

UNIDAD DE TRABAJO 1: INTRODUCCIÓN A LAS BASES DE DATOS

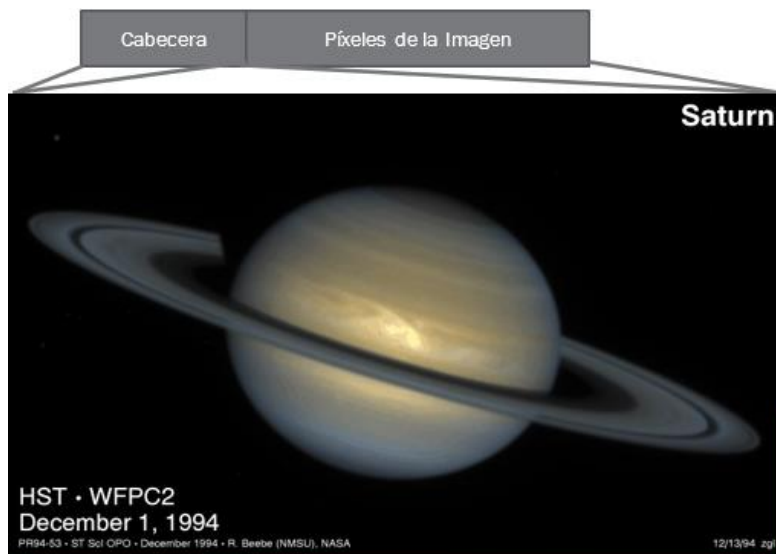
1. SISTEMAS DE ALMACENAMIENTO DE LA INFORMACIÓN.

A. FICHEROS

- Estructuras de información que crean los sistemas operativos de los ordenadores para poder almacenar datos.
- Tienen un *tipo* y un *formato* que determina la forma de interpretar la información que contiene.

Hay que tener en cuenta que todo lo que se almacena en los ficheros del ordenador son ristra de 0s y 1s por lo que hace falta interpretarlo. Esa interpretación dependerá del formato del fichero (la extensión) y del sistema operativo o la aplicación que intente “abrir” el fichero.

Por ejemplo, la imagen de abajo tiene la extensión .bmp por lo que es un fichero binario. Estos ficheros almacenan en un vector el color de cada pixel que compone la imagen. Además, se almacenará un tamaño.



B. TIPOS DE FICHEROS Y FORMATOS

Tradicionalmente, los ficheros se han clasificado de diferentes formas, según su contenido (texto o binario), según su organización (secuencial, directa o indexada) o según su utilidad (maestros, históricos o movimientos).

Vamos a centrarnos en clasificar los ficheros, según su contenido:

- Ficheros de texto: suelen llamarse “ficheros planos”.

- Asignan un valor numérico a cada carácter. El estándar es el código ASCII (American Standard Code for Information Interchange) es el más común, aunque hay otros estándares que también se usan.
 - No necesitan un formato para ser interpretado.
 - Son ficheros del tipo:
 - .ini, .inf, .conf → ficheros de configuración.
 - .sql, .c, .java → ficheros de código fuente.
 - .html, .php, .css, .xml → ficheros de páginas web.
 - .rft, .ps, .tex → formatos enriquecidos.
- **Ficheros binarios**: requieren de un formato para ser interpretados.
- Son ficheros del tipo:
 - Imágenes: .jpg, .gif, .bmp, ...
 - Videos: .mpg, .mov, .avi, .qt,...
 - Comprimidos: .zip, .Z, .gz, .tar, .lhz
 - Ejecutables o compiladores: .exe, .com, .cgi, .o, .a
 - Procesadores de texto: .doc, .docx, .odt

2. CONCEPTOS REFERENTES A LAS BASES DE DATOS.

Supongamos una pequeña empresa que se dedica a realizar ventas por correo. Ésta necesitará gestionar a sus clientes, tener un control del almacén y un sistema de facturas. La forma tradicional de resolver este problema consiste en utilizar una serie de aplicaciones (programas) en los que cada aplicación dispone de su propio conjunto de ficheros, que contienen los datos necesarios, y que están organizados de acuerdo a la forma que tiene la aplicación de tratarlos. Estas organizaciones de los ficheros serían las ya conocidas: secuencial, directa o indexada.

Si se decide cambiar la estructura de alguno de estos ficheros, será necesario también cambiar la propia aplicación. A la inversa, si se tiene que cambiar la aplicación, casi con toda seguridad habrá que cambiar el número de ficheros, su organización, tipo de campos, etc.

Además, y dado que las aplicaciones son para la misma empresa, es seguro que existirá un amplio grado de redundancia (repetición) de los datos entre los distintos ficheros de las aplicaciones, por ejemplo, datos de clientes repetidos en varios ficheros.

Si se reemplazan todos estos ficheros por una colección de datos apropiada, generalmente de gran tamaño, que sea accesible por todas las aplicaciones, que se consisten y que presente una redundancia mínima, entonces la colección de datos se transforma en una Base de Datos (BD). En el caso de la empresa de venta por correo la BD estará constituida por una única, colección de datos usada por las aplicaciones de clientes, almacén y facturas.

A. BASE DE DATOS.

Una definición formal de lo que es una BD sería la siguiente:

- Una BD es una colección interrelacionada de datos, almacenados en un conjunto sin redundancias innecesarias, cuya finalidad es la de servir a una o más aplicaciones de la manera más eficiente.

B. DATOS

- Es el componente fundamental de una BD. Están relacionados entre sí y se deben evitar al máximo posible las redundancias.
- Es un trozo de información concreta sobre algún concepto o suceso. Por ejemplo, 1996 es un número que representa el año de nacimiento de una persona.

Clientes			
	Id. de cliente	Nombre de compañía	Nombre del contact
+	ALFKI	Alfreds Futterkiste	Maria Anders
+	ANATR	Ana Trujillo Emparedados y helado	Ana Trujillo
+	ANTON	Antonio Moreno Taquería	Antonio Moreno
+	AROUT	Around the Horn	Thomas Hardy
+	BERGS	Berglunds snabbköp	Christina Berglund
+	BLAUS	Blauer See Delikatessen	Hanna Moos
+	BLONP	Blondel père et fils	Frédérique Citeaux

En esta tabla de clientes, los datos son cualquiera de los valores que se almacenan en las celdas de la tabla, cualquier valor de Nombre de compañía, contacto, etc....

C. TIPO DE DATOS.

Indica la naturaleza del campo. Podemos tener datos numéricos (números), alfanuméricos (caracteres), fechas, etc.

Cuando damos de alta en una tabla un nuevo campo, dependiendo del gestor de bases de datos que estemos utilizando podremos seleccionar un tipo de datos. En este caso, añadimos el campo email para la tabla de empleados usando Access.

Empleados	
Nombre del campo	Tipo de datos
email	Texto corto
	Texto largo
	Número
	Fecha/Hora
	Moneda
	Autonumeración
	Sí/No
	Objeto OLE
	Hipervínculo
	Datos adjuntos
	Calculado
	Asistente para búsqueda

General	
Tamaño del campo	255
Formato	
Máscara de entrada	
Título	
Valor predeterminado	
Regla de validación	
Texto de validación	
Requerido	No
Permitir longitud cero	Sí
Indexado	No
Compresión Unicode	Sí

D. CAMPO.

Es un identificador para toda una familia de datos que se corresponde con una columna de una tabla de la base de datos.

Pedidos			
Id. de pedido	Cliente	Empleado	Fecha de pedido
10248	Wilman Kala	5	04-07-19
10249	Toms Spezialitäten	6	05-07-19
10250	Hanari Carnes	4	08-07-19
10251	Victuailles en stock	3	08-07-19
10252	Suprêmes délicies	4	09-07-19

En la tabla pedidos los campos son: id de pedido, Cliente, Empleado y Fecha de pedido.

E. REGISTRO.

Es una recolección de datos referentes a un mismo concepto o suceso.

Clientes			
Id. de cliente	Nombre de compañía	Nombre del contact	Cargo del contacto
ALFKI	Alfreds Futterkiste	Maria Anders	Representante de ventas
ANATR	Ana Trujillo Emparedados y helado	Ana Trujillo	Propietario
ANTON	Antonio Moreno Taquería	Antonio Moreno	Propietario
AROUT	Around the Horn	Thomas Hardy	Representante de ventas
BERGS	Berglunds snabbköp	Christina Berglund	Administrador de pedidos
BLAUS	Blauer See Delikatessen	Hanna Moos	Representante de ventas
BLONP	Blondel père et fils	Frédérique Citeaux	Gerente de marketing

En la tabla de clientes un registro sería toda la información de una fila. Por eso un registro también se llama “fila” o “tupla”.

F. TABLA.

Es un conjunto de registros bajo un mismo nombre que representa el conjunto de todos ellos. Por ejemplo, toda la información de los proveedores de una empresa, se guardarán en la tabla Proveedores. La tabla de Clientes sería otro ejemplo.

G. CONSULTA.

Es una instrucción para hacer peticiones a una base de datos. Puede ser una búsqueda simple de un registro o una búsqueda de registros que cumplan varias condiciones (listar todos los varones entre 30 y 40 años que sean universitarios, por ejemplo). En cualquiera de los casos una búsqueda consiste en extraer información de una base de datos.

H. ÍNDICE.

Es una estructura que almacena los campos clave de una tabla. Funciona de forma similar al índice de un libro. (Nos ocuparemos de esto más adelante).

I. VISTA.

Es una transformación que se hace de una o más tablas para obtener una nueva tabla. Esta tabla es virtual, es decir, no se almacena.

J. INFORME.

Listado ordenado de cierta información solicitada por el usuario. Por ejemplo, informe de facturas impagadas del mes de enero ordenado por nombre de cliente.

K. GUIONES.

(Scripts). Conjunto de instrucciones que ejecutadas de forma ordenada, realizan operaciones avanzadas de mantenimiento de los datos almacenados en la base de datos.

L. PROCEDIMIENTOS.

Son un tipo especial de script que está almacenado en la propia base de datos.

M. ESQUEMA DE UNA BASE DE DATOS.

El esquema es la estructura donde se almacena la base de datos, contiene todo lo necesario para organizar la información mediante tablas, los registros y los campos. También el conjunto de objetos necesarios para el tratamiento de los datos como son vistas, índices, ect. Al esquema también se le llama “metainformación” o metadatos.

```
mysql> select table_schema, table_name, table_rows
-> from information_schema.tables
-> where table_schema='jardineria';
```

table_schema	table_name	table_rows
jardineria	Clientes	36
jardineria	DetallePedidos	295
jardineria	Empleados	32
jardineria	GamasProductos	0
jardineria	Oficinas	10
jardineria	Pagos	26
jardineria	Pedidos	115
jardineria	Productos	276

9 rows in set (0,01 sec)

Mediante esta consulta en MySQL obtenemos las filas (o número de registros) de una base de datos llamada “jardinería”

3. USOS DE LAS BASES DE DATOS.

- Bases de datos Administrativas: clientes, pedidos, facturas, etc.
- Bases de datos Contables: gestión de pagos, balances de pérdidas y ganancias, patrimonio, ...
- Bases de datos Científicas: datos climáticos y medioambientales, químicos, genómicos, ...
- Bases de datos sobre configuraciones: registro de Windows.

- Bibliotecas: información bibliográfica, como la biblioteca de un IES.
- Censos: guardan información demográfica de pueblos, ciudades y países.
- Virus: patrones de software malicioso.
- Otros: militares, videojuegos, deportes, etc.

4. EVOLUCIÓN DE LAS BBDD.

- **AÑOS 50:** se crearon las cintas magnéticas de acceso secuencial y ordenado a la información. Nacen las aplicaciones basadas en sistemas de ficheros pues ni siquiera existía el concepto de base de datos.
- **AÑOS 60:** aparecen los discos magnéticos y el acceso directo a cualquier parte de los ficheros. Surge la idea de las BBDD jerárquicas y en red.
- **AÑO 1969:** surge la primera bbdd en red, llamada CODASYL.
- **AÑOS 70:** Edgar Fran Codd (IBM), publica el primer modelo relacional usado mayoritariamente todavía en la actualidad. También aparece ORACLE.
- **AÑOS 80:** IBM, lanza su motor de bases de datos DB2. También crea el SQL.
- **AÑOS 90:** IBM y ORACLE incorporan a sus bases de datos la capacidad de manipular objetos.
- **ACTUALMENTE:** con la aparición de Internet, nacen las bases de datos distribuidas y el software de ayuda a la decisión.

5. SISTEMAS GESTORES DE BASES DE DATOS (SGBD)

Un SGBD es un conjunto de herramientas que facilitan la consulta, uso y actualización de una base de datos.



- Permiten almacenar datos, acceder y actualizarlos de forma sencilla.
- Garantizan la integridad de los datos, es decir, respetan las condiciones impuestas en la base de datos, evitando la grabación de datos incorrectos o incompletos.
- Junto con el Sistema Operativo, aseguran el acceso a los datos solo a usuarios con autorización.

- Proporcionan el esquema de la base de datos, es decir toda la información sobre la estructura de la misma: tablas, registros, campos, relaciones, permisos, usuarios, etc.
- Permiten el uso de transacciones, garantizando que todas las operaciones de la transacción se realicen correctamente y en caso de incidencias, deshacen los cambios sin ningún tipo de complicación.
- Ofrecen herramientas para elaborar estadísticas sobre operaciones realizadas, consultas, incidencias.
- Permiten concurrencia, es decir, que varios usuarios puedan estar trabajando sobre la misma base de datos. Control de conflictos.
- Ofrecen conectividad con el exterior. De esta manera, se puede replicar y distribuir bases de datos (protocolo ODBC).
- Permite realizar copias de seguridad y operaciones de restauración en caso de desastre. También la posibilidad de devolver a la base de datos a un estado anterior.

A. USUARIOS DE LOS SGBD (Sistemas Gestores de Bases de Datos)

Otro componente de un sistema **BD** son los usuarios. Existen tres tipos diferentes de usuarios, que se distinguen por el modo en que interactúan con el sistema:

- **Programadores de aplicaciones:** son los encargados de escribir los programas de aplicación que utilicen bases de datos. Utilizan llamadas al **SGBD** incrustadas en un programa escrito en un lenguaje huésped de alto nivel (Java, C#, ...)
- **Usuarios 'ingenuos':** son usuarios poco experimentados que utilizan las aplicaciones escritas previamente.
- **Usuarios casuales:** son usuarios que interactúan con el sistema sin escribir programas. Utilizan consultas formuladas en un lenguaje de consulta **BD**.
- El **Administrador** de la **BD (DBA)** es el encargado de controlar y manejar la **BD** Puede ser una sola persona o un equipo. Tiene las siguientes funciones específicas:
 - Describir el esquema de la **BD**, los campos, los registros y las relaciones entre ellos.
 - Definir la estructura física de almacenamiento y el método de acceso.
 - Modificar el esquema y la organización física.
 - Conceder autorizaciones para el acceso a los datos a los usuarios.
 - Especificar las reglas de integridad de los datos.

B. TIPOS DE SGBD

- **SGBD de Bases de Datos OFIMÁTICAS:** bases de datos pequeñas orientadas al uso doméstico o de pequeñas empresas. Ejemplo: ACCESS.
- **SGBD de Bases de Datos CORPORATIVAS:** destinados a empresas grandes o medianas con carga de datos, transacciones que requieren un servidor de grandes dimensiones (Unix o Windows Server). Estos gestores deben atender de forma rápida y eficiente a las peticiones de cientos de usuarios. Un ejemplo sería Oracle que junto con DB2 son los más demandados. En el medio, se quedaría MySQL.

C. 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 con el cual podemos realizar operaciones con el servidor. Es el lenguaje SQL. El SQL se divide en 4 sublenguajes:
 - **Lenguaje DML:** lenguaje de Manipulación de Datos. Nos aporta cuatro sentencias sencillas como son:
 - SELECT: para seleccionar determinados datos.
 - INSERT: insertar datos.
 - UPDATE: modificación de datos.
 - DELETE: borrar datos.
 - **Lenguaje DDL:** lenguaje de Definición de Datos. Nos permite crear toda la estructura de una base de datos (usuarios, tablas, ...). Ejemplo:
 - DROP: eliminar objetos.
 - CREATE: crear objetos.
 - **Lenguaje DCL:** lenguaje de control de datos, con comandos como GRANT Y REVOKE que permiten al administrador gestionar el acceso a los datos contenidos en la base de datos.
 - **Lenguaje TCL:** lenguaje para el control de transacciones. Este lenguaje permite ejecutar varios comandos a la vez si fuera posible mediante COMMIT y si en algún paso de la ejecución no es posible se procede a deshacer todas las acciones realizadas hasta ese momento mediante ROLLBACK.