

1.- Sistemas de almacenamiento de la información

[1.0.] Introducción

Según la RAE, la definición de sistema es “*Conjunto de cosas que ordenadamente relacionadas entre sí contribuyen a un determinado objeto*”.

En el caso de una **gestión electrónica de la información** (lo que actualmente se considera un **sistema de información electrónico**), los componentes son:

- o **Datos.** Se trata de la información relevante que almacena y gestiona el sistema de información. Ejemplos de datos son: *Sánchez, 12764569F, Calle Mayo 5, Azul...*
- o **Hardware.** Equipamiento físico que se utiliza para gestionar los datos: servidores, máquinas de los clientes, routers, switches, impresoras,...
- o **Software.** Aplicaciones informáticas que se encargan de la gestión de la base de datos y de las herramientas que facilitan su uso.
- o **Recursos humanos.** Personal que maneja el sistema de información.

Podemos hablar de dos tipos fundamentales de sistemas de información electrónicos:

[1.0.1] sistemas de gestión de ficheros

Este tipo de sistemas hace referencia a la forma que inicialmente se desarrolló en la informática para gestionar ficheros (y que aún se usa).

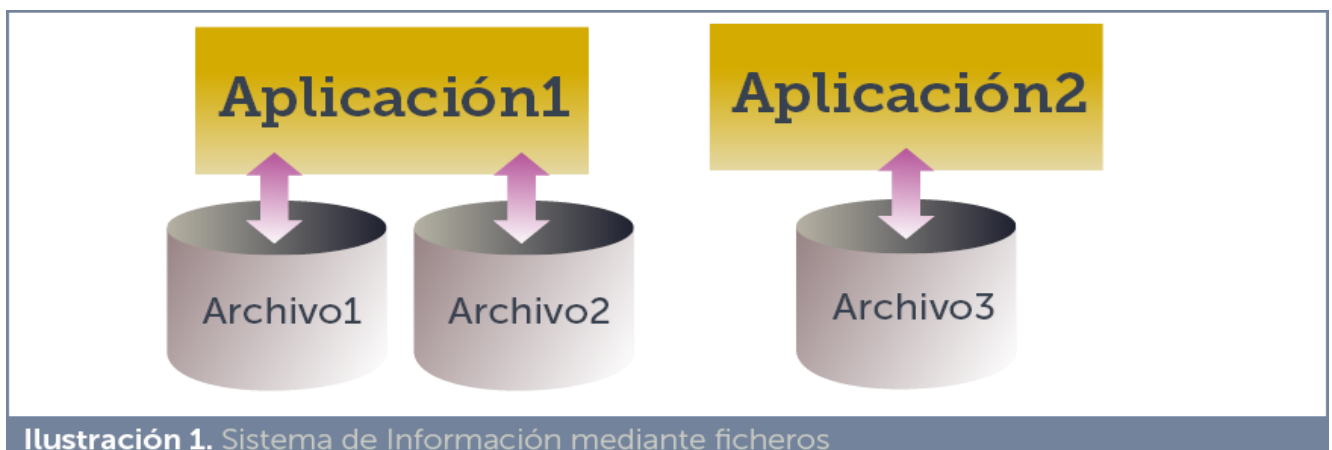


Ilustración 1. Sistema de Información mediante ficheros

Los datos se almacenan en ficheros y **se crean aplicaciones para acceder a los ficheros**. Hasta que una aplicación no deja de utilizar un fichero (lo cierra), este no puede ser usado por otra aplicación.

[1.0.2] sistemas de bases de datos

En este tipo de sistemas, los datos se centralizan en una **base de datos** común a todas las aplicaciones. Un software llamado **Sistema Gestor de Bases de Datos (SGBD)** es el que realmente accede a los datos y se encarga de gestionarlos.



Cuando una aplicación modifica un dato, la modificación será visible inmediatamente para el resto de aplicaciones.

[1.1.] Ficheros

[1.1.1] conceptos

Los **ficheros o archivos** son estructuras de información que crea el sistema operativo para poder almacenar datos en un soporte externo (disco duro, dvd, pen drive,...).

Para identificarlos se utiliza un nombre y una extensión que determina la forma de interpretar la información que guardan, ya que realmente lo único que se almacena en un fichero es una ristra de bits (ceros y unos).

Un fichero de datos está formado por un conjunto de datos, de igual o distinto tipo, agrupados en **registros lógicos**.

Registro lógico: conjunto organizado de datos que contienen información referente a un determinado objeto o concepto. A cada uno de estos datos se les llama **campo**. A cada campo se le debe asignar un nombre, un tipo y un tamaño.

Campo clave: permite identificar al registro diferenciándolo de todos los demás. No puede existir dos registros cuyo campo clave tengan los mismos valores.

Ejemplo:

Imaginemos que tenemos el fichero CLIENTES.DAT organizado de la siguiente manera:

n_cuenta	nombre	saldo	n_cuenta	Nombre	Saldo	
----------	--------	-------	----------	--------	-------	-------	--

Reg. Lógico 1

Reg. Lógico 2

Campo clave: n_cuenta.

[1 .1. 2] clasificación de ficheros

a) según su función en el tiempo

a.1. Permanentes: contienen información necesaria para el correcto funcionamiento de toda una aplicación informática. Su vida es tan larga como la de la aplicación.

Dependiendo de la frecuencia con la que se actualicen podemos distinguir:

- **Ficheros Maestros o de Situación:** son aquellos en los que la información está constantemente actualizada.
Ejemplo: un fichero de alumnos
- **Ficheros Constantes:** son aquellos cuyo contenido permanece invariable.
Ejemplo: fichero de asignaturas.
- **Ficheros Históricos:** contienen los datos acumulados a lo largo del tiempo de los archivos actualizados. Se utilizan para hacer estadísticas o informes.

a.2. Temporales: contienen información necesaria para el buen funcionamiento de un proceso o de un programa. Duran lo mismo que el proceso o el programa para el que se utiliza.

- **Fichero de movimiento o transacciones:** son los que contienen información necesaria para actualizar los ficheros maestros.

Ejemplo: fichero de biblioteca con los préstamos y devoluciones de los libros que se han hecho en un día completo.

- **Fichero de trabajo o intermedio:** almacenan datos o resultados auxiliares que van a ser utilizados por otros procesos o programas.

Ejemplo: si tenemos un fichero de préstamos y devoluciones desordenadas y queremos imprimir un listado ordenado por un campo.

b) según el modo de organizar sus registros

b.1 Secuenciales

Son aquellos en los que los registros se van almacenando en **posiciones físicas contiguas en el mismo orden en que fueron introducidos**.

Sólo permiten el **acceso secuencial**, es decir, para acceder a un determinado registro hay que leer todos los anteriores.

26	L.J.F	3000	17	R.S.T.	8000	20	F.M.P.	12000
----	-------	------	----	--------	------	----	--------	-------

b.2. Relativos

La posición física ocupada por cada registro en el soporte **depende del valor del campo clave**. Permiten el **acceso directo** a sus registros. Es decir, con solo indicar su clave se puede acceder directamente a un registro, sin necesidad de leer todos los anteriores. Lo único que necesitamos saber es el tamaño de cada registro, que en este tipo de ficheros debe de ser el mismo.



- **Organización directa:** el campo clave es numérico y coincide con la posición que ocupa ese registro en el soporte.

Direcciones	Datos	
.....		
17	R.S.T	8000
.....		
20	F.M.P	12000
21		
22		
.....		
26	L.J.F	3000

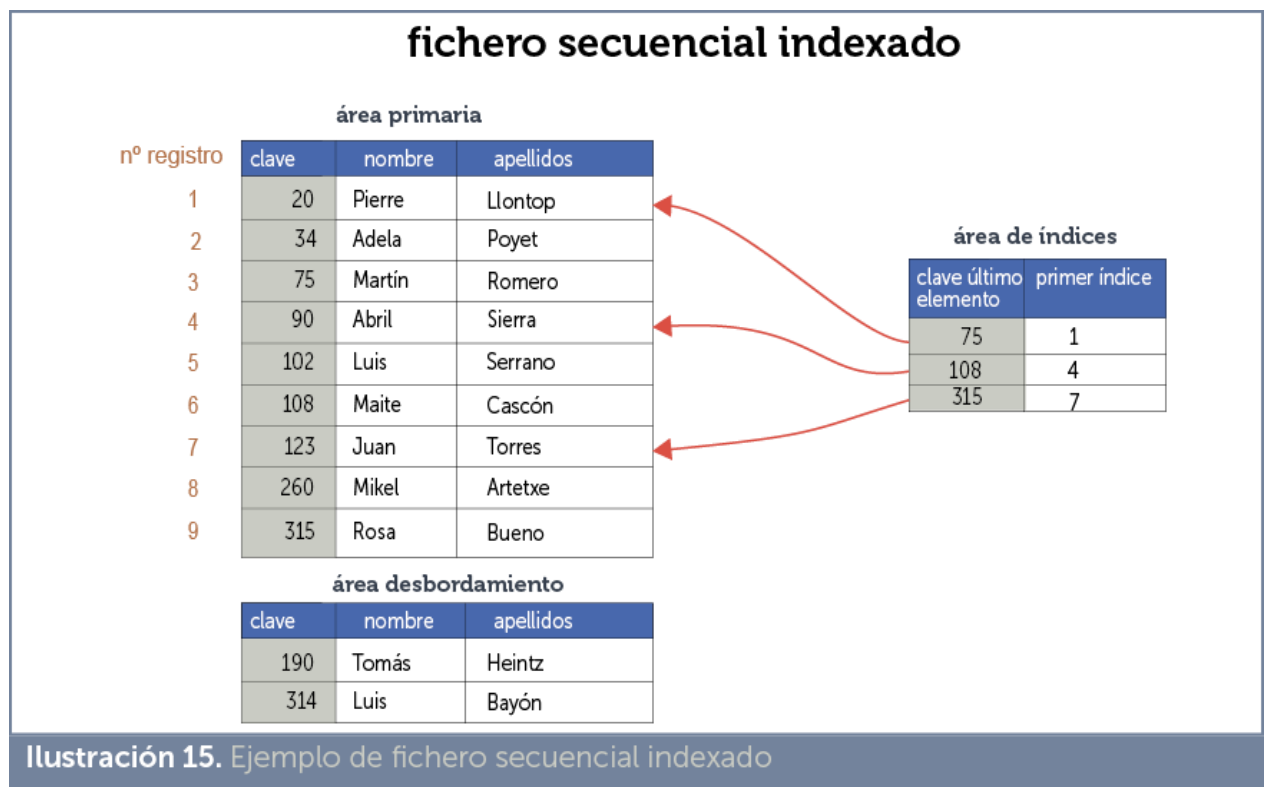
- **Organización aleatoria:** para obtener la dirección de un registro hay que realizar un algoritmo matemático que transforme su clave.

Direcciones	Clave	Datos	
1		
2		
3	20P	F.M.P	12000
4		
5	17F	R.S.T	8000
6		
7	26F	L.J.F	3000

b.3 Secuenciales Indexados

Su estructura se compone de tres áreas:

- o **Área Primaria.** Es un fichero secuencial ordenado que guarda los registros escritos durante la creación del fichero. Se encuentra dividido en bloques de n registros.
- o **Área de índices.** Es un fichero secuencial que contiene un registro por cada bloque del área primaria. Cada registro está formado por dos campos: el primero es la clave del último registro de cada bloque y el segundo almacena la dirección relativa de comienzo de cada bloque.
- o **Área de Overflow o Desbordamiento.** Contiene los registros almacenados posteriormente a la creación del fichero, por lo que no han sido incluidos en el área primaria. Este archivo está desordenado. Se utiliza sólo si se busca un registro y no se encuentra en el área primaria. En ese caso se recorre todo el archivo de overflow hasta encontrarlo. Para no tener demasiados archivos en overflow, cada cierto tiempo se reorganiza el archivo principal, ordenando los datos en el orden correcto y recalculando el archivo de índices.



Para acceder a un determinado registro se siguen estos pasos:

1. Se toma su clave y se comprueba en la tabla de índices entre que dos valores de los que figuran en la misma se encuentra dicha clave seleccionando aquella que sea superior a la clave dada. Se comprueba la dirección de almacenamiento asociada a dicho valor.
2. Se accede a esa dirección en el área primaria y se busca secuencialmente a partir de ella hasta localizar el registro o bien hasta que el número de clave sea la última del grupo que se ha obtenido de la tabla de índices.
3. Si la clave buscada no existe en el área primaria se busca secuencialmente en el área de overflow.

c) según su contenido

c.1 Ficheros de texto

Suelen también llamarse ficheros planos o ficheros ASCII.

ASCII (American Standard Code for Information Interchange) es un código estándar que asigna un valor numérico a cada carácter, con lo que se puede representar los documentos de texto plano, es decir, los que son directamente legibles.

Cada carácter se representa con un byte.(8 bits)

Con n bits se pueden generar 2^n combinaciones distintas de caracteres. Como 1 byte = 8 bits existen 256 (2^8) caracteres en la tabla de códigos ASCII numerados del 0 al 255.

Los ficheros de texto suelen tener extensiones que indican que tipo de texto se almacena. Por ejemplo:

- o **F. de configuración.** Contienen un texto sobre la configuración del sistema operativo o de alguna aplicación. Utilizan extensiones **.ini, .inf o .conf.**
- o **F. de código fuente.** Su contenido es el texto de un programa informático. Ejemplos: **.sql, .c, .java.**
- o **F. de páginas web.** Las páginas web son ficheros de texto con hipertexto (forma de escritura no secuencial que permite al lector elegir que secuencia seguir) que interpreta el navegador. Ejemplos: **.html, .php, .css, .xml.**
- o **Formatos enriquecido.** Textos que contienen códigos de control para visualizar un texto más elegante: **.rtf, .tex**

c.2 Ficheros binarios

Son todos los que no son de textos y precisan un formato para ser interpretados. Algunos tipos son:

- o *de imagen*: **.jpg, .gif, .bmp, .png**
- o *de video*: **.mpg, .avi, .mov, .qt**
- o *comprimidos*: **.zip, .tar, .gz**
- o *ejecutables*: **.exe, .com, .cgi**
- o *procesadores de texto*: **.doc, .odt**

Los ficheros que componen una base de datos son de tipo binario, ya que la información que almacenan debe tener una estructura lógica y organizada difícil de expresar mediante ficheros de texto. Su información se suele guardar en varios ficheros.

En MySQL, para cada tabla son necesarios 3 ficheros binarios con extensiones *frm*, *myd* y *myi*. Sin embargo en Access toda la información se almacena en un archivo con extensión *accdb*.

▪ [1.2.] Bases de Datos

Una **base de datos** puede definirse como una colección de datos interrelacionados, almacenados en una sola estructura global, independientes de los programas que los utilizan y sin redundancias innecesarias. Se puede acceder a sus datos de diversa forma y por múltiples usuarios al mismo tiempo.

Una base de datos está organizada mediante tablas que almacenan información sobre algún suceso u objeto. Entre todas ellas se crean unas relaciones que permiten mantener la información ordenada y sin contradicciones.

La estructura de una **tabla** es similar a una hoja de cálculo, pues se compone de filas y columnas. Cada columna almacena un campo y cada fila almacena un registro con tantos campos como columnas tenga la tabla.

[1.2.1] conceptos

- **Dato:** información concreta sobre algún concepto o suceso. Ejemplo: 1992 indica un año de nacimiento.
- **Tipo de dato:** indica la naturaleza del dato. Así hablamos de *datos numéricos* con los que se pueden realizar cálculos aritméticos y *datos alfanuméricos* que contienen caracteres y dígitos numéricos.
- **Campo:** identificador de una familia de datos que debe pertenecer a un tipo determinado. Ej: el campo FechaNacimiento representa las fechas de nacimiento de las personas que hay en la tabla. Pertenecen al tipo de datos Fecha.
Al campo también se le llama **columna**.
- **Registro o tupla:** recolección de datos referentes a un mismo concepto o suceso. Por ejemplo, los datos de una persona pueden ser su NIF, nombre, dirección, etc. También se le llama **fila**.
- **Campo clave :** campo especial que identifica de forma única a cada registro. Ejemplo: el NIF es único para cada persona.
- **Tabla:** conjunto de registros bajo un mismo nombre que representa el conjunto de todos ellos. Por ejemplo, los registros de los clientes de una base de datos se almacenan en una tabla llamada Clientes
- **Consulta:** instrucción para hacer peticiones a una BD. Puede ser una búsqueda de uno o varios registros que cumplan unas condiciones o una petición (query) de eliminación o inserción de registros.
- **Índice:** estructura que almacena los campos claves junto con su posición en la BD. Esto permite encontrar de una forma rápida un determinado registro.
- **Vista:** transformación que hace de una o más tablas para obtener una nueva tabla virtual (se almacena su definición pero no su contenido en un dispositivo de almacenamiento)
- **Informe:** listado ordenado de los registros y campos seleccionados (facturas impagadas del mes de mayo ordenado por nombre del cliente)
- **Guiones o scripts:** Conjunto de instrucciones que, ejecutadas de forma ordenada, realizan operaciones de mantenimiento de los datos almacenados en la base de datos.
- **Procedimientos:** tipo especial de script almacenado en la BD y que forma parte de su esquema.

[1 .2. 2] estructura de una BBDD

Una BBDD almacena la información a través de un *esquema*.

El **esquema** es la definición de la estructura donde se almacenan los datos y contiene todo lo necesario para organizar la información mediante tablas, registros y campos.

También contiene otros objetos necesarios para el tratamiento de los datos, tales como procedimientos, vistas o índices. Contiene todas las definiciones pero ningún dato.

Al esquema también se le llama metainformación (información sobre la información) o metadatos.

Una **instancia** es la información recopilada en una base de datos en un momento específico (también se conoce como **estado**). Cada vez que los datos se insertan o eliminan de la base de datos cambian el estado de la base de datos por lo que una instancia de la base de datos cambia con mucha frecuencia.

Diferencia clave entre el esquema y la instancia

Un esquema es la representación de diseño de una base de datos, mientras que instancia es la instantánea de una base de datos en un momento determinado.

[1 .2. 3] usos de las BBDD

Algunos de los usos más frecuentes son:

- o **Administrativas:** almacenan clientes, proveedores, facturas,.. de una empresa
- o **Contables:** pagos, declaraciones de hacienda, perdidas y ganancias,...
- o **Científicas:** datos climáticos y medioambientales, geológicos, ..
- o **Censos:** información de demográfica de pueblos, ciudades o países
- o **Bibliográficas:** biblioteca de un instituto
- o **Configuraciones:** datos de configuración de un sistema informáticos para motores de búsqueda. Google tiene una BD con información sobre todos los documentos de Internet. Los usuarios buscan en la BD de estos motores

[1.2.4] tipos de BBDD

a) según el modelo de datos

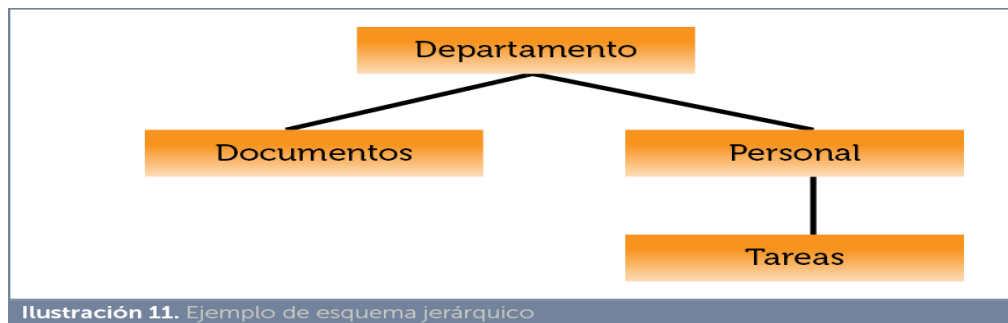
Un **modelo de datos** es un conjunto de conceptos que sirven para describir la estructura de una base de datos: los datos, las relaciones entre los datos y las restricciones que deben cumplirse sobre los datos.

a.1 Modelo jerárquico.

Utiliza árboles para la representación lógica de los datos.

Existe una tabla como punto de entrada a todos los datos (**nodo padre**) a partir de la que vamos ramificando nuestro camino para acceder al resto de datos que precisamos (**nodos hijos**).

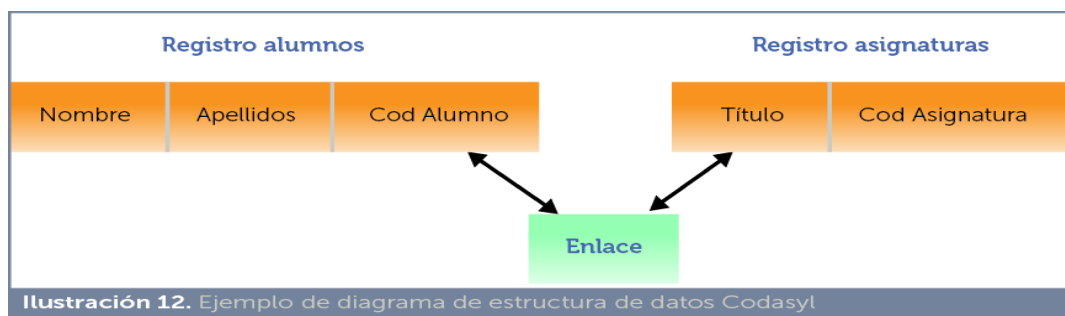
Cada nodo es una entidad o tabla y mantiene relaciones jerárquicas con sus nodos hijos que pueden variar en número.



a.2. Modelo en red

Utiliza estructuras tipo red para la representación lógica de los datos.

El modelo en red organiza la información en **nodos (registros)** y **enlaces**. En los *nodos* 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.



a.3 Modelo relacional

Los datos se organizan en **tablas** constituidas por filas y columnas. Existen **relaciones** entre las tablas.

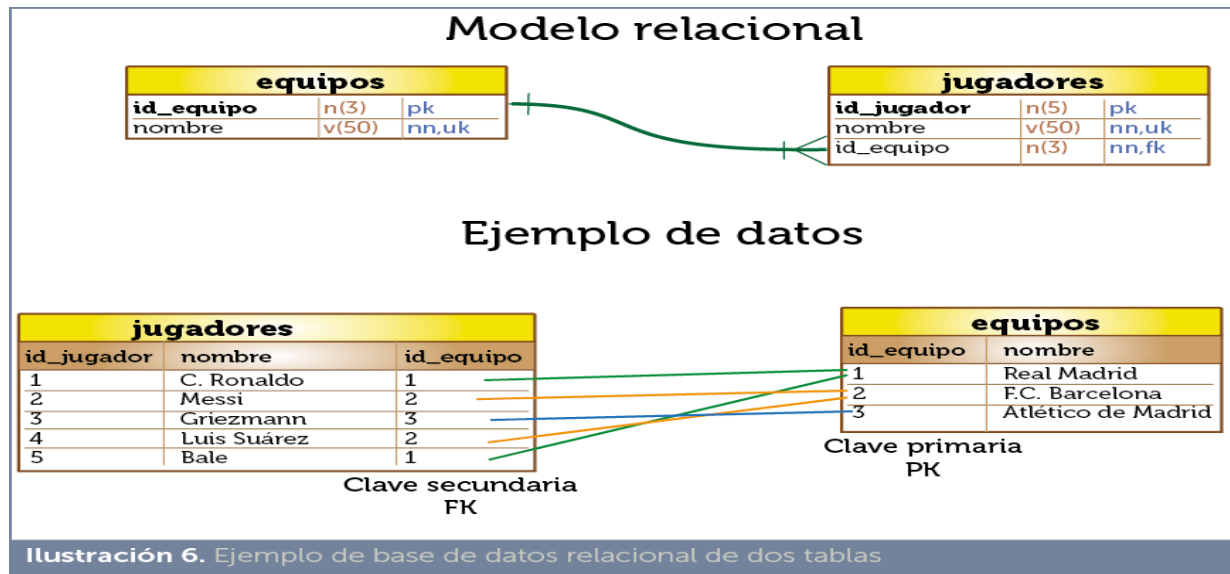


Ilustración 6. Ejemplo de base de datos relacional de dos tablas

a.4 Modelo orientado a objetos

En lugar de almacenar tablas y relaciones, se almacenan colecciones de objetos que serán procesados mediante programas orientados a objetos

Algunos SGBD relacionales existentes en el mercado han estado extendiendo sus modelos para incorporar conceptos orientados a objetos. A estos SGBD se les conoce como sistemas **objeto-relacionales**.

b) según la ubicación de la información

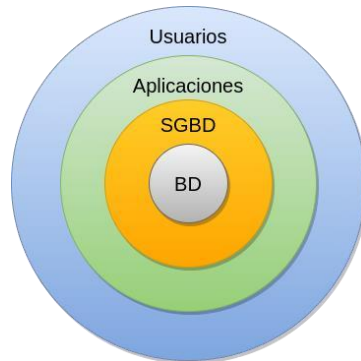
b.1 Bases de datos centralizadas: existe un único servidor en el que se instalan todos los servicios relacionados con la gestión de los datos, mientras que el resto de máquinas (*clientes*) usan dichos servicios.

b.2. Bases de datos distribuidas: En este tipo se multiplica el número de ordenadores que controlan una base de datos (*servidores*), intercambiando información y actualizaciones a través de la red.

• [1.3.] Sistemas Gestores de Bases de Datos

Un **SGBD** (**DBMS** = Data Base Management System) consiste en un conjunto de herramientas que facilitan la consulta, uso y actualización de una base de datos.

Todas las peticiones que se realizan a la base de datos se llevan a cabo de forma centralizada a través del SGBD por lo que podemos considerar a este como una interfaz entre las bases de datos y los usuarios.



[1 .3. 1]funciones de un SGBD

Para conseguir la eficiencia requerida un SGBD debe poseer las siguientes características:

- o Permitir a los usuarios almacenar, acceder y actualizar datos de forma sencilla y con gran rendimiento.
- o Garantizar la integridad de los datos, respetando las reglas y restricciones establecidas.
- o Proteger los datos contra accesos no autorizados.
- o Proporcionar un **diccionario de datos** que contenga el esquema de la base de datos y al que se pueda acceder de forma sencilla.
- o Permitir el uso correcto de transacciones, pudiendo deshacer los cambios en caso de incidencia. (Transacción: conjunto de operaciones relacionadas que se ejecutan como una unidad. O se ejecutan todas las operaciones o no se ejecuta ninguna.)
- o Suministrar mecanismos de seguimiento de las operaciones efectuadas, permitiendo así analizar posibles funcionamientos erróneos.
- o Solucionar los problemas de concurrencia.(varios usuarios trabajando a la vez sobre los mismos datos).
- o Independizar los datos de las aplicaciones que los utilizan, facilitando su migración a otras plataformas.
- o Ofrecer conectividad con el exterior para poder replicar y distribuir bases de datos.
- o Incorporar mecanismos para realizar copias de seguridad y restablecer un estado anterior de la base de datos.
- o Ofrecer herramientas para la importación y exportación automática de la información.

[1.3.2] componentes de los SGBD

Podemos catalogar a los componentes de un SGBD en toda su amplitud (es decir incluyendo también al elemento humano) de la siguiente manera:

- ✓ Núcleo y servidores de datos.
- ✓ Lenguajes.
- ✓ Utilidades
- ✓ Diccionario de datos
- ✓ Los recursos humanos.

[1.3.2.1] núcleo y servidores de datos

Conjunto de programas que coordinan y controlan el funcionamiento del SGBD.

- o El **núcleo** de la base de datos es el encargado de traducir todas las peticiones y prepararlas para su correcta interpretación por parte del sistema operativo
- o Los **servidores** poseen módulos que realizan funciones de sistema operativo, de forma que no haya que cerrar las aplicaciones para realizar acciones sobre los datos.

[1.3.2.1] lenguajes (SQL)

La principal herramienta de un SGBD es la interfaz de programación del usuario, que consiste en un lenguaje sencillo con el que el usuario pregunta al servidor y éste contesta a sus demandas.

Este lenguaje se denomina SQL (Structured Query Language) y está estandarizado por la ISO (International Organization for Standardization).

Se divide en 4 sublenguajes:

- o Lenguaje **DML** o **lenguaje de manipulación de datos** (Data Manipulation Language). Permite seleccionar, insertar, modificar y borrar datos
- o Lenguaje **DDL** o **lenguaje de definición de datos** (Data Definition Language) Permite crear toda la estructura de una BD
- o Lenguaje **DCL** o **lenguaje de control de datos** (Data Control Language). Permite al administrador gestionar el acceso a los datos de la BD
- o Lenguaje **TCL** o **lenguaje de control de transacciones**. Permite ejecutar varios comandos como si fuera un comando indivisible.

[1 .3. 2. 3] utilidades

Son aplicaciones que facilitan el trabajo a usuarios y programadores: generadores de menús, de informes y de formularios.

[1 .3. 2. 4] diccionario de datos

Contiene información sobre los datos de la base. Es lo que se llama una meta-base, ya que describe a las bases de datos almacenadas.

Entre otros incluye el esquema de la base, las tablas de usuarios con sus permisos, los programas de aplicación que utilizan, las operaciones que realizan dichos usuarios,...

[1 .3. 2. 5] recursos humanos

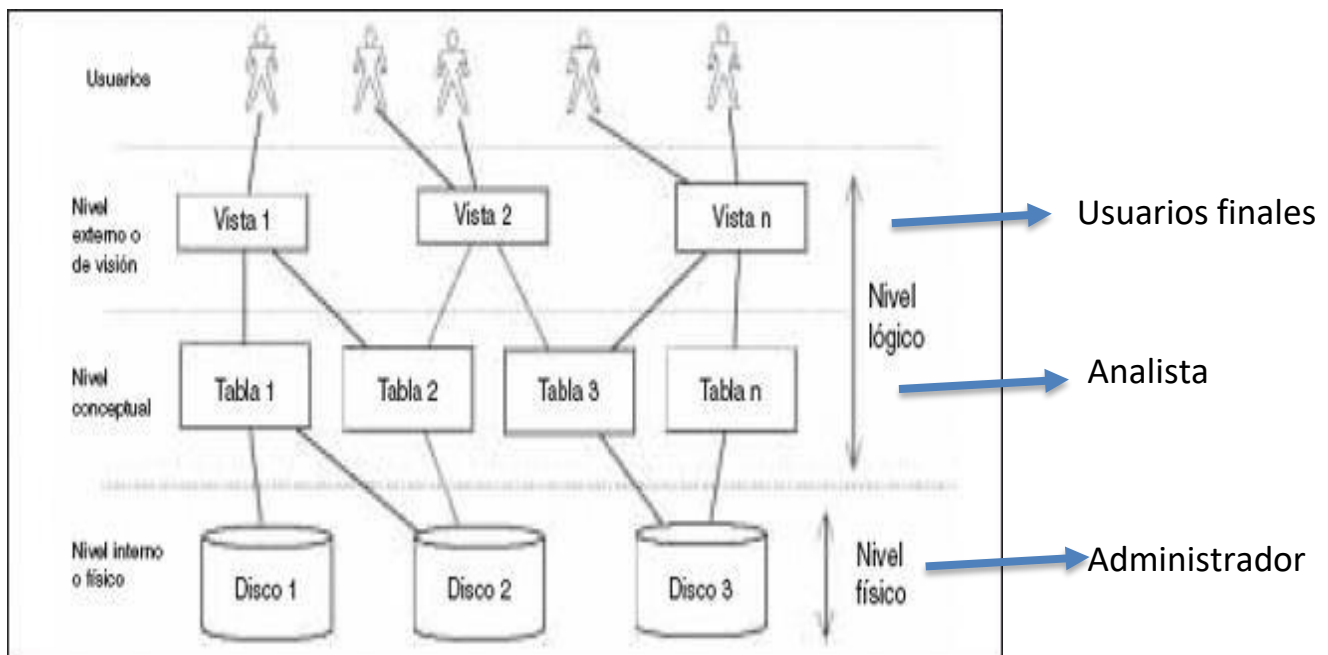
Entre las personas que intervienen en el desarrollo y uso de una BD se encuentran:

- o **Analistas**. Normalmente son los diseñadores de la base de datos y los directores de la programación de la misma.
- o **Desarrolladores o programadores**. Encargados de la realización de las aplicaciones de usuario de la base de datos.
- o **Administradores de la base de datos**. Encargados de crear el esquema interno de la base de datos. Sus tareas incluyen la planificación de copia de seguridad, gestión de usuarios y permisos y creación de los objetos de la base de datos.
- o **Usuarios finales** son los clientes de la base de datos. Podemos diferenciar distintas categorías: ☐
 - **Usuarios expertos**. Son usuarios que utilizan la base de datos para gestión avanzada de decisiones. Conocen y utilizan el lenguaje DML.
 - **Usuarios habituales**. Utilizan las aplicaciones creadas por los desarrolladores para consultar y actualizar los datos. Son los que trabajan en la empresa con la base de datos.
 - **Usuarios ocasionales**. Acceden a la base de datos a través de una aplicación que permite consultar ciertos datos.

[1.3.3] arquitectura de un SGBD

La arquitectura más habitual utilizada para los SGBD es la que hace una división en niveles de la base de datos de forma que se consideran tres niveles dependiendo de la perspectiva desde la que sea vista la información:

- o **Nivel interno:** Se refiere a la forma en la que realmente se almacena la información de la base de datos.
- o **Nivel conceptual:** Aquí se definen los datos que intervendrán en el sistema.
- o **Nivel externo:** Reproduce la visión individual del usuario o grupo de usuarios de la base de datos.



[1 .3. 4] tipos de SGBD

a) Según su capacidad y la potencia

- **Gestores de Bases de Datos ofimáticas.** Manipulan BD domésticas o de pequeñas empresas. Permiten construir pequeñas aplicaciones para manipular los datos de forma sencilla. Un ejemplo es Microsoft Access
- **Gestores de Bases de Datos corporativas** de grandes o medianas empresas que con una gran carga de transacciones y datos requieren un servidor de grandes dimensiones. Oracle es un ejemplo típico.
- **Gestores intermedios** entre gestores de BD ofimáticas y corporativas. MySQL es un gestor además de gratuito muy sencillo y capaz de manipular gran cantidad de datos.

b) Según el número de usuarios a los que da servicio

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

c) Según el propósito

Los SGBD pueden ser **de propósito general** o **de propósito específico**. Cuando el rendimiento es fundamental, se puede diseñar y construir un SGBD de propósito especial para una aplicación específica, y este sistema no sirve para otras aplicaciones. (sistemas de reservas de líneas aéreas)

▪ [1.4.] Bases de datos distribuidas

Cuando una base de datos está controlada por más de un servidor se dice que está distribuida. Estos servidores están interconectados mediante una red de telecomunicaciones y los SGBD proporcionan mecanismos para poder consultar la información independientemente del servidor y de la ubicación de los datos.

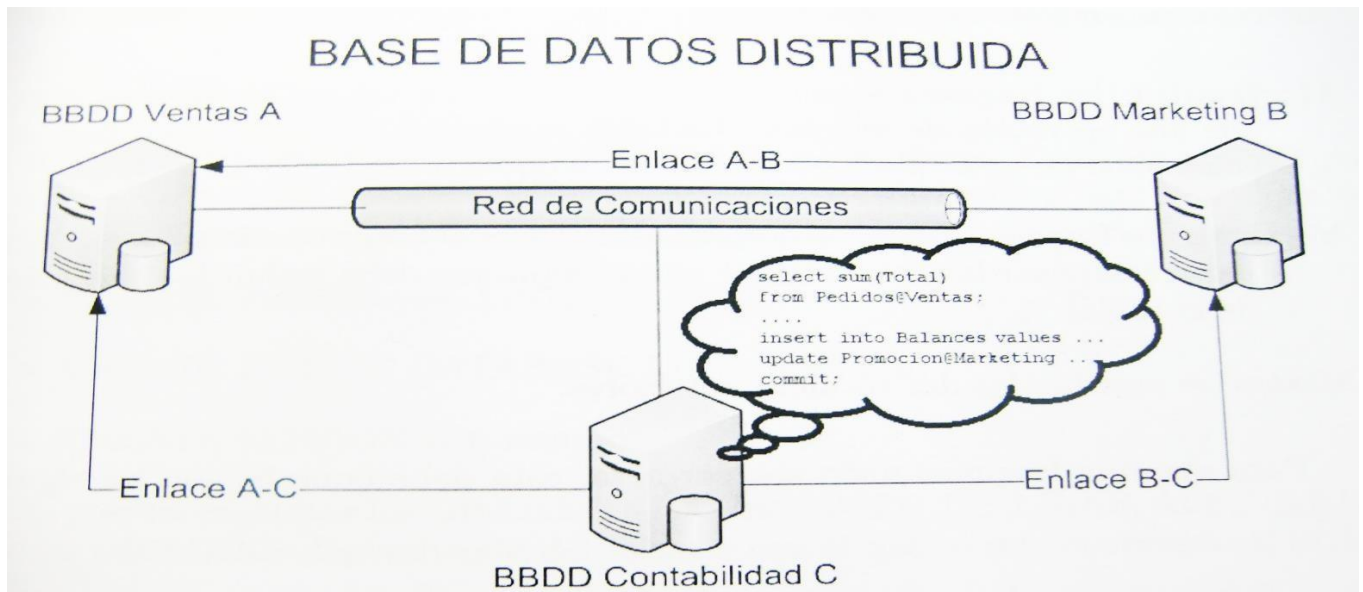
Las ventajas son:

- o **Mayor rapidez en el acceso a los datos y capacidad de almacenamiento.** Al aumentar el número de ordenadores donde se depositan los datos aumenta la capacidad de computación y, por tanto, la capacidad de búsqueda de información.
- o **Acceso más flexible.** El número de usuarios conectados que puede mantener el sistema se incrementa al haber más capacidad de cómputo y, gracias a la red de la BBDD distribuida, los usuarios pueden acceder desde más ubicaciones.
- o **Escalabilidad.** Distribuyendo una base de datos en distintas ubicaciones se puede incrementar fácilmente los recursos en cualquier momento según las necesidades.
- o **Fiabilidad y capacidad de tolerancia a fallos.** Realizando una buena política de fragmentación de la base de datos se consigue que, aunque la base de datos distribuida sufra algún percance, el resto del sistema siga funcionando con normalidad.

[1.4.1] componentes de una BBDD distribuida

Para poder crear una base de datos distribuida se necesitan estos componentes:

- o Sistemas gestores con capacidades distribuidas o un DSGBD distribuido en sí mismo.
- o Un sistema de bases de datos locales a cada uno de los servidores que compondrán la BBDD.
- o Una red de comunicaciones (basada en el protocolo TCP-IP) para poder aprovechar las infraestructuras existentes
- o Enlaces entre las BBDD locales, para que cualquier consulta pueda ser redirigida al servidor donde está la BBDD local.
- o Un diccionario de datos global, indicando donde están cada uno de los datos de la BBDD distribuida.



[1.4.2] técnicas de fragmentación

El contenido de una tabla puede separarse en varios fragmentos con los que podrá reconstruirse la tabla completa cuando sea necesario. Cada fragmento se encontrará en un nodo diferente, sin duplicarse los registros, de forma que sólo existirá una copia de cada elemento. Esto implica la disminución del coste de mantenimiento en detrimento de la disponibilidad y la fiabilidad de los datos en caso de caída de algún nodo.

La fragmentación puede ser:

- o **Horizontal:** los fragmentos son subconjuntos de una tabla y se definen a través de una operación de selección. La tabla original se reconstruye en base a una operación de unión de los fragmentos componentes.
- o **Vertical:** los fragmentos son subconjuntos de los atributos con sus valores. Para poder recomponer la tabla original cada fragmento debe incluir la clave primaria de la tabla.
- o **Mixta:** se mezclan las dos técnicas anteriores.

Para que una fragmentación sea correcta debe cumplir las siguientes reglas:

- ✓ Debe ser completa: cada elemento de la tabla debe estar en algún fragmento.
- ✓ Debe ser reconstruible: se puede conseguir una tabla completa a partir de la combinación de todos los fragmentos.
- ✓ Los fragmentos deben ser disjuntos: un elemento que está en un fragmento concreto no puede estar en ningún otro fragmento.
- ✓ Sistemas gestores con capacidades distribuidas o un DSGBD distribuido en si mismo.

▪ [1.5.] Legislación sobre seguridad

En la actualidad, el derecho que tiene cualquier ciudadano a conocer los datos que cualquier organización sabe de nosotros, a rectificarlos o a eliminarlos se recoge en la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

Esta ley nos indica qué podemos hacer con los datos de los trabajadores y usuarios de nuestros sistemas. Por ese motivo es fundamental conocerla desde la óptica de la seguridad informática.

El texto completo de la ley se puede consultar en este enlace [Ley Orgánica de Protección de Datos y garantía de los derechos digitales](#) y aquí el [Reglamento Europeo de Protección de Datos](#).

▪ [1.6.] Big Data

El término **Big Data** hace referencia a conjuntos de datos extremadamente grandes y complejos que superan la capacidad de las herramientas de software tradicionales para capturar, almacenar, administrar y procesar datos dentro de un tiempo razonable. El concepto de Big Data se caracteriza por las "tres V":

- **Volumen**: Se refiere a la cantidad masiva de datos generados o recopilados. Pueden ser datos de redes sociales, de registros de transacciones, de datos de sensores, etc...
- **Variedad**: Los datos pueden ser de diferentes tipos (texto, imágenes, videos, registros de eventos, etc)
- **Velocidad**: Los datos pueden generarse o actualizarse a una velocidad muy alta.

Durante los últimos años han surgido varias V más:

- **Veracidad**: La confiabilidad y precisión de los datos pueden variar. Algunos datos pueden ser inexactos.
- **Valor**: El objetivo final del análisis de Big Data es extraer información útil y conocimientos que sean valiosos para la toma de decisiones

El Big Data se encuentra en muchas áreas de la vida cotidiana:

- ✓ Redes Sociales: Generan enormes cantidades de datos continuamente. Las empresas analizan estos datos para comprender el comportamiento de los usuarios y personalizar la experiencia.
- ✓ Comercio Electrónico: Las tiendas on-line utilizan Big Data para rastrear el comportamiento de compra de los clientes y recomendar productos personalizados. También optimizan la gestión de inventario y la logística.
- ✓ Ciencia: Los proyectos científicos a gran escala, como la investigación en astrofísica, genómica y climatología, generan enormes cantidades de datos que se analizan para descubrir patrones y tendencias.

Trabajar con Big Data requiere de tres acciones clave:

1. Integración

Big data concentra datos de numerosas fuentes y aplicaciones distintas. Es necesario incorporar los datos, procesarlos y asegurarse de que estén formateados y disponibles para que puedan utilizarse.

2. Gestión

Big data requiere almacenamiento. Su solución de almacenamiento puede residir en la nube, en servidores o en ambos. Se suelen utilizar marcos de trabajo tales como Spark y Hadoop o bases de datos NoSQL

3. Análisis de Datos

Implica procesar y examinar las grandes cantidades de datos almacenados con el fin de identificar patrones, tendencias, relaciones y cualquier información relevante que pueda ser beneficiosa para la toma de decisiones y la generación de valor.

▪ [1.7.] Inteligencia de negocios

También conocida como **Business Intelligence (BI)**, se define como el conjunto de tecnologías, procesos y herramientas que permiten a las empresas recopilar, analizar, presentar y visualizar datos empresariales con el objetivo de tomar decisiones informadas y estratégicas. La inteligencia de negocios ayuda a convertir los datos brutos en información valiosa que puede utilizarse para mejorar la eficiencia operativa, optimizar procesos, identificar tendencias, entender el rendimiento empresarial y respaldar la toma de decisiones.

Business Intelligence y *Big Data* son dos tecnologías muy diferentes entre sí, pero totalmente complementarias en el mundo empresarial actual

Algunas **diferencias entre ambas tecnologías son:**

- El *Big Data* procesa todo tipo de datos, mientras el *Business Intelligence* únicamente trabaja con datos estructurados, es decir, almacenados en bases de datos SQL.
- En *Business Intelligence* se opera a través de bases de datos, mientras que con *Big Data* se ponen en marcha diversas tecnologías que permiten el tratamiento masivo de datos en paralelo y simultáneamente a través de sistemas distribuidos.
- Los expertos que se suelen requerir para trabajar en *Big Data* pasan por campos como las matemáticas, la estadística, la ingeniería o la informática. En el caso del *Business Intelligence*, los perfiles profesionales solicitados para este campo pertenecen a áreas como la administración de empresas, la economía o el marketing.