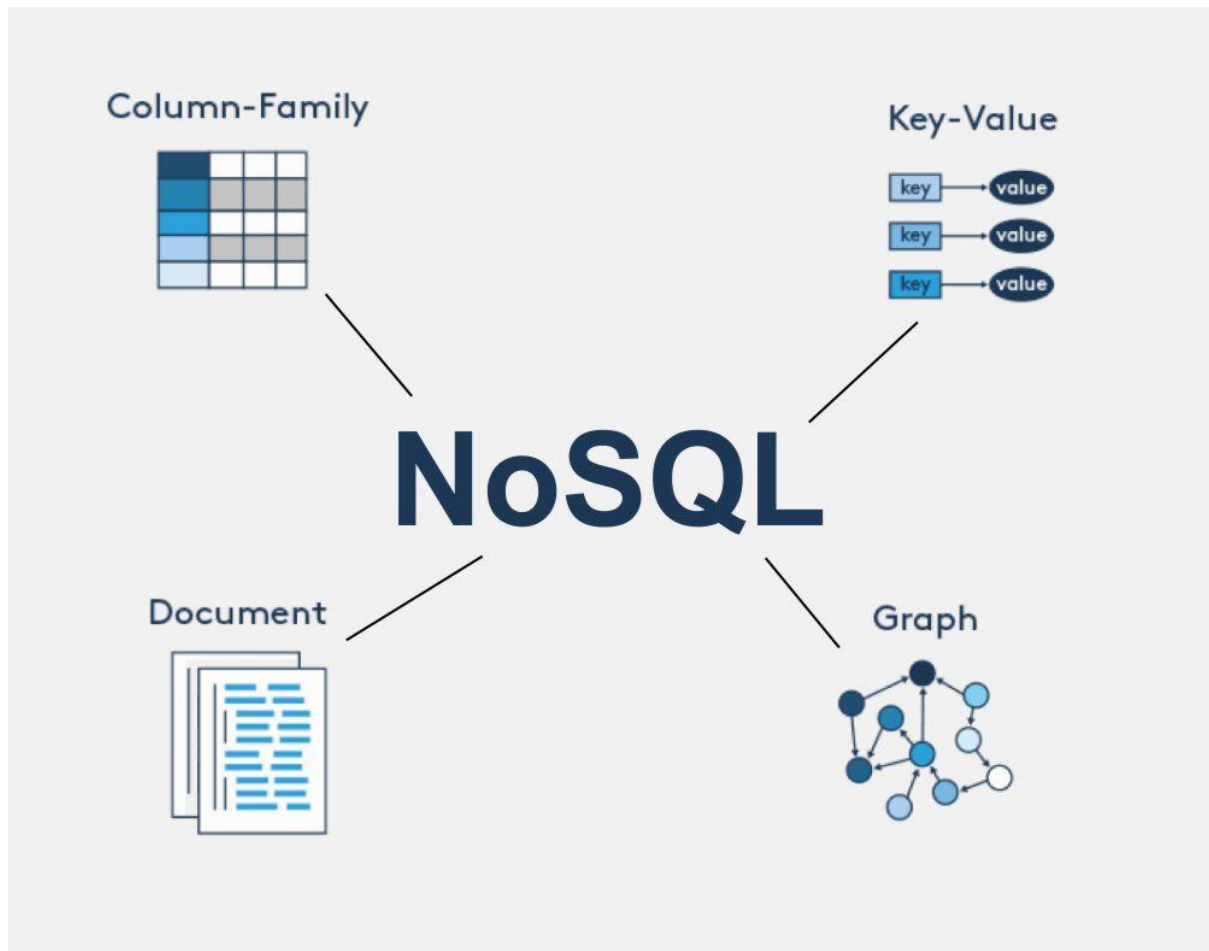


Base de datos NoSQL (Not Only SQL)

por: Lucas Castronuovo y Sebastián Roca



¿Qué son las bases de datos NoSQL?

Definición

El término NoSQL fue acuñado en 1998 por Carlo Trozzi y retomado en 2009 por Erick Evans, la diferencia es que Erick Evans sugiere referirse a esta familia de base de datos como big data.

Se generan a partir del uso de sistemas web y de respuesta inmediata, esto relacionado con que se consume un volumen de datos mayor así como la utilización de datos desde distintos usuarios y servicios.

Estas bases crecieron conjunto de las empresas líderes de servicios web como son: Google, Amazon, Facebook y Twitter, ya que estas tenían que enfrentarse a desafíos con el tratamiento de datos, a las cuales las tradicionales bases de datos no solucionaban

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.

Definición según

"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

Big data

En los últimos años, la cantidad de datos digitales que se generan en el mundo se ha multiplicado. Las redes sociales y el cada vez más fácil acceso a Internet del que disponemos las personas hacen que el volumen de tráfico y de datos que se generan sea cada vez mayor.

Con el surgimiento de las bases de datos relacionales las empresas encontraron el aliado perfecto para cubrir sus necesidades de almacenamiento, disponibilidad, copiado de seguridad y gestión de sus datos.

Pero debido a las tendencias actuales de uso de Internet, este tipo de sistemas han comenzado a experimentar dificultades técnicas, en algunos casos bloqueantes, que impiden el buen funcionamiento de los sistemas de algunas de las empresas más importantes de Internet. Este tipo de datos que son masivamente generados reciben un nombre: BigData, y un tipo de tecnología 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. Esta tecnología 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.

Definición según

"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"

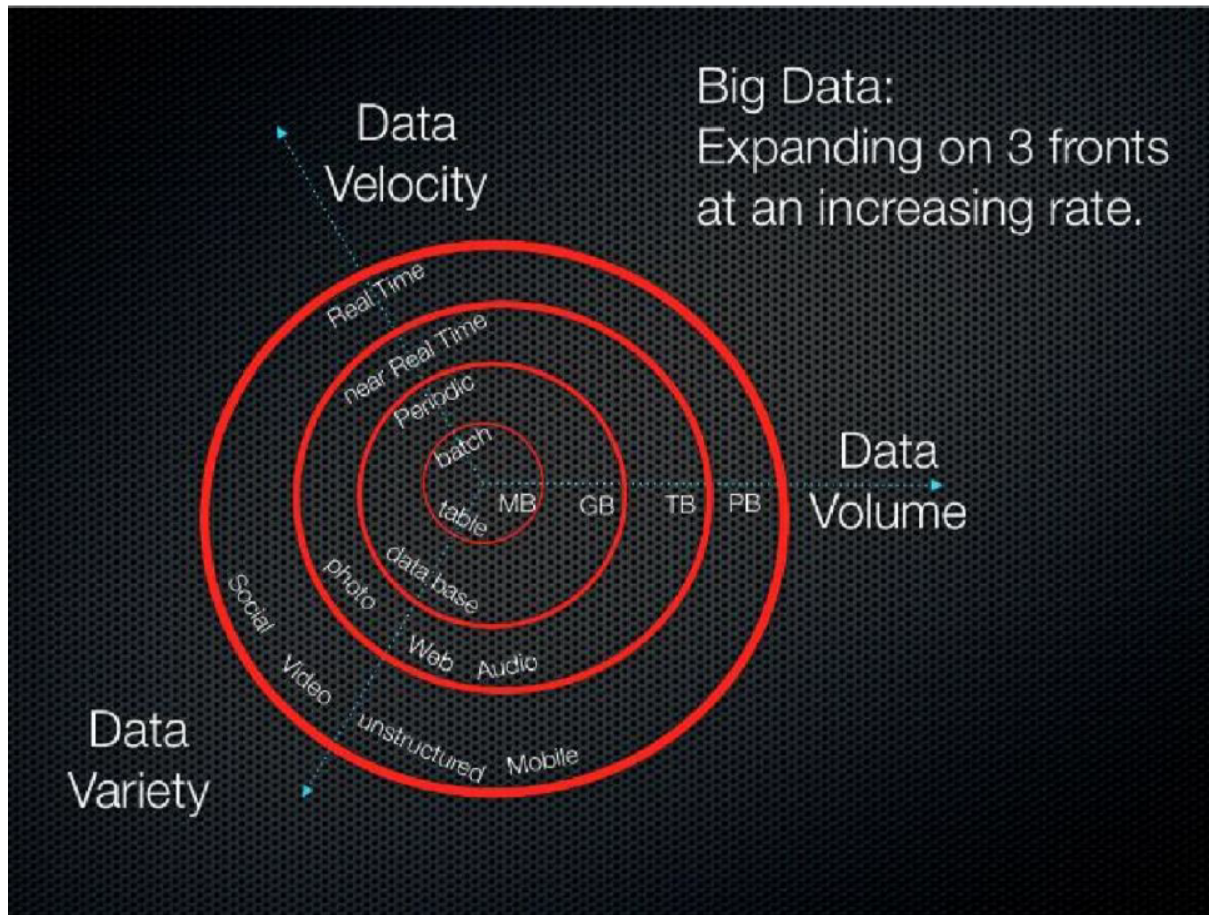
HowieT 2013

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

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. Se pueden destacar las 3V's:

- Volumen
- Velocidad
- Variedad



Arquitecturas NoSQL

Generalidades

Ofrecen garantías de consistencia débil, o transacciones restringidas a elementos de datos simples.

Emplean una arquitectura distribuida en donde los datos se guardan de modo redundante en distintos servidores. Esto es importante por que esto permite que las bases NoSQL puedan escalar horizontalmente y tener una copia completa en cada nodo.

Suelen ofrecer estructuras de datos sencillas como arrays asociativos o almacenes de pares clave-valor.

Bases de datos de Clave-Valor

Posee una estructura de datos sencilla. Está formada por tablas donde se almacenan filas o elementos. Estos últimos no tienen porque tener los mismos atributos ni la misma cantidad.

Bases de datos de Documentos

Almacena elementos en forma de documentos, codificados en formato JSON. Cada elemento es un conjunto de pares clave valor. Cada documento puede contener otros documentos.

Bases de Datos de Columnas Extendidas

Es una base de datos distribuida, persistente y ordenada en un mapa multidimensional. Una celda de esta base de datos viene definida por una clave primaria de fila (rowkey), por una clave de columna (columnkey) y una marca de tiempo (timestamp).

Bases de datos de Grafos

Es una base de datos basada en grafos, que son un conjunto de nodos y relaciones, usos comunes incluyen motores de recomendación y aplicaciones geoespaciales.

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
- Necesidad de Escalar en forma ilimitado
- Necesidad de Tiempos de Respuesta inmediatos
- Aplicaciones Web/Mobile 7x24 con alta disponibilidad
- Ser ACID no es mandatorio
- BASES DE DATOS diseñadas para SISTEMAS DISTRIBUIDOS

Comúnmente, las bases de datos NoSQL se utilizan para:

- Desarrollo de Redes Sociales
- Desarrollo Web o Móvil
- Big Data
- En aquellos sistemas donde las estructuras de los datos son variables
- Análisis de grandes cantidades de datos en modo lectura
- Captura y procesamiento de eventos

Comúnmente, las bases de datos NoSQL no se utilizan para:

- En aquellos sistemas donde los datos deben ser consistentes sin dar posibilidad de error
- Uso empresarial
- Rama de negocios

Bases de Datos de Clave-Valor: El almacenamiento de datos se da desde un método sencillo de clave-valor. Estos son almacenados en diccionarios. Los diccionarios contienen

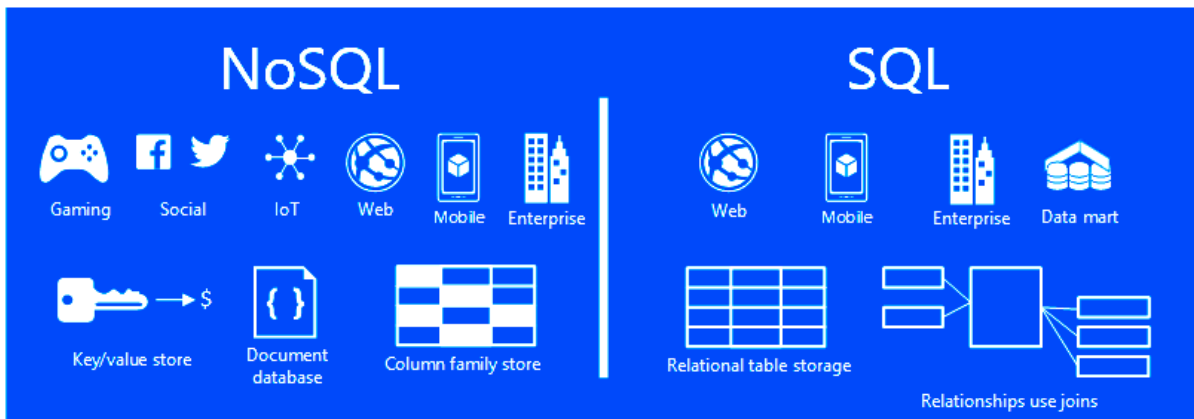
una colección de objetos referenciados con una clave que identifica los datos asociados de manera unívoca y que se utiliza para encontrar los datos en la base de datos. Esto determina que cada registro se define como un conjunto de una clave que referencia un valor.

Bases de Datos de Documentos: Está conformado por programas que almacenan, recuperan y gestionan datos de documentos, por lo que su diseño se determina desde una noción abstracta de documento. Un documento es una forma de organizar y almacenar la data como un set de pares de campo-valor, siendo el campo un identificador único para un datapoint y el valor es la data relacionada a un determinado identificador.

Bases de Datos de Columnas Extendidas: se encarga de almacenar los datos en columnas en vez de filas. Esto busca lograr que los datos puedan escribirse y leerse de manera eficiente en el almacenamiento del disco duro, disminuyendo el tiempo en que se puedan obtener las consultas requeridas. Por lo tanto, todos los valores de una columna “..están físicamente juntos..”. El almacenamiento de datos se da por orden de registro. Al fin y al cabo, permitirá que se accedan a elementos de datos individuales en columnas que forman un grupo. Las bases de datos columnares están formadas por un Keyspace que estructura la base de datos columnar, formando estas familias de columnas. En cada fila se encuentra una Row Key que es la clave única e identificador de la fila, la Columna donde dentro de ella estará el nombre, un valor y una marca de tiempo (Es el timestamp, indica la fecha y hora en la cual se insertan los datos).

Bases de Datos de Grafos: Se almacena la información como nodos de un grafo y sus respectivas relaciones con otros nodos, permitiendo así aplicar la teoría de grafos para recorrer la base de datos. Las relaciones pueden tener propiedades, que se definen por pares clave-valor, y añaden información adicional para poder aplicar algoritmos y realizar consultas. Cada nodo consta de un grado que indica el número de aristas que tiene, a su vez un grafo puede ser dirigido o no dirigido, dependiendo de si las aristas tienen nodos origen y nodos destino.

Comparación con otras bases de datos



Relacional

Una Base NoSQL es una base orientada a Documentos, es decir no es relacional. No permiten Joins. Pero al mismo tiempo se puede generar uno. No intentan garantizar ACID (ATOMICIDAD, CONSISTENCIA, AISLAMIENTO, DURABILIDAD). Pero soporta Escalabilidad Horizontal, es de una escalabilidad muy sencilla. Utilizando ampliamente los usos de memoria principales de los equipos para soportar los altos volúmenes de información.

Por otro lado, es Indexable: O sea, por ejemplo, nos va a permitir utilizar índices Geoespaciales como para hacer búsquedas por cercanía. Permiten colecciones de tamaño fijo, que van a facilitarnos con una velocidad más alta en acceso y modificación de datos. También se puede destacar el Teorema de CAP, donde se resalta que un sistema computacional distribuido no puede ofrecer Consistencia, Disponibilidad (Availability) y Tolerancia a la partición (Partition) simultáneamente. Las bases de datos Relacionales se dedican a respetar las propiedades ACID, mientras que las bases de datos NoSQL rescinden ACID para dar mayor prioridad a la Disponibilidad y Tolerancia a las particiones.

Orientada a Objetos

Las bases de datos NoSQL, al principio, solían ser bases de datos orientadas a objetos. Cuando las bases de datos NoSQL fueron más populares, los encargados de proveer bases de datos orientadas a objetos vieron la oportunidad de formar parte de este movimiento que se generaba.

Ventajas y Desventajas

Generalidades

Dado que no hay un único tipo de base de datos NoSQL, dependiendo de cual hablemos podemos encontrarnos con distintas ventajas y desventajas, pero en líneas generales estas cuentan con las siguientes ventajas:

- La escalabilidad y su carácter descentralizado. Soportan estructuras distribuidas.
- Suelen ser bases de datos mucho más abiertas y flexibles. Permiten adaptarse a necesidades de proyectos mucho más fácilmente que los modelos de Entidad Relación.
- Se pueden hacer cambios de los esquemas sin tener que parar bases de datos.
- Escalabilidad horizontal: son capaces de crecer en número de máquinas, en lugar de tener que residir en grandes máquinas.
- Se pueden ejecutar en máquinas con pocos recursos.
- Optimización de consultas en base de datos para grandes cantidades de datos.

Respecto a las desventajas de las bases de datos NoSQL en general, encontramos las siguientes:

- No todas las bases de datos NoSQL 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. Las nuevas bases de datos utilizan sus propias características en el lenguaje de consulta y no son 100% compatibles con el SQL de las bases de datos relacionales. El soporte a problemas con las queries de trabajo en una base de datos NoSQL es más complicado.
- Falta de estandarización. Hay muchas bases de datos NoSQL y aún no hay un estándar como si lo hay en las bases de datos relacionales.
- Soporte multiplataforma. Aún quedan muchas mejoras en algunos sistemas para que soporten sistemas operativos que no sean Linux.(Esto es algo que se deberá tomar en cuenta al elegir el motor)

Bases de datos de Clave-Valor

Este tipo se caracteriza por almacenar los datos en pares clave-valor, lo que le da ventajas tales como: Modelo dinámico, Estructuras genérica, Altamente flexible, permite agregar atributos en una etapa posterior al diseño. También como resultado de esto se crea la desventaja de no asegurar la integridad del tipo de dato.

Bases de datos de Documentos

Este tipo se caracteriza por almacenar documentos en la parte value del almacenamiento key-value, especificado en json. Las ventajas de este tipo son la posibilidad de usar estructuras anidadas, soportar arreglos y entender diferentes tipos de datos. Al igual que las claves-valor, tienen el problema de no asegurar la integridad del tipo de dato.

Bases de Datos de Columnas Extendidas

Este tipo se caracteriza por como es la unidad básica de almacenamiento, una simple columna. Esto le da flexibilidad para agregar columnas de diferentes filas. Aunque al igual que las anteriores no asegura la integridad del tipo de dato.

Bases de datos de Grafos

Este tipo se caracteriza por varias cosas como tener nodos relacionados entre sí. Cada uno tiene propiedades y los arcos tienen tipos y pueden tener múltiples propiedades. Por esto las estructuras internas son eficientes para almacenar y buscar en forma de estructuras de grafos.

Productos de Bases de Datos NoSQL

Base de dato - Empresas que lo usan - lanzamiento

En Base de Datos de Documentos:

- CouchDB (Apache Software Foundation) - Ubuntu, BBC, Credit Suisse, Meebo - abril 2005
- Couchbase (Couchbase, Inc) - Zynga, Cisco, Adobe, Mc Graw Hill, Mozilla foundation, Honda, Aol, LG - agosto 2010
- MarkLogic (Marklogic) - BBC, Boeing, Bancos líderes mundiales
- MongoDB (MongoDB Inc) - Foursquare, LinkedIn, Cisco, eBay, Bosch - 2007

En Base de Datos de Grafos:

- Neo4j (Neo Technology) - Walmart, eBay, UBS, Nomura, Cisco, HP, Telenor - 2007
- InfiniteGraph (Objectivity, Inc.) - Objectivity inc. - mayo 2021

En Base de Datos de Clave-Valor:

- Redis (Salvatore Sanfilippo) - Twitter, GitHub, Pinterest, Snapchat, StackOverflow, Trello - abril 2009
- Amazon DynamoDB (Amazon) - Samsung, Toyota, Capital One, Dropbox, Zoom, Disney - 2012
- Riak (Basho Technologies) - AT&T, Comcast, Github, Best buy, UK National Health Service, Riot Games - Agosto 2009

En Base de Datos Columnas Extendidas:

- Hypertable(Zvents inc.) - Baidu, GlusterFS, CloudStore - 2008
- Accumulo(NSA) - NSA(National Security Agency), DoD(Department of Defense) - 2012
- Apache Cassandra(Apache software foundation) - Twitter, Google, Amazon, Facebook, Instagram, Netflix, Apple - 2008
- Apache HBASE(Apache software foundation) - Facebook, Twitter, Yahoo, Adobe - 2006
- Amazon SimpleDB(Amazon) - Amazon, Google, Azure - diciembre 2007

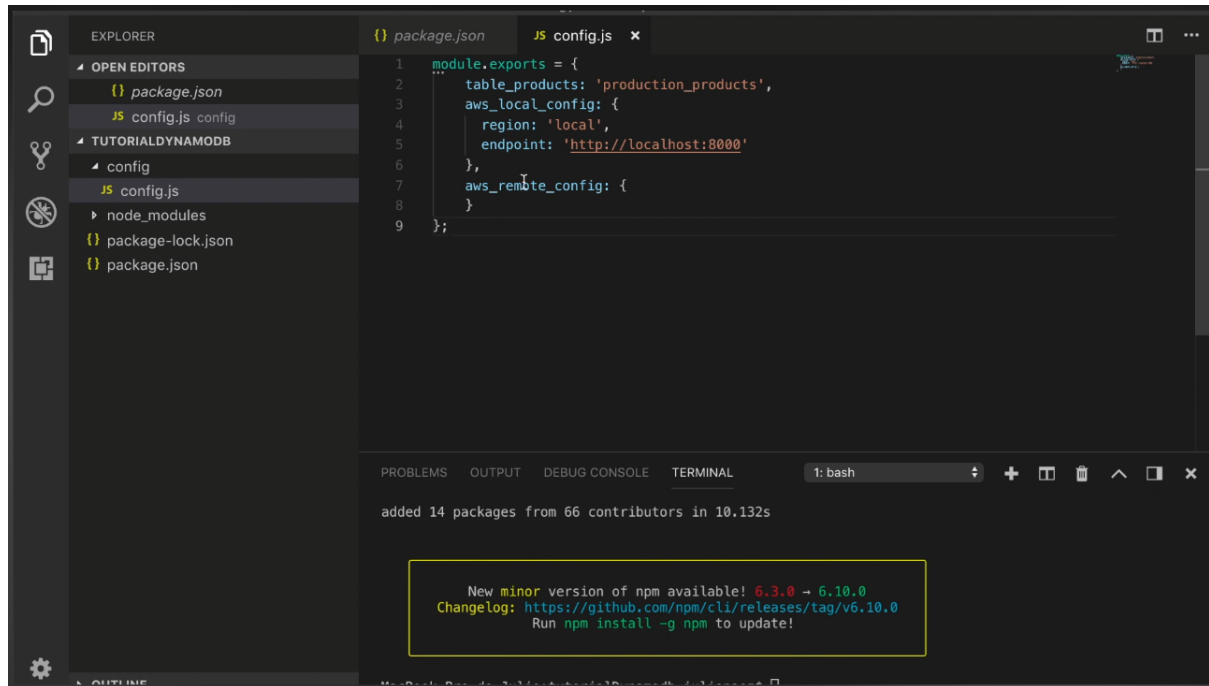
Como se usa una base de datos NoSQL?

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

Tutorial de instalación y uso

Link: https://www.youtube.com/watch?v=liTlg_aweBk

Configuración

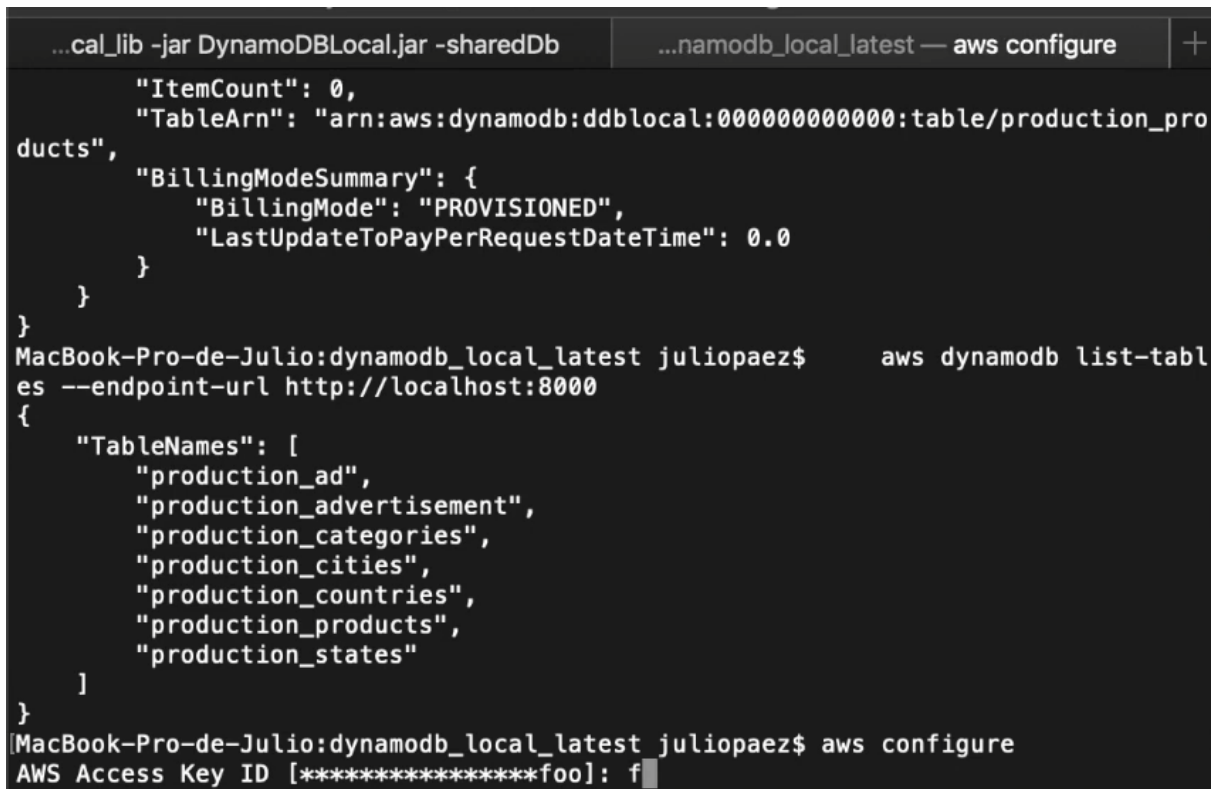
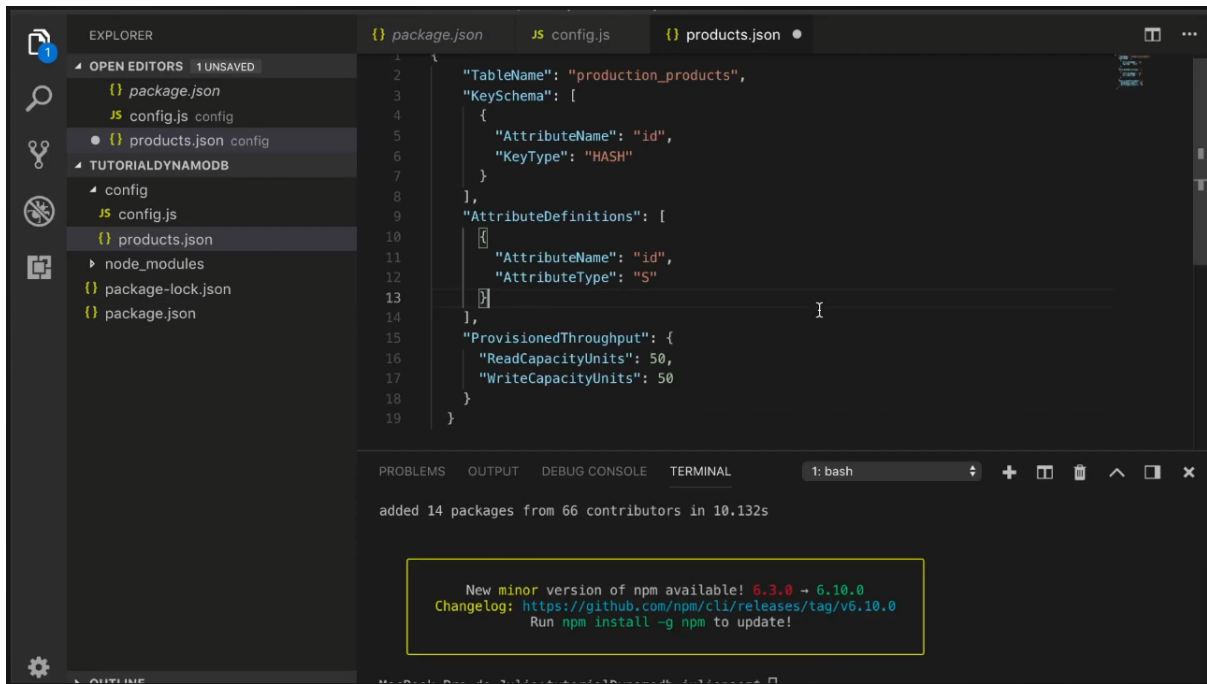


```
1 module.exports = {
2   table_products: 'production_products',
3   aws_local_config: {
4     region: 'local',
5     endpoint: 'http://localhost:8000'
6   },
7   aws_remote_config: {
8   }
9 };
```

added 14 packages from 66 contributors in 10.132s

New minor version of npm available! 6.3.0 → 6.10.0
Changelog: <https://github.com/npm/cli/releases/tag/v6.10.0>
Run npm install -g npm to update!

Tabla



Funciones

```
1 var AWS = require("aws-sdk");
2 const config = require('./config/config.js');
3 AWS.config.update(config.aws_local_config);
4 const docClient = new AWS.DynamoDB.DocumentClient();
5 let saveProducts = function(){
6     var object = {
7         "id": "1",
8         "name": "tennis Reebok",
9         "amount": "50",
10        "cost": "100 USD",
11        "created_at": new Date().toString()
12    };
13    const params = {
14        TableName: config.table_products,
15        Item: object
16    };
17    docClient.put(params, function(err, data){
18        if(err){
19            console.log("Este es el error: ", err);
20        }
21    });
22 }
```

added 14 packages from 66 contributors in 10.132s

New minor version of npm available! 6.3.0 → 6.10.0
Changelog: <https://github.com/npm/cli/releases/tag/v6.10.0>
Run `npm install -g npm` to update!

Cálculo del precio

Link: <https://calculator.aws/#/addService/DynamoDB>

Configurar Amazon DynamoDB

Descripción:

Elige un tipo de ubicación: Región

Elige las características de DynamoDB

☐ DynamoDB on-demand capacity ☒ DynamoDB provisioned capacity ☐ DynamoDB Accelerator (DAX) clusters

☐ DynamoDB Streams ☐ DynamoDB Backup and restore ☐ DynamoDB change data capture

☐ DynamoDB Data export to Amazon S3 ☐ DynamoDB Data Import from Amazon S3

DynamoDB provisioned capacity

Clase de tabla:

DynamoDB ofrece dos clases de tablas diseñadas para contribuir a la optimización de los costos. La clase de tabla DynamoDB Standard es la predeterminada y se recomienda para la gran mayoría de las cargas de trabajo. La clase de tabla DynamoDB Standard-Infrequent Access (DynamoDB Standard-IA) está optimizada para las tablas que almacenan datos a los que se accede con poca frecuencia, donde el almacenamiento es el principal costo. Cada clase de tabla ofrece diferentes precios para el almacenamiento de datos, así como para las solicitudes de lectura y escritura.

Costo inicial total: 180,00 USD
Costo total mensual: 26,14 USD

Mostrar detalles

Guardar y ver resumen Guardar y agregar servicio

Bibliografía

Material de la materia Laboratorio II (NoSql-dto2.pdf)

https://es.wikipedia.org/wiki/Amazon_SimpleDB

<https://es.wikipedia.org/wiki/CouchDB>

<https://couchdb.apache.org>

[https://es.wikipedia.org/wiki/Servidor de CouchBase](https://es.wikipedia.org/wiki/Servidor_de_CouchBase)
<https://www.couchbase.com>

<http://erecursos.uacj.mx/bitstream/handle/20.500.11961/3053/Proyecto%20de%20titulaci%C3%B3n%20base%20de%20datos%20no%20relacional%20couchbase%20y%20su%20impacto%20en%20las%20nuevas%20exigencias%20de%20manejo%20y%20control%20de%20datos%20en%20linea%20282%2029%20%282%29.Pdf?sequence=1&isAllowed=y#:~:text=Entre%20los%20clientes%20de%20Couchbase,%2C%20entre%20otros%20%5B1%5D>.

<https://www.tecnologias-informacion.com/nosql-empresas.html#:~:text=Los%20clientes%20de%20MarkLogic%20incluyen,y%20numerosos%20bancos%20I%C3%ADderes%20mundiales>.

<https://en.wikipedia.org/wiki/MarkLogic>
<https://www.marklogic.com>

<https://mappinggis.com/2014/07/mongodb-y-gis/#:~:text=La%20lista%20de%20organizaciones%20que,Otras%20son%20eBay%2C%20Expedia>.

https://go.neo4j.com/rs/710-RRC-335/images/Neo4j_CS_DieBayerische_ES.pdf

<https://es.wikipedia.org/wiki/Neo4j>
<https://github.com/neo4j/neo4j>
<https://neo4j.com>

<https://infinitegraph.com/about-us/>

<https://en.wikipedia.org/wiki/InfiniteGraph>
<http://infinitegraph.com>

<https://adrianalonsodev.medium.com/utilizando-redis-como-sistema-de-cache-en-symfony-437e449b8c0b#:~:text=Para%20que%20puedas%20entender%20el,%2C%20Snapchat%2C%20StackOverflow%20o%20Trello>.

<https://es.wikipedia.org/wiki/Redis>
<https://redis.io>

<https://platzi.com/tutoriales/1426-db-aws/5045-amazon-dynamodb-servicio-de-base-de-datos-nosql-rapido-y-flexible-para-cualquier-escala/#:~:text=Muchos%20de%20los%20negocios%20del,sus%20cargas%20de%20trabajo%20fundamentales>.

<https://calculator.aws/#/addService/DynamoDB>

<https://es.wikipedia.org/wiki/NoSQL>
<https://web.archive.org/web/20110217092700/http://www.nosql.es/blog/>
<https://hostingdata.co.uk/nosql-database/>

<https://www.mongodb.com/scale/types-of-nosql-database-management-systems>

<https://www.allthingsdistributed.com/2012/01/amazon-dynamodb.html>

https://es.wikipedia.org/wiki/Amazon_SimpleDB

https://es.wikipedia.org/wiki/Apache_HBase

<https://hbase.apache.org>

<https://github.com/apache/hbase>

https://es.wikipedia.org/wiki/Apache_Cassandra

https://cassandra.apache.org/_/index.html

<https://de.slideshare.net/grro/cassandra-by-example-the-path-of-read-and-write-requests>