

Introducción a Cassandra

Apache Cassandra es un sistema de gestión de bases de datos NoSQL distribuido, altamente escalable y de alto rendimiento. Fue diseñado para gestionar grandes cantidades de datos en múltiples servidores, proporcionando alta disponibilidad (**CAP**) sin un único punto de fallo (**CAP**).

Cassandra es un proyecto de **código abierto** desarrollado por Apache Software Foundation y escrito en Java. Es una base de datos gratuita con un modelo de negocio basado en servicios y soporte.

Se trata de una base de datos NoSQL **columnar** lo que significa que los datos se almacenan en columnas en lugar de filas. Esto se traduce en un acceso más eficiente a los datos cuando lo que se consulta es un subconjunto de las columnas de una tabla. Estos tipos de base de datos son especialmente apropiados para consultas analíticas.

Esta base de datos, como suele suceder con las bases de datos NoSQL, no garantiza los principios ACID (Atomicity Consistency Isolation Durability). En su lugar garantiza los principios BASE:

- **Basic Availability:** la base de datos siempre está disponible.
- **Soft state:** los datos pueden cambiar con el tiempo, incluso sin una entrada de datos.
- **Eventual consistency:** la base de datos llegará a un estado consistente en algún momento.

¿Cómo usaremos Cassandra?

La forma más sencilla de usar Cassandra es mediante el uso de contenedores Docker. Para ello podemos usar el siguiente comando:

```
docker run --name cassandra -p 9042:9042 -d cassandra:latest
```

Este comando lo usaremos la primera vez que queramos usar Cassandra. En las siguientes ocasiones podremos usar el siguiente comando:

```
docker start cassandra
```

y para parar el contenedor:

```
docker stop cassandra
```

Una vez que el contenedor esté en ejecución podremos conectarnos a él usando el siguiente comando:

```
docker exec -it cassandra cqlsh
```

de esta forma obtendremos una consola de Cassandra ejecutando una *shell* de CQL (cqlsh) desde la que podremos ejecutar comandos CQL.

¿Qué es CQL?

Cassandra Query Language (CQL) es un lenguaje de consulta similar a SQL que nos permitirá interactuar con Cassandra. Cassandra dispone de *drivers* para Java (JDBC), Python (DBAPI2), Node.js (Datastax), Go (gocql) y C++.

¿Qué es una base de datos columnar?

En una base de datos columnar los datos se almacenan en columnas en lugar de filas. Esto permite que las consultas sean más rápidas.

Estas bases de datos están concebidas para recuperar datos en forma de columnas. Esto es útil cuando se desea recuperar un subconjunto de columnas de una tabla que contiene un gran número de columnas. Estas consultas son típicas en aplicaciones analíticas.

Todos los valores de la misma columna se encuentran juntos en el disco. Esto permite que las consultas, cuando sólo nos interesa un subconjunto de columnas, sean más rápidas.

Características de Cassandra

Cassandra tiene las siguientes características que la diferencian de otras bases de datos:

- Base de datos distribuida.
- Tolerante a fallos.
- Escalable linealmente y de forma horizontal.
- No sigue el patrón **maestro-esclavo**.
- **Permite especificar el nivel de consistencia de las operaciones.**

En primer lugar Cassandra es una **base de datos distribuida**. Esto quiere decir que los datos se almacenan en múltiples nodos de un cluster. Esto permite que **los datos estén replicados** y que la base de datos sea **tolerante a fallos**.

Cassandra es **escalable linealmente**. Esto quiere decir que si añadimos más nodos a la base de datos el rendimiento de la misma aumentará de forma lineal. De esta forma podemos escalar la base de datos de forma horizontal.

No sigue el patrón maestro-esclavo. Todos los nodos son iguales y no hay un nodo maestro, es decir, es una base de datos **peer-to-peer**.

Desventajas de Cassandra

Para escalar Cassandra de forma horizontal es necesario añadir nuevos nodos a la base de datos. Esto puede ser complicado en algunos casos.

Para maximizar el rendimiento tendremos que conocer cuales serán las consultas que se realizarán a la base de datos con antelación a su creación. Esto puede ser complicado en algunos casos.