

# TEMA 1. INTRODUCCIÓN

Base de Datos CFGS DAW

Francisco Aldarias Raya paco.aldarias@ceedcv.es

2019/2020

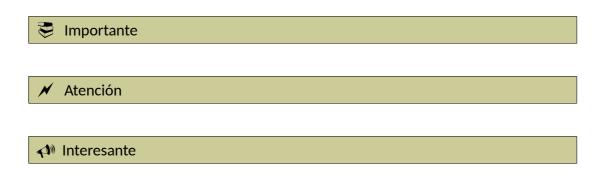
Versión:190917.1053

# Licencia

Reconocimiento - NoComercial - Compartirlgual (by-nc-sa): No se permite un uso comercial de la obra original ni de las posibles obras derivadas, la distribución de las cuales se debe hacer con una licencia igual a la que regula la obra original.

# Nomenclatura

A lo largo de este tema se utilizarán distintos símbolos para distinguir elementos importantes dentro del contenido. Estos símbolos son:



# ÍNDICE DE CONTENIDO

1. Almacenamiento de la información	
2. Ficheros	
2.1 Tipos de Ficheros según su contenido	
2.2 Tipos de Ficheros según su volatilidad	
2.3 Tipos de ficheros según el método de acceso	
3.Bases de datos	7
3.1 Conceptos	
3.2 Tipos de bases de datos según su localización	8
4. Sistema de gestión de bases de datos	11
4.1 Definición	
4.2 Objetivos de los SGBD	11
4.3 Arquitectura en niveles de las bases de datos	11
4.4 Componentes de un SGBD	12
4.5 Modelos de explotación de las bases de datos	
4.6 SGBD comerciales	14
4.7 SGBD libres	15
5.Arquitectura de una aplicación	
6.Bibliografía	

UD01. INTRODUCCIÓN

# 1. ALMACENAMIENTO DE LA INFORMACIÓN

En el mundo actual existe cada vez una mayor demanda de datos. Esta demanda siempre ha sido patente en empresas y sociedades, pero en estos últimos años todavía se ha disparado más debido al acceso multitudinario a Internet.

El propio nombre Informática (información + automática) hace referencia al hecho de ser una ciencia que tiene como objetivo automatizar el trabajo con información.



# Importante

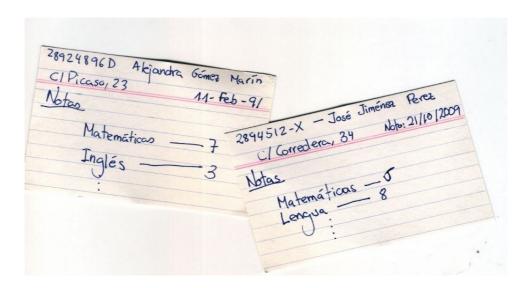
En informática se conoce como dato a cualquier elemento informativo que tenga relevancia para el sistema.

Para poder almacenar datos y cada vez más datos, el ser humano ideó nuevas herramientas como archivos, cajones, carpetas y fichas en las que almacenarlos.

Antes de la aparición del ordenador, el tiempo requerido para manipular esos datos era enorme. Sin embargo, el proceso de aprendizaje era relativamente sencillo ya que se usaban elementos que el usuario reconocía perfectamente.

Por esa razón, la informática adaptó sus herramientas para que los elementos que el usuario maneje en el ordenador, se parezcan a los que utilizaba manualmente. Así en informática se sigue hablado de ficheros, formularios, carpetas, directorios,....etc.

En los años 60, usando papel y "boli"



En los años 70-80 se utilizaban ficheros de texto donde se guardaba la información.

A partir de los años 90 y posteriores, empiezan a utilizarse los Sistemas Gestores de Bases de Datos (SGBD) donde se comienza a almacenar la información en forma de tablas.

```
Alumnos (DNI, Nombre, Dirección, Fecha nacimiento)
2894512X José Jiménez Perez C/ Corredera, 34 21-10-90
28924896D Alejandra Gómez Marín C/ Picaso, 23 11-02-91
Asignaturas (Código, Nombre)
001 Matemáticas
002 Lengua
003 Inglés
Notas(DNI, Código asignatura, nota)
2894512X 001 5
2894512X 002 8
28924896D 001 7
28924896D 003 3
```

Veamos por qué se ha producido esta evolución. ¿Qué ocurriría si quisiéramos conocer el número de alumnos de más de veinticinco años y con nota media superior a siete que además estén matriculados actualmente en la asignatura Bases de datos?

En un Sistema de Información (SI) sin informatizar, obtener está información podía requerir mucho tiempo y mucho trabajo, además de ser necesario la realización de cálculos (media,...) así como, ir revisando alumno por alumno.

En un SI con ficheros, podríamos crear un programa que fuera obteniendo la información del fichero y realizando los cálculos y devolviendo los resultados. Sin embargo, en un SI con bases de datos, esta consulta es trivial usando un lenguaje de consulta de datos.

#### 2. FICHEROS



# Importante

Un fichero (también denominado archivo) es un conjunto ordenado de datos que tienen entre sí una relación lógica y están almacenados en un soporte de información adecuado para la comunicación con el ordenador.

En un fichero se almacena información referente a un mismo tema de una forma estructurada con el fin de manipular los datos de manera individual. Un fichero está compuesto por estructuras

de datos más simples denominadas registros. Todos los registros de un fichero son del mismo tipo, es decir, un fichero está formado por un conjunto de registros homogéneos.

Cada registro está formado por campos que contienen información referente a un elemento o característica en particular dentro del fichero. Así, por ejemplo, los datos de alumnos de un instituto (tema) se pueden almacenar en un fichero, en el que cada registro contendría los campos o datos de cada alumno (número de matrícula, nombre, fecha de nacimiento, lugar de nacimiento, dirección, curso, grupo, aula, etc.).

# 2.1 Tipos de Ficheros según su contenido

Los tipos de ficheros según su contenido son:

- Ficheros de Texto. Se pueden abrir con editor de texto. Necesita separador de campos. Los ficheros de texto suelen llamarse también ficheros planos o ficheros ASCII.
- Binarios. No se puede ver el contenido directamente. No necesita separadores.

Generalmente los ficheros que componen una base de datos son de tipo binario, puesto que la información que hay almacenada en ellos debe tener una estructura lógica y organizada para que las aplicaciones puedan acceder a ella de manera universal, esto es, siguiendo un estándar.

Esta estructura lógica y organizada, generalmente es muy difícil de expresar mediante ficheros de texto, por tanto, la información de una base de datos se suele guardar en uno o varios ficheros:

- El software de gestión de base de datos Oracle guarda la información en múltiples tipos de ficheros, llamados "datafiles", "tempfiles", "logfiles", etc.
- Un tipo de tablas del gestor MySQL guarda su información en 3 ficheros de datos binarios, con extensión .frm, .myd y .myi.
- Access guarda toda la información de una base de datos con extensión "mdb"

# 2.2 Tipos de Ficheros según su volatilidad

Según la función que vaya a desempeñar los ficheros, éstos pueden ser clasificados de varias maneras.

- Permanentes o Fijos. 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.
- Temporales. Se utilizan para almacenar información útil para una parte de la aplicación, no para toda ella, y luego se borran. Son generados a partir de datos de ficheros permanentes. Tienen un corto periodo de existencia.

# 2.3 Tipos de ficheros según el método de acceso

• Ficheros secuenciales. 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.

- Ficheros de acceso directo. En este tipo de ficheros se puede acceder a un registro indicando la posición relativa del mismo dentro del archivo o, más comúnmente, a través de una clave que forma parte del registro como un campo más. Estos archivos deben almacenarse en dispositivos de memoria masiva de acceso directo, como son los discos magnéticos.
- Ficheros indexados. Se basan en la utilización de índices, que permiten el acceso a un registro del fichero de forma directa, sin tener que leer los anteriores. Estos índices son similares a los de los libros. Si nos interesa leer un capítulo concreto podemos recurrir al índice que nos dice en que página comienza, y abrimos el libro por esa página, sin tener que mirar en todas las páginas anteriores para localizarlo.

#### 3. BASES DE DATOS



# Importante

Una base de datos es un conjunto estructurado de datos que representa entidades y sus Interrelaciones.

Una base de datos es un conjunto de datos relacionados y organizados con cierta estructura. Según dicha organización distinguimos entre diferentes modelos de bases de datos como el relacional, jerárquico o en red.

El modelo de bases de datos más extendido es el relacional el cual almacena los datos en forma de tablas.

Estas tablas se relacionan formando vínculos o relaciones entre ellas, que ayudan a mantener la información de diversos objetos de forma ordenada y coherente (sin contradicciones). Cada una de estas tablas es una estructura que se parece a las hojas de cálculo, pues está dispuesta mediante filas y columnas. De este modo, cada fila almacena un registro con tantos campos como columnas tenga la tabla.

Por ejemplo, se podría tener una tabla de Producto, donde cada fila o registro es un producto y cada columna o campo representa un trozo de información sobre cada producto, por ejemplo su código o su descripción.

# 3.1 Conceptos

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. En una base de datos de una clínica, posibles entidades podrían ser: paciente, 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.

# 3.2 Tipos de bases de datos según su localización.

#### Bases de Datos distribuidas

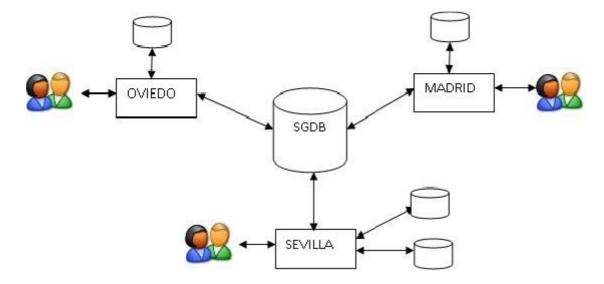
Una Base de Datos distribuida es un conjunto de bases de datos relacionadas lógicamente entre sí que se encuentran localizadas en diferentes ubicaciones físicas. Puede tratarse de dos máquinas conectadas entre sí mediante una o bien a través de Internet. Como consecuencia, dichas Bases de Datos pueden realizar tareas de forma autónoma o bien de forma distribuida. En el caso de las operaciones realizadas de forma distribuida, el usuario de cualquiera de las dos (o más) máquinas tendrá acceso a toda la información como si éstos estuvieran siendo accedidos de forma local.

#### **Ventajas**

- Es más barato crear una red de máquinas de bajo rendimiento que disponer de una sola máquina muy potente
- En cuanto a la disponibilidad, un fallo en cualquiera de los nodos no afecta al resto por lo que el resto de la Base de Datos seguirá accesible
- Puesto que los datos se encuentran distribuidos, éstos pueden ubicarse donde más interesa. Por ejemplo en el departamento con el que tienen relación. Además, este departamento dispondrá de la autonomía necesaria para controlar dicha información

# Desventajas

- Un sistema de Bases de Datos distribuida es complejo de poner en marcha y mantener
- En cuanto a la seguridad, al disponer de varios nodos, se dispone de varios sistemas diferentes que habrá que proteger
- Es un sistema bastante nuevo, por lo que existe poca experiencia ante fallos poco comunes
- Al disponer de varios nodos es posible que se necesite más mano de obra especializada en cada uno de ellos



### 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.

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.

#### Existen tres tipos de fragmentación:

- Fragmentación horizontal: La fragmentación horizontal se realiza sobre las tuplas 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 y la derivada.
- 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 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. 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).

Para representar los dos tipos de fragmentación, se utilizan los árboles

#### Bases de Datos centralizadas

Son el tipo de Bases de Datos más común. Normalmente, en grandes corporaciones, se dedica un único equipo para el almacenamiento y gestión de los datos utilizando algún tipo de SGBD. Toda la Base de Datos se encuentran ubicada lógica y físicamente en ese equipo. Es la solución más habitual puesto que normalmente es suficiente y además es mucho más sencilla de implementar.



#### **Ventajas**

- Fácil implementación. Al tratarse de una solución ampliamente adoptada es más sencillo de implementar y más fácil actuar ante posibles fallos
- Diseño sencillo. Al tratarse de una solución conocida, el diseño lógico es también más sencillo de realizar

# Desventajas

• En caso de fallo, al existir una sola ubicación física de la Base de Datos, éste afectará a todos los datos almacenados en ella

# 4. SISTEMA DE GESTIÓN DE BASES DE DATOS

#### 4.1 Definición

Un sistema de gestión de bases de datos (SGBD) (en inglés database management system, abreviado DBMS) es una coleccion de datos relacionados entre si estructurados y organizados y un conjunto de programas que acceden y gestionan esos datos.

### 4.2 Objetivos de los SGBD

- Permitir consultas no predefinidas y complejas.
- Ofrecer flexibilidad e independencia de datos.
- Minimizar redundancia.
- Garantizar integridad de los datos y referencial.
- Permitir concurrencia de usuarios.
- Proporcionar seguridad de la información.

# 4.3 Arquitectura en niveles de las bases de datos.

El cómite ANSI/SPARC define en 1975 una arquitectura para los sistemas gestores de bases de datos.

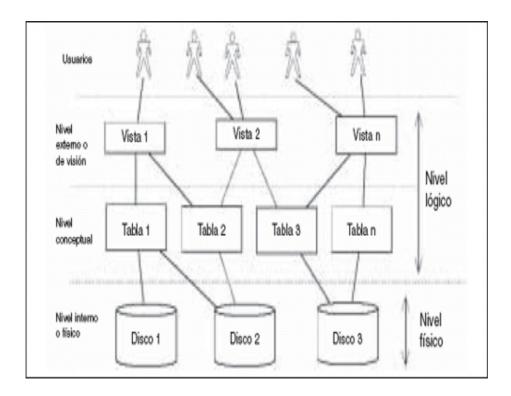
#### Consta de tres niveles:

- Nivel externo o de visión: Se compone de las distintas aplicaciones basadas en vistas de la base de datos. Es lo que ven los usuarios finales.
- Nivel conceptual: Se compone de las distintas tablas con sus atributos. Es el nivel que conocen los programadores.
- Nivel interno o físico: Define qué discos y archivos componen la base de datos y qué hay en cada uno de ellos. Sólo acceden a este nivel los administradores.

La ventaja de esta arquitectura en niveles es que proporciona independencia lógica y física de los datos respecto a las aplicaciones:

- Independencia lógica: Se pueden realizar cambios en el nivel conceptual (añadir tablas o atributos) sin que sea necesario reescribir todas las aplicaciones.
- Independencia física: Es posible modificar la ubicación de los ficheros que contienen los datos sin que se vean afectadas las aplicaciones.

Veamos esta arquitectura con una imagen:



# 4.4 Componentes de un SGBD.

Los SGBD se componen de:

- · Lenguajes.
- El diccionario de datos.
- Mecanismos de seguridad e integridad.
- Factor humano.

Veamos con detalle cada uno de ellos.

# Los lenguajes

Los lenguajes que tenga un SGBD deben permitir:

- Crear la estructura de la base de datos, incluyendo todos los objetos que puede incluir la misma (tablas, vistas, usuarios, procedimientos, funciones, triggers, etc.). Ej: DDL
- Consultar y manipular la información almacenada en la base de datos. Ej.: DML
- Asignar privilegios a usuarios, confirmar o abortar transacciones, etc. Ej.: DCL.
- En algunos casos, también incluyen un lenguaje de cuarta generación (4GL) para RAD (desarrollo rápido de aplicaciones). Ej: Asistentes de Access, Oracle Developer Suite

#### El diccionario de datos.

• El diccionario de datos contiene los metadatos (datos acerca de los datos) de la base de datos, esto es:

- La definición de todos los objetos existentes en la base de datos: tablas con sus columnas, vistas, procedimientos, triggers, índices, etc...
- La ubicación física de los objetos y el espacio asignado a los mismos.
- Los privilegios y roles asignados a los usuarios.
- Las restricciones de las tablas.
- Información de auditoría.
- Estadísticas de uso de la base de datos.
- Información del consumo de recursos actual.

# Mecanismos de seguridad e integridad.

Un SGBD debe proporcionar utilidades que permitan:

- La realización de copias de seguridad de los datos y la restauración de las mismas.
- Garantizar la protección de los datos ante accesos no autorizados.
- Implantar restricciones de integridad de los datos para evitar daños accidentales de los datos.
- Recuperar la base de datos hasta un estado consistente en caso de error del sistema o cualquier otro imprevisto.
- Controlar el acceso concurrente de los usuarios para evitar errores de integridad.

#### El factor humano

Un SGBD siempre va a tener distintas categorías de usuarios:

- Usuarios finales: Podrán acceder a la información sobre la que le hayan sido concedidos privilegios.
- Programadores: Realizan aplicaciones sobre los objetos de la base de datos para facilitar su trabajo a los usuarios finales.
- Administradores o DBAs: Garantizan el correcto funcionamiento de la base de datos y
  gestionan todos sus recursos. Tienen el nivel más alto de privilegios y responsabilidades
  legales en caso de que los datos tengan algún tipo de protección. Su objetivo es que la base
  de datos está siempre disponible y con un rendimiento óptimo.

# 4.5 Modelos de explotación de las bases de datos.

En nuestro entorno podemos encontrar los SGBD implantados de diferentes formas:

- Monopuesto: La base de datos se encuentra en una máquina y es explotada desde la misma máquina. Típico en SGBD de escritorio: Access, OpenBase, Sqlite, Mysql.
- Cliente/Servidor: El SGBD está en una máquina pero se accede a él desde muchas usando, por lo general, distintas aplicaciones.

• Grid de servidores: La base de datos está en distintas máquinas que trabajan colaborativamente para dar servicio a los clientes.

- BD distribuida: La información está en distintos servidores, pero no trabajan como una única máquina.
- Capas: Cliente  $\rightarrow$  Servidor web  $\rightarrow$  (Servidor de aplicaciones)  $\rightarrow$  Servidor de BD.

#### 4.6 SGBD comerciales

Se ha de tener en cuenta que sus aplicaciones deben estar sustentadas en un sistema que ofrezca garantías, pero que se ajuste a sus necesidades, dimensiones y presupuesto.

Actualmente, en el mercado de software existen multitud de sistemas gestores de bases de datos comerciales. Veremos los más importantes.

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 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
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	istema 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.
MICROSOFT SQL SERVER	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.

#### 4.7 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 verán los más utilizados actualmente:

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.
POSTGRES	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.
SQLITE	Sistema relacional, basado en una biblioteca escrita en C que interractua 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.

# 5. ARQUITECTURA DE UNA APLICACIÓN

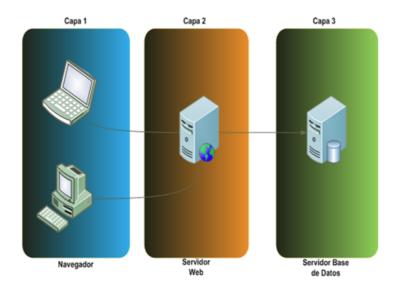
Una vez que hemos visto qué es un SGBD, una Base de Datos y los diferentes tipos con los que nos podemos encontrar justo con sus ventajas e inconvenientes, conviene echar un vistazo de una manera más global para ver cuál es la posición de un SGBD y la Base de Datos en el conjunto de una aplicación ya desarrollada.

Como ya se ha visto más arriba, muchos de los SGBD trabajan siguiendo lo que se conoce como arquitectura cliente-servidor. Siguiendo esta misma filosofía, existe la *programación por capas*, que consiste en una arquitectura en la que el principal objetivo es la separación de las diferentes capas lógicas de una aplicación (presentación, negocio y datos). Por ello, la arquitectura más conocida es la arquitectura de 3 capas, en la que la estructura de la aplicación se separa en tres niveles. Aunque no siempre las 3 capas están separadas físicamente, es conveniente que lo estén lógicamente. La separación física vendrá dada por la propia naturaleza de la aplicación (aplicación web) o bien por las propias necesidades en cuanto a rendimiento de la misma (uno o más equipos para dar soporte a la capa de negocio o datos).

- Capa de presentación Es la capa que ve el usuario. Presenta la aplicación al usuario, le permite interactuar con el mismo y le muestra la información que éste solicita. En definitiva es lo que se conoce como interfaz de usuario.
- Capa de negocio En esta capa se encuentra toda la lógica de la aplicación y será donde se realicen todos los procesos de tratamiento de la información obtenida en la capa de datos y cuyos resultados se muestran al usuario mediante la capa de presentación.

• Capa de datos En esta capa residen los datos. En ella se encuentran los SGBD y reciben peticiones para acceder o escribir datos desde la capa de negocio.

En aplicaciones de tamaño medio es bastante habitual encontrar un modelo lógico basado en 3 capas pero físicamente distribuido solamente en 2. En estos casos lo habitual será que el usuario cargue la capa de presentación en su propio equipo (ya sea ejecutando la aplicación o cargando la web en su navegador) y que las capas de negocio y datos se encuentren en otro equipo remoto (el servidor). A medida que la aplicación se haga más exigente las capas de negocio y datos pueden separarse en diferentes equipos (normalmente físicamente más cerca) y según se exige un mayor rendimiento se pueden añadir equipos a cualquiera de estas dos capas.





# 6. BIBLIOGRAFÍA

- 1. Administración de sistemas gestores de bases de datos. CFGS. Hueso Ibañez, Luis Editorial Rama. Año 2011. ISBN 978-84-9964-100-3.
- 2. Bases de Datos. Iván López, M.ª Jesús Castellano. John Ospino. 2017 Ed. Garceta.

3.