

Prueba Corta #2 y #4

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

R/ Como la base de datos guarda los datos de cada columna juntos en el disco, al hacer esta consulta, solo se necesitaría acceder a las columnas solicitadas, en vez de revisar todas las columnas de la tabla solo por la necesidad de revisar fila por fila. Además de esto, el columnar storage permite un procesamiento en paralelo de manera más eficiente.

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).

R/ Hay diversas maneras de lograr bajar esta presión sobre la base de datos, en primera instancia se podrían utilizar nodos read-only, los cuales serían nodos que se encarguen solo de ejecutar consultas tipo select en los datos; además de esto, se podría generar un caché de consultas, lo cual consistiría en almacenar en caché los resultados de las queries más frecuentes, así no se ejecutarían estas queries a la base de datos cada vez que sea necesario, simplemente se consulta al caché y este ya tendrá la respuesta y con mucha más velocidad.

3. 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).

R/

1. Utilizar un caché de consultas: tal y como se mencionó en la respuesta anterior, un caché de consultas podría ser de gran utilidad para ayudar a mitigar este problema, reduciendo la cantidad de consultas que llegan a la base de datos y, por ende, mejorando el rendimiento de esta. Como se está usando Azure Cloud, este caché de consultas podría ser implementado utilizando Azure Cache for Redis.
2. Utilizar particionamiento de datos y/o múltiples servidores: el particionamiento de datos es una técnica para dividir los datos en particiones más pequeñas, mientras que el llamado "escalado horizontal" es una técnica para distribuir la carga de trabajo en múltiples servidores y se pueden hacer a través de Azure Cloud con servicios como Cosmos. Esto puede provocar una mejora considerable de rendimiento, gracias a que aliviana la carga de la base de datos. Además de esto, estos múltiples servidores se pueden distribuir en múltiples ubicaciones, para así mejorar el rendimiento en otros continentes (availability zones).