

Dato cualquier **elemento informativo que tenga relevancia para un usuario.**

- Una **base de datos** se entenderá como una colección de datos relacionados entre sí y que tienen un significado implícito.
- Por **datos** queremos decir hechos conocidos que pueden registrarse y que tienen un significado implícito.

La definición presentada anteriormente hace referencia a dos elementos para que un conjunto de datos constituya una Base de Datos:

1. **Relaciones entre datos**, tema que se tratará en las secciones siguientes.
2. **Significado implícito** de los datos que se atribuye dependiendo del contexto en que se utilizan los mismos. Por ejemplo, el dato fecha en una base de datos de VENTAS puede referirse a la fecha de emisión de las facturas, mientras que si la base de datos es de MÚSICA quizás corresponda a la fecha en que se grabó un tema musical. Es decir, el significado de un dato, depende de la BD que lo contenga.

TIPO DE BASE DE DATOS

- Jerárquicos
- En red.
- **Relacionales.**
- Multidimensionales.
- De objetos.

Inconvenientes de los archivos tradicionales:

- **Datos redundantes.** Ya que se repiten continuamente.
- **Coste de almacenamiento elevado.** Al almacenarse varias veces el mismo dato en distintas aplicaciones, se requiere más espacio en los discos.
- **Tiempos de procesamiento elevados.** Al no poder optimizar el espacio de almacenamiento.
- **Probabilidad alta de inconsistencia en los datos.** Ya que un proceso cambia sus datos y no el resto. Por lo que el mismo dato puede tener valores distintos según qué aplicación acceda a él.
- **Difícil modificación en los datos.** Debido a la probabilidad de inconsistencia, cada modificación se debe repetir en todas las copias del dato (algo que normalmente es imposible).

Ventas de las bases de datos relacionales:

Sus **ventajas** son las siguientes:

- **Menor redundancia.** No hace falta tanta repetición de datos. Aunque, sólo los buenos diseños de datos tienen poca redundancia.
- **Menor espacio de almacenamiento.** Gracias a una mejor estructuración de los datos.
- **Acceso a los datos más eficiente.** La organización de los datos produce un resultado más óptimo en rendimiento.
- **Datos más documentados.** Gracias a los metadatos que permiten describir la información de la base de datos.
- **Independencia de los datos y los programas y procesos.** Esto permite modificar los datos sin modificar el código de las aplicaciones.
- **Integridad de los datos.** Mayor dificultad de perder los datos o de realizar i

Desventajas de las bases de datos relacionales:

- **Instalación costosa.** El control y administración de bases de datos requiere de un software y hardware potente.
- **Requiere personal cualificado.** Debido a la dificultad de manejo de este tipo de sistemas.
- **Implantación larga y difícil.** Debido a los puntos anteriores. La adaptación del personal es mucho más complicada y lleva bastante tiempo.
- **Un diccionario de datos o catálogo** accesible por el usuario que contiene la descripción de los datos de la base de datos.

Tipos de SGBD:

- **SGBD ofimáticos:** manipulan BD pequeños orientadas a almacenar datos domésticos o de pequeñas empresas. Ejemplos típicos son Microsoft ACCESS y LibreOffice Base.
- **SGBD corporativos:** tienen la capacidad de gestionar BD enormes, de medianas o grandes empresas con una carga de datos y transacciones que requieren de un servidor de gran capacidad. Un ejemplo típico de BD corporativas es ORACLE, actualmente junto de DB2 el servidor de BD más potente del mercado (también el más caro). Nosotros para nuestro aprendizaje utilizamos una versión gratuita con fines educativos, que aunque bastante limitada, nos sirve para introducirnos en la filosofía de ORACLE.

El **acceso a un registro** es el procedimiento que se utiliza para seleccionarlo. Los tipos de acceso son:

- **Secuencial:** los registros se leen uno detrás de otro desde el principio del fichero hasta localizar el registro buscado o hasta el final del fichero.
- **Directo:** permite seleccionar a un registro sin tener que leer los anteriores, accediendo directamente a él mediante su clave.
- **Indexado:** para seleccionar un registro consultamos previamente de forma secuencial en una tabla que contiene la clave más alta y la dirección de comienzo de cada bloque de registros. Una vez localizado se utiliza el acceso directo a ese bloque de registros y ,dentro del bloque, la lectura secuencial hasta localizarle.
- **Dinámico:** permite el acceso directo o por índice a un registro y a partir de ese se accede a los demás de forma secuencial.

Secuencial

Las ventajas de los ficheros con organización secuencial son:

- Rapidez en el acceso a un bloque de registros contiguos
- No se desperdicia espacio en el dispositivo de almacenamiento porque no hay huecos
- Se pueden utilizar cualquier tipo de registros: de longitud fija, variable o indefinida.

Los inconvenientes de este tipo de organización son:

- Para **acceder** al registro n hay que recorrer los $n-1$ registros anteriores. El acceso es secuencial, la lectura siempre se realiza hacia delante.
- Para realizar una **consulta** hay que crear un proceso en el que se compare el valor del campo que se pretende localizar con el valor del mismo campo correspondiente a cada registro leído del fichero.
- La **adición** de registros se realiza a continuación del último registro ya existente. No se pueden insertar nuevos registros.
- No se pueden **eliminar** registros. Para eliminar un registro se marca de modo que no se muestre, pero el registro existe y ocupa espacio en el dispositivo del almacenamiento. (borrado lógico)
- Para **mantener ordenado** y compactado el fichero, hay que crear un fichero nuevo a partir del existente.

Fichero secuencial encadenado

Son ficheros de organización secuencial gestionados mediante **punteros** que nos permiten tener los registros ordenados según un orden lógico diferente del orden físico en el que están grabados.

Los punteros son un campo adicional, por lo que aumentan el tamaño de los registros, y forman parte integrante de la estructura del registro, indicando cual es el siguiente o el anterior registro en secuencia lógica y no en secuencia física.

Las características de los punteros son:

- Estar en cada registro en una posición fija y definida para todos los registros
- Poseer una longitud constante

Acceso directo

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.

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

- Posicionamiento inmediato.
- Registros de longitud fija.
- 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.
- 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.

Indexados

En esta organización los registros con los datos se graban en un fichero secuencialmente, pero se pueden recuperar con acceso directo gracias a la utilización de un fichero adicional, llamado de **índices**, que contiene información de la posición que ocupa cada registro en el fichero de datos.

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.

En este tipo de ficheros se distinguen tres áreas:

Área primaria:	<p>En esta área se escriben los registros ordenados ascendentemente por el valor de su clave. Esta área del fichero está dividida en segmentos. Cada segmento almacena "n" registros consecutivos y almacenados en posiciones contiguas. Es un área de organización secuencial, donde el acceso a cada registro se realiza en una doble operación:</p> <ul style="list-style-type: none">• Acceso directo al segmento donde se haya ubicado el registro buscado• Acceso secuencial posteriormente, a los registros del segmento, hasta localizar el registro buscado o alcanzar el final del segmento, en caso de que no se halle.
Área de índices	<p>Es creada por el sistema al mismo tiempo en que se almacenan los datos. En esta área los registros están formados por dos campos:</p>

	<ul style="list-style-type: none"> • El valor del campo <i>clave del último registro</i> de un bloque o segmento. Los bloques están constituidos por un número fijo de registros consecutivos. • El segundo campo contiene la <i>dirección de comienzo</i> de cada uno de los segmentos en los que se halla dividida el área primaria
<i>Área de overflow:</i>	<p>O área de excedentes. Es la zona destinada a contener los registros almacenados posteriormente a la creación del fichero, por lo que no han sido incluidos en el área primaria. Estos registros tendrán claves intermedias a las de los registros previamente almacenados en el área primaria.</p>