

## **Prueba Corta #9 Bases de datos II**

**Isaac Araya Solano 2018151703**

**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)**

Al tener queries muy diferentes el uso de cache se puede volver muy ineficiente debido a que el propósito de los caches es tener la información de las consultas más frecuentemente realizadas, por lo que funcionan mejor cuando se suelen repetir mucho los queries. Si hay queries muy distintos y no hay patrones claros, los caches pueden tener baja precisión y hacer que disminuya el rendimiento. Por otra parte, si se quisieran meter más datos al caché para compensar el problema de la poca repetición, nos topamos el problema de la limitación de espacio que tiene el caché, causando que se puedan reemplazar los datos del caché y que cuando se busque un dato este ya no esté por lo que debe volver a accederse y esto nuevamente provoca una disminución de rendimiento.

**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)**

Ambas permiten la distribución de datos, lo cual hace que se pueda distribuir la carga de trabajo entre varios servidores, mejorando el rendimiento del sistema en general. Esto porque al tener varios servidores se puede permitir paralelismo, es decir, hacer consultas al mismo tiempo en cada partición y esto puede acelerar respuestas complejas y mejorar la capacidad de respuesta. Además, se pueden reducir los cuellos de botella ya que hace que los usuarios se distribuyan al hacer consultas y da una mayor tolerancia a fallos porque elimina la existencia de un único punto de fallo, haciendo que si una base o un shard se cae existan otros que puedan tomar su lugar.

**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)**

Los exclusive locks son muy necesarios en los sistemas de BD con Strong Consistency, pero pueden generar problemas de rendimiento. Es común que al tener exclusive locks se produzcan cuellos de botella al tener un workload muy alto, que es el caso. Esto porque al hacer que ciertos datos sean solo accesibles para un usuario puede provocar que otros que necesiten hacer una escritura en el mismo documento que ese usuario tengan que esperar mucho tiempo a que el lock sea liberado. Por otra parte, los locks retrasan las lecturas también, ya que los exclusive locks también bloquean la información para que solo un usuario pueda leerla, por lo que hace aun mas lentas este tipo de bases de datos. Además, si se fuera a realizar una actualización masiva, esto retrasaría todas las demás transacciones de la base al hacer un exclusive lock a gran escala.

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

Es muy importante realizar una buena selección de discos físicos a la hora de trabajar en bases de datos. Hay que tener en cuenta distintos aspectos de estos ya que, tanto para bases de datos SQL como NoSQL, los discos van a tener un impacto importante en el rendimiento. A la hora de seleccionar discos hay que tener en cuenta cosas como la velocidad de escritura y lectura, en una base de datos de un workload read-heavy tal vez sea importante escoger discos con una velocidad de lectura muy alta pero la velocidad de escritura puede no ser tan importante, y viceversa en caso de ser una base de datos con un workload de write-heavy. Al hacer consultas y lecturas puede ser importante optar por un disco SSD ya que estos tienen una lectura mucho más rápida que los HDD ya que trabajan por sectores en lugar de tener discos rotatorios, por lo que es mucho más rápido acceder a la información. De igual forma estos discos son más caros, por lo que se puede optar también por un hdd pero que a su vez tenga varios cabezales de lectura para que así sea lo más rápido posible. También hay que tener en cuenta la capacidad de manejo de I/O que tenga el disco, es decir, la cantidad de operaciones de lectura y escritura que puede manejar el disco al mismo tiempo, si una base de datos va a ser

accesada por muchos usuarios al mismo tiempo, puede ser importante invertir en discos con un manejo más alto de I/Os para que pueda dar abasto con el flujo de trabajo.