

BASES DE DATOS

Una base de datos es una colección de información ordenada e interrelacionada que es de importancia para una empresa.

La creación de una base de datos debe ser realizada cuidadosamente procurando cumplir con una serie de objetivos (entre otros) que se detallan a continuación :

??Permitir un fácil acceso a la información: El sistema debe ser facilitador de alto rendimiento, la velocidad es un factor esencial así como la consistencia de los datos

??Evitar redundancia de la información: La información almacenada ocupará irremediamente un espacio en memoria por lo cual es de vital importancia eliminar la posibilidad de almacenar datos repetidos ya que adicionalmente podrían llevarnos a inconsistencias en la información .

Un sistema de gestión de bases de datos (SGBD) consiste en una colección de datos interrelacionados y un conjunto de programas para acceder a dichos datos. La colección de datos, normalmente denominada base de datos, contiene información acerca de una empresa particular. El primer objetivo de un SGBD es proporcionar un entorno que sea tanto práctico como eficiente de usar en la recuperación y el almacenamiento de información de la base de datos.

Los SBD se diseñan para gestionar grandes cantidades de información. La gestión de los datos implica tanto la definición de estructuras para almacenar la información como la provisión de mecanismos para la manipulación de la misma. Los SBD deben proporcionar la fiabilidad de la información almacenada, a pesar de las caídas del sistema o del intento de acceso sin autorización.

Anteriormente se utilizaba el sistema de procesamiento de archivos en el cual toda la información de la empresa se almacenaba en archivos permanentes manejados por un sistema operativo convencional, se tenían que escribir diversos programas de aplicación para extraer registros y añadir registros a los archivos adecuados.

Los principales inconvenientes de la organización en un sistema de procesamiento de archivos son:

?? Redundancia e inconsistencia de datos. Como los programas y archivos son creados por diferentes personas a través del tiempo, estos pueden tener diferentes formatos y escritos en diferentes lenguajes de programación, la información puede estar duplicada en los diversos lugares. Esto conduce a un almacenamiento y coste de acceso más altos y puede conducir a inconsistencia de datos.

- ?? Dificultad en el acceso de datos. Algunas peticiones no previstas en el sistema original, no hay un programa de aplicación a la mano para hacerlo. Sin embargo hay programas para mostrar la lista de total y las opciones aquí son; primero, de esa lista sacar la información manualmente y segundo, mandar a un programador a hacer un programa que haga esa consulta. Y por cada consulta se debería hacer lo mismo.
- ?? Aislamiento de datos. La creación de nuevos programas de aplicación es muy difícil ya que los archivos pueden estar en diversos formatos y se tiene que recopilar la información dispersa en cada uno de los archivos.
- ?? Problemas de integridad. Los valores de los datos almacenados en la base de datos deben satisfacer ciertos tipos de ligaduras de consistencia. Los desarrolladores hacen cumplir esas ligaduras en el sistema añadiendo el código apropiado en los diversos programas de aplicación.
- ?? Problemas de atomicidad. Es necesario asegurar que en caso de fallo se regrese a un estado en el cual había un estado de consistencia. Y en estos casos pueden haber inconsistencias si se trabaja con archivos, y lo que debe de ocurrir es que si hay cambios se hagan completamente o no se haga nada en absoluto. Así se pueden evitar estas inconsistencias que no son seguras en uno que maneje archivos.
- ?? Problemas de seguridad. No todos los usuarios de un sistema de bases de datos deberían poder acceder a todos los datos. Como los programas de aplicación se añaden de una forma ad hoc, es difícil garantizar esta seguridad.

Visión de los datos

Un SGBD es una colección de archivos interrelacionados y un conjunto de programas que permiten acceder y modificar estos archivos. El propósito principal de un sistema de bases de datos es proporcionar a los usuarios una visión abstracta de los datos. Es decir, el sistema esconde ciertos detalles de cómo se almacenan y mantienen los datos.

Abstracción de datos

Para que el sistema sea útil, debe recuperar los datos eficientemente. Esto ha conducido al diseño de estructuras complejas para la representación de los datos en la base de datos. Los desarrolladores esconden la complejidad a través de los niveles de abstracción para simplificar la interacción de los usuarios con el sistema:

- ?? **Nivel físico.** El nivel más bajo de abstracción, describe como se almacenan realmente los datos. En el nivel físico se describen en detalle las estructuras de datos complejas de bajo nivel.
- ?? **Nivel lógico.** El siguiente nivel de abstracción describe que datos se almacenan en la base de datos y que relaciones existen entre esos datos.
- ?? **Nivel de vistas.** Este nivel solo describe parte de la base de datos completa. A muchos usuarios no les importará el nivel lógico, solo querrán ver parte de la información. Para la simplificación de la interacción del usuario se crea un nivel de vistas.

Las bases de datos van cambiando a lo largo del tiempo conforme se borran e insertan datos. La colección de información almacenada en la base de datos en un momento particular se llama ejemplar (instancia) de la base de datos. El diseño completo se llama esquema de la base de datos y estos rara vez son modificados (no es recomendable).

Independencia de datos

La capacidad para modificar una definición de esquema en un nivel sin que afecte a una definición de esquema en el siguiente nivel más alto se llama independencia de datos. Hay dos niveles de independencia.

- ?? **Independencia física de datos.** Es la capacidad para modificar el esquema físico sin provocar que los programas de aplicación tengan que reescribirse.
- ?? **Independencia lógica de datos.** Es la capacidad para modificar el esquema lógico sin causar que los programas de aplicación tengan que reescribirse. Las modificaciones en el nivel lógico son necesarias siempre que la estructura lógica de la base de datos se altere.

Modelos de datos

La parte esencial de la estructura de base de datos es el modelo de datos: una colección de herramientas conceptuales para describir los datos, las relaciones de datos, la semántica de los datos y las ligaduras de consistencia. Los diferentes modelos de datos que se han propuesto se clasifican en tres grupos diferentes: modelos lógicos basados en objetos, modelos lógicos basados en registros y modelos físicos.

Los modelos lógicos basados en objetos se usan para describir datos en los niveles lógico y de vistas. Proporcionan capacidades estructurales muy flexibles y permiten que las ligaduras se especifiquen explícitamente. Los más conocidos son:

- ?? Modelo entidad – relación.
- ?? Modelo orientado a objetos.
- ?? Modelo de datos semánticos.
- ?? Modelo de datos funcional.

Modelo entidad – relación

Esta basado en una percepción del mundo real que consta de una colección de objetos básicos, llamados entidades, y de las relaciones entre estos objetos. Una entidad es una cosa u objeto que es distinguible de otros objetos. Una relación es una asociación entre varias entidades. Se maneja la correspondencia de cardinalidades que expresa el número de entidades que pueden estar relacionadas con una entidad por medio de relaciones.

Modelo orientado a objetos

Está basado en una colección de objetos. Un objeto contiene valores almacenados en variables ejemplares dentro de este objeto. Contiene fragmentos de código que operan dentro del mismo y a éstos se les llama métodos. La única manera en que pueden acceder a la base de datos es a través del paso de mensajes a otro objeto.

Modelos lógicos basados en registros

Se usan para especificar la estructura lógica completa de la base de datos para proporcionar una descripción de alto nivel de la implementación. Este se estructura en registros de formato fijo de diferentes tipos. En cada tipo de registro se define un número de campos y cada uno tiene una longitud fija. Los principales modelos trabajados a través del tiempo son el modelo relacional, el de red y el jerárquico (el primero es el utilizado actualmente).

Modelo relacional

Se usa una colección de tablas para representar tanto los datos como las relaciones entre ellos. Cada tabla contiene varias columnas, y cada columna tienen un nombre único.

Modelo de red

Se representan por medio de colecciones de registros y las relaciones entre los datos se representan por medio de enlaces que se pueden ver como apuntadores. Los registros se organizan como colecciones de grafos dirigidos.

Modelo jerárquico

Es similar al modelo de redes, en el sentido en que los datos y relaciones entre los datos se representan mediante registros y enlaces. La diferencia es que en lugar de organizarse como grafos estos lo hacen como colecciones de árboles.

Modelo de datos físico

Se usa para describir datos en un nivel más bajo. Los más conocidos son el modelo de unificación y el modelo de memoria por marcos.

Un esquema de bases de datos se expresa mediante un conjunto de definiciones que se expresa en un lenguaje de definición de datos (LDD). Las instrucciones del LDD se compilan dando lugar a un conjunto de tablas que se almacenan en un archivo especial, el diccionario de datos contiene metadatos que son datos acerca de los datos.

Un lenguaje de manipulación de datos (LMD) es un lenguaje que permite a los usuarios acceder o manipular datos. Hay dos tipos: LMD procedimentales que requieren que se especifiquen los datos requeridos y como se buscarán, y los LMD no procedimentales que solo requiere que se especifique que datos se requieren.

El gestor de transacciones es el responsable de asegurar que la base de datos permanezca en un estado consistente a pesar de los fallos del sistema. El gestor de transacciones también se asegura que las transacciones ocurran sin conflictos.

Un gestor de almacenamiento es un módulo de programa que proporciona la interfaz entre los datos de bajo nivel almacenados en la base de datos y los programas de aplicación y las consultas enviadas al sistema. El gestor de almacenamiento es responsable de la interacción con los datos almacenados en el disco.

Estructura del sistema completo (Figura 1)

Un sistema de bases de datos se divide en módulos que se encargan de cada una de las responsabilidades del sistema completo. Algunas de estas fusiones del sistema de base de datos las pueden proporcionar el sistema operativo de la computadora. En la mayoría de los casos, los sistemas operativos de la computadora proporcionan solo los servicios más básicos y los sistemas de bases de datos deben constituirse sobre esta base. Así, el dueño de una base de datos debe incluir consideraciones de la interfaz entre el sistema de bases de datos y el sistema operativo.

Los componentes funcionales de un sistema de bases de datos se pueden dividir a grandes rasgos en componentes de procesamiento de consultas y componentes de gestión de almacenamiento. Los componentes de procesamiento de consultas incluyen:

- ?? Compilador del LMD, que traduce las instrucciones del LMD en lenguaje de consultas a instrucciones a bajo nivel que entiende el motor de evaluación de consultas.
- ?? Precompilador del LMD incorporado, que convierte las instrucciones del LMD incorporadas en un programa de aplicación en llamadas a procedimientos normales en el lenguaje anfitrión.
- ?? Intérprete del LDD, que interpreta las instrucciones del LDD y las registra en un conjunto de tablas que contienen metadatos.
- ?? Motor de evaluación de consultas que ejecuta las instrucciones a bajo nivel generadas por el compilador del LMD.

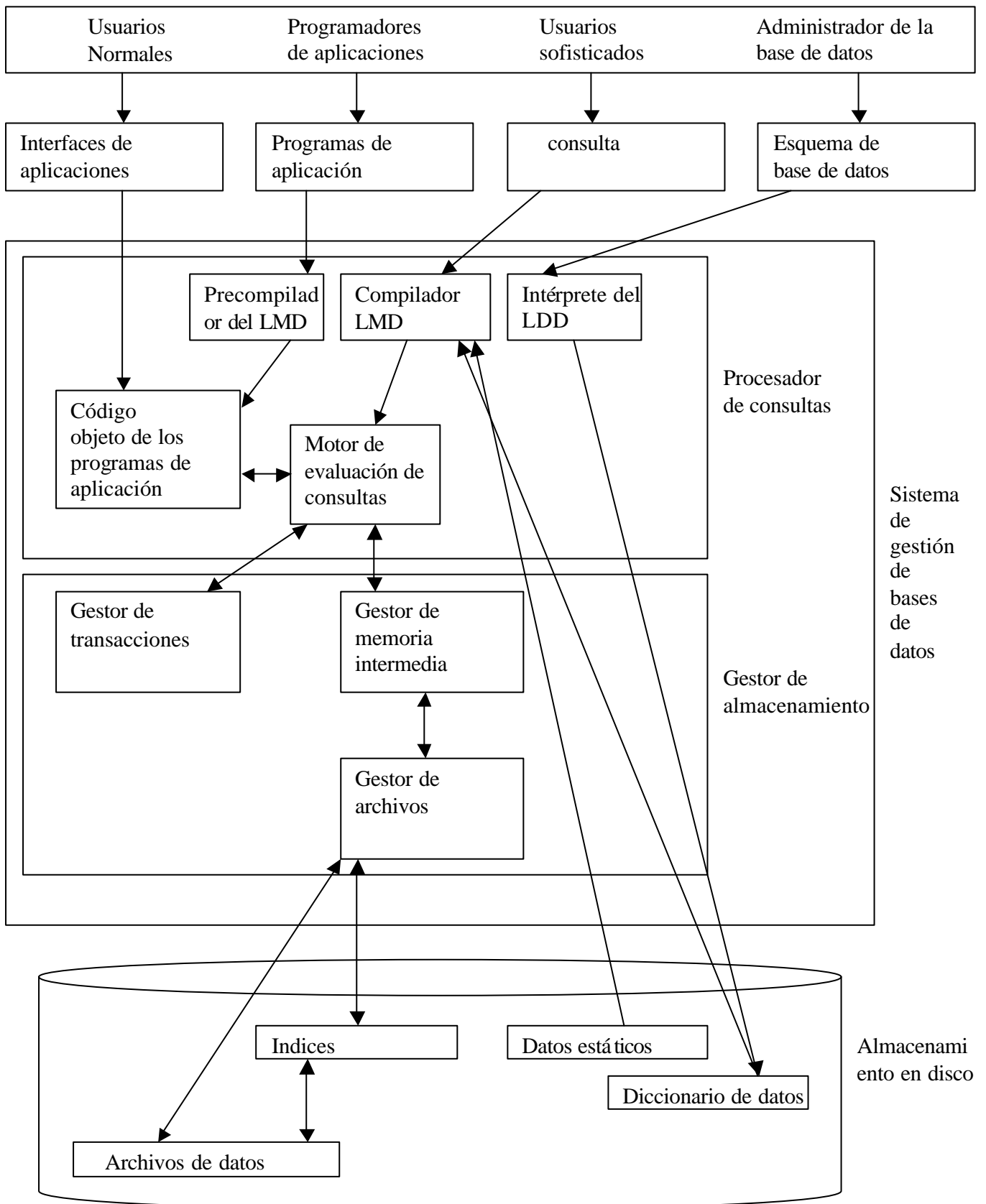


Figura 1. Esquema global de un sistema manejador de bases de datos.

Los componentes de gestión de almacenamiento proporcionan la interfaz entre los datos de bajo nivel almacenados en la base de datos y los programas de aplicación y envío de consultas al sistema. El gestor de almacenamiento tiene:

- ?? Gestor de autorización e integridad que comprueba que se satisfagan las ligaduras de integridad y la autorización de los usuarios para acceder a los datos.
- ?? Gestor de transacciones que asegura que la base de datos quede en un estado consistente a pesar de los fallos del sistema y que las ejecuciones de transacciones concurrentes ocurran sin conflictos.
- ?? Gestor de archivos que gestiona la reserva de espacio de almacenamiento de disco y las estructuras de datos usadas para representar la información almacenada en el disco.
- ?? Gestor de memoria intermedia que es responsable de traer los datos del disco de almacenamiento a memoria principal y decidir que datos traer a memoria caché.

También se necesitan diversas estructuras de datos como parte de la implementación física del sistema:

- ?? Archivos de datos que almacenan la base de datos en sí.
- ?? Diccionario de datos que almacena metadatos acerca de la estructura de la base de datos.
- ?? Índices que proporcionan acceso rápido a elementos de datos que tienen valores particulares.
- ?? Datos estáticos que almacenan información estadística sobre los datos en la base de datos, el procesador de consultas usa esta información para seleccionar las formas eficientes para ejecutar una consulta.

USUARIOS DE LA BASE DE DATOS

Hay cuatro diferentes tipos de usuarios de un Sistema de Base de Datos, diferenciados por la forma en que ellos esperan interactuar con el sistema.

- ?? Programadores de aplicaciones. Son profesionales informáticos que interactúan con el sistema a través de llamadas del LMD que están incluidas en un programa escrito en un lenguaje anfitrión (Cobol, Pascal, PL/I, SQL, VB, Java). Estos programas se llaman programas de aplicación.
- ?? Usuarios sofisticados. Interactúan con el sistema sin programas escritos. Ellos forman sus consultas en un lenguaje de consulta de la base de datos. Cada consulta se envía al procesador de consultas cuya función es transformar instrucciones LMD a instrucciones que el gestor de almacenamiento entienda.
 - Gestor de almacenamiento. Es un módulo de programa que proporciona la interfaz entre los datos de bajo nivel almacenados en la base de datos y los programas de aplicación y las consultas.

- ?? Usuarios especializados. Escriben aplicaciones de BD especializadas que no son adecuadas en el marco de procesamiento de datos tradicional. Entre estas aplicaciones están los sistemas de diseño asistido por computadora, sistemas de bases de conocimientos y expertos y sistemas de modelado del entorno.
- ?? Usuarios normales (finales). Son usuarios no sofisticados que interactúan con el sistema mediante la invocación de los programas de aplicación permanentes que se han escrito previamente.

Administrador de la base de datos (ABD, DBA): Es la persona que tiene control central del sistema. El ABD tiene las siguientes funciones:

- ?? Definición del esquema: El ABD crea el esquema original de la base de datos escribiendo un conjunto de definiciones que el compilador del LDD (lenguaje de definición de datos) traduce a un conjunto de tablas que son almacenadas permanente en el diccionario de datos.
- ?? Estructura de almacenamiento y definición del método de acceso: El ABD crea las estructuras de almacenamiento apropiadas y métodos de acceso escribiendo un conjunto de definiciones, que son traducidas por el compilador del lenguaje de definición y almacenamiento de datos.
- ?? Esquematación y modificación de la organización física: Los programadores llevan a cabo escasas modificaciones sobre el esquema de base de datos o la descripción de la organización de almacenamiento físico, para generar las modificaciones en las tablas correspondientes del sistema interno.
- ?? Concesión de la automatización para el acceso a datos: Esto permite al administrador determinar a qué partes de la base de datos pueden acceder los diferentes usuarios.
- ?? Especificación de las ligaduras de integridad: Los valores de los datos almacenados en la base de datos deben satisfacer ciertas ligaduras de integridad. Por ejemplo, quizás el número de horas que un empleado pueda trabajar en una semana no deba exceder de un límite especificado. Tales ligaduras deben ser especificadas explícitamente por el administrador de la base de datos.