

Examen

Estudiante: Celina Madrigal Murillo

Carné: 2020059364

Pregunta 1

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

R/ Realizar particiones o shards para así evitar que la base de datos se vuelva un cuello de botella cuando el volumen de datos se vuelva demasiado grande y muchos usuarios intenten ingresar o utilizar la aplicación al mismo tiempo. Con las particiones se permite el procesamiento paralelo de conjuntos de datos mas pequeños en todas las particiones.

- 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)

R/ Yo utilizaría alguna base de datos NoSQL porque son mas eficientes y tienen mas rendimiento que las SQL. Se podría utilizar por ejemplo MongoDB ya que soporta Hashed Sharding el cual distribuye los datos aplicando Hash, Ranged Sharding en el cual los datos se dividen de un punto a otro y Zoned Sharding el cual asegura que los datos estén cerca del usuario y cumpla con las regulaciones.

- 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)

R/ No es nada conveniente ya que la red social continua en crecimiento entonces necesita de mas seguridad, eficiencia, escalabilidad, control, entre otros, cosa que manteniendo la base en la casa de uno de los fundadores no va a tener. Al mover la base a algún Cloud Provider tendría todos estos beneficios y seguridades y si, por ejemplo, se llegara a dañar algún equipo no habría problemas ya que al almacenar nuestros datos en la nube garantiza que estén siempre disponibles.

- 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)

R/ Si al utilizar el índice invertido para recuperar las palabras solicitadas lo hacemos recuperándolas en su “manera Stemming”, ósea reducidas hasta su raíz se podría reducir el memory footprint ya que con una sola raíz nos podemos referir a muchas palabras parecidas entonces así no se utilizaría memoria repetidas veces en palabras muy similares.

Pregunta 2

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?

R/ Los índices pueden ayudar a buscar y ordenar los registros más rápidamente, pero estos también pueden ralentizar el rendimiento al agregar o actualizar los datos. Cuando se comienzan a hacer muchos includes y se hacen crean muchos índices el consumo de recursos crece mucho entonces puede pasar que por intentar mejorar el rendimiento hagamos totalmente lo contrario.

Pregunta 3

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?

R/ Los discos SSD son mucho más rápidos que los HDD por ende si se escoge uno HDD el rendimiento será mucho menor a diferencia si se escoge uno SSD. El almacenamiento en caché es una estrategia en la que se almacena una copia de los datos delante del almacén de datos principal. Las ventajas del almacenamiento en caché incluyen tiempos de respuesta más rápidos y la capacidad de servir los datos rápidamente, lo que puede mejorar la experiencia del usuario. Una base de datos en memoria permite lecturas/escrituras más rápidas en comparación con una base de datos tradicional basada en disco. Si el CPU tiene muy pocos cores se reduce el paralelismo y esto produce que aumente el coste en tiempo de una pregunta o cualquier función de recuperación de información. La red nos delimita como podemos sincronizar las bases de datos, cuantos clientes se pueden manejar, entre otros.

Pregunta 4

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?

R/ A nivel de base de datos contamos con CPU, disco y memoria los cuales tienen un impacto muy grande en el rendimiento de la base de datos, pero no considero que sean suficientes para lograr una escalabilidad automática adecuada, también se necesitan métricas a nivel de la aplicación para lograr esto, como un API que nos puede recolectar más información o datos importantes y necesarios como por ejemplo cantidad de usuarios que ingresan al sistema.

Referencias

¿Qué es la partición de bases de datos? - Explicación de partición de bases de datos - AWS.

(2022). Amazon Web Services, Inc. <https://aws.amazon.com/es/what-is/database-sharding/>

González, M. (2020, March 2). *Ventajas y Beneficios de los Servicios de Cloud Computing.*

CICE. <https://www.cice.es/blog/articulos/ventajas-cloud-computing/>

¿Qué es el Stemming? - Seobility Wiki. (2022). Seobility.net.

<https://www.seobility.net/es/wiki/Stemming>

Índice invertido - Google Arts & Culture. (2013). Google Arts & Culture; Google Arts &

Culture. <https://artsandculture.google.com/entity/m08t1ml?hl=es>

Definition of memory footprint. (2022). PCMAG; PCMag.

<https://www.pcmag.com/encyclopedia/term/memory-footprint>

Crear un índice y utilizarlo para mejorar el rendimiento - Soporte técnico de Microsoft. (2019).

Microsoft.com. <https://support.microsoft.com/es-es/office/crear-un-%C3%ADndice-y-utilizarlo-para-mejorar-el-rendimiento-0a8e2aa6-735c-4c3a-9dda-38c6c4f1a0ce#:~:text=Los%20%C3%ADndices%20pueden%20acelerar%20las,agrega%20o%20cambia%20un%20registro.>

v-aangie. (2022, September 22). *Almacenamiento de datos en caché para optimizar el rendimiento - Microsoft Azure Well-Architected Framework.* Microsoft.com.

<https://learn.microsoft.com/es-es/azure/architecture/framework/scalability/optimize-cache>

¿Qué es una base de datos en memoria? (2021). TIBCO Software.

<https://www.tibco.com/es/reference-center/what-is-an-in-memory-database>

Bases de datos en paralelo; a la sombra del desarrollo hardware. (1995, November 16).

Computerworld.es. <https://www.computerworld.es/archive/bases-de-datos-en-paralelo-a-la-sombra-del-desarrollo-hardware#:~:text=Para%20Mauricio%20del%20Castillo%2C%20responsable,se%20le%20hace%20al%20sistema%22>.

sudivate. (2022, September 22). *Observabilidad de los datos - Cloud Adoption Framework.*

Microsoft.com. <https://learn.microsoft.com/es-es/azure/cloud-adoption-framework/scenarios/cloud-scale-analytics/manage-observability>