

Bases de Datos

Conceptos Básicos

Apunte elaborado por:

- MCs. Norma Herrera
- Lic. Darío Ruano
- Prof. Paola Azar



Universidad
Nacional de San Luis

1. Introducción

Todos a diario interactuamos con bases de datos: cuando hacemos una operación en un cajero automático, cuando compramos un pasaje en una plataforma en línea o cuando consultamos el pronóstico del tiempo, estamos interactuando con una base de datos aunque no seamos conscientes de ello. En la actualidad, las bases de datos se usan tan ampliamente que se pueden encontrar tanto en grandes empresas y organismos de gobierno como en pequeños negocios e incluso en hogares.

Las bases de datos se usan para satisfacer las necesidades de información de una empresa u organismo en particular. Desde las grandes aplicaciones multiusuario de base de datos hasta las agendas electrónicas utilizan tecnología de bases de datos para asegurar la integridad de los datos y facilitar la labor tanto de usuarios de la misma como de los programadores que las desarrollan. El usuario final de una base de datos puede no saber que está interactuando con ella, dado que existe un sistema que permite la interacción con la base de datos, haciendo transparente los detalles de implementación de la misma.

Existen distintos tipos de bases de datos: relacionales, orientadas a objetos, textuales, multimedia, etc. En este curso nos centraremos en [bases de datos relacionales](#), también conocidas como [tradicionales](#) o [transaccionales](#).

2. Conceptos Básicos

Un **dato** es la mínima unidad de significado. Un dato es un hecho, representado con un valor, que tiene un significado asociado. Por ejemplo, si decimos *edad 25 años* tenemos un dato donde el valor es 25 y el significado asociado es *edad*. Otros ejemplos de datos:

DNI 11.111.111

Departamento Ventas

Año de Ingreso 2010

Notar que un dato por sí solo es irrelevante, no dice nada sobre ningún elemento específico. El dato *edad 25 años*, dice que hay un valor 25 cuyo significado es la edad, pero no dice quién es el que tiene esa edad. Los datos no tienen ningún valor informativo. La **información** surge cuando relacionamos dos o mas datos. Por ejemplo, si decimos:

El empleado cuyo DNI es 11.111.111 trabaja en el departamento Ventas y tiene 25 años

tenemos información porque se han relacionado 3 datos (DNI, departamento y edad) para decir algo específico sobre un empleado de una empresa. La información nos dice algo que no sabemos y que es de utilidad para la toma de decisiones.

Una **base de datos** es una colección de datos relacionados, por lo tanto en una base de datos lo que tenemos es información. La información que contiene una base de datos debe ser coherente, una colección aleatoria de datos no puede considerarse una base de datos. Una base de datos se diseña y se puebla con datos para un propósito específico, dirigida a un grupo de usuarios y con aplicaciones que son de interés para esos usuarios.

Una base de datos puede tener cualquier tamaño y complejidad y puede o no estar informatizada. Por ejemplo, las agendas de papel son bases de datos donde la carga de datos y las búsquedas se hacen manualmente. Manejar una base de datos manualmente tiene sentido cuando la cantidad de información que se maneja es pequeña, pero si el volumen de datos aumenta se necesitará informatizar la base de datos para que su operación sea eficiente.

Cuando informatizamos una base de datos tenemos un **Sistema de Base de Datos (SBD)**. Un sistema de bases de datos es la base de datos en sí más los programas que manejan esos datos. Estos programas los llamaremos programas de aplicación y son los que se encargan de insertar, eliminar y buscar información en la base de datos.

Un SBD proporciona una visión abstracta de la base de datos escondiendo detalles de implementación que no son de importancia para los usuarios finales. Además, un SBD debe tener las siguientes propiedades:

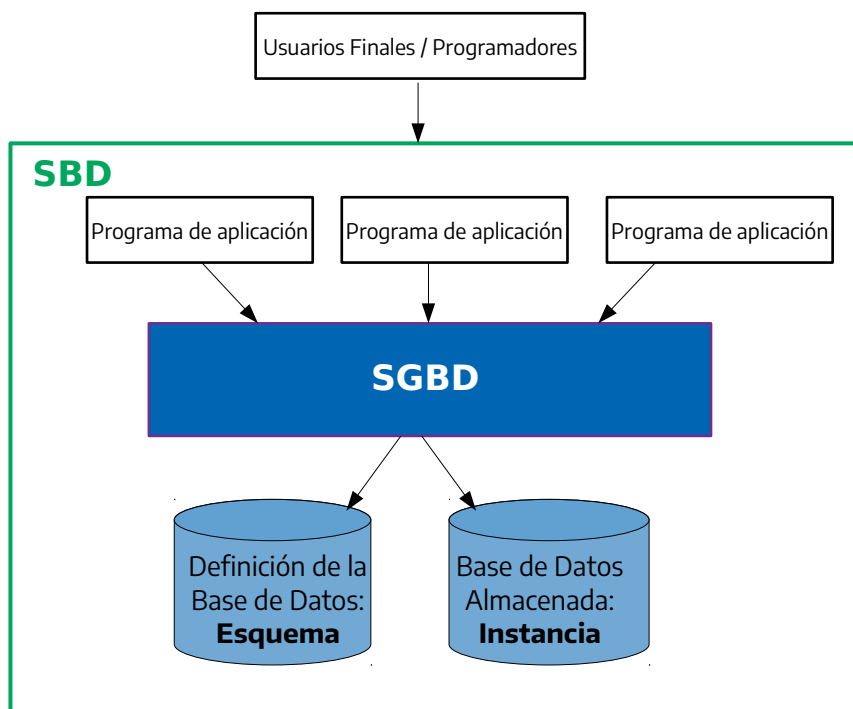
- **Independencia de datos:** los datos no dependen de los programas de aplicación que hacen uso de ellos.
- **Reducción de la Redundancia:** llamamos redundancia a la existencia de información duplicada. Reducir la redundancia implica realizar un mejor uso del espacio y evitar inconsistencias.
- **Seguridad:** un SBD debe permitir que tengamos un control sobre la seguridad de los datos.

Para informatizar una base de datos y crear el SBD contamos con los **Sistemas de Gestión de Base de Datos (SGBD o DBMS)** por sus siglas en inglés Database Management System). Un SGBD es un software de propósito general que permite crear y mantener una base de datos, facilitando los procesos de **definición, construcción y manipulación** de la base de datos para diferentes programas de aplicación.

Para **definir** una base de datos hay que especificar todos los datos que la conforman, los tipos de esos datos (numéricos, alfabéticos, etc), como se relacionan entre sí y las restricciones que existen entre ellos. Esto conforma lo que llamaremos **esquema de la base de datos**. Para **construir** la base de datos debemos cargar los datos en algún medio de almacenamiento permanente controlado por el SGBD. Los datos que conforman la base de datos en un instante dado de tiempo se denomina **instancia de la base de datos**. La **manipulación** de la base de datos se realiza a través de los programas de aplicación e implica operaciones tales como consultar datos, actualizar datos, ingresar nuevos datos, etc.

La distinción entre el esquema e instancia de la base de datos es muy importante. Cuando definimos una nueva base de datos, sólo especificamos su esquema, en ese momento la instancia de la base de datos es vacía. Cuando se cargan datos por primera vez, la base de datos pasa a la instancia inicial, pero su esquema no varía. De allí en adelante, siempre que se realice una operación de actualización de la base de datos, se tendrá una nueva instancia pero no se modifica el esquema.

Todos estos conceptos los podemos visualizar en la siguiente figura:



Entre los SGBD transaccionales mas usados podemos nombrar MySQL, PostgreSQL, Oracle, SQL Server, entre otros.

El término **motor de base de datos** generalmente se utiliza para referirse al software subyacente que un SGBD utiliza para definir, construir y manipular la bases de datos.

La mayoría de los SGBD proveen una herramienta gráfica para conectar con el motor y gestionar la base de datos. Por ejemplo, la herramienta gráfica de MySQL es MySQL Workbench, la de PostgreSQL es pgAdmin, la de Oracle es Oracle SQL Developer y la de SQL Server es SQL Server Management Studio. En este curso aprenderemos a usar **MySQL** a través de su **herramienta gráfica MySQL Workbench**.

También podríamos crear un SBD sin la asistencia de un SGB. En este caso deberíamos programar nuestro propio software encargado de realizar todas las operaciones hechas por el SGBD. Cualquiera sea el caso, siempre se necesita un software de gran capacidad para definir, construir y manipular una base de datos.

3. Análisis y Diseño de una Base de Datos: Modelo de Datos

Desde que se toma la iniciativa de informatizar una base de datos hasta que el SBD está operativo hay un proceso de desarrollo cuya complejidad depende de la complejidad de la base de datos que se quiera implementar. Básicamente podemos distinguir dos etapas:

1. Diseño
2. Implementación de procesos y carga de datos.

El proceso de diseño de una base de datos comienza con una descripción en lenguaje natural de la porción del mundo real que se quiere representar en la base de datos. Esta porción del mundo real se denomina **universo del discurso (UD)**. Por ejemplo, el siguiente relato describe un UD:

Un complejo de cabañas necesita armar una base de datos para administrar las reservas de las mismas. El complejo cuenta con un gran número de cabañas, cada una de las cuales tiene asignado un código que es único.

De cada cabaña se registra, además de su código, su nombre, la capacidad y la cantidad de habitaciones de la misma. Se sabe que no hay cabañas de más de 4 habitaciones y que la cantidad máxima de personas permitidas por cabaña es 10.

De los clientes se desea registrar su DNI, nombre, dirección y teléfono de contacto.

Cada vez que un cliente realiza una reserva se le asigna a la misma un código. De cada reserva realizada se registra además de su código, la cabaña reservada, el cliente que hizo la reserva, la fecha de reserva y la cantidad de días de la estadía. Se permite que un cliente haga más de una reserva, para la misma o para distintas fechas.

A partir de este relato podemos comenzar a identificar los datos que existirán en nuestra base de datos y las relaciones y/o restricciones que hay entre ellos. Por ejemplo: *código, nombre, capacidad y cantidad de habitaciones* son datos que se quiere registrar de las cabañas con la restricción de que no pueden existir dos cabañas con el mismo código (“...cada una de las cuales tiene asignado un código que es único...”). El problema es que esta descripción en lenguaje natural es ambiguo. Por ejemplo, en ningún lado del relato dice si los nombres de cabaña son únicos o si se pueden repetir. Es algo que solamente el dueño o administrador del complejo de cabañas nos debe decir. **Nunca debemos suponer nada sobre el UD que estamos trabajando.** Es una regla que siempre debe respetarse para evitar futuros inconvenientes en el SBD, cuando este ya está operativo.

Para evitar ambigüedades y posibles inconsistencias entre la base de datos y el UD que esa base de datos representa, se necesita un proceso de modelado que describa de manera formal y no ambigua el UD. Un **modelo** es un conjunto de herramientas conceptuales que nos permiten describir de manera formal y precisa el UD. El modelo nos permite describir el esquema de la base de datos sin que existan ambigüedades respecto de los datos y las relaciones y restricciones que hay entre ellos.

Habíamos visto que una de las características principales de los SBD es proporcionar cierto nivel de abstracción de datos, ocultando al usuario final detalles sobre el almacenamiento físico de los datos. Los modelos de datos son el instrumento principal para ofrecer esta abstracción.

Dado que una de las formas de implementar un SBD es usando un SGBD, los SGBD están basados en un modelo de datos. En la actualidad el modelo más ampliamente usado es el modelo relacional.

En esta materia aprenderemos a obtener el modelo relacional y a implementar el SBD usando un SGBD.

■

Bibliografía:

Fundamentos de Sistemas de Bases de Datos, 3a edición. R. Elmasri , S. Navathe. Capítulos 1 y 2 (páginas 3 a 31).

Fundamentos de Bases de Datos, 4ta edición. A. Silberschatz H. F. Korth, S. Sudarshan. Capítulo 1.