

Definición

Una base de datos consiste en un conjunto de datos relacionados entre sí. Por datos entendemos hechos conocidos que pueden registrarse y que tienen un significado implícito. Pero la acepción común del término *base de datos* suele ser más restringida. Una base de datos tiene las siguientes propiedades implícitas:

- Una base de datos representa algún aspecto del mundo real, en ocasiones llamadas **minimundo** o **universo de discurso**.
- 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 puebla con datos para un propósito específico. Está dirigida a un grupo de usuarios y tiene ciertas aplicaciones preconcebidas que interesan a dichos usuarios.

Objetivos de los sistemas de bases de datos

Entre los objetivos más importantes de los sistemas de bases de datos están:

- Disminuir la redundancia e inconsistencia de los datos: Puesto que los archivos y los programas de aplicaciones fueron creados por distintos programadores en un periodo largo, es posible que un mismo dato esté repetido en varios sitios (archivos). Esta redundancia aumenta los costos de almacenamiento y acceso, además de incrementar la posibilidad de que exista inconsistencia en la información.
- Reducir la dificultad para tener acceso a los datos: Supóngase que uno de los gerentes del banco necesita averiguar los nombres de todos los clientes que viven en cierta parte de la ciudad. El gerente llama al departamento de procesamiento de datos y pide que generen la lista correspondiente. Como ésta es una solicitud fuera de lo común no existe un programa de aplicaciones para generar semejante lista. Lo que se trata de probar aquí es que este ambiente no permite recuperar la información requerida en forma conveniente o eficiente.
- Evitar el aislamiento de los datos: Puesto que los datos están repartidos en varios archivos, y éstos pueden tener diferentes formatos, es difícil escribir nuevos programas de aplicaciones para obtener los datos apropiados.
- Corregir anomalías en el acceso concurrente: Para mejorar el funcionamiento del sistema y tener un tiempo de respuesta más corto, muchos sistemas permiten que varios usuarios actualicen la información simultáneamente. En un ambiente de este tipo, la interacción de las actualizaciones concurrentes puede resultar en información inconsistente. Para prevenir estas situaciones debe mantenerse alguna forma de supervisión en el sistema.
- Disminuir los problemas de seguridad: No es recomendable que todos los usuarios del sistema de base de datos pueda tener acceso a toda la información. Por ejemplo, en un sistema bancario, una persona que prepare los cheques de nómina sólo debe poder ver la parte de la base de datos que contenga información de los empleados. No puede consultar información correspondiente a las cuentas de los clientes.
- Disminuir los problemas de integridad: Los valores que se guardan en la base de datos debe satisfacer ciertos tipos de *limitantes de consistencia*. El sistema debe obligar al cumplimiento de estas limitantes. Esto puede hacerse agregando el código apropiado a los distintos programas de aplicaciones. El problema se complica cuando las limitantes implican varios elementos de información de distintos archivos.

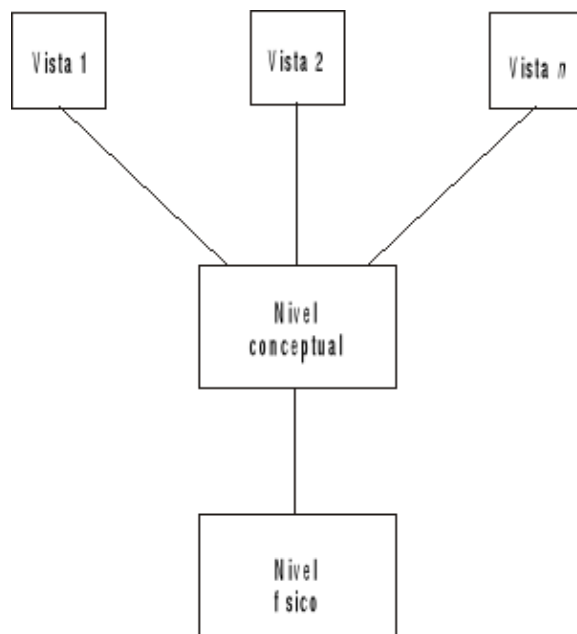
Abstracción de la información

Uno de los objetivos principales de un sistema de base de datos es proporcionar a los usuarios una visión *abstracta* de la información. Es decir, el sistema oculta ciertos detalles relativos a la forma como los datos se almacenan y mantienen.

Los sistemas de base de datos muchas veces son utilizados por personal que no cuenta con conocimientos de computación, esta complejidad debe estar escondida para los usuarios. Para ocultarla, se definen varios niveles de abstracción en los que puede observarse la base de datos.

- **Nivel Físico:** este es el nivel más bajo de abstracción, en el que se describe *cómo* se almacenan realmente los datos.
- **Nivel conceptual:** Este es el siguiente nivel más alto de abstracción, en el que se describe *cuáles* son los datos reales que están almacenados en la base de datos y qué relaciones existen entre los datos.
- **Nivel de visión:** Este es el nivel de abstracción más alto, en el cual se describe solamente una parte de la base de datos. Aunque en el nivel conceptual se utilizan estructuras más simples, todavía queda una forma de complejidad que resulta del gran tamaño de la base de datos.

La interrelación entre estos tres niveles de abstracción se muestra en la siguiente figura:



Modelos de datos

Para describir la estructura de una base de datos es necesario definir el concepto de *modelo de datos*. Éste es un grupo de herramientas conceptuales para describir los datos, sus relaciones, su semántica y sus limitantes. Se han propuesto varios modelos de datos diferentes, los cuales pueden dividirse en tres grupos: los modelos lógicos basados en objetos y en registros, y los físicos de datos.

Modelos lógicos basados en objetos. Se utilizan para describir los datos en los niveles conceptual y de visión. Se caracterizan por el hecho de que permiten una estructuración bastante flexible y hacen posible especificar claramente las limitantes de los datos. Algunos de los más conocidos son:

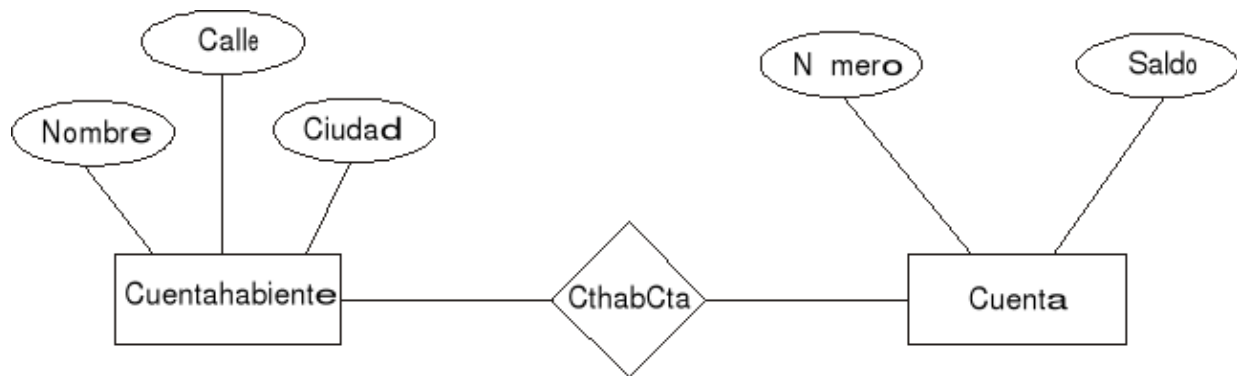
- El modelo entidad – relación
- El modelo binario
- El modelo semántico de datos
- El modelo infológico

La estructura lógica general de una base de datos puede expresarse gráficamente por medio de un diagrama entidad – relación que consta de los siguientes componentes:

- **Rectángulos**, que representan conjuntos de entidades.

- **Elipses**, que representan atributos.
- **Rombos**, que representan relaciones entre conjuntos de entidades.
- **Líneas**, que conectan los atributos a los conjuntos de entidades y los conjuntos de entidades a las relaciones.

Ejemplo:



Modelos lógicos basados en registros. Se utilizan para describir los datos en los niveles conceptual y de visión. A diferencia de los modelos de datos basados en objetos, estos modelos sirven para especificar tanto la estructura lógica general de la base de datos como una descripción en un nivel más alto de la implantación.

- **Modelo relacional:** Los datos y las relaciones entre los datos se representan por medio de una serie de tablas, cada una de las cuales tiene varias columnas con nombres únicos. Ejemplo:

Nombre

Calle

Ciudad

Número

Lowery

Maple

Queens

900

Shiver

North

Bronx

556

Shiver

North

Bronx

647

Hodges

Sidehill

Brooklyn

801

Hodges

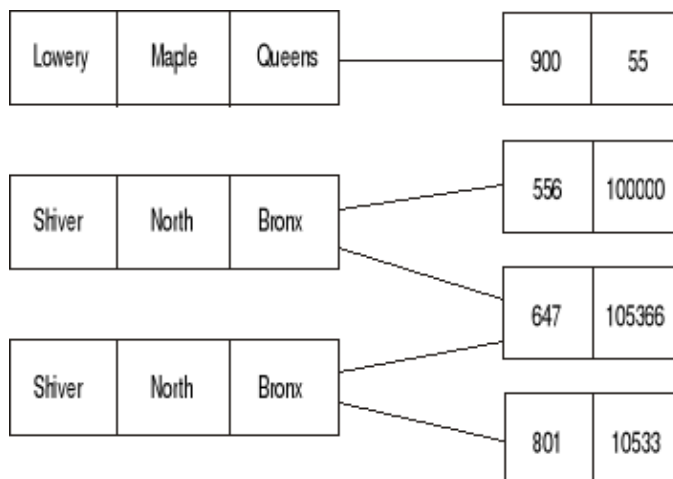
Sidehill

Brooklyn

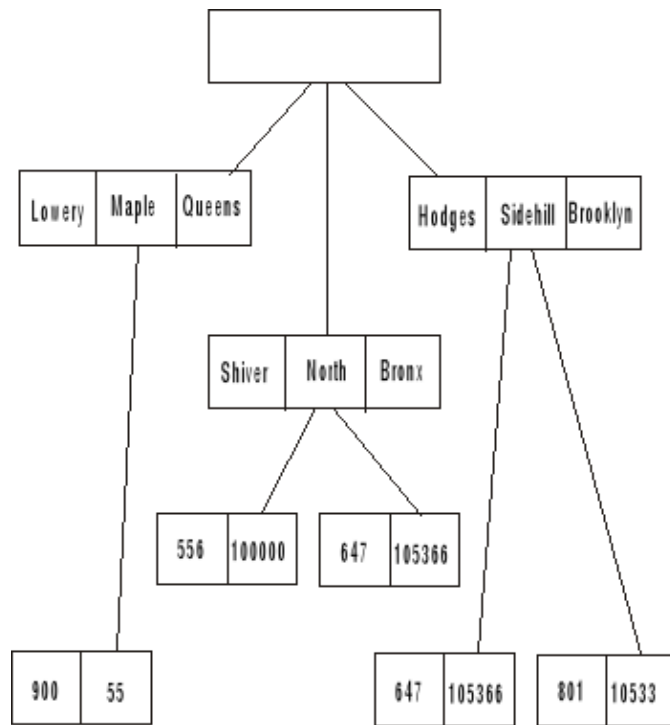
647

Número	Saldo
900	55
556	100000
647	105366
801	10533

- **Modelo de red:** Los datos se representan por medio de conjuntos de *registros* y las relaciones entre los datos se representan con *ligas*, que pueden considerarse como apuntadores. Ejemplo:



- **Modelo Jerárquico:** Difiere del de red en que los registros están organizados como conjuntos de árboles en vez de gráficas arbitrarias. Ejemplo.



Modelos físicos de los datos: Los modelos físicos sirven para describir los datos en el nivel más bajo. A diferencia de los modelos lógicos de los datos, son muy pocos los modelos físicos utilizados. Algunos de los más conocidos son:

- El modelo unificador
- La memoria de cuadros

Instancias y esquemas

Las bases de datos cambian con el tiempo al insertarse información en la base de datos y eliminarse de ella. El conjunto de información almacenado en la base de datos en cierto momento se denomina *instancia* en la base de datos. El diseño general de la base de datos se llama *esquema de la base de datos*. Los esquemas se alteran muy raras veces, o nunca.

Una variable de un determinado tipo tiene un valor específico en un momento dado. Así, el concepto del valor de una variable en los lenguajes de programación corresponde al concepto de una *instancia* del esquema de una base de datos.

Existen varios esquemas en la base de datos, y éstos se dividen de acuerdo con los niveles de abstracción mencionados anteriormente.

Independencia de los datos

La capacidad de modificar una definición de esquema en un nivel sin afectar la definición del esquema en el nivel inmediato superior se denomina *independencia de los datos*. Existen dos niveles de tal independencia:

- Independencia física: Es la capacidad de modificar el esquema físico sin obligar a que se vuelvan a escribir los programas de aplicaciones. En algunas ocasiones es necesario para mejorar el rendimiento.
- Independencia lógica: Es la capacidad de modificar el esquema conceptual sin obligar a que se vuelvan a escribir los programas de aplicaciones. Las modificaciones en el nivel conceptual son

necesarias siempre que se altera la estructura lógica de la base de datos.

Lenguaje de definición de datos

Un esquema de base de datos se especifica por medio de una serie de definiciones que se expresan en un lenguaje especial llamado *lenguaje de definición de datos* (en inglés: DDL, *data definition language*). El resultado de la compilación de las proposiciones en DDL es un conjunto de tablas que se almacena en un archivo especial llamado *diccionario* (o *directorio*) de datos.

Un directorio de datos es un archivo que contiene *metadatos*, es decir, datos acerca de los datos. Este archivo se consulta antes de leer o modificar los datos reales en el sistema de base de datos.

La estructura de almacenamiento y los métodos de acceso empleados por el sistema de base de datos se especifican por medio de un conjunto de definiciones de un tipo especial de DDL llamado *lenguaje de almacenamiento y definición de los datos*. El resultado de la compilación de estas definiciones es una serie de instrucciones que especifican los detalles de implantación de los esquemas de base de datos que normalmente no pueden ver los usuarios.

Lenguaje de Manipulación de datos

Los niveles de abstracción que se mencionaron anteriormente no sólo se aplican a la definición o estructuración de los datos, sino también al manejo de los datos; esta manipulación consiste en:

- La recuperación de información almacenada en la base de datos.
- La inserción de información nueva en la base de datos.
- La eliminación de información de la base de datos.

Un *lenguaje de manipulación de datos* (en inglés: DML, *data manipulation language*) permite a los usuarios manejar o tener acceso a los datos que están organizados por medio del modelo apropiado. Existen básicamente dos tipos de DML:

- De procedimientos: Necesitan que el usuario especifique *cuáles* datos quiere y *cómo* deben obtenerse.
- Sin procedimientos: Requieren que el usuario especifique *cuáles* datos quiere *sin* especificar cómo obtenerlos.

Manejador de base de datos

Generalmente las bases de datos requieren una gran cantidad de espacio de almacenamiento. Las bases de datos de las empresas comúnmente se miden en términos de *gigabytes* de información. Puesto que la memoria principal de la computadora no puede almacenar esta información, se guarda en discos. Ya que el movimiento de los datos en el disco es muy lento comparado con la velocidad de la unidad central de procesamiento de las computadoras, es imperativo que el sistema de base de datos estructure la información de tal manera que se reduzca la necesidad de transferir datos entre el disco y la memoria principal.

El objetivo de un sistema de base de datos es simplificar y facilitar el acceso a los datos. Las vistas de alto nivel ayudan a lograrlo. Si el tiempo de respuesta para una consulta es demasiado largo, el valor del sistema se reduce.

Un manejador de base de datos es un módulo de programa que constituye la interfaz entre los datos de bajo nivel almacenados en la base de datos y los programas de aplicaciones y las consultas hechas al sistema. El manejador de base de datos es responsable de las siguientes tareas:

- Interacción con el manejador de archivos: Los datos sin procesar se almacenan en el disco mediante el sistema de archivos. El manejador de base de datos traduce las diferentes proposiciones en DML a comandos de sistema de archivos de bajo nivel.
- Implantación de la integridad: Los valores de los datos almacenados en la base de datos deben satisfacer ciertos tipos de *limitantes de consistencia*. El administrador de la base de datos debe especificar estas limitantes de forma explícita. Si se especifican estas limitantes entonces el manejador de la base de datos puede verificar si las actualizaciones a la base de datos resultan en la violación de cualquiera de estas limitantes, y si así es, podrá realizar la acción apropiada.
- Puesto en práctica de la seguridad: Como se mencionó anteriormente, no es preciso que todos los usuarios de la base de datos tengan acceso a todo su contenido. Es labor del manejador de base de datos hacer que se cumplan estos requisitos de seguridad.
- Respaldo y recuperación: Es responsabilidad del manejador de la base de datos detectar fallas (como caídas de las cabezas lectoras de disco, interrupción del suministro de energía y errores de software) y restaurar la base de datos al estado que existía antes de presentarse la falla.
- Control de concurrencia: Es necesario que el sistema controle la interacción entre los usuarios concurrentes; lograr dicho control es una de las tareas del manejador de la base de datos.

Administrador de base de datos

Una de las razones principales para contar con sistema de manejo de base de datos es tener un control centralizado tanto de los datos como de los programas que tienen acceso a ellos. La persona que tiene este control centralizado sobre el sistema es el *administrador de base de datos* (en inglés: *DBA database administrator*). Las funciones del administrador de la base de datos son, entre otras:

- Definición de esquema: Es decir la creación del esquema original de la base de datos. Esto se logra escribiendo una serie de definiciones que el compilador de DDL traduce a un conjunto de tablas que se almacenan permanentemente en el *diccionario de datos*.
- Definición de la estructura de almacenamiento y del método de acceso: La creación de las estructuras de almacenamiento y métodos de acceso apropiados. Esto se lleva a cabo escribiendo una serie de definiciones que posteriormente son traducidas por el compilador del lenguaje de almacenamiento y definición de datos.
- Modificación del esquema y de la organización física: Estos cambios, aunque son relativamente poco frecuentes, se logran escribiendo una serie de definiciones utilizadas, ya sea por el compilador de DDL o por el compilador del lenguaje de almacenamiento y definición de datos para generar modificaciones a las tablas internas apropiadas del sistema.
- Concesión de autorización para acceso a los datos: Es decir, conceder diferentes tipos de autorización para acceso a los datos a los distintos usuarios de la base de datos. Esto permite al administrador de base de datos regular cuáles son las partes de la base de datos a la que van a tener acceso diversos usuarios.
- Especificación de las limitantes de integridad: Estas limitantes se conservan en una estructura especial del sistema que consulta el manejador de base de datos cada vez que se lleva a cabo una actualización en el sistema.

Usuarios de la base de datos