## **Bases de Datos**

# **Unidad de trabajo 1 Almacenamiento de la Información**



- Necesidad de almacenar información.
- El pasado de los archivos o ficheros
- Archivo o fichero digital
- Organización de ficheros
- Los soportes de la información
- Evolución de la gestión de los datos
- Problemática de la gestión de ficheros
- Uso de bases de datos

## 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.



Escritura cuneiforme sobre piedra o arcilla: registro administrativo de las actividades comerciales.



Base de datos de libros sobre la medicina tibetana antigua (provincia china de Qinghai)



Archivos persas que se ordenaban cronológicamente. Se empleaban pergaminos, papiros y tablillas de arcilla para los datos definitivos.



Tablillas de madera



Biblioteca de Alejandría



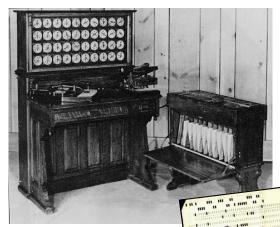
Archivo del inventario de una biblioteca



Base de datos musical



Archivo en papel



Maquina tabuladora o censadora, basada en tarjetas perforadas.



Cinta

magnética



magnético

Básicamente la forma de almacenar información difiere en:

- → El soporte que se ha utilizado (barro, madera, piel, papel, cinta...)
- → El sistema de codificación de la información (símbolos, escritura, agujeros, bits...)
- La forma de acceso a la información

Actualmente cualquier aspecto de nuestra sociedad y de nuestra vida cotidiana, casi todo lo que nos rodea, en alguna medida, está relacionado con los datos, su almacenamiento y su gestión.

Por ello la utilización de las bases de datos está ampliamente extendida.

El gran volumen de datos que actualmente manejamos y sus innumerables posibilidades requieren de la existencia de técnicos formados y capaces de trabajar con ellos.

¿Podríais enumerar situaciones reales en las que se utilizan bases de datos?

## El pasado de los archivos o ficheros

A principios y mediados del siglo pasado, los procesos básicos que se llevaban a cabo en una empresa se centraban en cuestiones relacionadas con contabilidad y facturación (clientes, proveedores, productos...).

Las necesidades de almacenamiento y gestión de información podían satisfacerse utilizando un número relativamente reducido de **archivos en papel agrupados y ordenados**. La mayor parte de las empresas utilizaban archivos de papel, donde guardaban datos de sus clientes en fichas escritas con máquinas de escribir o a mano.



 $Archivador \ \rightarrow archivos$ 

Fichero  $\rightarrow$  fichas

Memora Walls (Arthur Res)

Memora Walls (Arthur Res)

Memora Plat Soar

Friand Remank

United Associated Walls Wheelt

Modellis Wheelt

Modellis Wheelt

Modellis Carlshaves

Wallis Carlshaves

Wallis Carlshaves



Actualmente se usa de forma indistinta las palabras fichero y archivo

Al llevar a cabo una primera **informatización**, se pasó de tener los datos en formato papel a poder acceder a ellos de manera mucho más rápida a través del ordenador. En ese momento, la informática adaptó sus herramientas para que los elementos que el usuario maneja en el ordenador se parezcan a los que utilizaba manualmente.

Así en informática se sigue hablado de ficheros o archivos, formularios, carpetas, directorios,... El elemento que permitió llevar a cabo el almacenamiento de datos de forma permanente en dispositivos de memoria masiva fue el **fichero o archivo**.



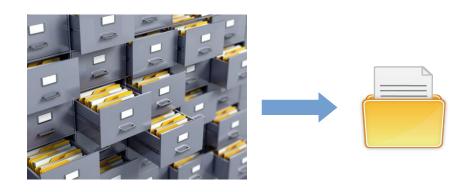


Observar el vídeo y responder:

- ¿En qué año se informatizó?
- ¿Cuál era su función principal?
- ¿Qué tecnología de almacenamiento empleaba?
- ¿Dónde están las pantallas?

## **Archivo o fichero digital**

Fichero o archivo digital: estructura de información creada por el sistema operativo para poder almacenar datos de forma permanente, de forma que puedan ser accedidos (se puedan añadir, suprimir, actualizar o consultar individualmente en cualquier momento).



#### Características:

- Dinámico: puede variar durante la ejecución de un proceso
- Su capacidad depende del soporte físico, no de la memoria
- Independiza los datos de los procesos

Los ficheros físicamente son una secuencia de ceros y unos (bits) almacenados en un soporte físico (disco duro, memoria extraíble, disco óptico...)





## Organización de ficheros

Sistema de archivos: modo de organización lógica de la información en 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.

Habitualmente 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.

- Absoluta: toma como origen el directorio principal, denominado raíz, del que cuelga el resto
- Relativa: toma como origen el directorio actual

```
acpi
asus-keyboard-backlight.sh
asus-wireless.sh
events
ibm-wireless.sh
undock.sh
adduser.conf
aliases
alsa
conf.d
alternatives
aliases.5.gz -> /usr/share/man/man5/aliases.sendmail.5.gz
animate -> /usr/bin/animate-im6.q16
animate.1.gz -> /usr/share/man/man1/animate-im6.q16.1.gz
```

## Los soportes de la información ¿dónde se guardan los datos?

Los ficheros se almacenan en soportes físicos manejados por dispositivos hardware del ordenador (periféricos), que permiten leer y grabar datos (dígitos binarios).

En primer lugar aparecen los soportes de acceso secuencial: es necesario leer todo el contenido, desde el principio, para acceder a un dato concreto. Esto conlleva un retardo considerable.



Cinta magnética

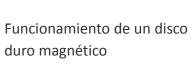
Actualmente se sigue desarrollando esta tecnología para copias de seguridad, en las que el tiempo de acceso no es tan crítico



The latest achievement has the potential to store 330 terabytes of uncc on a single on of your hand.

IBM Achieves the World's Highest Areal Recording Density for Magnetic Tape Storage (2:12)

Por otro lado existen los soportes de acceso aleatorio: puede accederse a cualquier dato almacenado sin tener que recorrer el soporte desde el principio. Es necesario que se implemente algún mecanismo para conocer a qué punto físico del soporte se desea acceder en cada momento.







Para que el ordenador pueda procesar el contenido de un fichero, es necesario que se copie su contenido del soporte físico a la memoria interna (RAM), para que la CPU pueda acceder a los datos.

Un fichero se divide en registros y éstos se agrupan en bloques. No es necesario que se copie todo el contenido del fichero a la memoria sino algunos bloques o una estructura de índice.



## Evolución de la gestión de los datos

Ejemplo: información sobre los alumnos matriculados en un instituto. Para cada alumno se almacenan sus datos personales, las asignaturas que cursa y las notas que ha obtenido en ellas.

Sistemas de información no informatizados "Lápiz y papel"



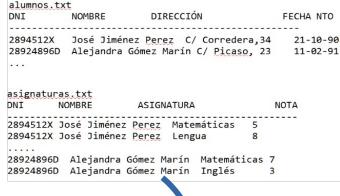
Primeros sistemas de información informatizados "Ficheros"





Fichas en papel

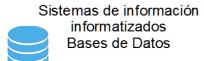
Ficheros informáticos



#### Base de datos

2894512X José Jiménez Pérez C/ Corredera, 34 21/10/90 002 Lengua 28924896D Alejandra Gómez Marín C/ Picasso, 23 11/02/91 003 Inglés	∯ DNI	∯ NOMBRE	∯ DIRECCION	⊕ FECHA_NACIM	001	Mateméticas
28924896D Alejandra Gómez Marín C/ Picasso, 23 11/02/91 003 Inglés	V	V	•	-		Lengua
	28924896D	Alejandra Gómez Marín	C/ Picasso, 23	11/02/91	003	Inglés

(} DNI		⊕ NOTA	
2894512x	001	5	
2894512X	002	8	
28924896D	001	7	



CODIGO 
 NOMBRE

## Problemática de la gestión de ficheros

Los ficheros permiten organizar y guardar conjuntos de datos con una determinada estructura, por o que pueden servir para la gestión de información de aplicaciones.

Pero si las aplicaciones, al ser diseñadas, 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.

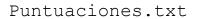
#### Ejemplo practico: puntuaciones de un juego

Supongamos que se quiere almacenar de forma permanente las puntuaciones de un juego, de forma que sea posible conocer las clasificaciones de los mejores jugadores.



¿Cómo podría hacerse?

Posible solución: a través de un fichero o archivo digital que almacene los datos.



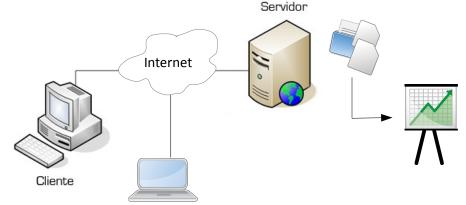


¿Qué ocurre si se quieren añadir las siguientes características?

• El juego es online (no está en nuestro dispositivo) y pueden jugar otros usuarios a través de internet.

#### Debe estar alojado en el servidor del juego

 No sólo se quieren conocer las máximas puntuaciones, también valores estadísticos: la media de jugadores que juegan cada día, cuántas veces juega un jugador, cómo un jugador mejora sus puntuaciones con cada nueva partida...



El fichero debe contener suficiente información. Cada vez que se termine una partida debe guardase el jugador, puntuación, hora y día de inicio y de fin de la partida, etc.

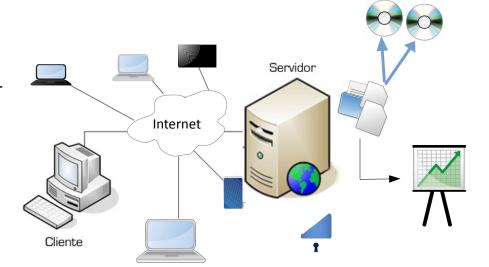
El fichero no debe bloquearse. Debe ser compartido y debe permitir escritura simultánea (concurrente) de varios usuarios

Debe implementar mecanismos de bloqueo.

Se quiere implementar un modelo profesional que reporte beneficios. ¿Qué ocurre si el juego se pone de moda y aumenta considerablemente el número de jugadores y partidas?



- Las posibles actualizaciones de software del juego o del sistema operativo 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 fácilmente
  - 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 puede 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 (usuario y contraseña)
- Habrá usuarios especiales que podrán gestionar los datos o acceder a información restringida. Se implementan perfiles basados en privilegios.
- No se quieren perder los datos (diponibilidad): replicación o copias de seguridad.



#### Conclusión:

Un gestor de base de datos se encarga de gestionar todos los aspectos anteriores de forma automática (concurrencia, disponibilidad, escalabilidad, seguridad, perfiles, replicación...).

Si el juego utiliza una base de datos, no necesitará gestionar los datos en su programación, sino que se centrará en los aspectos funcionales del propio juego.

De esta forma se asegura que la aplicación y la gestión de la información son independientes. La aplicación "usa" la base de datos para almacenar su información.

Además, proporciona independencia lógica y física sobre los datos: puede modificarse de forma independiente, sin afectar al resto, el sistema de almacenaje físico, la estructura de los datos y la programación del juego.



Hay que destacar que, a nivel físico, la base de datos sigue almacenando la información en ficheros como los que se han visto anteriormente, pero se encarga de su manejo por completo.

Para un usuario de la base de datos, los detalles y complejidad de la gestión de los ficheros están ocultos.

En resumen, el empleo de ficheros para la gestión de datos conlleva algunos problemas:

- Los ficheros son dependientes del programa y sistema operativo
- Problemas de concurrencia
- La modificación de la estructura de los datos puede ser compleja
- La modificación de los datos puede ser compleja
- Puede haber información duplicada en distintos archivos: ineficiencia
- Puede haber información duplicada en distintos archivos: incoherencia
- Las consultas de datos pueden ser ineficientes
- Problemas de seguridad: accesos no autorizados
- No se asegura la disponibilidad de los datos

Vamos a ver estos problemas con otro ejemplo

#### Ejemplo práctico: clientes y productos

Una empresa tiene un fichero con los datos de sus clientes, y otro con los datos de sus productos:







#### FICHERO CLIENTES

Nombre: Julián Pérez DNI: 45863217L

Teléfono: 610588963 Dirección: Calle Mayor 32

Nombre: Andrea Cosido

DNI: 45234523M Teléfono: 678500145 Dirección: Calle Cuesta 4

### FICHERO PRODUCTO-PRECIO

Referencia: 314

Descripción: cemento blanco

Precio: 33.25

Referencia: 502

Descripción: lámina roble

Precio: 12.5

Referencia: 503

Descripción: lámina pino

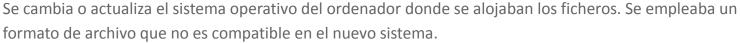
Precio: 12.5

Referencia: 021

Descripción: polímetro básico

Precio: 11

■ Problema: los ficheros son dependientes del programa y sistema operativo.





- Problema: concurrencia (acceso simultáneo a los datos)
  - 1. Julia abre el fichero de clientes. Esto es lo que ve:

#### **JULIA**

- 2. Julia comienza a escribir en el fichero un nuevo cliente (Miguel Cuervo).
  - 5. Julia termina de escribir y pulsa el botón "grabar".

#### FICHERO CLIENTES

Nombre: Julián Pérez DNI: 45863217L Teléfono: 610588963 Dirección: Calle Mayor 32

Nombre: Andrea Cosido DNI: 45234523M Teléfono: 678500145 Dirección: Calle Cuesta 4

Nombre: Miguel Cuervo DNI: 85265985L Teléfono: 610587552 Dirección: Ronda Alta 11 3. Elena abre el fichero Esto es lo que ve:

#### FICHERO CLIENTES

Nombre: Julián Pérez DNI: 45863217L Teléfono: 610588963 Dirección: Calle Mayor 32

Nombre: Andrea Cosido DNI: 45234523M Teléfono: 678500145 Dirección: Calle Cuesta 4

Nombre: Ignacio Martín DNI: 47892358H Teléfono: 658933680 Dirección: Calle Ronda 5

#### **ELENA**



- Elena comienza a escribir en el fichero un nuevo cliente (Ignacio Martín).
- 6. Elena termina de escribir y pulsa el botón "grabar".

Resultado: se ha perdido el cliente "Miguel Cuervo".

Problema: la modificación de la estructura de los datos puede ser compleja como al añadir o borrar nuevos campos.
 Además, requeriría cambiar las aplicaciones que lean esos ficheros.

Se quiere incorporar un descuento para algunos productos, sólo en algunos clientes.

Por ejemplo, se quiere hacer un descuento del 50% a Andrea Cosido en todas las láminas, aunque podrían haber otros descuentos para otros clientes.

#### FICHERO CLIENTES

Nombre: Julián Pérez DNI: 45863217L Teléfono: 610588963 Dirección: Calle Mayor 32

Nombre: Andrea Cosido DNI: 45234523M Teléfono: 678500145 Dirección: Calle Cuesta 4

#### FICHERO PRODUCTO-PRECIO

Referencia: 314

Descripción: cemento blanco

Precio: 33.25

Referencia: 502

Descripción: lámina roble

Precio: 12.5

Referencia: 503

Descripción: lámina pino

Precio: 12.5

Referencia: 021

Descripción: polímetro básico

Precio: 11

Hay que tener en cuenta que el descuento no depende del producto (un mismo producto no tiene el mismo descuento para todos los clientes) ni del cliente (un cliente no tiene el mismo descuento en todas sus compras), sino que depende del producto y del cliente concretos.

¿Cómo lo harías?

#### Solución con un único fichero:

se repite cada producto para cada cliente, poniendo el precio personalizado

- Permite reflejar toda la información
- Duplicidad: el mismo dato puede estar repetido muchas veces
- Difícil actualización de un dato, ya que puede estar en varios sitios

#### FICHERO CLIENTE-PRODUCTO

Nombre: Julián Pérez DNI: 45863217L

Teléfono: 610588963 Dirección: Calle Mayor 32

Referencia: 314

Descripción: cemento blanco

Precio: 33.25

Nombre: Julián Pérez DNI: 45863217L Teléfono: 610588963 Dirección: Calle Mayor 32

Referencia: 502

Descripción: lámina roble

Precio: 12.5

Nombre: Julián Pérez DNI: 45863217L

Teléfono: 610588963 Dirección: Calle Mayor 32

Referencia: 503

Descripción: lámina pino

Precio: 12.5

Nombre: Julián Pérez

DNI: 45863217L

Teléfono: 610588963 Dirección: Calle Mayor 32

Referencia: 021

Descripción: polímetro básico

Precio: 11

Nombre: Andrea Cosido DNI: 45234523M

Teléfono: 678500145 Dirección: Calle Cuesta 4

Referencia: 314

Descripción: cemento blanco

Precio: 33.25

Nombre: Andrea Cosido

DNI: 45234523M Teléfono: 678500145 Dirección: Calle Cuesta 4

Referencia: 502

Descripción: lámina roble

Precio: 6 25

Nombre: Andrea Cosido

DNI: 45234523M Teléfono: 678500145 Dirección: Calle Cuesta 4

Referencia: 503

Descripción: lámina pino

Precio: 6.25

Nombre: Andrea Cosido

DNI: 45234523M Teléfono: 678500145 Dirección: Calle Cuesta 4

Referencia: 021

Descripción: polímetro básico

Precio: 11

#### FICHERO CLIENTES

Nombre: Julián Pérez DNI: 45863217L Teléfono: 610588963 Dirección: Calle Mayor 32

Nombre: Andrea Cosido

DNI: 45234523M Teléfono: 678500145 Dirección: Calle Cuesta 4

#### FICHERO PRODUCTO

Referencia: 314

Descripción: cemento blanco

Referencia: 502

Descripción: lámina roble

Referencia: 503

Descripción: lámina pino

Referencia: 021

Descripción: polímetro básico

#### FICHERO CLIENTE-PRODUCTO

Nombre: Julián Pérez

Referencia: 314 Precio: 33.25

Nombre: Julián Pérez

Referencia: 502 Precio: 12.5

Nombre: Julián Pérez

Referencia: 503 Precio: 12.5

Nombre: Julián Pérez

Referencia: 021

Precio: 11

Nombre: Andrea Cosido

Referencia: 314 Precio: 33.25

Nombre: Andrea Cosido

Referencia: 502 Precio: 6.25

Nombre: Andrea Cosido

Referencia: 503 Precio: 6.25

Nombre: Andrea Cosido

Referencia: 021 Precio: 11

#### Solución con tres ficheros

- Permite reflejar toda la información
- Disminuye la duplicidad
- Posible incoherencia ante borrados o modificaciones







■ Problema: la modificación o actualización de los datos puede ser compleja, especialmente en modificaciones o selecciones masivas.

También puede resultar problemática cuando hay dependencias entre los datos (si se modifica un dato hay que tener en cuenta que hay que hacer modificaciones en otros datos relacionados)

- Por un error se quiere modificar el apellido de Andrea Cosido a Andrea Costa. Para ello es necesario acceder a todos los ficheros donde haya nombre de clientes, buscar "Andrea Cosido" y sustituirlo por "Andrea Costa"
- Se desea cambiar la descripción de los productos sustituyendo la palabra "lámina" por "laminado", solo para los productos cuya referencia sea menor que 300. Es necesario recorrer el fichero de productos, teniendo en cuenta la referencia del producto para decidir si hacer o no el cambio.

■ Problema: puede haber información duplicada en varios archivos, lo que es muy ineficiente.

Si la empresa tiene miles de clientes y de productos, se requiere mucho más espacio en disco, y mucho más cuidado al modificar. Los accesos pueden ralentizarse mucho.

■ Problema: puede haber información duplicada en varios archivos, lo que puede provocar incoherencia.

Por ejemplo, puede modificarse el apellido de un cliente en el fichero clientes, pero olvidar modificarlo en el fichero cliente-producto.

La consecuencia es que, cuando se busque el precio de un producto para ese cliente, no se encontrará.

#### FICHERO CLIENTES

Nombre: Julián Pérez DNI: 45863217L Teléfono: 610588963 Dirección: Calle Mayor 32

Nombre: Andrea Costa DNI: 45234523M Teléfono: 678500145 Dirección: Calle Cuesta 4

#### FICHERO PRODUCTO

Referencia: 314

Descripción: cemento blanco

Referencia: 502

Descripción: lámina roble

Referencia: 503

Descripción: lámina pino

Referencia: 021

Descripción: polímetro básico

#### FICHERO CLIENTE-PRODUCTO

Nombre: Julián Pérez Referencia: 314 Precio: 33.25

Nombre: Julián Pérez Referencia: 502 Precio: 12.5

Nombre: Julián Pérez Referencia: 503

Precio: 12.5

Nombre: Julián Pérez Referencia: 021 Precio: 11

Nombre: Andrea Cosido Referencia: 314

Precio: 33.25

Nombre: Andrea Cosido Referencia: 502

- Problema: consultas de datos ineficientes.
  - Para buscar la media del precio de los productos habría que recorrer todo el fichero y hacer cálculos numéricos.
  - Para conocer los datos del cliente con más descuento habría que recorrer varios ficheros, e ir comparando los descuentos.
- Problema de seguridad: accesos no autorizados.

Un cliente, aprovechando el descuido de un empleado, accede al fichero y se adjudica descuentos del 80% en todos los productos.



Para evitarlo, los ficheros deben estar protegidos 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).

■ Problema: no se asegura la disponibilidad de los datos.

Por un error un trabajador de la empresa borra el fichero CLIENTES, o parte de su contenido.

Para evitarlo, tiene que haber alguien que periódicamente haga copas 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:

