

# Unidad 1. Almacenamiento de la información

---

BASES DE DATOS

Ciclo Superior en Desarrollo de Aplicaciones Web  
CURSO 2022-2023

# Índice

1. Introducción.....	2
2. Ficheros de información .....	3
2.1. ¿Qué es un fichero? .....	3
2.2. Tipos de ficheros .....	4
2.3. Los soportes de información .....	4
3. Bases de datos.....	6
3.1. ¿Qué es una base de datos? .....	6
3.2. Usos de las bases de datos .....	8
3.3. Ubicación de la información .....	10
4. Modelos de bases de datos.....	12
4.1. Modelo jerárquico .....	12
4.2. Modelo en red .....	13
4.3. Modelo relacional .....	13
4.4. Modelo orientado a objetos .....	15
4.5. Otros modelos.....	15
5. Tipos de bases de datos .....	17
5.1. Bases de datos según su contenido .....	17
5.2. Bases de datos según su uso.....	17
5.3. Bases de datos según la variabilidad de la información .....	18
5.4. Bases de datos según la localización de la información .....	18
6. Sistemas gestores de bases de datos .....	19
6.1. Funciones de los SGBD .....	20
6.2. Componentes de los SGBD .....	21
6.3. Arquitectura de un SGBD .....	23
6.4. Tipos de SGBD .....	24
7. SGBD comerciales.....	26
8. SGBD libres .....	27
9. Bases de datos centralizadas .....	28
10. Bases de datos distribuidas .....	30
10.1. Fragmentación .....	31

# 1. Introducción

¿Te has preguntado alguna vez dónde y de qué manera se almacenan y gestionan los datos que utilizamos diariamente?

Si pensamos en cualquier acción de nuestra vida cotidiana, o si analizamos la mayoría de los ámbitos de actividad, nos encontramos que la utilización de las bases de datos está ampliamente extendida. Éstas, y los datos contenidos en ellas, serán imprescindibles para llevar a cabo multitud de acciones.

Piensa en las siguientes situaciones:

- Cuando seleccionamos nuestro canal favorito en la TDT.
- Al utilizar la agenda del móvil para realizar una llamada telefónica.
- Cuando operamos en el cajero automático.
- Al solicitar un certificado en un organismo público.
- Cuando acudimos a la consulta del médico.
- Al inscribirnos en un curso, plataforma OnLine, etc.
- Si utilizas un GPS.
- Cuando reservamos unas localidades para un evento deportivo o espectáculo.
- Si consumimos ocio digital.
- Cuando consultamos cualquier información en Internet. (Bibliotecas, enciclopedias, museos, etc.)
- Al registrarte en una página de juegos OnLine, redes sociales o foros.
- Incluso, si tienes coche, puede ser que éste incorpore alguna base de datos.

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

Este módulo profesional se centra en el estudio de las **bases de datos** y su uso en el desarrollo de aplicaciones. En esta primera unidad comenzaremos conociendo los primeros sistemas basados en ficheros para el almacenamiento y gestión de la información. Seguidamente, se desarrollarán los conceptos y definiciones básicas relacionadas con las bases de datos. Posteriormente, analizaremos sus modelos y tipos, y un poco más adelante, podremos conocer las características y capacidades de los sistemas gestores de bases de datos. Finalmente, identificaremos las herramientas reales con las que llevar a cabo la gestión de dichas bases.

## 2. Ficheros de información

### 2.1. ¿Qué es un fichero?

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 esta información era necesario establecer métodos adecuados para su almacenamiento. El elemento que permitió llevar a cabo el almacenamiento de datos de forma permanente en dispositivos de memoria masiva fue **el fichero o archivo**.

**Fichero o archivo:** conjunto de información relacionada, tratada como un todo y organizada de forma estructurada. Es una secuencia de dígitos binarios que organiza información relacionada con un mismo aspecto.

Los ficheros están formados por **registros lógicos** que contienen datos relativos a un mismo elemento u objeto (por ejemplo, los datos de usuarios de una plataforma educativa). A su vez, los registros están divididos en campos que contienen cada una de las informaciones elementales que forman un registro (por ejemplo, el nombre del usuario o su dirección de correo electrónico).

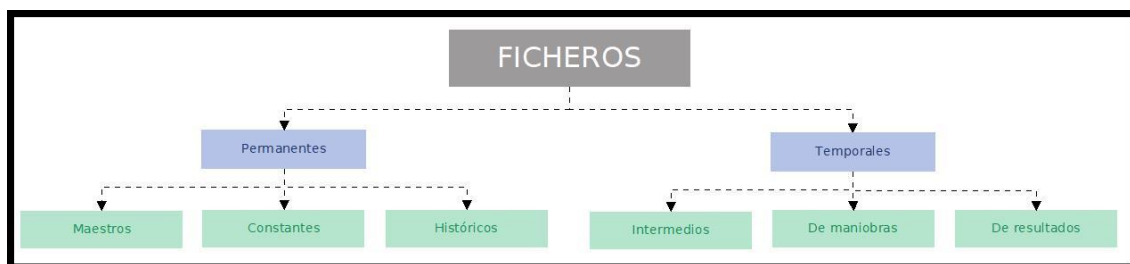
Hemos de resaltar que los datos están almacenados de tal forma que se puedan añadir, suprimir, actualizar o consultar individualmente en cualquier momento.

Como los ficheros suelen ser muy voluminosos, solo se pueden llevar a la memoria principal partes de ellos para poder procesarlos. La cantidad de información que es transferida entre el soporte en el que se almacena el fichero y la memoria principal del ordenador, en una sola operación de lectura/grabación, recibe el nombre de **registro físico o bloque**.

Normalmente en cada operación de lectura/grabación se transfieren varios registros del fichero, es decir un bloque suele contener varios registros lógicos. Al número de registros que entran en un bloque se le conoce con el nombre de **factor de blocaje**, y a esta operación de agrupar varios registros en un bloque se le llama **bloqueo de registros**.

## 2.2. Tipos de ficheros

Según la función que vayan 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. Es decir, los datos necesarios para el funcionamiento de ésta. Tienen un periodo de permanencia en el sistema amplio. 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.
  - **Ficheros constantes:** son aquellos que incluyen datos fijos para la aplicación. No suelen ser modificados y se accede a ellos para realización de consultas. Podría ser un archivo con códigos postales.
  - **Ficheros históricos:** contienen datos que fueron considerados como actuales en un periodo o situación anterior. Se utilizan para la reconstrucción de situaciones. 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.
  - **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.

## 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.

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

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

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

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

### 3. Bases de datos

Como hemos visto anteriormente, los ficheros permiten organizar y memorizar conjuntos de datos del mismo tipo o naturaleza con una determinada estructura, siendo un medio para el almacenamiento de los datos o resultados de una aplicación específica. El problema es que las aplicaciones que usan ficheros, al ser diseñadas, deben depender directamente de éstos y pierden independencia, lo que provoca serios inconvenientes: información duplicada, incoherencia de datos, fallos de seguridad, etc.

Para solucionar estos problemas, surge el concepto de base de datos. Una base de datos permitirá reunir toda la información relacionada en un único sistema de almacenamiento, de forma que cualquier aplicación pueda utilizarla de manera independiente y ofreciendo una mejora en el tratamiento de la información, así como una evolución para el desarrollo de aplicaciones.

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

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

#### 3.1. ¿Qué es una base de datos?

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

Entonces, ¿qué es una base de datos?

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

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

Por ejemplo, en una base de datos que almacena imágenes, los datos serían las propias imágenes. Cada imagen podría tener asociados una serie de metadatos que nos aportan información sobre ella, como podrían ser su tamaño o su formato (jpeg, png...).

Una base de datos constará de los siguientes elementos:

- **Entidades:** objeto real o abstracto con características diferenciadoras de otros, del que se almacena información en la base de datos. Dicho de otra forma, es algo acerca de lo cual se desea almacenar información. Por ejemplo, en una base de datos de un hospital, las entidades podrían ser: médico, paciente, consulta, etc.
- **Atributos:** son los datos que se almacenan de la entidad. Cualquier propiedad o característica de una entidad puede ser un atributo. Continuando con nuestro ejemplo, podrían ser atributos de un paciente: número de seguridad social, nombre, fecha nacimiento, 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 de cada paciente podría ser: *27/123056, Pablo Pérez Rodríguez, 23/03/2009*.
- **Campos:** donde se almacenan los atributos de cada registro. Teniendo en cuenta el ejemplo anterior, un campo podría ser la fecha de nacimiento.

Tal y como veremos en próximas unidades, en una base de datos relacional la información de las diferentes entidades se almacenará en tablas. Cada fila de una tabla es un registro, mientras que cada columna es un campo.

Las **ventajas fundamentales** que ofrece el uso de bases de datos se resumen a continuación:

- **Acceso múltiple:** diversos usuarios o aplicaciones podrán acceder a la base de datos 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 base de datos, 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 base de datos.



- **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 base de datos o a las aplicaciones que acceden a ellos.
- **Independencia lógica:** los cambios realizados en la base de datos 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 base de datos de manera sencilla y cómoda.
- **Consulta directa:** existe una herramienta para poder acceder a los datos interactivamente.

### 3.2. Usos de las bases de datos

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

#### ¿Quién utiliza las bases de datos?

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

En la siguiente tabla se muestran las funciones y características de cada uno de los usuarios de una base de datos:

Tipo	Funciones y características
<b>El administrador</b>	Es la persona encargada de la creación o implementación física de la base de datos. 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.
<b>Los diseñadores</b>	Son las personas encargadas de diseñar cómo será la base de datos. Llevarán a cabo la identificación de los datos, las relaciones entre ellos, sus restricciones, etc. Para ello han de conocer a fondo los

Tipo	Funciones y características
	datos y procesos a representar en la base de datos. Si estamos hablando de una empresa, será necesario que conozcan las reglas de negocio en la que ésta se mueve. Para obtener un buen resultado, el diseñador de la base de datos debe implicar en el proceso a todos los usuarios de la base de datos, tan pronto como sea posible.
<b>Los programadores de aplicaciones</b>	Una vez diseñada y construida la base de datos, los programadores se encargarán de implementar los programas de aplicación que servirán a los usuarios finales. Estos programas de aplicación ofrecerán la posibilidad de realizar consultas de datos, inserción, actualización o eliminación de los mismos. Para desarrollar estos programas se utilizan lenguajes de tercera o cuarta generación.
<b>Los usuarios finales</b>	Son los clientes finales de la base de datos. Al diseñar, implementar y mantener la base de datos se busca cumplir los requisitos establecidos por el cliente para la gestión de su información.

### ¿Para qué se utilizan las bases de datos?

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

- Banca: información de clientes, cuentas, transacciones, préstamos, etc.
- Líneas aéreas: información de clientes, horarios, vuelos, destinos, etc.
- Universidades: información de estudiantes, carreras, horarios, materias, etc.
- Transacciones de tarjeta de crédito: para comprar con tarjetas de crédito y la generación de los extractos mensuales.
- Telecomunicaciones: para guardar registros de llamadas realizadas, generar facturas mensuales, mantener el saldo de las tarjetas telefónicas de prepago y almacenar información sobre las redes.
- Medicina: información hospitalaria, biomedicina, genética, etc.
- Justicia 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.

### 3.3. Ubicación de la información

Utilizamos a diario las bases de datos, pero ¿dónde se encuentra realmente almacenada la información?

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

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

- **Discos SATA:** Es una interfaz de transferencia de datos entre la placa base y algunos dispositivos de almacenamiento, como puede ser el disco duro, lectores y regrabadores de CD/DVD/BD, Unidades de Estado Sólido u otros dispositivos. SATA proporciona mayores velocidades, mejor aprovechamiento cuando hay varias unidades, mayor longitud del cable de transmisión de datos y capacidad para conectar unidades al instante, es decir, insertar el dispositivo sin tener que apagar el ordenador.
- **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*).
- **RAID:** acrónimo de *Redundant Array of Independent Disks* o matriz de discos independientes, es un contenedor de almacenamiento redundante. Se basa en el montaje en conjunto de dos o más discos duros, formando un bloque de trabajo, para obtener desde una ampliación de capacidad a mejoras en velocidad y seguridad de almacenamiento. Según las características que queramos primar, se establecen distintos sistemas de RAID.
- **Sistemas NAS:** Es el acrónimo de *Network Attached Storage* o sistema de almacenamiento masivo en red. Estos sistemas de almacenamiento permiten compartir la capacidad de almacenamiento de un computador (Servidor) con ordenadores personales o servidores clientes a través de una red, haciendo uso de un sistema operativo optimizado para dar acceso a los datos a través de protocolos de comunicación específicos. Suelen ser dispositivos para almacenamiento masivo de datos con capacidades muy altas, de varios Terabytes, generalmente superiores a los discos duros externos. Además, se diferencian de estos al conectarse por red.

- **Sistemas SAN:** Acrónimo de *Storage Area Network* o red de área de almacenamiento. Se trata de una red concebida para conectar servidores, matrices (arrays) de discos y librerías de soporte. La arquitectura de este tipo de sistemas permite que los recursos de almacenamiento estén disponibles para varios servidores en una red de área local o amplia. Debido a que la información almacenada no reside directamente en ninguno de los servidores de la red, se optimiza el poder de procesamiento para aplicaciones comerciales y la capacidad de almacenamiento se puede proporcionar en el servidor donde más se necesite.

## 4. Modelos de bases de datos

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

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

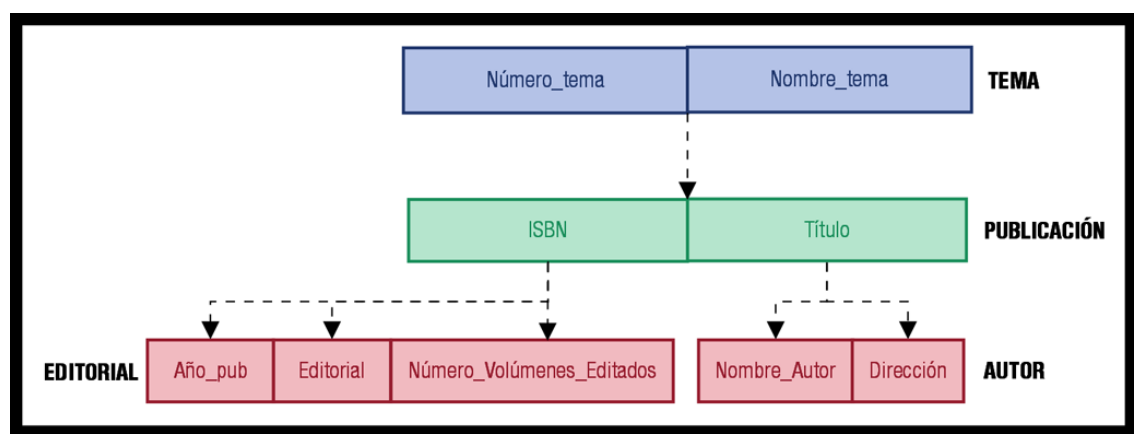
### 4.1. Modelo jerárquico

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. Un nodo puede tener 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. En el siguiente gráfico puedes observar la estructura de almacenamiento del modelo jerárquico:



En este ejemplo, el nodo padre es **TEMA**, que tiene un nodo hijo (**PUBLICACIÓN**). El nodo **TEMA** podría tener varios hijos **PUBLICACIÓN**, lo que indicaría que existen varias publicaciones que tratan del mismo tema. A su vez, cada publicación tiene asociada una editorial y un autor.

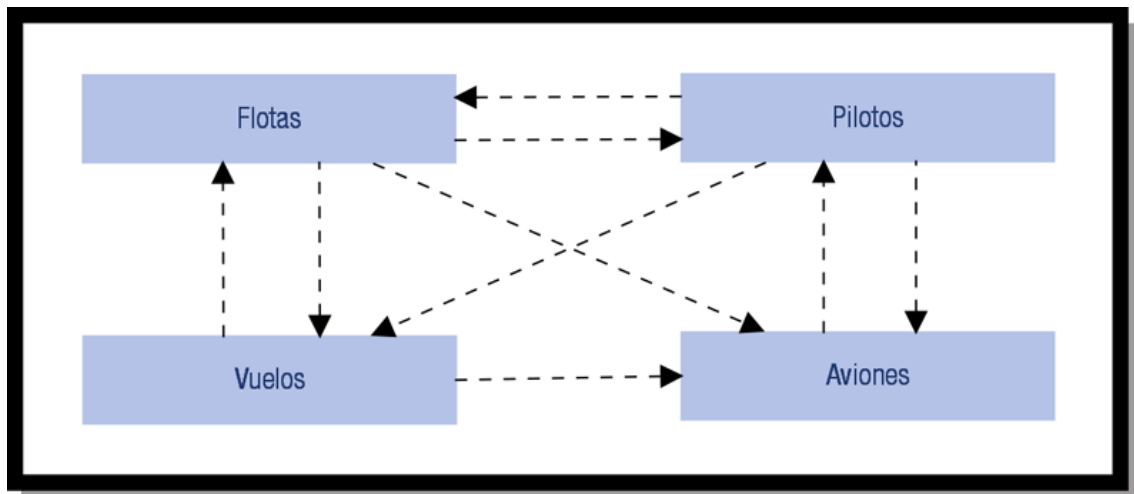
## 4.2. Modelo en red

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

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

En este modelo se pueden representar perfectamente cualquier tipo de relación entre los datos, pero hace muy complicado su manejo. Al no tener que duplicar la información se ahorra espacio de almacenamiento. El sistema de gestión de información basado en el modelo en red más popular es el sistema IDMS.

A continuación, se muestra un ejemplo de base de datos en red para un aeropuerto:



## 4.3. Modelo relacional

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

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

El modelo relacional utiliza **tablas bidimensionales** (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 deben ser 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.
- Cada registro o tupla es identificada por una clave que puede estar formada por uno o varios campos o atributos.

A continuación, puedes observar cómo es una relación con sus tuplas y atributos en el modelo relacional, en una base de datos de alumnos:



El lenguaje habitual para construir las consultas a bases de datos relacionales es **SQL** (*Structured Query Language*). Es un estándar implementado por los principales motores o sistemas de gestión de bases de datos relacionales, y con el que trabajaremos en las próximas unidades.

## 4.4. Modelo orientado a objetos

El modelo orientado a objetos define una base de datos 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**.

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, soporte multimedia, etc.

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

## 4.5. Otros modelos

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

### Modelo Objeto-Relacional

Las bases de datos pertenecientes a este modelo son un híbrido entre las bases del modelo relacional y el orientado a objetos. En una base de datos 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. La gran mayoría de las bases de datos relacionales clásicas de gran tamaño, como Oracle, SQL Server... son objeto-relacionales.

### Bases de datos transaccionales

Son bases de datos 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. Entre la más destacadas se encuentra Oracle.



**Modelo de bases de datos orientadas a documentos.** En este modelo el principal objeto de gestión es el documento. Estas bases de datos almacenan documentos que contienen datos que podrán estar almacenados en algún formato, como por ejemplo XML. Una de las bases de datos orientadas a documentos más relevantes es MongoDB.

Otra forma extendida de clasificar a las bases de datos es en bases de datos SQL y NoSQL. En el siguiente enlace tienes información sobre bases de datos NoSQL:

[Bases de datos NoSQL](#)

## 5. Tipos de bases de datos

Como hemos visto, existen diferentes tipos de bases de datos en función del modelo de datos que utilizan. Sin embargo, esta no es la única forma de diferenciar las bases de datos. A continuación, veremos diferentes criterios por los que también se pueden clasificar:

### 5.1. Bases de datos según su contenido

- **Bases de datos con información actual:** contienen información muy concreta y actualizada, normalmente, de tipo numérico (estadísticas, series históricas, resultados de encuestas...).
- **Directorios:** recogen datos sobre personas o instituciones especializadas en una actividad o materia concreta. Hay directorios de profesionales, de investigadores, de bibliotecas, de revistas científicas...
- **Bases de datos documentales:** En este último grupo, cada registro se corresponde con un documento, sea éste de cualquier tipo (una publicación impresa, un documento audiovisual, gráfico...). Dependiendo de si incluyen o no el contenido completo de los documentos que describen, podremos tener:
  - **Bases de datos de texto completo:** constituidas por los propios documentos en formato electrónico, por un volcado completo de su texto.
  - **Archivos electrónicos de imágenes:** constituidos por referencias que permiten un enlace directo con la imagen del documento original, sea éste un documento iconográfico (fotografías, imágenes de televisión...) o un documento impreso digitalizado en formato de imagen.
  - **Bases de datos referenciales:** sus registros no contienen el texto original sino tan sólo la información fundamental para describir y permitir la localización de documentos. Permiten obtener referencias sobre documentos que habrá que localizar posteriormente en otro servicio (archivo, biblioteca, fototeca, fonoteca...) o solicitar a un servicio de suministro de documentos.

### 5.2. Bases de datos según su uso

- **Base de datos individual:** Es una base de datos utilizada básicamente por una persona. El sistema administrador de la base de datos y los datos son controlados por el mismo usuario. Puede estar almacenada en la unidad de disco duro del usuario o en el servidor de archivos de una red de área local. Por ejemplo, un gerente de ventas podría contar con una base de datos para el control de sus vendedores y su desempeño.

- **Base de datos compartida:** Son bases de datos con múltiples usuarios y que muy probablemente pertenezcan a la misma organización, como la base de datos de una compañía. Se encuentra almacenada en una computadora potente y bajo el cuidado de un profesional en el área, el administrador de la base de datos. Los usuarios tienen acceso a la base de datos mediante una red de área local o una red de área extensa.
- **Base de datos de acceso público:** Son bases de datos accesibles por cualquier persona. Puede no ser necesario pagar un canon para hacer uso de los datos contenidos en ellas.
- **Base de datos propietarias o bancos de datos:** Se trata en general de bases de datos de gran tamaño, desarrolladas por una organización y que contienen temas especializados o de carácter particular. El público general puede tener acceso a estas bases a veces de forma gratuita y otras mediante el pago de una cuota. Pueden ofrecer información que va desde negocios, economía, inversión, técnica y científica hasta servicios de entretenimiento. Permiten encontrar en minutos lo que tardaría horas ojeando revistas.

### 5.3. Bases de datos según la variabilidad de la información

- **Bases de datos estáticas:** Son bases de datos de sólo lectura. Se utilizan para el almacenamiento de datos históricos que pueden ser analizados y utilizados para el estudio del comportamiento de un conjunto de datos a través del tiempo.
- **Bases de datos dinámicas:** Son bases de datos donde la información almacenada se modifica con el tiempo, permitiendo operaciones como actualización y adición de datos, además de las operaciones fundamentales de consulta.

### 5.4. Bases de datos según la localización de la información

- **Bases de datos centralizadas:** Se trata de bases de datos ubicadas en un único lugar, un único computador. Pueden ser bases de datos monousuario que se ejecutan en ordenadores personales o sistemas de bases de datos de alto rendimiento que se ejecutan en grandes sistemas.
- **Bases de datos distribuidas:** Según la naturaleza de la organización es probable que los datos no se almacenen en un único punto, sino que se sitúen en un lugar o lugares diferentes a donde se encuentran los usuarios. Una base de datos distribuida es la unión de las bases de datos mediante redes. Los usuarios se vinculan a los servidores de bases de datos distantes mediante una amplia variedad de redes de comunicación. Un ejemplo sería una compañía con diferentes oficinas regionales. La base de datos de la compañía se encuentra distribuida entre las diferentes oficinas (de forma que en cada oficina se almacenan los datos relacionados con ella). Sin embargo, mediante la conexión por red, los ejecutivos de la compañía pueden tener acceso a la información de todas las oficinas regionales.

## 6. Sistemas gestores de bases de datos

Para poder tratar la información contenida en las bases de datos se utilizan los sistemas gestores de bases de datos o SGBD, 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 bases de datos. Estos sistemas están diseñados para la manipulación de grandes bloques de información.

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

El SGBD permite a los usuarios la creación y el mantenimiento de una base de datos, facilitando la definición, construcción y manipulación de la información contenida en ésta. Definir una base de datos consistirá en especificar los tipos de datos, las estructuras y las restricciones que los datos han de cumplir a la hora de almacenarse en dicha base. Por otro lado, la construcción de la base de datos será el proceso de almacenamiento de datos concretos en algún medio o soporte de almacenamiento que esté supervisado por el SGBD. Finalmente, la manipulación de la base de datos incluirá la posibilidad de realización de consultas para recuperar información específica, la actualización de los datos y la generación de informes a partir de su contenido.

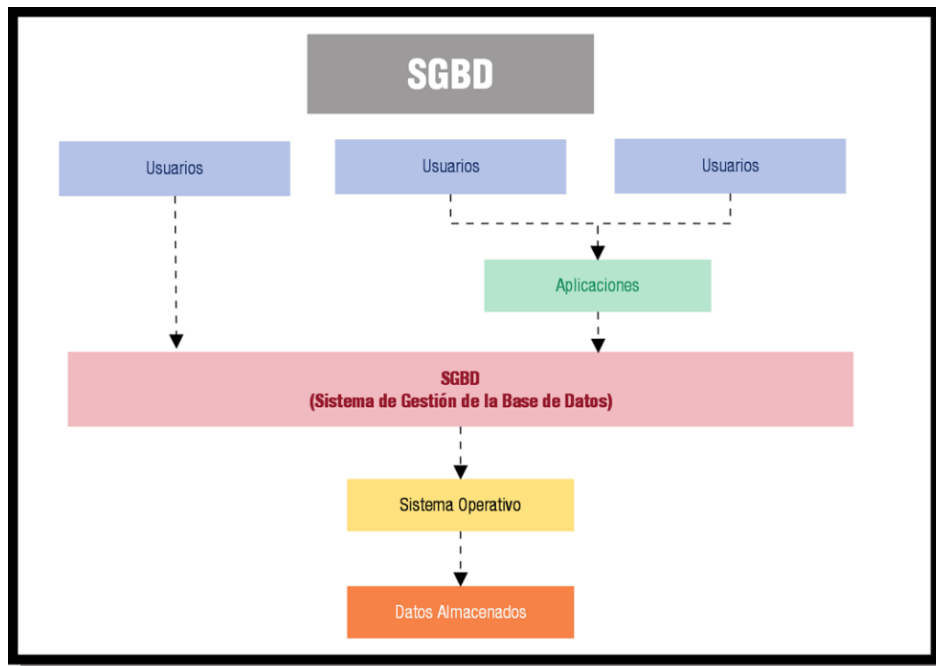
Las **ventajas** del uso de SGBD son:

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

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

manipulación de la información, basándose en las funciones y métodos propios del sistema operativo.

A continuación, se puede observar un gráfico que muestra la interacción del SGBD con los demás elementos del sistema:



## 6.1. Funciones de los SGBD

Un SGBD desarrolla tres funciones fundamentales: descripción, manipulación y control o utilización de los datos. A continuación, se detalla cada una de ellas:

### Función de descripción o definición

Permite al diseñador de la base de datos crear las estructuras apropiadas para integrar adecuadamente los datos. Esta función se realiza mediante el **lenguaje de descripción de datos (DDL)** y es la que permite definir las tres estructuras de la base de datos: estructura interna, estructura conceptual y estructura externa.

- **A nivel interno** (estructura interna), se ha de indicar el espacio de disco reservado para la base de datos, la longitud de los campos y su modo de representación.
- **A nivel conceptual** (estructura conceptual), se proporcionan herramientas para la definición de las entidades, las relaciones entre ellas, las restricciones de integridad...
- **A nivel externo** (estructura externa), se deben definir las vistas de los distintos usuarios a través del lenguaje para la definición de estructuras externas.

### 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.

Por manipulación de datos entenderemos:

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

### Función de control

Permite al administrador de la base de datos establecer mecanismos de protección de las diferentes visiones de los datos asociadas a cada usuario, proporcionando elementos de creación y modificación de dichos usuarios. Adicionalmente, incorpora sistemas para la creación de copias de seguridad, carga de ficheros, auditoría, protección de ataques, configuración del sistema, etc. El lenguaje que implementa esta función es el **lenguaje de control de datos (DCL)**.

¿Y a través de qué lenguaje podremos desarrollar estas tres funciones sobre la base de datos? Lo haremos utilizando el **Lenguaje Estructurado de Consultas (SQL)**.

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 pueden ser generadores de formularios, informes, interfaces gráficas, generadores de aplicaciones, etc.

## 6.2. Componentes de los SGBD

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 base de datos:** cualquier sistema gestor de base de datos 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:** permite almacenar metadatos para así describir 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 base de datos. Contiene las características lógicas de las estructuras que almacenan los datos, su nombre, descripción, contenido y organización. En una base de datos 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 base de datos:** 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 base de datos:** 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 base de datos** o Database Administrator (DBA), que será la persona o conjunto de ellas encargadas de la función de administración de la base de datos. Tiene el control centralizado de la base de datos y es el responsable de su buen funcionamiento. Es el encargado de autorizar el acceso a la base de datos, de coordinar y vigilar su utilización y de adquirir los recursos software y hardware que sean necesarios.
  - Los **usuarios de la base de datos**, que serán diferentes usuarios de la BD con diferentes necesidades sobre los datos, así como diferentes accesos y privilegios. Podemos establecer la siguiente clasificación:
    - Diseñadores.
    - Operadores y personal de mantenimiento.
    - Analistas y programadores de aplicaciones.
    - Usuarios finales: ocasionales, simples, avanzados y autónomos.
5. **Herramientas de la base de datos.** Son un conjunto de aplicaciones que permiten a los administradores la gestión de la base de datos, de los usuarios y

permisos, generadores de formularios, informes, interfaces gráficas, generadores de aplicaciones, etc.

### 6.3. Arquitectura de un SGBD

Un SGBD cuenta con una arquitectura a través de la que se simplifica a los diferentes usuarios de la base de datos su labor. El objetivo fundamental es separar los programas de aplicación de la base de datos 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 bases de datos orientadas a objetos). Tanto ANSI (EEUU) como ISO (resto del mundo) son el referente en cuanto a estandarización de bases de datos, conformando un único modelo de bases de datos.

La arquitectura propuesta 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:** en este nivel se describe la estructura física de la base de datos a través de un esquema interno encargado de detallar el sistema de almacenamiento de la base de datos 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:** en este nivel se describe la estructura completa de la base de datos 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:** en este nivel se describen las diferentes vistas que los usuarios percibirán de la base de datos. Cada tipo de usuario o grupo de ellos verá solo la parte de la base de datos que le interesa, ocultando el resto.

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

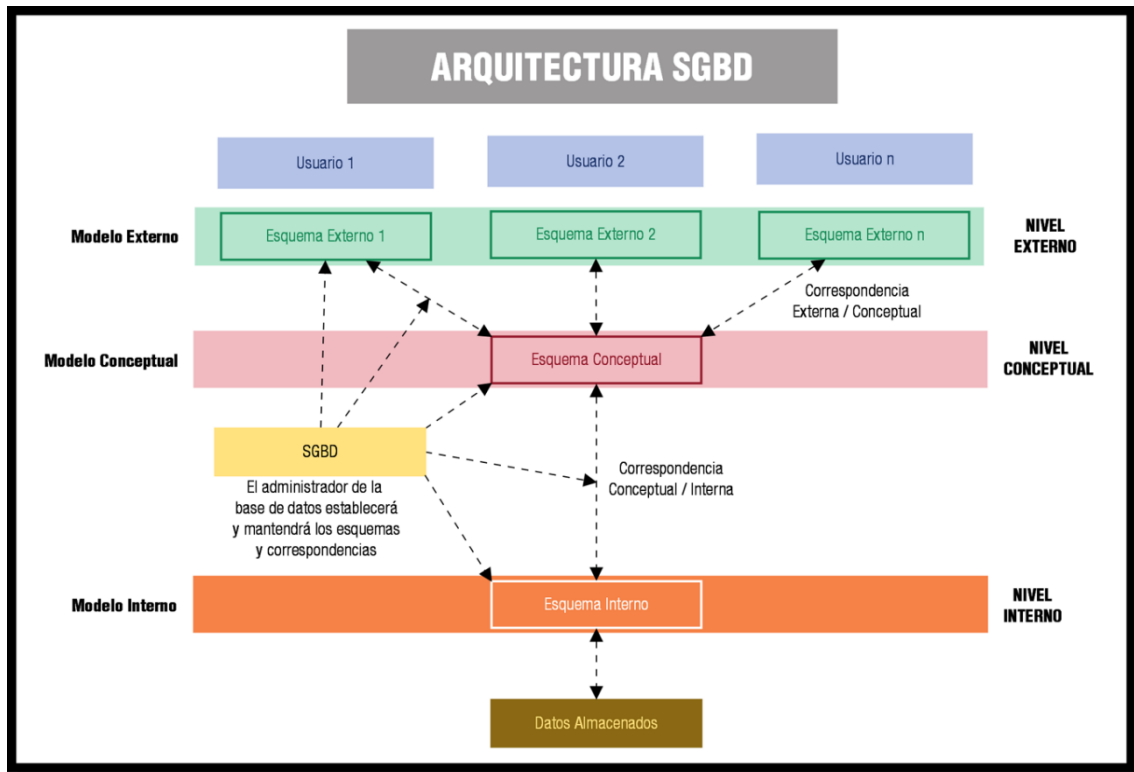
Gracias a esta arquitectura se consigue la **independencia de datos** a dos niveles:

- **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.

En el siguiente gráfico se puede apreciar la estructura de la que estamos hablando:



## 6.4. Tipos de SGBD

Podemos realizar diferentes clasificaciones de SGBD, atendiendo a diferentes criterios:

### Según el modelo lógico en que se basan

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

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

### Según 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.

#### **Según el número de sitios en los que está distribuida la base de datos**

- **Centralizados:** sus datos se almacenan en un solo computador. Los SGBD centralizados pueden atender a varios usuarios, pero el SGBD y la base de datos en sí residen por completo en una sola máquina.
- **Distribuidos:** la base de datos real y el propio software del SGBD pueden estar distribuidos en varios sitios conectados por una red.

#### **Según el propósito o finalidad**

- **Propósito general:** pueden ser utilizados para el tratamiento de cualquier tipo de base de datos y aplicación.
- **Propósito específico:** cuando el rendimiento es fundamental, se puede diseñar y construir un software de propósito especial para una aplicación específica, y este sistema no sirve para otras aplicaciones.

## 7. SGBD comerciales

Actualmente, en el mercado de software existen multitud de sistemas gestores de bases de datos comerciales. En este epígrafe se desglosan las características fundamentales de los más importantes y extendidos hasta la fecha.

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 y considerado "el mejor" puede no serlo para el tipo de proyecto que estemos desarrollando. Hemos de tener en cuenta qué volumen de carga debe soportar la base de datos, qué sistema operativo utilizaremos como soporte, cuál es nuestro presupuesto, plazos de entrega, etc.

A través de la siguiente tabla se exponen los SGBD comerciales más utilizados y sus características más relevantes:

SGBD	Descripción
<b>ORACLE</b>	Reconocido como uno de los mejores a nivel mundial. Es multiplataforma, potente a nivel transaccional, confiable y seguro. Es Cliente/Servidor. Basado en el modelo de datos relacional. De gran potencia, aunque con un precio elevado hace que se vea en empresas grandes y multinacionales. Ofrece versiones gratuitas.
<b>MYSQL</b>	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 bases de datos y plataformas web.
<b>DB2</b>	Multiplataforma, el motor de base de datos 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.
<b>INFORMIX</b>	Otra opción de IBM para el mundo empresarial que necesita un DBMS sencillo y confiable. Es un gestor de base de datos relacional basado en SQL. Multiplataforma. Consume menos recursos que Oracle, con utilidades muy avanzadas respecto a conectividad y funciones relacionadas con tecnologías de Internet/Intranet, XML, etc.
<b>Microsoft SQL SERVER</b>	Sistema Gestor de Base de Datos producido por Microsoft. Es relacional y sólo funciona bajo Microsoft Windows. Utiliza arquitectura Cliente/Servidor. Constituye la alternativa a otros potentes SGBD como son Oracle, PostgreSQL o MySQL.

## 8. SGBD libres

La alternativa a los sistemas gestores de bases de datos 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:

SGBD	Descripción
MySQL	Es un sistema de gestión de base de datos 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.
PostgreSQL	Sistema Relacional Orientado a Objetos. Considerado como la base de datos 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	Sistema Gestor de Base de Datos relacional, multiplataforma, con bajo consumo de recursos, excelente gestión de la concurrencia, alto rendimiento y potente soporte para diferentes lenguajes.
Apache Derby	Sistema Gestor 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.

## 9. Bases de datos centralizadas

Si nos preguntamos cómo es la arquitectura de un sistema de base de datos, 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 base de datos 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 sistemas operativos 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 bases de datos centralizadas son:

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

En la siguiente tabla se representan las ventajas e inconvenientes destacables de esta arquitectura de bases de datos:

Ventajas e inconvenientes de las bases de datos centralizadas.	
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 bases de datos centralizado falla, se pierde toda disponibilidad de procesamiento y sobre todo de información confiada al sistema.

## **Ventajas e inconvenientes de las bases de datos centralizadas.**

<b>Ventajas</b>	<b>Inconvenientes</b>
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 varias computadoras, ya que los trabajos siempre se ejecutarán en la misma máquina.
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.

## 10. Bases de datos 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 bases de datos. La respuesta a estas nuevas necesidades y evoluciones se materializa en los sistemas de bases de datos distribuidas.

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

**Sistema de bases de datos distribuida (SBDD):** es un sistema en el cual múltiples sitios de bases de datos 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.

**Sistema gestor de bases de datos distribuida (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. El término transparente significa que 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 base de datos local, un sistema de gestor de bases de datos 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.

En la siguiente tabla se representan las ventajas e inconvenientes destacables de las BDD:

Ventajas e inconvenientes de las bases de datos distribuidas.	
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.

Ventajas e inconvenientes de las bases de datos distribuidas.	
Ventajas	Inconvenientes
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 bases de datos se adapta más naturalmente a la estructura de las organizaciones. Permiten la incorporación de nodos de forma flexible y fácil.	El intercambio de mensajes y el cómputo adicional necesario para conseguir la coordinación entre los distintos nodos 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.

## 10.1. Fragmentación

Sabemos que en los sistemas de bases de datos distribuidas 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 bases de datos que se encuentran en diferentes servidores. El problema de fragmentación se refiere al particionamiento de la información para distribuir cada parte a los diferentes sitios de la red.

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.

Existen tres tipos de fragmentación:

- **Fragmentación horizontal:** la fragmentación horizontal se realiza sobre las tuplas (filas) de la relación, dividiendo la relación en subrelaciones que contienen un subconjunto de las tuplas que alberga la primera.
- **Fragmentación vertical:** la fragmentación vertical, en cambio, se basa en los atributos (columnas) de la relación para efectuar la división. El objetivo de la fragmentación vertical es particionar 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. Se puede realizar una fragmentación vertical tras realizar una fragmentación horizontal, o viceversa.