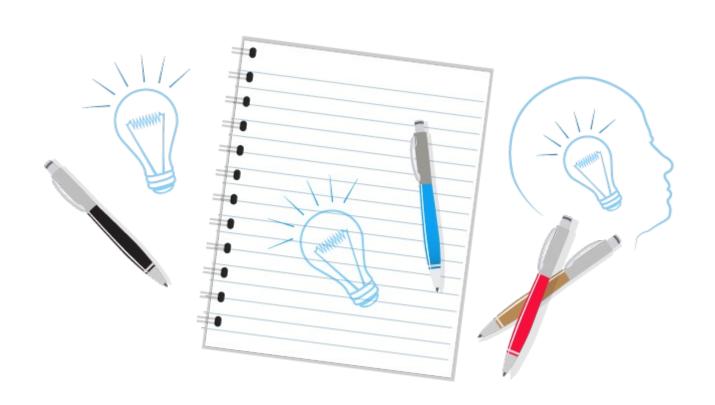
Tema 1 – Sistemas de almacenamiento de la información



¿Qué vamos a estudiar?

1.- Ficheros

- 1.1. Tipos de ficheros y formatos
- 1.2. Ficheros de texto
- 1.3. Ficheros binarios

2.- Bases de Datos

- 2.1. Conceptos
- 2.2. Estructura de una base de datos
- 2.3. Usos de las bases de datos
- 2.4. Modelos y tipos de base de datos
- 2.5. Evolución

3.- Sistemas Gestores de Base de Datos

- 3.1. Concepto de SGBD
- 3.2. Funciones de un SGBD
- 3.3. Lenguaje SQL
- 3.4. Tipos de SGBD

En un ordenador la información está almacenada en <u>dispositivos de</u> <u>almacenamiento</u>, esto es, discos duros, dvds, pen drives, etc. Para poder organizar la información en estos dispositivos, se utilizan los ficheros.

Los **ficheros** son estructuras de información que crean los sistemas operativos de los ordenadores para poder almacenar datos.

Suelen tener <u>nombre y extensión</u>, que determina el formato de la información que contiene.





Tema_1

Tema 1

1.1 – Tipos de ficheros y formatos

El **formato** y **tipo de fichero** determina la forma de interpretar la información que contiene, ya que, en definitiva, lo único que se almacena en un fichero es una ristra de bits (ceros y unos), de forma que es necesaria su interpretación para dar sentido a la información almacenada.

Los datos se ordenan según un formato, y el sistema operativo, debe conocer este formato para poder mostrarlos por pantalla en la forma correcta.

1.1 – Tipos de ficheros y formatos

El **contenido** de un fichero puede ser tratado como <u>texto</u>, o como <u>datos binarios</u>, es decir, los bits almacenados en un fichero pueden ser traducidos por el sistema operativo a caracteres alfabéticos y números que entiende el ser humano, o pueden ser tratados como componentes de estructuras de datos más complejas, como ficheros que almacenan sonido, vídeo, imágenes, etc.

La **organización** de un fichero indica la forma en que se han de acceder a los datos.

- Organización secuencial: los datos están organizados siguiendo una secuencia ordenada, es decir, unos detrás de otros. Hay que recorrer todos los datos para llegar a uno concreto.
- Organización directa: Permiten acceder a un dato en concreto sin necesidad de acceder a todos los anteriores.
- Organización indexada: permiten acceder a la información rápidamente, simulando la forma en que el índice de un libro facilita el acceso a sus contenidos.



Busca en internet las ventajas y desventajas de las distintas organizaciones de los ficheros y muestra una representación gráfica de cada uno de ellos.

1.1 – Tipos de ficheros y formatos

La **utilidad** de un fichero indica qué uso se va a hacer de él. Se pueden clasificar en:

- Ficheros maestros
- Ficheros de movimientos
- Ficheros históricos

Actualmente un sistema operativo trata un fichero desde dos puntos de vista:

- Según su contenido (texto o datos binarios)
- Según su tipo (imágenes, ejecutables, clips de videos, etc.)



Busca en internet la diferencia entre un fichero maestro, un fichero de movimiento y un fichero histórico.

1.2 – Ficheros de texto

Los **ficheros de texto** suelen llamarse también <u>ficheros planos o ficheros</u> <u>ascii</u>.

ASCII → American Standard Code for Information Interchange.

Es un estándar que asigna un valor numérico a cada carácter, con lo que se pueden representar los documentos llamados de Texto Plano, es decir, los que son directamente legibles por seres humanos.



Busca en internet la representación del código ASCII. Indica que valor numérico corresponde a los siguientes caracteres con símbolos y signos:

- > d
- **>** Ü
- > €
- > 4
- >
- ×Á
- > }
- > /

1.2 – Ficheros de texto

Los **ficheros de texto**, aunque <u>no necesitan un formato</u> para ser interpretado, suelen tener extensiones para conocer qué tipo de texto se halla dentro del fichero, por ejemplo:

- Ficheros de configuración: Son ficheros cuyo contenido es texto sobre configuraciones del sistema operativo o de alguna aplicación. Estos pueden tener extensión .ini, .inf , .conf
- Ficheros de código fuente: Su contenido es texto con programas informáticos. Ejemplos: .sql, .c, .java
- Ficheros de páginas web: Las páginas webs son ficheros de texto con hipertexto que interpreta el navegador: .html, .php, .css, .xml
- Formatos enriquecidos: Son textos que contienen códigos de control para ofrecer una visión del texto más elegante: .rtf

1.3 – Ficheros binarios

Los **ficheros binarios** son todos los que no son de texto, y requieren un formato para ser interpretado. A continuación, se muestran algunos tipos de formatos de ficheros binarios:

- De imagen: .jpg, .gif, .tiff, .bmp, .wmf, .png, .pcx; entre muchos otros
- De vídeo: .mpg, .mov, .avi, .qt
- Comprimidos o empaquetados: .zip, .Z, .gz, .tar, .lhz
- Ejecutables o compilados: .exe, .com, .cgi, .o, .a
- Procesadores de textos: .doc, .odt

2.1 – Conceptos

Una **Base de Datos** es una <u>colección de información</u> perteneciente a un mismo contexto (o problema), que está <u>almacenada</u> de forma organizada en <u>ficheros</u>.

Una base de datos está <u>organizada</u> mediante <u>tablas</u> que almacenan información concerniente a algún objeto o suceso. Estas tablas se relacionan formando <u>vínculos o relaciones</u> entre ellas, que ayudan a mantener la información de los 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.

Cada fila almacena un <u>registro</u> con tantos campos como columnas tenga la tabla.

2.1 – Conceptos

Ejemplo de tabla:

III Tabla1 : Tabla							
	ld	Nombre	Apellidos	Notas	Direction	Poblac	
•	4	vicente	Aranda	Es un director e	c/ Desideria	Turquia	
	11	juan	Benet	Es un viajante	c/ Puerto	Madrid	
	7	juan	Borbon	Es un personaje	c/ Republica	Barceloi	
	8	Aureliano	Buendia	Es un Militar	c/ Cementerio	Macond	
	1	George	Clooney	Es un actor de	C/ Estacion	Solaris	
	9	Paul	Deckard	Es un Policia	c/ Progreso	San Fra	
	2	Frank	Herbet	Es un escritor	c/ Rakis	Dune	

Esta tabla se podría relacionar con los pedidos que realiza el cliente de determinados productos.

2.1 – Conceptos

- Dato: es un trozo de información concreta sobre algún concepto o suceso. Por ejemplo, 1996 es un número que representa un año de nacimiento de una persona. Los datos se caracterizan por pertenecer a un tipo.
- Tipo de Dato: indica la <u>naturaleza del campo</u>. Así, se puede tener datos numéricos, que son aquellos con los que se pueden realizar cálculos aritméticos (sumas, restas, multiplicaciones...) y los datos alfanuméricos, que son los que contienen caracteres alfabéticos y dígitos numéricos.
- Campo: es un identificador para toda una familia de datos. Cada campo pertenece a un tipo de datos. Por ejemplo, el campo "FechaNacimiento" pertenece al tipo de dato Fecha. Al campo también se le llama columna.

2.1 – Conceptos

- Registro: Es una recolección de datos referentes a un mismo concepto o suceso. Por ejemplo, los datos de una persona pueden ser su NIF, año de nacimiento, su nombre, su dirección, etc. A los registros también se les llama tuplas o filas.
- Campo Clave: Es un campo especial que identifica de forma única a cada registro. Ejemplo: DNI
- ➤ **Tabla:** Es un conjunto de registros bajo un mismo nombre que representa el conjunto de todos ellos. *Por ejemplo, todos los clientes de una base de datos se almacenan en una tabla cuyo nombre es Clientes.*

2.1 – Conceptos

- Consulta: Es una instrucción para hacer <u>peticiones</u> a una base de datos. Puede ser una búsqueda simple de un registro especifico o una solicitud para seleccionar todos los registros que satisfagan un conjunto de criterios.
- Índice: Es una estructura que almacena los <u>campos clave</u> de una tabla, organizándolos para hacer más fácil encontrar y ordenar los registros de esa tabla. El índice tiene un funcionamiento similar al índice de un libro, guardando parejas de elementos: el elemento que se desea indexar y su posición en la base de datos. Para buscar un elemento que esté indexado, solo hay que buscar en el índice de dicho elemento para, una vez encontrado, devolver el registro que se encuentre en la posición marcada por el índice.

2.1 – Conceptos

- Vista: Es una transformación que se hace a una o más tablas para obtener una nueva tabla. Esta nueva tabla es una tabla virtual, es decir, no está almacenada en los dispositivos de almacenamiento del ordenador, aunque sí se almacena su definición. Ejemplo: una vista que tiene el nombre del cliente y el número de pedidos que ha realizado.
- Informe: Es un <u>listado ordenado de los campos y registros</u> seleccionados en un formato fácil de leer. Generalmente se usan como peticiones expresas de un tipo de información por parte de un usuario. *Por ejemplo, un informe de las facturas impagadas del mes de enero ordenado por nombre de cliente.*

2.1 – Conceptos

- Guiones o script: Son un conjunto de instrucciones, que, ejecutadas de forma ordenada, realizan operaciones avanzadas de mantenimiento de los datos almacenados en la base de datos. Ejemplo: script que ejecuta una copia de la BD.
- Procedimientos: Son un tipo especial de script que está almacenado en la base de datos y que forma parte de su esquema. Ejemplo: procedimiento que borra todos los productos que están dados de baja en la tabla de productos.

2.2 – Estructura de una base de datos

Esquema, metainformación o metadatos

- Una base de datos almacena los datos a través de un esquema.
- El esquema es la definición de la estructura donde se almacenan los datos, contiene todo lo necesario para organizar la información mediante tablas, registros (filas) y campos (columnas). Es un contenedor con nombre para los objetos de la base de datos, que permite agrupar objetos en espacios de nombres independientes.
- Los gestores de bases de datos modernos SQL Server, Oracle, MySQL y DB2, entre otros, almacenan el esquema de la base de datos en tablas.

Tabla △	Acción	Filas 🕢 Tipo	Cotejamiento	Tamaño
components	🚖 🔳 Examinar 🋂 Estructura	2 InnoDB	utf8_general_ci	16 KB
components_level	🚖 🔳 Examinar 🎉 Estructura	4 InnoDB	utf8_general_ci	16 KB
components_ship	🚖 🔳 Examinar 🋂 Estructura	2 InnoDB	utf8_general_ci	16 KB
user_register_meta	🙀 🔳 Examinar 📝 Estructura	19 InnoDB	utf8_general_ci	16 KB
14 tablas	Número de filas	1,874 MyISAM	utf8_general_ci	320 KB

2.3 – Usos de las bases de datos

Las bases de datos están en cualquier tipo de sistema informático, a continuación, se exponen solo algunos ejemplos de sus usos más frecuentes:

- Bases de datos Administrativas: Cualquier empresa necesita registrar y relacionar sus clientes, pedidos, facturas, productos, etc.
- Bases de datos Contables: También es necesario gestionar los pagos, balances de pérdidas y ganancias, patrimonio, declaraciones de hacienda...
- Bases de datos para motores de búsquedas: Por ejemplo, Google, tiene una base de datos gigantesca donde almacena información sobre todos los documentos de Internet. Posteriormente millones de usuarios buscan en la base de datos de estos motores.

2.3 – Usos de las bases de datos

- Científicas: Recolección de datos climáticos y medioambientales, químicos, genómicos, geológicos...
- Configuraciones: Almacenan datos de configuración de un sistema informático, como, por ejemplo, el registro de windows.
- Bibliotecas: Almacenan información bibliográfica, por ejemplo, la famosa tienda virtual Amazon o la biblioteca de un instituto.
- Censos: Guardan información demográfica de pueblos, ciudades y países.
- Virus: Los antivirus guardan información sobre todos los potenciales software maliciosos.
- Otros muchos usos: Militares, videojuegos, deportes, etc.



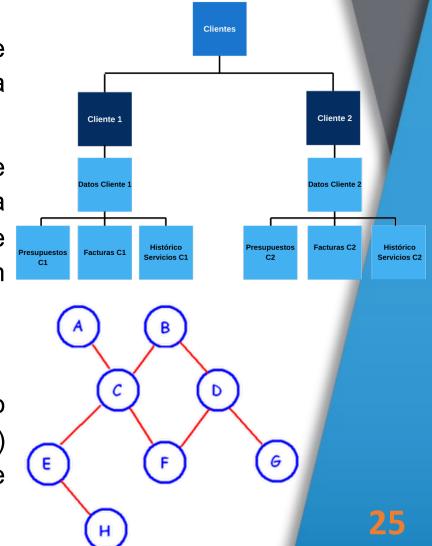
Busca en Internet las 10 bases de datos más grandes del mundo. Anota su nombre y su tamaño, y, en una hoja de cálculo, genera un gráfico que muestre la comparativa del tamaño de estas bases de datos.

2.4 – Modelos y tipos de base de datos

El modelo de base de datos es la arquitectura mediante la que se almacena e interrelaciona la información que se va a gestionar.

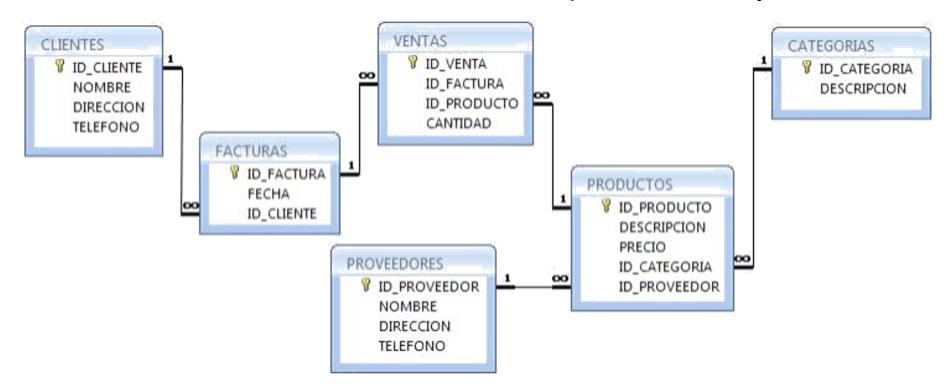
Modelo jerárquico: es el más antiguo. Refina la idea de fichero indexado, creando una estricta relación de jerarquía entre los datos de varios ficheros, motivo por el que presenta serias limitaciones semánticas. La implantación comercial más conocida es IMS de IBM.

Modelo en red: introduce mejoras respecto al modelo jerárquico (mayor independencia y flexibilidad de los datos) a costa de aumentar el nivel de complejidad. Algunas de sus implantaciones: CODASYL, IDMS/DB.



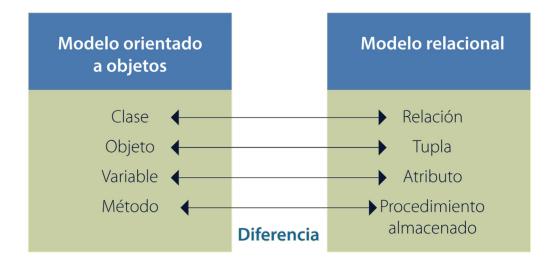
2.4 – Modelos y tipos de base de datos

Modelo relacional: representa la información en forma de entidades y relaciones entre ellas, huyendo de la rigidez de los modelos previos. Cada entidad y cada relación aparece en forma de tablas bidimensionales (filas y columnas). Es el modelo más extendido desde hace décadas. Implantaciones: MySQL, Oracle, ...



2.4 – Modelos y tipos de base de datos

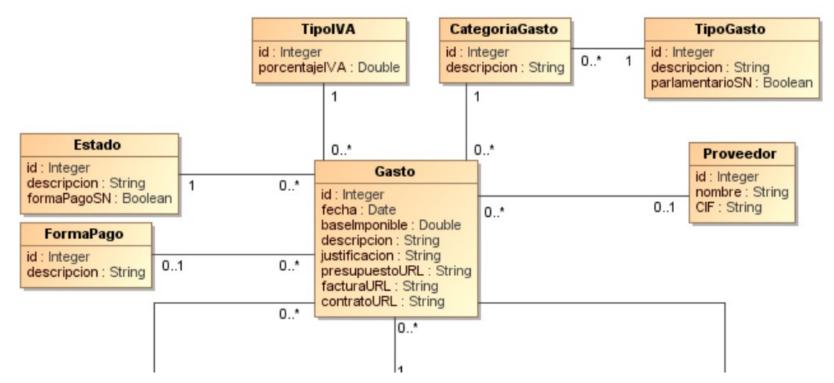
Orientado a objetos: aplica a los datos el paradigma de la orientación a objetos (POO). Irrumpió con fuerza en los años noventa debido a las nuevas necesidades de almacenamiento de las bases de datos relacionales (imágenes, documentos, ficheros de audio y vídeo). Implantaciones: Versant, db4o, InterSystems, Objectivity.



Clase	Objetos	Atributos/datos		
		Edad: 25		
	Juan Pérez	Puesto: Psicóloga social		
Franciscolo		Salario: 8000		
Empleado	María Suárez	Edad: 23		
		Puesto: Pedagoga		
		Salario: 15 000		

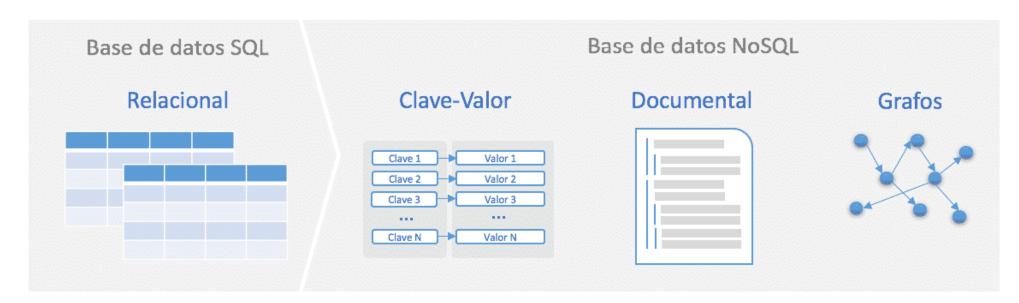
2.4 – Modelos y tipos de base de datos

Objeto-relacional: en los últimos años los fabricantes de bases de datos relacionales han incorporado a su software diversas capacidades de las bases de datos orientadas a objetos, creando este modelo híbrido. Ejemplos: Oracle, Microsoft SQL Server, IBM DB2, IBM informix, PostgreSQL.



2.4 – Modelos y tipos de base de datos

Modelo noSQL: Abarca una amplia gama de tecnologías y arquitecturas, busca resolver los problemas de escalabilidad y rendimiento de big data que las bases de datos relacionales no fueron diseñadas para abordar. Los datos almacenados no requieren estructuras fijas como tablas. Ejemplos: Cassandra, MongoDB



2.4 – Modelos y tipos de base de datos

Si atendemos al criterio de la <u>ubicación física</u> de la información, podemos diferenciar entre dos tipos de bases de datos

- Centralizadas: la base de datos reside en una sola máquina, típicamente el servidor de base de datos
- Distribuidas: la información se reparte en distintos servidores, generalmente alejados físicamente. Su implantación exige hacer un fuerte hincapié en aspectos de networking y seguridad. Existen 3 tipos de fragmentación:
 - → Fragmentación horizontal: se realiza sobre las tuplas o registros de la relación, y se agrupan por propiedades geográficas comunes, y contienen todos los atributos de la relación original
 - > Fragmentación vertical: se realiza sobre los atributos de la relación. Es más complicada
 - Fragmentación híbrida o mixta

2.5 – Evolución

La clasificación de las bases de datos en tipos está ligada a su evolución histórica. Según ha ido avanzando la tecnología, las bases de datos han mejorado cambiando la forma de representar y extraer la información.

- En la <u>década de 1950</u> se inventan las cintas magnéticas, que sólo podían ser leídas de forma secuencial y ordenadamente. Estas cintas, almacenaban ficheros con registros que se procesaban secuencialmente junto con ficheros de movimientos para generar nuevos ficheros actualizados.
- En la <u>década de 1960</u> se generaliza el uso de discos magnéticos, cuya característica principal es que se podía acceder de forma directa a cualquier parte de los ficheros, sin tener que acceder a todos los datos anteriores.
- ➤ En <u>1970</u> Edgar Frank Codd, científico informático inglés de IBM publicó en un artículo 'Un modelo relacional de datos para grandes bancos de datos compartidos', donde definió el modelo relacional.

2.5 - Evolución

- En la <u>década de 1980 IBM</u> lanza su motor de bases de datos DB2, para las plataformas MVS. Un año después, IBM crea el SQL (Structured Query Language), un potente lenguaje de consultas para manipular información de bases de datos relacionales
- A mediados de 1990, lanza una versión de DB2 (DB2 Parallel Edition) que es capaz de dividir una base de datos enorme en varios servidores comunicados por líneas de gran velocidad.
- A <u>finales de 1990</u> IBM y Oracle incorporan a sus bases de datos la capacidad de manipular objetos, creando las bases de datos orientadas a objetos.
- Con internet aparecen las bases de datos distribuidas, que intercambian información y actualizaciones a través de la red. El aumento de datos provoca la aparición del Software de ayuda a la decisión.



Busca en Internet la biografía de los siguientes personajes, y comenta su principal contribución a la evolución de las bases de datos:

- Edgar Frank Codd
- Bill Gates
- Larry Ellison
- Michael Monty Widenius
- Roger Kent Summit

3.1 – Concepto

Un Sistema Gestor de Bases de Datos, en adelante SGBD, se define como el conjunto de herramientas que facilitan la consulta, el uso y actualización de una base de datos.

Un ejemplo de software gestor de base de datos es Oracle 11g, que incorpora un conjunto de herramientas software que son capaces de estructurar en múltiples discos duros los ficheros de una base de datos, permitiendo el acceso a sus datos tanto a partir de herramientas gráficas como a partir de potentes lenguajes de programación (PL-SQL, php, C++,...)

3.2 – Funciones de un SGBD

- Permite a los usuarios **almacenar datos, acceder a ellos y actualizarlos** de forma sencilla y con un gran rendimiento, ocultando la complejidad y las características físicas de los dispositivos de almacenamiento
- Garantizan la integridad de los datos, respetando las reglas y restricciones que dicte el programador de la base de datos. Es decir, no permiten las operaciones que dejen cierto conjunto de datos incompletos o incorrectos
- Integran, junto con el sistema operativo, un **sistema de seguridad** que garantiza el acceso a la información exclusivamente a aquellos usuarios que dispongan de autorización.
- Proporcionan un diccionario de metadatos, que contiene el esquema de la BD, es decir, como están estructurados los datos en tablas, registros y campos, las relaciones entre los datos, usuarios, permisos, ... Este diccionario de datos debe ser accesible de la misma forma en la que se accede al resto de datos.

3.2 – Funciones de un SGBD

- Permite el uso de **transacciones**, garantizando que todas las operaciones de la transacción se realicen correctamente y de forma atómica, y en caso, de alguna incidencia, deshacen los cambios sin ningún tipo de complicación adicional.
- Ofrecen, mediante completas herramientas, estadísticas sobre el uso del gestor, registrando operaciones efectuadas, consultas solicitadas, operaciones fallidas y cualquier tipo de incidencia. Monitorizando el uso de las base de datos, y permitiendo analizar hipotéticos malfuncionamientos.
- Permiten concurrencia, es decir, varios usuarios trabajando sobre un mismo conjunto de datos. Además, proporcionan mecanismos que permiten arbitrar operaciones conflictivas en el acceso o modificación de un dato al mismo tiempo por parte de varios usuarios.

3.2 – Funciones de un SGBD

- Independizan los datos de la aplicación o usuario que esté utilizándolos, haciendo más fácil su migración a otras plataformas.
- Ofrecen conectividad con el exterior, pudiendo replicar y distribuir base de datos. El protocolo ODBC (Open Database Connectivity), es un estándar de conectividad muy extendido entre la base de datos y aplicaciones externas.
- Incorporan herramientas para la salvaguarda y restauración de la información en caso de desastre. Además, ofrecen sencillas herramientas para la importación y exportación automática de la información.



Busca en Internet las reglas de Codd para el funcionamiento de los Sistemas Gestores de Bases de Datos y establece una relación entre cada una de las leyes y las funciones que proporcionan los SGBD actuales.

3.3 – Lenguaje SQL

La principal herramienta de un gestor de base de datos es la interfaz de programación con el usuario. Este interfaz consiste en un lenguaje muy sencillo mediante el cual el usuario realiza preguntas al servidor, contestando este a las demandas del usuario. Este lenguaje se denomina SQL, *Structured Query Language*.

SQL está estandarizado por ISO (*International Organization for Standardarization*), es decir, todas las bases de datos que soporten SQL deben tener la misma sintáxis a la hora de aplicar el lenguaje. Su última revision fue en 2016.

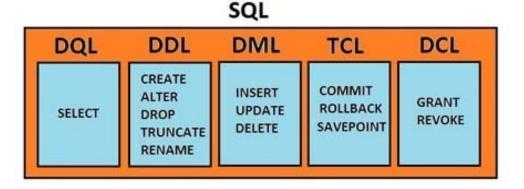
3.3 – Lenguaje SQL

SQL se divide en 5 sublenguajes, el total de todos ellos permite al SGBD cumplir con las funcionalidades requeridas por Codd:

- Lenguaje DDL o de definición de datos (*Data Definition Language*). Este lenguaje permite crear toda la estructura de una base de datos (desde tablas hasta usuarios). Sus cláusulas son de tipo DROP (elimina objetos) o CREATE (crea objetos).
- Lenguaje DML o de manipulación de datos (Data Manipulation Language). Este lenguaje permite al usuario insertar (INSERT), modificar (UPDATE) o eliminar datos (DELETE).
- **Lenguaje DQL** o de consulta de datos (Data Query Language). Este lenguaje permite al usuario consultar datos almacenados en la base de datos (SELECT).

3.3 – Lenguaje SQL

- Lenguaje DCL o lenguaje de control de datos (Data Control Language). Este lenguaje incluye comandos (GRANT y REVOKE) que permiten al administrador gestionar el acceso a los datos contenidos en la base de datos.
- Lenguaje TCL o lenguaje de control de transacciones (Transaction Control Language). Este lenguaje permite gestionar los cambios producidos por sentencias DML en forma de transacciones. Si es posible ejecutar todos los comandos se aplica la transacción (COMMIT), y si, en algún paso de la ejecución sucede algo inesperado, se pueden deshacer todos los paso dados (ROLLBACK)





Busca el significado del término *SQL Injection* e indica por qué un administrador debe protegerse frente a él.

3.4 – Tipos de SGBD

Los SGBD se pueden clasificar de muchas formas, por ejemplo, según el modelo de base de datos que gestionan. Como en la actualidad, la mayoría de los SGBD integran múltiples filosofías y tipos de funcionamiento, en este caso vamos a clasificar los SGBD según su capacidad y potencia

- SGBD Ofimáticas: almacenan bases de datos pequeñas orientadas a almacenar datos domésticos o de pequeñas empresas. Incluso estos gestores permiten construir pequeñas aplicaciones para ayudar al usuario inexperto a manipular los datos de una base de datos de forma sencilla e intuitiva. Ejemplo: Microsoft Access
- SGBD Corporativas: gestionan bases de datos enormes, de grandes o medianas empresas con una carga de transacciones que requieren un servidor de grandes dimensiones. Son capaces de manipular grandes cantidades de datos de forma muy rápida y eficiente para poder resolver la demanda de muchos usuarios. Ejemplo: Oracle, DB2, SQL Server, ...

3.4 – Tipos de SGBD

El elevado coste de los SGBD corporativos ha desencadenado que se haya recurrido a una solución intermedia entre gestores de bases de datos ofimáticas y corporativas.

Entre estas soluciones intermedias se encuentra MySQL-MariaDB, un SGBD que además de ser gratuito y sencillo, es capaz de manipular gran cantidad de datos cumpliendo todos los estándares de la arquitectura ANSI SPARC.



Realiza, con la ayuda de Internet, una lista de SGBD comerciales y libres y relaciónalos con los modelos de bases de datos que utilizan.



Debate sobre SGBD libres y comerciales.