



Instituto Tecnológico de Costa Rica.

CAMPUS TECNOLÓGICO CENTRAL CARTAGO

ESCUELA DE INGENIERÍA EN COMPUTACIÓN.

CURSO: BASES DE DATOS II IC-4302

I SEMESTRE 2023.

Prueba Corta 3 y 4

FECHA DE ENTREGA: 11/04/2023

Profesor:

Nereo Campos Araya.

Estudiante:

Diana Sanabria Calvo 2021436548

Índice

1. Pregunta 1	3
1.1. 1. Comente de que forma un query como: <code>SELECT col1, col2, ..., coloN FROM table; N</code> es menor al (número total de columnas /2). Se puede ver beneficiado por un columnar storage (25 pts)	3
2. Pregunta 2	4
2.1. 2. Una base de datos MariaDB o MySQL se encuentra bajo mucha presión debido a un alto número de lecturas (heavy read), de qué forma se podría lograr bajar está presión sobre la base de datos, justifique su respuesta (25 pts).	4
3. Pregunta 3	5
3.1. Durante la semana santa 2023 un grupo de estudiantes de Bases de Datos 2 decidieron desarrollar un juego retro de la plataforma Atari llamado Breakout, el mismo en su versión original es de un sólo jugador, pero se toma la decisión de implementar una versión multi-jugador, el backend se encuentra almacenado en una base de datos NoSQL que se encuentra ubicada en Azure Cloud en una región en Estados Unidos, el juego se volvió muy popular en menos de una semana, tanto que una gran cantidad de usuarios en otros continentes se han estado quejando de retrasos en el juego lo cual causa una mala experiencia, gracias a las herramientas de observabilidad configuradas, se ha descubierto que la naturaleza del workload es pocas escrituras y muchas lecturas a los mismos datos, presente dos recomendaciones (con su debida explicación) desde el punto de vista de base de datos que daría para mejorar el rendimiento de la misma (50 pts).	5

Bases de Datos II

Prueba Corta 3 y 4

Elaborado por Diana Sanabria Calvo.

1. Pregunta 1

- 1.1. 1. Comente de que forma un query como: **SELECT col1, col2, ..., coloN FROM table; N es menor al (número total de columnas /2).**
Se puede ver beneficiado por un columnar storage (25 pts)

El almacenamiento en columnas puede resultar ventajoso para las consultas que eligen menos de la mitad de las columnas de una tabla.

Cada fila de la tabla se almacena continuamente en el disco en un sistema de almacenamiento basado en filas. Por lo tanto, el motor de la base de datos debe leer todas las columnas de todas las filas para recuperar los datos necesarios al ejecutar una consulta que solo necesita una parte de las columnas de la tabla. En términos de tiempo y uso de recursos de almacenamiento, esto puede no ser eficiente.

Los valores de cada columna se almacenan continuamente en el disco en un sistema de almacenamiento en columnas, por el contrario. El motor de la base de datos solo necesita leer las columnas necesarias y evitar leer las columnas innecesarias al ejecutar una consulta que solo selecciona un subconjunto de las columnas de la tabla. Como resultado, la cantidad de tiempo y recursos necesarios para completar la consulta puede reducirse considerablemente.

2. Pregunta 2

- 2.1. 2. Una base de datos MariaDB o MySQL se encuentra bajo mucha presión debido a un alto número de lecturas (heavy read), de qué forma se podría lograr bajar esta presión sobre la base de datos, justifique su respuesta (25 pts).**

La implementación de un sistema de caché para mantener las respuestas a las consultas más populares en la memoria es una forma eficiente para reducir la cantidad de consultas a la base de datos que consumen mucho tiempo. Al reducir la cantidad de tiempo dedicado a buscar y ordenar datos, los índices pueden mejorar significativamente el rendimiento de la base de datos. Se tiene que verificar si las tablas tienen los índices correctos y si están configuradas para las consultas más comunes. Podemos hablar de la partición de datos y esta puede ayudar a que la base de datos funcione de manera más eficiente si la tabla en cuestión es muy grande. Al dividir la tabla en varias particiones más pequeñas, esto puede acelerar las consultas y reducir el tiempo de bloqueo. Por último tenemos lo que es la replicación, gracias a esta varias instancias de bases de datos pueden beneficiarse de la distribución de carga de lectura proporcionada por la replicación. Puede reducir las solicitudes de lectura en la base de datos principal y mejorar el rendimiento replicando la base de datos en servidores adicionales.

3. Pregunta 3

- 3.1. Durante la semana santa 2023 un grupo de estudiantes de Bases de Datos 2 decidieron desarrollar un juego retro de la plataforma Atari llamado Breakout, el mismo en su versión original es de un sólo jugador, pero se toma la decisión de implementar una versión multi-jugador, el backend se encuentra almacenado en una base de datos NoSQL que se encuentra ubicada en Azure Cloud en una región en Estados Unidos, el juego se volvió muy popular en menos de una semana, tanto que una gran cantidad de usuarios en otros continentes se han estado quejando de retrasos en el juego lo cual causa una mala experiencia, gracias a las herramientas de observabilidad configuradas, se ha descubierto que la naturaleza del workload es pocas escrituras y muchas lecturas a los mismos datos, presente dos recomendaciones (con su debida explicación) desde el punto de vista de base de datos que daría para mejorar el rendimiento de la misma (50 pts).**

Desde mi punto de vista como DBA aplicaría lo que es la replicación a nivel geográfico para reducir la latencia para los usuarios de otros continentes y así mejorar el rendimiento de las bases de datos en este caso NoSQL. Los datos se pueden duplicar en una región diferente de Azure Cloud que esté más cerca de los usuarios afectados si la base de datos NoSQL admite la replicación geográfica. Gracias a esto podemos mejorar considerablemente la experiencia del usuario y la velocidad de acceso a los datos. Por otro lado aplicaría lo que es el escalamiento horizontal, la naturaleza del workload es de pocas escrituras y muchas lecturas a los mismos datos, podemos agregar nodos de lectura a la base de datos NoSQL. Al distribuir la carga de lectura en varios nodos, se puede admitir más usuarios simultáneos sin sufrir penalizaciones en términos de rendimiento. En conclusión, el rendimiento de la base de datos NoSQL utilizada para almacenar el backend del juego Breakout se puede mejorar implementando la escala horizontal o la replicación geográfica. Al hacerlo, podemos disminuir la latencia y la carga de lectura se puede distribuir entre varios nodos, lo que permite que el juego admita más usuarios simultáneos.