

# Conceptos Fundamentales de las Bases de Datos

## Introducción

Una **base de datos** es un conjunto de datos relacionados entre si. Por datos entendemos hechos conocidos que pueden registrarse y que tienen un significado implícito. Por ejemplo, un número de teléfono, un número de documento, un nombre, etc.

Una base de datos tiene las siguientes propiedades implícitas:

- Una base de datos representa algún aspecto del mundo real, en ocasiones llamada minimundo. Las modificaciones del minimundo se reflejan en la base de datos.
- Una base de datos es un conjunto de datos lógicamente coherente, con cierto significado inherente. Una colección aleatoria de datos no puede considerarse propiamente una base de datos.
- Toda base de datos se diseña, construye y prueba con datos para un propósito específico. Está dirigida a un grupo de usuarios y tiene ciertas aplicaciones preconcebidas que interesan a dichos usuarios.

Tomemos el conjunto de toda la información que es relevante a la operación de una organización: para un banco, será información relativa a cuentas, clientes, empleados, equipos, etc. Toda esa información, apropiadamente organizada y codificada se colocará en la base de datos de la organización. Ningún programa de aplicación tendrá acceso directo a los archivos que componen la base de datos, sino que interpondremos entre estos archivos y los programas de aplicación un nuevo nivel de software, el **Sistema de Gestión de Base de Datos (SGBD)**. El SGBD provee acceso a la información a un alto nivel de abstracción: en lugar de manipular archivos, registros, índices, cilindros, el programa de aplicación se maneja enteramente en términos de clientes, cuentas, saldos, etc., que son traducidos por el SGBD a su implementación física.

Un **sistema de gestión de base de datos** es un conjunto de programas que permite a los usuarios crear y mantener una base de datos. Por lo tanto, un SGBD es un software de propósito general que facilita el proceso de definir, construir y manipular bases de datos para diversas aplicaciones.

- **Definir** la base de datos consiste en especificar los tipos de datos, las estructuras y las restricciones de los datos que se almacenarán en ella.
- El proceso de **construir** una base de datos consta de la carga de los datos mismos en algún medio de almacenamiento controlado por un SGBD.
- En la **manipulación** de una base de datos intervienen funciones como consultar la base de datos para obtener datos específicos, actualizar la base de datos para reflejar cambios en el minimundo y generar informes a partir de los datos.

## Tipos de Usuarios

Muchas personas participan en el diseño, uso y mantenimiento de una base de datos grande. Aquí las identificaremos.

- **Administradores:** Si consideramos a la base de datos y al SGBD como recursos del sistema de base de datos, debemos considerar a una persona que administre dichos recursos. El Administrador de la base de datos (DBA, en inglés) es quién se encarga de autorizar el acceso a la base de datos, de coordinar y vigilar su empleo, y de adquirir los recursos necesarios de software y hardware. El DBA es la persona responsable cuando surgen problemas como violaciones a la seguridad o una respuesta lenta del sistema.
- **Diseñadores de bases de datos:** se encargan de identificar los datos que se almacenarán en la base de datos y de elegir las estructuras apropiadas para representar y almacenar dichos datos.
- **Usuarios finales:** son las personas que necesitan tener acceso a la base de datos para consultarla, actualizarla y generar informes; la base de datos existe primordialmente para que ellos la usen.
- **Analistas y Programadores de Aplicaciones:** los analistas determinan los requerimientos de los usuarios finales y desarrollan especificaciones para transacciones programadas que satisfagan dichos requerimientos. Los programadores implementan estas especificaciones en forma de programas y luego prueban, depuran, documentan y mantienen esas transacciones programadas.

## Características del SGBD

Ahora analizaremos qué características son deseables en los SBGD y qué capacidades deben ofrecer.

### Control de redundancia

En la creación tradicional de programas de archivos, cada grupo de usuarios mantiene sus propios archivos para manejar sus aplicaciones de procesamiento de datos. Supongamos la situación para una base de datos de un supermercado. Los grupos involucrados serían, entre otros, el de finanzas y el de reposición. Con el enfoque tradicional, cada sector mantendría archivos independientes para los proveedores, cada uno con un objetivo distinto. Buena parte de los datos se almacenarían dos veces.

Generalmente, esta redundancia de datos provoca varios problemas:

- Es necesario realizar la misma actualización -introducir un nuevo proveedor- varias veces: una vez en cada archivo que contenga datos de proveedores. Esto implica duplicar el trabajo
- Se desperdicia espacio de almacenamiento al guardar los mismos datos en varios lugares
- Es posible que los archivos que representan los mismos datos se tornen inconsistentes

Con el enfoque de bases de datos, las vistas de los diferentes grupos de usuarios se integran durante el diseño de la base de datos. Para conservar la consistencia, debe crearse un diseño que almacene cada dato lógico -como la dirección del proveedor- en un solo lugar de la base de datos. Ello evita la inconsistencia y ahorra espacio de almacenamiento.

### Restricción de los accesos no autorizados

Cuando muchos usuarios comparten una misma base de datos, es probable que no todos tengan la autorización para tener acceso a toda la información que contiene. El SGBD debe contar con un subsistema de seguridad y autorización que permita al DBA crear cuentas y especificar restricciones para ellas.

### Cumplimiento de las restricciones de integridad

La mayor parte de las aplicaciones de base de datos tienen ciertas restricciones de integridad que deben cumplir los datos. El SGBD debe ofrecer recursos para definir tales restricciones y hacer que se cumplan. Por ejemplo, en la figura A, podemos especificar que el valor del elemento Ciclo dentro de cada registro de la tabla ESTUDIANTE debe ser un entero entre 1 y 5.

ESTUDIANTE			
Nombre	Matrícula	Ciclo	Carrera
Suárez	17	1	501
Álvarez	8	2	502

figura A: tabla ESTUDIANTE

Otro tipo de restricción especifica que los valores de los elementos de información sean únicos; por ejemplo, "cada registro de ESTUDIANTE debe tener un valor único de Matrícula". Estas restricciones se derivan de la **semántica** (o significado) de los datos y del minimundo que representa. Es responsabilidad de los diseñadores de la base de datos identificar las restricciones de integridad durante el diseño. Algunas restricciones se pueden especificar en el SGBD, el cual hará automáticamente que se cumplan; otras pueden requerir verificación mediante programas de actualización o en el momento en el que se introducen los datos.

### Respaldo y recuperación

Todo SGBD debe contar con recursos para recuperarse de fallos de hardware o de software. Para ello está el subsistema de **respaldo y recuperación** del SGBD. Por ejemplo, si el sistema falla mientras se está ejecutando un complejo programa de actualización, el subsistema de recuperación se encargará de asegurarse de que la base de datos se restaure al estado en el que estaba antes de que comenzara la ejecución del programa.

## Funciones de un SGBD

Las funciones provistas por un SGBD pueden agruparse en tres clases:

- Consulta y actualización de datos

- Mantenimiento de esquemas
- Manejo de transacciones

### Consulta y Actualización

Ésta es la clase más básica de funciones y la única que es visible "desde afuera". Consiste en un conjunto de herramientas que permite a los distintos tipos de usuarios del SGBD extraer, manipular y modifica la información almacenada en la base de datos.

Algunos sistemas proveen una única interfaz, llamada generalmente lenguaje de consulta, para ser usada por todos los tipos de usuario. Por ejemplo, el lenguaje **SQL** permite ejecutar consultas y actualizaciones en una notación vagamente similar al lenguaje natural. La expresión:

```
select matricula, nombre
from alumnos
where carrera = "Ingeniería" and edad < 21
```

produce como resultado una tabla con los números de matrícula y los nombres de todos los alumnos menores de edad de la carrera de Ingeniería.

### Mantenimiento de esquemas

El *esquema* de la base de datos es la descripción de la estructura de la información almacenada en ella. Por ejemplo, para un sistema basado en tablas, el esquema puede consistir en una lista de tablas en uso, los campos que contienen, el tipo de datos de cada campo, descripciones en lenguaje natural del propósito de cada tabla y cada campo, y restricciones sobre los valores admisibles en cada campo.

Así como los usuarios necesitan acceder, agregar y modificar datos, también necesitan acceder, agregar y modificar el esquema de datos. Por ejemplo, un usuario que se acerca por primera vez a una base de datos querrá saber antes que nada qué información contiene ésta, un programador puede escribir programas que definan y creen nuevos tipos de entidades, o eliminen algunos preexistentes; el DBA necesita controlar qué usuarios tienen accesos a qué información, formulando reglas de seguridad que se hacen parte del esquema.

### Manejo de Transacciones

Una de las áreas principales de aplicación de los sgbd's es lo que se llama procesamiento de transacciones. Una transacción es un programa de aplicación, generalmente de duración breve, que accede y actualiza una parte también generalmente pequeña de la base de datos. Típicos ejemplos son un depósito o extracción de una cuenta bancaria, o una reservación en un vuelo, o una verificación de una tarjeta de crédito.

El manejo de transacciones consiste en controlar múltiples transacciones ejecutando el *paralelo* sobre una misma base de datos corriendo en un sistema que puede *fallar*. Los objetivos del gestor de transacciones del sgbd son: evitar que las transacciones interfieran unas con otras al ejecutar en paralelo, y garantizar que la base de datos no sea dañada en forma irreparable por caídas, ya sea del sistema en sí o de alguna de las transacciones. El primero de los objetivos da lugar a lo que se llama *control de paralelismo*; el segundo, a *técnicas de recuperación*.

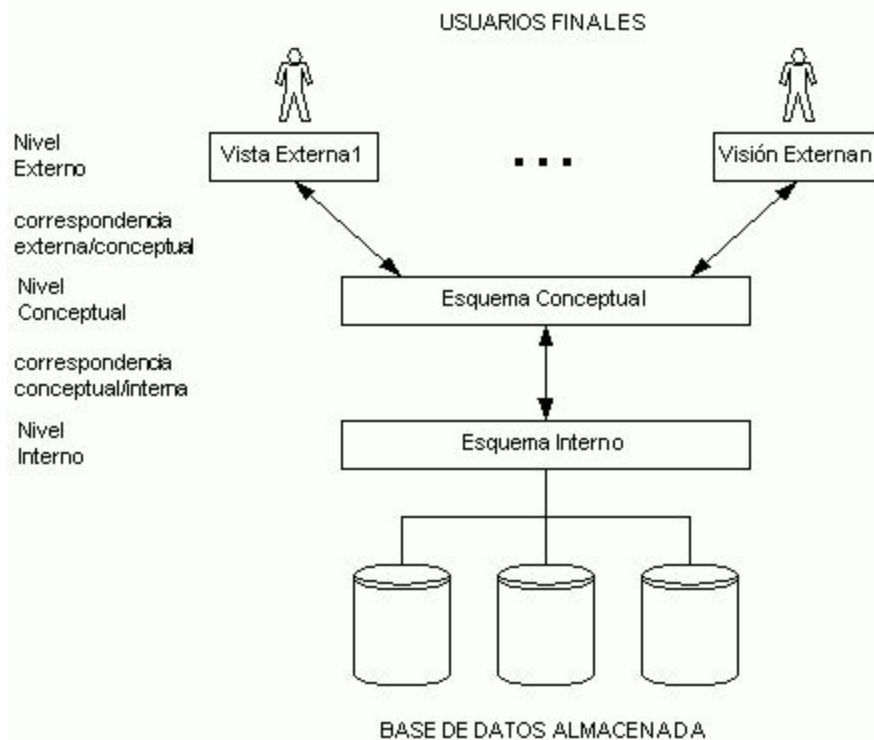
## Arquitectura del SGBD e independencia con respecto a los datos

Hay tres características importantes inherentes al enfoque de las bases de datos, que son la separación de los programas y los datos, manejo de múltiples vistas de usuario y el empleo de un catálogo para almacenar la descripción (esquema) de los datos. Aquí especificaremos una arquitectura para los sistemas de bases de datos, denominada **arquitectura de tres esquemas**, propuesta como ayuda para contar con estas características.

### **La arquitectura de tres capas**

El objetivo de la arquitectura de tres esquemas consiste en formar una separación entre las aplicaciones del usuario y la base de datos física. En esta arquitectura, los esquemas se pueden definir en los tres niveles siguientes:

- El **nivel interno** tiene un **esquema interno**, que describe la estructura física de almacenamiento de la base de datos. El esquema interno emplea un modelo físico de los datos y describe todos los detalles para su almacenamiento, así como los caminos de acceso para la base de datos
- El **nivel conceptual** tiene un **esquema conceptual**, que describe la estructura de toda la base de datos para una comunidad de usuarios. El esquema conceptual oculta los detalles de las estructuras físicas de almacenamiento y se concentra en describir entidades, tipos de datos, vínculos, operaciones de los usuarios y restricciones. En este nivel podemos usar un modelo de datos de alto nivel o uno de implementación
- El **nivel externo** o **de vistas** incluye varios **esquemas externos** o **vistas de usuario**. Cada esquema externo describe la parte de la base de datos que interesa a un grupo de usuarios determinado, y oculta a ese grupo el resto de la base de datos. En este nivel podemos usar un modelo de datos de alto nivel o uno de implementación



Los tres esquemas no son más que descripciones de los datos; los únicos datos que existen realmente están en el nivel físico. En un SGBD basado en la arquitectura de tres esquemas, cada grupo de usuarios hace referencia exclusivamente a su propio esquema externo; por tanto, el SGBD debe transformar una solicitud expresada en términos del esquema conceptual, y luego a una solicitud en el esquema interno que se procesará sobre la base de datos almacenada. Si la solicitud es una obtención de datos, será preciso modificar el formato de la información extraída de la base de datos almacenada para que coincida con la vista externa del usuario. El proceso de transformar solicitudes y resultados de un nivel a otro se denomina **correspondencia** o **transformación** (*mapping*).

### Independencia con respecto a los datos

La arquitectura de tres esquemas puede servir para explicar el concepto de **independencia con respecto a los datos**, que podemos definir como la capacidad para modificar el esquema de un nivel del sistema de base de datos sin tener que modificar el esquema del nivel inmediato superior. Podemos definir dos tipos de independencia con respecto a los datos:

1. La **independencia lógica con respecto a los datos** es la capacidad de modificar el esquema conceptual sin tener que alterar los esquemas externos ni los programas de aplicación. Podemos modificar el esquema conceptual para ampliar la base de datos (añadiendo un nuevo tipo de registro o un elemento de información), o para reducir la base de datos (eliminando un tipo de registro o un elemento de información). En el segundo caso, la modificación no deberá afectar los esquemas externos que solo se refieran a los datos restantes.

2. La **independencia física con respecto a los datos** es la capacidad de modificar el esquema interno sin tener que alterar el esquema conceptual (o los externos). Tal vez sea preciso modificar el esquema interno por la necesidad de reorganizar ciertos archivos físicos -por ejemplo, al crear estructuras de datos adicionales- a fin de mejorar el rendimiento de las operaciones de obtención o actualización. Si la base de datos aún contiene los mismos datos, no deberá ser necesario modificar el esquema conceptual.

## Lenguajes de Bases de Datos

El SGBD debe proveer a cada tipo de usuario de un lenguaje acorde a su categoría

### Lenguaje de definición de datos

Una vez que se ha completado el diseño de una base de datos y se ha elegido un SGBD para su implementación, el primer paso será especificar los esquemas conceptual e interno de la base de datos y cualquier correspondencia entre ambos. En muchos SGBD en los que no se mantiene una separación estricta entre niveles, el DBA y los diseñadores de la base de datos utilizan un mismo lenguaje, el **lenguaje de definición de datos** (DDL, en inglés), para definir ambos esquemas. El SGBD contará con un compilador de DDL cuya función será procesar enunciados escritos en el DDL para identificar las descripciones de los elementos de los esquemas y almacenar la descripción del esquema en el catálogo del SGBD.

### Lenguaje de manipulación de datos

Una vez que se han compilado los esquemas de la base de datos y que en ésta se han introducido datos, los usuarios requerirán algún mecanismo para manipularla. Las operaciones de manipulación más comunes son la obtención, la inserción, la eliminación y la modificación de datos. El SGBD ofrece un **lenguaje de manipulación de datos** (DML, en inglés) para estos fines