

UT1. Almacenamiento de la información

¿Dónde y cómo se almacenan y gestionan los datos que utilizamos diariamente?

Si pensamos en cualquier actividad de nuestra vida cotidiana veremos que el uso de las bases de datos está ampliamente extendido y que su uso y el de los datos contenidos en ellas son imprescindibles para llevar a cabo multitud de acciones:

- Ver los canales favoritos de la TDT
- Seguir viendo una serie en una plataforma
- Utilizar los contactos del móvil para hacer una llamada
- Al sacar dinero de un cajero automático
- Al registrarnos en una red social
- Al utilizar el GPS
- Cuando compramos entradas para un concierto, cine, teatro,...

Como vemos el mundo que nos rodea está relacionado con los datos, su almacenamiento y su gestión.

En este tema:

- conoceremos los primeros sistemas basados en ficheros para el almacenamiento y gestión de la información,
- luego veremos los conceptos y las definiciones básicas relacionadas con las bases de datos
- analizaremos sus modelos y tipos
- conoceremos las capacidades y características de los Sistemas Gestores de Bases de Datos (en adelante SGBD)
- identificaremos las herramientas con las que llevaremos a cabo la gestión de las Bases de Datos

1. Ficheros de información

1.1. ¿Qué es un fichero?

En los años 70, los procesos básicos que se llevaban a cabo en una empresa eran la facturación y la contabilidad. La necesidad de almacenar y gestionar la información se satisfacía utilizando un número relativamente reducido de archivos en papel agrupados y ordenados, los típicos ficheros clásicos.

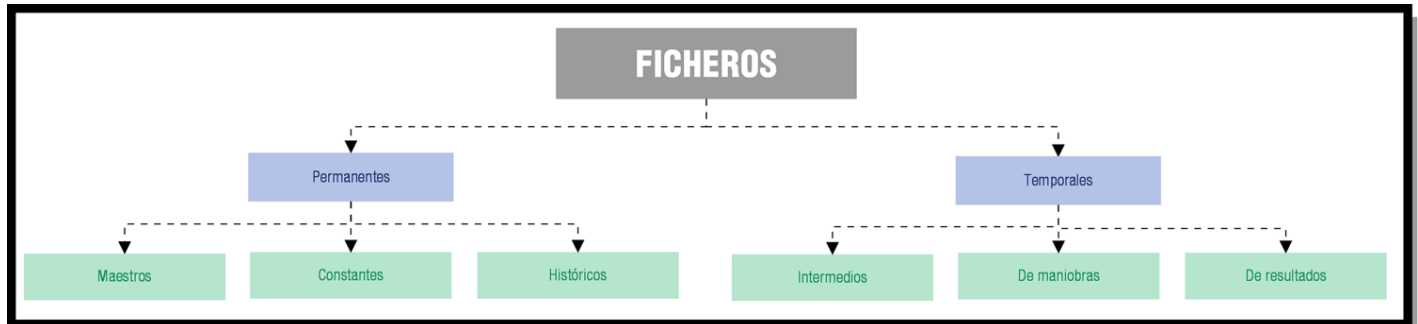
Al llevar a cabo la primera informatización se pasó de tener los datos en formato papel a poder acceder a ellos de manera mucho más rápida a través del ordenador. La informática adaptó sus herramientas para que los elementos que maneja el usuario en el ordenador se parezcan a los que utiliza de manera manual. Por eso en la informática se sigue hablando de archivos, ficheros, formularios, carpetas, directorios,...

La información que se trasladaba del papel al formato digital debía ser almacenada para posteriormente poder recuperarla, consultarla y procesarla. 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 o archivo** es un conjunto de información relacionada, tratada como un todo y organizada de forma estructurada. Es una secuencia de dígitos binarios que organiza información relacionada con un mismo aspecto.

1.2. Tipos de ficheros

Según la función que vayan a desempeñar los ficheros se pueden clasificar de varias formas, una podría ser la siguiente:



Ficheros permanentes contienen información relevante para la aplicación, es decir, los datos necesarios para que ésta funcione. Tienen un periodo de permanencia en el sistema amplio. Se dividen en:

- o **Ficheros maestros:** contienen los datos actuales que pueden modificarse desde la aplicación. Son el núcleo. Podría ser un archivo con los datos de los usuarios de una plataforma educativa
- o **Ficheros constantes:** son los que contienen datos fijos para la aplicación, no suelen ser modificados y se accede a ellos para realizar consultas. Podría ser un archivo con códigos postales
- o **Ficheros históricos:** contienen los datos que en algún momento fueron actuales. Se utilizan para la reconstrucción de situaciones. Podría ser un archivo con los datos de los usuarios que se han dado de baja de la plataforma educativa

Ficheros temporales se utilizan para almacenar información útil para una parte de la aplicación, pero no para toda ella. Son generados a partir de datos de ficheros permanentes. Tienen un corto periodo de existencia. Se dividen en:

- o **Ficheros intermedios:** almacenan resultados de una aplicación que serán utilizados por otra, dentro de una misma tarea.
- o **Ficheros de maniobras:** se utilizan para almacenar datos propios de un programa que no pueden ser mantenidos en la memoria principal, por falta de espacio. Normalmente se encuentran en programas de cálculo numérico, editores y compiladores. Su vida es inferior al tiempo de ejecución del programa.
- o **Ficheros de resultados:** almacenan los datos que van a ser transferidos a un dispositivo de salida, por ejemplo, un archivo de impresión, contiene los datos que van a ser transferidos a una impresora.

Ejemplo

Supongamos que queremos almacenar y procesar los datos relativos a la ocupación y cobros de un hotel. Para ello necesitaremos almacenar una serie de datos fijos como las características de las habitaciones, las tarifas, etc...

Esta información la almacenaremos en ficheros, que por su naturaleza serán **permanentes y constantes**. Los registros de estos ficheros pueden tener la siguiente estructura:

FicheroCaracteristicasHabitaciones:

- Número de habitación
- Número de camas
- Categoría
- Exterior/Interior

FicheroPrecios:

- Categoría de la habitación
- Número de camas
- Precio por noche en temporada alta
- Precio por noche en temporada baja

Por otro lado, tendremos que crear archivos que contendrán datos variables. Debe existir un fichero con la información actual sobre la ocupación del hotel, que será un fichero **permanente** y también el fichero **maestro** de la aplicación. Podría tener la siguiente estructura:

FicheroOcupacionHotel:

- Número de habitación
- Inservible (Si/No)
- Ocupada (Si/No)
- Número de personas
- DNI del huésped
- Nombre y apellidos del huésped
- Fecha de entrada
- Fecha prevista de salida

Cada vez que un cliente abandona el hotel se creará una factura que se cobrará al cliente y será un archivo de **resultados**. Este archivo se transferirá a la impresora y después será borrado. Por tanto será un archivo temporal. Los datos de ocupación de habitaciones del archivo **FicheroOcupacionHotel**, una vez el cliente abandone el hotel, pasarán a un archivo **histórico**.

1.3. Soportes de información

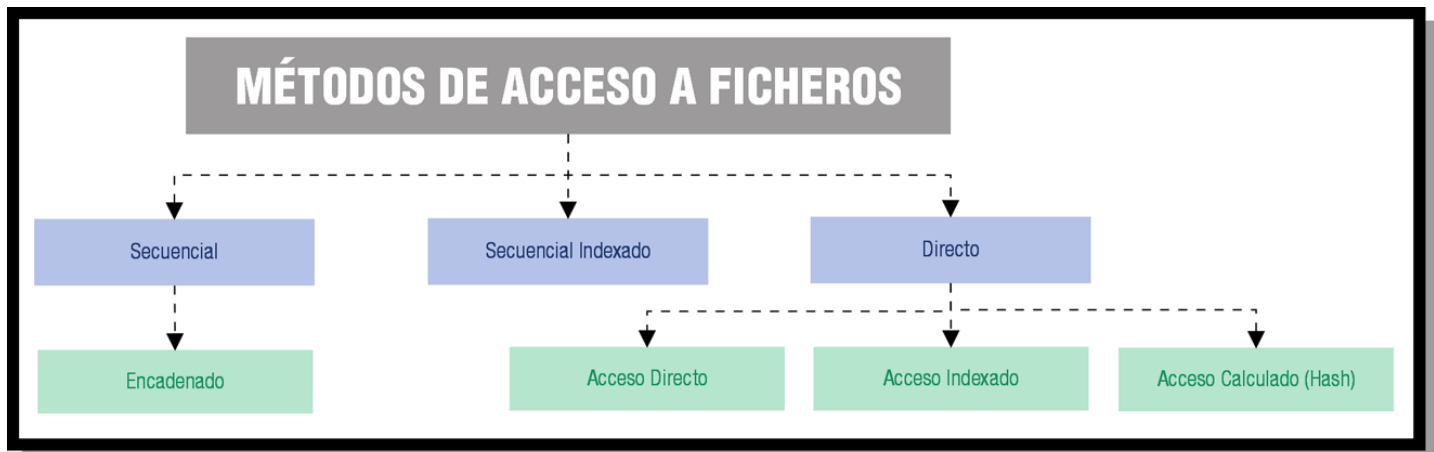
Los ficheros se almacenan en soportes de información manejados por dispositivos periféricos del ordenador, que permiten leer y grabar datos en el soporte. Según el modo de acceso a la información, podemos distinguir 2 tipos de soportes:

- **Soportes de acceso secuencial:** por ejemplo las cintas magnéticas, si queremos leer un dato que se encuentra en la mitad de la cinta, tendremos que leer todo hasta llegar a esa posición
- **Soportes de acceso directo:** por ejemplo, discos, el acceso a los datos puede hacerse de forma directa, pudiendo colocarnos en la posición que nos interesa y leer a partir de ella.

1.4. Métodos de acceso

A medida que la tecnología ha ido evolucionando, principalmente por los avances en hardware, el acceso a la información contenida en los distintos ficheros también ha cambiado. Los principales objetivos de estas modificaciones pueden resumirse en:

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



1.5. Ficheros secuenciales

La principal característica de un fichero con organización secuencial es que sus registros están almacenados de forma contigua, de manera, que la única forma de acceder a uno en concreto, 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.

Este tipo de ficheros pueden utilizar dispositivos de acceso secuencial, como son las cintas magnéticas de almacenamiento de datos. También se utiliza en los CD de audio y los DVD de vídeo, en los que la música o las imágenes se almacenan a lo largo de una espiral continua.

Los registros almacenados se identifican por medio de una información ubicada en uno de sus campos, a este campo se le denomina **clave o llave**. Si se ordena un archivo secuencial por su clave, es más rápido realizar cualquier operación de lectura o escritura.

Otras características relevantes de los ficheros secuenciales son:

- La lectura siempre se realiza hacia delante.
- Son ficheros monousuario, no permiten el acceso simultáneo de varios usuarios.
- 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 direccionables.
- 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



Imagen 1. Estructura de un fichero secuencial

1.6. Ficheros de acceso directo

En este tipo de ficheros se puede acceder a un registro 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.

Campo clave: campo que permite identificar y localizar un registro de manera ágil y organizada.

- 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.



- La clave nos indica la dirección
- La clave debe ser numérica
- La clave debe estar comprendida dentro del rango de valores de las direcciones
- Es el método más rápido

Las principales características de los ficheros de acceso directo son:

- Posicionamiento inmediato
- Registros de longitud fija
- Apertura de ficheros para lectura y escritura
- Permiten múltiples usuarios
- Los registros se borran colocando un cero en la posición que ocupan
- Permiten el uso de algoritmos para la compactación de huecos
- Los archivos se crean con un tamaño predefinido, es decir, con un máximo de registros
- Sólo es posible en soportes direccionables

- 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.

1.7. Ficheros indexados

Se basan en la utilización de **índices**, que permiten el acceso a un registro del fichero de forma directa, sin tener que leer los anteriores. Estos índices son similares a los de los libros:

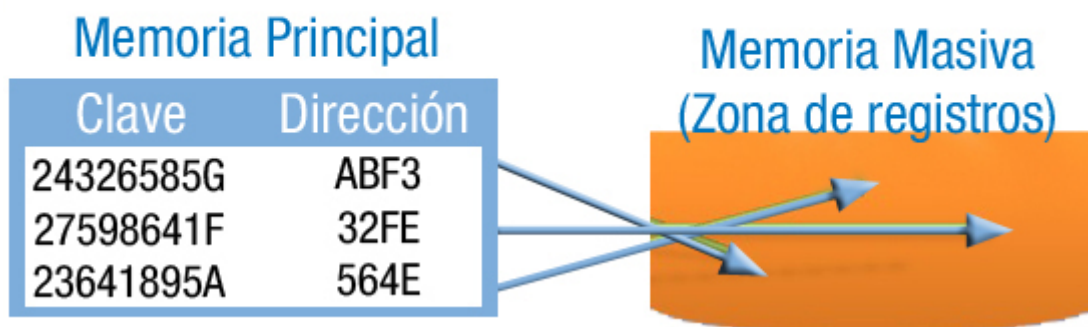
Si nos interesa leer un capítulo concreto de un libro, vamos al índice miramos en que página comienza, y abrimos el libro por esa página, sin tener que mirar en todas las páginas anteriores para localizarlo.

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.

FICHEROS INDEXADOS



Las principales características de un fichero indexado, son las siguientes:

- El diseño del registro tiene que tener un campo o una 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**.
Un mismo fichero puede tener más de un campo clave, pero al menos uno de ellos no admitirá valores duplicados y se le llama clave **primaria**. A las restantes se les llama claves **alternativas**.
- 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 grabando.

(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.

1.8. Parámetros de utilización

En función del uso que se vaya a dar al fichero, serán adecuados distintos tipos de organización. Utilizando **parámetros de referencia**, podremos determinar el uso de un fichero. Estos parámetros son:

- **Capacidad o volumen:** es el espacio, en caracteres, que ocupa el fichero. La capacidad podrá calcularse multiplicando el número previsto de registros por la longitud media de cada registro.
- **Actividad:** permite conocer la cantidad de consultas y modificaciones que se realizan en el fichero. Para poder especificar la actividad se deben tener en cuenta:
 - o **Tasa de consulta o modificación:** que es el porcentaje de registros consultados o modificados en cada tratamiento del fichero, respecto al número total de registros contenidos en él.
 - o **Frecuencia de consulta o modificación:** número de veces que se accede al fichero para hacer una consulta o modificación en un periodo de tiempo fijo.
- **Volatilidad:** mide la cantidad de inserciones y borrados que se efectúan en un fichero. Para determinar la volatilidad es necesario conocer:
 - o **Tasa de renovación:** es el tanto por ciento de registros renovados en cada tratamiento del fichero, respecto al número total de registros contenidos en él.
 - o **Frecuencia de renovación:** es el número de veces que se accede al fichero para renovarlo en un periodo de tiempo fijo.
- **Crecimiento:** es la variación de la capacidad del fichero y se mide con la tasa de crecimiento, que es el porcentaje de registros en que aumenta el fichero en cada tratamiento.