



Ingeniería en Computación

Base de Datos II

Examen

Profesor(a):
Nereo Campos

Estudiantes:
Sebastian Campos Zuñiga
2016140230

12/11/2022

Pregunta 1 (60 pts)

- **Dar una solución detallada de cómo podría mejorar el rendimiento de la base de datos actual, reduciendo el Down time al mínimo, esto permitirá ganar tiempo para dar una solución mucho más duradera con la mínima afectación a los usuarios. (10 pts.)**

Para evitar el Down time al mínimo, se tiene pocas opciones para aumentar el rendimiento, una opción para hacer más rápidas las consultas es segmentar las tablas por país, normalmente las personas solo les interesan los posts de sus amigos y estos son del mismo país, al segmentar la tabla por países la búsqueda sería más rápida.

Al segmentar las tablas de esta manera se puede indexar las tablas por país para mejorar la velocidad de búsqueda un poco más.

- **Dar una recomendación detallada de que tipo de base de datos se debería utilizar para abordar este problema, además debe recomendar algunas de las bases de datos SQL o NoSQL estudiadas durante el curso tanto en lecturas, así como las utilizadas en proyectos o ejemplos en clase. Tome en cuenta que sería posible utilizar más de una base de datos para optimizar el almacenamiento de los datos de la tabla post, amigos y usuario, tome en cuenta que tan fácil es escalar la base de datos en su recomendación, debe dar prioridad a servicios managed services y SaaS, no olvide la localidad y naturaleza de los datos. (30 pts)**

Para este caso de uso se va a recomendar usar 3 bases de datos para almacenar estos mismos, para los posts que son la mayor cantidad de datos que se están ingresando se recomienda utilizar una base de datos NOSQL como aws DynamoDB que permite tener con una llave que en este caso el usuario que publica un mensaje y los atributos que serían los mensajes de esta forma se tendría todos los datos en la nube con las ventajas que esta trae como son, alta disponibilidad y constancia

Para los usuarios se podría seguir usando Maria Db para guardar los datos de cada uno de los usuarios.

Por último, para los amigos utilizar las nuevas tecnologías como son las bases de datos de grafos para tener una mayor facilidad de rastreo y seguimiento de los amigos de los usuarios, así como poder hacer recomendaciones a los mismos analizando las conexiones existentes en el grafo.

La recomendación es tener las 3 bases de datos en la nube para así pagar por lo que se necesita y tener todas las ventajas que se tiene en los servicios en la nube. Para un acceso más rápido se aconseja el uso de discos SSD y se puede tener las bases de datos lo más cercano a los usuarios para un tiempo de respuesta más rápido.

- **Comente acerca de que tan conveniente es mantener la base de datos actual en la casa de uno de los fundadores, comparado con mover ésta algún Cloud Provider como AWS. (10 pts.)**

Si es por motivos económicos, nada que hacer, si se dispone de recursos, emigrar la base de datos a un cloud provider tiene muchos beneficios, como la alta disponibilidad, la escalabilidad y pagar por lo que realmente se ocupa además de poder reasignar recursos dependiendo de la necesidad de la aplicación, en momentos de alto o bajo tráfico de usuarios.

• **Basándose en el funcionamiento de un índice invertido el cual fue estudiado en clase y es utilizado por motores como Elasticsearch y el concepto de Natural Language Processing (NLP) llamado Stemming el cual también fue discutido en clase, comente ¿Cómo se podría reducir el memory footprint de la base de datos actual? (10 pts.)**

Al ser una red social de una palabra, el índice puede llegar a ser tan grande como palabras tenga un idioma. La recomendación para hacer que el memory footprint sea mas bajo es hacer el índice invertido solo con verbos y adjetivos, así el tamaño se vería reducido de manera muy grande.

Pregunta 2 (10 pts.) Comente, ¿Cómo afectan los índices en el rendimiento de las bases de datos relacionales?, enfoque su respuesta tanto en como benefician el rendimiento así la forma en la cual lo impactan de forma negativa. Suponiendo que el hardware no es un problema (se puede comprar cuanto se necesite), ¿Podemos crear cuantos índices queramos o estos no tendrán mayor impacto en el rendimiento?

Los índices mejoran el rendimiento en las bases de datos en la forma que cargan los datos a un cache de búsqueda en memoria, haciendo que las consultas no tenga que ir a disco a buscar los datos, ir a disco es mas lento que leer desde la memoria principal, lo que sucede es que si existen mucho índices y estos no están en el cache el motor de bases de datos en el momento que se da cuenta que el dato no esta en el cache actual debe ir a buscar este índice y cargar la pagina en la que se encuentre siendo contraproducente ya que alarga el tiempo de consulta. Por lo mismo el crear mucho índices termina siendo peor para el rendimiento.

Pregunta 3 (20 pts) Figura 1 El rendimiento de todo sistema de base de datos puede verse afectado por muchos factores, uno de ellos es el ambiente en el cual se ejecuta, este se encuentra compuesto por los componentes de hardware y el sistema operativo y otros programas de usuario compitiendo por los recursos del computador. Comente de forma clara y concisa, ¿Cómo afecta el rendimiento de una base de datos los componentes ilustrados en la Figura 1?

El primer componente que puede ser afectado es donde se guardan los datos, este puede ser un disco duro que son mas lento pero mejores para y mas baratos para conservar los datos por mas tiempo o un disco duro de estado solido que son más rápidos, pero mas costosos. La conexión que se tiene el disco también afecta a la base de datos ya que puede ser por internet o directa, además de donde se encuentre el servidor si es por internet si está lejos del donde se realiza la consulta esta puede ser más lenta.

El sistema operativo también tiene una influencia muy importante en la base de datos, si en este es enfocado a servidor o si esta en una computadora que además corre otros programas este será mas lento, el servidor especializado es lo mejor para tener un mejor rendimiento.

Por último, el CPU también tiene un rol primordial en el rendimiento de una base de datos dependiendo de los núcleos que tenga esta afectara al rendimiento porque esto influye directamente a la capacidad de ejecutar 1 o varios procesos en el momento. Además, si es una escritura o una lectura esto influye si la base de datos tiene alta consistencia en los datos. El cache es donde se cargan los índices, si este es muy pequeño se ira muchas veces a disco a cargar los índices afectando la velocidad.

Pregunta 4 (10 pts) La escalabilidad automática es una característica muy deseada en los sistemas de bases de datos tanto SQL como NoSQL, la misma permite mediante la obtención de métricas en tiempo real interpretar el comportamiento actual para predecir el comportamiento futuro, con esto se puede ajustar tanto el hardware como la configuración de las bases de datos, para poder atender el workload de un sistema. Comente la importancia de la Observabilidad tanto a nivel de aplicación como de base de datos para lograr una escalabilidad automática adecuada, ¿Considera que las métricas de memoria, CPU y disco son suficientes para lograr ésta?

Tener estas lecturas son fundamentales, las métricas de memoria funcionan para ver la cantidad de espacio que se esta utilizando al momento en la base de datos, tener una capacidad cerca del limite puede llevar a que se pierdan datos. Del CPU el uso y el tiempo que tiene puede llegar para analizar si se ocupan más núcleos para así llevar a cabo mas o menos tareas en el momento. Los de disco para analizar las cantidades de lecturas y escrituras que se llevan a cabo. Con estos datos se puede llegar a concluir se se ocupan mas recursos, si el sistema esta un cloud service pedir mas recursos si es necesario o al contrario reducir los recursos así tener un mejor servicio a los usuarios y mantener los costos a el limite necesario. Otro dato importante es la cantidad de conexiones de usuario que se tienen y la cantidad de consultas que le llegan a la base de datos para poder hacer una predicción más exacta.