

1. Explique cómo afectan los siguientes componentes el rendimiento de un sistema de bases de datos:

Disco: el disco es un tipo de almacenamiento que es económico, de gran capacidad y baja velocidad. Aquí estarán todos los archivos que se van a utilizar, pero hay que tener cuidado con cómo se va a hacer. Afecta negativamente a las bases de datos por su baja velocidad. Pero también tiene un efecto positivo en las bases de datos porque permite almacenar grandes cantidades de datos y utilizar estrategias de almacenamiento para brindar alta disponibilidad y garantizar que no haya pérdida de datos.

Memoria virtual: esta memoria entra en juego porque la capacidad que tienen las memorias no es muy alta y es eficiente utilizar una parte del disco para que funcione parecido a la memoria. Los datos de la BD deseablemente van a estar en memoria, pero si se llega al límite una parte de los datos se va a ir a la virtual. El problema en este caso es que se deben ir cambiando datos y registros de memoria virtual a memoria principal, muchos context switch, una situación que ralentiza considerablemente el tiempo y rendimiento. Se debe priorizar evitar el uso de memoria virtual en el manejo de sistemas de bases de datos. Los cloud providers no recomiendan su uso.

Memoria: es la memoria principal, caracterizada por ser la más rápida de todas y con gran diferencia. Su velocidad se ve reflejada en el precio y también en las limitaciones del cpu o tarjeta madre que se use. También es uno de los componentes más importantes en una base de datos, si los registros se conservan en memoria todas las consultas y trabajo que se haga en la base será muy rápido. La transmisión de datos es más eficiente. El procesador tiene un acceso veloz a la memoria lo que añade una ventaja más para la memoria.

Caché de CPU: si un procesador tiene muchos núcleos e hilos se dice que es potente, pero por lo general no se observa mucho el papel que tiene la cache. La cantidad que tenga el procesador responderá a su potencia y capacidad de multitarea y control de procesos. Es muy importante para las bases de datos teniendo relación con el manejo rápido de información y ejecución eficiente. La cache manda las instrucciones al CPU, y comparte información que es necesaria para la ejecución de las aplicaciones. Cuando el CPU no tiene la información que ocupa este le pide a la cache que se la brinde y esta al SO.

CPU: este es el componente más rápido y más caro de la computadora, sin tomar en cuenta la GPU. En comparación con la memoria principal es más rápido. Obviamente afecta el rendimiento de las bases de datos porque es el que ejecuta todas las instrucciones de la computadora. Si el procesador tiene muchos threads y cores tiene un mayor paralelismo.

Es importante mencionar que siempre es recomendado tener hardware dedicado para la BD y buena conexión a internet.

2. ¿De qué forma se benefician las aplicaciones del uso de caches? Explique.

Como se mencionó anteriormente la cache es bastante rápida y las aplicaciones pueden aprovecharse de esto. Cuando la cache tiene la capacidad suficiente el procesador es muy rápido. Las aplicaciones pueden mandar la información necesaria para que el procesador pueda trabajar, y la cache se encarga de que el CPU no tenga que estar esperando por recursos. Los beneficios son un mejor rendimiento, baja latencia, ayuda al paralelismo, y la posibilidad de que muchas aplicaciones puedan estar trabajando a la vez.

3. Desde el punto de vista de Elasticsearch, ¿qué es un índice?

Un índice es una colección optimizada de documentos, y estos documentos son un conjunto de campos de datos. Estos campos tienen estructuras de datos dedicadas dependiendo del tipo de dato.

4. ¿Qué es un mapping en Elasticsearch?

El mapeo es cómo se va a almacenar e indexar los documentos y sus campos en Elasticsearch. Puede dejarse automático o configurarse. Si se configura permite análisis específico del idioma y usar formas personalizadas y datos que no se detectan automáticamente. Un mismo campo se puede indexar de diferentes formas dependiendo de su propósito. Elastic puede reconocer una gran cantidad de tipos de datos (booleanos, enteros, punto flotante, ...)