

# Tema 1: Introducción a las bases de datos.

## 1.1. Evolución histórica de las bases de datos.

Las bases de datos empezaron a utilizarse a partir de la década de los 70 del siglo XX. Antes de emplearse las bases de datos para almacenar la información se utilizaban archivos o ficheros, por lo que vamos a comenzar explicando el concepto de archivo o fichero.

La parte del ordenador en la que se almacena información se denomina memoria y podemos hablar de dos tipos de memoria fundamentalmente: la memoria principal o memoria RAM y la memoria secundaria.

La memoria RAM es de tipo volátil, lo que quiere decir que la información contenida en ella desaparece al desconectarse el ordenador. Por este motivo, se hace necesaria la existencia de una memoria secundaria en la cual permanezca la información, aunque se apague el ordenador.

La información depositada en la memoria secundaria está organizada en archivos o ficheros, por lo que podríamos decir que un fichero consiste en un conjunto de bytes almacenados de forma organizada en un dispositivo de almacenamiento secundario (disco, CD, DVD, pen drive, etc.).

En los ficheros de datos la información se almacena en unas unidades llamadas registros, cada uno de los cuales a su vez consta de varios campos. Así, por ejemplo, en una empresa que se dedica a la comercialización de productos o servicios se podría disponer de un fichero de clientes. Este fichero constaría de muchos registros, uno por cada uno de los clientes de que dispone la organización. A su vez, cada registro se descompondría en varios campos. Por ejemplo, podríamos almacenar por cada cliente su NIF, nombre, apellidos, dirección, localidad, provincia, e-mail y teléfono. Cada una de estas unidades de información es un campo. En la figura 1 se representa la estructura de registro Cliente y en la figura 2, un ejemplo de información almacenada en el fichero de clientes.

NIF	Nombre	Apellidos	Dirección	Localidad	Provincia	e-mail	Teléfono
-----	--------	-----------	-----------	-----------	-----------	--------	----------

Figura 1: Registro Cliente

12345678C	Ana	Gil Ruiz	Avda. , 5 2º A	Getxo	Vizcaya	<a href="mailto:agr@r.es">agr@r.es</a>	948789989
88776655X	Luis	Gómez García	Mayor, 20	Madrid	Madrid	<a href="mailto:lgg@r.es">lgg@r.es</a>	911111111
...	...	...	...	...	...	...	...
00000000A	María	Pérez España	Menor, 15	Estepona	Málaga	<a href="mailto:mbe@r.es">mbe@r.es</a>	666868900

Figura 2: Contenido del fichero Clientes

Vamos a ver a continuación la forma tradicional de gestionar y almacenar los datos mediante el empleo de ficheros. Supongamos una empresa que necesita mantener información acerca de los clientes a los que atiende y acerca de los productos que vende. Sobre los clientes y los productos será necesario realizar diferentes tratamientos, para cada uno de los cuales se dispondrá de una aplicación.

Pues bien, cada una de estas aplicaciones dispondrá de su propio conjunto de ficheros que contendrán los datos necesarios, los cuales estarán organizados de acuerdo con la forma en que son tratados por la correspondiente aplicación. Los ficheros son diseñados, por tanto, para una determinada aplicación y, en consecuencia, la organización de los datos en los ficheros es totalmente dependiente de la aplicación que los trata.

De esta forma, si es necesario cambiar la estructura de alguno de los ficheros, será también necesario modificar la aplicación que los usa. Por otro lado, si es necesario cambiar una aplicación, casi con total seguridad será necesario modificar el número de ficheros que utiliza, su organización, los campos, etc.

Por otro lado, es casi seguro que existirá una alta redundancia de los datos, es decir, que existirán datos repetidos en diversos ficheros, por ejemplo, podrían estar los nombres de los clientes repetidos en varios ficheros. Esto implica que se está ocupando memoria de manera innecesaria, pero tiene otro inconveniente mucho más importante, que es la posibilidad de que se generen inconsistencias por el siguiente motivo: si un dato está repetido en varios ficheros, una modificación del mismo se debe llevar a cabo sobre varios ficheros, de manera que si no se hace así, pueden generarse inconsistencias, es decir, puede ocurrir que el mismo dato tenga diferentes valores en distintos ficheros. Supongamos el caso de que la dirección de un cliente esté almacenada en varios ficheros. Pues bien, si un cliente modifica su dirección, habrá que actualizarla en varios ficheros y si olvidamos actualizarla en uno de ellos, llegará el momento en que no sepamos cuál es su dirección real: la almacenada en un fichero o la guardada en el otro.

Esta manera de trabajar recibe el nombre de sistema orientado al proceso porque la manera en que están almacenados y organizados los datos es totalmente dependiente del proceso que se lleva a cabo con ellos.

Las desventajas de los sistemas tradicionales que trabajaban con ficheros desembocaron en la aparición del concepto de base de datos.

Una base de datos se puede definir como (Piattini et al, 2006): Una colección o depósito de datos integrados con redundancia controlada y con una estructura que refleje las interrelaciones y restricciones existentes en el mundo real. Los datos, que han de ser compartidos por diferentes usuarios y aplicaciones, deben mantenerse independientes de estas, y su definición y descripción únicas para cada tipo de dato, han de estar almacenadas junto con los mismos. Los procedimientos de actualización y recuperación, comunes y bien determinados, habrán de ser capaces de conservar la seguridad (integridad, confidencialidad y disponibilidad) del conjunto de los datos.

## 1.2. Ventajas e inconvenientes de las bases de datos.

La sustitución de un conjunto de ficheros por una base de datos proporciona las siguientes ventajas:

- Independencia de los datos respecto a los tratamientos y viceversa: esto quiere decir que un cambio en los tratamientos o programas no va a conllevar una modificación en la base de datos, un nuevo diseño de la misma. Por otra parte, la realización de modificaciones sobre la base de datos, como inclusiones o modificaciones de informaciones o cambios en la manera en que se almacenen los datos, no va a implicar casi nunca una modificación en los programas que acceden a los datos de la base de datos.
- Consistencia de los datos: eliminando o controlando las redundancias de datos se reduce el riesgo de que haya inconsistencias. Si la redundancia es mínima y está controlada por el sistema, como ocurre en las bases de datos, el propio sistema se encargará de garantizar que las copias se mantienen consistentes. A diferencia de lo que ocurre con los sistemas tradicionales de ficheros, en una base de datos no hay datos redundantes. Más bien, en una base de datos son solo redundantes o están repetidos únicamente aquellos datos que es imprescindible que lo estén para que en la base de datos se puedan reflejar las relaciones existentes entre los datos. Esto quiere decir que, por ejemplo, en una base de datos nunca aparecerá repetida la dirección de un cliente, pero sí puede aparecer su NIF repetido en varias tablas para reflejar, por ejemplo, que ese cliente ha realizado un pedido.
- Compartición de datos: en los sistemas de ficheros, estos pertenecen a las personas o departamentos que hacen uso de ellos. Cuando se trabaja con una base de datos, esta pertenece a la empresa y puede ser compartida por todos los usuarios que tienen autorización para ello. Para ello, cada usuario deberá conocer su nombre de usuario y

contraseña y en función del perfil del usuario podrá realizar determinadas operaciones sobre la base de datos. Será el administrador de la base de datos (DBA) el encargado de asignar a cada usuario un nombre de usuario y contraseña. Además, también será labor del DBA especificar los privilegios que un usuario tiene sobre los objetos de la base de datos: puede ocurrir que solo pueda consultar ciertas tablas, que otras tablas las pueda consultar y modificar, etc. Para facilitar esta tarea se suelen definir roles, los cuales agrupan un conjunto de privilegios sobre los objetos de la base de datos, de manera que si a un usuario se le asigna un rol, adquiere todos los privilegios del rol en cuestión.

- Mayor valor informativo: en la base de datos se almacenan los datos junto con las interrelaciones existentes entre ellos, por lo que el valor informativo del conjunto (de la base de datos) es mayor que la suma del valor informativo de los elementos individuales.
- Mejora en la accesibilidad a los datos: los sistemas gestores de bases de datos (en adelante, SGBD) incluyen lenguajes, como el SQL, que permiten a los usuarios con pocos conocimientos informáticos realizar consultas sobre los datos sin necesidad de escribir programas para tal fin. Analistas y programadores pueden generar programas para que los usuarios que desconozcan el lenguaje SQL puedan acceder a la base de datos y realizar operaciones sobre ella.
- Mejora en la integridad de los datos: la integridad de la base de datos, que se refiere a la consistencia y validez de los datos almacenados, se expresa mediante determinadas condiciones o restricciones que se deben cumplir, que suelen hacer referencia a los valores permitidos para un atributo. Pues bien, el subsistema de integridad del SGBD se encarga de asegurar que estas condiciones se cumplan. Esto se debe a que en una base de datos no solo se almacenan los datos propiamente dichos, sino también la semántica de los mismos.
- Control de la concurrencia: en algunos sistemas de ficheros, si hay varios usuarios que acceden simultáneamente a un mismo fichero, es posible que se pierda información o que afecte a la integridad de los datos. Sin embargo, los SGBD gestionan el acceso concurrente a la base de datos y permiten que no ocurra este tipo de problemas.
- Reducción del espacio de almacenamiento: la desaparición o disminución de la redundancia en las bases de datos conlleva una ocupación menor de memoria secundaria.

El trabajo con bases de datos se ha generalizado desde hace ya varias décadas debido a sus ventajas frente al trabajo con ficheros. No obstante, los sistemas de bases de datos también presentan algunos inconvenientes que se van a indicar a continuación:

- Instalación costosa: la implantación de un sistema de bases de datos puede implicar elevados costes en adquisición de hardware y software, entre los cuales es de destacar el coste de adquisición y mantenimiento del sistema gestor de la base de datos (SGBD).
- Personal especializado: va a ser necesaria la contratación de personal especializado que se encargue del diseño y administración de la base de datos.
- Falta de rentabilidad a corto plazo: la implantación de un sistema de base de datos, por los costes que conlleva tanto en personal como en equipos y por el tiempo que tarda en estar operativo, no resulta rentable a corto plazo, aunque sí lo sea a medio y largo plazo.

### **1.3. El sistema de gestión de la base de datos (SGBD/DBMS).**

Un sistema de gestión de bases de datos (SGBD), en inglés *Database Management System* (DBMS), es una colección de programas que facilitan la labor de gestionar la base de datos en su conjunto. En general, como indican Oltra et al. (2006), “el SGBD se encargará de gestionar el correcto funcionamiento interno de la base de datos, en lo que se refiere al control de la concurrencia y de la integridad, además de facilitar a los usuarios la creación, el mantenimiento y, en ocasiones, el diseño de dicha base de datos”.

### **1.4. Anomalías del acceso concurrente.**

Las bases de datos se emplean fundamentalmente en entornos multiusuario, lo que quiere decir que lo habitual es que una base de datos sea utilizada por varios usuarios, incluso simultáneamente. El uso concurrente o simultáneo de una base de datos por múltiples usuarios origina problemas de concurrencia. Así, mientras un usuario está modificando un registro de una tabla, no se debe permitir que otros usuarios realicen ninguna operación con dicho registro. Para llevar a cabo esto, existe un componente en el sistema gestor de la base de datos (SGBD) que se encarga de controlar que los accesos concurrentes sobre la base de datos no lleven a estados inconsistentes. Para ello, existen diferentes esquemas de control de concurrencia, como el basado en bloqueos.

Siguiendo a Silberschatz, Korth y Sudarshan (2002), un elemento de datos se puede bloquear en dos modos:

- En modo compartido: si una transacción bloquea un elemento de datos en modo compartido, solo lo puede leer, es decir, no lo puede escribir.
- En modo exclusivo: si una transacción obtiene un bloqueo sobre un elemento de datos en modo exclusivo, puede leerlo y escribirlo.

Cuando una transacción desea realizar una determinada operación sobre un elemento de datos, debe pedir un bloqueo de uno de estos dos tipos según la operación que se desee realizar. Esta petición es enviada al llamado gestor de bloqueos, de manera que la transacción solo podrá realizar la operación después de que le sea concedido el bloqueo solicitado.

Hay que tener en cuenta, como indican Silberschatz, Korth y Sudarshan (2002), que el modo compartido es compatible con otro modo compartido, pero no con el modo exclusivo. Esto quiere decir que puede haber varios bloqueos compartidos simultáneos pedidos por diversas transacciones sobre un mismo elemento de datos. Si sobre ese mismo elemento otra transacción solicita un bloqueo exclusivo, deberá esperar con anterioridad a que se liberen todos los bloqueos en modo compartido.

## **1.5. Administración de los datos y administración de bases de datos.**

Una base de datos se emplea para almacenar la información que se maneja en una empresa, en una institución, etc. Por este motivo es necesario que exista una persona en la organización responsable de los datos que se manejan en la misma. Esta persona debe entender y conocer los datos que se manejan en la organización y las necesidades de la organización en relación con dichos datos. Es responsabilidad de este individuo, al que nos podemos referir como administrador de los datos, decidir cuáles son los datos que se deben almacenar en la base de datos y establecer políticas para mantener y manejar estos datos de la mejor manera posible.

Por otro lado, debe existir otra figura en la empresa, que es el administrador de la base de datos, que se va a encargar de implementar informáticamente las decisiones tomadas por el administrador de datos. El administrador de la base de datos (DBA) debe ser, por tanto, un individuo con los conocimientos informáticos suficientes como para crear la base de datos en un determinado SGBD e implementar los controles necesarios para hacer cumplir las políticas establecidas por el administrador de datos. El administrador de la base de datos también tendrá otras responsabilidades, a las que se hará referencia en la sección 1.11.