



Bases de Datos NoSQL (Not Only SQL)

Lucas Castronuovo y Sebastián Roca



¿Qué son las BD NoSQL?

NoSql es utilizado para describir un subconjunto de bases de datos que difiere en distintos modos de bases de datos tradicionales. Son útiles cuando se necesita acceder y analizar gran cantidad de datos no estructurados o datos que se almacenen de forma remota en varios servidores virtuales o en la nube. Resuelve problemas de almacenamiento masivo y alto desempeño. No usa SQL como principal lenguaje. De ahí viene su nombre NotOnly SQL.

Breve Historia

El término NoSQL fue acuñado en 1998 por Carlo Trozzi, Not Only SQL.

Erick Evans retomó el término en 2009, la diferencia es que Erick Evans sugiere referirse a esta familia de base de datos como big data.

Definiciones



*"NoSQL abarca una amplia variedad de **diferentes tecnologías** de bases de datos que se desarrollaron en respuesta a las **demandas** presentadas en la construcción de aplicaciones **modernas**"*

MongoDB

*"Un ambiente de base de datos NoSQL es, simplemente, un **sistema** de base de datos **no***

***Relacional y ampliamente distribuido** que permite una organización **rápida** y ad hoc y análisis de tipos de datos muy **dispares**. Las bases de datos NoSQL a veces se denominan bases de datos en la nube, bases de datos no relacionales, bases de datos Big Data y una miríada de otros términos y se desarrollaron en respuesta al gran volumen de datos generados, almacenados y analizados por usuarios modernos (datos generados por el usuario) y sus aplicaciones (datos generados por máquina)".*

Datastax

*"La intención original ha sido desarrollar las modernas bases de datos de la **web**. El movimiento comenzó a principios de 2009 y está creciendo rápidamente. A menudo más características se aplican, tales como: libertad de esquemas, fácil soporte de replicación, API simple, consistencia eventual (no ACID), una gran cantidad de datos y mucho más. Así que el término engañoso "nosql" (la comunidad ahora lo traduce principalmente como "no sólo sal") debe ser visto como un alias a algo parecido a la definición de arriba".*

NoSql-database.org

Arquitecturas NoSQL

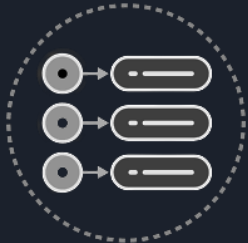
- Generalidades
 - Consistencia débil o transacciones restringidas
 - Arquitectura Distribuida
 - Estructura de datos sencillas



Arquitecturas NoSQL

BD de Clave-Valor

- Estructura de datos sencilla
- Formada por tablas donde se almacenan filas o elementos



BD de Documentos

- Se almacena en forma de documentos
- Conjunto de pares clave-valor
- Se puede contener otros documentos



BD de Columnas Extendidas

- BD ordenada en un mapa multidimensional
- Definida por
 - Rowkey
 - Columnkey
 - Timestamp

Nombre	Dirección	Ciudad
Maria	Calle Mallorca	Madrid
José	Avenida Castellana	Sevilla
Luisa	Calle Aragón	Barcelona

Base de datos orientada a filas

Nombre	Maria	José	Luisa
Dirección	Calle Mallorca	Avenida Castellana	Calle Aragón
Ciudad	Madrid	Sevilla	Barcelona

Base de datos orientada a columnas

BD de Grafos

- Basada en grafos
- Usado en motores de recomendación y aplicaciones geoespaciales



Big Data y su relación con NoSQL

Big Data se refiere a datos que son masivamente generados. Un tipo de tecnología que ha surgido para tratar de poner solución a muchos de los problemas de los que adolecen los sistemas de almacenamiento tradicionales cuando intentan manejar este tipo de datos masivos. Son las tecnología que se conoce como NoSql.

Definición

Big Data es la ingente cantidad de información, en su mayor parte desestructurada, que hoy en día generamos toda la sociedad como consecuencia de nuestra actividad tanto en Internet como fuera de ella.

Definiciones



"Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of your database architectures. To gain value from this data, you must choose an alternative way to process it."

O'Really 2012

En el blog oficial de Microsoft Enterprise se puede leer:

"Big data is the term increasingly used to describe the process of applying serious computing power — the latest in machine learning and artificial intelligence — to seriously massive and often highly complex sets of information"

HowieT 2013

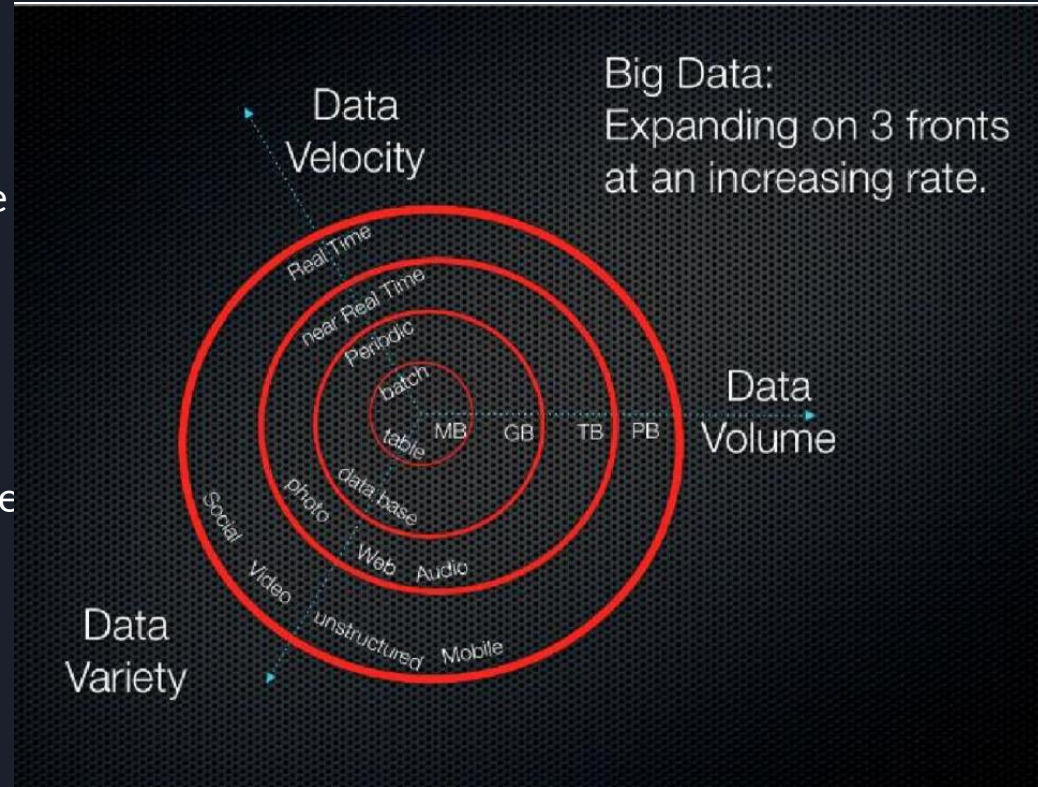
Adrian Mery, de la revista Teradata expone la siguiente definición: *"Big data exceeds the reach of commonly used hardware environments and software tools to capture, manage and process it within a tolerable elapsed time for its user population"*

Merv 2011

3V's : Volumen, Velocidad y Variedad

Volumen, Velocidad y Variedad son los principales factores que intentan mejorarse con las B.D. NoSQL.

En los últimos 10 años, el crecimiento del Volumen de Datos No Estructurados que se generan a gran velocidad llevó a que se desarrollen nuevos paradigmas de almacenamiento.



Funcionalidad

- Las bases de datos NoSQL son requeridas cuando se cumplen estos puntos:
 - Mayor Velocidad en que los datos se generan
 - Mayor Volumen de Datos
 - Necesidad de almacenar datos de formatos Variados
 - Ser ACID no es mandatorio
- Las bases de datos NoSQL se utilizan para:
 - Desarrollo de Redes Sociales
 - Desarrollo Web o Móvil
 - Big Data
- Las bases de datos NoSQL no se utilizan para
 - Sistemas consistentes, donde la posibilidad de error debe ser baja
 - Uso empresarial
 - Rama de negocios



Funcionalidad

BD de Clave-Valor

- Almacenamiento en diccionarios
- Cada registro se define como un conjunto de una clave que referencia a un valor

BD de Documentos

- Conformado por programas que almacenan, recuperan y gestionan datos de documento
- Diseño basado en una noción abstracta de un documento

BD de Columnas Extendidas

- Almacenamiento de datos en columnas en vez de filas
- Los datos se leen y se escriben de manera eficiente en el disco duro
- El almacenamiento de datos se da por orden de registro

BD de Grafos

- Se almacena la información como nodos de un grafo y sus respectivas relaciones con otros nodos
- Las relaciones pueden tener propiedades, definidas por pares clave-valor



Generalidades

Ventajas

La escalabilidad y su carácter descentralizado.

Soportan estructuras distribuidas.

Abiertas y flexibles.

Cambios de los esquemas sin tener que parar bases de datos.

Escalabilidad horizontal

Se pueden ejecutar en máquinas con pocos recursos.

Optimización de consultas en base de datos para grandes cantidades de datos.

Desventajas

No contemplan la atomicidad de las instrucciones y la integridad de los datos.

Soportan lo que se llama consistencia eventual.

Problemas de compatibilidad entre instrucciones SQL.

Falta de estandarización.

No hay soporte multiplataforma.

Ventajas y desventajas específicas

BD de Clave-Valor

- Ventajas: Modelo dinámico, Estructuras genérica, Altamente flexible, permite agregar atributos en una etapa posterior al diseño.
- Desventaja: No asegura la integridad del tipo de dato.

BD de Documentos

- Ventajas: La posibilidad de usar estructuras anidadas, soportar arreglos y entender diferentes tipos de datos.
- Desventaja: No asegura la integridad del tipo de dato.

BD de Columnas Extendidas

- Ventajas: Flexibilidad, puede agregar columnas de diferentes filas.
- Desventaja: No asegura la integridad del tipo de dato.

BD de Grafos

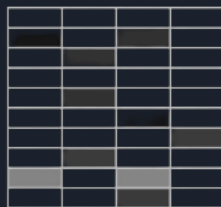
- Ventajas: las estructuras internas son eficientes para almacenar y buscar en forma de estructuras de grafos.

Comparación con otras Bases de Datos

Base de datos NoSQL y Base de datos Relacional

Una Base de Datos NoSQL se diferencia de una Base de datos Relacional teniendo en cuenta que:

- No permite joins.
- No garantizan ACID
- Indexable
- Colecciones de tamaño fijo
- Teorema CAP



Comparación con otras Bases de Datos

Base de datos NoSQL y Base de datos Orientada a Objetos

- Las bases de datos NoSQL, al principio, solían ser bases de datos orientadas a objetos
- La Popularidad de las bases de datos NoSQL influenció a los proveedores de las bases de datos orientada a objetos



Productos de Bases de Datos NoSQL

BD de Clave-Valor

- Redis
- Amazon DynamoDB
- Riak

BD de Documentos

- CouchDB
- Couchbase
- Marklogic
- MongoDB

BD de Columnas Extendidas

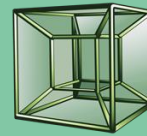
- Hypertable
- Accumulo
- Apache Cassandra
- Apache HBASE
- Amazon SimpleDB

BD de Grafos

- Neo4j
- InfiniteGraph



redis



HYPERTABLE^{INC}



¿Como se usan las BD NoSQL?

Dado la variedad de base de datos NoSQL, decidimos elegir Dynamodb como un ejemplo particular.

Tutorial de instalación y uso https://www.youtube.com/watch?v=liTlg_aweBk

