

Capítulo 1 Introducción

Fundamentos de Bases De Datos

Silberchatz, Korth and Sudarshan

Cuarta Edición

Ed. Mc Graw Hill

Un *sistema gestor 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 es denominada *base de datos* y contiene información relevante para una empresa.

Los sistemas de bases de datos 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 información. Deben proporcionar fiabilidad ante cualquier imprevisto como alguna caída del sistema o el ingreso de intrusos sin autorización.

APLICACIONES DE LOS SISTEMAS DE BASES DE DATOS

Las bases de datos hoy en día son ampliamente usadas a continuación enumeraremos algunos de los campos donde es más cotidiano su uso.

- Banca: Para información de clientes, números de cuentas, transacciones y préstamos.
- Líneas Aéreas: Para las reservaciones, horarios de vuelos y destinos.
- Universidades: Para información de los estudiantes, matriculas, asignaturas y cursos.
- Transacciones de Tarjeta de Crédito: Para compras, y generación mensual de extractos.
- Telecomunicaciones: Para registros de llamadas, facturación e información de redes.
- Finanzas: Información sobre empresas, ventas, compras, bolsa y bonos.
- Ventas: Información de clientes, productos y compras.
- Recursos Humanos: Información sobre empleados, salarios, impuestos y beneficios.

A lo largo de las últimas décadas el uso de las bases de datos ha crecido enormemente, puesto que ya es muy común encontrarlas en todo momento de nuestras vidas a veces inconscientemente las estamos usando, a tal grado que empresas que se dedican únicamente al desarrollo de software de bases de datos como Oracle han crecido casi como Microsoft o IBM.

HISTORIA DE LOS SISTEMAS DE BASES DE DATOS

Una de las tareas más importantes de la computadora es el procesamiento de datos, es esta característica la que ha impulsado su crecimiento. Desde mucho antes de que aparecieran las computadoras comerciales ya teníamos el primer predecesor a las bases de datos actuales, estas son las tarjetas perforadas que desarrollara Hollerith, las cuales fueron utilizadas al principio del siglo XX para un censo en los Estados Unidos, en ese tiempo eran manipuladas por sistemas mecánicos, sin embargo después fueron usadas ampliamente como medio para introducir datos a las computadoras.

Las técnicas de almacenamiento de datos han evolucionado a lo largo de los años.

Década de los 50's: Se desarrollaron las cintas magnéticas para el almacenamiento de datos. Las tareas de procesamiento de datos fueron automatizadas con los datos almacenados en cintas, este consistía en leer datos de una o más cintas y escribir datos en una nueva cinta, los datos también podían leerse desde las tarjetas perforadas e incluso ser impresos.

Década de los 60 y 70: El uso de los discos duros cambió en gran medida el escenario del procesamiento de datos, ya que los discos duros permitieron el acceso directo a los datos sin importar la ubicación física que tuvieran, se liberaron de lo tedioso de la secuencialidad.

Década de los 80's: Las bases de datos relacionales no pudieron competir con el rendimiento de las bases de datos de red y jerárquicas existentes. Esto cambió con la aparición de System R, un proyecto innovador de IBM Research que desarrolló técnicas para la construcción de un sistema de bases de datos relacionales eficiente.

Década de los 90's: El lenguaje SQL se diseñó fundamentalmente para las aplicaciones de ayuda a la toma de decisiones, que son intensivas en consultas cuando se tienen grandes cantidades de información. Muchos vendedores de bases de datos introdujeron muchos productos de bases de datos paralelas en este periodo, así como también comenzaron a ofrecer bases de datos relacionales orientadas a objetos.

Finales de los 90's: El principal acontecimiento fue el crecimiento explosivo del world wide web, las bases de datos se implantaron mucho más extensivamente que antes. Los sistemas de bases de datos actuales tienen capacidad para sistemas de transacciones muy altas, así como alta fidelidad y disponibilidad.

SISTEMAS DE BASES DE DATOS CONTRA SISTEMAS DE ARCHIVOS

Una manera que usaban antes para almacenar la información era conservarla en archivos del sistema operativo. Para que los usuarios manipularan dicha información se tenían que crear programas con una aplicación específica como:

- Un programa para efectuar cargos o abonos.
- Un programa para añadir una cuenta nueva.
- Un programa para calcular el saldo de una cuenta.
- Un programa para generar operaciones mensuales.

Mantener la información almacenada de esta forma presentaba algunos inconvenientes.

Redundancia e inconsistencia de datos: Debido a que los archivos y programas son creados por diferentes programadores en un largo periodo de tiempo, los diversos archivos tienen probablemente diferentes formatos y los programas pueden estar escritos en diferentes lenguajes de programación, la misma información puede estar duplicada en diversos lugares.

Dificultad en el acceso de datos: Con el paso del tiempo puede pasar que cierto usuario necesita algún tipo de información que no fue previsto en el diseño del programa, el usuario no tenía más que buscar la información manualmente.

Aislamiento de Datos: Al estar los datos dispersos en varios archivos, era muy común que ciertos datos quedaran inmersos en archivos que ya no se abrían.

Problemas de Integridad: Los valores de algunos datos almacenados en los archivos no podían satisfacer ciertos tipos de restricciones de consistencia.

Problemas de Atomicidad: La computadora como cualquier sistema mecánico o eléctrico está sujeta a fallos, en este caso no se podía asegurar que cuando ocurriera un fallo los datos se pudieran restaurar al estado que se encontraban antes del fallo.

Anomalías en el acceso concurrente. Podía ocurrir que hubiera errores de actualización cuando dos usuarios accedieran a los mismos datos simultáneamente.

Problemas de Seguridad: Existían muy pocas restricciones de seguridad, casi cualquier usuario podía acceder a la información sin mucho esfuerzo.

ABSTRACCION DE DATOS

Como muchos usuarios de las bases de datos no están muy familiarizados con las computadoras, los desarrolladores esconden la complejidad a los usuarios a través de varios niveles de abstracción para simplificar el entorno usuario-sistema:

Nivel Físico: Describe como se almacenan realmente los datos, se describen con detalle las estructuras de datos complejas de bajo nivel

Nivel Lógico: Describe que datos se almacenan en la base de datos y que relaciones existen entre esos datos.

Nivel de Vistas: Es básicamente el entorno del usuario de base de datos amigable para aquellos que no tienen mucha experiencia con las computadoras.

MODELOS DE DATOS

Bajo la estructura de la base de datos se encuentra el modelo de datos, es una colección de herramientas conceptuales para describir los datos, las relaciones, la semántica y las restricciones de consistencia.

Modelo Entidad Relación (E-R).

Esta basado en una percepción del mundo real que consta de una colección de objetos básicos, llamados *entidades* “cosas” y de relaciones entre estas cosas, ellos de describen en la base de datos mediante un conjunto de *atributos*, se le llama *relación* a la asociación entre varios entes. La estructura lógica de una base de datos se puede representar con un diagrama E-R que consta de :

Rectángulo: Representa entidades

Elipse: Representa atributos de las entidades

Rombos: Representa relaciones entre conjunto de entidades.

Líneas: Sirven para unir a los anteriores.

Modelo Relacional.

En este modelo se utilizan un grupo de tablas para representar los datos y las relaciones entre ellos. Cada tabla esta compuesta por varias columnas y cada columna tiene un nombre único.

Modelo de Datos Orientado a Objetos

Es como una extensión del modelo E-R con las nociones de encapsulación, métodos, e identidad de objeto.

Modelo de Datos Relacional Orientado a Objetos.

Combina las características del modelo de datos orientado a objetos y el modelo de datos relacional.

LENGUAJES DE BASES DE DATOS

Un sistema de bases de datos proporciona un *lenguaje de definición de datos* para especificar el *esquema* de la base de datos y un *lenguaje de manipulación de datos* para expresar las consultas a la base de datos y modificaciones.

Lenguaje de Definición de Datos

Un esquema de base de datos se especifica mediante un conjunto de definiciones expresadas mediante un lenguaje especial llamado *Lenguaje de Definición de Datos (LDD)*

Lenguaje de Manipulación de Datos (LMD)

Es un lenguaje que permite a los usuarios acceder o manipular los datos organizados mediante el modelo de datos apropiado. Existen 2 tipos básicamente:

Procedimentales: Requieren que el usuario especifique qué datos se necesitan y cómo obtenerlos.

Declarativos: Requieren que el usuario especifique qué datos se necesitan sin especificar como obtenerlos.

Acceso a la Base de Datos desde Programas de Aplicación

Los programas de aplicación son programas que se usan para interactuar con la base de datos, se escriben en C++, Cobol, Java. Los programas necesitan comunicarse con el programa anfitrión por medio de estándares por ejemplo:

ODBC (Open Data Base Connectivity), es el estándar para acceder en lenguaje C

JDBC (Java Data Base Connectivity) es el estándar para acceder en Java

USUARIOS Y ADMINISTRADORES DE LA BASE DE DATOS

Las personas que trabajan con una base de datos se pueden catalogar como usuarios de bases de datos o como administradores de bases de datos.

Usuarios de Bases de Datos:

Hay 4 tipos de usuarios de un sistema de bases de datos, diferenciados por la forma en que ellos esperan interactuar con el sistema.

Usuarios normales: Son los usuarios no sofisticados que interactúan con el sistema mediante la innovación de alguno de los programas de aplicación permanentes que se han escrito previamente.

Programadores de Aplicaciones: Son profesionales informáticos que escriben programas de aplicación.

Los Usuarios Sofisticados: Interactúan con el sistema sin programas escritos. En su lugar, ellos forman sus consultas en un lenguaje de consultas de bases de datos.

Usuarios Especializados: Usuarios sofisticados que escriben aplicaciones de bases de datos especializadas que no son adecuadas para el marco de procesamiento de datos tradicional.

Administrador de la Base de Datos

Es la persona que tiene el control central sobre el sistema y sus funciones son:

- Definición del esquema
- Definición de la estructura y el método de acceso
- Modificación del esquema y de la organización física
- Concesión de autorización para el acceso de los datos
- Mantenimiento rutinario

ESTRUCTURA DE UN SISTEMA DE BASES DE DATOS

Gestor de Almacenamiento

Es un módulo de programa que proporciona la interfaz entre los datos de bajo nivel en la base de datos y los programas de aplicación, es responsable de la interacción con el gestor de archivos. Los datos en bruto son almacenados en un disco usando un sistema de archivos determinado por el sistema operativo. Es responsable del almacenamiento, recuperación y actualización de los datos. Sus componentes son:

Gestor de Autorización e Integridad: Comprueba que se satisfagan las restricciones de integridad.

Gestor de Transacciones: Se asegura que la base de datos quede en un estado consistente a pesar de los fallos del sistema.

Gestor de Archivos: Gestiona la reserva de espacio en disco, y representa la información ya almacenada.

Gestor de memoria intermedia: Trae los datos del disco de almacenamiento a memoria principal.

El gestor de almacenamiento implementa varias estructuras de datos como parte de la implementación física del sistema.

Archivos de Datos: Son los que almacenan la base de datos

Diccionario de Datos: Almacena meta datos acerca de la estructura de la base de datos

Índices: proporcionan un acceso rápido a elementos de datos que tienen valores particulares.

Procesador de Consultas

Sus componentes son:

Interprete del LDD: Interpreta las instrucciones del LDD y registra las definiciones en el diccionario de datos.

Compilador del LMD: Traduce las instrucciones del LMD en un lenguaje de consultas

Motor de Evaluación de Consultas: Ejecuta las instrucciones de bajo nivel generadas por el compilador.