BASES DE DATOS UD-1

ALMACENAMIENTO DE LA INFORMACIÓN (I)

FICHEROS

ÍNDICE

[NECESIDAD DE ALMACENAR INFORMACIÓN 3](#_Toc116110412)

[FICHEROS 3](#_Toc116110413)

[A. EL PASADO DE LOS ARCHIVOS O FICHEROS 3](#_Toc116110414)

[B. ARCHIVO O FICHERO DIGITAL 4](#_Toc116110415)

[C. ORGANIZACIÓN DE FICHEROS 4](#_Toc116110416)

[D. LOS SOPORTES DE LA INFORMACIÓN 5](#_Toc116110417)

[E. EVOLUCIÓN DE LA GESTIÓN DE LOS DATOS 6](#_Toc116110418)

[F. PROBLEMÁTICA DE LA GESTIÓN DE FICHEROS 6](#_Toc116110419)

[G. USO DE BASES DE DATOS 10](#_Toc116110420)

# NECESIDAD DE ALMACENAR INFORMACIÓN

Una base de datos es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso.

Desde siempre ha habido una necesidad de plasmar la información en algún medio, así tenemos como ejemplos:

* La escritura cuneiforme sobre piedra o arcilla.
* El uso de pergaminos o papiros.
* El uso de tablillas de madera.
* Etc.

De la misma forma siempre ha habido necesidad de almacenar esa información, por ejemplo:

* La Biblioteca de Alejandría.
* Un archivo de inventario de una biblioteca.
* Una base de datos musical.
* Un archivo de papel.
* Una cinta magnética, un disco magnético, un disco SSD…

La forma de almacenar la información difiere en:

1. El soporte utilizado: barro, madera, papel, cinta…
2. El sistema de codificación de la información: símbolos, agujeros, bits…
3. La forma de acceso a la información.

Actualmente cualquier aspecto de nuestra sociedad y de nuestra vida está relación con los datos, su almacenamiento y su gestión. Es debido a esto que la utilización de bases de datos está ampliamente extendida.

# FICHEROS

## EL PASADO DE LOS ARCHIVOS O FICHEROS

A principios y mediados del siglo XX, las necesidades de almacenamiento y gestión de la información podían satisfacerse utilizando un número relativamente reducido de **archivos en papel agrupados y ordenados**, estos archivos a su vez estaban compuestos de **fichas**.

Al llevar a cabo una primera informatización, se pasó de tener esos datos en papel a tenerlos en el ordenador y poder acceder más fácilmente a ellos. La informática adaptó sus herramientas para que los elementos que el usuario maneja en el ordenador se parezcan a los que utilizaba manualmente.

El elemento que permitió llevar a cabo el almacenamiento de datos de forma permanente en dispositivos de memoria masiva fue el **fichero o archivo.**

## ARCHIVO O FICHERO DIGITAL

Es la estructura de información creada por el SO para poder almacenar datos de forma permanente, para que puedan ser accedidos.

* + Son **dinámicos**: pueden variar durante la ejecución de un proceso.
  + Su capacidad **depende del soporte físico**, no de la memoria.
  + **Independiza los datos de los procesos**.

## ORGANIZACIÓN DE FICHEROS

* + **Sistema de archivos**: Modo de organización lógica de la información de un ordenador.
  + **Directorio**: Tipo de archivo especial que se emplea como contenedor de otros archivos, almacenando información sobre ellos. Es una forma de organizar ficheros. Actualmente se emplea una estructura en árbol o jerárquica, en la que los directorios pueden contener archivos y otros directorios.
  + **Ruta**: Forma de referenciar a un fichero dentro de una estructura de directorios de un sistema de archivos. Dos tipos:
    - **Absoluta**: Toma como origen el directorio principal, denominado raíz, del que cuelga el resto.

C:\carpeta1\carpeta2\archivo1.doc

* + - **Relativa**: Toma como origen el directorio actual.

carpeta2\archivo1.doc

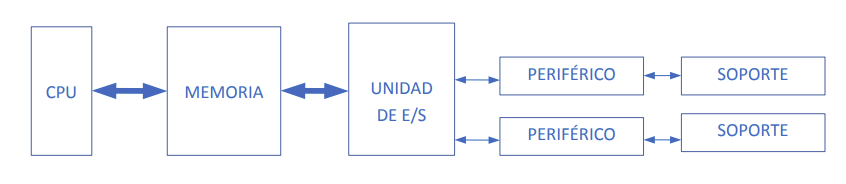
## LOS SOPORTES DE LA INFORMACIÓN

Los ficheros se almacenan en soportes físicos manejados por dispositivos hardware del ordenador que permiten leer y grabar datos.

Dos tipos:

* + **Soportes de acceso secuencial**: Requieren leer todo el contenido para acceder a un dato concreto. Son las cintas magnéticas. Actualmente se usan para las copias de seguridad debido a que el tiempo de acceso en ese caso no es tan importante.
  + **Soportes de acceso aleatorio**: Permiten acceder a cualquier dato almacenado sin tener que recorrer el soporte desde el principio. Requieren de un mecanismo que reconozca el punto físico del soporte al que se quiere acceder. Son los discos duros, los DVDs, los CDs, los pendrives, las tarjetas SD, las memory stick, los discos duros portátiles, los disquetes…



Para que el ordenador pueda procesar el contenido de un fichero es necesario que se copie su contenido del soporte físico a la RAM para que la CPU pueda acceder a los datos. Se sigue el siguiente esquema:

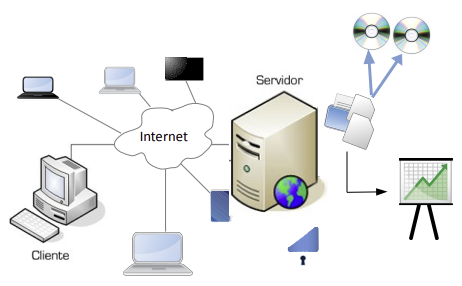
Un fichero se divide en **registros** y estos se agrupan en **bloques** por lo que no es necesario que se copie todo el contenido del fichero sino sólo algunos bloques o una estructura de índice.

## EVOLUCIÓN DE LA GESTIÓN DE LOS DATOS

1. **Sistemas de** **información no informatizados** 🡪 Fichas en papel.
2. **Primeros sistemas de información informatizados** 🡪 Ficheros informáticos
3. **Sistemas de información informatizados** 🡪 Bases de datos.

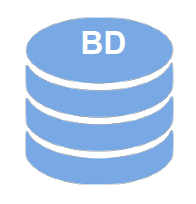
## PROBLEMÁTICA DE LA GESTIÓN DE FICHEROS

Los ficheros permiten organizar y guardar conjuntos de datos por lo que pueden servir para la gestión de información de aplicaciones.

Pero si las aplicaciones deben depender directamente de sus ficheros o archivos, **se pierde independencia entre el programa y los datos** y surgen inconvenientes: información duplicada, incoherencia de datos, fallos de seguridad, etc.

Suponiendo un ejemplo práctico de un fichero de puntuaciones de un juego nos vamos a encontrar una serie de problemas.

* + Si el juego es online 🡪 El fichero deberá estar alojado en el servidor del juego.
  + No sólo se va a almacenar las máximas puntuaciones, sino también valores estadísticos (media de jugadores, cuántas veces juega un jugador, etc.) 🡪 El fichero debe contener suficiente información y, además, debe actualizar cada vez que termina una partida.
  + El fichero no puede bloquearse, debe permitir la escritura simultanea de varios usuarios 🡪 Debe implementar mecanismos de bloqueo.
  + Las posibles actualizaciones de software del juego o del SO no deben afectar al almacenaje de los datos 🡪 Se requiere **independencia** entre el juego y la gestión de los datos.
  + Se debe poder cambiar la estructura de los datos sin tener que cambiar la programación del juego 🡪 Requiere **independencia** **lógica** de los datos.
  + Se debe poder cambiar el almacenaje físico de los datos de forma fácil y sencilla 🡪 Requiere **independencia** **física** de los datos.
  + Debe buscarse un sistema muy eficiente de acceso simultáneo a los datos (**concurrencia**). Mientras un usuario está escribiendo una puntuación otros pueden estar escribiendo otras.
  + El servidor donde está el juego debe ser potente y con gran capacidad de almacenaje de datos. Debe ser sencillo el poder ampliar la capacidad (**escalabilidad**).
  + No se quiere que nadie no autorizado acceda a los datos guardados: se requieren mecanismos de **seguridad**.
  + Habrá usuarios especiales que podrán gestionar los dados o acceder a información restringida 🡪 Implementar **perfiles** basados en **privilegios**.
  + No se quieren perder los datos (**disponibilidad**): **replicación** o **copias de seguridad**.

Es por esto que el uso de **gestores de bases de datos** son una mejor opción. Estos gestores se encargan de gestionar todos los aspectos anteriores de forma automática.

Permiten, por tanto, que **la aplicación y la gestión de la información sean independientes**. La aplicación usa la base de datos para almacenar su información.

Hay que destacar que la base de datos sigue usando ficheros como almacenamiento de la información, pero se encarga de su manejo por completo.

Por tanto, el empleo de ficheros para la gestión de datos conlleva varios problemas:

* Son **dependientes** del programa y del sistema operativo. Por ejemplo, al cambiar el SO del ordenador donde se alojaban los ficheros se corrompen los ficheros al no ser compatibles con el nuevo SO.
* **Problemas de concurrencia**. Por ejemplo:
* La **modificación** de la estructura de los datos puede ser **compleja** como el añadir o borrar nuevos campos. Además, requeriría cambiar las aplicaciones que lean esos ficheros. Por ejemplo, si queremos incorporar un descuento para algunos productos, pero sólo para algunos clientes tenemos varias opciones:
  + Solución con un único fichero: Se usa un único “fichero cliente-producto” y se repite cada producto para cada cliente, poniendo el precio personalizado.
    - Permite reflejar toda la información.
    - El mismo dato puede estar repetido muchas veces (duplicidad).
    - Difícil actualización de un dato, ya que puede estar en varios sitios.
  + Solución tes ficheros: Se usan tres ficheros (“clientes”, “producto” y “cliente-producto”).
    - Permite reflejar toda la información.
    - Disminuye la duplicidad.
    - Posible incoherencia ante borrados o modificaciones.
* La **modificación** de los datos puede ser **compleja**, especialmente en modificaciones o selecciones masivas. También puede resultar problemática cuando hay dependencia entre los datos (si se modifica un dato hay que tener en cuenta que hay que hacer modificaciones en otros datos relacionados). Ejemplos serían:

* + Por un error se quiere modificar el apellido de un cliente, para ello habría que acceder a todos los ficheros donde haya nombres de clientes y buscar al cliente y modificar su apellido.
  + Se quiere cambiar la descripción de un tipo de producto, pero sólo aquellos que tengan determinado número de referencia, para ello será necesario recorrer todo el fichero de productos y tener en cuenta la referencia del producto para decidir si se realiza el cambio o no.
* Puede haber **información duplicada** en distintos archivos (**ineficiencia** por la pérdida de espacio por esos datos duplicados e **incoherencia** porque en dos ficheros podría haber datos contradictorios). Ejemplos:
  + De ineficiencia: Si la empresa tiene miles de clientes y de productos, se requiere más espacio en disco y mucho más cuidado al modificar haciendo que los accesos se ralenticen mucho.
  + De incoherencia: A la hora de realizar un cambio de apellido de un cliente en un fichero, no tener en cuenta que ese mismo cliente puede aparecer en otros archivos.
* Las consultas pueden ser **ineficientes**. Por ejemplo, si se quiere conocer la media del precio de todos los productos habría que recorrer todo el fichero y hacer cálculos numéricos.
* Problemas de seguridad: **accesos no autorizados**. Por ejemplo, un cliente accede al fichero y se adjudica descuentos del 80% en todos los productos.

Para evitar esto, los ficheros deben estar protegidas con algún sistema de seguridad (**autenticación**, **roles**, **permisos** y **perfiles**). Además, es interesante llevar un registro de qué cambios se han hecho y quién los ha hecho (**logs**).

* **No se asegura la disponibilidad** de los datos. Por ejemplo, por un error un trabajador de la empresa borra el fichero “clientes” o parte de su contenido.

Para evitar que esto suceda, alguien debe realizar periódicamente copias de los ficheros (***backups*** o **copias de seguridad** y **métodos de restauración**).

## USO DE BASES DE DATOS

Hoy en día las bases de datos se utilizan en todos los ámbitos de la sociedad, entre otros:

* + Banca.
  + Tarjetas de crédito.
  + Justicia y seguridad.
  + Líneas aéreas.
  + Comercio y empresa.
  + Medicina.
  + Cultura.
  + Etc.