

Encontrado origen en :  
[https://sarreplec.caib.es/pluginfile.php/9796/mod\\_resource/content/2/BD01\\_Contenidos\\_Imprimible/index.html](https://sarreplec.caib.es/pluginfile.php/9796/mod_resource/content/2/BD01_Contenidos_Imprimible/index.html)

## 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. 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.)
- Al registrarte en una página de juegos OnLine, redes sociales o foros.
- Incluso, si tienes coche, puede ser que éste incorpore alguna base de datos.

Casi todo lo que nos rodea, en alguna medida, está relacionado con los datos, su almacenamiento y su gestión. El gran volumen de datos que actualmente manejamos y sus innumerables posibilidades requieren de la existencia de técnicos perfectamente formados y capaces de trabajar con ellos.

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. Seguidamente, se desarrollarán los conceptos y definiciones básicas relacionadas con las bases de datos, posteriormente analizaremos sus modelos y tipos, un poco más adelante, podremos conocer las características y capacidades de los sistemas gestores de bases de datos y finalmente, identificaremos las herramientas reales con las que llevar a cabo la gestión dichas bases.

## 2. Los ficheros de información

### 2.1. ¿Qué es un fichero? 1.A 2.1 Perspectiva fichero by GPTo1

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**.

**Fichero o archivo: 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.**

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

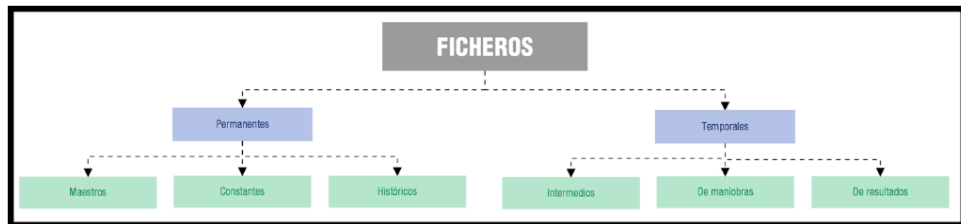
- Añadir,
- Suprimir,
- Actualizar o
- Consultar individualmente en cualquier momento.

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 bloqueaje**, y a esta operación de agrupar varios registros en un bloque se le llama **bloqueo de registros**.

#### 2.1.1 Tipos de ficheros

Según la función que vaya a desempeñar los ficheros, éstos pueden ser clasificados de varias maneras. En la siguiente imagen puedes observar una posible clasificación.



**Ficheros permanentes:** contienen información relevante para la aplicación. Son datos necesarios para el funcionamiento de ésta. Se dividen en:

- **Maestros.** Contienen el **estado actual** de los datos que pueden modificarse desde la aplicación.  
*Por ejemplo el archivo de datos de usuarios de una plataforma educativa.*
- **Constantes:** Incluyen datos fijos para la aplicación, no suelen ser modificados y se accede a ellos para consultar.  
*Por ejemplo un archivo con códigos postales.*
- **Históricos.** Contienen datos que fueron considerados como actuales en un periodo o momento anterior. Se utilizan para reconstruir situaciones.  
*Por ejemplo un fichero con los usuarios que se dieron de baja de una plataforma educativa.*

**Ficheros temporales:** almacenan información útil para una parte de la aplicación. Se generan a partir de datos de ficheros permanentes. Tienen un corto periodo de existencia, se dividen en:

- **Intermedios.** Almacenan resultados de una aplicación que *serán utilizados por otra.*
- **De maniobra.** Almacenan datos de una aplicación que *no pueden ser mantenidos en memoria principal* por falta de espacio.
- **De resultados.** Almacenan datos que serán *transferidos a un dispositivo de salida.*

## 2.2. Los 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. Los soportes más utilizados para almacenar los ficheros son las cintas magnéticas y los discos (magnéticos, ópticos, o magneto-ópticos). Dentro de estos dos tipos de soporte existen en el mercado una gran variedad de modelos.

Inicialmente, los primeros sistemas de almacenamiento físico eran tambores de cinta magnética. Tenían unas dimensiones parecidas a los discos de vinilo. Estos tambores funcionaban de manera similar a los antiguos casetes, pero sus mayores dimensiones les permitían almacenar gran cantidad de datos en formato digital, es decir en ceros y unos, en orden secuencial.

Posteriormente, los sistemas de almacenamiento de información comenzaron a cambiar de la mano de los avances en el hardware, en concreto con la aparición del disquete y del disco duro. Eran dispositivos de acceso aleatorio, no siendo necesario en ellos pasar por todos los datos desde el inicio hasta la zona donde se encuentra la información que nos interesa.

Por tanto, se distinguen dos tipos de soportes para el almacenamiento de datos:

- **Soportes de Acceso Directo a los datos** (Por ejemplo: discos). Son los más empleados y el acceso a los datos puede hacerse de forma directa, pudiendo colocarnos en la posición que nos interesa y leer a partir de ella.
- **Soportes de Acceso Secuencial** (Por ejemplo: cintas magnéticas). Se suelen usar en copias de seguridad y si deseamos leer un dato que está en la mitad de la cinta, tendremos que leer todo lo que hay hasta llegar a esa posición.

Para saber más:

Cintas magnéticas de almacenamiento de datos

Discos magnéticos

Unidades de estado sólido

Discos ópticos.

Disco magneto-ópticos.

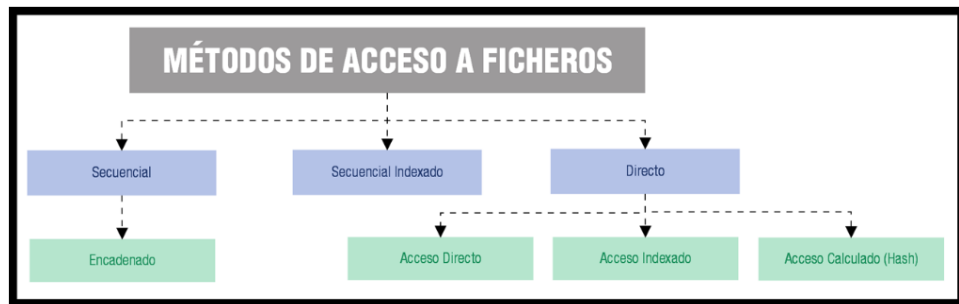
## 2.3. Métodos de acceso

A medida que la tecnología ha ido evolucionando, atendiendo principalmente a los avances hardware, el acceso a la información contenida en los diferentes tipos de ficheros ha variado mucho.

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.

Las distintas formas de organizar un fichero en un soporte de memoria o, lo que se conoce también por métodos de acceso a los ficheros se detallan en el siguiente gráfico.



## 2.4. Ficheros secuenciales

Un fichero con organización secuencial se caracteriza porque sus registros están **almacenados de forma contigua**, de manera, que la única forma de acceder a él, 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 o soportes no direccionables o 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*.

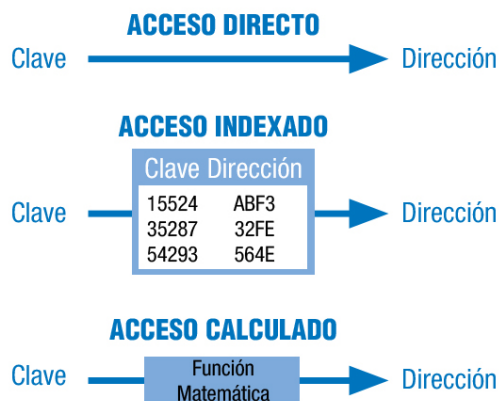
## 2.5. Ficheros de acceso directo

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.

A través de una transformación específica aplicada a la clave, se obtendrá la dirección física en la que se encuentra el registro. Según la forma de realizar esta transformación, existen diferentes modos de acceso:



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.

La medida básica de posicionamiento del puntero en el fichero es el byte, dependiendo del tipo de codificación de caracteres que empleemos (Unicode, ANSI) se utilizarán 1 o 2 bytes por carácter respectivamente. Teniendo esto en cuenta, el puntero avanzará de uno en uno o de dos en dos bytes para poder leer o escribir cada carácter.

Otras características fundamentales de los ficheros de acceso directo o aleatorio son:

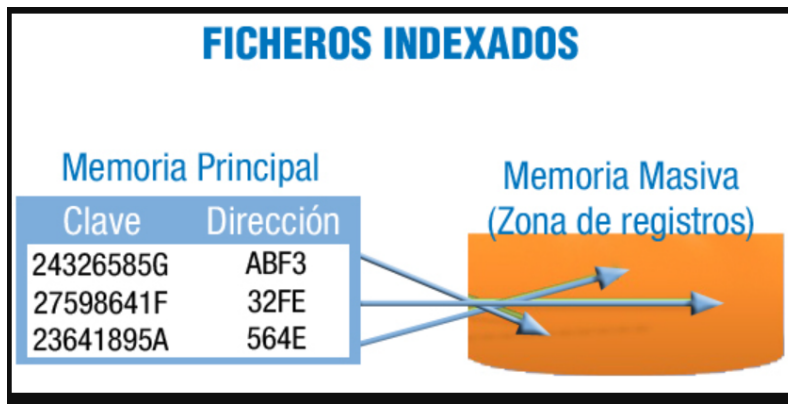
- Posicionamiento *inmediato*.
- Registros de *longitud fija*.
- Apertura del fichero en *modo mixto, para lectura y escritura*.
- Permiten *múltiples usuarios* utilizándolos.
- Los registros *se borran colocando un cero en la posición que ocupan*.
- Permiten la utilización de *algoritmos de compactación 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 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*.

## 2.6. 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 podemos recurrir al índice que nos dice 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*.



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. Un mismo fichero puede tener mas 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.

## 2.7. Otros (secuenciales indexados, hash,...)

### Ficheros secuenciales indexados

También llamados parcialmente indexados, al igual que en los ficheros indexados existe una **zona de índices** y otra **zona de registros de datos**, pero esta última se encuentra dividida en **segmentos** (bloques de registros) ordenados.

En la tabla de índices, cada fila hace referencia a cada uno de los segmentos. La clave corresponde al último registro y el índice apunta al registro inicial. Una vez que se accede al primer registro del segmento, dentro de él se localiza (de forma secuencial) el registro buscado.

Esta organización es muy utilizada, tanto para procesos en los que intervienen pocos registros como para aquellos en los que se maneja el fichero completo.

Las principales características son:

- Permite el acceso secuencial. Esto es muy interesante cuando la tasa de actividad es alta. En el acceso secuencial, además, los registros se leen ordenados por el campo clave.
- Permite el acceso directo a los registros. Realmente emula el acceso directo, empleando para ello las tablas de índices. Primero *busca la clave en el área de índices y luego va a leer al área de datos en la dirección que le indica la tabla*.
- Se pueden actualizar los registros en el mismo fichero, sin necesidad de crear un fichero nuevo de copia en el proceso de actualización.
- *Ocupa mas espacio en el disco que los ficheros secuenciales*, debido al uso del área de índices.
- Solo se puede utilizar soportes direccionables.
- Obliga a una *inversión económica mayor*, por la necesidad de programas y, a veces, hardware mas sofisticado.

### Ficheros de Acceso Calculado o Hash:

Cuando utilizamos ficheros indexados es necesario siempre tener que consultar una tabla para obtener la dirección de almacenamiento a partir de la clave. La técnica del acceso calculado o **hash**, permite accesos más rápidos, ya que en lugar de consultar una tabla, se utiliza una transformación o función matemática (función de hashing) conocida, que a partir de la clave genera la dirección de cada registro del archivo. Si la clave es alfanumérica, deberá previamente ser transformada en un número.

El **mayor problema** que presenta este tipo de ficheros es que a partir de diferentes claves se obtenga la misma dirección al aplicar la función matemática o transformación. A este problema se le denomina **colisión**, y las claves que generan la misma dirección se conocen por **sinónimos**. Para resolver este problema se aplican diferentes métodos, como tener un bloque de excedentes o zona de sinónimos, o crear un archivo de sinónimos, etc.

Para llevar a cabo la transformación existen multitud de métodos, siendo algunos:

- **Módulo**: La dirección será igual al resto de la división entera entre la clave y el número de registros.
- **Extracción**: La dirección será igual a una parte de las cifras que se extraen de la clave.

Una buena transformación o función de hash, será aquella que produzca el menor número de colisiones. En este caso hay que buscar una función, a ser posible biunívoca, que relacione los posibles valores de la clave con el conjunto de números correlativos de dirección. Esta función consistirá en realizar una serie de cálculos matemáticos con el valor de la clave hasta obtener un número entre 1 y n, siendo n el número de direcciones que tiene el fichero.

## 2.8. Parámetros de utilización

En función del uso que se vaya a dar al fichero, serán adecuados unos tipos u otros de organización. Mediante la utilización de 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:

- **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
- **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:
  - **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.
  - **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.