

BASES DE DATOS

Desarrollo de Aplicaciones Web

GESTIÓN DE BASES DE DATOS

Administración de Sistemas Informáticos en Red

Apuntes

Luis Dorado Garcés
Basado en el trabajo de Jose Luis Comesaña

ÍNDICE

1	Introducción.....	1
2	Los ficheros de datos.....	2
2.1	Concepto de fichero de datos.....	2
2.2	¿Qué es un fichero de registros?	2
2.3	Tipos de ficheros de datos estructurados.....	4
2.3.1	Según su soporte.....	4
2.3.2	Según su método de acceso	5
2.3.3	<i>Ampliación: Según su vigencia.....</i>	6
2.4	Ficheros secuenciales.....	6
2.5	Ficheros de acceso directo o aleatorio	7
2.6	Ficheros directos-indexados	8
2.7	Ficheros de Acceso Calculado o Hash	10
2.8	<i>Ampliación: Ficheros Secuenciales Indexados.....</i>	11
2.9	<i>Ampliación: Parámetros de utilización</i>	11
3	Bases de datos.....	11
3.1	Conceptos	11
3.2	Usos.....	12
3.2.1	¿Para qué se utilizan las bases de datos?	12
3.2.2	¿Quién utiliza las bases de datos?	12
3.3	Ubicación de la información	13
4	Modelos obsoletos de bases de datos.....	15
4.1	Modelo jerárquico	15
4.2	Modelo en red	16
5	Modelo relacional	17
5.1	<i>Ampliación: Componentes del modelo relacional.....</i>	17
5.2	<i>Ampliación: Modelo orientado a objetos.....</i>	17
5.3	<i>Ampliación: Otros modelos</i>	17
5.3.1	<i>Modelo Objeto-Relacional</i>	17
5.3.2	<i>Modelo de bases de datos deductivas</i>	17
5.3.3	<i>Bases de datos multidimensionales</i>	17
5.3.4	<i>Bases de datos transaccionales</i>	17
6	<i>Ampliación: Modelo NoSQL.....</i>	18
7	<i>Ampliación: Tipos de bases de datos</i>	18
8	Sistemas gestores de bases de datos.....	18
8.1	Funciones	19
8.2	<i>Ampliación: Componentes</i>	20
8.3	<i>Ampliación: Arquitectura</i>	20
8.4	Tipos.....	20
8.5	<i>Ampliación: Funcionamiento</i>	21
9	<i>Ampliación SGBD comerciales.....</i>	21
10	<i>Ampliación SGBD libres</i>	21
11	<i>Ampliación: Bases de datos centralizadas</i>	21
12	<i>Ampliación: Bases de datos distribuidas</i>	21
12.1	<i>Ampliación: Fragmentación.....</i>	21

Caso práctico

Ada sabe bien que BK Programación deberá hacer frente a retos importantes que requerirán del dominio adecuado de múltiples disciplinas. Tiene claro que el desarrollo de sus proyectos ha de estar apoyado sobre unas bases firmes, y una de ellas será la gestión adecuada de los datos.

Considera que **Juan y María** deben conocer la evolución que han experimentado las técnicas de almacenamiento de información, destacando que el dominio de las bases de datos es fundamental para garantizar un funcionamiento óptimo de las aplicaciones que BK Programación va a tener que desarrollar.

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:

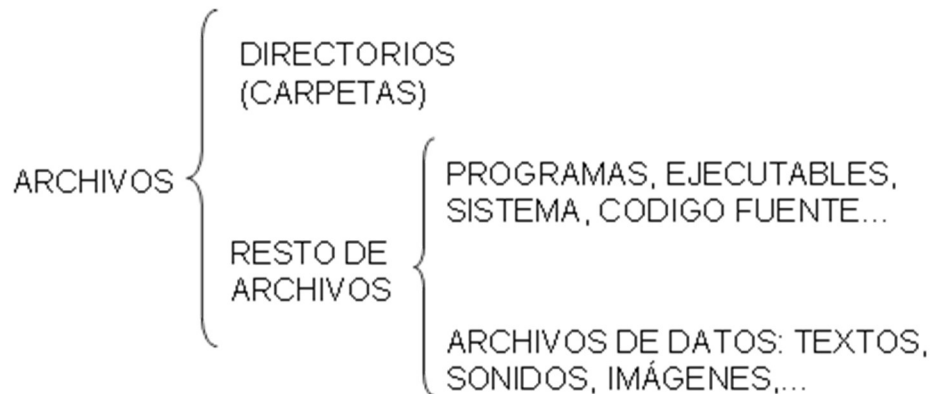
- ✓ Cuando seleccionamos nuestro canal favorito en la TDT.
- ✓ Al utilizar la agenda del móvil para realizar una llamada telefónica.
- ✓ Cuando operamos en el cajero automático.
- ✓ Al solicitar un certificado en un organismo público.
- ✓ Cuando acudimos a la consulta del médico.
- ✓ Al inscribirnos 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.)
- ✓ 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.

Suponemos que no es necesario que continuemos más para darnos cuenta de que 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 datos

2.1 Concepto de fichero de datos



La información de un ordenador está almacenada en lo que se llaman archivos o ficheros. Normalmente los archivos están formados por un nombre, un punto y una extensión (p.e. PROGRAMA.EXE).

El nombre nos sirve para diferenciar unos archivos de otros y la extensión para atribuirle unas propiedades concretas. Estas propiedades asociadas o "tipo de archivo" vienen dadas por las letras que conforman la extensión. Normalmente su máximo son tres letras aunque existen algunas excepciones (.jpeg, .html, .java, etc.). Cada uno de estos pequeños grupos de caracteres está asociado a un tipo de archivo.

Podemos dividir los archivos en dos grandes grupos.

- Ficheros ejecutables: También llamados programas o aplicaciones. Contienen instrucciones que procesará la CPU para realizar una tarea concreta. Procesar texto, ejecutar un videojuego, etc. En ellos reside la **funcionalidad**.
- Ficheros de datos: Almacenan información que tendrá que ser procesada con ayuda de algún programa.

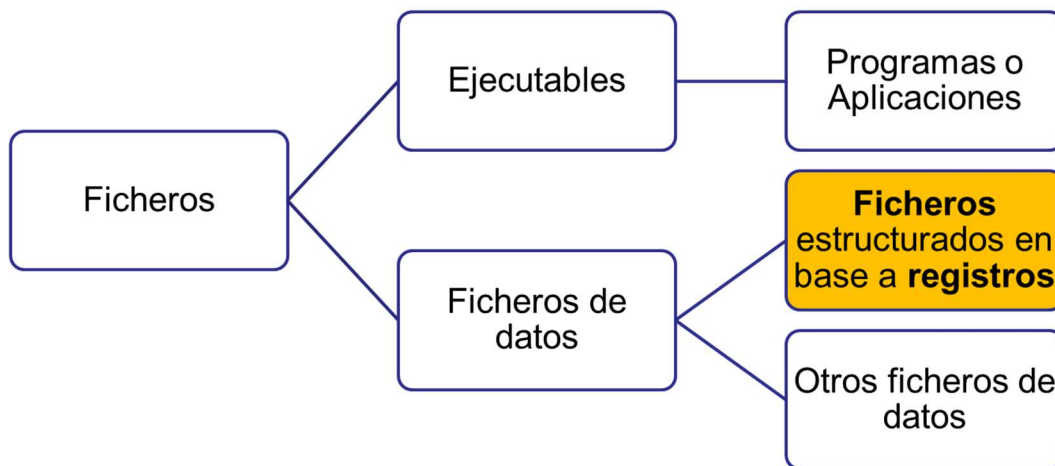
Ejemplos:

Fichero Ejecutable	Fichero de datos
Visualizador, editor gráfico...	Imagen (JPG, PNG...)
Reproductor de audio	Fichero de sonido (MP3, FLAC, WAV...)
Reproductor multimedia	Fichero de vídeo (MP4, AVI, MKV...)
Excel, Power Point, Word...	Hojas de cálculo (XLSX), Presentación (PPTX), Documento de texto (DOCX)...

De todos modos, la mayoría de los programas llevan otros archivos que resultan necesarios aparte del ejecutable. Suelen tratarse de ficheros de configuración (almacenan la configuración del programa). Estos archivos adjuntos que requieren los programas son necesarios para su buen funcionamiento, y aunque puedan tener formatos distintos, no pueden ser separados de su programa original. En Windows tenemos ejemplos de estos ficheros de configuración tales como DLL (bibliotecas de vínculos dinámicos) o DAT (datos de programa de diversos tipos).

2.2 ¿Qué es un fichero de registros?

Básicamente es un fichero de datos específico que almacena información estructurada en base a registros.



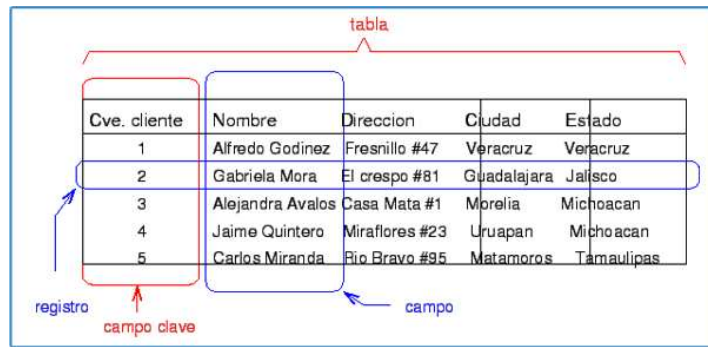
En la década de los setenta, los procesos básicos que se llevaban a cabo en una empresa se centraban en cuestiones relacionadas con contabilidad y facturación. Las necesidades de almacenamiento y gestión de información podían satisfacerse utilizando un número relativamente reducido de archivos en papel 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 de datos (ficheros a secas a partir de ahora): 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 trabajadores en un fichero Excel**). 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).



Características de un registro:

- ✓ Los **registros de datos** son conjuntos de datos almacenados en un sistema.
- ✓ Se suelen representar gráficamente como **filas de una tabla** estructurado en campos.
- ✓ Todos los registros tienen la misma estructura.
- ✓ Normalmente, todos los registros tienen un **campo cuyo valor** es único para cada registro. Este campo se llama **clave**.

ID Empl	Apellido	Nombre	Empleo	Departamento	Sección	Salario	Fch comienzo
1011	Gorton	Hazel	Auxiliar Contable	Contabilidad	Copiadoras	\$27.598	03/02/1986

1 Ejemplo de registro

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.

2.3 Tipos de ficheros de datos estructurados

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.

2.3.1 Según su soporte

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.

Conoce más sobre las características de cintas y discos a través de los enlaces que te proponemos:

http://es.wikipedia.org/wiki/Cinta_magn%C3%A9tica_de_almacenamiento_de_datos

<http://www.infodata.es/tour/disco-duro.html>

http://es.wikipedia.org/wiki/Disco_%C3%B3ptico

<http://www.consumer.es/web/es/tecnologia/hardware/2006/01/26/148862.php>

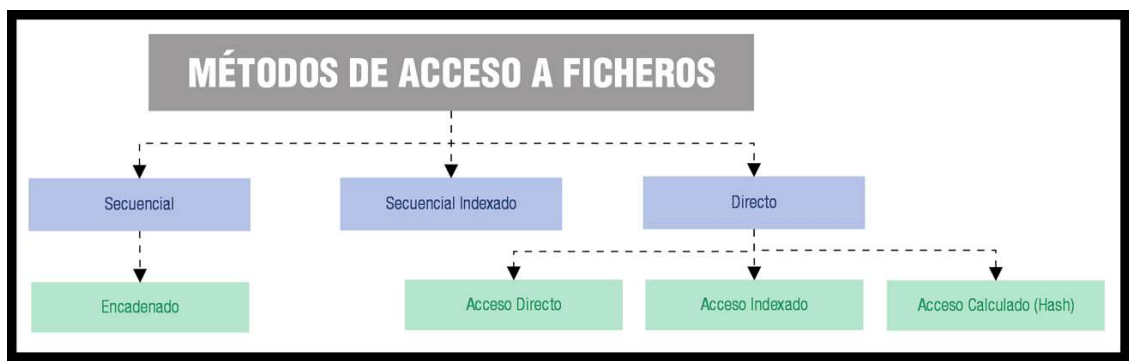
<http://www.infodata.es/tour/magneto-opticos.html>

2.3.2 Según su método 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.



Las organizaciones secuenciales, de acceso aleatorio o directo y de acceso indexado son las más comunes. En los siguientes epígrafes se detallarán las características de cada uno de los métodos de acceso a los ficheros.

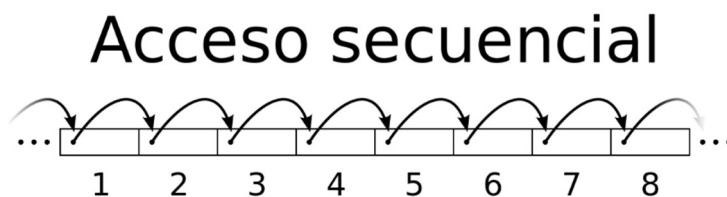
Relaciona los diferentes métodos de acceso a los ficheros.

Método de acceso	Relación	Tipo de acceso
Encadenado.	2	1. Directo.
Indexado.	3	2. Secuencial.
Calculado o Hash.	1	3. Directo o secuencial.

Como ves el método de acceso Indexado existe en los dos tipos de acceso: directo y secuencial.

2.3.3 Ampliación: Según su vigencia**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.

Características-ventajas

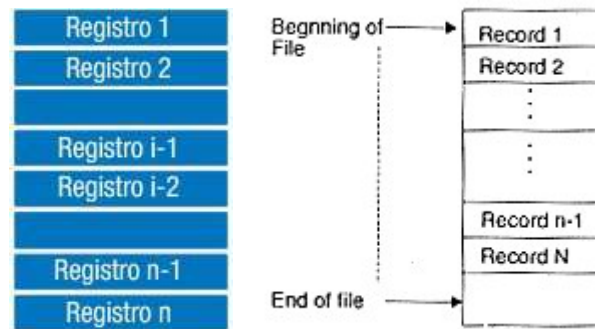
- ✓ **Rapidez** en el acceso a un bloque de registros **contiguos**.
- ✓ Estructura versátil. Se pueden utilizar **cualquier tipo de registros**: de longitud fija o variable.
- ✓ Como no permite borrados, aprovechan al máximo el soporte de almacenamiento al **no dejar huecos vacíos**.
- ✓ Puede ser usado cualquier tipo de soporte, **tanto en secuenciales como direccionables**.

Características-inconvenientes

- ✓ El acceso es secuencial, la lectura siempre se realiza **hacia delante**. Para acceder al registro n hay que **recorrer los n-1 registros anteriores**.
- ✓ Son ficheros **monousuario**, no permiten el acceso simultáneo de varios usuarios.
- ✓ El modo de apertura del fichero es exclusivo: **o lectura o escritura**.
- ✓ **No permite borrados**. Para eliminar un registro se **marca como inactivo/borrado**, pero el registro existe y ocupa espacio en el dispositivo del almacenamiento (borrado lógico).
- ✓ La **adición** de registros se realiza **a continuación del último** registro ya existente. No se pueden insertar registros entre los que ya están grabados.
- ✓ Para mantener **ordenado y compactado** el fichero, hay que **crear un fichero nuevo** a partir del existente.

Dirección de memoria	Marca de borrado	Nombre	Apellidos	Teléfono
1200		Alfredo	Bárcena	768334472
1300	*	Isabel	De los Ríos	987335612
1400		Carmen	Sierra	955347612
1500		Fernando	Ruiz	674992455
1600		Juan Carlos	Abad	573982277

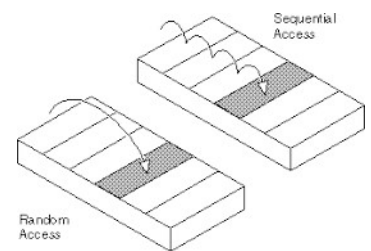
En el siguiente gráfico se observa la estructura de un fichero secuencial.



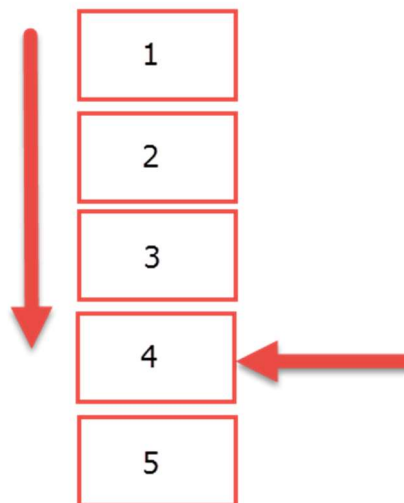
2.5 Ficheros de acceso directo o aleatorio

En este tipo de ficheros se puede acceder a un registro indicando la posición relativa del mismo dentro del archivo (la dirección).

Estos archivos deben almacenarse en dispositivos de memoria masiva de acceso directo, como son los discos magnéticos o las memorias Flash.



With **sequential access**, elements #1,2, 3 must be processed before element #4 can be processed.



With **direct access**, element #4 in the list can be accessed without having to process the elements before it.

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**, no siendo necesario recorrer todo el fichero para encontrar un determinado registro.

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.

Características fundamentales

- ✓ **Posicionamiento inmediato.** Se accede directamente al registro buscado, disminuyendo el tiempo de acceso sensiblemente.
- ✓ **Registros de longitud fija.**
- ✓ Permiten **múltiples usuarios** utilizándolos.
- ✓ Los registros se borran colocando una marca de borrado en la posición que ocupan. Cuando se necesita **añadir un registro serán sobrescritos.**
- ✓ Esta organización sólo es posible en **soportes direccionables** (soportes de acceso directo a través de una dirección).
- ✓ Permiten realizar procesos de actualización en tiempo real, es decir, el usuario no percibe retardo al actualizar ficheros.

2.6 Ficheros directos-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.



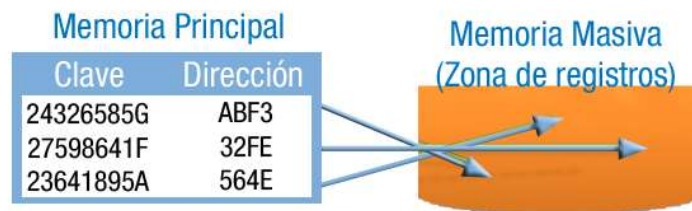
Por tanto, existirá una **zona de registros** en la que se encuentran los datos del archivo y una **zona de índices**, que contiene una tabla con las claves de los registros y las posiciones donde se encuentran los mismos. La tabla de índices estará ordenada por el campo clave.

1. La tabla de índices **será cargada en memoria principal (RAM)** para realizar en ella la búsqueda **secuencial** de la **clave del registro** a encontrar, obteniéndose así **la dirección** donde se encuentra el registro.
2. Una vez localizada la dirección, sólo hay que acceder a la zona de registros en el soporte de almacenamiento (disco duro) 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.

Índice
Registro 0
Registro 1
Registro 2
Registro 3
Registro 4
Registro 5
⋮

FICHEROS INDEXADOS



1. La **consulta** se realiza mediante una consulta secuencial de la tabla de índices para determinar donde se encuentra (dirección) el registro buscado y realizar un acceso directo al registro.
2. El **borrado**: la eliminación de los registros debe realizarse mediante marcas. Se generan huecos que realmente son posiciones de memoria ocupadas por registros marcados pero que no han sido eliminados físicamente del fichero. Estos registros marcados serán ocupados cuando se realicen adiciones.
3. La **inserción** de registros se realiza creando un nuevo registro tanto en el índice como en la tabla de registro. En caso de haber registros marcados como borrados se usarán estos huecos.

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

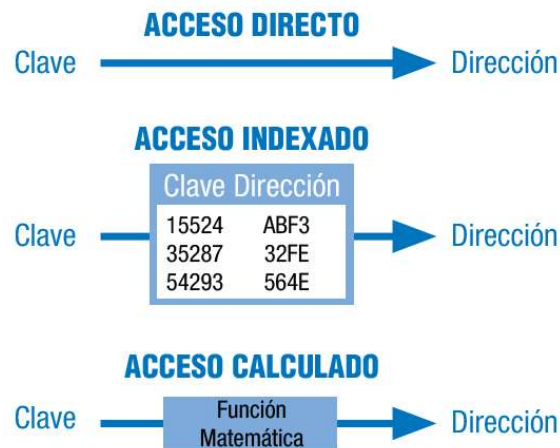
- ✓ **Posicionamiento inmediato** una vez consultada la tabla índice.
- ✓ **Distribución aleatoria** dentro del soporte de almacenamiento
- ✓ Registros de **longitud fija**.
- ✓ El fichero contará con una **tabla índice de pares clave-dirección** que, a partir de una clave, devolverá su dirección.
- ✓ Permiten utilizar dispositivos de acceso **secuencial y directo**.
- ✓ **Multiusuario**.
- ✓ Aunque el acceso a la **tabla índice** es secuencial, se halla en RAM y el **tiempo de acceso a memoria principal es despreciable**.

Ejemplo

Índice cargado en memoria RAM	
Clave	Dirección
1056	1
1079	10
1290	5
1368	6
1557	2
1676	3
1793	7
1908	8
1961	9
1967	4

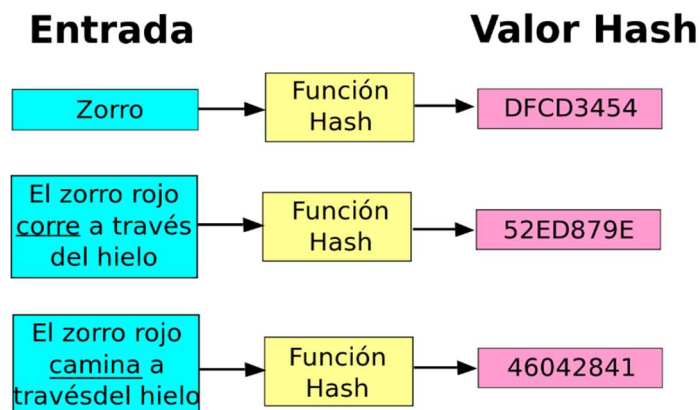
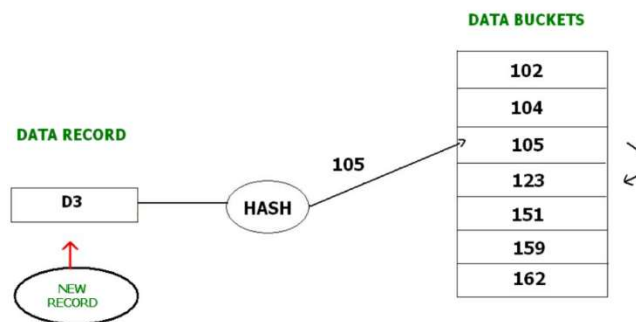
Fichero ubicado en disco duro			
Dirección	ID Empl	Apellido	Nombre
1	1056	Gonzales	Joe
2	1557	Bates	Lisa
3	1676	Wells	Jason
4	1967	Aruda	Felice
5	1290	Cooper	Linda
6	1368	Wu	Tammy
7	1793	Able	Aaron
8	1908	Zostoc	Melissa
9	1961	Mueller	Kris
10	1079	Price	Ellen

2.7 Ficheros de Acceso Calculado o Hash

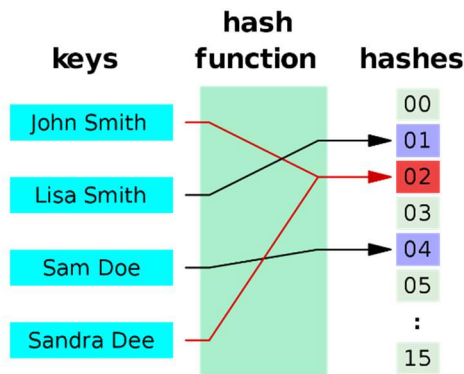


Cuando utilizamos ficheros indexados es necesario siempre tener que consultar una tabla para obtener la dirección de almacenamiento a partir de la clave.

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.



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.



2.8 Ampliación: Ficheros Secuenciales Indexados

2.9 Ampliación: Parámetros de utilización

3 Bases de datos

Si *BK Programación* va a desarrollar aplicaciones para diferentes ámbitos, deberá documentarse adecuadamente para poder seleccionar qué sistema de base de datos debe utilizar en cada situación. Para ello, todos sus miembros tendrán que recordar, actualizar o aprender gran cantidad de interesantes conocimientos relacionados con este campo de la informática.

Como hemos visto anteriormente, los ficheros permiten organizar y memorizar conjuntos de datos del mismo tipo o naturaleza con una determinada estructura, siendo un medio para el almacenamiento de los datos o resultados de una aplicación específica. Pero si las aplicaciones, al ser diseñadas, deben depender directamente de sus ficheros o archivos, se pierde independencia y surgen serios inconvenientes: como información duplicada, incoherencia de datos, fallos de seguridad, dificultad para establecer relaciones complejas, etc.

Estos problemas debían ser solucionados, es cuando aparece el concepto de base de datos. Una base de datos permitirá reunir toda la información relacionada en un único sistema de almacenamiento, pudiendo cualquier aplicación utilizarla de manera independiente y ofreciendo una mejora en el tratamiento de la información, así como una evolución para el desarrollo de aplicaciones.

La gestión de las bases de datos ha experimentado gran cantidad de cambios, partiendo de aplicaciones especializadas hasta llegar a convertirse en el núcleo de los entornos informáticos modernos. Con la llegada de Internet en los noventa, el número de usuarios de bases de datos creció exponencialmente, y aunque muchos de ellos no sean conscientes de ello, el acceso a dichas bases forma parte de la vida cotidiana de muchos de nosotros.

Conocer los sistemas que gestionan las bases de datos, sus conceptos fundamentales, el diseño, lenguajes y la implementación de éstas, podemos considerarlo imprescindible para alguien que se está formando en el campo de la informática.

3.1 Conceptos

A finales de los setenta, la aparición de nuevas tecnologías de manejo de datos a través de los sistemas de bases de datos supuso un considerable cambio. Los sistemas basados en ficheros separados dieron paso a la utilización de sistemas gestores de bases de datos, que son sistemas software centralizados o distribuidos que ofrecen facilidades para la definición de bases de datos, selección de estructuras de datos y búsqueda de forma interactiva o mediante lenguajes de programación.

Llegados a este punto, te preguntarás... ¿Qué es una base de datos?

Base de datos: Es una colección de datos relacionados lógicamente entre sí, con una definición y descripción comunes y que están estructurados de una determinada manera. Es un conjunto estructurado de datos que representa entidades y sus interrelaciones, almacenados con la mínima redundancia y posibilitando el acceso a ellos eficientemente por parte de varias aplicaciones y usuarios.

La base de datos no sólo contiene los datos de la organización, también almacena una descripción de dichos datos. Esta descripción es lo que se denomina **metadatos**, se almacena en el **diccionario de datos o catálogo** y es lo que permite que exista **independencia de datos** lógica-física.

3.2 Usos

Ya sabemos lo que es una base de datos y sus características principales, pero es necesario conocer quien las usa y para qué.

3.2.1 ¿Para qué se utilizan las bases de datos?

Enumerar todos y cada uno de los campos donde se utilizan las bases de datos es complejo, aunque seguro que quedarán muchos en el tintero, a continuación se recopilan algunos de los ámbitos donde se aplican.

- ✓ Banca: información de clientes, cuentas, transacciones, préstamos, etc.
- ✓ Líneas aéreas: información de clientes, horarios, vuelos, destinos, etc.
- ✓ Universidades: información de estudiantes, carreras, horarios, materias, etc.
- ✓ Transacciones de tarjeta de crédito: para comprar con tarjetas de crédito y la generación de los extractos mensuales.
- ✓ Telecomunicaciones: para guardar registros de llamadas realizadas, generar facturas mensuales, mantener el saldo de las tarjetas telefónicas de prepago y almacenar información sobre las redes.
- ✓ Medicina: información hospitalaria, biomedicina, genética, etc.
- ✓ Justicia: casos, sentencias, investigaciones, etc.
- ✓ Legislación: normativa, registros, etc.
- ✓ Organismos públicos: datos ciudadanos, certificados, etc.
- ✓ Sistemas de posicionamiento geográfico.
- ✓ Hostelería y turismo: reservas de hotel, vuelos, excursiones, etc.
- ✓ Ocio digital: juegos online, apuestas, etc.
- ✓ Cultura: gestión de bibliotecas, museos virtuales, etc.
- ✓ Etc.

3.2.2 ¿Quién utiliza las bases de datos?

Existen cuatro tipos de personas que pueden hacer uso de una base de datos: el administrador, los diseñadores de la base de datos, los programadores de aplicaciones y los usuarios finales.

¿Quién utiliza las bases de datos?	
Tipo	Funciones y características
El administrador	<p>Es la persona encargada de la creación o implementación física de la base de datos.</p> <p>Es quien escoge los tipos de ficheros, los índices que hay que crear, la ubicación de estos, etc.</p> <p>En general, es quien toma las decisiones relacionadas con el funcionamiento físico del almacenamiento de información. Siempre teniendo en cuenta las posibilidades del sistema de información con el que trabaje.</p> <p>Junto a estas tareas, el administrador establecerá la política de seguridad y de acceso para garantizar el menor número de problemas.</p>
Los diseñadores	<p>Son las personas encargadas de diseñar cómo será la base de datos. Llevarán a cabo la identificación de los datos, las relaciones entre ellos, sus restricciones, etc.</p> <p>Para ello han de conocer a fondo los datos y procesos a representar en la base de datos.</p> <p>Si estamos hablando de una empresa, será necesario que conozcan las reglas de negocio en la que esta se mueve. Para obtener un buen resultado, el diseñador de la base de datos debe implicar en el proceso a todos los usuarios de la base de datos, tan pronto como sea posible.</p>
Los programadores de aplicaciones	<p>Una vez diseñada y construida la base de datos, los programadores se encargarán de implementar los programas de aplicación que servirán a los usuarios finales. Estos programas de aplicación ofrecerán la posibilidad de realizar consultas de datos, inserción, actualización o eliminación de los mismos. Para desarrollar estos programas se utilizan lenguajes de tercera o cuarta generación.</p>
Los usuarios finales	<p>Son los clientes finales de la base de datos. Al diseñar, implementar y mantener la base de datos se busca cumplir los requisitos establecidos por el cliente para la gestión de su información.</p>

3.3 Ubicación de la información

Utilizamos a diario las bases de datos, pero ¿Dónde se encuentra realmente almacenada la información? Las bases de datos pueden tener un tamaño muy reducido (1 MegaByte o menos) o bien, ser muy voluminosas y complejas (del orden de Terabytes). Sin embargo todas las bases de datos normalmente se almacenan y localizan en discos duros y otros dispositivos de almacenamiento, a los que se accede a través de un ordenador. Una gran base de datos puede necesitar servidores en lugares diferentes, y viceversa, pequeñas bases de datos pueden existir como archivos en el disco duro de un único equipo.

A continuación, se exponen los sistemas de almacenamiento de información más utilizados para el despliegue de bases de datos, comenzaremos por aquellos en los que pueden alojarse bases de datos de tamaño pequeño y mediano, para después analizar los sistemas de alta disponibilidad de grandes servidores.

- ✓ **Discos SATA:** Es una interfaz de transferencia de datos entre la placa base y algunos dispositivos de almacenamiento, como puede ser el disco duro, lectores y regrabadores de CD/DVD/BD, Unidades de Estado Sólido u otros dispositivos. SATA proporciona mayores velocidades, mejor

aprovechamiento cuando hay varias unidades, mayor longitud del cable de transmisión de datos y capacidad para conectar unidades al instante, es decir, insertar el dispositivo sin tener que apagar el ordenador. La primera generación específica en transferencias de 150 Megabytes por segundo, también conocida por SATA 150 MB/s o Serial ATA-150. Actualmente se comercializan dispositivos SATA II, a 300 MB/s, también conocida como Serial ATA-300 y los SATA III con tasas de transferencias de hasta 600 MB/s.

- ✓ **Discos SCSI:** Son interfaces preparadas para discos duros de gran capacidad de almacenamiento y velocidad de rotación. Se presentan bajo tres especificaciones: SCSI Estándar (Standard SCSI), SCSI Rápido (Fast SCSI) y SCSI Ancho-Rápido (Fast-Wide SCSI). Su tiempo medio de acceso puede llegar a 7 milisegundos y su velocidad de transmisión secuencial de información puede alcanzar teóricamente los 5 MB/s en los discos SCSI Estándares, los 10 MBps en los discos SCSI Rápidos y los 20 MBps en los discos SCSI Anchos-Rápidos (SCSI-2). Un controlador SCSI puede manejar hasta 7 discos duros SCSI.
- ✓ **RAID:** acrónimo de Redundant Array of Independent Disks o matriz de discos independientes, es un contenedor de almacenamiento redundante. **Se basa en el montaje en conjunto de dos o más discos duros**, formando un bloque de trabajo, para obtener desde una ampliación de capacidad a mejoras en velocidad y seguridad de almacenamiento. Según las características que queramos primar, se establecen distintos sistemas de RAID.
- ✓ **Sistemas NAS:** Es el acrónimo de Network Attached Storage ó sistema de almacenamiento masivo en red. Estos sistemas de almacenamiento permiten compartir la capacidad de almacenamiento de un computador (Servidor) con ordenadores personales o servidores clientes a través de una red, haciendo uso de un sistema operativo optimizado para dar acceso a los datos a través de protocolos de comunicación específicos. Suelen ser dispositivos para almacenamiento masivo de datos con capacidades muy altas, de varios Terabytes, generalmente superiores a los discos duros externos y además se diferencian de estos al conectar por red.
- ✓ **Sistemas SAN:** Acrónimo de Storage Area Network o red de área de almacenamiento. Se trata de una red concebida para conectar servidores, matrices (arrays) de discos y librerías de soporte. La arquitectura de este tipo de sistemas permite que los recursos de almacenamiento estén disponibles para varios servidores en una red de área local o amplia. Debido a que la información almacenada no reside directamente en ninguno de los servidores de la red, se optimiza el poder de procesamiento para aplicaciones comerciales y la capacidad de almacenamiento se puede proporcionar en el servidor donde más se necesite.

Puedes ampliar más información sobre algunos de los sistemas de almacenamiento vistos, además de tendencias y curiosidades en almacenamiento, a través de los siguientes enlaces:

<http://es.wikipedia.org/wiki/RAID>

<http://www.ideasgeek.net/2010/05/12/nas-dispositivo-de-almacenamiento-externo-masivo-conectado-por-red/> <https://tihuilo.wordpress.com/2010/05/27/sistemas-de-almacenamiento/> <http://www.oracle.com/technetwork/server-storage/general/3d-demos-333955.html>

<http://databaseandtech.wordpress.com/2009/06/22/fusion-tables-google-trae-la-base-de-datos-a-la-nube-del-internet/> <http://www.cosasquecontar.com/2011/04/bigtable-como-google-almacena-los-datos/>

4 Modelos obsoletos de bases de datos

La clasificación tradicional de las bases de datos establece tres modelos de bases de datos: jerárquico, en red y relacional. En la actualidad el modelo de bases de datos más extendido es el relacional. Aunque, hay que tener en cuenta que dos de sus variantes (modelo de bases de datos distribuidas y orientadas a objetos) son las que se más se están utilizando en los últimos tiempos.

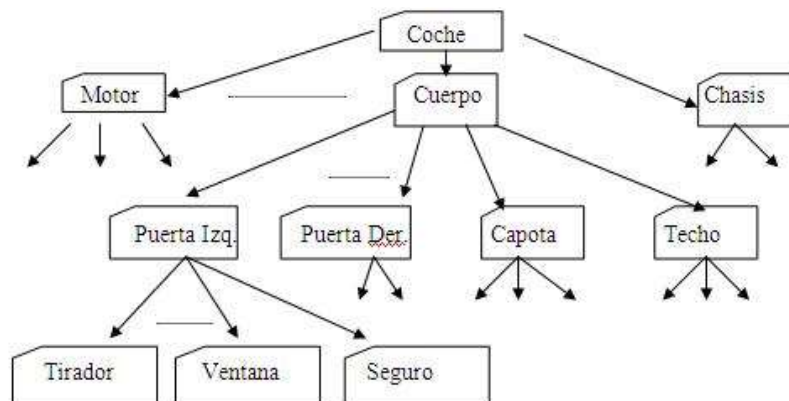
En los siguientes epígrafes analizaremos cada uno de ellos, así como otros modelos de bases de datos existentes.

Conoce las características generales y graba en tu memoria fotográfica los gráficos que representan a cada uno de los modelos expuestos en el siguiente artículo:
<http://es.kioskea.net/contents/bdd/bddtypes.php3>

4.1 Modelo jerárquico

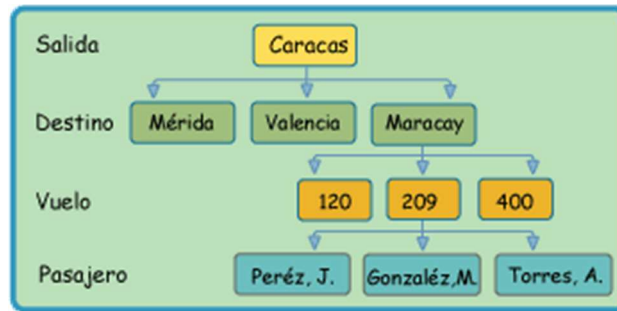
Cuando IBM creó su Sistema Administrador de Información o IMS, se establecieron las bases para que la gran mayoría de sistemas de gestión de información de los años setenta utilizaran el modelo jerárquico. También recibe el nombre de modelo en árbol, ya que utiliza una estructura en árbol invertido para la organización de los datos.

La información se organiza con una jerarquía en la que la relación entre las entidades de este modelo siempre es del tipo padre/hijo. De tal manera que existen nodos que contienen atributos o campos y que se relacionarán con sus nodos hijos, pudiendo tener cada nodo más de un hijo, pero un nodo siempre tendrá un sólo padre.



La forma visual de este modelo es de árbol invertido, en la parte superior están los padres y en la inferior los hijos.

Hoy en día, debido a sus limitaciones, el modelo jerárquico está en desuso. En el siguiente gráfico puedes observar la estructura de almacenamiento del modelo jerárquico.



4.2 Modelo en red

El modelo de datos en red aparece a mediados de los sesenta como respuesta a limitaciones del modelo jerárquico en cuanto a representación de relaciones más complejas. Podemos considerar a IDS (Integrated Data Store) de Bachman como el primer sistema de base de datos en red. Tras él se intentó crear un estándar de modelo de red por parte de CODASYL, siendo un modelo que tuvo gran aceptación a principios de los setenta.

El modelo en red organiza la información en registros (también llamados nodos) y enlaces. En los registros se almacenan los datos, mientras que los enlaces permiten relacionar estos datos. Las bases de datos en red son parecidas a las jerárquicas sólo que en ellas puede haber más de un padre.

En este modelo se pueden representar perfectamente cualquier tipo de relación entre los datos, pero hace muy complicado su manejo. Al no tener que duplicar la información se ahorra espacio de almacenamiento.

El sistema de gestión de información basado en el modelo en red más popular es el sistema IDMS.

Si deseas completar tus conocimientos acerca de este modelo, te proponemos los siguientes enlaces:

http://es.wikipedia.org/wiki/Modelo_de_red

http://ddd.uab.cat/pub/elies/elies_a2000v9/4-2-2.htm

5 Modelo relacional

Este modelo es posterior a los dos anteriores y fue desarrollado por Codd en 1970. Hoy en día las bases de datos relacionales son las más utilizadas.

En el modelo relacional la base de datos es percibida por el usuario como un conjunto de tablas. Esta percepción **es sólo a nivel lógico**, ya que a nivel físico puede estar implementada mediante distintas estructuras de almacenamiento.

El modelo relacional utiliza **tablas bidimensionales** para la representación lógica de los datos y las relaciones entre ellos. Cada tabla posee un nombre que es único y contiene un conjunto de columnas.

Se llamará **registro o tupla** a cada fila de la tabla y **campo o atributo** a cada columna de la tabla.

A los conjuntos de valores que puede tomar un determinado atributo, se le denomina **dominio**.

Una **clave** será un atributo o conjunto de atributos que identifique de forma única a una tupla.

Las tablas deben cumplir una serie de requisitos:

- ✓ Todos los registros de una table son del mismo tipo, es decir, tienen los mismos atributos.
- ✓ No existen campos o atributos repetidos.
- ✓ No existen registros duplicados.
- ✓ No existe orden en el almacenamiento de los registros.
- ✓ Cada registro o tupla es identificada por una **clave** que puede estar formada por uno o varios campos o atributos.

A continuación puedes observar cómo es una relación con sus tuplas y atributos en el modelo relacional.



El lenguaje habitual para construir las operaciones con las bases de datos relacionales es SQL, Structured Query Language o Lenguaje Estructurado de Consultas, un estándar implementado por los principales motores o sistemas de gestión de bases de datos relacionales.

Durante su diseño, una base de datos relacional pasa por un proceso al que se conoce como normalización de una base de datos.

5.1 Ampliación: Componentes del modelo relacional

5.2 Ampliación: Modelo orientado a objetos

5.3 Ampliación: Otros modelos

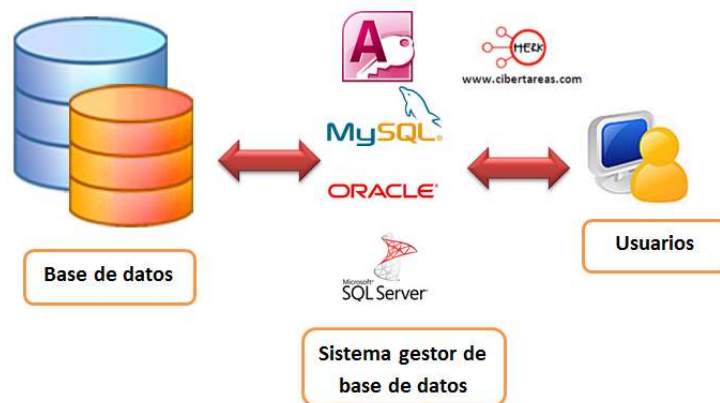
6 Ampliación: Modelo NoSQL

7 Ampliación: Tipos de bases de datos

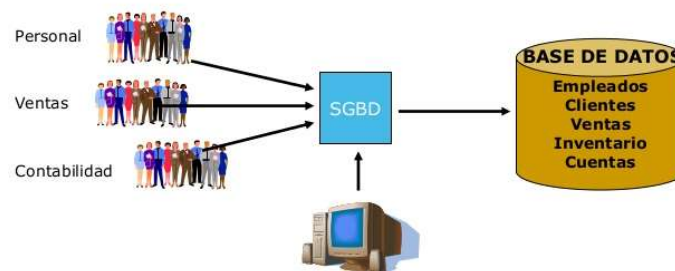
8 Sistemas gestores de bases de datos

Para poder tratar la información contenida en las bases de datos se utilizan los sistemas gestores de bases de datos o SGBD, también llamados DBMS (DataBase Management System), que ofrecen un conjunto de programas que permiten crear, acceder y gestionar dichos datos.

El objetivo fundamental de los SGBD es proporcionar eficiencia y seguridad a la hora de recuperar o insertar información en las bases de datos. Estos sistemas están diseñados para la manipulación de grandes bloques de información.



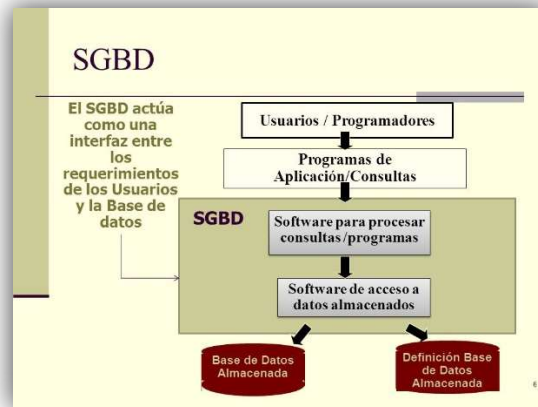
Sistema Gestor de Base de Datos: Conjunto coordinado de programas, procedimientos, lenguajes, etc., que suministra, tanto a los usuarios no informáticos, como a los analistas programadores, o al administrador, los medios necesarios para describir y manipular los datos contenidos en la base de datos, manteniendo su integridad, confidencialidad y seguridad.



El SGBD permite a los usuarios la creación y el mantenimiento de una base de datos, facilitando la definición, construcción y manipulación de la información contenida en éstas.

- **Definir una base de datos** consistirá en especificar los tipos de datos, las estructuras y las restricciones que los datos han de cumplir a la hora de almacenarse en dicha base.
- Por otro lado, **la construcción de la base de datos** será el proceso de almacenamiento de datos concretos en algún medio o soporte de almacenamiento que esté supervisado por el SGBD.
- Finalmente, **la manipulación de la base de datos** incluirá la posibilidad de realización de consultas

para recuperar información específica, la actualización de los datos y la generación de informes a partir de su contenido.



Las ventajas del uso de SGBD son:

- ✓ Proporcionan al usuario una visión abstracta de los datos, ocultando parte de la complejidad relacionada con cómo se almacenan y mantienen los datos.
- ✓ Ofrecen Independencia física, es decir, la visión que tiene de la información el usuario, y la manipulación de los datos almacenados en la Base de Datos, es independiente de cómo estén almacenados físicamente.
- ✓ Disminuyen la redundancia y la inconsistencia de datos.
- ✓ Aseguran la integridad de los datos.
- ✓ Facilitan el acceso a los datos, aportando rapidez y evitando la pérdida de datos.
- ✓ Aumentan la seguridad y privacidad de los datos.
- ✓ Mejoran la eficiencia.
- ✓ Permiten compartir datos y accesos concurrentes.
- ✓ Facilitan el intercambio de datos entre distintos sistemas.
- ✓ Incorporan mecanismos de copias de seguridad y recuperación para restablecer la información en caso de fallos en el sistema.

El SGBD interacciona con otros elementos software existentes en el sistema, concretamente con el sistema operativo (SO). Los datos almacenados de forma estructurada en la base de datos son utilizados indistintamente por otras aplicaciones, será el SGBD quien ofrecerá una serie de facilidades a éstas para el acceso y manipulación de la información, basándose en las funciones y métodos propios del sistema operativo.

8.1 Funciones

Un SGBD desarrolla tres funciones fundamentales como son las de descripción, manipulación y utilización de los datos. A continuación se detallan cada una de ellas.

1. **Función de descripción o definición:** Permite al diseñador de la base de datos crear las estructuras apropiadas para integrar adecuadamente los datos. Esta función es la que permite definir las tres estructuras de la base de datos: Estructura interna, Estructura conceptual y Estructura externa. (Estos conceptos se verán más adelante en el epígrafe sobre arquitectura del SGBD). Esta función se realiza mediante el **lenguaje de definición de datos** o **DDL**. Mediante ese lenguaje: se definen las estructuras de datos, se definen las relaciones entre los datos y se definen las reglas

(restricciones) que han de cumplir los datos.

Ampliación

Se especificarán las características de los datos a cada uno de los tres niveles.

- **A nivel interno** (estructura interna), se ha de indicar el espacio de disco reservado para la base de datos, la longitud de los campos, su modo de representación (lenguaje para la definición de la estructura externa).
- **A nivel conceptual** (estructura conceptual), se proporcionan herramientas para la definición de las entidades y su identificación, atributos de las mismas, interrelaciones entre ellas, restricciones de integridad, etc.; es decir, el esquema de la base de datos (lenguaje para la definición de estructura lógico global).
- **A nivel externo** (estructura externa), se deben definir las vistas de los distintos usuarios a través del lenguaje para la definición de estructuras externas. Además, el SGBD se ocupará de la transformación de las estructuras externas orientadas a los usuarios a las estructuras conceptuales y de la relación de ésta y la estructura física.

2. **Función de manipulación:** permite a los usuarios de la base buscar, añadir, suprimir o modificar los datos de la misma, siempre de acuerdo con las especificaciones y las normas de seguridad dictadas por el administrador. Se llevará a cabo por medio de un **lenguaje de manipulación de datos (DML)** que facilita los instrumentos necesarios para la realización de estas tareas. También se encarga de definir la **vista externa** de todos los usuarios de la base de datos o vistas parciales que cada usuario tiene de los datos definidos con el DDL. Por manipulación de datos entenderemos:
 - La recuperación de información almacenada en la base de datos, lo que se conoce como **consultas**.
 - La **inserción** de información nueva en la base de datos.
 - El **borrado** de información de la base de datos.
 - La **modificación** de información almacenada en la base de datos.
3. **Función de control:** permite al administrador de la base de datos establecer mecanismos de protección de las diferentes visiones de los datos asociadas a cada usuario, proporcionando elementos de creación y modificación de dichos usuarios. Adicionalmente, incorpora sistemas para la creación de copias de seguridad, carga de ficheros, auditoría, protección de ataques, configuración del sistema, etc. El lenguaje que implementa esta función es el **lenguaje de control de datos o DCL**.

¿Y a través de qué lenguaje podremos desarrollar estas funciones sobre la base de datos? Lo haremos utilizando el **Lenguaje Estructurado de Consultas (SQL: Structured Query Language)**. Este lenguaje proporciona sentencias para realizar operaciones de DDL, DML y DCL. SQL fue publicado por el ANSI en 1986 (American National Standard Institute) y ha ido evolucionando a lo largo del tiempo. Además, los SGBD suelen proporcionar otras herramientas que complementan a estos lenguajes como generadores de formularios, informes, interfaces gráficas, generadores de aplicaciones, etc.

8.2 Ampliación: Componentes

8.3 Ampliación: Arquitectura

8.4 Tipos

¿Qué tipos de SGBD existen? Para responder a esta pregunta podemos realizar la siguiente clasificación,

atendiendo a diferentes criterios:

- a. El primer criterio que se suele utilizar es **por el modelo lógico en que se basan**. Actualmente, el modelo lógico que más se utiliza es el **relacional**. Los modelos en red y jerárquico han quedado obsoletos. Otro de los modelos que más extensión está teniendo es el modelo **orientado a objetos**. Por tanto, en esta primera clasificación tendremos:

- ✓ Modelo Jerárquico.
- ✓ Modelo de Red.
- ✓ Modelo Relacional.
- ✓ Modelo Orientado a Objetos.

(Para recordar los modelos de bases de datos vistos, sitúate en el epígrafe 4 de esta Unidad de Trabajo y analiza su contenido.)

- b. El segundo criterio de clasificación se centra en el **número de usuarios** a los que da servicio el sistema:

- ✓ **Monousuario**: sólo atienden a un usuario a la vez, y su principal uso se da en los ordenadores personales.
- ✓ **Multiusuario**: entre los que se encuentran la mayor parte de los SGBD, atienden a varios usuarios al mismo tiempo.

- c. El tercer criterio se basa en el **número de sitios en los que está distribuida la base de datos**:

- ✓ **Centralizados**: sus datos se almacenan en un solo computador. Los SGBD centralizados pueden atender a varios usuarios, pero el SGBD y la base de datos en sí residen por completo en una sola máquina.
- ✓ **Distribuidos (Homogéneos, Heterogéneos)**: la base de datos real y el propio software del SGBD pueden estar distribuidos en varios sitios conectados por una red. Los sistemas homogéneos utilizan el mismo SGBD en múltiples sitios. Una tendencia reciente consiste en crear software para tener acceso a varias bases de datos autónomas preexistentes almacenadas en sistemas distribuidos heterogéneos. Esto da lugar a los SGBD federados o sistemas multibase de datos en los que los SGBD participantes tienen cierto grado de autonomía local.

- d. El cuarto, y último, criterio establece su clasificación **según el propósito**:

- ✓ **Propósito General**: pueden ser utilizados para el tratamiento de cualquier tipo de base de datos y aplicación.
- ✓ **Propósito Específico**: Cuando el rendimiento es fundamental, se puede diseñar y construir un software de propósito especial para una aplicación específica, y este sistema no sirve para otras aplicaciones. Muchos sistemas de reservas de líneas aéreas son de propósito especial y pertenecen a la categoría de **sistemas de procesamiento de transacciones en línea**, que deben atender un gran número de transacciones concurrentes sin imponer excesivos retrasos.

8.5 Ampliación: Funcionamiento

9 Ampliación SGBD comerciales

10 Ampliación SGBD libres

11 Ampliación: Bases de datos centralizadas

12 Ampliación: Bases de datos distribuidas