02-91
...
Asignaturas (Código, Nombre)
001 Matemáticas
002 Lengua
003 Inglés
Notas(DNI, Código_asignatura, nota)
2894512X 001 5
2894512X 002 8
28924896D 001 7
28924896D 003 3

```

Veamos por qué se ha producido esta evolución. ¿Qué ocurriría si quisiéramos conocer el número de alumnos de más de veinticinco años y con nota media superior a siete que además estén matriculados actualmente en la asignatura Bases de datos I?

En un Sistema de Información (SI) sin informatizar, obtener esta información podía requerir mucho tiempo y mucho trabajo, además de ser necesario la realización de cálculos (media,...) así como ir revisando alumno por alumno.

En un SI con ficheros, podríamos crear un programa que fuera obteniendo la información del fichero y realizando los cálculos y devolviendo los resultados.

Sin embargo, en un SI con bases de datos, esta consulta es trivial usando un lenguaje de consulta de datos.

2 BASES DE DATOS

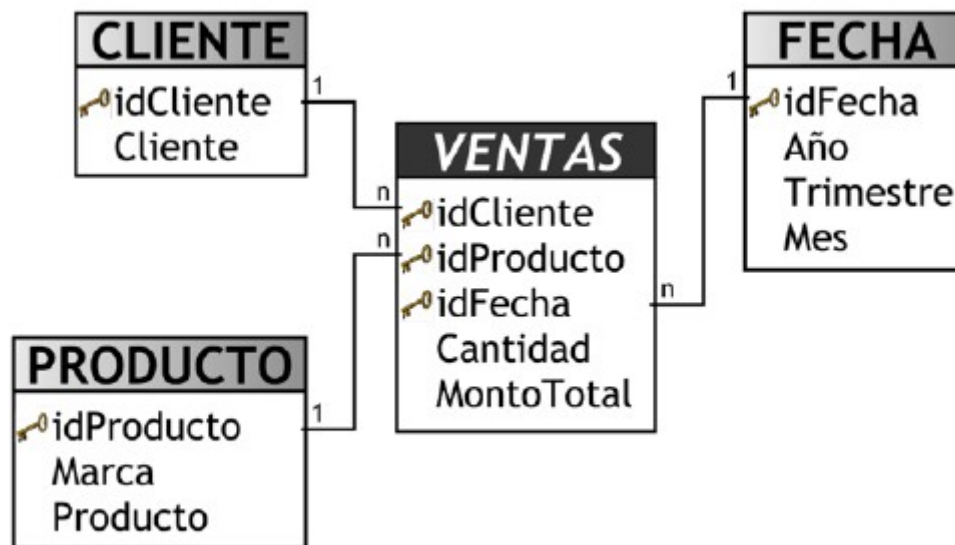
Una Base de Datos (BD) es una colección de información perteneciente a un mismo contexto, que está almacenada de forma organizada en tablas.

Estas tablas se relacionan formando vínculos o relaciones entre ellas, que ayudan a mantener la información de diversos objetos de forma ordenada y coherente (sin contradicciones).

Cada una de estas tablas es una estructura que se parece a las hojas de cálculo, pues está dispuesta mediante filas y columnas. De este modo, cada fila almacena un registro con tantos campos como columnas tenga la tabla.

Por ejemplo, se podría tener una tabla de Cliente, donde cada fila o registro es un cliente de la empresa y cada columna o campo representa un trozo discreto de información sobre cada cliente, por ejemplo su número identificativo o el nombre.

Ejemplo de tablas relacionadas entre sí



Ejemplo de tablas relacionadas entre sí

2.1 Conceptos

Uno de los grandes problemas al que se enfrentan los informáticos cuando comienzan su aprendizaje, es el gran número de términos desconocidos que debe asimilar, incluyendo el enorme número de sinónimos y siglas que se utilizan para nombrar la misma cosa.

Tratando, a modo de resumen, de aclarar algunos de los componentes que se pueden encontrar en una base de datos, y que se verán en próximos capítulos, se definen los siguientes conceptos:

- **Dato:** El dato es un trozo de información concreta sobre algún concepto o suceso. Por ejemplo, 1996 es un número que representa un año de nacimiento de una persona. Los datos se caracterizan por pertenecer a un tipo.

- **Tipo de Dato:** El tipo de dato indica la naturaleza del campo. Así, se puede tener datos numéricos, que son aquellos con los que se pueden realizar cálculos aritméticos (sumas, restas, multiplicaciones...) y los datos alfanuméricos, que son los que contienen caracteres alfabéticos y dígitos numéricos. Estos datos alfanuméricos y numéricos se pueden combinar para obtener tipos de datos más elaborados.

Por ejemplo, el tipo de dato Fecha contiene tres datos numéricos, representando el día, el mes y el año de esa fecha.

- **Campo:** Un campo es un identificador para toda una familia de datos. Cada campo pertenece a un tipo de datos.

Por ejemplo, el campo FechaNacimiento representa las fechas de nacimiento de las personas que hay en la tabla. Este campo pertenece al tipo de dato Fecha. Al campo también se le llama columna.

- **Registro:** Es una recolección de datos referentes a un mismo concepto o suceso.

Por ejemplo, los datos de una persona pueden ser su NIF, año de nacimiento, su nombre, su dirección, etc. A los registros también se les llama tuplas o filas.

- **Campo Clave:** Es un campo especial que identifica de forma única a cada registro. Así, el NIF es único para cada persona, por tanto es campo clave. Hay varios tipos de campos clave como se explicará más adelante.

- **Tabla:** Es un conjunto de registros bajo un mismo nombre que representa el conjunto de todos ellos.

Por ejemplo, todos los clientes de una base de datos se almacenan en una tabla cuyo nombre es Clientes.

- **Consulta:** Es una instrucción para hacer peticiones a una base de datos. Puede ser una búsqueda simple de un registro específico o una solicitud para seleccionar todos los registros que satisfagan un conjunto de criterios. Aunque en castellano, consulta tiene un significado de extracción de información, en inglés query, una consulta es una petición, por tanto, además de las consultas de búsqueda de información, que devuelven los campos y registros solicitados, hay consultas (peticiones) de eliminación o inserción de registros, de actualización de registros, cuya ejecución altera los valores de los mismos.

- **Índice:** Es una estructura que almacena los campos clave de una tabla, organizándolos para hacer más fácil encontrar y ordenar los registros de esa tabla. El índice tiene un funcionamiento similar al índice de un libro, guardando parejas de elementos: el elemento que se desea indexar y su posición en la base de datos. Para buscar un elemento que esté indexado, solo hay que buscar en el índice de dicho elemento para, una vez encontrado, devolver el registro que se encuentre en la posición marcada por el índice.

- **Vista:** Es una transformación que se hace a una o más tablas para obtener una nueva tabla.

Esta nueva tabla es una tabla virtual, es decir, no está almacenada en los dispositivos de almacenamiento del ordenador, aunque sí se almacena su definición.

- **Informe:** Es un listado ordenado de los campos y registros seleccionados en un formato fácil de leer. Generalmente se usan como peticiones expresas de un tipo de información por parte de un usuario.

Por ejemplo, un informe de las facturas impagadas del mes de enero ordenado por nombre de cliente.

2.2 Estructura de una base de datos

Una base de datos almacena los datos a través de un esquema.

□ El esquema es la definición de la estructura donde se almacenan los datos, contiene todo lo necesario para organizar la información mediante tablas, registros (filas) y campos (columnas).

También contiene otros objetos necesarios para el tratamiento de los datos (procedimientos, vistas, índices, etc.) y que se estudiarán más adelante. Al esquema también se le suele llamar metainformación, es decir, información sobre la información o metadatos.

Los gestores de bases de datos modernos Oracle, MySQL y DB2, entre otros, almacenan el esquema de la base de datos en tablas, de tal manera que el propio esquema de la base de datos se puede tratar como si fueran datos comunes de la base de datos.

2.3 Usos de las bases de datos

Las bases de datos podemos encontrarlas en cualquier tipo de sistema informático, a continuación se exponen solo algunos ejemplos de sus usos más frecuentes:

- Bases de datos Administrativas: Cualquier empresa necesita registrar y relacionar sus clientes, pedidos, facturas, productos, etc.
- Bases de datos Contables: También es necesario gestionar los pagos, balances de pérdidas y ganancias, patrimonio, declaraciones de hacienda.
- Bases de datos para motores de búsquedas: Por ejemplo Google o Altavista. Tienen una base de datos gigantesca donde almacenan información sobre todos los documentos de Internet. Posteriormente millones de usuarios buscan en la base de datos de estos motores.
- Científicas: Recolección de datos climáticos y medioambientales, químicos, genómicos, geológicos.
- Configuraciones: Almacenan datos de configuración de un sistema informático como por ejemplo, el registro de Windows.
- Bibliotecas: Almacenan información bibliográfica, por ejemplo, la famosa tienda virtual Amazon o la biblioteca de un instituto.
- Censos: Guardan información demográfica de pueblos, ciudades y países.
- Virus: Los antivirus guardan información sobre todos los potenciales software maliciosos.
- Otros muchos usos: Militares, videojuegos, deportes, etc.

3 SISTEMAS GESTORES DE BASE DE DATOS

3.1 Concepto de Sistema Gestor de Base de Datos

Se define un Sistema Gestor de Base de Datos, en adelante SGBD, en inglés, DBMS (DataBase Management System), como el conjunto de herramientas que facilitan la consulta, uso y actualización de una base de datos.

Un ejemplo de software Gestor de Base de Datos es Oracle 11g, que incorpora un conjunto de herramientas software que son capaces de estructurar en múltiples discos duros los ficheros de una base de datos, permitiendo el acceso a sus datos tanto a partir de herramientas gráficas como a partir de potentes lenguajes de programación (PL-SQL, php, c++. . .).

3.2 Funciones de un SGBD

Los SGBD del mercado cumplen con casi todas las funciones que a continuación se enumeran:

1. Permiten a los usuarios almacenar datos, acceder a ellos y actualizarlos de forma sencilla y con un gran rendimiento, ocultando la complejidad y las características físicas de los dispositivos de almacenamiento.
2. Garantizan la integridad de los datos, respetando las reglas y restricciones que dicte el programador de la base de datos. Es decir, no permiten operaciones que dejen cierto conjunto de datos incompletos o incorrectos.

Un problema que nos encontramos a la hora de almacenar datos es la redundancia, que se define como la cantidad de datos repetidos en la información guardada. Uno de los principales objetivos de una base de datos es reducir todo lo posible la redundancia. Con ello conseguimos dos cosas:

1. Que la información ocupe menos espacio
2. Que sea lo más coherente posible.

La inconsistencia de los datos se produce cuando un dato redundante es diferente en dos o más sitios.

3. Integran, junto con el sistema operativo, un sistema de seguridad que garantiza el acceso a la información exclusivamente a aquellos usuarios que dispongan de autorización.
4. Proporcionan un diccionario de metadatos, que contiene el esquema de la base de datos, es decir, cómo están estructurados los datos en tablas, registros y campos, las relaciones entre los datos, usuarios, permisos, etc. Este diccionario de datos debe ser también accesible de la misma forma sencilla que es posible acceder al resto de datos.
5. Permiten el uso de transacciones, garantizan que todas las operaciones de la transacción se realicen correctamente, y en caso de alguna incidencia, deshacen los cambios sin ningún tipo de complicación adicional.
6. Ofrecen, mediante completas herramientas, estadísticas sobre el uso del gestor, registrando operaciones efectuadas, consultas solicitadas, operaciones fallidas y cualquier tipo de incidencia. Es posible de este modo, monitorizar el uso de la base de datos, y permiten analizar hipotéticos funcionamientos erróneos.
7. Permiten la concurrencia, es decir, varios usuarios trabajando sobre un mismo conjunto de datos. Además, proporcionan mecanismos que permiten arbitrar operaciones conflictivas en el acceso o modificación de un dato al mismo tiempo por parte de varios usuarios.
8. Independizan los datos de la aplicación o usuario que esté utilizándolos, haciendo más fácil su migración a otras plataformas.

9. Ofrecen conectividad con el exterior. De esta manera, se puede replicar y distribuir bases de datos. Además, todos los SGBD incorporan herramientas estándar de conectividad. El protocolo ODBC4 está muy extendido como forma de comunicación entre bases de datos y aplicaciones externas.

10. Incorporan herramientas para la salvaguarda y restauración de la información en caso de desastre. Algunos gestores, tienen sofisticados mecanismos para poder establecer el estado de una base de datos en cualquier punto anterior en el tiempo. Además, deben ofrecer sencillas herramientas para la importación y exportación automática de la información.

3.3 El lenguaje SQL

La principal herramienta de un gestor de base de datos es la interfaz de programación con el usuario. Este interfaz consiste en un lenguaje muy sencillo mediante el cuál el usuario realiza preguntas al servidor, contestando este a las demandas del usuario.

Este lenguaje comúnmente se denomina SQL, Structured Query Language, está estandarizado por la ISO, es decir, todas las bases de datos que soporten SQL deben tener la misma sintaxis a la hora de aplicar el lenguaje.

3.4 Tipos de SGBD

Se pueden clasificar los SGBD de muchas formas, por ejemplo, según las bases de datos que gestionan, clasificando los SGBD según traten bases de datos relacionales, bases de datos orientadas a objetos, etc. Puesto que en la actualidad, la mayoría de los SGBD integran múltiples filosofías y tipos de funcionamiento, en este curso vamos a clasificar los de gestores de bases de datos según la capacidad y potencia del propio gestor:

Los Gestores de Bases de Datos ofimáticas son aquellos que manipulan bases de datos pequeñas (ofimáticas) orientadas a almacenar datos domésticos o de pequeñas empresas. Incluso estos gestores permiten construir pequeñas aplicaciones para ayudar a un usuario inexperto a manipular los datos de una base de datos de forma sencilla e intuitiva.

Un ejemplo de un SGBD ofimático es Microsoft Access y su equivalente aplicación de LibreOffice Base, que poseen tanto una interfaz de usuario muy sencilla como un potente lenguaje de programación (VBA=Visual Basic for Applications) para ofrecer a usuarios avanzados otras posibilidades de gestión mucho más específicas.

Los Gestores de bases de datos Corporativas son aquellas que tienen la capacidad de gestionar bases de datos enormes, de grandes o medianas empresas con una carga de datos y transacciones que requieren un servidor de grandes dimensiones (generalmente un Servidor Unix, o un Windows 2000 Server con altas prestaciones).

Estos gestores son capaces de manipular grandes cantidades de datos de forma muy rápida y eficiente para poder resolver la demanda de muchos (cientos) de usuarios.

Un ejemplo típico de servidor de base de datos Corporativas es el antes comentado Oracle, actualmente, junto con DB2, el servidor de base de datos más potente del mercado (también el más caro).

Precisamente, ese coste tan alto es el que ha desencadenado que se haya recurrido a una solución intermedia entre gestores de base de datos ofimáticas y corporativas. Entre estas soluciones intermedias se encuentra MySQL, un gestor de base de datos que, además de ser gratuito y sencillo, es capaz de manipular gran cantidad de datos cumpliendo prácticamente todos los estándares de la arquitectura ANSI SPARC. Aunque implementa SQL, no tiene un lenguaje de programación propio como SQL Server u Oracle (aunque está en desarrollo), pero a

cambio se integra fácilmente en las típicas soluciones XAMPP, que son paquetes que incluyen, además de MySQL, una versión del servidor Web Apache y varios lenguajes de script (php, perl. . .) que dotan a MySQL de potentes herramientas para acceso y publicación de los datos.

3.5 La inteligencia artificial en las base de datos

El uso de la inteligencia artificial (IA) en las bases de datos(BBDD) está cada vez más presente. El elemento más relevante es la ayuda para realizar consultas a personas inexpertas, y para ellos se utiliza la inteligencia artificial, que permite generar consultas SQL para que el usuario pueda introducirlas en las bases de datos.

Una web que permite el uso de la IA en BBDD mediante la generación de SQL es <https://www.sqlai.ai/>

4 BIBLIOGRAFÍA

- Bases de Datos. Iván López, M.^a Jesús Castellano. John Ospino. Ed. Garceta