

2. Bases de datos

Como hemos visto anteriormente, los ficheros permiten organizar y almacenar conjuntos de datos del mismo tipo con una determinada estructura. Estos ficheros pueden ser utilizados por una aplicación o ser resultado. Pero podemos encontrarnos con información duplicada, incoherencia de datos, fallos de seguridad, etc ...

2.1. Conceptos

Para solucionar estos problemas aparece el concepto de **Base de Datos**, que permitirá reunir toda la información relacionada en un único sistema de almacenamiento, de manera que cualquier aplicación pueda utilizarla de manera independiente, mejorando así el tratamiento de la información y por tanto participando en la evolución del desarrollo de aplicaciones.

Base de datos: es una colección de datos relacionados lógicamente entre sí, con una definición y descripción comunes y que están estructurados de una determinada manera. Este conjunto estructurado de datos representa entidades y sus interrelaciones, almacenados con la mínima redundancia y posibilitando el acceso a ellos de manera eficiente por parte de varias aplicaciones y usuarios.

Una base de datos consta de los siguientes elementos:

Entidades son los objetos del mundo real de los que se almacena información en la base de datos. Por ejemplo, en una base de datos de una clínica veterinaria podríamos tener como entidades: *mascotas, doctores, etc.,...*

Atributos son los datos que se almacenan de la entidad. Cualquier propiedad o característica de la entidad puede ser un atributo. Por ejemplo, en la entidad *mascotas*, podríamos tener los atributos: *especie, raza, color, nombre, etc.,...*

Registros donde se almacena la información de cada entidad. Es un conjunto de atributos que contienen los datos de una entidad. Siguiendo con nuestro ejemplo, un registro podría ser: *perro, galgo, negro, Jack,...*

Campos donde se almacenan los atributos de cada registro, en nuestro ejemplo, un campo podría ser el valor "galgo"

Mascotas			
Especie	Raza	Color	Nombre
Perro	Galgo	Negro	Jack

Éstas son las ventajas fundamentales del uso de las bases de datos:

- **Acceso múltiple:** varios usuarios o aplicaciones pueden acceder a la base de datos, sin que existan problemas en el acceso o los datos.
- **Utilización múltiple:** cada uno de los usuarios o aplicaciones puede disponer de una visión particular de la estructura de la base de datos, de tal manera que cada uno de ellos accederá sólo a la parte que le corresponde.
- **Flexibilidad:** la forma de acceder a la información puede ser establecida de diferentes maneras, ofreciendo tiempos de respuesta muy reducidos.
- **Confidencialidad y seguridad:** el control del acceso a los datos puede ser establecido para que unos usuarios o aplicaciones puedan acceder a unos datos y a otros no, impidiendo a los usuarios no autorizados la utilización de la base de datos.
- **Protección contra fallos:** en caso de errores en la información, existen mecanismos bien definidos que permiten la recuperación de los datos de forma fiable.
- **Independencia física:** un cambio de soporte físico de los datos (por ejemplo: el tipo de disco), no afectaría a la base de datos o a las aplicaciones que acceden a ellos.

- **Independencia lógica:** los cambios realizados en la base de datos no afectan a las aplicaciones que la usan.
- **Redundancia:** los datos se almacenan, por lo general, una única vez. Aunque si es necesario, podríamos repetir información de manera controlada.
- **Interfaz de alto nivel:** mediante la utilización de lenguajes de alto nivel puede utilizarse la base de datos de manera sencilla y cómoda.
- **Consulta directa:** existe una herramienta para poder acceder a los datos interactivamente.

2.2. Usos

¿Quién utiliza las bases de datos?

Existen **cuatro tipos de personas** que pueden hacer uso de una base de datos:

- **el administrador** encargado del control técnico de todo el sistema, es la persona encargada de la creación o implementación física de la base de datos. Es quien toma las decisiones relacionadas con el funcionamiento físico del almacenamiento de información. Establecerá la política de seguridad y el acceso para garantizar el menor número de problemas
- **los diseñadores de la base de datos** Son las personas encargadas de diseñar cómo será la base de datos. Llevarán a cabo la identificación de los datos, las relaciones entre ellos, sus restricciones, etc. Para ello han de conocer a fondo los datos y procesos a representar en la base de datos. Si estamos hablando de una empresa, será necesario que conozcan las reglas de negocio en la que esta se mueve. Para obtener un buen resultado, el diseñador de la base de datos debe implicar en el proceso a todos los usuarios de la base de datos.
- **los programadores de aplicaciones** Una vez diseñada y construida la base de datos, los programadores se encargarán de implementar los programas de aplicación que servirán a los usuarios finales. Estos programas de aplicación ofrecerán la posibilidad de realizar consultas de datos, inserción, actualización o eliminación de los mismos. Para desarrollar estos programas se utilizan lenguajes de tercera o cuarta generación.
- **los usuarios finales** Son los clientes finales de la base de datos, utilizan aplicaciones desarrolladas para interactuar con la base de datos.

¿Para qué se utilizan las bases de datos?

Las bases de datos se utilizan en infinidad de campos:

- **Banca:** información de clientes, cuentas, transacciones, préstamos, ...
- **Líneas aéreas:** información de clientes, horarios, vuelos, destinos, ...
- **Universidades:** información de estudiantes, carreras, horarios, materias, ...
- **Transacciones de tarjeta de crédito:** para comprar con tarjetas de crédito y la generación de los extractos mensuales.
- **Telecomunicaciones:** para guardar registros de llamadas realizadas, generar facturas mensuales, mantener el saldo de las tarjetas telefónicas de prepago y almacenar información sobre las redes.
- **Medicina:** información hospitalaria, ...
- **Justicia y Seguridad:** delinquentes, casos, sentencias, investigaciones, ...
- **Organismos públicos:** datos ciudadanos, certificados, ...
- **Sistemas de posicionamiento geográfico.**
- **Hostelería y turismo:** reservas de hotel, vuelos, ...
- **Ocio digital:** juegos online, apuestas, ...
- **Cultura:** gestión de bibliotecas, museos virtuales, ...

2.3. Ubicación

Utilizamos a diario las bases de datos, pero ¿dónde se encuentra realmente almacenada la información?

Las bases de datos pueden tener un tamaño muy reducido (<1 Mb) o bien, ser muy voluminosas y complejas (del orden de Terabytes). Sin embargo todas las bases de datos normalmente se almacenan y localizan en discos duros y otros dispositivos de almacenamiento, a los que se accede a través de un ordenador. Una gran base de datos puede necesitar servidores en lugares diferentes, y viceversa, pequeñas bases de datos pueden existir como archivos en el disco duro de un único equipo.

Las principales **características** que definen a un sistema de almacenamiento de datos son:

- **Capacidad:** mide la cantidad de datos que puede almacenar, es medida en bytes (Gigabytes o Terabytes, habitualmente, aunque con el Big Data se manejan incluso Petabytes).
- **Rendimiento:** cómo de rápido y eficiente es el sistema de almacenamiento de datos.
- **Fiabilidad:** es la disponibilidad de los datos cuando son solicitados, así como el hecho de disponer de una baja tasa de errores o fallos.
- **Recuperabilidad:** mide la capacidad del sistema para recuperar datos tras una pérdida, borrado, corrupción o cualquier otro incidente que impida el acceso a los mismos.

Los **dispositivos de almacenamiento** son la clave para configurar un sistema de almacenamiento óptimo y seguro y poder garantizar el acceso, la integridad y la privacidad de la información. Los dispositivos de almacenamiento más utilizados en la actualidad son:

Discos

Uno de los principales medios para almacenar información es mediante el uso de discos. En la actualidad existen dos tipos de discos:

- **Discos duros:** Los **discos HDD** (*Hard Drive Disk*) son dispositivos mecánicos que utilizan el magnetismo para grabar los datos. Se componen de varios discos rígidos que giran a gran velocidad y cuyo cabezal de escritura/lectura se encarga de grabar o leer la información
- **Discos de estado sólido:** Las unidades de estado sólido SSD (*Solid State Drive*) almacenan la información en chips con memorias flash interconectadas (memorias NAND que mantienen la información cuando se corta el suministro eléctrico). Dentro de este tipo de dispositivos se pueden diferenciar tres tipos según su conexión: SATA, M.2 y PCIe NVME

El sistema de disco duro **NAS** (*Network Attached Storage*) permite colocar tantas unidades como se necesite. Es decir, cuando se termina la capacidad, basta con añadir un disco duro más.

Los sistemas NAS son en sí mismos pequeños ordenadores conectados a una red que permiten una gestión de datos muy eficiente y garantizan la seguridad en caso de un ciberataque.

Cintas magnéticas

Este tipo de dispositivo de almacenamiento graba los datos en pistas sobre una banda plástica que cuenta con material magnético. Existen distintos tipos de cintas de almacenamiento según su composición química o formatos de grabación que utilicen.

Cada sistema de almacenamiento en cinta dispone de sus propias características: su tamaño, tipo de contenedor del sistema, capacidad de almacenamiento o las características magnéticas de la cinta.

A pesar de que se trata de un sistema de almacenamiento de datos antiguo, actualmente tiene una gran utilidad cuando se manejan datos masivos. La gran durabilidad y su reducido coste las han convertido en un medio de almacenamiento que es aún utilizado para copias de seguridad.

Almacenamiento en red

Las redes han evolucionado y en la actualidad una red empresarial tiene una capacidad de transferencia de al menos 1000 Mbps (en caso de red de fibra óptica, de 10 Gbps), lo que permite transferir mucha información en poco tiempo. Esta velocidad de transferencia ha hecho populares métodos de almacenamiento en red como SAN (cuyo uso principal es en servidores de aplicaciones) o NAS (destinados sobre todo a almacenamiento empresarial o personal).

Sistemas SAN (*Storage Area Network o red de área de almacenamiento*): Se trata de una red concebida para conectar servidores, matrices (arrays) de discos y librerías de soporte. La arquitectura de este tipo de sistemas permite que los recursos de almacenamiento estén disponibles para varios servidores en una red de área local o amplia. Debido a que la información almacenada no reside directamente en ninguno de los servidores de la red, se optimiza el poder de procesamiento para aplicaciones comerciales y la capacidad de almacenamiento se puede proporcionar en el servidor donde más se necesite.

Sistemas NAS (*Network Attached Storage ó sistema de almacenamiento masivo en red*). Estos sistemas de almacenamiento permiten compartir la capacidad de almacenamiento de un computador (Servidor) con ordenadores personales o servidores clientes a través de una red, haciendo uso de un sistema operativo optimizado para dar acceso a los datos a través de protocolos de comunicación específicos. Suelen ser dispositivos para almacenamiento masivo de datos con capacidades muy altas, de varios Terabytes, generalmente superiores a los discos duros externos y además se diferencian de estos al conectar por red.

Almacenamiento en la nube

Hoy en día los sistemas de almacenamiento de datos tienden a migrar al almacenamiento cloud. Se trata de contratar un espacio externo (en los servidores de un proveedor de almacenamiento en la nube) donde es posible almacenar toda la información y acceder a ella independientemente del sistema operativo, ubicación, hora o dispositivo que se utilice.

Las ventajas del almacenamiento en la nube en cuanto a seguridad, portabilidad y accesibilidad lo han convertido en el sistema de almacenamiento de datos más utilizado por las empresas de éxito en la actualidad.

El almacenamiento en la nube implica disponer de una constante conexión a Internet por lo que es recomendable utilizar distintos tipos de almacenamiento para garantizar el acceso a la información en todo momento.

3. Sistemas gestores de bases de datos

El sistema gestor de bases de datos, o SGBD actúa como interfaz entre los datos y las aplicaciones que utilizan esos datos, y proporciona mecanismos para definirlos, almacenarlos de manera óptima, mantenerlos y manipularlos. Además ofrece potentes herramientas para la administración orientada a la gestión de usuarios y el almacenamiento de los datos, gestión de instancias, control de transacciones, recuperación y copias de seguridad.

Las principales funciones de un SGBD son:

- **Definición y control centralizado:** definición de los datos y las relaciones entre las estructuras que los almacenan.
- **Manipulación de los datos:** mecanismos para consultar y modificar datos almacenados o añadir otros nuevos.
- **Seguridad e integridad:** gestión de accesos autorizados y control de las restricciones para garantizar la integridad y consistencia de los datos
- **Tareas de administración:** copias de seguridad, recuperación ante fallos, roles, control de privilegios, etc
- **Control de recurrencia:** garantiza el acceso simultáneo a los datos y la veracidad de éstos

3.1. SGBD Comerciales

Actualmente, en el mercado de software existen multitud de sistemas gestores de bases de datos comerciales:

SGBD	Descripción
Oracle	Reconocido como uno de los mejores a nivel mundial. Es multiplataforma, confiable y seguro. Es Cliente/Servidor. Basado en el modelo de datos Relacional. De gran potencia, aunque con un precio elevado hace que sólo se vea en empresas muy grandes y multinacionales. Ofrece una versión gratuita Oracle Database 11g Express Edition.
MySQL	Sistema muy extendido que se ofrece bajo dos tipos de licencia, comercial o libre. Para aquellas empresas que deseen incorporarlo en productos privativos, deben comprar una licencia específica. Es Relacional, Multihilo, Multiusuario y Multiplataforma. Su gran velocidad lo hace ideal para consulta de bases de datos y plataformas web.
DB2	Multiplataforma, el motor de base de datos relacional integra XML de manera nativa, lo que IBM ha llamado pureXML, que permite almacenar documentos completos para realizar operaciones y búsquedas de manera jerárquica dentro de éste, e integrarlo con búsquedas relacionales.
INFORMIX	Otra opción de IBM para el mundo empresarial. Es un gestor de base de datos relacional basado en SQL. Multiplataforma. Consume menos recursos que Oracle, con utilidades muy avanzadas respecto a conectividad y funciones relacionadas con tecnologías de Internet/Intranet, XML, etc.
Microsoft SQL SERVER	Sistema Gestor de Base de Datos producido por Microsoft. Es relacional, sólo funciona bajo Microsoft Windows, utiliza arquitectura Cliente/Servidor. Constituye la alternativa a otros potentes SGBD como son Oracle, PostgreSQL o MySQL.
SYBASE	Un DBMS con bastantes años en el mercado, tiene 3 versiones para ajustarse a las necesidades reales de cada empresa. Es un sistema relacional, altamente escalable, de alto rendimiento, con soporte a grandes volúmenes de datos, transacciones y usuarios, y de bajo costo.

3.4. SGBD Libres

SGBD	Descripción
MySQL	Es un sistema de gestión de base de datos relacional, multihilo y multiusuario con más de seis millones de instalaciones. Distribuido bajo dos tipos de licencias, comercial y libre. Multiplataforma, posee varios motores de almacenamiento, accesible a través de múltiples lenguajes de programación y muy ligado a aplicaciones web.
PostgreSQL	Sistema Relacional Orientado a Objetos. Considerado como la base de datos de código abierto más avanzada del mundo. Desarrollado por una comunidad de desarrolladores que trabajan de forma desinteresada, altruista, libre y/o apoyados por organizaciones comerciales. Es multiplataforma y accesible desde múltiples lenguajes de programación.
Firebird	Sistema Gestor de Base de Datos relacional, multiplataforma, con bajo consumo de recursos, excelente gestión de la concurrencia, alto rendimiento y potente soporte para diferentes lenguajes.
Apache Derby	Sistema Gestor escrito en Java, de reducido tamaño, con soporte multilenguaje, multiplataforma, altamente portable, puede funcionar embebido o en modo cliente/servidor.
SQLite	Sistema relacional, basado en una biblioteca escrita en C que interactúa directamente con los programas, reduce los tiempos de acceso siendo más rápido que <i>MySQL</i> o <i>PostgreSQL</i> , es multiplataforma y con soporte para varios lenguajes de programación.

4. Ciclo de vida de una base de datos

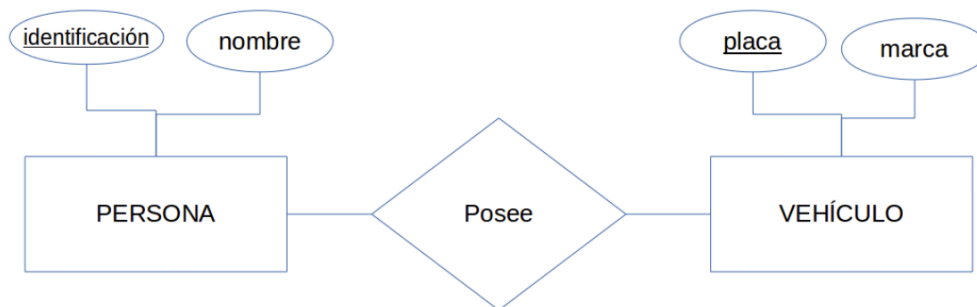
El ciclo de vida de las BBDD están totalmente integrado en la ingeniería del software y sus fases están relacionadas con ellas.



Imagen 1. Ciclo de vida de las BBDD

Podemos distinguir las siguientes fases:

- **Planificación:** qué vamos a hacer, con qué recursos contamos, qué técnicas se usan para la recopilación de los datos, formato, cómo se hará el diseño, ...
- **Definición:** se especifica el ámbito y los límites de aplicación de la base de datos, con qué sistemas actúa, se determinan los usuarios,...
- **Recogida y análisis de requisitos:** se recogen y analizan los requerimientos de los usuarios y de las áreas de aplicación, a través de entrevistas, observación del funcionamiento de la organización, ...
- **Diseño de la base de datos:** está dividido en 3 partes
 - o **Diseño conceptual:** es un esquema conceptual que parte de los requisitos del usuario y es independiente del SGBD



- o **Diseño lógico:** tras refinar el esquema conceptual, se obtiene el esquema lógico, en esta fase debemos pensar cómo normalizar nuestras tablas para evitar duplicar información y ahorrar espacio. También es independiente del SGBD

Persona	Posee	Vehículo
Identificación		Placa
Nombre		marca

- o **Diseño físico:** En esta última fase ya debemos ver qué tipos de datos utilizaremos, sus dominios (qué valores va a permitir), qué índices debemos crear para optimizar las consultas, entre otros. Aquí ya escribimos nuestro SQL para plasmar todo nuestro diseño en el SGBD elegido.

```
CREATE TABLE PERSONA (  
  Identificación INT,  
  Nombre varchar(150)  
)
```

- **Diseño de la aplicación:** diseño de los programas que interactuarán con la base de datos
- **Prototipado:** un prototipo que el usuario podrá utilizar para ver qué aspectos son o no adecuados. Este paso ayuda a detectar posibles errores en la toma de requisitos.
- **Implementación:** usando el lenguaje de definición de datos se crea el esquema de la base de datos, se implementan procesos, etc,...
- **Conversión y carga de datos:** puede que la base de datos deba crearse por la actualización de una aplicación, debemos reemplazar los datos antiguos por los nuevos incluidos en distintas estructuras
- **Prueba y mantenimiento:** se prueban y validan los requisitos del usuario, se diseñan test con datos reales y se toman medidas sobre la fiabilidad y calidad del software desarrollado.