

Base de Datos NoSQL

Parte A

Fernando Thul

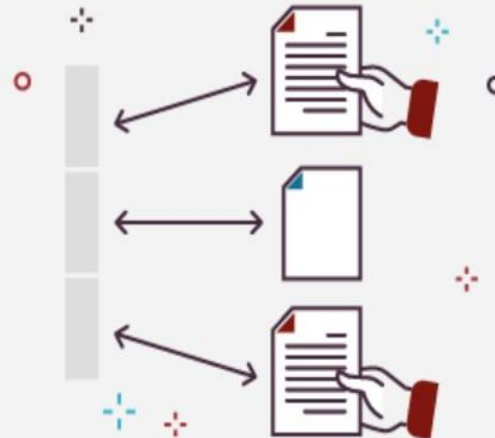
fernando.thul@gmail.com

Objetivos de aprendizaje

Objetivos de aprendizaje



Identificar las distintas **bases de datos NoSQL** y reconocer sus **ventajas y limitaciones**



Familiarizarse con las **bases de datos documentales** y sus usos



Utilizar **MongoDB** para manipular, clasificar y resumir los datos y obtener información

¿ Que son las Bases de Datos ?

Las bases de datos son sistemas organizados para almacenar y administrar grandes cantidades de información de manera eficiente y estructurada. Se trata de una colección organizada de datos relacionados o no que se almacenan en un formato electrónico y pueden ser accedidos, actualizados y gestionados fácilmente.

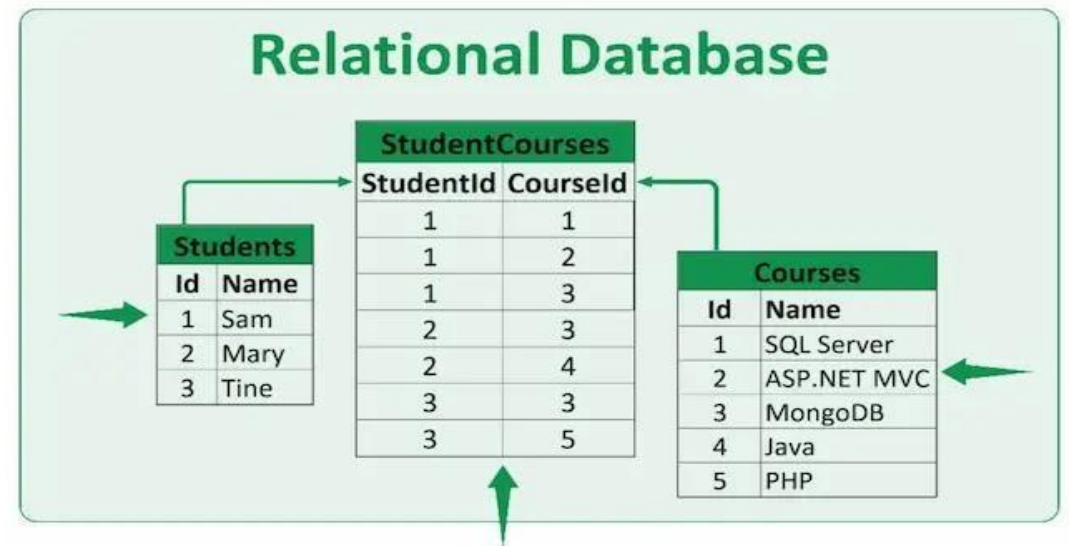


¿ Que tipos de Base de Datos existen ?

1. **Bases de Datos Relacionales (RDBMS):**

Son las bases de datos más tradicionales y ampliamente utilizadas. Almacenan datos en tablas con filas y columnas, y utilizan el lenguaje SQL para gestionar y consultar la información.

Ejemplos famosos son MySQL, PostgreSQL y Microsoft SQL Server.



¿ Que tipos de Base de Datos existen ?

2. **Bases de Datos NoSQL:**

Estas bases de datos se desarrollaron para abordar limitaciones de las bases de datos relacionales en ciertos escenarios. NoSQL significa "No Solo SQL".

Vienen en diferentes formas, como bases de datos de documentos, columnas, clave-valor y gráficos. Ejemplos incluyen MongoDB (documentos), Cassandra (columnas) y Redis (clave-valor).

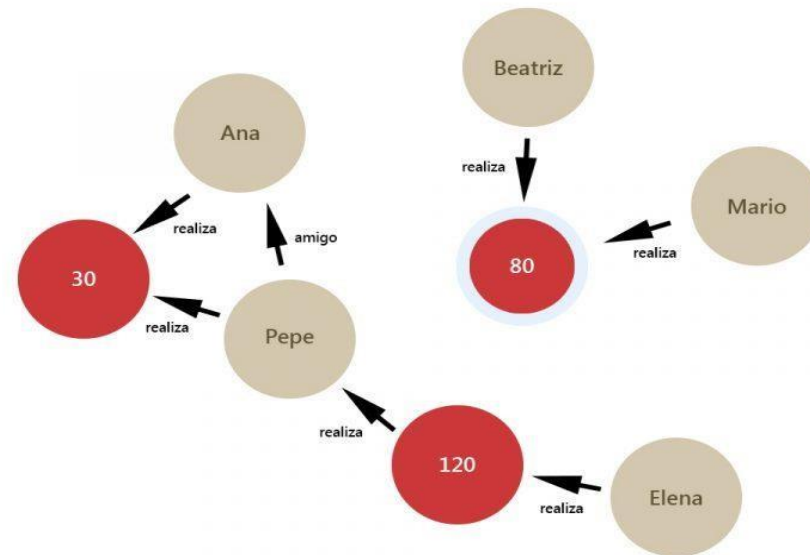


¿ Que tipos de Base de Datos existen ?

3. **Bases de Datos de Grafos:**

Están diseñadas para almacenar y gestionar datos interconectados, como redes sociales o relaciones complejas. Utilizan estructuras de grafos para representar los datos y permiten consultas eficientes sobre esas relaciones.

Ejemplos son Neo4j y Amazon Neptune.



¿ Que tipos de Base de Datos existen ?

4. **Bases de Datos Columnares:**

Organizan los datos en columnas en lugar de filas, lo que es especialmente eficiente para consultas que involucran agregaciones y análisis. Son comunes en el análisis de datos y la inteligencia empresarial.

Ejemplos incluyen Apache HBase y Google Bigtable.

Column family (Table)

partition key columns ...

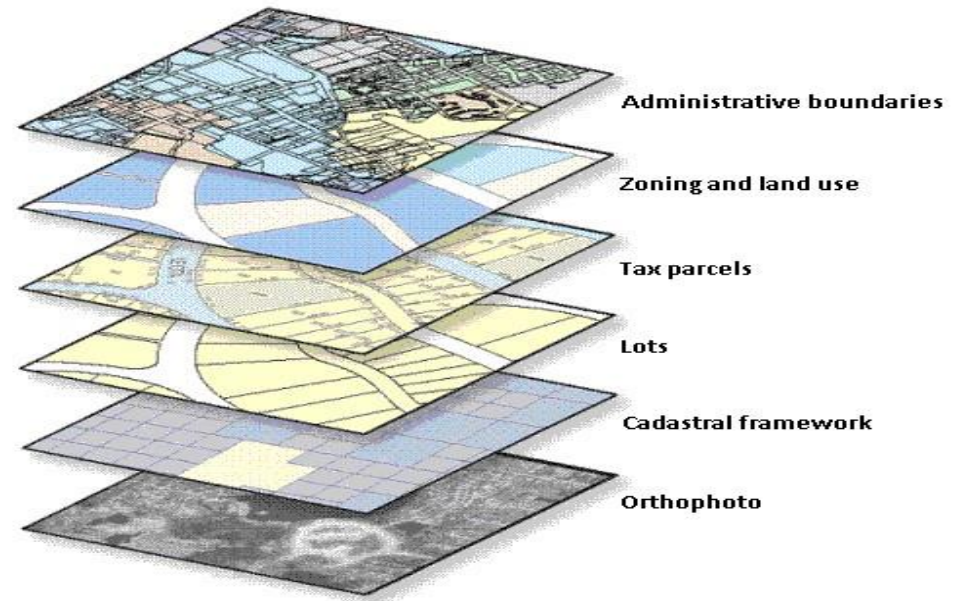
101	email	name	tel	
	ab@c.to	otto	12345	
103	email	name	tel	tel2
	karl@a.b	karl	6789	12233
104	name			
	linda			

¿ Que tipos de Base de Datos existen ?

5. **Bases de Datos Espaciales:**

Estas bases de datos se centran en el almacenamiento y la manipulación de datos geográficos y espaciales, como mapas y coordenadas GPS. Son cruciales en aplicaciones de mapas y sistemas de información geográfica (GIS).

PostGIS es un ejemplo popular.



¿ Que tipos de Base de Datos existen ?

6. **Bases de Datos de Documentos**

Las bases de datos documentales están diseñadas para almacenar datos semiestructurados, como documentos, normalmente en formato JSON o XML.

Las bases de datos documentales permiten expresar los datos en su forma natural. MongoDB es la más difundida



Aplicaciones de Base de Datos Relacionales

3. **Salud y Medicina:**

En la atención médica, las bases de datos almacenan historiales médicos, diagnósticos, tratamientos y programación de citas, facilitando la atención y el seguimiento de pacientes.



Aplicaciones de Base de Datos Relacionales

4. **Educación:**

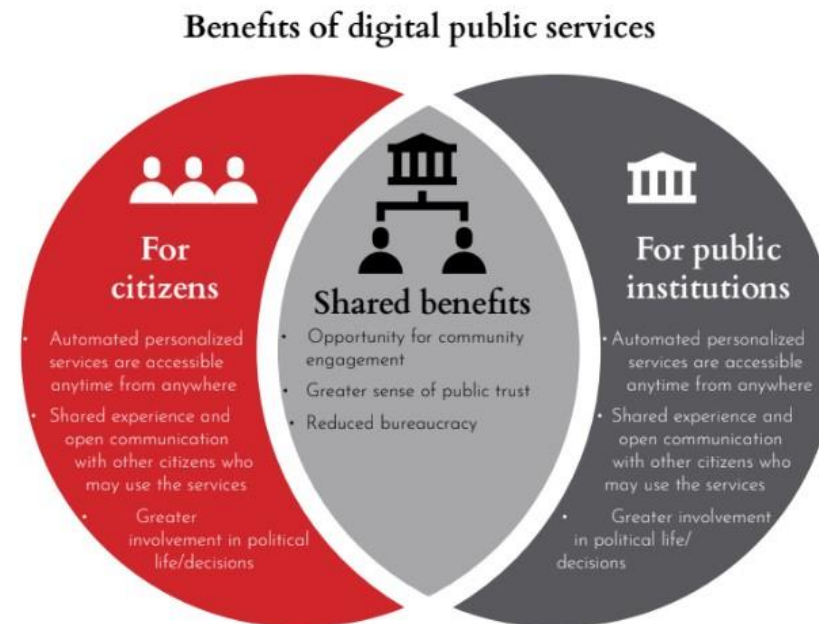
Las bases de datos se utilizan en sistemas de gestión educativa para gestionar registros estudiantiles, horarios, calificaciones y recursos educativos.



Aplicaciones de Base de Datos Relacionales

5. **Gobierno y Administración Pública:**

Las bases de datos se utilizan para administrar registros públicos, censos, permisos y otra información gubernamental.



En qué casos podemos elegir BD NoSQL

Las aplicaciones que pueden soportar fallos (pérdida de datos menores, retraso en la replicación, lectura de escritura sucia, etc.) pueden utilizar bases de datos NoSQL

Las aplicaciones para las redes sociales, los juegos móviles y web, y el streaming eligen NoSQL por su necesidad de escalabilidad frente a la consistencia, estos tipos de aplicaciones funcionan bien con la consistencia eventual

Aplicaciones en las que las interacciones de los usuarios pueden dividirse en combinaciones de transacciones en línea y fuera de línea

Hablemos sobre bases de datos NoSQL

RECUERDE



Lo que ya sabemos sobre las bases de datos relacionales: son prácticas (modelo de datos simple, SQL), multiusuario, seguras, constantes, fiables, sólidas y eficientes.

Ventajas:

- Se utilizan en distintos ámbitos
- Datos normalizados
- Consultas exhaustivas
- Alta compatibilidad

Desventajas

- Esquemas fijos
- Combinación de tablas costosa
- Estructuras de datos limitadas
- Difícil de escalar

TENGA EN CUENTA



Clases de base de datos NoSQL

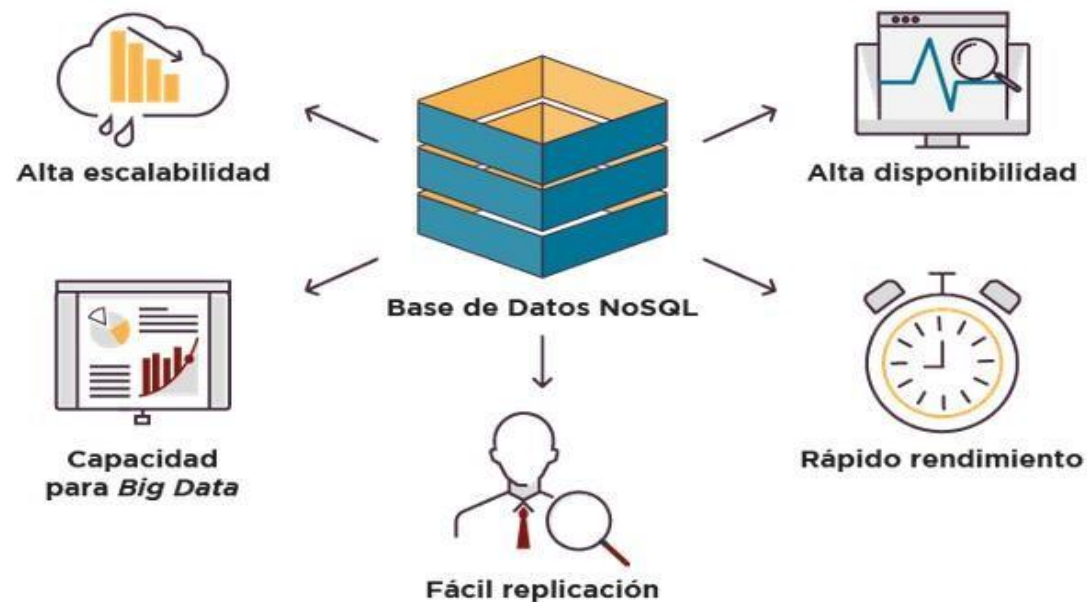
Como vimos en el módulo 2, existen cuatro clases de bases de datos no relacionales:

- Base de datos de clave-valor: almacena una estructura de datos diccionario (o *hash*). Cada valor está asociado a una clave única. Útil para almacenar datos en caché.
- Base de datos de grafos: las entidades se almacenan en grafos, nodos, aristas y propiedades.
- Orientada a documentos: los datos se almacenan en forma de documentos (múltiples pares clave-valor).
- Familia de columnas: las entidades se almacenan en columnas y no en filas. Los datos están desnormalizados.

Hablemos sobre bases de datos NoSQL

Las **bases de datos NoSQL** se utilizan cuando se trabaja con **grandes conjuntos de datos distribuidos**. Operan en aplicaciones web en tiempo real, y su uso se ha incrementado con el tiempo. Una base de datos NoSQL tiene un diseño simple, un buen control de la disponibilidad y un escalado horizontal simple a clústeres de máquinas. Las operaciones de las NoSQL son más rápidas que las de las bases de datos relacionales debido al uso de distintas estructuras de datos. El problema reside en el factor decisivo que supone la idoneidad de la base de datos.

Propiedades de una Base de Datos NoSQL



Hablemos sobre bases de datos NoSQL

Atributos de los sistemas NoSQL

Alternativa a los SGBDR tradicionales:

- Esquema flexible
- Configuración más rápida y barata
- Gran escalabilidad (cantidad de datos y eficiencia)
- Coherencia distendida
- Mayor rendimiento y disponibilidad

Lenguaje de consulta no declarativo: más programación para encontrar patrones para las tareas relacionadas con la ciencia de los datos.

Ventajas de las bases de datos NoSQL

Escalado flexible

- El SGBDR puede escalar cargas mayores en servidores más grandes.
- Los NoSQL aumentan la escala distribuyendo, perfectamente, los datos en múltiples *hosts*.

Especialistas en administradores de BD

- Los SGBDR necesitan expertos bien entrenados para monitorizar la BD.
- Los NoSQL requieren menos gestión debido a la reparación automática y a unos modelos de datos más sencillos.

Ventajas de las bases de datos NoSQL

Big Data

- Gran aumento de datos SGBDR: mayor capacidad de datos y límite de volumen
- NoSQL diseñado para *big data* (replicación/distribución/datos estructurados)

Modelos de datos flexibles

- Cambios en la gestión del esquema, puesto que el SGBDR debe administrarse con cuidado.
- Las bases de datos NoSQL son más distendidas en cuanto a la estructura de los datos.
- Los cambios del esquema de la base de datos no se tienen que gestionar como una unidad de cambio complicada.
- La aplicación está diseñada para abordar un esquema amorfo.

Desventajas de las bases de datos NoSQL

Asistencia

- Los proveedores de SGBDR proporcionan un alto nivel de asistencia a sus clientes.
- Extraordinaria reputación
- Los NoSQL son proyectos de código abierto con empresas emergentes que los apoyan.
- Reputación aún no establecida

Madurez

- El SGBDR es un producto maduro, estable y fiable.
- Los NoSQL continúan implementando su conjunto de funciones básicas.

Desventajas de las bases de datos NoSQL

Administración

- El administrador de los SGBDR tiene un papel bien definido.
- NoSQL no necesita un administrador. Sin embargo, mantener los NoSQL requiere esfuerzo.

Falta de expertos

- Existe una población activa de desarrolladores de SGBDR capacitados y experimentados.
- NoSQL sigue reclutando desarrolladores.

Desventajas de las bases de datos NoSQL

Análisis e inteligencia empresarial

- Los SGBDR están diseñados para este sector.
- Los NoSQL están diseñados para satisfacer las necesidades de las aplicaciones Web 2.0, no para consultar datos *ad hoc*.
- Se están elaborando herramientas para abordar esta necesidad.

Datos soportados por las bases de datos

Datos estructurados —

Datos que pueden almacenarse en tablas con filas y columnas.

Datos que tienen una clave relacional que puede asignarse fácilmente a campos prediseñados.

Datos que abarcan varias tablas y que pueden vincularse mediante operaciones de combinación.

Los datos estructurados representan solamente entre el 5% y el 10% de los datos.

Datos semiestructurados —

Los datos semiestructurados son información que no reside en una base de datos relacional, pero sí posee propiedades organizacionales que facilitan el análisis. Tras un tratamiento de la información, estos datos pueden almacenarse en una base de datos relacional.

XML y JSON son algunos ejemplos de formato.

Datos no estructurados —

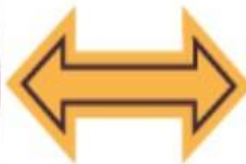
Los datos no estructurados suponen alrededor del 80% de los datos, y a menudo constan de texto y de contenido multimedia. Algunos ejemplos son los mensajes de correo electrónico, los documentos de procesamiento de textos, los vídeos, las fotos, los archivos de audio, las presentaciones, las páginas web y muchos otros documentos empresariales.

Características de ambos modelos

De ACID a BASE

ACID se centra en la
Coherencia y en la
Disponibilidad

- Atomicidad
- Coherencia
- Aislamiento
- Durabilidad



BASE se centra en la
tolerancia a la fragmentación
y en la disponibilidad

- Básicamente disponible
- Estado flexible
(El estado del sistema puede
variar con el tiempo)
- Consistencia eventual
(Propagación asíncrona)