

TEMA 1

ALMACENAMIENTO DE LA INFORMACIÓN

1.Introducción

¿Te has preguntado alguna vez dónde y de qué manera se almacenan y gestionan los datos que utilizamos diariamente?

Si pensamos en cualquier acción de nuestra vida cotidiana, o si analizamos la mayoría de los ámbitos de actividad, nos encontramos que la utilización de las bases de datos está ampliamente extendida. Éstas, y los datos contenidos en ellas, serán imprescindibles para llevar a cabo multitud de acciones.

¿Crees que no es para tanto? Piensa en las siguientes situaciones:

- Cuando seleccionamos nuestro canal favorito en la TDT.
- Al utilizar la agenda del móvil para realizar una llamada telefónica.
- Cuando operamos en el cajero automático.
- Al solicitar un certificado en un organismo público.
- Cuando acudimos a la consulta del médico.
- Al inscribirnos en un curso, plataforma OnLine, etc.
- Si utilizas un GPS.
- Cuando reservamos unas localidades para un evento deportivo o espectáculo.

- Si consumimos ocio digital.
- Cuando consultamos cualquier información en Internet.
(Bibliotecas, enciclopedias, museos, etc.)

Este módulo profesional se centra en el estudio de las Bases de Datos y su uso en el desarrollo de aplicaciones. En esta primera unidad comenzaremos conociendo los primeros sistemas basados en ficheros para el almacenamiento y gestión de la información

2. Ficheros de información

2.1. ¿Qué es un fichero?

En la **década de los setenta**, los procesos básicos que se llevaban a cabo en una empresa **se centraban** en cuestiones relacionadas con **contabilidad y facturación**. Las necesidades de almacenamiento y gestión de información podían satisfacerse utilizando un número relativamente reducido de archivos en papel agrupados y ordenados, los típicos ficheros clásicos.

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, formularios, carpetas, directorios,...

La **información debía ser trasladada desde el papel al formato digital** y por lo general, era **necesario almacenarla** para su **posterior recuperación, consulta y procesamiento**. De este modo, para llevar a cabo un **tratamiento eficiente** de ésta era **necesario establecer métodos adecuados para su almacenamiento**. El elemento que permitió llevar a cabo el almacenamiento de datos de forma permanente en dispositivos de memoria masiva **fue el fichero o archivo**.

Un fichero es un conjunto de información relacionada, tratada como un todo y organizada de forma estructurada. Es una secuencia de números binarios que organiza información relacionada a un mismo aspecto.

Los **ficheros** están **formados por registros lógicos** que **contienen datos relativos a un mismo elemento u objeto** (por ejemplo, los datos de usuarios de una plataforma educativa). A su vez, los **registros** están **divididos en campos** que **contienen** cada una de las **informaciones elementales que forman un registro** (por ejemplo, el nombre del usuario o su dirección de correo electrónico).

Hemos de resaltar que los **datos están almacenados** de tal forma que se **puedan añadir, suprimir, actualizar o consultar individualmente en cualquier momento**.

Como los ficheros suelen ser muy voluminosos, solo se pueden llevar a la memoria principal partes de ellos para poder procesarlos. La cantidad de información que es transferida entre el soporte en el que se almacena el fichero, y la memoria principal del ordenador, en una sola operación de lectura/grabación, recibe el nombre de registro físico o bloque.

Normalmente en cada operación de lectura/grabación se transfieren varios registros del fichero, es decir un bloque suele contener varios registros lógicos. Al número de registros que entran en un bloque se le conoce con el nombre de factor de blocaje, y a esta operación de agrupar varios registros en un bloque se le llama bloqueo de registros.

2.2. Tipos de ficheros

Según su contenido, se clasifican en:

- Texto plano

- Almacenan secuencias de caracteres correspondientes a una codificación determinada (ASCII, Unicode, ...). Son legibles mediante un software de edición de texto como el bloc de notas

Ejemplos: ficheros .txt, html

- Binarios

- Contienen información codificada en binario para su procesamiento por parte de aplicaciones. Su contenido es ilegible en un editor de texto.

Ejemplos: archivos ejecutables (.exe), documentos de aplicaciones (.pdf, .docx), ficheros de imagen, audio y video (.jpg, .gif, .mp3, .avi), archivos del sistema (.dll)

Según la función que vaya a desempeñar los ficheros, éstos pueden ser clasificados de varias maneras.

- Ficheros Maestros

- Ficheros con datos fundamentales para una organización (datos clientes)

- Ficheros de Movimientos

- Ficheros con variaciones (altas, bajas o modificaciones). Luego reemplazan a los maestros para incorporar las modificaciones.

- Ficheros Históricos

- Ficheros con datos que no son necesarios diariamente (Clientes dados de baja)

- **Ficheros temporales**
 - Se utilizan para almacenar información útil para una parte de la aplicación, no para toda ella. Son generados a partir de datos de ficheros permanentes. Tienen un corto periodo de existencia. (clientes que son de Alicante)

2. Soportes de información.

Los ficheros se almacenan en soportes de información manejados por dispositivos periféricos del ordenador, que permiten leer y grabar datos en el soporte. Históricamente tenemos los siguientes soportes de información:

- Soportes perforados

Las primeras tarjetas perforadas aparecen en Francia cuando Joseph-Marie Jackard (1753-1834), invento un telar controlado por medio de tarjetas perforadas. Este telar operaba de la siguiente manera: las tarjetas se perforan estratégicamente y se acomodan en cierta secuencia para indicar un diseño de tejido en particular.

En 1843 se sugirió la idea de que las tarjetas perforadas pudieran adaptarse de manera que propiciaran que el motor analítico de Babbage repitiera ciertas operaciones.

Las tarjetas perforadoras no son más que simples cartulinas en las que se disponen 12 filas y 80 columnas. La presencia o ausencia de perforación en los diversos puntos es lo que define la información almacenada en la ficha o tarjeta.

Para la memorización de la información en una ficha perforada, se utilizan unos dispositivos especiales llamados perforadores de tarjetas. Posteriormente, en su lectura se utiliza otro dispositivo, el lector de tarjetas, que mediante las células fotoeléctricas efectúa la lectura detectando presencia o ausencia de perforación.

Las fichas perforadas son soportes cuya reutilización no es factible. Una vez perforados unos datos, no se pueden perforar otros nuevos. Este procedimiento de memorización se utilizaba cuando la información era poco voluminosa.

El problema de este soporte de información era el gran volumen de tarjetas que eran necesarios sólo para realizar simples cálculos. Además era muy común que se produjeran fallos al no colocar estas en el orden correcto. Otro de los problemas que se tenía es que este soporte era sólo de lectura y para comprobar que el programa había sido cargado correctamente había que ejecutarlo, algo que llevaba mucho tiempo.

- Dispositivos magnéticos

Estos dispositivos utilizan propiedades magnéticas para almacenar información.

discos duros, discos de 3 1/2", cintas de audio o casetes.

Capacidad para almacenar grandes cantidades de información en espacios reducidos con el consiguiente bajo costo por byte almacenado.

Es memoria "no volátil", es decir, guarda la información aunque se retire el suministro de energía eléctrica.

El acceso a la información puede ser secuencial o directo según el soporte.

- Dispositivos ópticos

Los discos ópticos aparecieron a fines de la década de los 80', siendo utilizado como un medio de almacenamiento de información para la televisión. Su alta capacidad y su fácil transportabilidad, hicieron que este dispositivo se popularice y comience a comercializarse en 1988 y a utilizarse en las computadoras.

Un disco óptico es lo que se conoce como una unidad de almacenamiento digital en formato circular, capaz de almacenar archivos de manera segura a través de codificación realizada con láser.

Este tipo de discos se encargan de guardar todo tipo de archivos, tanto los datos que requieran los usuarios, como contenidos de audio y vídeo, para posteriormente ser reproducidos bajo distintas reglas de uso. Por ejemplo, la codificación permite almacenar y grabar vídeos en formato DVD o Blu-ray, o simplemente documentos para su posterior lectura en un ordenador tradicional.

Una de las principales características de los discos ópticos son la clara utilidad de la que disponen para utilizarse en forma de copias de seguridad. Para ello los usuarios sólo deben realizar un volcado de datos desde la memoria de su ordenador y obtendrán, guardado de forma segura en un disco óptico, una copia de los archivos que tuvieran en el equipo. Otra de las características de los discos ópticos es su precio, reducido y de compra sencilla para los usuarios. En los discos ópticos la información se guarda de una forma secuencial en una espiral que comienza en el centro del disco. Además de la capacidad, estos discos presentan ventajas como la fiabilidad, resistencia a los arañazos, la suciedad y a los efectos de los campos magnéticos

.Entre ejemplos de discos ópticos tenemos el cd, dvd.

- memoria flash (USB)

Las memorias USB son dispositivos para el almacenamiento de cualquier tipo de información digital que generalmente utilizan memorias flash, memorias que pueden borrarse y reescribirse.

Bien, como decíamos antes, son dispositivos para almacenar datos digitales como por ejemplo: todo tipo de documentos, programas, canciones, fotografías digitales, películas, etc. Y llevártelas a cualquier parte de forma sencilla ya que son transportables, podemos llevarlas en nuestros bolso lo que hace que estos dispositivos nos sean muy útiles,

casi imprescindibles a día de hoy.

Estas memorias, además, tienen la gran característica de ser muy resistentes (algunas incluso son resistentes al agua) y baratas dejándonos guardar información durante muchos años (teóricamente unos 20 años) y puedes eliminar y sustituir los datos almacenados hasta un millón de veces aproximadamente.

Las memorias USB se clasifican según dos características principales: Su capacidad de almacenamiento y su velocidad para transmitir datos.

Podemos encontrar memoria desde 16GB hasta 1TB

Según su velocidad de transmisión de datos se clasifican en:

- USB 1.0 de velocidad hasta 1,5 Mbps
- USB 1.1 de velocidad hasta 12 Mbps
- USB 2.0 de velocidad hasta 480 Mbps
- USB 3.0 de velocidad hasta 5 Gbps
- USB 3.1 de velocidad hasta 10 Gbps
- USB 3.2 de velocidad hasta 20Gbps

Sus principales ventajas son las siguientes:

1. Bajo coste y garantía para el servicio que ofrecen
2. Resistentes al polvo, ralladuras,
3. Son muy cómodas de transportar por su pequeño tamaño.
4. Duran hasta 20 años.
5. Pueden funcionar prácticamente en cualquier dispositivo.
6. Gran capacidad de almacenamiento de memoria con respecto a disquetes, cd's, dvd's...etc
7. Pueden sobrescribirse hasta un millón de veces.
8. Podemos almacenar todo tipo de datos: programas, archivos, documentos, fotografías digitales, canciones, películas..etc
9. Son muy fáciles de utilizar, simplemente conectándolas a un puerto USB ya podemos almacenar datos en ellas.

Sus principales desventajas son:

1. Pueden dejar de funcionar repentinamente por diferentes accidentes: variaciones de voltaje mientras están conectadas a una fuente de alimentación, por usarlas prolongadamente durante muchos años, por que se nos puedan caer (desde alturas superiores a 1 metro).

2. Por tener un tamaño tan pequeño podemos perderlas fácilmente.
3. Al ser tan fáciles de utilizar en cualquier ordenador son susceptibles de poder almacenar virus ya que no tienen ningún sistema propio de seguridad (no disponen de antivirus propio).
4. Privacidad de datos: si nuestras memorias USB caen en “malas manos” porque las podamos perder, o nos las puedan robar, cualquier puede tener acceso a los datos que hayamos almacenado en ellas.

4. Métodos de acceso a lo ficheros

A medida que la tecnología ha ido evolucionando, el acceso a la información contenida en los diferentes tipos de ficheros ha ido variando.

Los objetivos fundamentales de estas modificaciones pueden resumirse en los siguientes puntos:

- Proporcionar un acceso rápido a los registros.
- Conseguir economizar el almacenamiento.
- Facilitar la actualización de los registros.
- Permitir que la estructura refleje la organización real de la información.

Los principales métodos de acceso a los ficheros son los siguientes:

- **Ficheros Secuenciales**

- En estos ficheros, los registros se almacenan de forma continua en el orden en el que fueron grabados.
- La única forma de acceder a un registro, es leyendo un registro tras otro desde el principio hasta el final. En los ficheros secuenciales suele haber una marca indicativa del fin del fichero, que suele denominarse **EOF** (End of File). Para detectar el final del fichero sólo es necesario encontrar la marca EOF.

Otras características relevantes de los ficheros secuenciales son:

- La lectura siempre se realiza hacia delante.
- Tienen una estructura rígida de campos. Todos los registros deben aparecer en orden, es decir, la posición de los campos de cada registro siempre ha de ser la misma.
- El modo de apertura del fichero, condiciona la lectura o escritura.
- Aprovechan al máximo el soporte de almacenamiento, al no dejar huecos vacíos.
- Se pueden grabar en cualquier tipo de soporte, tanto en secuenciales como direccionales.

- Todos los lenguajes de programación disponen de instrucciones para trabajar con este tipo de ficheros.
- No se pueden insertar registros entre los que ya están grabados.

Ejemplo fichero secuencial:

```
00789521T#Paula#Sanz#González#619554687$50687452Y#José Luis#García#Viñals#
667859621$38546998X#Javier#Peinado#Martín#666932541$09653801B#Ruth#Lázaro#
Cardenal#689330247%
```

% sería marca fin de fichero
separa los datos
\$ separa registros

◦ **Ficheros de acceso directo**

En este tipo de ficheros se puede acceder a un registro en concreto sin necesidad de acceder a todos los anteriores, indicando la posición relativa del mismo dentro del archivo o, más comúnmente, a través de una clave que forma parte del registro como un campo más. Estos archivos deben almacenarse en dispositivos de memoria masiva de acceso directo, como son los discos magnéticos.

Cada uno de los registros se guarda en una posición física, que dependerá del espacio disponible en memoria masiva, de ahí que la distribución de los registros sea aleatoria dentro del soporte de almacenamiento. Para acceder a la posición física de un registro se utiliza una dirección o índice, no siendo necesario recorrer todo el fichero para encontrar un determinado registro.

A través de una transformación específica aplicada a la clave, se obtendrá la dirección física en la que se encuentra el registro.

En el acceso directo la clave coincide con la dirección, debiendo ser numérica y comprendida dentro del rango de valores de las direcciones. Es el método más rápido.

Otras características fundamentales de los ficheros de acceso directo son:

- Posicionamiento inmediato.
- Registros de longitud fija.
- Apertura del fichero en modo mixto, para lectura y escritura.
- Permiten la utilización de algoritmos de compactación de huecos.

- Los **archivos** se crean con un **tamaño definido**, es decir, con un **máximo de registros establecido durante la creación**.
- Esta organización **sólo es posible en soportes direccionales**.
- Se usan cuando el **acceso a los datos de un registro se hace siempre empleando la misma clave** y la **velocidad de acceso a un registro es lo que más nos importa**.
- Permiten la **actualización de los registros en el mismo fichero, sin necesidad de copiar el fichero**.
- Permiten realizar **procesos de actualización en tiempo real**.

Ejemplo de fichero de acceso directo

103	→	103	M-0568-CF	Citroen C4	Cambio de aceite
		104	0454-CNK	Seat Ibiza	Cambio de filtros
105	→	105	1563-BPM	Renault Megane	Sustituir correa

- **Ficheros de acceso indexado**

Aquí tenemos el fichero de datos que **contiene la información** y un **fichero de índices** que permiten el **acceso directo a los datos y contiene una lista de claves y la posición donde esta almacenada la información**.

La idea es **como el índice de un libro** donde aparecen las palabras que forma parte del índice (**claves**) y el número de página (**posición del registro**)

Por tanto, **existirá una zona de registros** en la que se encuentran **los datos del archivo** y una **zona de índices**, que contiene una tabla con las claves de los registros y las posiciones donde se encuentran los mismos. **La tabla de índices estará ordenada por el campo clave**.

La **tabla de índices será cargada en memoria principal** para realizar en ella la **búsqueda de la fila correspondiente a la clave del registro a encontrar**, **obteniéndose así la dirección donde se encuentra el registro**. Una vez localizada la dirección, sólo hay que **acceder a la zona de registros en el soporte de almacenamiento** y posicionarnos en la dirección indicada. Puesto que la tabla debe prever la inclusión de **todas las direcciones posibles del archivo**, su **principal**

inconveniente resulta determinar su tamaño y mantenerla ordenada por los valores de la clave.

Las características más relevantes de un fichero indexado, son las siguientes:

- El diseño del registro tiene que tener un campo, o combinación de campos, que permita identificar cada registro de forma única, es decir, que no pueda haber dos registros que tengan la misma información en él. A este campo se le llama campo clave y es el que va a servir de índice.
- Permiten utilizar el modo de acceso secuencial y el modo de acceso directo para leer la información guardada en sus registros.
- Para acceder a este tipo de ficheros utilizando el modo de acceso directo se hace conociendo el contenido del campo clave del registro que queremos localizar. Con esa información el sistema operativo puede consultar el índice y conocer la posición del registro dentro del fichero.
- Para acceder a este tipo de ficheros utilizando el modo de acceso secuencial los registros son leídos ordenados por el contenido del campo clave, independientemente del orden en que se fueron grabados (el orden lógico no es igual al orden físico), debido a que el acceso a los datos se hace a través del índice, que para hacer más fácil la búsqueda de los registros, permanece siempre ordenado por el campo clave.
- Solamente se puede grabar en un soporte direccionable. Por ejemplo, un disco magnético. Si esto no fuera así, no podría emplear el acceso directo.

Ejemplo de fichero de acceso indexado ordenado por número de habitantes

		PAÍS	CONTINENTE	HAB. (mill)
5	→	Noruega	Europa	5
5	→	Argentina	Amer. Sur	41
6	→	España	Europa	47
11	→	Italia	Europa	60
15	→	Dinamarca	Europa	5
21	→	Alemania	Europa	81
30	→	Camerún	África	21
41	→	Ecuador	Amer. Sur	15
47	→	Inglaterra	Europa	53
53	→	Nicaragua	Amer. Sur	6
60	→	México	Amer. Cent.	112
81	→	Perú	Amer. Sur	30
112	→	Cuba	Amer. Cent.	11