**BASES DE DATOS: CONCEPTO**

Una base de datos es un repositorio estructurado para datos cuyo propósito principal es reservar los datos para algún conjunto de objetivos de la organización. Tradicionalmente, estos objetivos solían ser principalmente administrativos. La mayoría de los sistemas de bases de datos se construyen para contener los datos que son necesarios en las actividades diarias de una organización.

**SISTEMAS DE GESTION DE BASE DE DATOS (SGBD)**

Un sistema gestor de base de datos o SGBD (del inglés: Data Base Management System o DBMS) es un software que permite administrar una base de datos. Proporciona el método de organización necesario para el almacenamiento y recuperación flexible de grandes cantidades de datos. Algunos ejemplos de SGBD son MySQL, MariaDB, PostgreSQL, Microsoft SQL Server, Oracle Database y Microsoft Access.

**UN POCO DE HISTORIA**

**Sistemas de navegación (1960)**

Con la emergente aparición de los sistemas a mediados de los 60, Charles Bachman fundó el Database Task Group dentro del grupo creador de COBOL, CODASYL, para luego en 1971 publicar su estándar conocido como “Aproximación CODASYL”. A su vez, en 1968, IBM también tenía su SGBD propio, conocido como IMS. Se trataba de un software desarrollado para el programa Apolo sobre System/360.

**Sistemas relacionales (1970)**

Edgar Codd, un trabajador de IBM, escribió múltiples artículos con aproximaciones variadas, impulsado por su descontento en la falta de operación de búsqueda. Su investigación culminó en el documento conocido como "A Relational Model of Data for Large Shared Data Banks".

Reconciliar toda la información es la clave de este sistema. En el modelo relacional, una parte de la información se usa como clave, identificando de manera biunívoca un registro concreto. Cuando se recopila información acerca de un usuario, se accederá a la información de las tablas optativas buscando mediante esa clave. Por ejemplo, si el nombre de usuario es único, la dirección y número de teléfono de ese usuario será guardada con el nombre de usuario como clave. La recopilación de esta información en un solo registro es algo para lo que los lenguajes tradicionales no están pensados.

Así como el enfoque de navegación requiere programas que realicen bucles para recolectar registros, el enfoque relacional también los requerirá. La solución de Codd para los necesarios bucles se basa en un lenguaje orientado a conjuntos, una sugerencia que más tarde cristalizaría en el ubicuo SQL. Planteó el uso de una rama del álgebra llamada cálculo de tuplas, y demostró que con ella se podrían realizar todas las operaciones típicas sobre una base de datos, además de extraer conjuntos de datos de una forma sencilla.

El artículo de Codd cayó en manos de dos personas en Berkeley, Eugene Wong y Michael Stonebraker. Ellos comenzaron un proyecto llamado INGRES con fondos asignados a un proyecto de base de datos geográfica programada por los estudiantes. Comenzando en 1973, INGRES produjo sus primeras versiones de prueba que estuvieron listas para uso general en 1979. INGRES era muy similar a System R de IBM en varios aspectos, incluyendo un lenguaje para acceso a los datos, conocido como QUEL. Con el paso del tiempo, INGRES adoptó el estándar SQL.

IBM realizó una implementación de prueba del modelo relacional —PRTV— y una de producción —Business System 12— ambas descontinuadas. Honeywell escribió MRDS para Multics, y aparecen también dos nuevas implementaciones: Alphora Dataphor y Rel. La mayoría de las demás implementaciones de SGBD llamados relacionales son en realidad SGBD SQL.

En la década de 1970, la Universidad de Míchigan comenzó el desarrollo del MICRO Information Management System basado en el modelo teórico de datos de D. L. Childs. Micro fue utilizado para gestionar gran cantidad de datos en el Departamento de Trabajo del gobierno de EUA. Corría en mainframe usando Michigan Terminal System. Estuvo en producción hasta 1998.

Sistemas SQL (finales de década de 1970)

Artículo principal: SQL

IBM empezó a trabajar a principios de 1970 en un prototipo lejanamente basado en los conceptos de Codd llamándolo System R. La primera versión estuvo lista en 1974 o 1975, y comenzó así el trabajo en sistemas multitabla, en los que los datos podían disgregarse de modo que toda la información de un registro (alguna de la cual es opcional) no tiene que estar almacenada en un único trozo grande. Las versiones multiusuario siguientes fueron probadas por los usuarios en 1978 y 1979, tiempo por el que un lenguaje SQL había sido estandarizado. Las ideas de Codd se revelaron como operativas y superiores a las de CODASYL, lanzando a IBM al desarrollo de una verdadera versión de producción de System R, conocido como SQL/DS, y posteriormente como Database 2 (DB2).

Muchos de los técnicos de INGRES estaban seguros del éxito comercial del sistema, y formaron sus propias compañías para comercializar el desarrollo pero con una interfaz SQL. Sybase, Informix, NonStop SQL y la misma INGRES se vendían como derivados del INGRES original en los años 1980. Incluso el SQL Server de Microsoft está basado en Sybase, y por consiguiente en INGRES. Solo Larry Ellison —el fundador de Oracle— comenzó un nuevo camino basado en el artículo de IBM sobre System R, y aventajó a IBM sacando al mercado su primera versión en 1978.

Stonebraker aplicó las lecciones de INGRES al desarrollo de una nueva base de datos —Postgres— conocida ahora como PostgreSQL. PostgreSQL se utiliza para muchas aplicaciones críticas (los registros de dominios .org y .info lo usan para su almacenamiento primario, así como grandes compañías e instituciones financieras).

En Suecia, el artículo de Codd generó la base de datos Mimer SQL4​ en la Universidad de Uppsala. En 1984 este proyecto se consolidó en una compañía independiente. A principios de 1980, Mimer introdujo la gestión de transacciones para dar robustez a las aplicaciones, una idea que fue recogida en muchos otros SGBD.

Sistemas orientados a objetos (1980)

Artículo principal: Base de datos orientada a objetos

Durante la década de 1980 el auge de la programación orientada a objetos influyó en el modo de manejar la información de las bases de datos. Programadores y diseñadores comenzaron a tratar los datos en las bases de datos como objetos. Esto quiere decir que si los datos de una persona están en la base de datos, los atributos de la persona como dirección, teléfono y edad se consideran que pertenecen a la persona, no son datos extraños. Esto permite establecer relaciones entre objetos y atributos, más que entre campos individuales.

Otro gran foco de atención durante la década fue el incremento de velocidad y fiabilidad en el acceso. En 1989, dos profesores de la Universidad de Wisconsin publicaron un artículo en una conferencia ACM en el que exponían sus métodos para mejorar las prestaciones de las bases de datos. La idea consistía en replicar la información importante —y más solicitada— en una base de datos temporal de pequeño tamaño con enlaces a la base de datos principal. Esto implicaba que se podía buscar mucho más rápido en la base de datos pequeña que en la grande. Su mejora de prestaciones llevó a la introducción de la indización, incorporado en la totalidad de los SGBD.

Sistemas NoSQL (2000)

Artículo principal: NoSQL

El siglo XXI trajo una nueva tendencia en las bases de datos: el NoSQL. Esta tendencia introducía una línea no relacional significativamente diferentes de las clásicas. No requieren por lo general esquemas fijos, evitan las operaciones join almacenando datos desnormalizados y están diseñadas para escalar horizontalmente. La mayor parte de ellas pueden clasificarse como almacenes clave-valor o bases de datos orientadas a documentos.

Recientemente ha habido una gran demanda de bases de datos distribuidas con tolerancia a particiones, pero de acuerdo con el teorema CAP no es posible conseguir un sistema distribuido que simultáneamente proporcione consistencia, disponibilidad y tolerancia al particionado. Un sistema distribuido puede satisfacer solo dos de las tres restricciones a la vez. Por dicha razón muchas de las bases de datos NoSQL usan la llamada consistencia eventual para proporcionar disponibilidad y tolerancia al particionado, con un nivel máximo de consistencia de datos.

Entre las aplicaciones más populares encontramos MongoDB, MemcacheDB, Redis, CouchDB, Hazelcast, Apache Cassandra y HBase, todas ellas de código abierto.

Sistemas XML (2010)

Las bases de datos XML forman un subconjunto de las bases de datos NoSQL. Todas ellas usan el formato de almacenamiento XML, que está abierto, legible por humanos y máquinas y ampliamente usado para interoperabilidad.

En esta categoría encontramos: BaseX, eXist, MarkLogic Server, MonetDB/XQuery, Sedna**.**