

INTRODUCCIÓN A LAS BASES DE DATOS.



1. Introducción a las Bases de Datos

Las bases de datos son el pilar fundamental para la organización y gestión de la información en el mundo digital. Se definen como conjuntos estructurados de datos, interrelacionados entre sí, que permiten almacenar y recuperar información de forma eficiente.

Su uso abarca desde pequeñas agendas personales hasta sistemas complejos que gestionan millones de registros en empresas, gobiernos y organizaciones de todo tipo. Las bases de datos nos permiten:

Almacenar grandes cantidades de información de forma organizada y segura.

Acceder a la información de forma rápida y eficiente, utilizando diferentes criterios de búsqueda.

Analizar los datos para obtener información útil y tomar decisiones estratégicas.

Compartir la información con otros usuarios de forma segura y controlada.

Modificar, eliminar y actualizar datos según requerimientos.

En general las bases de datos son herramientas esenciales para la gestión de la información en la era digital. Dominar sus conceptos básicos permitirá aprovechar al máximo las posibilidades que ofrece este mundo, desde organizar tus tareas personales hasta contribuir al desarrollo de grandes proyectos.



Los altos volúmenes de datos hoy son el insumo para la Inteligencia Artificial Generativa.

Operaciones que se hacen con los datos:

- Recolección
- Clasificación
- Búsqueda
- Conservación
- Transformación

Con las bases de datos se realizan operaciones comunes llamadas CRU:D: (Create, Read, Update y Delete).

1.1-Porque guardar la informacion?

Para retomar y acumular datos históricos que garanticen la trazabilidad de procesos, evidencias, guardar la historia de los hombres, las empresas y la sociedad, además de ser valiosos para el entrenamiento de sistemas de Inteligencia artificial.

Para consolidar y sacar resultados generales, tendencias y análisis de datos.

En gran volumen son el insumo de la inteligencia artificial.

1.2-Que es una BD

Es una colección organizada de información o datos que se almacenan de manera estructurada. Es una estructura de datos compleja con gran capacidad de almacenamiento y lenguaje que permite operaciones sobre ella.

1.3-Como guardar la información

Las bases de datos se guardan con sistemas especiales para la gestion de bases de dato llamadas “Sistemas de gestion de Bases de datos” (DBMS).



Algunos DBMS:

Una estructura de datos es una forma de organizar información para que pueda ser utilizada de manera recurrente y eficiente. Las bases de datos constituyen la memoria de la información que manejamos. Nos referimos a esto como persistencia de datos.

1.4- Memoria volátil vs No volátil

Se refiere a estructuras de datos temporales o permanentes para manipular y procesar o guardar resultados finales en medios persistentes.

1.4.1 Memoria volátil o temporal:

RAM (Random Access Memory), es un tipo de memoria informática que almacena temporalmente datos e instrucciones que están siendo utilizados activamente por un computador. Volátil- porque su contenido no se conserva cuando se apaga el dispositivo o PC.

En programación la memoria volátil se maneja en estructuras de datos básicas como arreglos, pila, colas, y árboles binarios entre otros.

Estas memorias son volátiles se borran al apagar el computador o dispositivo. Algunas denominaciones RAM, DRAM, SRAM

1.4.2 Memoria no volátil:

Es cualquier tipo de sistema de memoria de computadora que almacena datos sin necesidad de tener encendido el equipo o dispositivo. Los datos son almacenados en



estructuras de datos complejas que guardan la información y permiten realizar operaciones sobre ellas, la información es almacenada en: Cintas magnéticas, Discos (flexibles y duros), memorias externas (usb, cd, CD, micro sd,), memoria de estado sólido (SSD) , y hoy en diferentes lugares, pc locales o en sitios remotos corporativos o globales como los servicios en la nube.

2. Breve historia de las bases de datos

1. Primeros pasos (1884 - 1960):

- **1884:** Herman Hollerith crea el tabulador electromagnético de tarjetas perforadas, utilizado para el censo de Estados Unidos.
- **1950s:** Las primeras computadoras UNIVAC I utilizan cintas magnéticas para almacenar datos.
- **1960s:** Se desarrolla el modelo jerárquico y de red para organizar datos en estructuras de árbol y punteros.

2. Nacimiento de las bases de datos relacionales (1970 - 1980):

- **1970:** Edgar Frank Codd define el modelo relacional, basado en la teoría de conjuntos y en la organización de datos en tablas con relaciones entre sí.
- **1979:** Se lanza SQL (Structured Query Language), el lenguaje estándar para interactuar con bases de datos relacionales.

3. Expansión y diversificación (1990 - actualidad):

- **1990s:** Surgen las bases de datos orientadas a objetos para gestionar datos complejos.
- **2000s:** Aparece NoSQL como alternativa a las bases de datos relacionales para grandes volúmenes de datos no estructurados.
- **2010s:** Crecimiento de la nube y la computación distribuida, impactando el desarrollo de las bases de datos.

Hitos importantes:

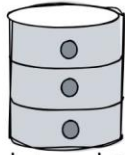
- **1963:** Se utiliza por primera vez el término "base de datos" en un simposio en California.
- **1976:** Se lanza el primer SGBD (DBMS) relacional comercial, System R de IBM.
- **1980s: 1990s** Inicios de las bases de datos relacionales y aparición de productos comerciales con productos como Oracle, Sybase, DB2.
- **2008:** Google lanza Bigtable, una base de datos NoSQL para grandes volúmenes de datos.
- **2010s:** Crecimiento de las bases de datos en la nube como Amazon RDS y Microsoft Azure SQL Database.

Tipos de bases de datos:

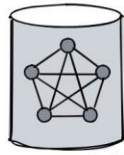
- **Bases de datos relacionales:** Almacenan datos en tablas con relaciones entre sí. Ejemplo: MySQL, Oracle, Postgress, Sql Server.
- **Bases de datos NoSQL: (Not Only SQL)** Diseñadas para grandes volúmenes de datos., como manejo de documentos columnas y otras estructuras. Ejemplo: MongoDB, Casandra
- **Bases de datos en la nube:** Soluciones de bases de datos alojadas en la nube. Ejemplo: Amazon RDS. MongoDB, Cassandra.



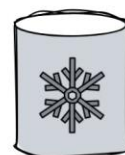
CIENCIA DE DATOS - IA



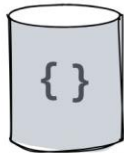
Relational DB



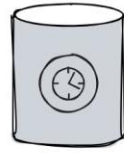
Graph DB



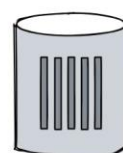
Wide-Column Store



Document DB



Time Series Database

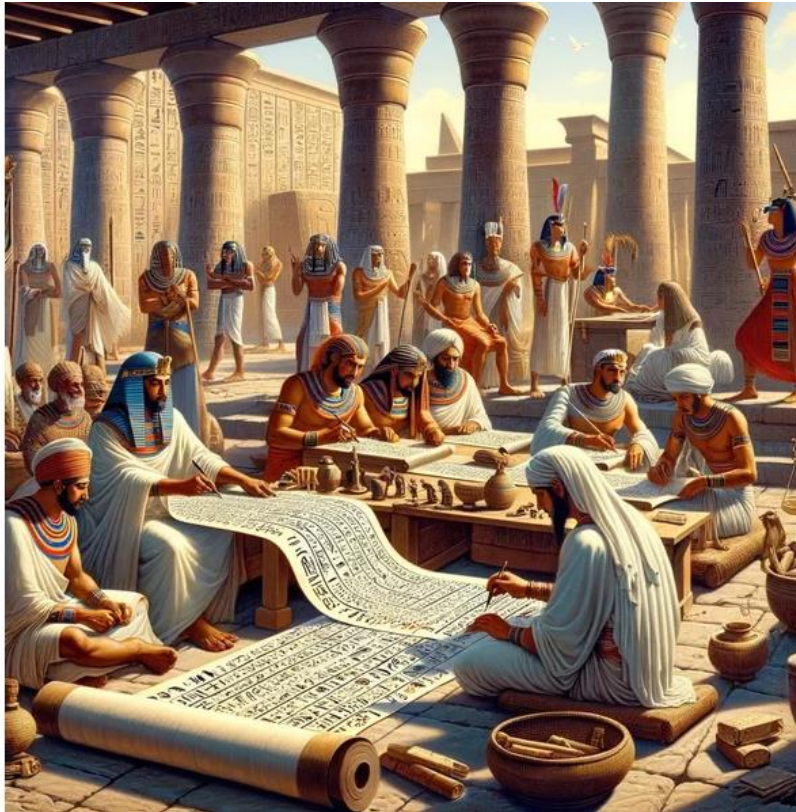


Columnar DB

.

Un poco de Historia

Las primeras civilizaciones utilizaban métodos rudimentarios de registro de datos. Los eventos generaban información valiosa que fue recopilada en diferentes medio según la época.



Estos hitos han jugado un papel crucial en la evolución de las bases de datos hasta lo que son hoy.

Desarrollo de SABRE (1960): IBM y American Airlines desarrollaron SABRE, un sistema operativo que manejaba las reservas de vuelos, transacciones e informaciones sobre los pasajeros de la compañía American Airlines.

El Uso del término "Base de Datos" (1963)**: se usó por primera vez en un simposio celebrado en California.

En 1884, los censos se realizaban de forma manual, hasta que Herman Hollerith inventó la máquina automática de perforación de tarjetas

1950, El inglés Oberlin Smith inicio automatizacion de la información de nóminas, a través de cintas magnéticas, que a su vez respaldaban dicha información.

CODASYL (también escrito Codasyl) es el acrónimo para "Conference on Data Systems Languages", un consorcio de industrias informáticas formado en 1959 con el objeto de regular el desarrollo de un lenguaje de programación estándar que pudiera ser utilizado en multitud de ordenadores. De todos estos esfuerzos resultó el lenguaje COBOL

2. **Invención de la máquina de tarjetas perforadas (1884)**: Herman Hollerith inventó la máquina automática de perforación de tarjetas, que se usó en el censo de los Estados Unidos, mejorando significativamente el proceso de terminación¹⁴.

Década de los 70 Creación del modelo relacional (1970): Edgar Frank Codd definió el modelo relacional, que propuso una forma estructurada de organizar la información en tablas, facilitando su acceso y manipulación.

Lawrence "Larry" Ellison, pudo desarrollar el Relational Software System o sistema de datos ORACLE, un sistema de administración de Base de Datos relacionados, el cual se destacaba por su estabilidad, escalabilidad, transacciones y multiplataforma.

IBM desarrolló unas técnicas para construir un sistema de bases de datos relacionales llamado System R.

El Sistema R utilizó el lenguaje de consulta Secuela, el cual ha contribuido al desarrollo de SQL / DS, DB2, Allbase, Oracle y SQL

Ingres DBMS en la Universidad de California (1974-1977) el lenguaje llamado QUEL. que sirvió de base para la creación de Ingres Corporación, MS SQL Server, Sybase, PACE Wang, y Britton.

En la década de los años 80', se desarrolló el SQL (Structured Query Language), un lenguaje de consultas que permite consultar, con el fin de recuperar información de una base de datos y manipularla de manera simple.

3. ****Desarrollo de SABRE (1960)**:** IBM y American Airlines desarrollaron SABRE, un sistema operativo que manejaba las reservas de vuelos, transacciones e informaciones sobre los pasajeros de la compañía American Airlines².

Creación del modelo relacional (1970):** Edgar Frank Codd definió el modelo relacional, que propuso una forma estructurada de organizar la información en tablas, facilitando su acceso y manipulación¹³.

Adopción de SQL como estándar de la industria (1986):** SQL comenzó a ser el estándar de la industria, ya que las bases de datos relacionales con su sistema de tablas pudieron competir con las bases jerárquicas y de red².

En la década de los años 80', se desarrolló el SQL (Structured Query Language), un lenguaje de consultas que permite consultar, valga la redundancia, con el fin de recuperar información de una base de datos y a su vez, hacer cambios sobre esa misma base, de forma sencilla

Em los 90' la investigación en bases de datos giró en torno a las bases de datos orientadas a objetos. Así se desarrollaron herramientas como Excel y Access del paquete de Microsoft Office que marcan el inicio de las bases de datos orientadas a objetos.

En 2000' Con la llegada de Internet y la WWW (Word Wide Web) las bases de datos se hicieron más accesibles y evolucionaron poco a poco hasta lo que conocemos

hoy en día nuevos paradigmas como DBMS, almacenamiento en la nube, Data Mining, BigData, DataWarehousing, IA.

De 2010 hasta hoy han surgido, Diversidad de bases de datos SQL y NoSQL, manipulación de DataSet, Formatos diferentes: XML, JSON; Analítica de Datos, Inteligencia Artificial.



Las bases de datos han evolucionado constantemente desde sus inicios, adaptándose a las necesidades cambiantes del mundo digital. Su desarrollo ha impactado significativamente la forma en que se gestiona y analiza la información, siendo una herramienta fundamental para el funcionamiento de empresas, gobiernos y organizaciones en la era digital.

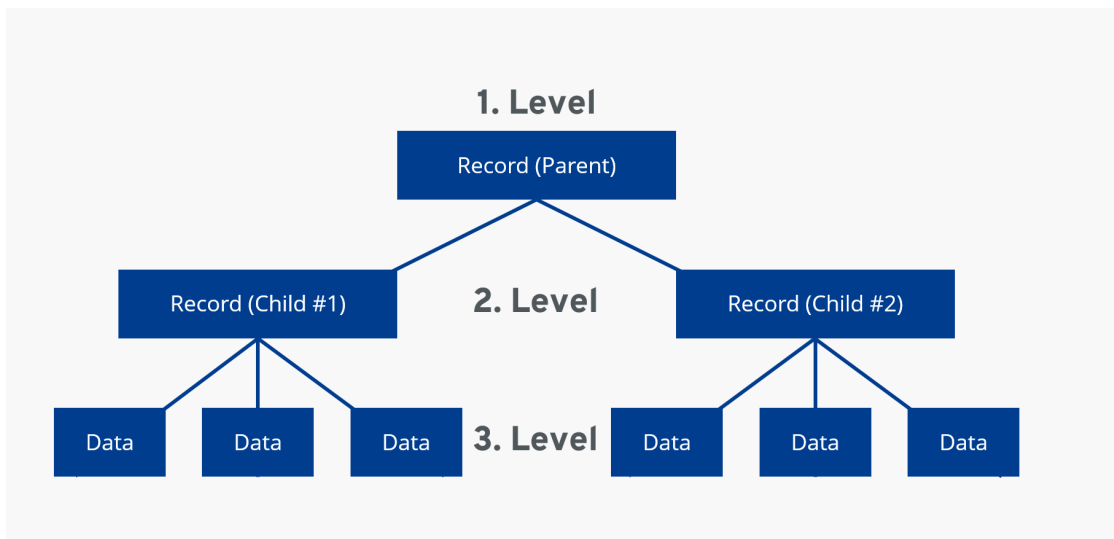
2.1 BASES DE DATOS JERARQUICAS

son bases de datos que almacenan la información en una estructura jerarquizada, concretamente los datos son organizados de forma de un árbol. El nodo superior (la raíz) contiene la información más general, y de él se desprenden nodos secundarios (hijos) con información más específica relacionada con el nodo padre. Cada nodo padre puede tener varios nodos hijos, pero un nodo hijo solo puede tener un nodo padre.

Acceso a la información: Para acceder a un registro específico, se debe seguir la ruta a través de la jerarquía, comenzando desde la raíz y bajando por los nodos hijos hasta llegar al registro deseado. Esto puede ser un poco inflexible comparado con otros modelos de bases de datos.

Actualmente, las bases de datos jerárquicas han sido en gran parte reemplazadas por las bases de datos relacionales que ofrecen mayor flexibilidad y capacidad para modelar relaciones complejas entre datos. Sin embargo, todavía se pueden encontrar en algunos sistemas heredados o en aplicaciones específicas donde la estructura jerárquica es muy adecuada para la información que se maneja.

Un ejemplo es la base de datos IMS de IBM.



En grafos y árboles se utilizan algoritmos de búsqueda y recorrido BFS y DFS.

Las principales características de las bases de datos jerárquicas:

- Se organizan en forma de árbol invertido, con un nodo raíz (0), nodos padre e hijos.

- El árbol se organiza en niveles.
- Los arcos (enlaces) son asociaciones jerárquicas entre dos nodos,
- Uno nodo padre puede tener un número ilimitado de nodos hijos, pero a un nodo hijo solo le puede un padre.
- Solo pueden existir relaciones de uno a uno de uno a varios.
- El árbol siempre se recorre en un orden prefijado.
- La estructura del árbol no se puede modificar cuando ha quedado establecida.

Ventajas BD Jerarquía

- Las conexiones dentro del árbol son fijas y hace que la navegación por ellas sea rápida.
- Muestra una estructura de la base de datos fácil de ver y comprender.
- Permite predefinir relaciones, simplificando las variaciones futuras.
- Permite mantener la integridad e independencia de la información.

Desventajas BD Jerarquía

- Escasa independencia entre los registros (nodos), puesto que, para acceder a un registro, se debe pasar por los padres, algo que quita
- Implica una mala gestión de la redundancia de datos, puesto que cuando un registro tiene relación con dos o más registros, debe almacenarse varias veces, puesto que un hijo no puede tener varios padres.
- Lo anterior implica un mayor volumen de datos y posibles problemas en la integridad y coherencia de los datos, puesto que, si se modifica una de las copias de un registro, se deben modificar también las restantes.
- Sin embargo, modificar este tipo de bases de datos resulta complejo por su rigidez.

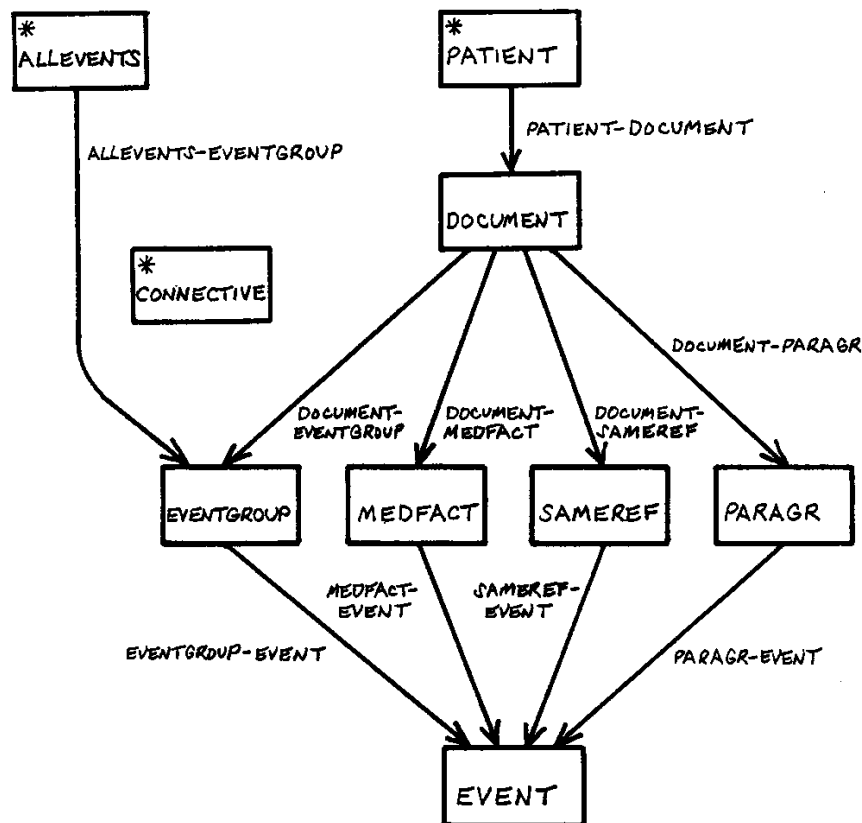
2.2 BASE DE DATOS EN RED

Una base de datos de red es un tipo de base de datos que almacena datos en una estructura de red. Cada registro es un nodo que está conectado a uno o más nodos mediante enlaces. Esta estructura permite relacionar los datos de una manera más flexible que en una base de datos jerárquica.

Un ejemplo de esta es CODASYL es el estándar para los sistemas de bases de datos de red. Este estándar se basa en la estructura jerárquica de la información.

Dio origen al Lenguaje COBOL. originalmente para trabajar con el modelo de datos jerárquico CODASYL.

Ejemplos Modelo Codasyl:



Ventajas BD Red

- Flexibilidad en la organización de los datos.
- Permite relaciones múltiples entre los datos.
- Capacidad para recuperar datos de manera más rápida y eficiente.

Desventajas BD Red

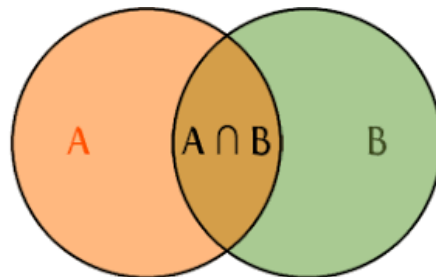
- Mayor complejidad en el diseño y la implementación.
- Mayor costo en términos de hardware y software.
- Mayor tiempo de respuesta y complejidad de diseño.

2.3- BASES DE DATOS RELACIONALES

Existían muchas dificultades con las BD red y las Jerárquicas por su complejidad creciente, no tenían un conocimiento sólido que permitiera crear un estándar de fácil difusión, entendimiento y aplicación a nivel global. Después de muchos modelos, las bases de datos encontraron su fundamento en la teoría de Conjuntos que dio origen al modelo de bases de datos relacionales.

2.3.1 Teoría de conjuntos

Se dedica a estudiar las características de los conjuntos y las operaciones que pueden efectuarse entre ellos. Esta teoría alcanza consenso mundial y es conocida en todo lugar, constituye la base para el modelo relacional. Donde cada conjunto hace referencia a un volumen de datos referentes a una entidad.



2.3.2 Las operaciones entre conjuntos

Son la esencia de las consultas de información en una base de datos relacional.

Unión: Combina dos conjuntos en uno solo, eliminando duplicados.

Intersección: Encuentra los elementos que son comunes a dos conjuntos.



Diferencia: Determina los elementos que están en un conjunto, pero no en otro.

Complemento: Conformar un conjunto con todos los elementos del conjunto universal que no pertenecen al conjunto original.

Diferencia simétrica: Selecciona los elementos de un conjunto que no están en el otro, y los elementos del otro conjunto que no están en el primero.

Producto cartesiano: Genera un nuevo conjunto a partir de todos los pares ordenados posibles que se pueden formar tomando el primer elemento del par del primer conjunto y el segundo elemento del par del segundo conjunto.

Estas operaciones y otras hacen parte de lo cotidiano con el manejo de bases de datos.

Las bases de datos relacionales son conocidas también como bases de datos SQL.

Referencias

Historia de las Bases de Datos: <https://informatecdigital.com/bases-de-datos/historia-de-las-bases-de-datos-una-guia-completa/>

Que es una base de datos: <https://www.ionos.es/digitalguide/hosting/cuestiones-tecnicas/bases-de-datos/>

Fecha Creación	Enero 25 2024
Responsable	Plinio Neira Vargas
Revisado por	Sonia Escobar
Fecha Revisión	Febrero 10 2024



Taller 00:

Introducción a las Bases de datos

1. Consulta de 6 bases de datos (DBMS) SQL y No SQL, productos y fabricantes.
2. Consultar 3 sistemas de hojas electrónicas.
3. Consultar el algoritmo DFS

<https://www.techiedelight.com/es/depth-first-search/>

- Implementarlo en Python
- Analizar el código
- Hacer pruebas -ejecutarlo

Ver video DSF, sugerencias:

<https://www.youtube.com/watch?v=UAIDAxof3kA>

https://www.youtube.com/watch?v=_Yf8tneauJ8