

Cassandra DB

Natalia González Palacios

March 10, 2023



**UNIVERSIDAD
CENTRAL**

¿Qué es?

- Es un SGBD no relacional (NoSQL).
- Base de datos distribuida (Divide en fragmentos, en diferentes ubicaciones físicas), masivamente escalable.
- Código abierto de nivel superior, gratuito y modificable por la comunidad de desarrolladores.
- Esta soportada por la Apache Software Foundation.

Origen

- Fue desarrollada inicialmente por Facebook
- El nombre "Cassandra" fue elegido por los desarrolladores en honor a la princesa troyana Cassandra
- Se publicó como software de código abierto en julio de 2008.
- En 2009, Apache Software Foundation asumió el control del proyecto y lo convirtió en un proyecto de código abierto de nivel superior.



Cassandra

Principales características

- Escalabilidad horizontal
- Modelo de datos columnar
- Alta disponibilidad
- Bajo tiempo de latencia
- No tiene un solo punto de falla
- Flexibilidad
- CQL (Cassandra Query Language)

Modelo columnar

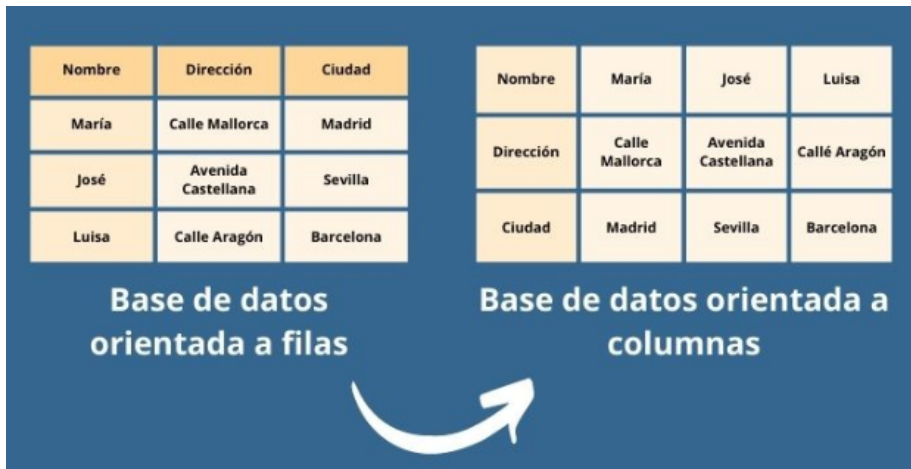
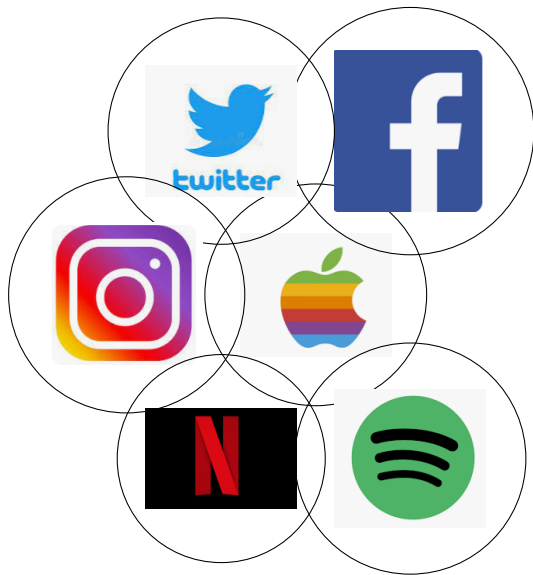


Figure: Comparación modelo tradicional vs modelo columnar.

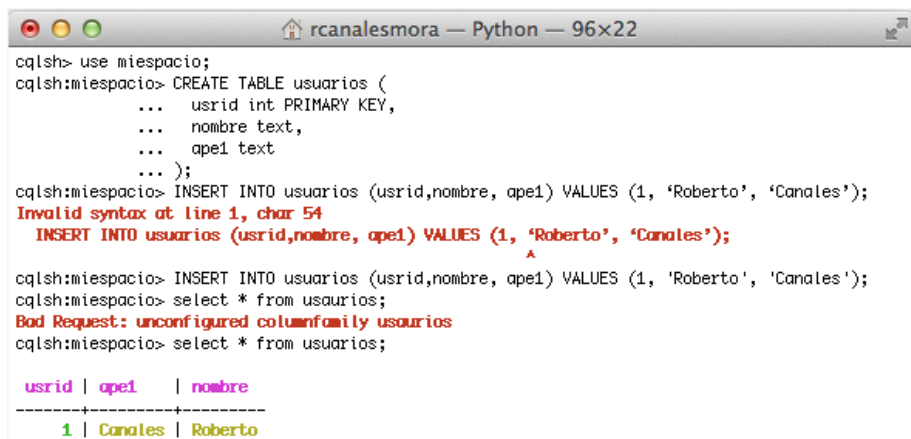
Para qué sirve

Se utilizan en aplicaciones en línea de alta disponibilidad, sistemas de comercio electrónico, sistemas de control de inventario y en cualquier aplicación donde la escalabilidad y la tolerancia a fallos son esenciales.

Principales empresas



- Cassandra introduce CQL (Cassandra Query Language)
- Cassandra utiliza un lenguaje diferente al SQL tradicional porque se basa en un modelo de datos distribuido, en lugar de un modelo de datos relacional.
- Cassandra permite que el sistema se adapte mejor a la arquitectura de datos distribuida.
- CQL también proporciona una interfaz más simple y fácil de usar para los desarrolladores.
- Se puede ejecutar CQL en la consola de comandos de Cassandra "cqlsh".
- O también en herramientas de terceros como DataStax Studio, Apache Spark y Apache NiFi.



```
cqlsh> use miespacio;
cqlsh:miespacio> CREATE TABLE usuarios (
    ... usrid int PRIMARY KEY,
    ... nombre text,
    ... ape1 text
    ... );
cqlsh:miespacio> INSERT INTO usuarios (usrid,nombre, ape1) VALUES (1, 'Roberto', 'Canales');
Invalid syntax at line 1, char 54
    INSERT INTO usuarios (usrid,nombre, ape1) VALUES (1, 'Roberto', 'Canales');
                                     ^
cqlsh:miespacio> INSERT INTO usuarios (usrid,nombre, ape1) VALUES (1, 'Roberto', 'Canales');
cqlsh:miespacio> select * from usaurios;
Bad Request: unconfigured columnfamily usaurios
cqlsh:miespacio> select * from usuarios;
```

usrid	ape1	nombre
1	Canales	Roberto

Figure: Interfaz gráfica "cqlsh".

Ejemplos CQL

```
CREATE KEYSPACE mikeyspace WITH replication = {'class': 'SimpleStrategy', 'replication_factor': '3'} AND durable_writes = true;

SELECT * FROM mikeyspace.mitable WHERE id = 1;

CREATE TABLE mikeyspace.mitable (id INT PRIMARY KEY, name TEXT, age INT);

INSERT INTO mikeyspace.mitable (id, name, age) VALUES (1, 'Juan', 25);

UPDATE mikeyspace.mitable SET age = 26 WHERE id = 1;
```

Figure: Ejemplo códigos Cassandra”.

- Puede incluir cláusulas como JOIN, WHERE, GROUP BY, ORDER BY

Ventajas y desventajas

Ventajas:

- Alta disponibilidad (Se replica la información en los nodos Cluster).
- Tolerancia a Fallos.
- Código abierto, gratuito, modificable y adaptable a necesidades específicas.
- Eficiente (Distribución de las cargas de trabajo en los nodos)

Desventajas:

- La gestión de una base de datos distribuida es más compleja.
- Cassandra requiere un hardware potente y bien configurado para funcionar correctamente, lo que puede aumentar los costos.
- A diferencia de los SGBD relacionales, Cassandra no es óptimo para consultas complejas que involucran varias tablas.
- Aunque CQL es fácil de usar, es menos expresivo que SQL y no admite todas las características de SQL.

- Apache Cassandra. (2021). Retrieved September 9, 2021, from <https://cassandra.apache.org/>
- DataStax. (2021). What is Cassandra? Retrieved September 9, 2021, from <https://www.datastax.com/what-is-cassandra>
- Lakshman, A., Malik, P. (2010). Cassandra: A decentralized structured storage system. ACM SIGOPS Operating Systems Review, 44(2), 35-40.
- Hewitt, C., Sahu, S. (2016). Cassandra High Performance Cookbook: Over 150 recipes to design and optimize large scale Apache Cassandra deployments. Packt Publishing Ltd.
- The Apache Software Foundation. (2021). Apache Cassandra - Features. Retrieved September 9, 2021, from <https://cassandra.apache.org/features/>