

# 1. Introducción

Una base de datos es el almacén para guardar datos de forma organizada.

Han existido tres tipos de bases de datos a lo largo de la historia:

- Papel y bolígrafo
- Sistemas de ficheros: word, excel, etc.
- Sistemas de gestión de bases de datos

## 2. Conceptos Básicos

- Sistema de gestión de bases de datos: es un conjunto de herramientas, que permiten manipular y gestionar la información de una base de datos.

Estos sistemas han de tener una serie de características:

- Permisos: se han de establecer niveles de acceso a la información.
  - Seguridad de la información.
  - Organización de los datos.
- Sistemas de Información:
    - Bit: la unidad de información más pequeña (valores de 0 o 1).
    - Byte: conjunto de 8 bits (palabra).
    - Campo: cada dato al que puede hacerse referencia en una base de datos.
    - Registro: conjunto de campos que definen a un objeto de la base de datos.
    - Archivo: todos los registros del mismo tipo de objetos.
    - Base de datos: conjunto de archivos, compone la información del sistema.
  - El archivo: lugar donde se almacena la información de una base de datos. La base de datos se divide en distintos archivos.
    - Al residir en soportes externos, su existencia no está limitada al tiempo de ejecución del programa, sino que permanece cuando éste termina.
    - Los datos pueden transportarse de un equipo a otro.
    - A efectos prácticos, tienen capacidad de almacenamiento ilimitada, ya que, aunque un soporte dispone de capacidad limitada, se pueden adquirir más.

Existen distintos tipos de archivos según su función:

- Permanentes: sus registros varían poco en el tiempo.
  - Constantes: permanecen prácticamente invariables.
  - De situación: reflejan el estado actual de una empresa. Se actualizan periódicamente para adaptarlos a una nueva situación.
  - Históricos: se obtienen de los anteriores cuando estos dejan de utilizarse. Se utilizan para estudios estadísticos o de consulta.
- De movimiento: almacenan temporalmente la información que se utiliza para actualizar los archivos de situación.
- De maniobra: archivos temporales creados durante la ejecución de un programa y borrados al terminar el proceso.

### **3. Operaciones en Archivos**

.CREATE      .OPEN      .READ      .RENAME      .DELETE

### **4. Acceso a la Información**

- Acceso secuencial: para leer o escribir un dato es necesario pasar por todos los registros ya almacenados.
- Acceso directo: permite acceder directamente a una ubicación del dispositivo.
- Acceso índice: tenemos un índice donde podemos buscar la ubicación que nos interesa. La lectura del índice se realiza de forma secuencial. Una vez tenemos la ubicación que nos interesa, la lectura o escritura se realiza de forma directa.

### **5. Gestión de la Información**

Existen dos tipos de soportes:

- Secuenciales: los datos se graban unos a continuación de otros, es decir, para leer o escribir un dato debemos recorrer todos los datos anteriormente almacenados (escritura CD, escritura DVD, etc.).
- Direccionables: el espacio de almacenamiento se divide en espacios direccionables, pudiendo acceder a un dato por la dirección en que está almacenado (USB, etc.)

A la hora de almacenar la información, a cada archivo se le asignan bloques.

Hay tres métodos de asignación:

- Asignación contigua: todos los bloques deben estar en posiciones contiguas.
  - Acceso secuencial: sencillo.
  - Acceso directo: sencillo.
  - Antes de guardar un archivo, es necesario determinar el número de bloques que ocupará el archivo en disco.
  - Si los bloques vacíos no son contiguos → fragmentación externa
- Asignación enlazada: fragmenta el archivo en bloques que pueden estar distribuidos aleatoriamente por el disco. El sistema operativo guarda la dirección del primer bloque y cada bloque almacena en sus últimos bytes la dirección del siguiente bloque en secuencia.
  - Acceso secuencial: sencillo.
  - Acceso directo: imposible.
- Asignación indexada: el bloque índice es una tabla de entradas, donde la entrada "x" tiene la dirección del bloque respectivo del archivo.
  - Acceso secuencial: sencillo (facilitado por los índices).
  - Acceso directo: sencillo (facilitado por los índices).

## 6. SGBD

Sus componentes más importantes son el hardware, los datos, los usuarios y los procedimientos encargados del uso correcto de la información (software).

El sistema de control de base de datos es el encargado de recoger las peticiones que se realicen y devolver los datos cuando sea posible.

El administrador del sistema organiza y controla los recursos del sistema:

- Define la estructura de la base de datos con su lenguaje.
- Controla el acceso a la base de datos, concediendo permisos a los usuarios.
- Define estrategias de recuperación frente a posibles fallos.

Los objetivos principales del SGBD son:

- Garantizar la independencia de los datos frente a las aplicaciones.
- Redundancia mínima de información.
- Versatilidad en las búsquedas de información.
- Denegar accesos no autorizados.
- Controlar la integridad de la información.
- Dotar de mecanismos para realizar copias de seguridad.
- Solucionar problemas de concurrencia.

Los componentes del SGBD son:

- Núcleo: encargado de coordinar y controlar el funcionamiento.
- Lenguajes: permiten la definición de la base de datos y el manejo de sus datos.
  - Lenguaje de descripción de datos (DDL): define la estructura de la base de datos, además de los procedimientos de seguridad. Para administradores.
  - Lenguaje de manipulación de datos (DML): gestiona la información (añadir, eliminar, modificar y recuperar información). Utilizado por los usuarios.
- Utilidades: aplicaciones que facilitan el trabajo a los usuarios y programadores. Incorporan un interfaz fácil de entender, con menús que guían al usuario.
- Diccionario de datos: guarda toda la información referente a las “normas” del sistema de gestión de base de datos y formato de las órdenes.

Si la base de datos es compleja, el diccionario puede dividirse en tres capas:

- Global: información común a todos los usuarios, incluido el administrador.
- Intermedia: representa las percepciones individuales que tiene cada usuario.
- Local: se representan los datos específicos de la base de datos.

En 1975, ANSI-SPARC propuso una arquitectura de tres niveles. Su objetivo es el de separar los programas de aplicación de la base de datos física.

Para lograr esta separación, se utilizan tres niveles:

- Externo: Es el conjunto de percepciones individuales de la base de datos. Cada visión individual de un usuario se denomina vista. Al crear una vista, el usuario puede seleccionar los campos que le interesen.

- Interno: describe la estructura física de la base de datos mediante un esquema interno, que especifica todos los detalles para el almacenamiento y el acceso.
- Conceptual: se describe la estructura enfocada a la empresa y usuarios. Este esquema oculta los detalles de las estructuras y se concentra en describir entidades, atributos, relaciones, operaciones de los usuarios y restricciones.

La independencia de datos es la capacidad para modificar el esquema en un nivel del sistema sin tener que modificar el esquema del nivel inmediato superior.

Existen dos tipos de independencia de datos:

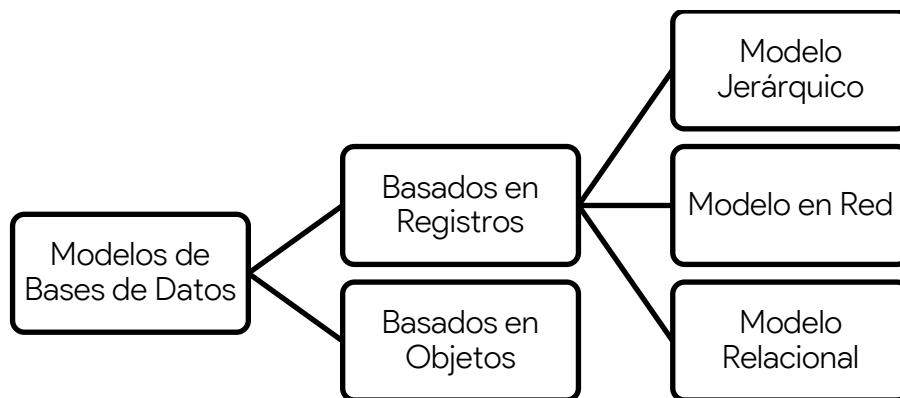
- Lógica: capacidad de modificar el esquema conceptual sin tener que alterar los esquemas externos ni los programas de aplicación.
- Física: capacidad de modificar el esquema interno sin tener que alterar el esquema conceptual (o los externos).

Estos son los principales Sistemas de gestión de Bases de Datos:

- MySQL (Oracle):
  - Gratuito.
  - Código abierto.
  - Apoyo de la comunidad: más seguridad, abundante documentación, etc.
- ACCESS (Microsoft):
  - Para principiantes y pequeñas empresas.
  - Interfaz clara y con abundantes funcionalidades.
  - Ofrece una versión gratuita.
- SQL Server (Microsoft):
  - Para programadores y grandes empresas.
  - Más potente.
  - Se posiciona como la alternativa a Oracle.
  - Ofrece una versión de prueba de 180 días.
- DB2 (IBM):
  - Ofrece la mejor seguridad.
  - Gran compatibilidad con distintos sistemas operativos.
  - Ofrece una versión gratuita.

- Oracle (Oracle):
  - El más utilizado.
  - Índices denominados “Tablespaces”.
  - Ofrece una versión gratita en su web.

## 7. Tipos de Bases de Datos



- Modelos basados en registros: la base de datos está estructurada en registros de varios tipos, cada uno con un número fijo de campos.

Existen 3 tipos de modelos basados en registros:

- Jerárquico: es el primer modelo que se implantó (años 70), y muchos de los modelos actuales se basan en él. Utiliza estructuras con forma de árbol.
  - En red: modelo propio de los años 70. Consiste en una estructura de nodos conectados, con la diferencia de que un nodo puede tener varios padres.
  - Relacional: alternativa al modelo jerárquico y en red, que ofrece más flexibilidad en el tratamiento de los datos. Sus características son:
    - Los datos se almacenan en tablas o entidades.
    - Utilizan un esquema predefinido que define la estructura.
    - Utilizan el lenguaje SQL para la gestión de la base de datos.
    - La integración se asegura con relaciones coherentes entre las tablas.
- Modelos orientados a objetos: en desarrollo, pretende sustituir al relacional. Sustituye el acceso a un dato concreto, por un conjunto de datos, los objetos.

Presenta ventajas como la facilitación tanto de las funciones de los SGBD, como del uso de lenguajes de programación actuales.

- Objetos: cada entidad del mundo real es modelada como un objeto.
- Objetos complejos: los atributos de un objeto pueden ser objetos.
- Encapsulamiento: cada objeto tiene definidos los métodos y la interfaz para tener acceso y manipular dicho objeto, es decir, restricciones. Por ejemplo, establecer que los números de teléfono sean de nueve dígitos.
- Clases: todos los objetos definidos por los mismos atributos y métodos forman una clase. Cada objeto de una clase es una instancia de la clase.
- Herencia: una subclase heredará los atributos de otras clases existentes.

Modelo Relacional	Modelo Orientado a Objetos
Entidades	Objetos
Atributos	Objetos complejos
Restricciones	Encapsulamiento

Los avances han impulsado un tipo de base de datos que ha roto con el esquema tradicional de las bases de datos relacionales, las no relacionales (NoSQL).

- Son bastante más rápidas que las relacionales.
- Permiten almacenar grandes cantidades de información.
- Aportan mayor flexibilidad a la hora de modificar el esquema de la información.
- No hacen uso del lenguaje SQL tradicional.

## 8. Bases de Datos Distribuidas

Una base de datos distribuida (BDD) es una colección de datos distribuidos en diferentes nodos de una red de computadoras. Representan una estructura geográficamente descentralizada, aumentan la disponibilidad de los datos y reducen el tráfico de comunicación.

Sus ventajas son:

- El acceso a los datos es más rápido debido a que se localizan más cercanos al lugar donde se utilizan.
- El procesamiento es más rápido debido a que varios nodos intervienen en el procesamiento de una carga de trabajo.
- En caso de que un nodo sufra una avería, el sistema no se ve afectado.
- Mayor tolerancia a los fallos.

Sus desventajas son:

- Es complejo el control y la manipulación de los datos.
- Es complejo asegurar la integridad de la información en presencia de fallos.
- El control de concurrencia y los mecanismos de recuperación son mucho más complejos que en un sistema centralizado.
- Requieren una configuración previa para su correcto funcionamiento.

Hay dos tipos de bases de datos distribuidas:

- Homogéneas: todos los sitios tienen la misma base de datos y cooperan entre ellos en el procesamiento de las solicitudes.
- Heterogéneas: cada sitio puede tener distintas bases de datos.

La etapa diferenciadora entre el diseño de una centralizada y una distribuida es el diseño de la distribución, que consta de dos actividades:

- Fragmentación: decidir cómo se divide la base de datos y en qué partes.
- Asignación: decidir dónde se ubica cada parte, así como si se tienen réplicas de los datos.

## 9. Modelo Entidad-Relación

Esquema a de una base de datos, independiente de su estructura final.

Se rige por 3 características:

- Refleja tan sólo la existencia de los datos, no lo que se hace con ellos.
- Es independiente de las bases de datos y sistemas operativos concretos.
- No tiene en cuenta restricciones de espacio, almacenamiento, ni de tiempo.

Sus elementos son los siguientes:

- Entidades: todo objeto real o abstracto del que se quiere obtener información. Representa el nombre de cada tabla.
- Atributos: información propia de las entidades. Hay dos tipos:
  - Identificadores
  - Descriptores



- Relaciones: son las asociaciones entre objetos del mundo real. También pueden contener atributos propios. Cada una puede unir una o varias entidades.

Tienen sus propias tablas y dentro ellas, además de contener sus atributos, la relación toma las claves identificadoras de las entidades a las que relaciona.

La participación define el mínimo y máximo de ocurrencias que le pueden corresponder a una ocurrencia de la otra relación: (0,1) , (1,1) , (0,n) , (1,n)

La cardinalidad representa el número máximo de ocurrencias asociadas al número máximo de ocurrencias del resto de las entidades relacionadas.

## 10. Modelo Relacional

- Cardinalidad 1:1
  - Participación (1,1) (1,1): cada tabla tendrá como FK el PK de la otra tabla.
  - Participación (0,1) (1,1): se propaga la PK de la (1,1) a la de (0,1).
  - Participación (0,1) (0,1): tres tablas, una para cada entidad y otra para la relación, que contendrá las PK de las dos entidades (ambas como PK y FK).
- Cardinalidad 1:N
  - Participación (1,1) (X,N): se propaga la PK de (1,1) a la entidad de N.
  - Participación (0,1) (X,N): tres tablas, una para cada entidad y otra para la relación, cuya PK será la de N, y las FK de ambas.
- Cardinalidad N:M
  - Participación (X,N) (X,M): tres tablas, una para cada entidad y otra para la relación, cuyas PK y FK serán las de ambas.
- Relaciones Reflexivas

*Tabla [PK, atributos, PK\_Modificada (FK de tabla)].*
- Relaciones Ternarias
  1. Participación (X,N) (X,M) (X,P): la relación tiene tres PK y FK.
  2. Participación (X,N) (X,M) (X,1): la relación tiene 2 PK (de N y M) y 3 FK.
  3. Participación (X,N) (X,1) (X,1): la relación tiene 2 PK (de N y una de 1) y 3 FK.
  4. Participación (X,1) (X,1) (X,1): la relación tiene 2 PK de 1 y 3 FK.

## 11. Normalización

Consiste en aplicar una serie de reglas a las relaciones obtenidas al pasar del modelo entidad-relación al relacional.

Se normaliza para evitar la redundancia (datos duplicados), evitar problemas de actualización de los datos en las tablas, proteger la integridad de los datos, etc.

No siempre es la forma más eficiente de representar la información, ya que la creación de nuevas tablas hace que las consultas de datos sean más lentas.

- 1FN: se requieren datos atómicos en las tablas.
  1. Todos los registros de la tabla deben tener una clave primaria como mínimo.
  2. Los campos de la tabla no deben contener multivaluados.
  3. La tabla no ha de contener datos redundantes.
  4. La tabla no ha de contener atributos con valores null.
- 2FN: todos los atributos no clave deben depender de toda la clave primaria.
- 3FN: todos los atributos no primarios han de ser dependientes de una clave. Cada uno de los atributos ha de estar totalmente correspondido con la clave de la tabla donde se encuentre.
- 4FN: las multivaluadas han de estar representadas de la manera más eficiente.