

# CASSANDRA: UNA BASE DE DATOS DISTRIBUIDA IDEAL PARA EL MODELADO DEL BIG DATA.



**Integrantes:** Equipo "Thunderbolts"

Joshua Ramirez Alfaro B76136  
Daniel Sancho Varela B66676  
Mario Viquez González B78443

# UN POCO DE HISTORIA...

En 2008, se publicó como proyecto de código abierto, comenzó en Facebook, donde se diseñó para impulsar la función de búsqueda en la bandeja de entrada.

En febrero de 2010, se convirtió en un proyecto de primer nivel de la Apache Software Foundation.

El 17 de febrero de 2010 se graduó como proyecto de alto nivel.

Está inspirado e influenciado por los trabajos de 2007 de Amazon Dynamo y 2006 de Google Bigtable. Actualmente, es mantenido y desarrollado por Datastax.



# ¿POR QUÉ UTILIZAR APACHE CASSANDRA PARA BIG DATA?

Gestiona los datos en forma de clusters interconectados a miles de nodos repartidos por los centros de datos.

Se le conoce como una "base de datos orientada a columnas" en NoSQL, en la que la información se almacena columna por columna, en contraste con el enfoque basado en filas de los sistemas de bases de datos tradicionales.

Debido a esto tiene menos operaciones de E/S para poder almacenar la información.

Pilares de Cassandra:

- Escalabilidad
- Baja probabilidad de fallos por la distribución de la información.

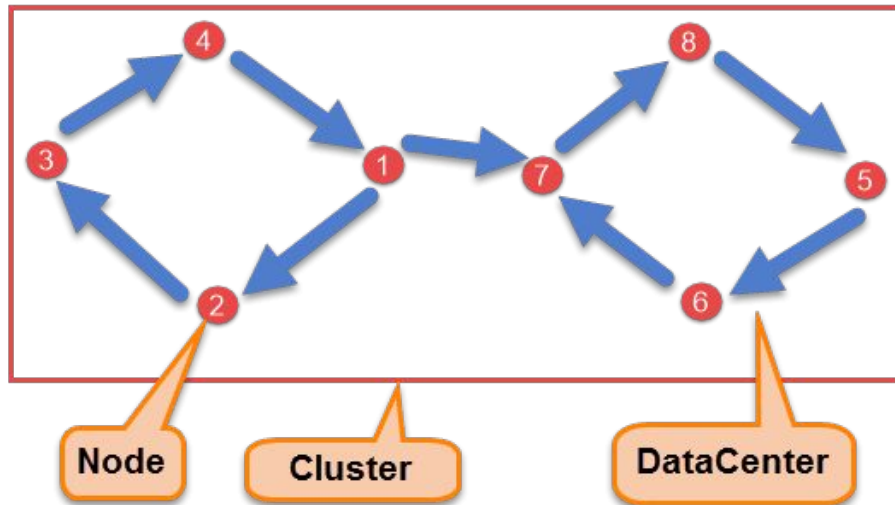


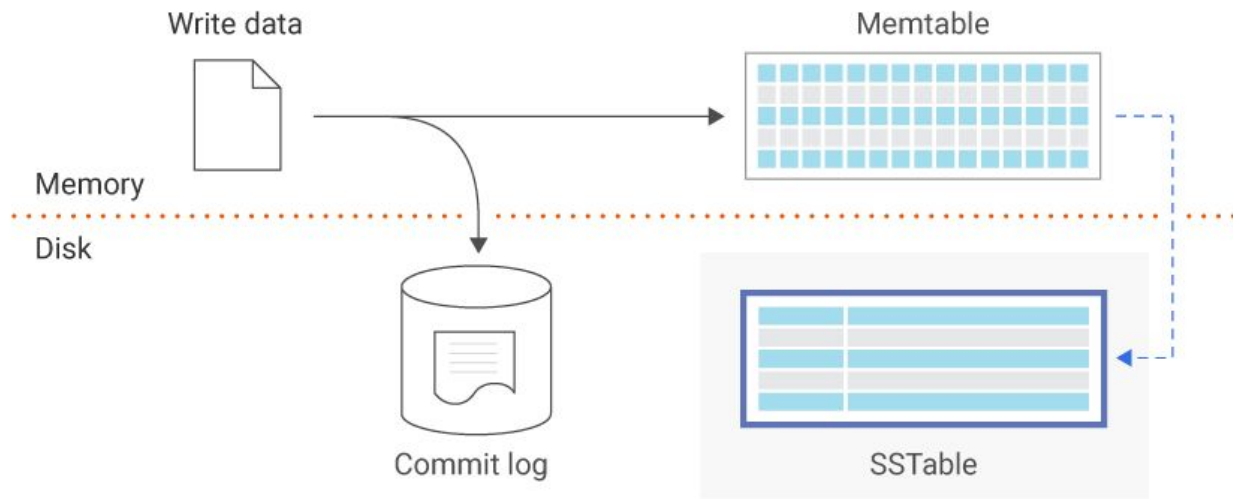
# ¿QUÉ HACE ÚNICO A APACHE CASSANDRA?

- Servicio altamente disponible y sin un punto único de fallo
- Verdadera consistencia en acceso y disponibilidad
- Es capaz de escribir de manera rápida grandes cantidades de datos sin afectar la efectividad de las lecturas
- Ofrece gran precisión y velocidad sin dependencia en el volumen de los datos.

# ARQUITECTURA

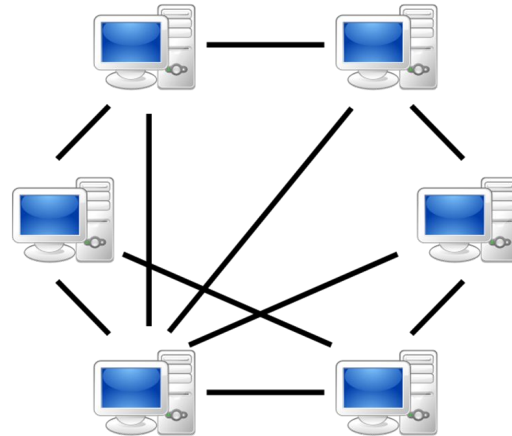
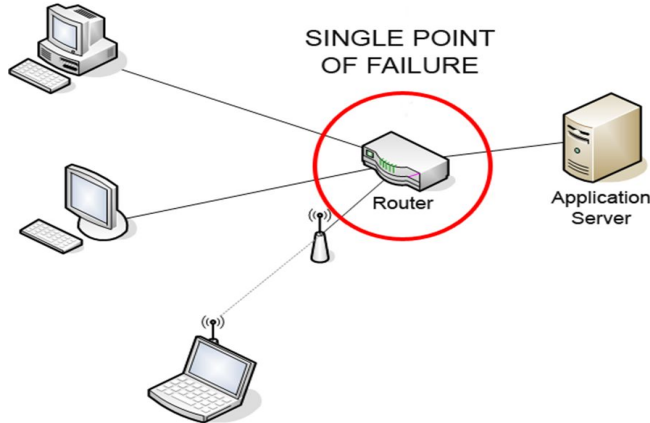
- **Nodo:** Almacena datos.
- **Centro de datos:** Colección de nodos.
- **Clúster:** Contenedor de centro de datos.
- **Commit log:** Mecanismo de recuperación.
- **Mem-table:** Estructura de datos.
- **SSTable:** Archivo en disco.



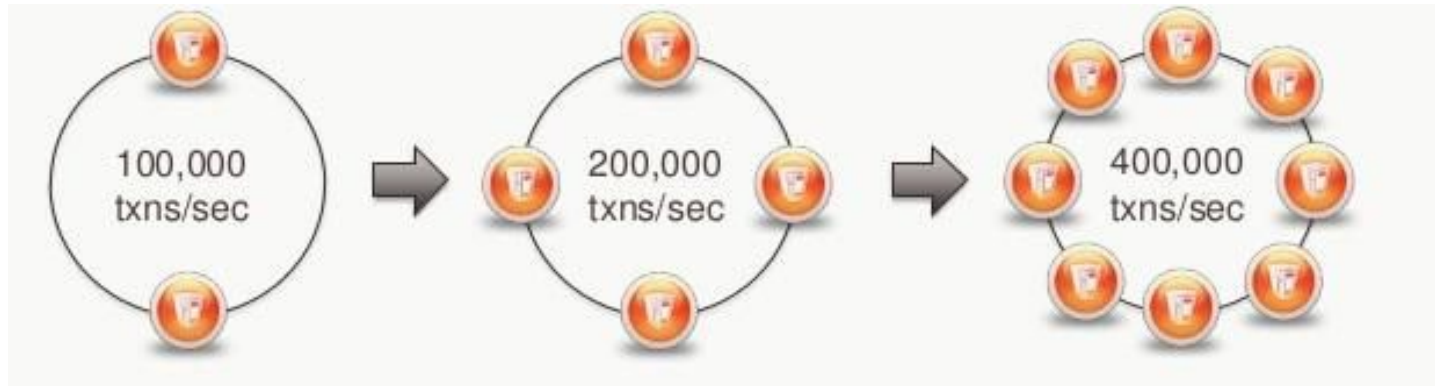


# CARACTERÍSTICAS: DISTRIBUIDA

- Cassandra distribuye sus datos entre varios nodos
- Es muy útil para clusters.
- Realiza comunicación peer-to-peer(p2p). No single point of failure.

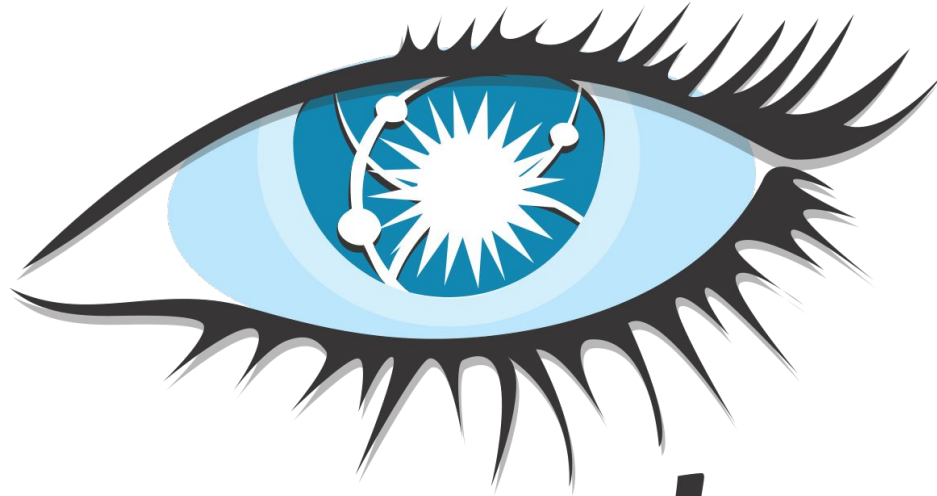


# CARACTERÍSTICAS: ESCALA LINEALMENTE



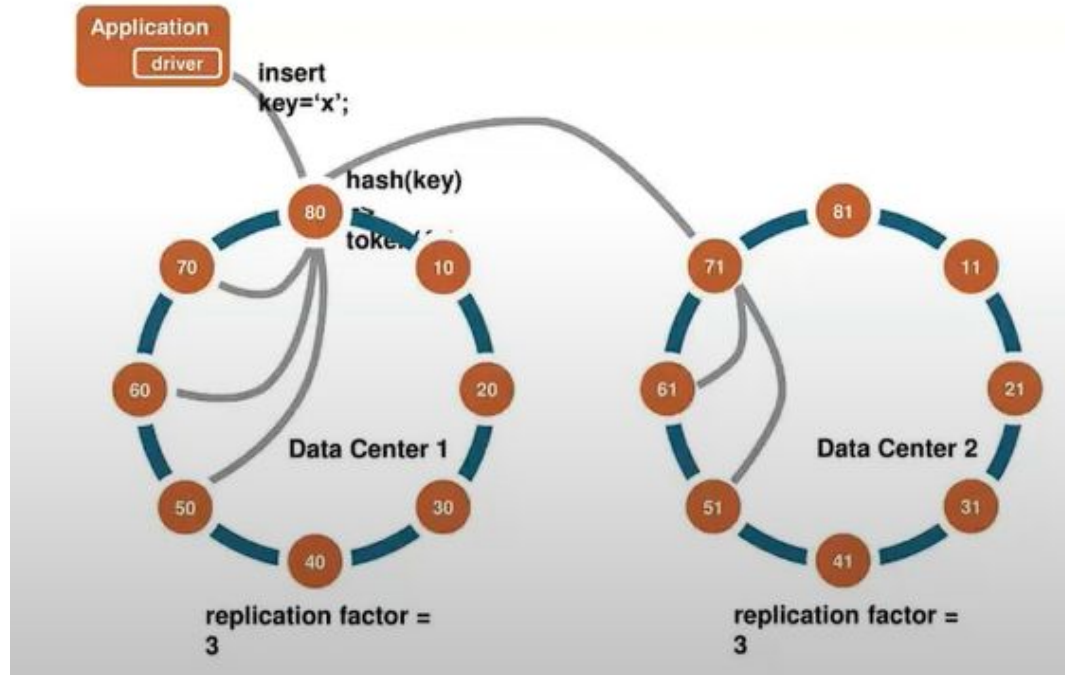


FUNCIONES PRINCIPALES



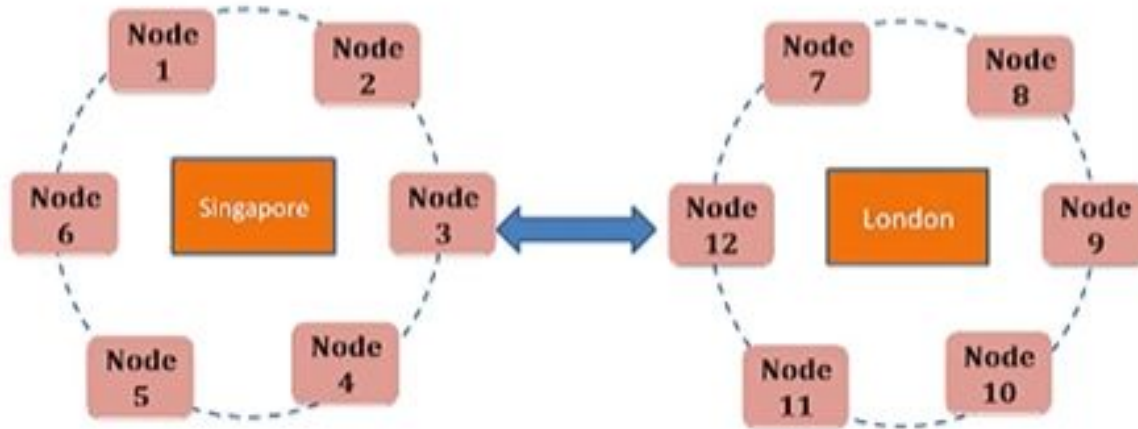
***cassandra***

# PERMITE LA REPLICACIÓN EN VARIOS DATA CENTER



# ES TOLERANTE A FALLOS

- Los nodos que presentan fallos pueden recuperarse sin necesidad de detener o inactivar la aplicación.



# ESCENARIOS APROPIADOS

- Las escrituras superan a las lecturas por un gran margen.
- Los datos se actualizan raramente.
- El acceso de lectura es por una clave primaria conocida.
- Los datos se pueden particionar mediante una clave que permite repartir la base de datos entre varios nodos.
- No es necesario realizar uniones o agregados.

# CASSANDRA



VS

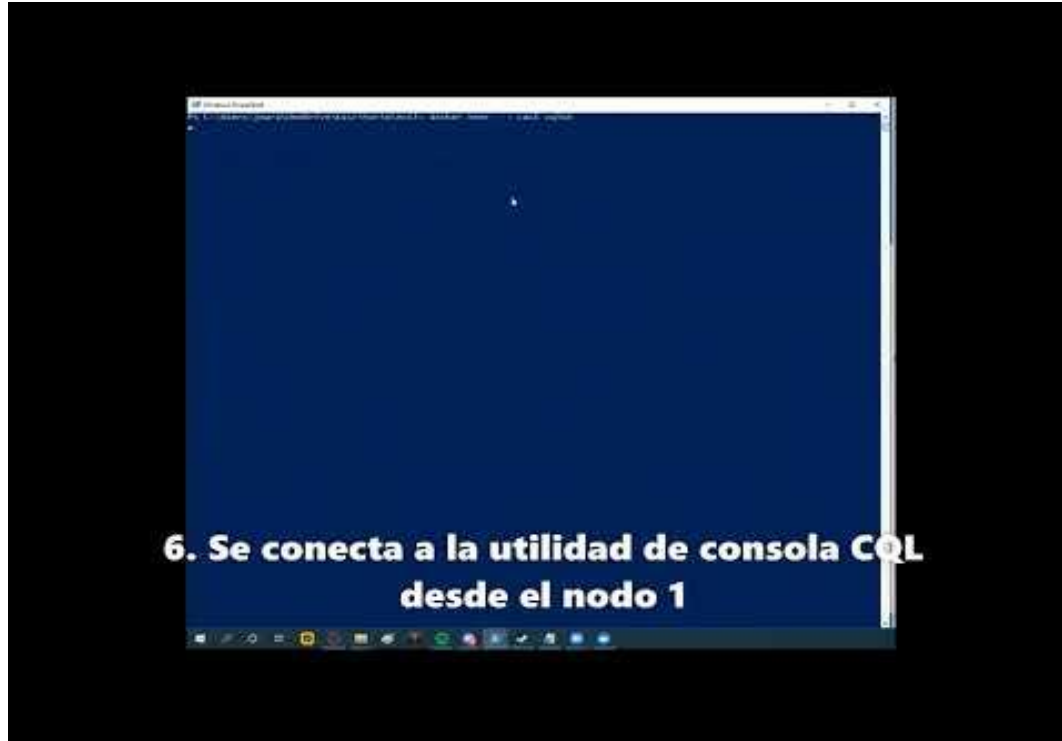


# MONGODB

- Arquitectura de almacenamiento columnar y almacena los datos en forma de filas y columnas tradicionales.
- Múltiples nodos maestro.
- Escalabilidad de escritura.

- Almacena los datos en un formato similar a JSON en documentos que se almacenan en una colección.
- Único nodo maestro.
- Escalabilidad de escritura limitada.

# EJEMPLO PRÁCTICO - CLUSTER EN APACHE CASSANDRA



# REFERENCIAS BIBLIOGRÁFICAS

[1]. Cassandra - Introducción — DATA SCIENCE. (2020). Recuperado el 16 de Mayo del 2022, de página web: <https://datascience.eu/es/programacion/cassandra-introduccion-2/>

[2]. Welcome to Apache Cassandra's documentation! Apache Cassandra Documentation. (2022). Apache Cassandra. <https://cassandra.apache.org/doc/latest/>

[3]. Cassandra Introduction: What is Apache Cassandra?. (2022). Recuperado el 6 de Junio del 2022, de página web: <https://www.bmc.com/blogs/apache-cassandra-introduction/>

[4]. What is Cassandra and why are big tech companies using it? | Ubuntu. (2022). Recuperado el 6 de Junio del 2022, de página web: <https://ubuntu.com/blog/apache-cassandra-top-benefits>

[5]. técnicas, C., & Cassandra, A. (2020). Apache Cassandra: gestión distribuida de grandes bases de datos. Recuperado el 6 de Junio del 2022, de página web: <https://www.ionos.es/digitalguide/hosting/cuestiones-tecnicas/apache-cassandra/>



# REFERENCIAS BIBLIOGRÁFICAS

[6]. Cassandra: NoSQL, BigData, escalación y replicación. (2018). Recuperado el 6 de Junio del 2022, de página web:  
<https://www.panel.es/bigdata-cassandra-escalacion-y-replicacion/>

[7]. Apache Cassandra: Introducción - Aprender BIG DATA desde cero. (2019). Recuperado el 6 de Junio del 2022, de página web:  
<https://aprenderbigdata.com/introduccion-apache-cassandra/>

[8]. Top 5 reasons to use the Apache Cassandra Database. (2018). Recuperado el 6 de Junio del 2022, de página web:  
<https://fedakv.medium.com/top-5-reasons-to-use-the-apache-cassandra-database-d541c6448557>

[9]. Cassandra, la dama de las bases de datos NoSQL . (2022). Recuperado el 6 de Junio del 2022, de página web:  
<https://www.paradigmadigital.com/dev/cassandra-la-dama-de-las-bases-de-datos-nosql>

[10]. Apache Cassandra | Apache Cassandra Documentation. (2022). Recuperado el 6 de Junio del 2022, de página web: [https://cassandra.apache.org/\\_/index.html](https://cassandra.apache.org/_/index.html)





# ACCESO AL REPOSITORIO DE GITHUB



<https://github.com/MarioVGz/Apache-Cassandra-Investigacion>

*¡Gracias!*