

Introducción a las bases de datos



Fran Sánchez, Kike Rueda y Fran Ramírez

1 Sistemas de almacenamiento.

Un almacenamiento de datos es un repositorio central de información que se puede analizar para tomar decisiones mejor informadas. Los datos fluyen hacia un almacenamiento de datos desde sistemas transaccionales, bases de datos relacionales y otros orígenes, normalmente a una cadencia regular. Los analistas empresariales, los ingenieros de datos, los científicos de datos y los responsables de la toma de decisiones obtienen acceso a los datos mediante herramientas de inteligencia empresarial (BI), clientes SQL y otras aplicaciones de análisis. Los datos y el análisis se han vuelto fundamentales para que las empresas mantengan la competitividad. Las empresas utilizan informes, paneles y herramientas de análisis para extraer información de los datos, monitorear el desempeño de la empresa y respaldar la toma de decisiones. Los almacenamientos de datos son fundamentales para estos informes, paneles y herramientas de análisis. Guardan los datos de manera eficaz para reducir las operaciones de entrada y salida (E/S) de datos y entregar resultados de consultas rápidamente a cientos y miles de usuarios de manera simultánea.

La arquitectura de un almacenamiento de datos está compuesta por capas. El nivel superior es el cliente frontend que presenta los resultados mediante informes, análisis y herramientas de minería de datos. El nivel intermedio contiene el motor de análisis que se utiliza para obtener acceso a los datos y para analizarlos. El nivel inferior de la arquitectura es el servidor de la base de datos, donde los datos se cargan y almacenan. Los datos se almacenan de dos diferentes maneras: 1) los datos a los que se obtiene acceso de forma frecuente se almacenan en almacenamiento muy rápido (como unidades SSD); y, 2) los datos a los que se obtiene acceso de forma poco frecuente se almacenan en un almacenamiento de objetos rentable, como Amazon S3. El almacenamiento de datos se encarga automáticamente de que los datos a los que se obtiene acceso de forma poco frecuente se trasladen al almacenamiento “rápido” a fin de optimizar la velocidad de consulta.

El almacenamiento de datos puede contener varias bases de datos. Dentro de cada base de datos, los datos se organizan en tablas y en columnas. Dentro de cada columna, se puede definir una descripción de los datos, como un número entero, un campo de datos o una cadena. Las tablas se pueden organizar dentro de esquemas, que se pueden concebir como carpetas. Cuando los datos se incorporan, se almacenan en varias tablas descritas por el esquema. Las herramientas de consulta usan el esquema para determinar a qué tablas de datos obtendrán acceso y cuáles de ellas analizarán.

2 ¿Qué son las bases de datos?

Una base de datos es una recopilación organizada de información o datos estructurados, que normalmente se almacena de forma electrónica en un sistema informático. Normalmente, una base de datos está controlada por un sistema de gestión de bases de datos (DBMS). En conjunto, los datos y el DBMS, junto con las aplicaciones asociadas a ellos, reciben el nombre de sistema de bases de datos, abreviado normalmente a simplemente base de datos.

Los datos de los tipos más comunes de bases de datos en funcionamiento actualmente se suelen utilizar como estructuras de filas y columnas en una serie de tablas para aumentar la eficacia del procesamiento y la consulta de datos. Así, se puede acceder, gestionar, modificar, actualizar, controlar y organizar fácilmente los datos. La mayoría de las bases de datos utilizan un lenguaje de consulta estructurada (SQL) para escribir y consultar datos.

Aunque las bases de datos pueden contener muchos tipos de datos, algunos de ellos se encuentran mutuamente protegidos por las leyes de varios países. Por ejemplo en España, los datos personales se encuentran protegidos por la Ley Orgánica de Protección de Datos de Carácter Personal (LOPD)

3 Evolución de las bases de datos.

3.1- El principio de las bases de datos

Tiene lugar en la antigüedad cuando se instauró un **sistema manual** para registrar y organizar la información de las cosechas recogidas, los primeros censos y de los registros. Pero el proceso era extremadamente lento, sobre todo cuando se quería encontrar algún dato registrado anteriormente.

Un nuevo avance llegó de la mano del inventor Herman Hollerith en el año 1884, él fue responsable de idear el primer tabulador electromagnético que funcionaba con tarjetas perforadas lo que ayudó a recopilar toda la información.

Años después, en la década de los 50, la llegada de las cintas magnéticas **simplificó el registro y el almacenamiento** de la información, en 1956 apareció el primer **disco duro** de la mano de la multinacional IBM que en contraste con las cintas magnéticas ofrecía acceso aleatorio lo que permitía llegar a un dato en específico con **más rapidez**.

• Años 60

La industria informática seguía avanzando durante la década de 1960 y gracias a este hecho se contaba con una **base de datos computarizada** que coincidió con la aparición de los primeros ordenadores personales.

Durante esta época, continuaba el progreso de la evolución de las bases de datos con dos modelos que dieron de qué hablar. Se trata del modelo de base de **datos de red** ideado por Charles Bachman cuyo propósito era presentar un registro flexible donde los datos se conectaban entre sí a través de enlaces en un sistema de red. Por su parte, en el modelo de bases de **datos jerárquicas** los datos se distribuyen en una figura similar a la de un árbol.

• Años 70

Edgar Frank Codd ideó la base de datos relacional en los 70's. Este científico informático británico marcó la evolución de las **bases de datos** gracias al desarrollo de teorías relacionadas con el modelado de datos.

Larry Ellison diseñó, gracias a los descubrimientos de Edgar Frank Codd, un **sistema de gestión de datos** conocido como Oracle

En 1974 apareció SQL, un tipo de lenguaje de programación, lo que facilitó la estructuración y recuperación de información de bases de datos relacionales ya que permite concretar varios tipos de procesamientos de transacciones.

• Años 80 y 90

En los 80's SQL era el más usado en lenguaje de consultas, y se estandarizó por el Instituto Nacional Estadounidense de Estándares (ANSI) como de la Organización Internacional de Normalización (ISO). SQL comenzó a ser el estándar, ya que las bases de datos relacionales con su sistema de tablas pudieron competir con las bases jerárquicas y de red

Microsoft lanza su sistema de gestión de bases de datos llamado Microsoft Access, ideado para reunir datos e información desde otras plataformas, como Excel.

El lanzamiento de la red informática mundial conocida como World Wide Web (WWW), ha facilitado consultar diferentes bases de datos.

- **Siglo XXI**

Los más conocidos en el mercado de base de datos siguen siendo IBM, Oracle y Microsoft. Aún así, la informática y el comercio electrónico siguen evolucionando para poder competir con las bases de datos más utilizadas y conocidas.

Las bases de datos se han convertido en una herramienta muy útil y se adapta a cualquier sistema operativo. También los sistemas administradores de bases de datos son importantes para dar efectividad a una base de datos, ya que ofrecen accesos inmediatos y son útiles para almacenar, modificar y administrar los datos. La creación de equipos y sistemas cada vez más **sofisticados y seguros** sirvieron para que hoy en día podamos tener acceso a las mejores bases de datos, entre otras muchas cosas.

4 Tipos de bases de datos

Existen diversos tipos de base de datos, pueden clasificarse de diversas maneras. A continuación veremos las principales.

Según su flexibilidad de modificación

Bases de datos dinámicas

Son aquellas donde los **datos pueden actualizarse** o incluso modificarse. La mayoría puede ser actualizada en tiempo real.

Bases de datos estáticas

Son bases de datos de consulta cuyos **datos no pueden modificarse**.

Según su forma de organización

Bases de datos jerárquicas

Las bases de datos jerárquicas son aquellas organizadas en forma de un árbol al revés. Almacenan la información en forma de registros dentro de una estructura jerárquica, es de aquí que proviene su nombre.

Cada registro de este "árbol" es llamado nodo. Nodos son registros que contienen alguna información de interés y a partir del nodo raíz son enlazados los otros nodos descendientes: padres e hijos. Cada nodo padre puede tener varios nodos hijos, pero cada nodo hijo solo puede tener un solo nodo padre.

Este tipo de base de datos se recomienda para administrar grandes volúmenes de informaciones, pero actualmente no se utiliza mucho.

Las principales características de la base de datos jerárquica son:

- Diferentes usuarios pueden acceder y compartir información.
- Los datos son independientes.
- Es una estructura rígida por lo que es difícil modificarla.
- Requiere gran conocimiento de las unidades de información.
- Los nodos distantes del nodo raíz son de difícil acceso por lo que se requiere tiempo.

Bases de datos de red

Esta base de datos es una variación de la anterior. La diferencia está en que en la base de datos jerárquica un nodo hijo no puede tener varios padres y aquí sí.

Las características de estas bases de datos son semejantes a las de las bases de datos jerárquicas, aunque estas son mucho más potentes y complejas.

Bases de datos relacionales

Las bases de datos relacionales son las más usadas actualmente para administrar datos de forma dinámica. Permite crear todo tipo de datos y relacionarlos entre sí.

Los datos son almacenados en registros que son organizados en tablas, de esta forma pueden asociarse los elementos entre sí muy fácilmente, además se pueden cruzar sin ninguna dificultad.

Sus principales características son:

- Pueden ser utilizadas por cualquier persona.
- Son de fácil gestión.
- Se pueden acceder a los datos con rapidez.
- Garantiza la total consistencia de los datos, sin posibilidad de error.
- No son muy eficaces con datos gráficos, CAD, sistema de información geográfica ni datos en multimedia.

Bases de datos deductivas

Conocidas también como bases de datos lógicas. Se utilizan generalmente en buscadores, pero pueden usarse de otras formas.

Permiten almacenar los datos y consultarlos a través de búsquedas que utilizan reglas y normas previamente almacenadas.

Sus principales características son:

- Permite expresar consultas por medio de reglas lógicas.
- Soporta conjuntos de datos complejos.
- Se puede inferir información a través de los datos almacenados.
- Usan fórmulas matemáticas o algoritmos lógicos.

Bases de datos multidimensionales.

Estas bases de datos utilizan conceptualmente la idea de un cubo de datos. Donde las informaciones se almacenan en la intersección de tres o más atributos. Esta concepción puede ser algo compleja pero su uso es bastante simple.

Algunas de sus principales características son:

- No emplean ninguna jerarquía.
- Facilita tanto la búsqueda como la modificación posterior.
- Utiliza un espacio menor de almacenamiento.
- Tiene acceso a grandes cantidades de información.

5 Ejemplos de sistemas gestores de base de datos más conocidos.

Existen los siguientes modelos de bases de datos:

- Relacional
- Jerárquica
- De red
- Orientada a objetos
- Orientada a documentos

El más común y popular es el modelo de base de datos relacional, en el que los datos se estructuran en filas de tabla. La ventaja de este modelo radica en la posibilidad de crear diferentes relaciones entre las filas y presentarlas en columnas. El procedimiento es diferente al del modelo de base de datos jerárquico, donde los diferentes datos se organizan en relaciones padre-hijo, en una estructura similar a la de un árbol.

Otros enfoques para organizar los datos son el modelo de base de datos de red, donde los datos, como el nombre indica, se estructuran en forma de red, o el modelo de bases de datos orientada a objetos, en el que no solo importa la relación entre los registros de datos, sino también el concepto de la herencia: esto significa que los objetos pueden transferir algunos de sus atributos a otros objetos, lo que se regula a través del SGBD.

Por su parte, el modelo de base de datos orientado a documentos permite almacenar los registros de datos en diferentes documentos.