

Prueba Corta # 9

1. Suponiendo que un sistema de bases de datos relacional que presenta un read-heavy workload y los queries son muy diferentes, explique detalladamente ¿porque el uso de caches puede afectar el rendimiento del sistema de forma negativa? (30 pts)

Primero, los datos pueden cambiar y, si el cache no se actualiza a tiempo, los datos dentro de este último pueden volverse obsoletos, por ende, el usuario vería datos desactualizados. Además de esto, hay que pensar en el mantenimiento de este cache y los recursos que este mismo consuma. El cache se puede volver ineficiente si las consultas son muy diferentes una de otra, porque, si cada consulta solicitada tiene datos diferentes, esto causa que los datos cacheados puedan no volverse a utilizar; otra forma de hacer el cache ineficiente es si las consultas no acceden datos que están cerca físicamente o temporalmente, ya que los datos escritos en cache son desplazados antes de que se puedan reutilizar.

2. El particionamiento de tablas en bases de datos relacionales es un concepto muy parecido al de shards en bases de datos NoSQL, explique detalladamente ¿Cómo afecta el particionamiento y el sharding en el rendimiento de bases de datos SQL y NoSQL? (30 pts)

El particionamiento puede mejorar el rendimiento de consultas, ya que este reduce la cantidad de datos que deben ser analizados en una consulta dada. Además, puede hacer la indexación y respaldos más eficientes con relaciones en paralelo. Por otro lado, el sharding puede mejorar el rendimiento, ya que este distribuye la carga de trabajo entre múltiples servidores y permite realizar las operaciones en paralelo. Aparte de esto, también puede ayudar a superar las limitaciones de tamaño de los sistemas de almacenamiento porque este distribuye los datos en más espacio de almacenamiento. Habiendo dicho esto, el particionamiento y el sharding también pueden tener un impacto negativo y complicar la administración de base de datos. Estas pueden sobrecomplicar y hacer menos eficientes las consultas que abarcan múltiples particiones o shards, además de aumentar la complejidad y consistencia de transacciones y datos.

3. En un sistema de bases de datos con Strong Consistency cuyo workload es de read-heavy y write-heavy, ¿Cómo afectan los exclusive locks el rendimiento de las bases de datos NoSQL? (20 pts)

En las bases de datos write-heavy, las operaciones de escritura pueden bloquear las operaciones de lectura y escritura concurrentes en el mismo registro, causando mayor latencia. Los exclusive locks hacen que cada operación tenga que esperar si el registro que se quiere leer está siendo actualizado por una operación de escritura, en read-heavy, tiene como consecuencia una cantidad significativa de tiempo de espera, realentizando las operaciones de lectura y degradando el rendimiento general del sistema. Además de esto, como el sistema posee los dos, esto puede llevar a la contention. Si muchas operaciones de lectura y escritura están intentando acceder al mismo registro simultáneamente, estas operaciones podrían terminar esperando a que se libere el exclusive lock, aumentando significativamente la latencia y reduciendo el rendimiento del sistema.

4. Explique detalladamente, ¿Cómo afecta la selección de discos físicos el rendimiento de una base de datos SQL y NoSQL? (20 pts)

Hay muchos factores que pueden tener un impacto significativo en el rendimiento de una base de datos:

- Velocidad de lectura/escritura: esto puede afectar directamente a la velocidad en la que se pueden leer y escribir datos. Por ejemplo, hay una diferencia considerable entre un SSD y un HDD, lo que puede resultar en mejor rendimiento.
- IOPS: A mayor capacidad de IOPS, mejor capacidad de base de datos para manejar gran cantidad de lecturas y escrituras concurrentes.
- Capacidad de almacenamiento: Si la base de datos crece hasta el límite del disco, será necesario agregar o reemplazar discos, lo cual implica costos adicionales y tiempo de inactividad.
- Durabilidad: los discos pueden tener baja durabilidad y fiabilidad, lo cual puede causar fallos. A mayor cantidad de fallos, mayor cantidad de interrupciones. Usando uno de los ejemplos anteriores, los SSD tienden a ser más duraderos y confiables que los discos HDD, a pesar de ser más costosos.
- Localización física: en el caso de que la base de datos sea distribuida, la ubicación física puede tener un impacto en rendimiento, ya que, ahora también tendría un impacto la latencia de red.