

UNIDAD 4

Bases de datos NoSQL

Bases de Datos NoSQL

Contenido de la Unidad

- **Bases de datos NoSQL - Conceptos**
- Contraste entre los conceptos ACID, CAP y BASE
- Historia y evolución
- Tipos de Bases de Datos NoSQL (key-value, column-family, graph)
- Laboratorio Base de Datos NoSQL
- Resumen

Bases de Datos NoSQL

Conceptos

¿Qué son las bases de datos NoSQL (no Relacionales)?

Son muchas las aplicaciones web que utilizan algún tipo de bases de datos para funcionar.

Hasta ahora estábamos acostumbrados a utilizar bases de datos **SQL** como son **MySQL, Oracle** o **MS SQL**.

Desde hace ya algún tiempo han aparecido otras que reciben el nombre de **NoSQL (Not only SQL – No sólo SQL)** y que han llegado con la intención de hacer frente a las bases relacionales utilizadas por la mayoría de los usuarios.

Bases de Datos NoSQL

Conceptos

¿Qué son las bases de datos NoSQL (no Relacionales)?

Algunas Particularidades:

- Diseñadas específicamente para modelos de datos específicos.
- Poseen esquemas flexibles para crear aplicaciones modernas.
- Fáciles de desarrollar, por su funcionalidad y el rendimiento a escala.

Bases de Datos NoSQL

Conceptos

¿Cómo funciona una base de datos NoSQL?

Las bases de datos NoSQL utilizan una variedad de modelos de datos para acceder y administrar datos.

Estos tipos de bases de datos están optimizados específicamente para aplicaciones que requieren **grandes volúmenes de datos, baja latencia y modelos de datos flexibles**, lo que se logra mediante la flexibilización de algunas de las restricciones de coherencia de datos en otras bases de datos.

Bases de Datos NoSQL

Conceptos

¿Por qué debería utilizar una base de datos NoSQL?

Las bases de datos **NoSQL** se adaptan perfectamente a muchas aplicaciones modernas, como dispositivos móviles, web y juegos, que requieren bases de datos flexibles, escalables, de alto rendimiento y altamente funcionales para proporcionar excelentes experiencias de usuario.

Bases de Datos NoSQL

Conceptos

¿Por qué debería utilizar una base de datos NoSQL?

- **Flexibilidad:** Ofrecen esquemas flexibles que permiten un desarrollo más rápido y más iterativo. El modelo de datos flexible hace que las bases de datos NoSQL sean ideales para datos semiestructurados y no estructurados.
- **Escalabilidad:** Están diseñadas para escalar usando clústeres distribuidos de hardware en lugar de escalar añadiendo servidores caros y sólidos.
- **Alto rendimiento:** Están optimizadas para modelos de datos específicos y patrones de acceso que permiten un mayor rendimiento que el intento de lograr una funcionalidad similar con bases de datos relacionales.
- **Altamente funcional:** Proporcionan API altamente funcionales y tipos de datos que están diseñados específicamente para cada uno de sus respectivos modelos de datos.

Bases de Datos NoSQL

Conceptos

Ventajas de las bases de datos NoSQL

La forma de almacenar la información ofrece ciertas ventajas sobre los *modelos relacionales*, las ventajas más significativas son:

- Se ejecutan en máquinas con pocos recursos.
- Escalabilidad Horizontal.
- Manejo de gran cantidad de datos.
- No se generan cuellos de botella.

Bases de Datos NoSQL

Conceptos

Principales diferencias con las bases de datos Relacionales

Algunas de las diferencias que podemos identificar entre las bases de datos NoSQL y las bases de datos Relacionales:

- **No utilizan el lenguaje SQL para sus consultas:** Evitan utilizar este lenguaje o lo utilizan como lenguaje de apoyo.
- **No utilizan estructuras fijas como tablas para el almacenamiento de datos:** Permiten hacer uso de otros tipos de modelos de almacenamiento de información como sistemas de clave-Valor, objetos o gráficos.

Bases de Datos NoSQL

Conceptos

Principales diferencias con las bases de datos Relacionales

Algunas de las diferencias que podemos identificar entre las bases de datos NoSQL y las bases de datos Relacionales:

- **No suelen permitir Operaciones JOIN:** Al disponer de un volumen de datos tan extremadamente grande suele resultar deseable evitar los JOIN. Las soluciones más directas consisten en desnormalizar los datos, o bien realizar el JOIN mediante software, en la capa de aplicación.
- **Arquitectura distribuida:** Las bases de datos relacionales suelen estar centralizadas en una única máquina o bien en una estructura máster-esclavo, sin embargo en los casos NoSQL la información puede estar compartida en varias máquinas mediante mecanismos de tablas Hash distribuidas.

Bases de Datos NoSQL

Bases de datos NoSQL - Conceptos

Dudas / Preguntas



Bases de Datos NoSQL

Contenido de la Unidad

- Bases de datos NoSQL - Conceptos
- **Contraste entre los conceptos ACID, CAP y BASE**
- Historia y evolución
- Tipos de Bases de Datos NoSQL (key-value, column-family, graph)
- Laboratorio Base de Datos NoSQL
- Resumen

Bases de Datos NoSQL

Transacciones

CAP y Sistemas distribuidos

El teorema de CAP dice que en un sistema distribuido con datos compartidos no se puede tener Consistencia, Disponibilidad y Tolerancia a Particiones.

Por lo que tenemos que optar por dos de estas tres características, de acuerdo a las necesidades puntuales de nuestro sistema distribuido.

Bases de Datos NoSQL

Transacciones

CAP y Sistemas distribuidos

Debemos optar por:

- **Consistencia:** Todos los nodos en la red leen la información más reciente de la red. Si no pueden acceder a la información más reciente, error de consistencia.
- **Disponibilidad:** Todos los nodos en la red pueden leer la información que hay disponible (no necesariamente la más reciente) sin errores de consistencia.
- **Tolerancia a particiones:** El sistema continúa funcionando a pesar que la comunicación entre los nodos se pierda o sea lenta.

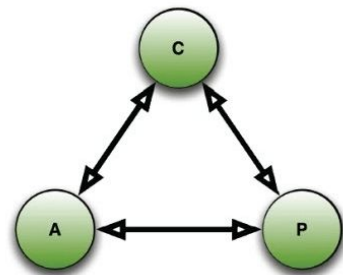
Bases de Datos NoSQL

Transacciones

CAP y Sistemas distribuidos (cont.)

Aunque las características mencionadas anteriormente son agradables el teorema de CAP dice que solo es posible aplicar dos de las tres características, siempre debemos "sacrificar" una:

- Si apuntamos a la **Consistencia** se debe sacrificar la disponibilidad. Es necesario que los nodos en el sistema se "pongan al día" con la información más reciente antes de hacer cualquier cosa. Este enfoque se denomina "**CAP-Consistent**".



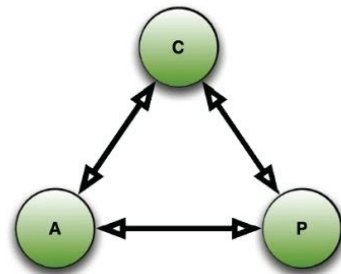
- CA: Consistencia y Disponibilidad.
- CP: Consistencia y tolerancia a particiones.
- AP: Disponibilidad y tolerancia a particiones.

Bases de Datos NoSQL

Transacciones

CAP y Sistemas distribuidos (cont.)

- Si apuntamos a **Consistencia** y **Disponibilidad**, el sistema no será distribuido, estará alojado en un solo nodo.
- Si apuntamos a la **Disponibilidad**, se debe sacrificar la consistencia corriendo el riesgo de que algunos servicios no usen la información actualizada para hacer sus funciones cuando quedan incomunicados con otros nodos. Este enfoque se denomina **"CAP-Available"**.



- CA: Consistencia y Disponibilidad.
- CP: Consistencia y tolerancia a particiones.
- AP: Disponibilidad y tolerancia a particiones.

Bases de Datos NoSQL

Transacciones

Sistemas ACID (Atomicity, Consistency, Isolation, Durability)

Las transacciones ACID tienen las siguientes características:

- **Atomicidad:** Evita que la transacción quede “a mitad de camino”, las operaciones se aplican de principio a fin si todo sale bien (commit) o en caso de error, la base de datos queda como si nada hubiera pasado (rollback).
- **Consistencia:** Asegura que la información ingrese de acuerdo a una serie de validaciones como revisión de campos únicos, llaves foráneas, condiciones entre otros mecanismos.

Bases de Datos NoSQL

Transacciones

Sistemas ACID (Atomicity, Consistency, Isolation, Durability)

Las transacciones ACID tienen las siguientes características:

- **Aislamiento:** Procura que la transacción se ejecute sin la interferencia de otras transacciones, existen varios niveles de Aislamiento.
- **Durabilidad o persistencia:** Protege la información ante fallos del software de base de datos.

Bases de Datos NoSQL

Transacciones

Sistemas Base

A diferencia de los sistemas **ACID**, los sistemas **BASE** ganan disponibilidad perdiendo parte de la consistencia, otorgando un estado de los datos más flexible.

Características:

- **Disponibilidad.**
- **Flexibilidad.**
- **Eventual Consistencia** (puede perderla en algún momento).

Bases de Datos NoSQL

Contraste entre los conceptos ACID, CAP y BASE

Dudas / Preguntas



Bases de Datos NoSQL

Contenido de la Unidad

- Bases de datos NoSQL - Conceptos
- Contraste entre los conceptos ACID, CAP y BASE
- **Historia y evolución**
- Tipos de Bases de Datos NoSQL (key-value, column-family, graph)
- Laboratorio Base de Datos NoSQL
- Resumen

Bases de Datos NoSQL

Historia y Evolución

- El acrónimo **NoSQL** corresponde a la frase "**Not only SQL**", aunque originalmente era "**No SQL**". Este término hace referencia a cualquier sistema de gestión de datos que no sigue el modelo de las RDBMS y que no aplica el lenguaje SQL para realizar las consultas.
- Este tipo de bases de datos comenzaron a originarse en el año **1998** y no fueron creadas por una empresa específica, sino que fueron concebidas por distintas empresas y grupos independientes que buscaban soluciones específicas a sus problemas. El auge de este tipo de base de datos se produjo por la necesidad de cubrir las carencias de las RDBMS.

Bases de Datos NoSQL

Historia y Evolución

- El término **NoSQL** aparece con la llegada de la web 2.0 ya que hasta ese momento sólo subían contenido a la red aquellas empresas que tenían un portal, pero con la llegada de aplicaciones como **Facebook**, **Twitter** o **Youtube**, cualquier usuario podía subir contenido, provocando así un crecimiento exponencial de los datos.
- Comienzan a aparecer problemas de gestión de toda esa información almacenada en bases de datos relacionales. En un principio, para solucionar estos problemas de accesibilidad, las empresas optaron por utilizar un mayor número de máquinas pero pronto se dieron cuenta de que esto no solucionaba el problema, además de ser una solución muy cara. Otra solución era la creación de sistemas pensados para un uso específico que con el paso del tiempo han dado lugar a soluciones robustas, apareciendo así el movimiento NoSQL

Bases de Datos NoSQL

Historia y Evolución

Algunas bases de datos NoSQL:



Bases de Datos NoSQL

Contenido de la Unidad

- Bases de datos NoSQL - Conceptos
- Contraste entre los conceptos ACID, CAP y BASE
- Historia y evolución
- **Tipos de Bases de Datos NoSQL (key-value, column-family, graph)**
- Laboratorio Base de Datos NoSQL
- Resumen

Bases de Datos NoSQL

Tipos de Bases de Datos NoSQL

- **Key-value (Redis, Aerospike, Riak, etc.):** almacenan los datos en una estructura “clave-valor”, conocida como “diccionario” y cuyos valores no requieren un esquema fijo. En este tipo de base de datos no existe el concepto de relaciones, están pensadas para almacenar información básica y que pueda ser consultada de forma muy rápida.
- **Document oriented (MongoDB, eXist, CouchDB, etc.):** almacenan datos de tipo documento. Cabe destacar que no debe confundirse el término *documento* con ficheros Excel, Word, etc. El término *documento* se refiere a una estructura de datos (clave, valor) representada en formato XML, JSON o BSON, que puede llegar a tener una estructura de datos anidados.

Bases de Datos NoSQL

Tipos de Bases de Datos NoSQL

- **Graphs (Neo4j, OrientDB, GraphDB):** almacenan los datos en formato de grafos, son excelentes para la gestión de datos interconectados y superan el concepto de “relación” de las RDBMS. Su principal ventaja es que, a diferencia de las RDBMS, los grafos persisten tal cual (en modo binario) en la base de datos, con sus propiedades y relaciones, con lo cual las consultas son mucho más rápidas comparadas con RDBMS.
- **Column based (Cassandra, ScyllaDB, etc.):** almacenan los datos en forma de columnas (familias), a diferencia de las RDBMS, que almacenan los datos en forma de fila. Es decir, cada registro (fila) puede contener una o más columnas y cada columna puede contener distintas estructuras de datos y además de forma versionada. Este tipo de bases de datos están pensadas para la consulta de datos de tipo histórico (logs, eventos, etc.), con el objetivo de obtener información agregada.

Bases de Datos NoSQL

Tipos de Bases de Datos NoSQL

Key Value / Tuple Store:

En un sistema clave-valor existen contenedores, también se les llama cabinets. En cada contenedor es posible tener datos de la misma naturaleza (por ejemplo productos, pedidos, clientes, etc.) o totalmente diferentes (puede haber un contenedor por cliente), todo depende de los desarrolladores de la aplicación.

Ejemplo de Contenedor:

[users.cab]

john=myp4ssw0rd
sony=4f5h0r8vn0

[users_data.cab]

john_name=Jonh Clax
john_email=jonh@yahoo.com
john_country=Canada
john_birthdate=04/05/1960
sony_name=Sony Sand
sony_email=sony@yahoo.com
sony_country=England
sony_birthdate=04/05/1948

Bases de Datos NoSQL

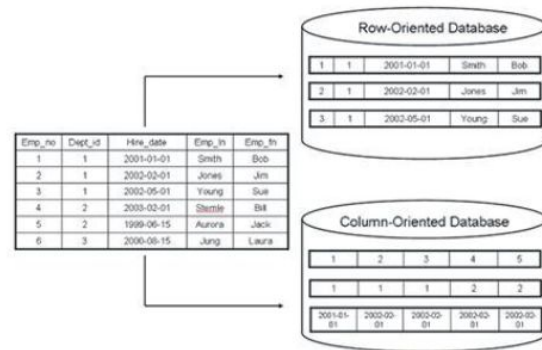
Tipos de Bases de Datos NoSQL

Wide Column Store / Column Families

A primera vista son muy similares a las bases de datos relacionales, pero en realidad son muy diferentes:

- Sistema Relacional => Almacenamiento por Filas
- Sistema Orientado a Columnas => Almacenamiento en columnas

Poseen tablas de datos donde las columnas de los valores de datos representan el almacenamiento estructural, los datos son almacenados como “Secciones” de las columnas en lugar de filas de datos:



Bases de Datos NoSQL

Tipos de Bases de Datos NoSQL

Document Store

Diseñada para gestionar información orientada a documentos o datos semi-estructurados. Constituye una de las principales categorías de las llamadas bases de datos NoSQL.

- Sistema Relacional => Tablas.
- Document Store => Diseñados en torno a la definición abstracta de documentos.

Posee enormes ventajas en cuanto a Rendimiento y Escalabilidad. Ejemplo:

```
Atlas atlas-v0uuxz-shard-0 [primary] MoviesDB> db.movies.find().pretty()
[
  {
    _id: 'Titanic_2001',
    name: 'Titanic',
    year: 2001,
    Genre: [ 'romantic', 'drama' ],
    rating: 7,
    Description: 'Película romantica sobre la historia del hundimiento del titanic',
    country: 'Argentina'
  },
  {
    _id: 'Ronin_2010',
    name: 'Ronin',
    year: 2010,
    Genre: [ 'Action' ],
    rating: 8,
    Description: 'Película de accion de espionaje',
    country: 'Argentina'
  },
]
```

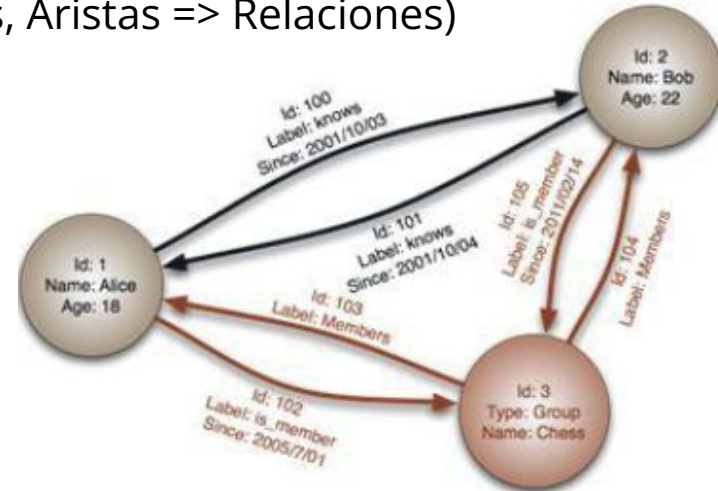
Bases de Datos NoSQL

Tipos de Bases de Datos NoSQL

Graph Database

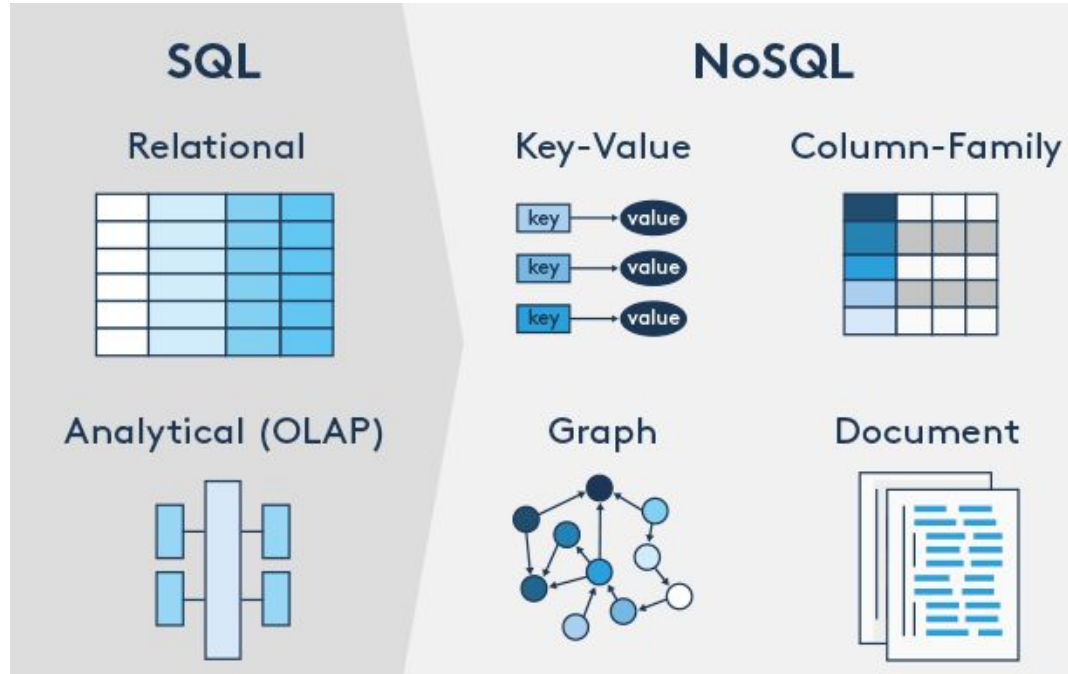
Representan la información como nodos de un grafo y sus relaciones con las aristas del mismo, de manera que se pueda usar teoría de grafos para recorrer la base de datos y describir atributos (nodos => Entidades, Aristas => Relaciones)

Debe estar absolutamente **normalizada**



Bases de Datos NoSQL

Tipos de Bases de Datos NoSQL



Bases de Datos NoSQL

Tipos de Bases de Datos NoSQL

Algunas bases de datos NoSQL y sus tipos:

- **MongoDB** → Orientada a Documentos
- **Cassandra** → Wide Column Store
- **Riak** → Key Value
- **Redis** → Key Value (in-Memory)
- **neo4j** → Graph Database



Bases de Datos NoSQL

Tipos de bases de datos NoSQL

Dudas / Preguntas



Bases de Datos NoSQL

Contenido de la Unidad

- Bases de datos NoSQL - Conceptos
- Contraste entre los conceptos ACID, CAP y BASE
- Historia y evolución
- Tipos de Bases de Datos NoSQL (key-value, column-family, graph)
- **Laboratorio Base de Datos NoSQL**
- Resumen

Bases de Datos NoSQL

Contenido de la Unidad

- Bases de datos NoSQL - Conceptos
- Contraste entre los conceptos ACID, CAP y BASE
- Historia y evolución
- Tipos de Bases de Datos NoSQL (key-value, column-family, graph)
- Laboratorio Base de Datos NoSQL
- **Resumen**

Bases de Datos NoSQL

Resumen de la Unidad

Durante esta unidad hemos aprendido:

- Introducción y conceptos de las bases de datos NoSQL
- Definición y contraste entre los conceptos de Diseño de los distintos tipos de bases de datos
- Introducción y desarrollo del marco histórico y evolución de las bases de datos NoSQL
- Recorrido por los distintos tipos de bases de datos NoSQL y algunos conceptos que las diferencian.

Bases de Datos NoSQL

Bibliografía Recomendada



- Making Sense of NoSQL: A Guide for managers and the rest of us - Dan McCreary, Ann Kelly
- NoSQL and SQL Data Modeling: Bringing Together Data, Semantics, and Software - Ted Hills
- Seven Databases in Seven Weeks: A Guide to Modern Databases and the NOSQL Movement - Luc Perkins, Eric Redmond, Jim Wilson
- NoSQL for Mere Mortals - Dan Sullivan Sullivan
- NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence - Pramod Sadalage, Martin Fowler

