

Instituto Tecnológico de Costa Rica

Campus Tecnológico Central Cartago

Escuela de Ingeniería en Computación

Bases de Datos II

IC-4302

Resumen 8

Fecha de entrega: 23/05/2023

I-Semestre 2023

Profesor:

Nereo Campos Araya.

Elaborado por:

Diana Sanabria Calvo 2021436548.

Bigtable: A Distributed Storage System for Structured Data

Bigtable se creó para manejar grandes volúmenes de datos estructurados.

Las aplicaciones internas de Google producen y manejan enormes cantidades de datos que los sistemas de bases de datos convencionales no pueden manejar. Bigtable se desarrolló para abordar estos problemas y ofrecer una solución confiable y escalable.

Las tablas, las columnas y las versiones constituyen la base de la arquitectura de Bigtable. Las tablas de Bigtable están organizadas en filas y la clave de fila de cada fila las identifica de manera única. Las columnas se clasifican en familias, y cada fila tiene un máximo de tres columnas. Varias iteraciones de un valor están presentes en cada celda de una tabla y cada iteración tiene una marca de tiempo.

Bigtable distribuye los datos en múltiples servidores llamados "tablets". Cada tablet es responsable de un rango de claves de fila y se asigna a un servidor en el clúster de Bigtable. El maestro de Bigtable es responsable de asignar los tablets a los servidores y manejar la reasignación de tablets cuando sea necesario. Esto permite que Bigtable maneje conjuntos de datos muy grandes dividiéndolos en unidades más pequeñas y distribuyéndolos en varios servidores.

Para acceder y administrar los datos almacenados, Bigtable proporciona una interfaz de programación sencilla. Los clientes se comunican con Bigtable mediante comandos "put", "get" y "delete". También admite operaciones más complejas como filtros y escaneo de rango de filas. Los clientes pueden beneficiarse de la replicación y la tolerancia a fallas del sistema al conectarse a Bigtable de forma remota.

El uso de Bigtable en varias aplicaciones internas de Google, incluidas Google Earth y Google Analytics, también se analiza en el artículo. Además, se proporcionan estadísticas de rendimiento y escalabilidad para mostrar qué tan bien maneja Bigtable las cargas de trabajo pesadas.

Bigtable utiliza métodos como el particionamiento y la replicación de datos para lograr una alta disponibilidad y rendimiento, se basa en una arquitectura distribuida. Es una herramienta efectiva para aplicaciones que necesitan un almacenamiento efectivo de datos estructurados debido a su sencilla interfaz de programación y funciones sofisticadas.