

# T01: Introducción a bases de datos. Elementos.

## 1. Introducción.

En el entorno actual se acumula una mayor demanda de datos que nunca antes, por lo tanto tenemos la necesidad de **gestionarlos, la información es poder**.

En informática se conoce como **dato** a cualquier **elemento informativo que tenga relevancia para el usuario**. Antiguamente la información se almacenaba en ficheros, cajones, carpetas y fichas de cartón. El proceso era enorme.

Con la aparición de la **informática** se adaptaron sus herramientas para que el usuario tuviera más **comodidad y orden** al gestionar la **información**. Los sistemas de información actuales se basan en **Bases de Datos y Sistemas Gestores de Bases De Datos** que se han vuelto imprescindibles en la sociedad moderna.

## 2. Definición de Base de Datos.

La mayoría de nosotros nos encontramos con actividades que **requieren** usos de **Bases De Datos** (ingresar dinero en el banco, comprar entradas a un concierto...) Estos son ejemplos de lo que se llama **aplicaciones tradicionales de Bases De Datos**, las cuales trabajan con **información numérica o de texto**.

Existen **otras** Bases De Datos como:

- **Base de Datos Multimedia:** sus contenidos consisten en imágenes, fragmentos de videos, sonidos (Tik Tok).
- **Base de Datos Espaciales o Geográficas:** almacenan mapas y símbolos que representan superficies geográficas (Google Maps).
- **Base de Datos Documentales:** permite crear índices de texto para poder realizar búsquedas entre un texto de gran tamaño de forma más rápida.
- **Bases de Datos Deductivas:** almacenar hechos y reglas por las que se rigen dichos hechos, lo cual permite mediante procedimientos de inferencia extraer nuevos hechos.

Podemos definir **datos** como cualquier hecho conocido que **puede registrarse** y que tiene un **significado implícito**.

Y por **Base de Datos** a una colección de datos pertenecientes al **mismo contexto** que se almacenarán en **ficheros**, estas se organizan en **tablas** relacionadas las cuales recopilan datos de **cada objeto o suceso** del que trata el **Sistema de Información**.

Estas definiciones hacen referencia a **dos elementos** para que un conjunto de datos se convierta en una Base de Datos:

- **Relaciones** entre tablas.
- **Significado implícito** de los datos que se atribuyen dependiendo del contexto en que se utilizan los mismos.

### *3. Evolución histórica y tipos de bases de datos.*

#### *3.1 Introducción.*

Los **predecesores** de los (SGBD) fueron los **Sistemas Gestores de Ficheros** o Tradicionales.

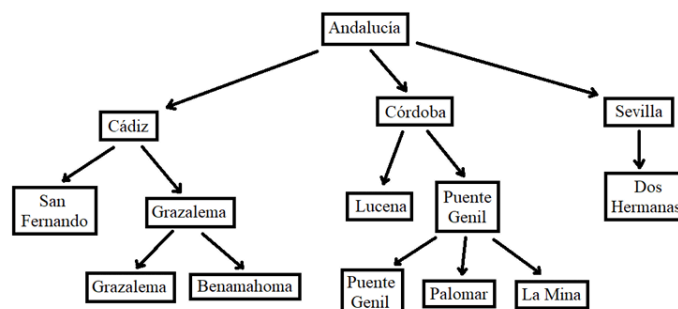
1. **Ficheros Tradicionales:** Consiste en **almacenar datos en ficheros individuales, exclusivo** para cada **aplicación particular**, en este sistema los datos **pueden repetirse innecesariamente** y **actualizar** dichos datos es muy **lento**.
2. **Base de Datos:** Es un almacenamiento de datos formalmente **definido, controlado** para servir a **múltiples y diferentes aplicaciones**.

En el tradicional la información está **más dispersa** entre diferentes ficheros y existen cierto número de programas que los **recuperan y agrupan**, aunque los sistemas de ficheros supusieron un gran avance sobre los sistemas manuales tenían bastantes **inconvenientes** que **se solucionaron con las Bases De Datos**.

#### *3.2 Evolución y tipos de Bases De Datos.*

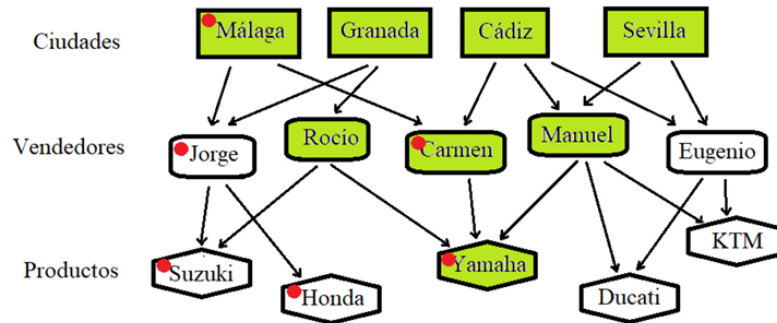
##### ★ **Base de datos Jerárquica**

Fue usada en las **primeras** Bases de Datos. Las relaciones entre registros forman una **estructura de árbol**, debajo observamos un ejemplo sobre la estructura de alguna población de Andalucía.



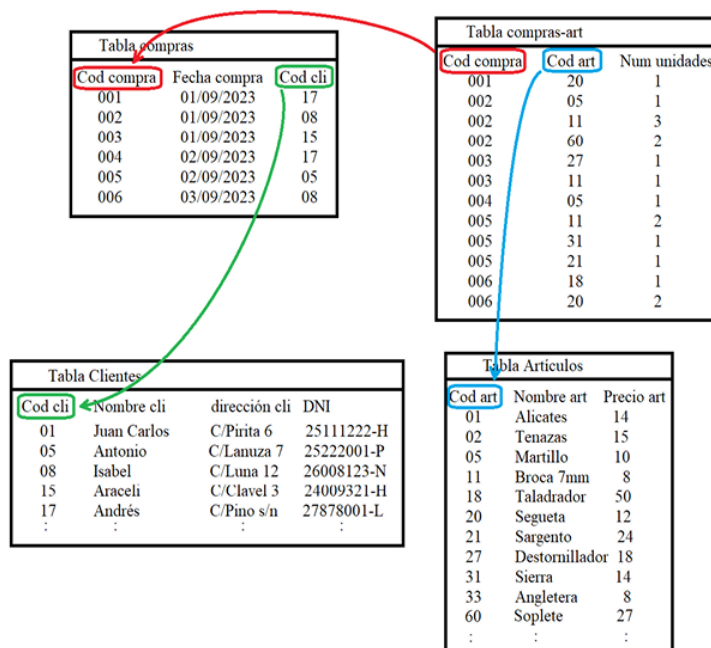
## ★ Base de datos Red.

Contiene relaciones **más complejas** que las jerárquicas, admite **relaciones** de los elementos **de un determinado nivel** con más de uno **de nivel superior**, permitiendo así **caminos alternativos**. Su inventor fue **Charles Bachman** y el estándar fue publicado en **1969** por **CODASYL**, ahora veremos un ejemplo de el.



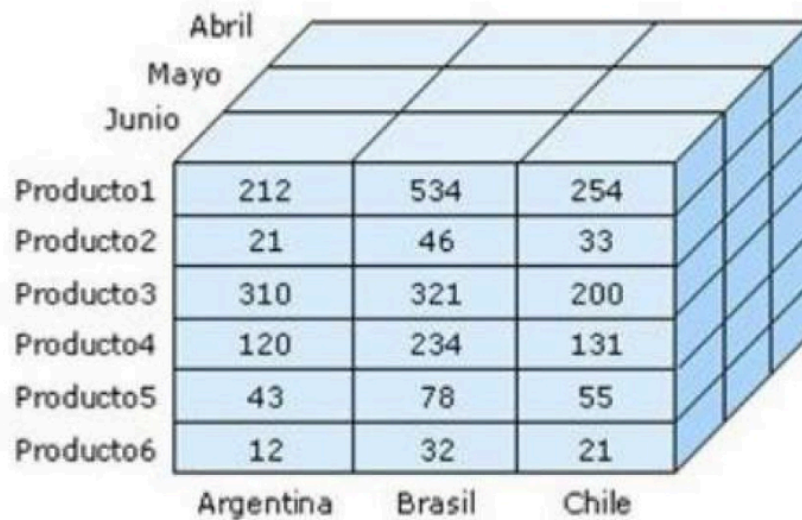
## ★ Base de datos Relacional.

Es la **más extendida hoy en día**, almacena los datos en filas o registros (**tuplas**) y columnas o campos (**atributos**). Estas pueden estar conectadas entre sí por **claves comunes**.



### ★ Base de datos con estructura Multidimensional.

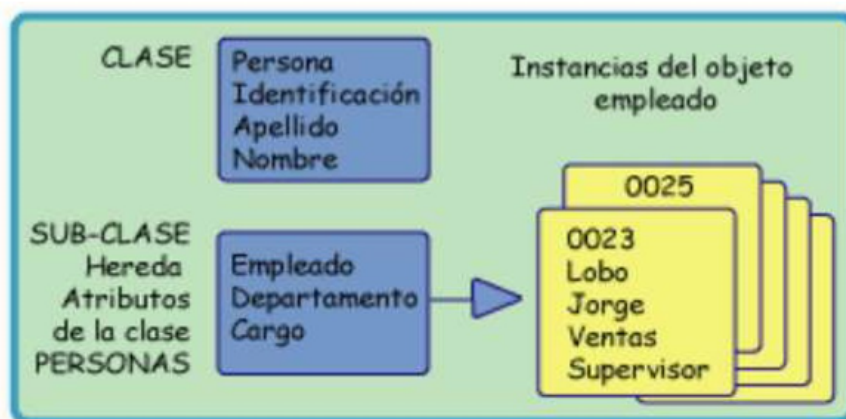
Tiene parecidos al **modelo relacional** pero en vez de dos dimensiones (filas-columnas), tiene **N dimensiones**, se le denomina **cubo**, fue una solución dado al problema de trabajar con **enorme volumen de datos** resultantes de años de información, esta se puede ver así.



	Abril		
	Mayo		
	Junio		
Producto1	212	534	254
Producto2	21	46	33
Producto3	310	321	200
Producto4	120	234	131
Producto5	43	78	55
Producto6	12	32	21
	Argentina	Brasil	Chile

### ★ Bases de Datos Orientada a Objetos.

Está fue diseñada para el siguiente el paradigma de los **lenguajes orientados a objetos**, de este modo soporta los tipos de **datos gráficos, imágenes, voz y texto de manera natural**, tiene gran difusión en **aplicaciones web** para **aplicaciones multimedia**.



### 3.3 Sistemas de fichero Tradicionales

Cada programa almacena y utiliza sus propios datos, su unica ventaja es que son procesos independientes, con lo cual la modificación no afecta al resto, pero estos tienen grandes inconvenientes:

- **Datos redundantes:** Se repiten continuamente en cada aplicación que se utilice.
- **Coste de almacenamiento elevado:** Al almacenarse varias veces, se requiere mas espacios en los discos.
- **Tiempos de procesamiento elevados:** Al no poder optimizar el almacenamiento, si existen varias copias hay que tenerlas todas en cuenta.
- **Probabilidad alta de inconsistencia en los datos:** Por errores humanos un datos al modificarse puede hacer que tenga valores distintos en dos copias.
- **Difícil modificación en los datos:** Cada modificación se debe repetir en todas las copias del dato.

Aquí un ejemplo de información basada en ficheros:

JUGADORES		PORTEROS		BARCELONA		REAL_MADRID		JUG_LIGA_PASADA		JUG_LIGA_ACTUAL	
PK	Codjugador	PK	Codjugador	PK	Codjugador	PK	Codjugador	PK	Codjugador	PK	Codjugador
	Club jugador dorsal codpaís codDem		club jugador dorsal CodPaís NombDem		Club jugador dorsal codpaís codDem		Club jugador dorsal codpaís codDem		Club jugador dorsal codpaís codDem		Club jugador dorsal codpaís codDem
CLUBES		CLUB_JUG		ESTADIOS		EQUIPACIÓN		SOCIOS		ESPONSOR	
PK	CodClub	PK	CodJugador	PK	CodEstadio	PK	CodEquip	PK	Numsocio	PK	CodEspon
	Club Nombre Dirección Población Provincia CosPostal Tífono Colores Himno Fax AñoFundación presupuesto Presidente Vicepresidente CodEquip Sponsor		CodClub		CodEquipo Estadio Dirección CodPostal Población Provincia Capacidad Sentados Inauguración Dimensiones		NombreEquip Encasa		CodClub Codpost Nombre Apellidos Dirección Provincia FechaAlta CuotaActual		Esponsor
		ENTRENADORES				PAÍSES					
		PK	CodEntren			PK	CodPaís				
			CodClub Entrenador FechaNcmto Población Provincia				Nombrepais Comunitario				
		ENTREN_CLUBES		CAMPEONATO		PALMARÉS		DEMARCACIÓN			
		PK	codCluB	PK	CodTrofeo	PK	CodClub CodTrofeo	PK	CodDem		
			CodEntrenador		NombreTrofeo		Año		Demarcación		

### ★ Fichero secuencial.

Cada registro se graba en el espacio dedicado al fichero según va surgiendo, el fichero empieza en el primer registro que se escribió y termina en el último registro guardado.

Características:

- Todo el **espacio** ocupado corresponde a **registros**.
- Para acceder a un determinado registro es necesario **empezar desde el principio** y leer **uno a uno**.
- Es difícil de **mantener** ya que para **añadir, eliminar o modificar** registros la mayoría de veces hay que **reescribir** el fichero.

### ★ Fichero indexado.

Se caracteriza por tener dos áreas diferenciadas:

- **Área Índice:** contendrá sólo los códigos de los registros ordenados y un enlace al lugar exacto del disco duro donde está almacenado el registro completo. (se solicita **clave**, se busca en el **índice** y se va directamente mediante el **enlace**).
- **Área de Datos:** Contiene los datos según van apareciendo (permite acceso **secuencial** y **aleatorio**, utiliza más espacio al depender de un índice), solo se puede utilizar en discos magnéticos.

### ★ Fichero directo.

Se ha de elegir un **campo clave** para acceder a los registros, a través de esa clave se utiliza un **algoritmo** para indicar **la posición del fichero** (Ej. letra Dni [ejercicio 2.12 prog]).

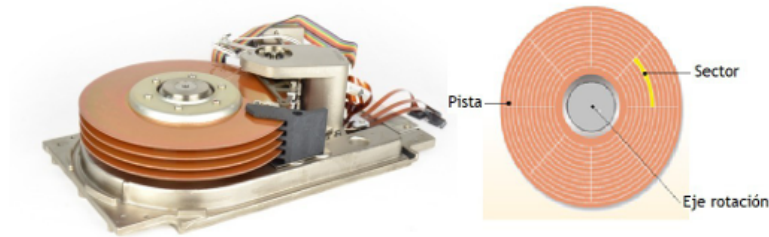
Se caracteriza por, tener **dos áreas definidas (datos y colisiones)**, **estimar** un número de registros que se esperan y **reservar espacio para sus zonas de colisión**, sólo se permite el **acceso directo** y que a **mayor cantidad de colisiones, más huecos** habrá en la zona de datos y esto hará que se **relentice**.

### ★ Cinta Magnética.

Similar a un **casete**, se escribe una **marca de inicio** de fichero, su **contenido** y luego una **marca final**, esto **permite escribir más de un fichero** en la misma cinta.

## ★ Disco Magnético.

Se organiza así: pista, superficie y sectores.



El brazo de lectura selecciona el sector del disco, ya que este anda girando constantemente, esto crea un acceso aleatorio, que rápidamente se posiciona, lee o escribe.

### *3.5 Sistema de Base de Datos Relacional.*

#### **Ventajas:**

- **Menor redundancia** (no hace falta tanta repetición de datos).
- **Menor** espacio de **almacenamiento**.
- Acceso a los datos más **eficiente**.
- **Documentación** de datos.
- **Independencia** de los datos respecto a programas y procesos.
- **Integridad** de los datos.
- Mayor **seguridad** en los datos.

#### **Desventajas:**

- **Instalación costosa** ya que se requiere un **software** y **hardware potente**.
- **Personal cualificado**, debido a la dificultad de manejo de este tipo de sistemas
- **Implantación** larga y difícil.

## *4. Conceptos básicos de una Base De Datos.*

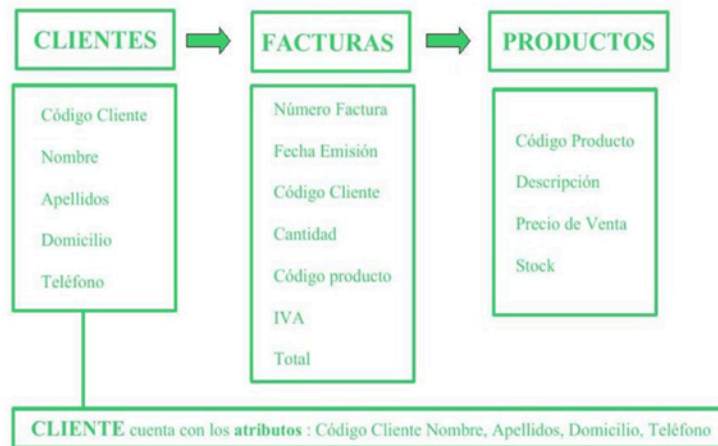
### *4.1 Datos.*

Los datos son hechos conocidos que deben registrarse y que tienen un significado implícito.

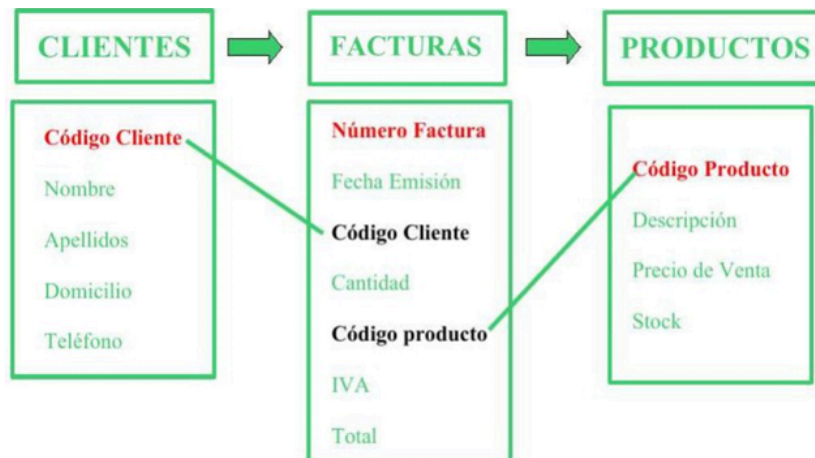


## 4.2 Entidades.

Es todo aquello de lo que interesa guardar datos. por ejemplo: clientes, facturas, productos...



## 4.3 Claves primarias y claves foráneas. Relaciones.



- **Claves Primarias:**

- **Código cliente** es la clave primaria de **CLIENTES** (a cada cliente se le asocia un código y a cada código le corresponde un cliente).
- **Número Factura** es clave primaria de **FACTURAS**.
- **Código Producto** es clave primaria de **PRODUCTOS**.

- **Claves Foráneas:**

- En **FACTURAS**, son claves foráneas **Código Cliente** y **Código Producto**. **CLIENTES** se relaciona con **FACTURAS** a través del **Código Cliente** que figura en ambas tablas y con **PRODUCTOS** mediante el **Código Producto**.



#### 4.4 Restricción de integridad referencial.

Cuando **dos tablas** están **relacionadas**, se cumple la **restricción de integridad referencial** cuando **todos** los **valores** que toman el atributo que es **clave foránea** en una de ellas, está representado como clave primaria en la otra tabla.

#### 4.5 Metadatos.

Son **datos** acerca de los datos **presente** en la **Base de Datos**.

- Qué **tipo de datos** se van a **almacenar** (texto, números, fechas).
- Qué **nombre** se le asigna a cada dato.
- Cómo están **agrupados**.
- Cómo se **relacionan**.
- Qué **valores** puede tomar.

El tipo de datos puede ser Numérico, alfabético, fecha, lógico (Si /NO).

Dato	Tipo	Longitud
Num	Numérico	4
Nombre	Alfabético	20
.....	.....	.....

La longitud indica la cantidad máxima de caracteres que puede asumir el dato.

**Ejemplo de Restricción de Dominio:**  
Num >0 y <9999

#### 4.6 Otros conceptos de Bases de Datos.

CLIENTES				
Código cliente	Nombre	Apellidos	Domicilio	teléfono
12	Juan	García López	C/Pino 27	952 121212
13	Inés	Martínez Durán	C/Camino del pinar 7	636 077000
14	Luis	Moreno Sánchez	C/Pirita 31	956 090990

→ Tupla (registro)

Atributo (campo)

Dato

- **Tabla (fichero):**

Conjunto de filas y columnas bajo el mismo nombre que representa el conjunto de valores almacenados para una serie de instancias de una entidad.

- **Atributo (campo):**

Es cada una de las columnas de una tabla.

- **Tupla (registro):**

Corresponde a cada una de las filas de la tabla.

- **Tipo de Dato:**

Indica **la naturaleza del campo**, es decir, qué **contenidos** puede almacenar y qué **operaciones** se pueden hacer con ellos, podemos tener datos **numéricos, aritméticos, caracteres, fechas ...**

- **Consulta:**

Es una **instrucción** para hacer **peticiones** a una Base de Datos.

- **Índice:**

Estructura que **almacena** los campos claves de una tabla, **organizándolos** para hacer más fácil **encontrar y ordenar** los **registros**.

- **Vista:**

Se obtienen al **guardar** una **consulta** de una o varias tablas, de esta forma se obtiene una **tabla visual** la cual no está almacenada en los dispositivos de almacenamiento del ordenador.

- **Informe:**

**Listado ordenado** de los campos y registros seleccionados en un formato **fácil de leer**.

- **Guiones o Scripts:**

**Conjunto de instrucciones**, que ejecutadas de forma ordenada, realizan operaciones avanzadas o de mantenimiento de los datos.

- **Procedimientos:**

Son un tipo especial de **Script** que están **almacenados** en la Base de Datos y forman parte de sus esquemas.

## *5. Sistemas de Gestión de Bases de Datos.*

### *5.1 Sistema Gestor de Base de Datos.*

Es una **aplicación** que permite a los usuarios **definir, crear y mantener** una Base de Datos.

- Permite la **definición** de la Base de Datos mediante el lenguaje de definición de datos (Data Description Language). Este lenguaje permite especificar la **estructura, tipo y restricciones** de los datos.
- Permite la **inserción, actualización, eliminación y consulta** de datos mediante el lenguaje de manejo o manipulación de datos (Data Manipulation Language).
- Acceso **controlado** mediante:
  - **Sistema de seguridad.**
  - Sistema de **integridad y consistencia** de datos.
  - Sistema de **control de concurrencia y gestión de transacciones.**
  - Sistema de **control de recuperación.**
- Diccionario de **datos** o **catálogo** contiene la descripción de los datos de la Base de Datos y es accesible para los usuarios.

### *5.2 Tipos de SGBD.*

Se pueden **clasificar** según el **tipo** de Base de Datos que **gestionen**:

- **SGBD ofimáticos:** **manipulan** Bases de Datos pequeñas orientadas a almacenar **datos domésticos** o de **pequeñas empresas.**
- **SGBD corporativos:** tienen la capacidad de **gestionar** Bases de Datos **enormes de medianas o grandes empresas**, con una carga de datos y transacciones que requieren un **servidor de gran capacidad.**
- **SGBD intermedios:** Es **gratuito, fácil de instalar**, permite trabajar con **volúmenes considerables** de datos y está **integrado** en paquetes XAMPP (windows,linux...).

# Ejercicios

**(1) Para cada una de las siguientes cuestiones elige razonadamente las respuestas correctas.**

I ¿Cuáles de los siguientes puntos representan inconvenientes de los Sistemas de Ficheros?

- a. Redundancia e Inconsistencia.
- b. Sistema de Gestión de Datos independiente de la máquina y del SO.
- c. Control de concurrencia.
- d. Dificil modificación de los datos. (v)

II Los sistemas orientados a BD presentan las siguientes ventajas...

- a. Integridad de los datos.
- b. Redundancia.
- c. Cada aplicación maneja sus propios datos.
- d. Independencia entre los datos y las aplicaciones que los usan. (v)

III Los datos son...

- a. Todo aquello de lo cual interesa guardar información.
- b. Hechos conocidos que pueden registrarse y que tienen un significado implícito. (v)
- c. Información acerca de los metadatos.
- d. Las claves primarias y foráneas de cada entidad.

IV Un SGBD ...

- a. Está formado por datos acerca de los datos presentes en la Base de Datos.
- b. Es una aplicación que permite a los usuarios definir, crear y mantener una Base de Datos, y proporciona acceso controlado a la misma.
- c. Permite a los usuarios tener acceso a la BD completa impidiendo restricciones.
- d. Permite la inserción, actualización, eliminación y consulta de datos mediante el lenguaje de manejo o manipulación de datos. (v)

**(2) Explica brevemente los antecedentes de las BD actuales.**

Los antecedentes a las BD actuales fueron los Sistemas Gestores de Ficheros, los cuales consistían en almacenar datos en ficheros individuales, exclusivos para cada aplicación particular, los caracteriza un proceso lento y la repetición de datos.

**(3) Enumera y explica brevemente los inconvenientes que presentan los antiguos sistemas de Archivos o de Ficheros.**

1. **Datos redundantes:** Se repiten continuamente en cada aplicación.
2. **Coste de almacenamiento elevado:** Al almacenarse varias veces se requiere mas espacio.
3. **Tiempo de procesamiento elevado:** Al no poder optimizarse al existir varias copias hay que tenerlas todas en cuenta.
4. **Inconsistencia de datos:** Errores humanos pueden causar que en dos copias distintas un mismo dato tenga distintos valores.
5. **Difícil modificación de datos:** Cada modificación se debe repetir en todas las copias del dato.

**(4) Tipos de organizaciones de los ficheros y principales características.**

1. **Fichero Secuencial:** Cada registro se graba en el espacio dedicado al fichero según va surgiendo, empieza en el primer registro que se escribe y termina en el último registro guardado.
2. **Fichero Indexado:** Contiene dos áreas diferenciadas: Área Índice (se solicita con clave y funciona con enlaces) y el Área de datos (contiene los datos según aparecen).
3. **Fichero Directo:** Se caracteriza por, tener dos áreas definidas (datos y colisiones), estimar un número de registros que se esperan y reservar espacio para sus zonas de colisión.

**(5) ¿Qué significa que los datos de una BD tienen un significado implícito? Por tres ejemplos, diferentes de los expuesto en el tema, que ilustren tu explicación.**

El significado implícito quiere decir que cada dato adquiere valor dentro de un contexto:

- Un número, por sí solo, no dice nada.
- Pero si ese número está en la columna Edad de la tabla Clientes, ya sabemos que representa la edad de una persona.

**(6) Soportes de almacenamiento. Explica la organización de un disco duro. Elementos que aparecen y descríbelos brevemente.**

1. **Cinta Magnética:** Se escribe una marca de inicio de fichero, su contenido y luego una marca final permitiendo escribir más de un fichero en la misma cinta.
2. **Disco Magnético:** Se organiza en pista, superficie y sectores, su brazo de lectura selecciona el sector del disco (ya que gira constantemente), crea un acceso aleatorio que rápidamente se posiciona y lee o escribe.

**(7) Definir que es un SGBD.**

Es una aplicación que permite definir, crear y mantener una base de datos, permite su definición, la inserción, actualización, eliminación y consulta de datos, maneja un acceso controlado y contiene un diccionario de datos.

**(8) Indica las principales ventajas de las BD frente a los antiguos sistemas de ficheros.**

1. Menor redundancia (no hace falta tanta repetición de datos).
2. Menor espacio de almacenamiento.
3. Acceso a los datos más eficiente.
4. Documentación de datos.
5. Independencia de los datos respecto a programas y procesos.
6. Integridad de los datos. Mayor seguridad en los datos.

**(9) Nombra los distintos tipos de Bases de Datos que existen según el modelo que siguen.**

1. Base de datos Jerárquica .
2. Base de datos Red.
3. Base de datos Relacional.
4. Base de datos con estructura Multidimensional.
5. Bases de Datos Orientada a Objetos.

**(10) Describe el significado de las siguientes siglas: DDL, DML y DCL. Explica la utilidad de cada una.**

1. **(Data Description Language).** Este lenguaje permite especificar la estructura, tipo y restricciones de los datos.
2. **(Data Manipulation Language).** Permite la inserción, actualización, eliminación y consulta de datos.
3. **(Data Control Language).** Conjunto de instrucciones que sirven para controlar el acceso y la seguridad de la base de datos.

**(11) ¿Qué es un script o guión? Define los siguientes conceptos:**

Conjunto de instrucciones, que ejecutadas de forma ordenada, realizan operaciones avanzadas o de mantenimiento de los datos.

- a. **Dato.** Cualquier elemento informativo que tenga relevancia para el usuario.
- b. **Tipo de Dato.** Indica la naturaleza del campo, es decir, qué contenidos puede almacenar y qué operaciones se pueden hacer con ellos
- c. **Campo.** Es cada una de las columnas de una tabla.

- d. **Registro.** Cada fila de una tabla, que agrupa los valores de los campos para una instancia concreta de la entidad.
- e. **Tabla.** Conjunto de registros organizados en filas y columnas bajo un mismo nombre.
- f. **Relación.** Asociación entre dos o más tablas mediante claves primarias y foráneas.
- g. **Consulta.** Instrucción que permite recuperar información de una Base de Datos.
- h. **Procedimiento.** Son un tipo especial de Script que están almacenados en la Base de Datos y forman parte de sus esquemas.

**(12) ¿Qué es el diccionario de datos?**

Es un catálogo que contiene la descripción de los datos de la Base de Datos y es accesible para los usuarios.

**(13) ¿Qué quiere decir que una Base de Datos permita la concurrencia?**

Significa que varios usuarios o procesos pueden acceder y trabajar con los mismos datos al mismo tiempo, sin que se produzcan errores, inconsistencias o pérdidas de información.

## 7. Planteamiento práctico

(pag 19 documento de clase)

**(1) Completa una tabla buscando ejemplos en el enunciado de cada uno de los inconvenientes estudiados en los sistemas de ficheros. Deberás completar la respuesta con una breve explicación que ilustre el porqué dicho ejemplo presenta el inconveniente en cuestión.**

INCONVENIENTE	EJEMPLO	EXPLICACIÓN
Coste de almacenamiento elevado	Utilizar un almacén para guardar y organizar los datos.	Se necesita un lugar físico ya que todos los datos se almacenan en papel y el costo del alquiler del almacén es más caro que tenerlo online.
Datos redundantes	Se añade un nuevo bloque la para vender.	Se almacenan datos de más al repetir datos antiguos aumentando su espacio de almacenamiento.
Probabilidad alta de inconsistencia de datos	El nombre de un bloque y dirección está mal escrito.	Puede ser que se deba a que los datos al ser escritos por personas cabe la posibilidad de error.
Difícil modificación de los datos	Tenemos que reescribir todo el fichero antiguo porque un bloque ya no es posible venderlo.	Se debería volver a modificar todo ya que la información deberá estar actualizada.
Tiempo de procesamiento elevado	Estos datos tardan en reescribirse.	El hecho de reescribirlos cada vez que se actualizan lo hace muy lento.

**2. Supongamos ahora que quisiéramos resolver el problema real usando el enfoque actual de BD. Detecta en el enunciado ejemplos de los siguientes elementos fundamentales de una BD:**

- Datos: (Nombre de un cliente: “*Juan García*”, Precio de alquiler: “*750 € mensuales*”...)
- Entidades: (Oficinas, Propietarios, Anuncios)
- Claves principales: (Código de oficina → clave primaria de la tabla *OFICINAS*, Código de empleado → clave primaria de la tabla *PLANTILLA*...)



- Claves foráneas: (En *INMUEBLES*, el código de empleado actúa como clave foránea para indicar qué trabajador gestiona ese inmueble.)
- Integridad referencial: (Si en la tabla *CONTRATOS* aparece un código de cliente = 1225, ese cliente debe existir previamente en la tabla *INQUILINOS*.)
- Metadatos: (Tipo de dato: “*precio del alquiler en euros (numérico)*”, Restricciones: “*la duración mínima de un contrato es de tres meses y la máxima de un año*”...)