



Universidad Galileo

Carrera: Técnico universitario en desarrollo Full Stack

Curso: Matemática para desarrolladores.

Sede: Virtual

Investigación base de datos

Nombre: Maidellin Suset Alvarado Cayax

Carné: 24011377

Sección: T

Fecha de entrega: 12/11/2024

BASE DE DATOS



Una base de datos es una recopilación organizada de información o datos. Se almacenan de forma electrónica en un sistema informático

La base de datos esta controlada por un sistema de gestión de base de datos. Se almacenan en forma de filas y columnas. La mayoría de estos datos utilizan lenguaje de consulta estructurada SQL para escribir y consultar datos.

Importancia de gestionar y organizar datos de forma estructurada

Al estructurar los datos, es mucho más fácil acceder a ellos y recuperar la información específica. Mejoramos la seguridad y la gestión de los datos.

Al tener los datos organizados, se pueden descubrir patrones y estadísticas también puntos de inflexión y graficas.



Una estructura bien gestionada reduce que los datos se guarden dos veces, lo cual reduce el riesgo de errores y mejora la precisión de la información.

Bases de Datos Relacionales (SQL)

Organiza los datos mediante filas y columnas. Se debe definir un esquema (estructura de datos) antes de almacenar datos. Esto asegura que todos los registros tengan el mismo formato.

Se suelen escalar agregando más recursos al servidor (escalabilidad vertical), aunque también existen algunas opciones de escalabilidad horizontal.



Donde se usa?

- ✓ Aplicaciones bancarias y financieras:
- ✓ Sistemas de gestión empresarial
- ✓ Sistema de gestión de inventario

Bases de Datos No Relacionales (NoSQL)

Almacena los datos sin un esquema personalizado o definido, y son ideales para trabajar con datos no estructurados. No dependen de tablas y permiten almacenar datos en diferentes formatos, como documentos, gráficos, columnas.

Tipos de bases de datos NoSQL:

Almacenan datos en documentos JSON o similares. Ejemplo: MongoDB.

Almacenan datos en pares clave-valor, ideales para almacenamiento en caché. Como por ejemplo Redis.

Organizan datos por columnas en lugar de filas, eficaces para análisis de datos. Como por ejemplo Cassandra.

Almacenan datos en grafos para analizar relaciones complejas entre entidades por ejemplo: Neo4j.



Características:

Se pueden escalar de forma distribuida en múltiples servidores, lo cual es ideal para grandes volúmenes de datos. También está diseñada para operaciones de lectura y escritura rápidas, siendo estas más lentas.

Como usarla?:

- ✓ Aplicaciones web y móviles en tiempo real
- ✓ Blogs, plataformas de contenido multimedia o comercio electrónico, donde la estructura de datos varía.

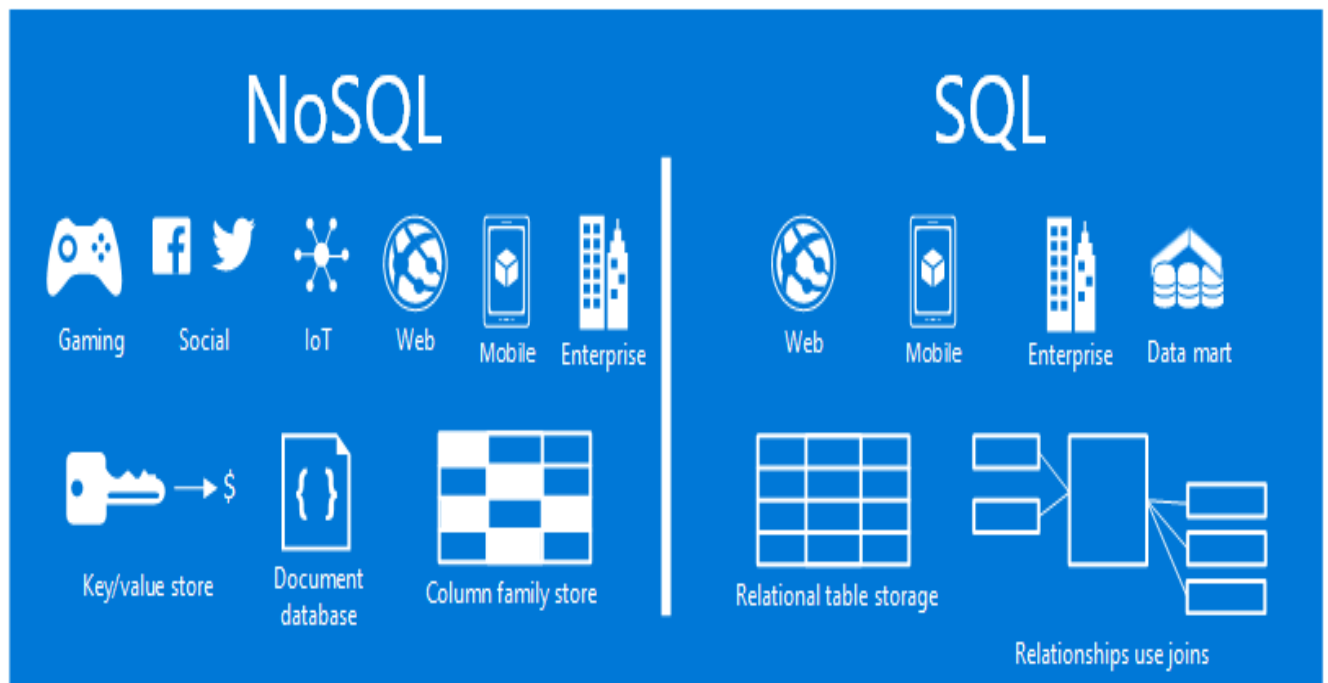
Comparación entre SQL y NoSQL:

SQL organiza los datos en tablas que se relacionan entre sí mediante claves, lo que permite estructurar los datos de forma organizada y de manera rápida, mientras que NoSQL son flexibles y están diseñadas para almacenar grandes volúmenes de datos sin un esquema fijo, sacrificando la velocidad.

SQL	NoSQL
Los datos se organizan en tablas con filas y columnas.	No requieren un esquema fijo
Registros precisos y realizar consultas rápidas	Lenta en búsqueda
Escalable vertical	Escalable horizontal.
Finanzas, ERP, inventarios	Aplicaciones web, Big Data, análisis en tiempo real

Se escalan principalmente agregando más potencia al servidor en lugar de distribuirse entre varios.

Dado que la estructura de datos puede variar, permiten una gran flexibilidad.



Bases de datos relacionales más comunes

MySQL

Es una base de datos relacional de código abierto y su facilidad de uso, su velocidad en consultas básicas y su adaptabilidad para aplicaciones de mediano y gran tamaño. Fácil de configurar y rápida para operaciones de consulta de datos, especialmente en aplicaciones de lectura intensiva.

Es adecuada para aplicaciones que crecen en número de usuarios y volumen de datos.



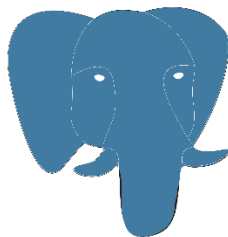
Aplicaciones comunes:

- Es muy popular en el desarrollo de aplicaciones web y sitios dinámicos, como los desarrollados en conjunto con PHP o JavaScript. Es utilizado por plataformas como WordPress, Joomla y Drupal para almacenar contenido de sitios web.
- MySQL, aunque no está diseñado como un sistema de almacén de datos, puede servir en entornos de análisis ligeros, donde la integración con herramientas de BI como Power BI o Tableau permite consultas y visualización de datos.

- Permite manejar inventarios, catálogos de productos y transacciones, siendo usada en plataformas como Magento y WooCommerce.
- Permite realizar análisis simples de datos mediante consultas SQL, siendo útil en pequeñas empresas que no requieren una infraestructura analítica avanzada.
- Aplicaciones de redes sociales utilizan MySQL para perfiles de usuario, publicaciones y conexiones entre usuarios.

PostgreSQL

- Su enfoque está en la conformidad con los estándares SQL y en el soporte para datos estructurados y semi-estructurados.
- Asegura consistencia y confiabilidad en operaciones complejas. Diseñado para gestionar grandes volúmenes de datos y múltiples operaciones simultáneas.



PostgreSQL

Aplicaciones comunes:

- PostgreSQL es una elección común para almacenar datos de aplicaciones web y móviles debido a su compatibilidad con ORMs (Object-Relational Mappers) como Sequelize, Django etc.
- Es común en sistemas de comercio electrónico debido a su rendimiento en transacciones financieras y consultas complejas.

- PostgreSQL se puede usar junto con herramientas de análisis de datos como Apache Spark y herramientas de BI como Tableau, para realizar consultas analíticas y reportes.
- Procesa grandes volúmenes de datos y la posibilidad de trabajar con estructuras complejas lo hacen útil para llevar registros contables y realizar análisis de riesgos financieros.
- Es útil para la preparación de datos en ciencia de datos, ayuda a limpiezas y agregaciones directamente en la base de datos antes de usar los datos en modelos de machine learning.

Oracle

Es conocida por su seguridad, confiabilidad y capacidades de gestión de datos. Incluye avanzadas opciones de respaldo, recuperación y encriptación de datos. Es escalable, permitiendo tanto escalabilidad vertical como horizontal, y cuenta con opciones de base de datos en la nube. Oracle es ideal para aplicaciones de análisis de datos en tiempo real.

ORACLE®
D A T A B A S E



Casos comunes de uso:

- Lo usan los bancos y sistemas financieros para manejar transacciones y operaciones de gran volumen y en tiempo real por su capacidad de manejar datos complejos y transacciones complejas en tiempo real.

- Oracle se usa para aplicaciones de investigación médica y desarrollo farmacéutico, donde se requiere un alto nivel de procesamiento y análisis por obvias razones.
- Oracle se usa en la administración de infraestructuras de servicios públicos, como redes de agua y energía, donde se requiere una base de datos confiable para almacenar datos operativos y realizar análisis.

Tipos de bases de datos NoSQL

Documentales:

Son bases de datos que almacenan información en forma de documentos, tenemos formatos como JSON, BSON o XML.

Los documentos pueden tener estructuras de datos variables, es decir, no todos los documentos tienen que seguir el mismo esquema, lo que permite agregar o modificar campos sin afectar a otros documentos.

Son muy adecuadas para datos semi-estructurados y estructuras jerárquicas, ya que permiten mayor flexibilidad en la forma de los datos. No son eficientes como las bases de datos relacionales para operaciones que requieren muchas relaciones o uniones complejas entre datos porque acá manejamos textos y cadenas no claves.

Se usan en aplicaciones web que manejan datos de usuario, productos, perfiles, como tiendas en línea, redes sociales, y plataformas de mensajería.

Ejemplos de servicios de base de datos NoSQL: MongoDB, CouchDB

Clave-Valor:

Almacenan datos como pares clave-valor, donde una clave única se asocia a un valor específico. Funcionan de forma similar a un diccionario

Son rápidas e ideales para búsquedas directas mediante claves, y se utilizan frecuentemente en aplicaciones que requieren almacenar grandes cantidades de datos con consultas rápidas, como en cachés de datos.

No es un esquema rígido y no es difícil, el acceso a los datos mediante clave es muy rápido y eficiente.

Este tipo de almacenamiento está asociado a la clave. Puede ser un dato simple (como un número o texto) o un objeto complejo (como un JSON o un blob de datos binarios), dependiendo de la base de datos y el caso de uso.

Ejemplos de servicios: Redis, DynamoDB. Amazon DynamoDB

Columnar:

Los datos columnar almacenan datos por columnas. Esto hace que las consultas que afectan a un conjunto de columnas específicas sean mucho más rápidas y eficientes.

Ideales para manejar grandes volúmenes de datos y realizar análisis de datos masivos, como en entornos de big data.

Las bases de datos columnar leen solo los datos necesarios para una consulta específica. Esto reduce la cantidad de datos leídos del disco y acelera las consultas. Son escalables, ya que están diseñadas para distribuir los datos entre múltiples nodos o servidores.

Ejemplos: Apache Cassandra, HBase.

Grafos:

Se utilizan principalmente para casos de uso que requieren entender relaciones y conexiones complejas, como redes sociales, sistemas de recomendación y análisis de redes.

Utilizamos grafos para recomendar productos o contenidos basados en trends o en recomendaciones de otros usuarios. En telecomunicaciones, los grafos ayudan a la creación de redes de conexiones y encontrar patrones o detectar fallos.

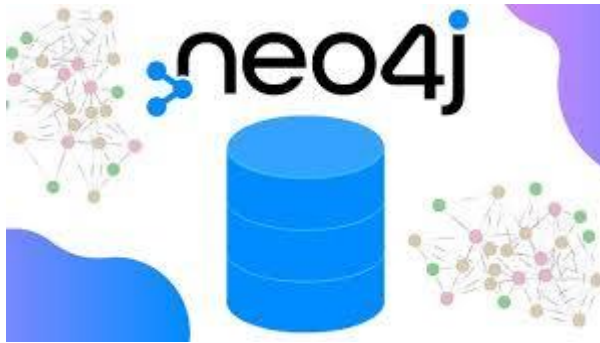
Ejemplos de servicios: Neo4j, ArangoDB.

¿Qué es Neo4j?

Neo4j utiliza un modelo de grafos que organiza los datos en nodos y aristas (o relaciones), lo que permite una representación estructurada de cómo los datos están interrelacionados

Nos muestra entidades, como personas, lugares, cosas, etc. Cada nodo puede tener propiedades asociadas de esta manera se guardan.

Las relaciones pueden tener direcciones y también pueden contener propiedades. En Neo4j, las relaciones están integradas en la estructura del grafo, lo que hace que las consultas sean mucho más rápidas, incluso cuando los datos son grandes y las relaciones son complejas.



Neo4j Relacionados con la Matemática:

Podemos buscando conexiones, recomendaciones o influencers (análisis de centralidad). Analizar patrones de compra.

También las conexiones entre transacciones y personas para detectar patrones inusuales que podrían indicar fraude. En este caso montos que excedan el límite o fondos que no se encuentren en la cuenta del que desea retirar.

En sistemas de transporte o logística, donde es necesario encontrar la ruta más eficiente entre nodos (como ciudades o almacenes).

Modelar interacciones entre proteínas, genes, o procesos biológicos complejos como redes metabólicas. En conclusiones nos ayuda a modelar matemáticamente patrones, además a analizar la coherencia de los datos, recomendaciones a base de repetición etc.