



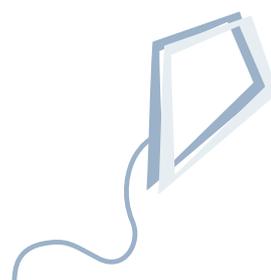
GOBIERNO  
DE ESPAÑA

MINISTERIO  
DE EDUCACIÓN

# DISEÑO DE BASES DE DATOS RELACIONALES

Varios autores

PROGRAMACIÓN



AULA  
MENTOR

[educacion.es](http://educacion.es)



NIPO: 820-09-112-X

Autores:

Dolores Cuadra

Elena Castro

Paloma Martínez

Revisión técnica y ampliación de contenidos:

Emilia Carballo

Susana Pérez

Coordinación Pedagógica:

M<sup>a</sup> Cinta Cascales

Diseño gráfico de portada:

Lorena Gordo López

## Módulo A

Unidad didáctica 1: Introducción a las Bases de Datos	3
Unidad didáctica 2: Metodologías de desarrollo de Bases de Datos	19

## Módulo B

Unidad didáctica 1: Fase de análisis de requisitos – Modelo E/R	33
Unidad didáctica 2: Modelo Relacional	113

## Módulo C

Unidad didáctica 1: Álgebra relacional	189
Unidad didáctica 2: SQL – Lenguaje de consulta estructurado	199

# MÓDULO A

## UNIDADES DIDÁCTICAS:

1. Introducción a las Bases de Datos
2. Metodologías de Desarrollo de Bases de Datos

# Introducción a las Bases de Datos

## UNIDAD DIDÁCTICA 1

### Índice de la unidad:

1. La Base de Datos como un componente de los sistemas de información
2. Definición de la Base de Datos
3. Sistema de Gestión de Bases de Datos
4. Arquitectura de Bases de Datos a tres niveles
5. Lenguajes de SGBD
6. Herramientas de SGBD
7. Algunas arquitecturas de Sistemas de Bases de Datos

### En esta unidad aprenderás:

- Qué es una base de datos y sus principales ventajas
- Qué es un sistema gestor de bases de datos y sus principales funciones
- Arquitectura de una base de datos según los tipos de usuarios
- Qué es el lenguaje SQL
- Diferencias entre Sistemas centralizados, cliente-servidor, sistemas paralelos y sistemas distribuidos

## 1. La Base de Datos como un componente de los Sistemas de Información

Desglosando en detalle los componentes de un SI nos encontramos con cinco grandes componentes:

- El Contenido, es decir, los datos con su correspondiente descripción, almacenados en un soporte de ordenador (por ejemplo, en unos grandes almacenes se tendrían los datos de clientes, ventas, productos, etc.).
- Equipo físico (hardware) formado por la unidad central de proceso y los equipos periféricos (discos, terminales, impresoras, redes, ...).
- Equipo lógico (software) compuesto por los programas, documentación, lenguajes de programación, etc. que debe gestionar los datos (creación, consulta, recuperación y mantenimiento) así como controlar las comunicaciones y dar soporte a tratamientos específicos (por ejemplo, gestión de personal, facturación, etc.).
- El Administrador, encargado de asegurar la calidad de los datos almacenados y de permitir su uso correcto y permanente. El administrador o administradores debe controlar la disponibilidad, la confidencialidad y la integridad de los datos. La disponibilidad se refiere a que los datos deben estar accesibles en todo momento, es decir, que ante cualquier tipo de catástrofe o fallo, se tengan los mecanismos adecuados de recuperación para que el sistema siga funcionando; la confidencialidad se encarga de no desvelar datos a usuarios no autorizados y la integridad asegura que los datos no se falseen, es decir, que sean correctos, válidos y precisos.
- Un conjunto de Usuarios formado por las personas que acceden al sistema de información y que pueden ser de dos tipos: informáticos (analistas y programadores encargados de desarrollar las aplicaciones, bases de datos, etc.) y los usuarios finales con pocos conocimientos de informática que requieren consultas y actualizar los datos mediante interfaces adecuados a sus características.

El componente que nos interesa en este libro, la Base de Datos, se encuentra en el software del sistema de información. A continuación, vamos a dar una definición más precisa de este componente que desglosaremos en todos sus aspectos. Posteriormente, estudiaremos la necesidad de las Bases de Datos en los sistemas de información actuales en contraposición a los sistemas de ficheros.

## 2. Definición de Base de Datos

**Definición:** Una Base de Datos (BD) es una colección o depósito de datos integrados, almacenados en soporte secundario (no volátil) y con redundancia controlada. Los datos, que han de ser compartidos por diferentes usuarios y aplicaciones, deben mantenerse independientes de ellos y su definición (estructura de la BD), única y almacenada junto con los datos, se ha de apoyar en un modelo de datos, el cual ha de permitir captar las interrelaciones y restricciones existentes en el mundo real. Los procedimientos de actualización y recuperación, comunes y bien determinados, facilitarán la seguridad del conjunto de los datos. De Miguel et. al (1999)

Veamos en qué consiste cada uno de los aspectos mencionados en esta definición de Base de Datos que no son más que distintas definiciones según distintas perspectivas.

- La Base de Datos es un conjunto de datos relativos a una determinada parcela del mundo real (por ejemplo, una biblioteca, una empresa petroquímica, una universidad, etc.,) que se almacenan en un soporte informático no volátil (es decir, dispositivos de memoria secundaria como discos, cintas, etc. que hacen que los datos no desaparezcan "cuando no se están usando").
- Además, no debe existir redundancia (el término redundancia es sinónimo de datos repetidos), es decir, no deben existir duplicidades perjudiciales ni innecesarias (a ser posible un determinado tipo de dato, por ejemplo, los datos de un cliente de una empresa, sólo deben aparecer en un sitio en la BD). En ocasiones, es necesaria cierta redundancia (a nivel de almacenamiento físico<sup>1</sup>) que mejora la eficiencia de la BD, por ejemplo, ante determinados tipos de consultas de datos. Sin embargo, esta redundancia siempre debe ser controlada por el sistema para que no se produzcan inconsistencias; piense el lector qué sucedería si los datos de los clientes de una empresa se repiten en varias

---

<sup>1</sup> En secciones posteriores estudiaremos que se habla de tres niveles (conceptual, lógico y físico) en la arquitectura de una BD. Por ahora, nos basta saber que el nivel físico concierne a cómo se almacenan los datos en los ficheros de la BD.

partes de la BD y no se controlara: puede ocurrir que si un cliente cambia de dirección postal y sólo se actualiza esta información en uno de los sitios, entonces la BD quedaría en estado inconsistente (el cliente aparece con datos distintos en distintas partes de la BD). Posteriormente, se volverá al tema de la redundancia cuando se estudien los modelos de datos como herramientas de diseño de Bases de Datos.

Por otro lado, las BD han de atender a múltiples usuarios de la organización (informáticos que desarrollan programas de acceso a la BD, administrativos usuarios de las aplicaciones, etc.) así como a distintas aplicaciones (por ejemplo, aplicaciones de contabilidad, de facturación, etc., todas ellas accediendo a los datos contenidos en la BD de la empresa).

Esta visión unificada de los datos se contrapone con los sistemas tradicionales de ficheros. Aunque no entraremos aquí en detalles de en qué consisten los sistemas de ficheros que se utilizaban con anterioridad a la aparición de las BD sí indicaremos algunos aspectos relevantes que los diferencian de las BD; estos aspectos son:

- Independencia de datos y procesos: es el objetivo fundamental de las BD, mantener separados los datos de los tratamientos que los utilizan. En los sistemas basados en ficheros cada fichero se diseñaba para responder a las necesidades de una aplicación determinada y apenas podían utilizarse por otra aplicación. En las BD, los datos se encuentran en un único almacén y son accedidos por todas las aplicaciones.
- Descripción de los datos junto con los datos: la definición y descripción<sup>2</sup> del conjunto de datos contenidos en la BD deben ser únicas y estar integradas con los mismos datos. Posteriormente se verá que los datos están interrelacionados y estructurados de acuerdo a un modelo capaz de recoger el máximo contenido semántico. En los sistemas de ficheros, los datos se encuentran en distintos ficheros diseñados expresamente (ad-hoc) para cada tipo de aplicación, y la descripción de los datos se encuentra junto con los programas de la aplicación. Recuérdese, por ejemplo, un programa en un lenguaje de programación en el que al principio del código se define de qué tipo son las variables y las

---

<sup>2</sup> En las siguientes secciones se verá que la descripción de los datos es lo que se denomina estructura o esquema de la BD.

estructuras de datos que se van a manejar en el programa que accede a los ficheros.

- Procesos de actualización y recuperación bien establecidos: la recuperación y actualización de los datos se realiza de acuerdo a procesos bien determinados que se incluyen en el Sistema de Gestión de Bases de Datos<sup>3</sup> (SGBD). El SGBD también proporcionará los instrumentos necesarios para el mantenimiento de la seguridad (confidencialidad, disponibilidad e integridad) del conjunto de datos.

---

<sup>3</sup> En el siguiente apartado se estudiará que es un SGBD. Por ahora, sólo decir que el conjunto de programas y herramientas que nos permite crear, actualizar, manipular y mantener una BD.

### 3. Sistemas de Gestión de Bases de Datos

En la sección anterior se ha estudiado qué es una BD y se ha mencionado alguna de las funcionalidades de un SGBD. En esta sección nos centraremos en describir en profundidad qué es un SGBD y cuál es la funcionalidad que debe proporcionar.

**Definición:** Un Sistema de Gestión de Bases de Datos (SGBD<sup>4</sup>) es un conjunto coordinado de programas, procedimientos, lenguajes, herramientas, etc., que suministra, tanto a los usuarios no informáticos como a los analistas, programadores o administradores de una BD, los medios necesarios para describir y manipular los datos integrados en la BD, manteniendo su integridad, confidencialidad y disponibilidad. De Miguel et al. (1999).

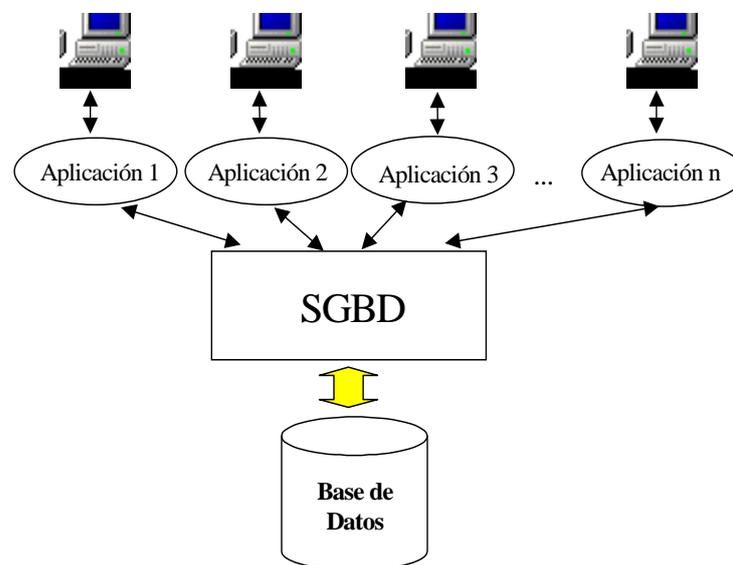


Figura 1.1: Sistema de Bases de Datos

<sup>4</sup> En inglés, Database Management System (DBMS).

Se denomina Sistema de Bases de Datos a la unión de una BD, un SGBD más las aplicaciones que acceden a la BD. La Figura 1.1 muestra la arquitectura de un sistema de BD; en ella se observa que el SGBD hace de interfaz entre los usuarios que acceden a la BD mediante las aplicaciones y la Base de Datos que contiene toda la información.

Ya hemos mencionado en la sección anterior que existen distintos tipos de usuarios (administradores, programadores, usuarios finales) con necesidades diferentes. Para poder dar soporte a estos usuarios el SGBD debe proporcionar una serie de funciones que se describen a continuación:

- Función de definición: permite a los diseñadores de la BD describir los elementos de datos, su estructura y las relaciones que existen entre ellos; como se estudiará más adelante, el SGBD proporciona un lenguaje para la definición las tablas, los atributos que la componen, las restricciones semánticas así como las características de tipo físico o almacenamiento.
- Función de manipulación: permite a los usuarios de la BD añadir, suprimir o modificar los datos de la misma siempre y cuando se respeten los aspectos de seguridad que haya establecido el administrador de la BD.
- Función de control: esta función aúna los interfaces que requieren los distintos tipos de usuarios para comunicarse con la BD así como las herramientas necesarias para el administrador para establecer los mecanismos de seguridad y mantenimiento de la BD.

Para que el SGBD pueda llevar a cabo estas funciones se necesita un lenguaje que permita especificar lo que cada tipo de usuario necesita en su comunicación con la BD. En las BD relacionales se emplea el SQL (Standard Query Language).

#### 4. Arquitectura de BD a tres niveles

A continuación, vamos a definir cuál es la arquitectura de una BD según la visión que de ella tienen los distintos tipos de usuario. En una BD se identifican tres capas de estructuración según tres niveles de abstracción. Así, se distingue un nivel externo, un nivel lógico y un nivel físico.

- El nivel externo se corresponde con la visión de la BD que cada usuario tiene en particular. Esto significa que no todos los usuarios necesitar conocer la BD completa sino que únicamente necesitan una vista parcial de ella (la que le permita llevar a cabo su trabajo); por ejemplo, un administrativo que trabaje elaborando las nóminas de los empleados de una empresa no necesita conocer los datos relativos a las ventas de productos de esa empresa.
- El nivel lógico se corresponde con la visión total de la empresa; esta vista global se interpone entre el nivel externo y el nivel físico siendo independiente tanto del equipo como de cada usuario en particular; por ejemplo, el administrador de la BD si necesita tener una vista completa de la BD de la empresa para llevar a cabo su trabajo.
- El nivel físico se corresponde con la vista del soporte físico informático en cuanto a que se refiere a la forma en que se organizan los datos en el almacenamiento físico (índices o punteros, longitud de los campos, caminos de acceso a los datos, particionamientos de memoria, etc.).

La gestión de estos tres niveles debe estar soportada en cualquier SGBD.

## 5. Lenguajes de un SGBD

De acuerdo a las funciones a las que debe dar soporte un SGBD estudiadas en el apartado 3 y a los distintos niveles de estructuración de una BD vistos en el apartado 4, los SGBD deben proporcionar un lenguaje para que los distintos tipos de usuario puedan comunicarse con la BD. Así, en los SGBD relacionales se tiene el lenguaje SQL que de acuerdo a su función se descompone en:

- Lenguaje de Definición de Datos (LDD): utilizado para definir la estructura lógica de la BD (nivel lógico), la estructuras externas requeridas para el desarrollo de las diferentes aplicaciones (nivel externo) así como la estructura interna (nivel físico).
- Lenguaje de Manipulación de Datos (LMD): una vez se ha descrito la BD, ésta ya está preparada para cargar los datos en las estructuras definidas y para su utilización. Así, el LMD permite añadir, suprimir, modificar y buscar datos en la BD. Es el SGBD el que se encarga de acceder al correspondiente soporte físico para localizar los datos con los que se harán las operaciones especificadas.
- Lenguaje de Control: el administrador de la BD utiliza este lenguaje para especificar los aspectos de seguridad física (copias de seguridad, re arranque de la BD en caso de caída, etc.) así como de protección frente a accesos no permitidos (autorizaciones y contraseñas, perfiles de usuarios, etc.). El lenguaje de control también se requiere para definir los interfaces que necesitan los distintos usuarios para comunicarse con la BD.

## 6. Herramientas de un SGBD

Aparte de los lenguajes vistos en el apartado 5, los SGBD proporcionan otro tipo de herramientas de gran utilidad en el desarrollo de aplicaciones de Bases de Datos. Entre otras, existen:

- herramientas de ayuda al desarrollo (CASE<sup>5</sup>) en las fases de análisis, diseño e implementación de BD que generalmente incluyen diagramadores para esquemas conceptuales y lógicos de bases de datos, generadores de código SQL, etc.
- generadores de informes y pantallas que facilitan la presentación de los datos recuperados de la BD.
- generadores de aplicaciones basados en lenguajes de cuarta generación (4GL) que permiten a los usuarios desarrollar aplicaciones sin tener que programar en lenguajes convencionales
- facilidades de usuario para facilitar la consulta de los datos (menús, interfaces gráficas, etc.).

---

<sup>5</sup> Computer Aided Software Engineering

## 7. Algunas arquitecturas de Sistemas de Bases de Datos

En este apartado revisaremos brevemente algunos conceptos relacionados con las distintas arquitecturas de Sistemas de BD. En una arquitectura se reflejan aspectos como la conexión en red, el paralelismo y la distribución:

- Red: permite que algunas tareas se ejecuten en un sistema servidor y que otras se ejecuten en los clientes (son lo que se denominan sistemas de BD cliente-servidor).
- Paralelismo: acelerar la ejecución de tareas (transacciones, etc.) de acuerdo al sistema informático subyacente (sistemas de BD paralelos).
- Distribución: Datos situados donde se han generado o donde son más necesarios pero accesibles desde todos los sitios (sistemas de BD distribuidos).

Según estos aspectos distinguimos sistemas centralizados, sistemas cliente-servidor, sistemas paralelos y sistemas distribuidos.

En *los sistemas centralizados* existe un único sistema informático sin interacción con otros ordenadores. En estos sistemas podemos diferenciar entre:

- Sistema monousuario formado por un ordenador personal o por una estación de trabajo con una única CPU y un sistema operativo monousuario (no permite que varios usuarios puedan acceder simultáneamente a la BD)
- Sistema multiusuario formado por varias CPU y con sistema operativo multiusuario con terminales conectados al sistema servidor; estos terminales no poseen ninguna funcionalidad propia aparte de la de visualizar el resultado de los procesos que se ejecutan en el servidor.

En *los sistemas cliente-servidor* existe un reparto de funcionalidades, es decir, los terminales se sustituyen por ordenadores personales que gestionan el interfaz de usuario SQL, interfaz de formularios, diseñadores de informes e

interfaz gráfica. Los sistemas servidores satisfacen las peticiones generadas por los sistemas clientes.

Se distinguen dos tipos de servidores:

- Servidores de transacciones (servidores de consultas) con un interfaz mediante el que los clientes envían peticiones para realizar una acción que el servidor ejecutará y cuyos resultados se devuelven al cliente.
- Servidores de datos (el servidor envía los datos a las máquinas clientes en las que se realiza el procesamiento enviando después los datos de vuelta).

Respecto a los *sistemas paralelos* representan una solución al manejo de BD muy grandes o con un gran volumen de transacciones por segundo. El objetivo es realizar operaciones simultáneamente mediante el uso de varios procesadores y varios discos en paralelo. Los modelos de arquitecturas para máquinas paralelas son:

- Memoria compartida: Todos los procesadores comparten una memoria común.
- Disco compartido: Los procesadores comparten un disco común (cada procesador con su memoria)
- Sin compartimiento: No hay compartición ni de disco ni de memoria
- Jerárquico: modelo híbrido de los anteriores.

Por último, en los *sistemas distribuidos* la BD se almacena en varios ordenadores que no comparten ni memoria ni discos pero que se comunican mediante redes de alta velocidad o líneas telefónicas. Estos ordenadores se encuentran en varios lugares geográficos distintos. En un sistema distribuido se dan dos tipos de transacciones:

- Transacciones locales: Acceso a datos del ordenador en el que se inició la transacción
- Transacciones globales: Acceso a datos de un ordenador distinto o acceso a datos de varios ordenadores distintos.

Las ventajas que proporcionan los sistemas distribuidos frente a los sistemas centralizados son la compartición de datos (acceso a datos en distintos sitios), por ejemplo, dos sucursales bancarias pueden compartir datos entre sí; la autonomía en cuanto a que cada administrador controla su BD y, por último, la disponibilidad de los datos pues si un ordenador falla, están los demás para poder seguir trabajando, en particular, si hay duplicación de datos.



- **Una Base de Datos (BD)** es una colección o depósito de datos integrados, almacenados en soporte secundario (no volátil) y con redundancia controlada. Las ventajas que presentan son las siguientes:
  - Independencia de datos y procesos:
  - Descripción de los datos junto con los datos
  - Procesos de actualización y recuperación bien establecidos
- En una BD se identifican tres capas de estructuración según tres niveles de abstracción.
  - El nivel externo
  - El nivel lógico
  - El nivel físico
- **Un Sistema de Gestión de Bases de Datos (SGBD)** es un conjunto coordinado de programas, procedimientos, lenguajes, herramientas, etc., que suministra, tanto a los usuarios no informáticos como a los analistas, programadores o administradores de una BD, los medios necesarios para describir y manipular los datos integrados en la BD, manteniendo su integridad, confidencialidad y disponibilidad. Sus principales funciones son las de definición, manipulación y control.
- En una **arquitectura** de bases de datos se reflejan aspectos como la conexión en red, el paralelismo y la distribución. Según estos aspectos distinguimos sistemas centralizados, sistemas cliente-servidor, sistemas paralelos y sistemas distribuidos.

Sistemas Centralizados	Monousuario Multiusuario
Sistemas Cliente-Servidor	Servidores de transacciones Servidores de datos
Sistemas paralelos	Memoria Compartida - Disco Compartido
Sistemas distribuídos	Sin compartimiento - Jerárquico Transacciones Locales