

Instituto Tecnológico de Costa Rica

Jesús Andrés Cortés Álvarez

Carné: 2021579439

Lectura 4

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

- Una de las diferencias que presenta el sistema Big Table es la capacidad de escalar grandes cantidades de datos. Caso contrario con base de datos SQL que, por su estructura de modelo de datos relacional, no se permite tanta cantidad de datos.
- A pesar de que posee algunas características similares a las bases de datos relacionales como escalabilidad y alto rendimiento, Big Table presenta una interfaz diferente, ya que proporciona a los clientes un modelo simple que admite el control dinámico sobre el diseño y formato de datos.
- Los parámetros del esquema de Big Table permiten a los clientes, controlar dinámicamente si sirven datos desde la memoria o desde disco.

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

- Un ejemplo mencionado en la lectura es el hecho de conservar una copia de una colección de páginas web e información relacionada que podría llegar a ser muy utilizada en muchos proyectos diferentes, mediante el URL como clave de una fila, otros aspectos de la página web como el nombre de la columna y se almacenaría el contenido de las páginas web en la columna de contenido.
- Otra de las decisiones de diseño mencionadas es que facilita a los clientes a razonar sobre el comportamiento del sistema en presencia de datos concurrentes a actualizaciones en la misma fila.
- En el apartado de familias de columnas, se menciona otro ejemplo que aumenta el rendimiento del sistema, como lo es el control de acceso y la contabilidad tanto del disco como de la memoria, ya que llevar este control permite administrar varios tipos diferentes de aplicaciones, como agregar nuevos datos a la base, leer datos base y la creación de familias de columnas derivadas, incluso al finalizar, observar todas las familias existentes por motivos de privacidad.

3. ¿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?

- Tal vez no podría cumplir el papel de Prometheus tan completo, pero tomando en cuenta la funcionalidad de Google Analytics donde menciona que en cada fila se almacena una tupla, que contiene el nombre del sitio web y la hora en la que se creó la sesión, esto garantiza que las sesiones que se visitan al mismo sitio web son contiguos y se encuentran ordenados

cronológicamente, pero a pesar de esta funcionalidad no es una herramienta específica de observabilidad como Prometheus.

4. Explique en detalle la organización de tablets en Big Table

- Se define de esta forma a cada rango de filas, que también funciona como la unidad de distribución y equilibrio de carga. De esta forma genera las lecturas en rangos de filas cortos son más eficientes y requieren comunicación con una pequeña cantidad de máquinas. También cabe mencionar que cuando una tableta crece, se divide automáticamente en varias tabletas.

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

- Según la lectura algunas de las fallas en los sistemas distribuidos son los siguientes:
 - Particiones de red estándar.
 - Fallas debido a corrupción de memoria y red.
 - Desfase del reloj.
 - Otras fallas como máquinas bloqueadas.
 - Particiones de red extendidas y asimétricas.
 - Desbordamiento de cuotas GFS.
 - Mantenimiento de hardware planificado y no planificado.