

1.1. Introducción

Las bases de datos y su tecnología están teniendo un mayor impacto con el creciente uso de los computadores. En realidad las bases de datos desempeñan un papel fundamental en casi todas las áreas donde se utilizan computadores, incluyendo negocios, ingeniería, medicina, leyes, educación y biblioteconomía, por nombrar algunas. La palabra *base de datos* tiene un uso tan habitual que debemos empezar definiéndola. Nuestra primera definición es bastante general.

Una **base de datos** es una colección de datos relacionados. Por **datos**, queremos decir hechos conocidos que pueden registrarse y que tienen un significado implícito. Por ejemplo, considérese los nombres, números de teléfono y direcciones de gente que usted conoce. Usted podría registrar dichos datos en un libro de direcciones indexado, o podría almacenarlos en un disquete, utilizando un computador personal y un software como DBASE IV o V, Microsoft ACCESS o EXCEL. Éste es una colección de datos relacionados con un significado implícito y por tanto es una base de datos.

La definición precedente de *base de datos* es bastante general; por ejemplo podemos considerar que la colección de palabras que componen esta página de texto tienen datos relacionados y por tanto constituyen una base de datos. Sin embargo, el uso habitual del término *base de datos* suele ser más restrictivo. Una base de datos tiene las siguientes propiedades implícitas:

- Una base de datos representa algunos aspectos del mundo real, en ocasiones denominado **minimundo** o **Universo del Discurso (UdD)**. Los cambios en el minimundo se reflejan en la base de datos.
- Una base de datos es una colección coherente de datos con significados inherentes. Un conjunto aleatorio de datos no puede considerarse como una base de datos.
- Una base de datos se diseña, construye y puebla con datos para un propósito específico. Está destinada a un grupo de usuarios concreto y tiene algunas aplicaciones preconcebidas en las cuales están interesados dichos usuarios.

En otras palabras, una base de datos tiene alguna fuente de la cual provienen los datos, algún grado de interacción con los sucesos del mundo real, y una audiencia que está activamente interesada en el contenido de la base de datos.

Una base de datos puede tener cualquier tamaño y complejidad. Por ejemplo, la lista de nombres y direcciones mencionada anteriormente, puede constar de unos pocos cientos de registros, cada uno con una estructura sencilla. Por otro lado, el catálogo de fichas de una gran biblioteca puede contener medio millón de fichas almacenadas bajo diferentes categorías (por apellido del autor, por materia, por título del libro) con cada categoría ordenada alfabéticamente. Las autoridades fiscales mantienen incluso una base de datos de mayor tamaño y complejidad para el control de las declaraciones de la renta presentadas por los contribuyentes. Si suponemos que en Estados Unidos hay 100 millones de contribuyentes y que cada contribuyente rellena una media de cinco formularios por declaración, con aproximadamente 200 caracteres de información por formulario,

obtendremos una base de datos de $100 * (10^6) * 200 * 5$ caracteres (bytes) de información. Si suponemos que las autoridades fiscales mantienen las tres últimas declaraciones de la renta de cada contribuyente además de la actual, podríamos encontrarnos con una base de datos de $4 * (10^{11})$ bytes (400 gigabytes). Esta gran cantidad de información debe organizarse y gestionarse de forma que los usuarios puedan localizar, recuperar y actualizar los datos según lo necesiten.

Una base de datos puede crearse y mantenerse manualmente o puede estar informatizada. El catálogo de fichas de una biblioteca es un ejemplo de una base de datos que puede crearse y mantenerse manualmente. Una base de datos informatizada puede crearse y mantenerse bien mediante un conjunto de programas de aplicación diseñados específicamente para dicha tarea o bien mediante un sistema de gestión de base de datos.

Un **sistema de gestión de base de datos (SGBD)**, en inglés *database management system* o *DBMS*) es una colección de programas que permiten a los usuarios crear y mantener una base de datos. El SGBD es por tanto un *sistema software de propósito general* que facilita los procesos de *definición, construcción y manipulación* de bases de datos para distintas aplicaciones. La **definición** de una base de datos consiste en especificar los tipos de datos, las estructuras y restricciones para los datos que se van a almacenar en dicha base. La **construcción** de la base de datos es el proceso de almacenar los datos concretos sobre algún medio de almacenamiento controlado por el SGBD. La **manipulación** de la base de datos incluye funciones tales como consultar la base de datos para recuperar unos datos específicos, actualizar la base de datos para reflejar los cambios ocurridos en el minimundo, y generar informes a partir de los datos.

No es preciso utilizar software de SGBD de propósito general para implementar una base de datos informatizada. Podríamos codificar nuestro propio conjunto de programas para crear y mantener la base de datos, es decir, crear nuestro propio software de SGBD de *propósito específico*. En cualquier caso, usemos o no un SGBD de propósito general, normalmente tendremos que emplear gran cantidad de software para manipular la base de datos. Denominaremos **sistema de base de datos** al conjunto formado por la base de datos más el SGBD. La Figura 1.1 ilustra estos conceptos.