



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 9

FECHA DE ENTREGA: 31/05/2023

Profesor:

Nereo Campos Araya.

Estudiante:

Diana Sanabria Calvo 2021436548

Índice

1. Pregunta 1	3
1.1. 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?	3
2. Pregunta 2	4
2.1. 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?	4
3. Pregunta 3	5
3.1. 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?	5
4. Pregunta 4	6
4.1. Explique detalladamente, ¿Cómo afecta la selección de discos físicos el rendimiento de una base de datos SQL y NoSQL?	6

Bases de Datos II

Prueba Corta 9

Elaborado por Diana Sanabria Calvo.

1. Pregunta 1

- 1.1. 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?**

Podemos decir que el uso de caches puede afectar el rendimiento del sistema de forma negativa porque cuando las consultas diferentes es difícil predecir cuales datos serán accedidos con mayor frecuencia y cuáles serán requeridos en el futuro y puede resultar lo que conocemos como "cache misses". Cuando la caché no contiene los datos solicitados. Entonces el sistema debe acceder a los datos del almacenamiento principal, lo que puede resultar costoso en términos de tiempo de acceso. Las caches tienen un tamaño muy limitado y solamente pueden contener cierta cantidad de datos. También podemos hablar de la consistencia de datos, porque en un sistema de bases de datos es importante mantener la consistencia de los datos para evitar ciertos problemas, por ejemplo problemas de lectura. Al usar una caché, existe la posibilidad de que los datos queden obsoletos o desactualizados en comparación con los datos del almacenamiento principal. Como resultado, la confiabilidad y la integridad de los datos pueden verse afectadas por resultados de consulta incorrectos o inconsistentes.

2. Pregunta 2

2.1. 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?

Con respecto al particionamiento en bases de datos relacionales y el sharding en bases de datos NoSQL son conceptos similares en términos de distribuir y dividir los datos de una base de datos en múltiples fragmentos. Aunque los enfoques difieren ligeramente en su implementación, ambos tienen un impacto en el rendimiento de las bases de datos SQL y NoSQL.

En las bases de datos relacionales, la partición implica dividir una tabla grande en unidades más pequeñas llamadas particiones para distribuir los datos en varias ubicaciones físicas o servidores. Las siguientes son algunas formas en que el particionamiento de la base de datos SQL puede tener un impacto en el rendimiento.

Los datos se dispersan en varias ubicaciones físicas mediante la partición de una tabla en fragmentos para la distribución de la carga. Al distribuir la carga de trabajo entre varios servidores, es posible reducir la congestión del servidor y aumentar el rendimiento del sistema en su conjunto. Además, se puede acceder a varias particiones a la vez cuando se ejecutan consultas, lo que mejora la capacidad de procesamiento en paralelo y acelera los tiempos de respuesta.

Acceso efectivo a los datos: mediante el uso de particiones, puede minimizar el tamaño de cada partición, lo que reduce el volumen de información que se debe escanear o acceder en una sola consulta. Como resultado, el sistema puede operar de manera más rápida y efectiva porque solo necesita usar una parte de los datos totales para cada consulta.

La capacidad de escalar horizontalmente o agregar nuevos servidores o nodos de base de datos para manejar aumentos en la carga de trabajo es una característica de la partición en bases de datos relacionales.

El sharding en bases de datos NoSQL implica dividir y distribuir los datos en diferentes shards o fragmentos, donde cada shard es responsable de almacenar un subconjunto de datos.

Mediante la distribución de datos entre varios fragmentos, la fragmentación en bases de datos NoSQL permite una escalabilidad masiva. Para permitir el procesamiento de consultas en paralelo y una mayor capacidad de almacenamiento, cada fragmento se puede asignar a un servidor o nodo diferente. A medida que aumentan el tamaño y la carga de trabajo de la base de datos, esto mejora el rendimiento.

También se puede lograr una latencia de consulta reducida dividiendo los datos en fragmentos y distribuyéndolos entre varios servidores. Las consultas pueden ejecutarse más rápidamente si un fragmento está más cerca del cliente o tiene una carga de trabajo más liviana, lo que mejora la experiencia del usuario y el rendimiento del sistema.

En las bases de datos NoSQL, la fragmentación puede aumentar la disponibilidad y la tolerancia a fallas. Mediante la distribución de datos entre varios fragmentos, es posible que otros fragmentos continúen funcionando sin interrupciones, incluso si falla un fragmento o un servidor. Además, la administración y el mantenimiento de la base de datos en entornos de alto rendimiento se simplifican mediante la administración de fragmentos individuales, que es más simple y escalable.

3. Pregunta 3

3.1. 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?

1. Retraso en el acceso a los datos: los bloqueos exclusivos aseguran que solo un proceso pueda acceder a un recurso a la vez. Otros procesos que requieren acceso al recurso de la base de datos deben esperar cuando se le aplica un bloqueo exclusivo. Si varios procesos intentan acceder a los mismos recursos a la vez, esto puede ralentizar el sistema.

2. Los bloqueos exclusivos pueden dificultar que la base de datos maneje una gran cantidad de operaciones simultáneas, lo cual es un problema de escalabilidad. Puede haber congestión si numerosos procesos intentan acceder y modificar los mismos recursos al mismo tiempo, lo que hace necesario que los procesos esperen antes de obtener acceso. La escalabilidad del sistema puede verse obstaculizada y como resultado el rendimiento puede verse afectado.

3. Tiempos de espera prolongados: cuando un proceso solicita un bloqueo exclusivo y otro proceso ya lo tiene, el primer proceso debe esperar hasta que se libere el bloqueo. La espera puede prolongarse, lo que reduce los tiempos de respuesta y el rendimiento general del sistema según el tiempo que demore la operación actual y el tiempo que se mantenga el bloqueo.

4. Los bloqueos exclusivos pueden dar lugar a puntos muertos o situaciones en las que varios procesos se bloquean entre sí, así como a condiciones de espera. El interbloqueo se produce cuando un proceso tiene un bloqueo en un recurso y otros procesos necesitan acceder a él. La disponibilidad y el rendimiento del sistema pueden verse afectados negativamente por estas circunstancias.

4. Pregunta 4

4.1. Explique detalladamente, ¿Cómo afecta la selección de discos físicos el rendimiento de una base de datos SQL y NoSQL?

El rendimiento de las bases de datos SQL y NoSQL se ve significativamente afectado por la elección de los discos físicos. En ambas situaciones, seleccionar unidades más rápidas, como SSD, mejora la velocidad de acceso a los datos, así como el tiempo de respuesta de consultas y operaciones de escritura. Además, las unidades de mayor capacidad permiten almacenar grandes volúmenes de datos, lo cual es esencial para las bases de datos que administran grandes conjuntos de información.

La selección correcta del disco también es importante para la escalabilidad de las bases de datos NoSQL porque los discos más rápidos pueden manejar más operaciones simultáneas y una mayor carga de trabajo. Además, las unidades con funciones de redundancia y recuperación, como RAID, mejoran la tolerancia a fallas.

En pocas palabras, la elección de los discos físicos adecuados es esencial para maximizar la eficacia y el rendimiento de las bases de datos SQL y NoSQL. Las unidades más rápidas mejoran los tiempos de respuesta y el acceso a los datos, mientras que las unidades de mayor capacidad brindan suficiente almacenamiento.