

Unidad 1: Almacenamiento de la información.

Índice de contenidos

1	Introducción.....	2
2	Los ficheros de información.....	2
2.1	¿Qué es un fichero de datos?	2
2.2	Tipos de ficheros.....	3
2.3	Los soportes de información.....	4
2.4	Parámetros de utilización.....	5
2.5	Organización de ficheros.....	5
2.5.1	Organización secuencial.....	5
2.5.2	Organización secuencial encadenada.....	6
2.5.3	Organización secuencial indexada.....	6
2.5.4	Organización directa o aleatoria.....	7
2.6	Inconvenientes de un sistema de gestión de archivos.....	9
3	Bases de datos.....	10
3.1	¿Por qué son necesarias las BD?.....	10
3.2	Conceptos.....	11
3.3	Usos.....	12
3.4	Ubicación de la información.....	13
4	Evolución de los modelos de BD.....	14
4.1	Modelo jerárquico.....	14
4.2	Modelo en red.....	14
4.3	Modelo relacional.....	15
4.4	Modelo orientado a objetos.....	15
4.5	Modelo orientado a documentos.....	16
4.6	Otros modelos.....	16
4.6.1	Modelo Objeto-Relacional.....	16
4.6.2	Modelo de BD deductivas o lógicas.....	17
4.6.3	BD multidimensionales.....	17
4.6.4	BD transaccionales.....	17
4.7	Tabla comparativa de los principales modelos.....	18
5	Sistemas Gestores de Bases de Datos (SGBD)	18
5.1	Funciones.....	19
5.2	Componentes.....	20
5.3	Arquitectura.....	22
5.4	Tipos.....	23
6	SGBD comerciales y libres.....	24
7	BD centralizadas.....	25
8	BD distribuidas.....	26
8.1	Fragmentación.....	27

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 on-line, 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 on-line, 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 (BD) 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 (SGBD)** y finalmente, identificaremos las herramientas reales con las que llevar a cabo la gestión dichas BD.

2 Los ficheros de información.

2.1 ¿Qué es un fichero de datos?



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 hablando 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 de forma que se permita la búsqueda de datos individuales. Son identificados por un nombre y la descripción de la carpeta o directorio que lo contiene.

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 por el Sistema Operativo 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 bloaje, y a esta **operación de agrupar varios registros en un bloque** se le llama bloqueo de registros.

El Sistema Operativo (SO) realiza, además, la transformación de la dirección lógica (posición relativa que ocupa el registro en el fichero) usada en los programas de usuario en la dirección física (posición real o efectiva donde se encuentra el registro en el soporte de información) con la que se direcciona en el soporte.

Una clave o identificativo es un **campo (o conjunto de campos)** que **identifica cada registro del fichero**. En un mismo fichero puede haber una, varias o ninguna clave. Cuando una **clave** se **utiliza como campo de localización en el fichero** se denomina llave (key).

Los **ficheros** pueden estar **formados por registros** de uno de los siguientes tipos:

- ✓ **Longitud fija.**
- ✓ **Longitud variable:** el sistema reserva una palabra al comienzo de cada registro para anotar su longitud.
- ✓ **Delimitados:** el sistema incluye un carácter especial al final de cada registro.
- ✓ **Indefinido:** en este caso el SO no realiza ninguna gestión sobre la longitud de los registros del fichero. Será el propio programa del usuario el que se encargue de localizar el principio y el final de cada registro.

2.2 Tipos de ficheros.

En una aplicación informática se pueden utilizar ficheros para realizar funciones diversas. Conocer la función que va a desempeñar un fichero concreto es fundamental a la hora de decidir cómo se va a organizar éste.

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 una aplicación** (los datos necesarios para el

funcionamiento de ésta). **Su vida es larga** y generalmente no puede generarse de una forma inmediata a partir de otros ficheros. Estos se subdividen en:

- ✓ **Ficheros maestros**: contienen el estado actual de los **datos que pueden modificarse desde la aplicación**. Es la **parte central de la aplicación**, su núcleo. Podría ser un archivo con los datos de los usuarios de una plataforma educativa, un archivo de clientes de un banco, etc.
- ✓ **Ficheros constantes**: son aquellos que **incluyen datos fijos para la aplicación**. No suelen ser modificados (o suele ser infrecuente) y se accede a ellos para realización de consultas. Podría ser un archivo con códigos postales, la ubicación de estantes en una biblioteca, una tabla de números primos, etc.
- ✓ **Ficheros históricos**: contienen **datos que fueron considerados como actuales en un periodo o situación anterior**. Se utilizan para la reconstrucción de situaciones (actual o anteriores). Podría ser un archivo con los usuarios que han sido dados de baja en la plataforma educativa.
- **Ficheros temporales**: se utilizan **para almacenar información útil para una parte de la aplicación**, no para toda ella. Son **generados a partir de datos de ficheros permanentes**. Tienen **un corto periodo de existencia**. Estos se subdividen en:
 - ✓ **Ficheros intermedios**: almacenan **resultados de una aplicación que serán utilizados por otra**, dentro de una misma tarea.
 - ✓ **Ficheros de maniobras**: almacenan **datos de una aplicación que no pueden ser mantenidos en memoria principal** por falta de espacio.
 - ✓ **Ficheros de resultados**: almacenan **datos que van a ser transferidos a un dispositivo de salida**.



Autoevaluación

Supongamos una aplicación informática para gestionar una biblioteca, existirá un fichero con el catálogo de libros disponibles, otro con las editoriales, otro con información sobre libros que se han quedado obsoletos, etc. ¿A cuál de los siguientes tipos correspondería el fichero que almacena las editoriales?

- Fichero maestro.
- Fichero constante.
- Fichero intermedio.

2.3 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. **El tamaño de los ficheros está limitado por el tamaño de los dispositivos que los albergan**.

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 secuencial o no direccionables** (Ej.: 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**.
- ✓ **Soportes de acceso directo o direccionables** (Ej.: 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 (no es necesario recorrer o leer otros bloques).

2.4 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:

- a) **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.
- b) **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.
- c) **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.
- d) **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.



Autoevaluación

La volatilidad de un fichero es un parámetro que indica:

- La variación del volumen del fichero.
- La cantidad de veces que se abre o cierra el fichero.
- El peso de los procesos de inserción y borrado en dicho fichero (frecuencia de renovación).

2.5 Organización de ficheros.

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 cambiado 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.

Hay **distintas formas de estructurar u organizar los ficheros sobre un soporte de información**. Las características de utilización del fichero dependen de la organización que se adopte. Se debe, pues, **optar por una u otra organización, atendiendo a la forma en que se va a usar el fichero**. Las principales organizaciones de ficheros se detallan en los apartados siguientes.

2.5.1 Organización secuencial.

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.

En la imagen que se muestra en el lateral se puede ver la estructura de un fichero secuencial.

Registro 1
Registro 2
Registro i-1
Registro i-2
Registro n-1
Registro n

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.

No es posible realizar fácilmente las operaciones de inserción, modificación o borrado sobre un fichero secuencial. Si se necesita actualizar un archivo con organización secuencial se debe crear de nuevo el archivo (esto se realiza mediante un programa escrito para tal fin).

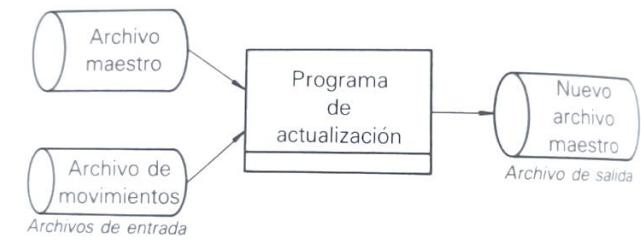
Tan sólo es posible añadir un nuevo registro de forma fácil al final del fichero.

La consulta se realiza en orden secuencial.

2.5.2 Organización secuencial encadenada.

En un fichero con organización secuencial encadenada, **junto a cada registro se almacena un puntero (enlace) con la dirección del registro siguiente, según el orden lógico del fichero**. Los ficheros con esta organización sólo pueden ser gestionados en soportes direccionables.

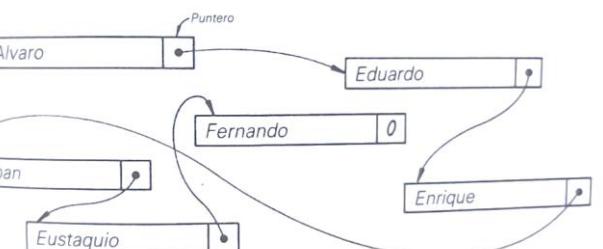
La **consulta** es secuencial, cada vez que se lee un registro se lee la posición del siguiente, lo que permite seguir la secuencia lógica del archivo.



Para **insertar un registro** es necesario, en primer lugar, **localizar la posición en que se desea insertar** (entre qué dos registros se quiere que aparezca). El **registro se crea en una zona libre y se modifican los punteros** (enlaces) para que el registro quede insertado entre los dos registros deseados.

Para **borrar un registro** se asigna al puntero del registro anterior la dirección del registro siguiente al que se desea borrar. El Sistema Operativo puede o no liberar el espacio ocupado por el registro borrado.

Al **modificar**, si ésta **no implica un aumento de longitud del registro**, éste **puede reescribirse en el mismo espacio**. En el caso de que el registro aumente de longitud se debe **insertar el registro y posteriormente borrar la versión anterior a la modificación**.



La principal ventaja de esta organización es su flexibilidad, y su principal inconveniente es su limitación a consulta secuencial.

2.5.3 Organización secuencial indexada.

Un archivo con organización secuencial indexada está **formado por dos estructuras o zonas distintas**: zona de registros y zona de índices.

En la zona de registros se encuentran los registros ordenados según el valor de una clave. En esta zona se pueden direccionar los registros.

La zona de índices se procesa sólo de forma secuencial. Su estructura es la de un archivo secuencial puro en que cada registro contiene dos campos: clave y dirección de un registro.

La zona de registros está dividida en tramos lógicos. Cada tramo está formado por una serie de registros consecutivos. Por cada tramo de la zona de registros hay un registro en la zona de índices. En este registro, se encuentra el valor mayor de clave del tramo.

Para permitir actualizaciones es necesario incluir en la estructura una zona de desbordamientos, en esta zona los registros están desordenados y cada nuevo registro se escribe al final en esta zona.

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 más espacio en el disco que los ficheros secuenciales:** debido al uso del área de índices.
- ✓ **Solo se pueden utilizar soportes direccionables.**
- ✓ Obliga a una inversión económica mayor, por la necesidad de programas y, a veces, de hardware más sofisticado.

2.5.4 Organización directa o aleatoria.

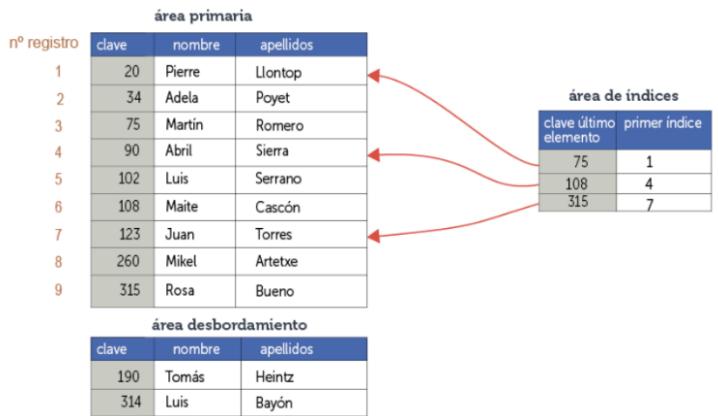
En este tipo de ficheros se puede acceder a un registro indicando la posición relativa del mismo dentro del archivo o, más comúnmente, a través de una clave que forma parte del registro como un campo más. Estos archivos deben almacenarse en dispositivos de memoria masiva de acceso directo, como son los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

Algunas 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**.
- ✓ Esta organización sólo es posible en soportes **direccionables**.



- ✓ 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**.

A través de una transformación específica aplicada a la clave, se obtendrá la dirección física en la que se encuentra el registro. El problema fundamental de esta organización es la elección de la transformación que se ha de usar, y en algunos casos aparecen dos situaciones no deseadas: **direcciones que no corresponden a ninguna clave** (zonas de disco sin utilizar) y **direcciones que corresponden a más de una clave** (da lugar a lo que se conoce como sinónimos).

Hay **dos formas de resolver el problema de los sinónimos**, siempre a costa de complicar la estructura del fichero:

1. Cuando se asocia a una clave una dirección ya ocupada por un registro distinto, **se busca en el archivo, por algún procedimiento, hasta encontrar una posición libre donde escribir el registro**.
2. **Se reserva una zona de desbordamiento donde se escribirán los registros que no se puedan escribir en la posición que les corresponde** según la transformación.

Según la forma de realizar esta transformación, existen diferentes **modos de acceso**:

2.5.4.1 Direccionamiento directo.

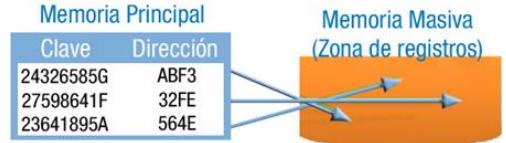
Es **factible cuando la clave es numérica y su rango de valores no es mayor que el rango de direcciones en un archivo**. **Se utiliza como dirección la propia clave**.

En algunos casos **pueden quedar lagunas de direcciones sin utilizar**, en lugares conocidos de antemano.

Un archivo aleatorio con direccionamiento directo **siempre está ordenado**.

2.5.4.2 Direccionamiento asociado.

Se puede utilizar para cualquier tipo de clave. Si se utiliza este método **debe construirse una tabla en la que figurarán todas las claves y la dirección donde se encuentra el registro correspondiente**.



Al añadir nuevos registros las claves se colocan al final de la tabla, normalmente desordenada. Por tanto, habrá que localizar las claves en ella por lectura secuencial, lo que ralentiza el acceso, a menos que la tabla esté residiendo en memoria principal o se ordene.

2.5.4.3 Direccionamiento calculado (hashing).

Cuando utilizamos ficheros con direccionamiento asociado 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.
- ✓ **Elevación al cuadrado**: se eleva al cuadrado la clave y se toman los dígitos centrales.
- ✓ **Plegamiento**: se descomponen la clave en segmentos de cifras contiguas del mismo tamaño, y se suman.

Una **buenas** 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.



Autoevaluación

En los ficheros de acceso directo los registros siempre se encuentran en posiciones contiguas dentro del soporte de almacenamiento.

Verdadero

Falso



Autoevaluación

En un fichero con acceso calculado:

- Se utiliza la dirección como clave.
- Hay una tabla en la que está cada clave con la dirección del registro correspondiente.
- La dirección se obtiene a partir de la clave mediante un algoritmo.

2.6 Inconvenientes de un sistema de gestión de archivos.

Antes de aparecer los SGBD, como hemos visto, la información se trataba y se gestionaba utilizando los típicos sistemas de gestión de archivos. Los datos se guardaban en archivos y los programas manejaban esos archivos para obtener la información. Si la estructura de los datos de los archivos cambia, todos los programas que los manejan se deben modificar. En estos sistemas de gestión de archivos, la definición de los datos se encuentra codificada dentro de los programas en lugar de almacenarse de forma independiente, y además el control del acceso y la manipulación de los datos viene impuesto por los programas.

Esto supone un gran inconveniente a la hora de tratar grandes volúmenes de información. Surge así la idea de separar los datos contenidos en los archivos de los programas que los manipulan (es decir, que se pueda modificar la estructura de los datos de los archivos sin que por ello se tenga que modificar los programas con los que trabajan).

Inconvenientes de un sistema de gestión de archivos:

- ✓ **Redundancia e inconsistencia de los datos:** los archivos son creados por distintos programas y van cambiando con el tiempo, por lo que pueden tener distintos formatos y los datos pueden estar duplicados.
- ✓ **Dependencia de los datos física-lógica:** la estructura física de los datos (definición de archivos y registros) se encuentra codificada en los programas.
- ✓ **Dificultad para tener acceso a los datos:** cada vez que se necesite una consulta no prevista es necesario codificar el programa correspondiente.
- ✓ **Separación y aislamiento de los datos:** al estar repartidos en varios archivos con diferentes formatos, es difícil escribir nuevos programas que aseguren la manipulación correcta de los datos.
- ✓ **Dificultad para el acceso concurrente:** es complicado que los usuarios actualicen los datos simultáneamente.
- ✓ **Dependencia de la estructura del archivo con el lenguaje de programación:** la estructura se define dentro de los programas, la incompatibilidad entre archivos generados por distintos lenguajes hace que los datos sean difíciles de procesar.
- ✓ **Problemas en la seguridad de los datos:** difícil implantar restricciones de seguridad pues las aplicaciones se van añadiendo al sistema según se van necesitando.
- ✓ **Problemas de integridad de datos:** los valores almacenados en los archivos deben cumplir con restricciones de consistencia (no se puede crear un pedido si el cliente no existe).

Todos estos inconvenientes hacen posible el fomento y desarrollo de las BD y los SGBD.

3 Bases de datos.

Como hemos visto anteriormente, los ficheros permiten organizar y almacenar 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, 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.

Las bases de datos son instrumentos de gran utilidad para gestionar grandes ficheros y facilitar la consulta de información. En muchas, además, puede definirse un esquema de permisos que establece qué personas o programas pueden acceder a los datos, y a cuáles, con el objetivo de presentar el contenido de forma adecuada y clara.

Los distintos sistemas de BD se diferencian conceptualmente entre sí y tienen, por lo tanto, sus propias ventajas y desventajas. Pero, antes que nada, es conveniente diferenciar entre la base de datos en sí y el sistema que la gestiona. Como BD se designa al **conjunto de los datos que se ha de almacenar**, mientras que el SGBD es **responsable de su administración**, determinando así su estructura, el orden, los permisos de acceso, las dependencias, etc.

La gestión de las BD 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 BD creció exponencialmente, y aunque muchos de ellos no sean conscientes de ello, el acceso a dichas BD forma parte de la vida cotidiana de muchos de nosotros.

Conocer los sistemas que gestionan las BD, 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 ¿Por qué son necesarias las BD?

Para aumentar la eficiencia estructural del tratamiento electrónico de los datos, ya en la década de los 60, se empezó a desarrollar el concepto de la BD electrónica como capa separada de software entre el SO y el programa de aplicación. Esto fue el resultado de la experiencia del día a día, pues tanto manipular los archivos como supervisar y repartir los permisos adquirió tal complejidad que el procesamiento electrónico de los datos no significó un avance real. Así, la idea del sistema de BD electrónico se convirtió en una de las innovaciones más relevantes en el desarrollo del ordenador.

Los primeros modelos que se desarrollaron fueron las BD en red y jerárquicas, si bien pronto demostraron ser demasiado simples y estar limitadas técnicamente. IBM fue la empresa que revolucionó el sector, con el desarrollo del modelo relacional de BD en los años setenta, con mucho el más potente, que pronto encontró un campo de cultivo favorable en el mundo laboral. Los productos que más éxito tuvieron en este momento, fueron el lenguaje de consultas a BD SQL de Oracle y los sucesores de IBM, SQL/DS y DB2.

Hasta bien entrados los años 2000, cuando algunos proyectos de código libre metieron algo de aire fresco al sector, el mercado del software de BD estuvo gobernado por los pesos pesados. Entre los sistemas libres más populares se encuentran MySQL y PostgreSQL. La tendencia iniciada en 2001 hacia los sistemas NoSQL también contribuyó a la debilitación de la posición de los sistemas de BD de los grandes fabricantes.

Hoy, los sistemas de BD son imprescindibles en numerosos campos. **Cualquier tipo de software concebido para las empresas se basa en robustas BD** con un gran número de opciones y herramientas para los administradores del sistema. La seguridad de los datos, además, ha ido ganando importancia con el tiempo, y es que en las BD electrónicas se almacenan y cifran contraseñas, datos personales e incluso divisa digital.

El sistema financiero moderno, no es más que una red de BD, en la cual, la mayor parte de las cuantías monetarias solo existen como unidades electrónicas de información, cuya protección, por medio de BD seguras es una de las tareas principales de las instituciones financieras. Aunque no solo por esto son cruciales las BD electrónicas para la civilización moderna.

3.2 Conceptos.

A finales de los setenta, la aparición de nuevas tecnologías de manejo de datos a través de los sistemas de BD supuso un considerable cambio. Los sistemas basados en ficheros separados dieron paso a la utilización de **SGBD**, 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**.

Base de datos (BD): 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 BD 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.

Una BD constará de los siguientes elementos:

- ✓ **Entidades:** objeto real o abstracto con características diferenciadoras de otros, del que se almacena información en la BD. En una BD de una clínica veterinaria, posibles entidades podrían ser: ejemplar, doctor, consulta, etc.
- ✓ **Atributos:** son los **datos que se almacenan de la entidad**. Cualquier propiedad o característica de una entidad puede ser atributo. Continuando con nuestro ejemplo, podrían ser atributos: raza, color, nombre, número de identificación, etc.
- ✓ **Registros:** donde se almacena la información de cada entidad. Es un **conjunto de atributos que contienen los datos que pertenecen a una misma repetición de entidad**. En nuestro ejemplo, un registro podría ser: 2123056, Sultán, Podenco, Gris, 23/03/2009.
- ✓ **Campos:** donde se **almacenan los atributos de cada registro**. Teniendo en cuenta el ejemplo anterior, un campo podría ser el valor Podenco.

Las ventajas fundamentales que ofrece el uso de BD se resumen a continuación:

- ✓ **Acceso múltiple:** diversos usuarios o aplicaciones podrán acceder a la BD, sin que existan problemas en el acceso o los datos.
- ✓ **Utilización múltiple:** cada uno de los **usuarios o aplicaciones podrán disponer de una visión particular de la estructura de la BD**, de tal manera que cada uno de ellos accederá sólo a la parte que realmente le corresponde.
- ✓ **Flexibilidad:** la forma de acceder a la información puede ser establecida de diferentes maneras, ofreciendo tiempos de respuesta muy reducidos.
- ✓ **Confidencialidad y seguridad:** el control del acceso a los datos podrá ser establecido para que unos usuarios o aplicaciones puedan acceder a unos datos y a otros no, impidiendo a los usuarios no autorizados la utilización de la BD.
- ✓ **Protección contra fallos:** en caso de errores en la información, existen mecanismos bien definidos que permiten la recuperación de los datos de forma fiable.
- ✓ **Independencia física:** un cambio de soporte físico de los datos (por ejemplo: el tipo de discos), no afectaría a la BD o a las aplicaciones que acceden a ellos.
- ✓ **Independencia lógica:** los cambios en la BD no afectan a las aplicaciones que la usan.
- ✓ **Redundancia:** los datos se almacenan, por lo general, una única vez. Aunque si es necesario, podríamos repetir información de manera controlada.

- ✓ **Interfaz de alto nivel:** mediante la utilización de lenguajes de alto nivel puede utilizarse la BD de manera sencilla y cómoda.
- ✓ **Consulta directa:** existe una herramienta para poder acceder a los datos interactivamente.



Autoevaluación

Una BD es:

- Un programa para gestionar archivos muy grandes.
- El conjunto de datos de los usuarios almacenados en un único disco duro.
- Conjunto de datos de distinto tipo relacionados entre sí, junto con un programa de gestión de dichos datos.

3.3 Usos.

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

¿Quién utiliza las BD?

Tipo	Funciones y características
El administrador	Es la persona encargada de la creación o implementación física de la BD . Es quien escoge los tipos de ficheros, los índices que hay que crear, la ubicación de éstos, etc. 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. Junto a estas tareas, el administrador establecerá la política de seguridad y de acceso para garantizar el menor número de problemas.
Los diseñadores	Son las personas encargadas de diseñar cómo será la BD . Llevarán a cabo la identificación de los datos, las relaciones entre ellos, sus restricciones, etc. Para ello han de conocer a fondo los datos y procesos a representar en la BD. 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 BD debe implicar en el proceso a todos los usuarios de la BD, tan pronto como sea posible.
Los programadores de aplicaciones	Una vez diseñada y construida la BD, 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.
Los usuarios finales	Son los clientes finales de la BD . Al diseñar, implementar y mantener la BD se busca cumplir los requisitos establecidos por el cliente para la gestión de su información.

¿Para qué se utilizan las BD?

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**.

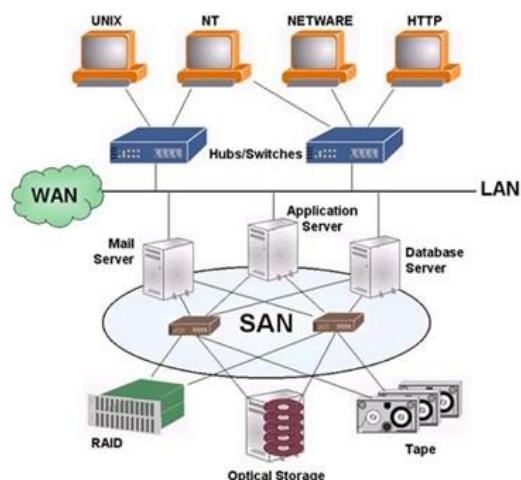
- ✓ **Banca:** información de clientes, cuentas, transacciones, préstamos, etc.
- ✓ **Líneas aéreas:** información de clientes, horarios, vuelos, destinos, etc.
- ✓ **Universidades:** información de estudiantes, carreras, horarios, materias, etc.
- ✓ **Transacciones de tarjeta de crédito:** para comprar 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 y Seguridad:** delincuentes, 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.

3.4 Ubicación de la información.

Utilizamos a diario las BD, pero ¿Dónde se encuentra realmente almacenada la información? Las BD 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 BD 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 BD puede necesitar servidores en lugares diferentes, y viceversa, **pequeñas BD 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 BD**, comenzaremos por aquellos en los que pueden alojarse BD 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 especifica 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, se establecen distintos sistemas de RAID.
- ✓ **Sistemas NAS:** es el acrónimo de Network Attached Storage o **sistema de almacenamiento masivo en red**. Estos sistemas de almacenamiento **permiten compartir la capacidad de almacenamiento de un ordenador (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, se optimiza el poder de procesamiento para aplicaciones comerciales y la capacidad de almacenamiento se puede proporcionar en el servidor donde más se necesite.



Autoevaluación

Rellena los huecos con los conceptos adecuados.

Un tipo de red donde se optimiza el poder de procesamiento para aplicaciones comerciales, pudiendo proporcionarse la capacidad de almacenamiento en el servidor donde más se necesite, se denomina sistema

4 Evolución de los modelos de BD.

Las diferencias entre los modelos de BD más habituales es resultado de la evolución técnica de la transmisión electrónica de datos, que no solo perseguía la eficiencia y la manejabilidad, sino también, el empoderamiento de los fabricantes más renombrados.

Modelar: definir un mundo abstracto y teórico tal que las conclusiones que se pueden sacar de él coinciden con las manifestaciones aparentes del mundo real.

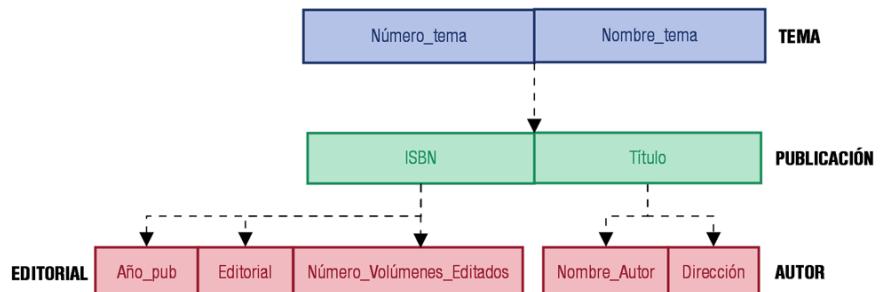
Modelo: instrumento que se aplica a una parcela del mundo real para obtener una estructura de datos a la que denominamos esquema. El modelo será un conjunto de conceptos, reglas y convenciones que nos permiten describir los datos del mundo real.

Modelo de datos: conjunto de conceptos o herramientas conceptuales que sirven para describir la estructura de una BD (los datos, las relaciones y las restricciones que se deben cumplir sobre los datos).

Esquema de la BD: descripción de una BD mediante un modelo de datos.

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.

Los datos de este modelo se almacenan en estructuras lógicas llamadas **segmentos**. Los segmentos se relacionan entre sí utilizando **arcos**. La forma visual de este modelo es de árbol invertido, en la parte superior están los padres y en la inferior los hijos.

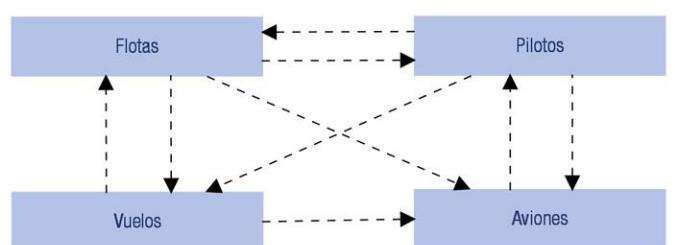
Hoy en día, debido a sus limitaciones, el **modelo jerárquico está en desuso**.

4.2 Modelo en red.

El modelo de datos en red **aparece** a mediados de los sesenta como respuesta a las 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 BD 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



almacenar los datos, mientras que los enlaces permiten relacionar estos datos. Las BD en red son parecidas a las jerárquicas sólo que en ellas puede haber más de un parente.

En este modelo se puede 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.

4.3 Modelo relacional.

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

En el modelo relacional la BD 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 (relaciones) para la representación lógica de los datos y las relaciones entre ellos. Cada relación (tabla) posee un nombre que es único y contiene un conjunto de columnas.

Se llamará registro, entidad 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 son del mismo tipo.
- ✓ La tabla sólo puede tener un tipo de registro.
- ✓ No existen campos o atributos repetidos.
- ✓ No existen registros duplicados.
- ✓ No existe orden en el almacenamiento de los registros.

El lenguaje habitual para construir las consultas a BD 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 BD relacional pasa por un proceso al que se conoce como normalización de una BD.

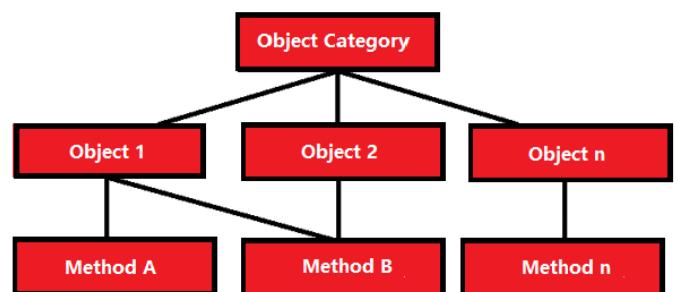
4.4 Modelo orientado a objetos.

El modelo orientado a objetos define una BD en términos de **objetos, sus propiedades y sus operaciones**. Los **objetos con la misma estructura y comportamiento pertenecen a una clase, y las clases se organizan en jerarquías**. Las operaciones de cada clase se especifican en términos de procedimientos predefinidos denominados métodos. Algunos sistemas existentes en el mercado, basados en el modelo relacional, han sufrido evoluciones incorporando conceptos orientados a objetos. A estos modelos se les conoce como sistemas objeto-relacionales.

El **objetivo** del modelo orientado a objetos es **cubrir las limitaciones del modelo relacional**. Gracias a este modelo se incorporan mejoras como la herencia entre tablas, los tipos definidos por el usuario, disparadores almacenables en la base de datos (triggers), soporte multimedia, etc.

Los conceptos más importantes del paradigma de objetos que el modelo orientado a objetos incorpora son:

- ✓ **Encapsulación:** propiedad que permite **ocultar la información al resto de los objetos**, impidiendo así accesos incorrectos o conflictos.



- ✓ **Herencia:** propiedad a través de la cual los **objetos heredan comportamiento** dentro de una jerarquía de clases.
- ✓ **Polimorfismo:** propiedad de una operación mediante la cual puede ser aplicada a distintos tipos de objetos.

Desde la aparición de la Programación Orientada a Objetos (POO u OOP) se empezó a pensar en BD adaptadas a estos lenguajes. Este modelo es considerado como el fundamento de las BD de tercera generación, siendo consideradas las BD en red como la primera y las BD relacionales como la segunda generación. Aunque **no han reemplazado a las BD relacionales**, si son el **tipo de BD que más está creciendo en los últimos años**.

4.5 Modelo orientado a documentos.

En este modelo, **los documentos son la unidad básica para el almacenamiento de datos**. Estas unidades son las que estructuran los datos y no deben confundirse con los documentos de los programas de procesamiento de texto. Aquí, **los datos se guardan en los llamados pares clave-valor**, comprendiendo así, una “clave” y un “valor”. Como **no están definidos ni la estructura ni el número de pares**, los documentos que integran una BD orientada a documentos pueden resultar muy dispares entre sí. Cada documento es una unidad cerrada en sí misma y establecer relaciones entre documentos no resulta fácil, pero en este modelo no es necesario.

Key	First Name	Surname	Zipcode ID
1	Walker	McClain	1
2	Blain	Muller	2
3	Jack	Schmidt	2
4	Greg	Cohn	4

Zipcode ID	City	State	Zipcode
1	Bend	Oregon	97701
2	Niwot	Colorado	80503
3	Spartenburg	South Carolina	29301
4	Syracuse	New York	13201


```
{
  "ID":3,
  "First Name": "Jack",
  "Surname": "Schmidt",
  "Zipcode": "80503",
  "City": "Niwot",
  "State": "Colorado"
}
```

JSON

En el modelo relacional (arriba representado con las dos tablas), varias relaciones (tablas) se conectan entre sí para seleccionar un registro común. En el modelo documental, un único documento basta para guardar toda la información. Aquí no se está obligado a utilizar un determinado esquema porque, mientras se use siempre el mismo lenguaje de BD, **este modelo está conceptualmente libre de esquemas**.

Una idea fundamental de las BD documentales es que **los datos que guardan relación entre sí siempre se guardan juntos en un lugar (en el documento)**. Mientras que las BD relacionales suelen representar y mostrar la información relacionada conectando varias tablas, en el modelo que nos ocupa es suficiente con consultar un solo documento. Esto reduce el número de procedimientos necesarios para consultar la BD.

Estos sistemas son especialmente interesantes para las aplicaciones web, puesto que permiten guardar formularios HTML completos. Fue sobre todo con el avance de la web 2.0 cuando estas BD vieron aumentar su popularidad. Con todo, es necesario remarcar que **entre los diversos sistemas basados en documentos se dan diferencias notables, desde la sintaxis hasta la estructura interna, por lo que no todas las BD orientadas a documentos son apropiadas para cualquier escenario**. Es debido a estas diferencias por lo que hoy disponemos de algunos sistemas de BD orientados a documentos de la reputación de Lotus Notes, Amazon SimpleDB, MongoDB, CouchDB, Riak, ThruDB, OrientDB, etc.

4.6 Otros modelos.

Además de los modelos vistos hasta el momento, vamos a detallar a continuación las particularidades de otros modelos de BD existentes y que, en algunos casos, son una evolución de los anteriores.

4.6.1 Modelo Objeto-Relacional.

Las BD pertenecientes a este modelo, **son un híbrido entre las BD del modelo relacional y el orientado a objetos**. El mayor inconveniente de las BD orientadas a objetos radica en los costes de la conversión de las BD relacionales a BD orientadas a objetos.

En una BD objeto-relacional (BDOR) siempre se busca obtener lo mejor del modelo relacional, incorporando las mejoras ofrecidas por la orientación a objetos. En este modelo se siguen almacenando tuplas, aunque la estructura de las tuplas no está restringida, sino que las relaciones pueden ser definidas en función de otras, que es lo que denominamos herencia directa.

El estándar en el que se basa este modelo es SQL99. Este estándar ofrece la posibilidad de añadir a las BD relacionales procedimientos almacenados de usuario, triggers, tipos definidos por el usuario, consultas recursivas, bases de datos OLAP, tipos LOB, ...

Otra característica a destacar es la capacidad para incorporar funciones que tengan un código en algún lenguaje de programación como, por ejemplo: SQL, Java, C, etc.

La gran mayoría de las BD relacionales clásicas de gran tamaño, como Oracle, SQL Server, etc., son objeto-relacionales.

4.6.2 Modelo de BD deductivas o lógicas.

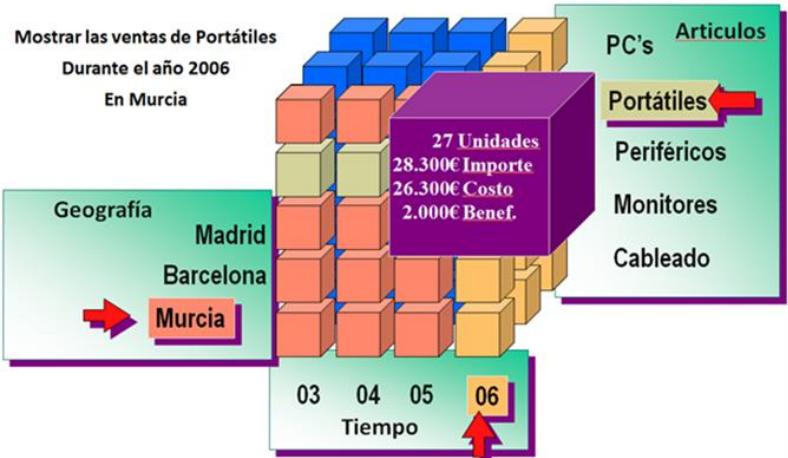
En este modelo las BD almacenan la información y permiten realizar deducciones a través de inferencias. Es decir, se derivan nuevas informaciones a partir de las que se han introducido explícitamente en la BD por parte del usuario.



Las BD deductivas son también llamadas BD lógicas, al basarse en lógica matemática. Surgieron para contrarrestar las limitaciones del modelo relacional para la respuesta a consultas recursivas y la deducción de relaciones indirectas entre los datos almacenados.

4.6.3 BD multidimensionales.

Son BD ideadas para desarrollar aplicaciones muy concretas. Básicamente almacena sus datos con varias dimensiones, es decir que, en vez de un valor, encontramos varios dependiendo de los ejes definidos o una BD de estructura basada en dimensiones orientada a consultas complejas y alto rendimiento. En una BD multidimensional, la información se representa como matrices multidimensionales, cuadros de múltiples entradas o funciones de varias variables sobre conjuntos finitos. Cada una de estas matrices se denomina cubo. Eso facilita el manejo de grandes cantidades de datos dentro de empresas, dándole a esto una amplia aplicación dentro de varias áreas y diferentes campos del conocimiento humano.



4.6.4 BD transaccionales.

Son BD caracterizadas por su velocidad para gestionar el intercambio de información, se utilizan sobre todo en sistemas bancarios, análisis de calidad y datos de producción industrial. Son bases de datos muy fiables, ya que en ellas cada una de las operaciones de inserción, actualización o borrado se realizan completamente o se descartan.

4.7 Tabla comparativa de los principales modelos.

Modelo de BD	Desarrollo	Ventajas	Inconvenientes	Ámbitos de aplicación	Marcas
Jerárquico	Década de 1960	Acceso de lectura muy rápido, estructura clara, técnicamente simple	Estructura fija en árbol que no permite conexiones entre árboles	Bancos, compañías de seguros, sistemas operativos	IMS/DB
En red	Principios década de 1970	Admite varias formas de acceder a un registro, sin jerarquía estricta	En BD más grandes no se tiene una vista general	Grandes ordenadores	UDS (Siemens), DMS (Sperry Univac)
Relacional	1970	Simple, flexible. Creación y edición fácil y flexible, fácil de ampliar, rápida puesta en marcha, contexto de competencia muy dinámica	Inmanejable con cantidades grandes de datos, segmentación deficiente, atributos de clave artificiales, interfaz de programación externa, no refleja bien las propiedades y la conducta de los objetos	Control de gestión, facturación, sistemas de control de inventario, sistemas de gestión de contenido, etc.	MySQL, PostgreSQL, Oracle, SQLite, DB2, Ingres, MariaDB, Microsoft Access
Orientado a objetos	Final de la década de 1980	Soporta mejor los lenguajes de programación orientados a objetos, permite almacenar contenido multimedia	El rendimiento empeora con grandes volúmenes de datos, pocas interfaces compatibles	Inventario (museos, comercio minorista)	db4o
Orientado a documentos	Década de 1980	Los datos relacionados se guardan de forma centralizada en documentos independientes, estructura libre, concepción multimedia	El trabajo de organización es relativamente alto, a menudo requiere conocimientos de programación	Aplicaciones web, buscadores, bases de datos de texto	Lotus Notes, Amazon SimpleDB, MongoDB, CouchDB, Riak, ThruDB, OrientDB

5 Sistemas Gestores de Bases de Datos (SGBD).

Para poder tratar la información contenida en las BD 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 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 BD**. 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 BD, manteniendo su integridad, confidencialidad y seguridad.

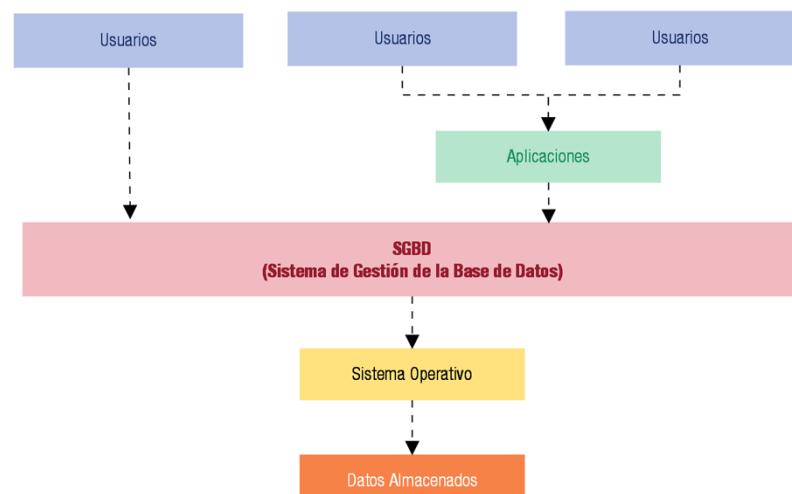
El SGBD permite a los usuarios la creación y el mantenimiento de una BD, facilitando la definición, construcción y manipulación de la información contenida en éstas. Definir una BD consistirá en especificar los tipos de datos, las estructuras y las restricciones que los datos han de cumplir a la hora de almacenarse en dicha BD. Por otro lado, la

construcción de la BD 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 BD 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 BD, 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 SO. Los datos almacenados de forma estructurada en la BD 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 SO.



5.1 Funciones.

El SGBD es una aplicación que permite a los usuarios definir, crear y mantener la BD y proporciona un acceso controlado a la misma. Debe prestar los siguientes servicios:

- ✓ **Creación y definición de la BD:** especificación de la estructura, el tipo de los datos, las restricciones y relaciones entre ellos mediante lenguajes de definición de datos (toda esta información se almacena en el diccionario de datos).
- ✓ **Manipulación de los datos:** realizando consultas, inserciones y actualizaciones utilizando lenguajes de manipulación de datos.
- ✓ **Acceso controlado a los datos de la BD:** mediante mecanismos de seguridad de acceso de usuarios.
- ✓ **Mantener la integridad y consistencia de los datos:** utilizando mecanismos para evitar que los datos sean perjudicados por cambios no autorizados.
- ✓ **Acceso compartido a la BD:** controlando la interacción entre usuarios concurrentes.
- ✓ **Mecanismos de respaldo y recuperación:** para restablecer la información en caso de fallos del sistema.

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 BD crear las estructuras apropiadas para integrar adecuadamente los datos. Esta función es la que permite definir las tres estructuras de la BD: **estructura interna, estructura conceptual y estructura externa**.

Esta función se realiza mediante el lenguaje de descripción 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.

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 BD, 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 BD (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 de datos 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 BD 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 BD, lo que se conoce como consultas.
- La inserción de información nueva en la BD.
- El borrado de información de la BD.
- La modificación de información almacenada en la BD.

3. **Función de control:** permite al administrador de la BD 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 BD? 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.



Autoevaluación

El DDL de una BD sirve para:

- La introducción de los datos en una BD.
- Definir la estructura lógica de la BD.
- Interrogar a la BD (consultar la información de dicha BD).

5.2 Componentes.

Una vez descritas las funciones que un SGBD debe llevar a cabo, imaginarás que **un SGBD es un paquete de software complejo que ha de proporcionar servicios relacionados con el almacenamiento y la explotación de los datos de forma eficiente**. Para ello, cuenta con una serie de **componentes** que se detallan a continuación:

1. **Lenguajes de la BD:** cualquier sistema gestor de BD ofrece la posibilidad de utilizar lenguajes e interfaces adecuadas para sus diferentes tipos de usuarios. A través de los lenguajes se pueden especificar los datos que

componen la BD, su estructura, relaciones, reglas de integridad, control de acceso, características físicas y vistas externas de los usuarios. Los lenguajes del SGBD son: **Lenguaje de Definición de los Datos (DDL)**, **Lenguaje de Manejo de Datos (DML)** y **Lenguaje de Control de Datos (DCL)**.

2. **El diccionario de datos:** descripción de los datos almacenados. Se trata de información útil para los programadores de aplicaciones. Es el lugar donde se deposita la información sobre la totalidad de los datos que forman la BD. Contiene las características lógicas de las estructuras que almacenan los datos, su nombre, descripción, contenido y organización. En una BD relacional, el diccionario de datos aportará información sobre:

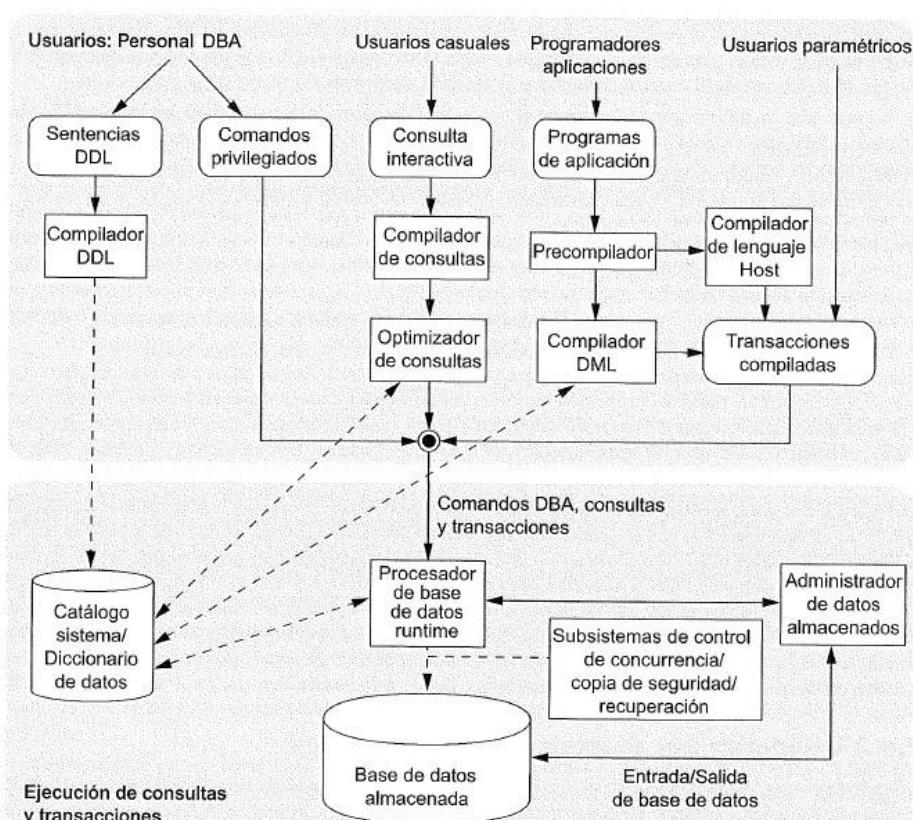
- ✓ Estructura lógica y física de la BD.
- ✓ Definición de tablas, vistas, índices, disparadores, procedimientos, funciones, etc.
- ✓ Cantidad de espacio asignado y utilizado por los elementos de la BD.
- ✓ Descripción de las restricciones de integridad.
- ✓ Información sobre los permisos asociados a cada perfil de usuario.
- ✓ Auditoría de acceso a los datos, utilización, etc.

3. **El gestor de la BD:** es la parte de software encargada de garantizar el correcto, seguro, íntegro y eficiente acceso y almacenamiento de los datos. Este componente es el encargado de proporcionar una interfaz entre los datos almacenados y los programas de aplicación que los manejan. Es un intermediario entre el usuario y los datos. Es el encargado de garantizar la privacidad, seguridad e integridad de los datos, controlando los accesos concurrentes e interactuando con el sistema operativo.

4. **Usuarios de la BD:** en los SGBD existen diferentes perfiles de usuario, cada uno de ellos con una serie de permisos sobre los objetos de la BD. Generalmente existirán:

- ✓ El **administrador de la BD** o **Database Administrator (DBA)**, que será la persona o conjunto de ellas encargadas de la función de administración de la BD. Tiene el control centralizado de la BD y es el responsable de su buen funcionamiento. Es el encargado de autorizar el acceso a la BD, de coordinar y vigilar su utilización y de adquirir los recursos software y hardware que sean necesarios.
- ✓ Los **usuarios de la BD**, que serán diferentes usuarios de la BD con diferentes necesidades sobre los datos, así como diferentes accesos y privilegios. Podemos establecer la siguiente clasificación:
 - Operadores y personal de mantenimiento.
 - Analistas y programadores de aplicaciones.
 - Usuarios finales: ocasionales, simples, avanzados y autónomos.
 - Diseñadores.

5. **Herramientas de la BD:** son un conjunto de aplicaciones que permiten a los administradores la gestión de la BD, de los usuarios y permisos, generadores de formularios, informes, interfaces gráficas, generadores de aplicaciones, etc.



5.3 Arquitectura.

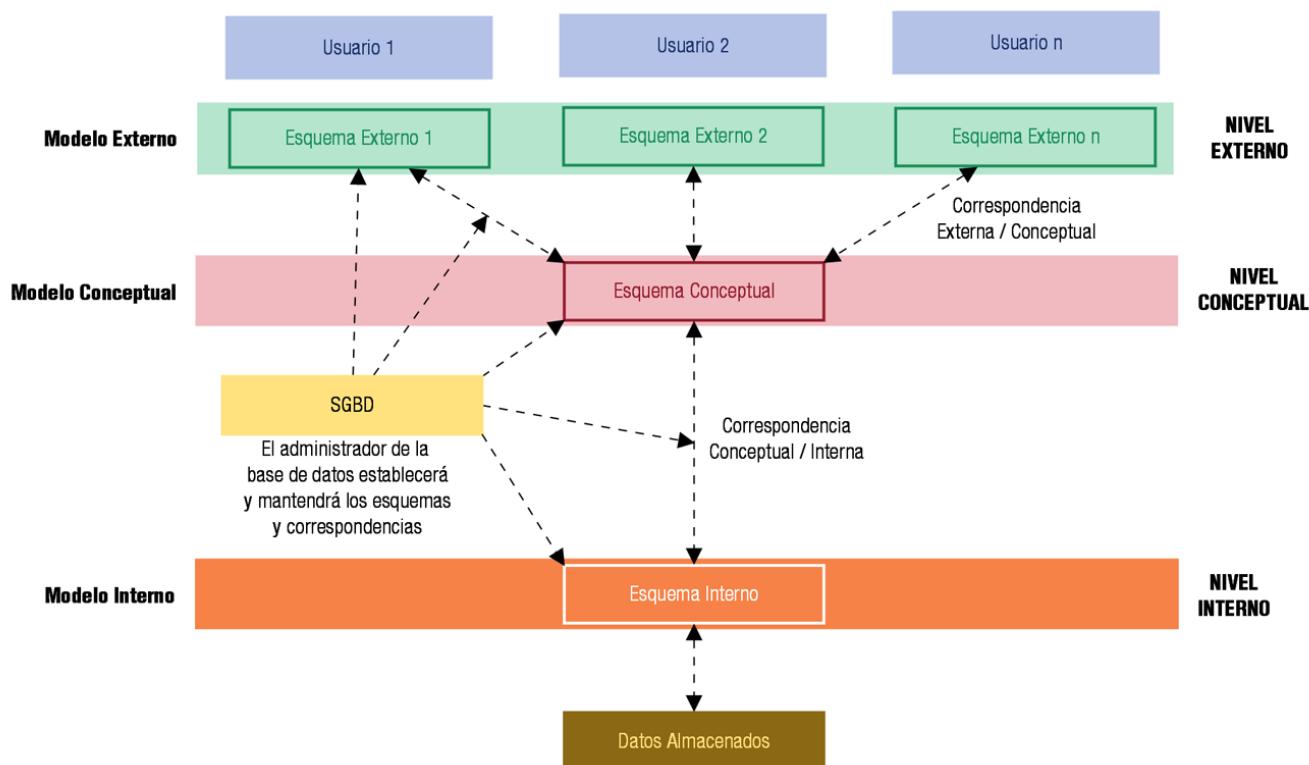
Un SGBD cuenta con una arquitectura a través de la que se simplifica a los diferentes usuarios de la BD su labor. El **objetivo fundamental es separar los programas de aplicación de la BD física**.

Encontrar un estándar para esta arquitectura no es una tarea sencilla, aunque los tres estándares que más importancia han cobrado en el campo de las bases de datos son ANSI/SPARC/X3, CODASYL y ODMG (éste sólo para las BD orientadas a objetos). Tanto ANSI (EEUU), como ISO (Resto del mundo), son el referente en cuanto a estandarización de BD, conformando un único modelo de BD.

La **arquitectura propuesta por ANSI-SPARC** proporciona **tres niveles de abstracción**: nivel interno o físico, nivel lógico o conceptual y nivel externo o de visión del usuario. A continuación, se detallan las características de cada uno de ellos:

- ✓ **Nivel interno o físico (perspectiva del almacenamiento físico)**: en este nivel se **describe la estructura física de la BD a través de un esquema interno encargado de detallar el sistema de almacenamiento de la BD y sus métodos de acceso**. Es el nivel **más cercano al almacenamiento físico**. A través del esquema físico se **indican**, entre otros, los archivos que contienen la información, su organización, los métodos de acceso a los registros, los tipos de registros, la longitud, los campos que los componen, las unidades de almacenamiento, etc.
- ✓ **Nivel lógico o conceptual (perspectiva del administrador)**: en este nivel se **describe la estructura completa de la BD a través de un esquema que detalla las entidades, atributos, relaciones, operaciones de los usuarios y restricciones**. Los detalles relacionados con las estructuras de almacenamiento se ocultan, permitiendo realizar una abstracción a más alto nivel.
- ✓ **Nivel externo o de visión del usuario (perspectiva del programador/usuario)**: en este nivel se **describen las diferentes vistas que los usuarios percibirán de la BD**. Cada tipo de usuario o grupo de ellos verá sólo la parte de la BD que le interesa, ocultando el resto.

Para una BD, sólo existirá un único esquema interno, un único esquema conceptual y podrían existir varios esquemas externos definidos para uno o varios usuarios.



Los SGBD basados en esta arquitectura permiten que cada grupo de usuarios haga referencia a su propio esquema externo. El SGBD debe de transformar cualquier petición de usuario (esquema externo) a una petición expresada en términos de esquema conceptual para finalmente ser una petición expresada en el esquema interno que se procesará

sobre la BD almacenada. El **proceso de transformar peticiones y resultados de un nivel a otro se denomina correspondencia o transformación**, el SGBD es capaz de interpretar una solicitud de datos y realiza los siguientes pasos:

- El usuario solicita unos datos y crea una consulta.
- El SGBD verifica y acepta el esquema externo para ese usuario.
- Transforma la solicitud al esquema conceptual.
- Verifica y acepta el esquema conceptual.
- Transforma la solicitud al esquema físico o interno.
- Selecciona la o las tablas implicadas en la consulta y ejecuta la consulta.
- Transforma del esquema interno al conceptual, y del conceptual al externo.
- Finalmente, el usuario ve los datos solicitados.

Gracias a esta arquitectura se consigue la independencia de datos a dos niveles:

- ✓ **Independencia lógica:** podemos modificar el esquema conceptual sin alterar los esquemas externos ni los programas de aplicación.
- ✓ **Independencia física:** podemos modificar el esquema interno sin necesidad de modificar el conceptual o el externo. Es decir, se puede cambiar el sistema de almacenamiento, reorganizar los ficheros, añadir nuevos, etc., sin que esto afecte al resto de esquemas.



Autoevaluación

El esquema conceptual de la totalidad de la BD puede obtenerse de la unión de todos los esquemas externos definidos para cada usuario de la BD.

- Verdadero
- Falso

5.4 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.

Ver punto 4 de la unidad para tener acceso a la clasificación.

- 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 BD**:
 - ✓ **Centralizados:** sus datos se almacenan en un solo ordenador. Los SGBD centralizados pueden atender a varios usuarios, pero el SGBD y la BD en sí residen por completo en una sola máquina.
 - ✓ **Distribuidos (Homogéneos, Heterogéneos):** la BD 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 BD 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 criterio toma como referencia el **coste**. La mayor parte de los paquetes cuestan entre 10.000 y 100.000 €. Los sistemas monousuario más económicos para microordenadores cuestan entre 0 y 3.000 €. En el otro extremo, los paquetes más completos cuestan más de 100.000 €.
- e) El quinto, y último, criterio establece su clasificación **según el propósito**:
- ✓ **Propósito General:** pueden ser utilizados para el tratamiento de cualquier tipo de BD 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.

6 SGBD comerciales y libres.

Actualmente, en el mercado de software existen multitud de SGBD comerciales. En este epígrafe se desglosan las características fundamentales de los más importantes y extendidos hasta la fecha. Pero, como podrás observar, **la elección de un SGBD es una decisión muy importante a la hora de desarrollar proyectos**. A veces, el sistema más avanzado, "el mejor" según los entendidos, puede no serlo para el tipo de proyecto que estemos desarrollando. **Hemos de tener en cuenta qué volumen de carga debe soportar la BD, qué SO utilizaremos como soporte, cuál es nuestro presupuesto, plazos de entrega, etc.**

En la siguiente tabla se exponen los **SGBD comerciales más utilizados y sus características más relevantes**:

SGBD	Descripción
ORACLE	Reconocido como uno de los mejores a nivel mundial . Es multiplataforma, confiable y seguro . Es Cliente/Servidor. Basado en el modelo de datos Relacional. De gran potencia, aunque con un precio elevado hace que sólo se vea en empresas muy grandes y multinacionales . Ofrece una versión gratuita Oracle Database Express Edition.
MYSQL	Sistema muy extendido que se ofrece bajo dos tipos de licencia, comercial o libre . Para aquellas empresas que deseen incorporarlo en productos privativos, deben comprar una licencia específica . Es relacional, multihilo, multiusuario y multiplataforma . Su gran velocidad lo hace ideal para consulta de BD y plataformas web.
DB2	Multiplataforma, el motor de BD relacional integra XML de manera nativa , lo que IBM ha llamado pureXML, que permite almacenar documentos completos para realizar operaciones y búsquedas de manera jerárquica dentro de éste, e integrarlo con búsquedas relacionales .
INFORMIX	Otra opción de IBM para el mundo empresarial que necesita un DBMS sencillo y confiable . Es un gestor de BD relacional basado en SQL. Multiplataforma . Consuma menos recursos que Oracle, con utilidades muy avanzadas respecto a conectividad y funciones relacionadas con tecnologías de Internet/Intranet, XML, etc.
Microsoft SQL SERVER	SGBD producido por Microsoft . Es relacional, sólo funciona bajo Microsoft Windows , utiliza arquitectura Cliente/Servidor. Constituye la alternativa a otros potentes SGBD como son Oracle, PostgreSQL o MySQL.
SYBASE	Un SGBD con bastantes años en el mercado, tiene 3 versiones para ajustarse a las necesidades reales de cada empresa. Es un sistema relacional, altamente escalable, de alto rendimiento, con soporte a grandes volúmenes de datos, transacciones y usuarios, y de bajo costo .

Otros SGBD comerciales importantes son: DBASE, ACCESS, INTERBASE y FOXPRO.

La alternativa a los SGBD comerciales la encontramos en los **SGBD de código abierto o libres, también llamados Open Source**. Son sistemas **distribuidos y desarrollados libremente**. En la siguiente tabla se relacionan los cinco más utilizados actualmente, así como sus principales características y enlaces a sus páginas web:

SGBD	Descripción
MySQL	Es un SGBD relacional, multihilo y multiusuario con más de seis millones de instalaciones. Distribuido bajo dos tipos de licencias, comercial y libre . Multiplataforma, posee varios motores de almacenamiento, accesible a través de múltiples lenguajes de programación y muy ligado a aplicaciones web .

SGBD	Descripción
PostgreSQL	Sistema Relacional Orientado a Objetos. Considerado como la BD de código abierto más avanzada del mundo . Desarrollado por una comunidad de desarrolladores que trabajan de forma desinteresada, altruista, libre y/o apoyados por organizaciones comerciales. Es multiplataforma y accesible desde múltiples lenguajes de programación .
Firebird	SGBD relacional, multiplataforma , con bajo consumo de recursos , excelente gestión de la conurrencia, alto rendimiento y potente soporte para diferentes lenguajes.
Apache Derby	SGBD escrito en Java, de reducido tamaño , con soporte multilenguaje, multiplataforma, altamente portable , puede funcionar embebido o en modo cliente/servidor .
SQLite	Sistema relacional, basado en una biblioteca escrita en C que interactúa directamente con los programas , reduce los tiempos de acceso siendo más rápido que MySQL o PostgreSQL, es multiplataforma y con soporte para varios lenguajes de programación.

7 BD centralizadas.

Si nos preguntamos cómo es la arquitectura de un sistema de BD, hemos de saber que todo depende del sistema informático que la sustenta. Tradicionalmente, la arquitectura centralizada **fue la que se utilizó inicialmente, aunque hoy en día es de las menos utilizadas**.

Sistema de BD centralizado: es aquella estructura en la que el SGBD está implantado en una sola plataforma u ordenador desde donde se gestiona directamente, de modo centralizado, la totalidad de los recursos. Es la arquitectura de los centros de proceso de datos tradicionales. Se basa en tecnologías sencillas, muy experimentadas y de gran robustez.

Los sistemas de los años sesenta y setenta eran totalmente centralizados, como corresponde a los SO de aquellos años, y al hardware para el que estaban hechos: un gran ordenador para toda la empresa y una red de terminales sin inteligencia ni memoria.

Las principales **características** de las BD centralizadas son:

- ✓ **Se almacena completamente en una ubicación central**, es decir, todos los componentes del sistema residen en un solo ordenador o sitio.
- ✓ **No posee múltiples elementos de procesamiento ni mecanismos de intercomunicación** como las BD distribuidas.
- ✓ Los **componentes** de las BD centralizadas son: **los datos, el software de gestión de BD y los dispositivos de almacenamiento** secundario asociados.
- ✓ Son **sistemas en los que su seguridad puede verse comprometida más fácilmente**.

Ventajas	Inconvenientes
Se evita la redundancia debido a la posibilidad de inconsistencias y al desperdicio de espacio.	Un mainframe en comparación de un sistema distribuido no tiene mayor poder de cómputo.
Se evita la inconsistencia. Ya que si un hecho específico se representa por una sola entrada, la no-concordancia de datos no puede ocurrir.	Cuando un sistema de BD centralizado falla , se pierde toda disponibilidad de procesamiento y sobre todo de información confiada al sistema.
La seguridad se centraliza.	En caso de un desastre o catástrofe, la recuperación es difícil de sincronizar.
Puede conservarse la integridad.	Las cargas de trabajo no se pueden difundir entre varios ordenadores, ya que los trabajos siempre se ejecutarán en la misma máquina.

Ventajas	Inconvenientes
El procesamiento de los datos ofrece un mejor rendimiento.	Los departamentos de sistemas retienen el control de toda la organización.
Mantenimiento más barato. Mejor uso de los recursos y menores recursos humanos.	Los sistemas centralizados requieren un mantenimiento central de datos.

8 BD distribuidas.

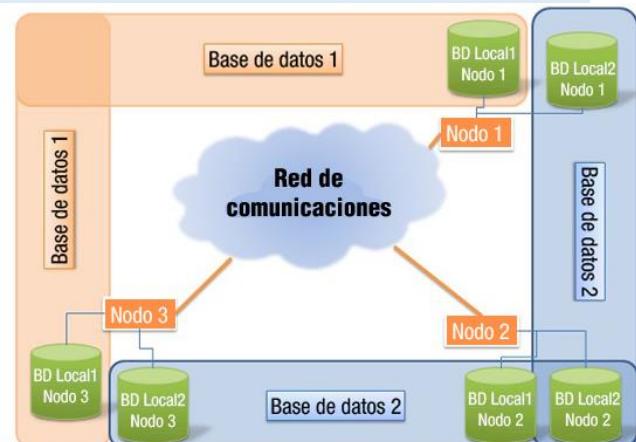
La necesidad de integrar información de varias fuentes y la evolución de las tecnologías de comunicaciones, han producido cambios muy importantes en los sistemas de BD. La respuesta a estas nuevas necesidades y evoluciones se materializa en los sistemas de BD distribuidas.

BD distribuida (BDD): es un conjunto de múltiples BD lógicamente relacionadas las cuales se encuentran distribuidas entre diferentes nodos interconectados por una red de comunicaciones.

Sistema de BD distribuida (SBDD): es un sistema en el cual múltiples sitios de BD están ligados por un sistema de comunicaciones, de tal forma que, un usuario en cualquier sitio puede acceder los datos en cualquier parte de la red exactamente como si los datos estuvieran almacenados en su sitio propio.

SGBD distribuido (SGBDD): es aquel que se encarga del manejo de la BDD y proporciona un mecanismo de acceso que hace que la distribución sea transparente a los usuarios (la aplicación trabajaría, desde un punto de vista lógico, como si un solo SGBD ejecutado en una sola máquina, administrara esos datos).

Un SGBDD desarrollará su trabajo a través de un conjunto de sitios o nodos, que poseen un sistema de procesamiento de datos completo con una BD local, un SGBD e interconectados entre sí. Si estos nodos están dispersos geográficamente se interconectarán a través de una red de área amplia o WAN, pero si se encuentran en edificios relativamente cercanos, pueden estar interconectados por una red de área local o LAN. Este tipo de sistemas es utilizado en: organizaciones con estructura descentralizada, industrias de manufactura con múltiples sedes (automoción), aplicaciones militares, líneas aéreas, cadenas hoteleras, servicios bancarios, etc.



Ventajas	Inconvenientes
El acceso y procesamiento de los datos es más rápido ya que varios nodos comparten carga de trabajo.	La probabilidad de violaciones de seguridad es creciente si no se toman las precauciones debidas.
Desde una ubicación puede accederse a información alojada en diferentes lugares.	Existe una complejidad añadida que es necesaria para garantizar la coordinación apropiada entre los nodos.
Los costes son inferiores a los de las bases centralizadas.	La inversión inicial es menor, pero el mantenimiento y control puede resultar costoso.
Existe cierta tolerancia a fallos. Mediante la replicación, si un nodo deja de funcionar el sistema completo no deja de funcionar.	Dado que los datos pueden estar replicados, el control de concurrencia y los mecanismos de recuperación son mucho más complejos que en un sistema centralizado.
El enfoque distribuido de las BD se adapta más naturalmente a la estructura de las organizaciones. Permiten la incorporación de nodos de forma flexible.	El intercambio de mensajes y el cómputo adicional necesario para conseguir la coordinación entre los distintos nodos

Ventajas	Inconvenientes
	constituyen una forma de sobrecarga que no surge en los sistemas centralizados.
Aunque los nodos están interconectados, tienen independencia local.	Dada la complejidad del procesamiento entre nodos es difícil asegurar la corrección de los algoritmos, el funcionamiento correcto durante un fallo o la recuperación.

8.1 Fragmentación.

Sabemos que **en los sistemas de BDD la información se encuentra repartida en varios lugares**. La forma de extraer los datos consultados puede realizarse mediante la fragmentación de distintas tablas pertenecientes a distintas BD que se encuentran en diferentes servidores. **El problema de la fragmentación se refiere al particionamiento de la información para distribuir cada parte a los diferentes sitios de la red.**

Pero hay que tener en cuenta el **grado de fragmentación** que se aplicará, ya que éste es un **factor determinante a la hora de la ejecución de consultas**. Si no existe fragmentación, se tomarán las relaciones o tablas como la unidad de fragmentación. Pero también **puede fragmentarse a nivel de tupla (fila o registro) o a nivel de atributo (columna o campo) de una tabla**. No será adecuado un grado de fragmentación nulo, ni tampoco un grado de fragmentación demasiado alto. El **grado de fragmentación deberá estar equilibrado y dependerá de las particularidades de las aplicaciones que utilicen dicha base de datos**. Concretando, **el objetivo de la fragmentación es encontrar un nivel de particionamiento adecuado en el rango que va desde tuplas o atributos hasta relaciones completas**.

Cuando se lleva a cabo una fragmentación, existen **tres reglas fundamentales a cumplir**:

- ✓ **Compleitud:** si una relación R se descompone en fragmentos R1, R2, ..., Rn, **cada elemento de datos que pueda encontrarse en R deberá poder encontrarse en uno o varios fragmentos Ri**.
- ✓ **Reconstrucción:** si una relación R se descompone en una serie de fragmentos R1, R2, ..., Rn, **la reconstrucción de la relación a partir de sus fragmentos asegura que se preservan las restricciones definidas sobre los datos**.
- ✓ **Disyunción:** si una relación R se descompone verticalmente, sus **atributos primarios clave normalmente se repiten en todos sus fragmentos**.

Existen **tres tipos de fragmentación**:

- ✓ **Fragmentación horizontal:** la fragmentación horizontal **se realiza sobre las tuplas o filas de la relación, dividiendo la relación en subrelaciones** que contienen un subconjunto de las tuplas que alberga la primera. Existen **dos variantes** de la fragmentación horizontal: la **primaria** (particionar las tuplas de una relación en subconjuntos, donde cada subconjunto tenga propiedades comunes) y la **derivada** (consiste en dividir una relación partiendo de las condiciones definidas sobre alguna otra).
- ✓ **Fragmentación vertical:** la fragmentación vertical, en cambio, **se basa en los atributos de la relación para efectuar la división**. Una **relación R produce fragmentos R1, R2, ..., Rr**, **cada uno de los cuales contiene un subconjunto de los atributos de R, así como la llave primaria de R**. El objetivo de la fragmentación vertical es partitionar una relación en un conjunto de relaciones más pequeñas de manera que varias de las aplicaciones de usuario se ejecutarán sobre un fragmento. En este contexto, una fragmentación óptima es aquella que produce un esquema de fragmentación que minimiza el tiempo de ejecución de las consultas de usuario. La fragmentación vertical es más complicada que la horizontal, ya que existe un gran número de alternativas para realizarla.
- ✓ **Fragmentación híbrida o mixta:** podemos **combinar ambas**, utilizando por ello la denominada fragmentación mixta. Si tras una fragmentación vertical se lleva a cabo otra horizontal, se habla de la **fragmentación mixta HV**. Para el caso contrario, estaremos ante una **fragmentación VH**.



TAREA

- Imagina que vas a desarrollar una aplicación para llevar el control de las ventas de ordenadores de una tienda de informática y en principio no vas a utilizar ningún gestor de BD ¿Podrías crear dicha aplicación? ¿de qué elemento se dispone en un Sistema Operativo para guardar la información? ¿qué problemas crees que va a plantear este desarrollo? ¿de qué tipos pueden ser los elementos utilizados para guardar la información? Pon un ejemplo de la información que guardarías en cada uno de los tipos de la pregunta anterior.
- Crea una tabla con las ventajas e inconvenientes que presentan las distintas formas de organizar los archivos o ficheros para BD.
- Cita las ventajas de las BD sobre los sistemas de gestión de datos basados en ficheros.
- Teniendo en cuenta los distintos modelos de BD, ¿qué tipo de BD es el más utilizado en la actualidad? ¿qué tipos de BD no se utilizan actualmente? ¿qué tipo de BD noSQL están apareciendo en la actualidad y en qué tipo de aplicaciones se suelen utilizar?
- ¿Cuántos modelos conceptuales se pueden tener de una base de datos? ¿y externos? ¿qué diferencia hay entre ellos?
- Dada la imagen siguiente, describe los pasos que sigue el SGBD para interpretar una solicitud de datos de un usuario determinado. ¿Qué significa el nivel de abstracción?



- ¿Qué funciones realiza un SGBD? Describe cada una de ellas.
- En función de la ubicación de la información ¿qué tipos de almacenamiento podemos tener? ¿qué ventajas e inconvenientes presenta cada uno de ellos?
- ¿Cuáles son los componentes de un SGBD? ¿qué función realiza cada uno de ellos? Describe y comenta la imagen de la página 21.
- ¿Qué permiten las BD distribuidas? Enumera sus ventajas e inconvenientes.
- Considera la siguiente relación, ¿qué fragmentos se obtendrían al realizar una fragmentación horizontal por ESCUELA?
- Considera las siguientes relaciones, ¿qué fragmentos principales y derivados se obtendrían al realizar una fragmentación horizontal por ESCUELA?

Jno	NOMBRE	NOTA	ESCUELA
J1	LUIS YANEZ	8	CIME
J2	ERIKA QUIROZ	8	CIME
J3	DANIEL MURILLO	9	EISIC
J4	MARIA JOSE MENDEZ	10	EISIC
J5	ADONIS PABON	9	EISIC

ID	MATERIA	ESCUELA	CRÉDITOS	NIVEL
1	Análisis Matemático	EISIC	4	1
2	Sistemas Operativos	EISIC	6	3
3	Programación II	CIME	6	2
4	Tecnología Eléctrica	EISIC	4	2
5	Técnicas de Aprendizaje	CIME	4	1
6	Dibujo Mecánico	CIME	6	3

Jno	NOMBRE	MATERIA	NOTA
J1	LUIS YANEZ	Sistemas Operativos	8
J2	ERIKA QUIROZ	Dibujo Mecánico	8
J3	DANIEL MURILLO	Técnicas de Aprendizaje	9
J4	MARIA JOSE MENDEZ	Análisis Matemático	10
J5	ADONIS PABON	Programación II	9

- Considera la relación del ejercicio 11. Realiza una fragmentación vertical por nota y por nombre/escuela.

Vídeos en YouTube

- Archivos y tipos de archivos: <https://www.youtube.com/watch?v=ZbhcqdnH0Q8A>
- Archivos binarios: <https://www.youtube.com/watch?v=LZEjh0AtOH4>
- Estructuras de archivos para bases de datos: https://www.youtube.com/watch?v=5lyPv_SeFls
- Modelos de bases de datos: <https://www.youtube.com/watch?v=KUwkCysHZ9g>
- Sistemas de gestión de bases de datos: <https://www.youtube.com/watch?v=O1uXxI4KteA>
- Gestión de datos: <https://www.youtube.com/watch?v=Uj3pf9kH3cl>
- Arquitectura de un SGBD: <https://www.youtube.com/watch?v=h5FggYdjdMA>
- Bases de datos SQL vs NoSQL:
 - <https://www.youtube.com/watch?v=0xx3fcYcb28>
 - <https://www.youtube.com/watch?v=zmXI2dOGWL8>