

Base de Datos II (IC4302) – Semestre 2, 2023

Lectura 4 y 5 – Consideraciones en bases de datos NoSQL

David Suárez Acosta – 2020038304

¿Cómo se diferencia el modelo de datos de Big Tables de una base de datos SQL?

El modelo de datos de Big Tables consiste en el almacenamiento de datos en filas y columnas sin un esquema definido, donde existen familias de columnas que se refiere a una unión de columnas y donde cada celda puede contener múltiples versiones de un mismo dato indexadas por un *timestamp*.

Las bases de datos SQL almacenan los datos en tablas, conformadas por columnas de datos, con un esquema claro y con relaciones entre las tablas.

¿Cuáles decisiones de diseño en Big Table aumenta el rendimiento del sistema? Explique.

Las decisiones de diseño en Big Table que aumentan el rendimiento en el sistema son todas aquellas relacionadas con el almacenamiento por columnas, es decir las *familias de columnas*.

Las *familias de columnas* son el "access control" tanto para el disco como para la memoria se realizan a nivel de familias de columnas, es decir que son la unidad base del access control, por lo que se reduce el tiempo de búsqueda de columnas relacionadas ya que se encuentran en el mismo lugar físico.

¿Considera que Big Table podría cumplir el papel de Prometheus en un sistema de Observabilidad? En caso de responder No, explique detalladamente, en caso de responder si, ¿utilizarían versiones de timestamps para cada métrica y recolectarían cada métrica como un row separado?

Pienso que Big Table no podría cumplir el papel de Prometheus en un sistema de Observabilidad ya que, aunque se podría intentar adaptar Big Table para que se utilice en el sistema, nunca lograría llegar a remplazar a Prometheus ya que Prometheus está diseñada para procesamiento en tiempo real mientras que Big Table no, por lo que no sería igual de eficiente.

Explique en detalle la organización de tablets en Big Table

Una *tablet* son los rangos de filas en una tabla de Big Table, estas tablets son la unidad de distribución y load balancing y están ordenadas lexicográficamente. Las tablets también se organizan en servidores de tablets, los cuales son parte de un cluster de servidores de tablets que permite un mejor acceso y mantenimiento de los datos almacenados en Big Table.

Comente los tipos de fallas de sistemas distribuidos en bases de datos que se mencionan en la lectura.

Los diferentes tipos de fallas de sistemas distribuidos en bases de datos mencionados en la lectura son:

- Fallas por corrupción de memoria.
- Fallas por corrupción de red.
- Fallas por problemas con el reloj.
- Hung machines.
- Fallos por particiones asimétricas de la red.
- Bugs en otros sistemas.

- Fallo por Overflow.
- Fallo por mantenimiento de hardware, planeado y no planeado.