

1. Explique en que consiste un clustered index y cuál es la diferencia entre este y un índice non-clustered que utiliza INCLUDE para agregar columnas al índice.

Un clustered index es un índice que ordena las filas de una tabla de base de datos en función de un valor de columna específico y define la manera en que los datos de la tabla se almacenan en el disco. La diferencia entre un clustered index y un non-clustered index que utiliza INCLUDE radica en que este último almacena los punteros a los datos en las hojas a más bajo nivel del árbol, permitiendo que el índice sea más pequeño. [1]

2. Explique el concepto de memory footprint y como afecta este la creación de índices. ¿Cuál es la relación entre un memory footprint alto y la paginación a disco?

Memory footprint corresponde a la cantidad de memoria que el software hace referencia o utiliza cuando se ejecuta. También es importante notar que los programas también pueden reservar memoria mientras se ejecutan. Cuando se tiene un memory footprint alto, En relación con lo mencionado anteriormente, cuando tenemos un memory footprint muy alto, se tiene también una gran cantidad de operaciones de entrada y salida. [2] Cuando esto sucede, el sistema recurre a una paginación de disco, donde una parte del disco actúa como un pedazo de memoria adicional. [3]

3. FASTantic Inc es una empresa especializada en optimización de búsquedas sobre datos, está a sido contratada por la empresa TooSlow para ayudarle a organizar 40 billones de registros, los registros tienen las siguientes columnas:

1. country: este es un código de país
2. city: está es una ciudad en un país específico.
3. date: está es la fecha en que el registro fue agregado a los datos.
4. payload: es un documento JSON que contiene el evento.

FASTantic Inc debe optimizar la búsqueda sobre las columnas country, city y date. Explique la mejor forma de organizar los datos para incrementar la velocidad de búsqueda, actualmente se hace un scan sobre todos los datos. Asuma que no existe una base de datos mencione estructuras de datos que utilizará. ¿Que tipo de base de datos recomendaría a TooSlow para almacenar sus datos?

Considero que sería apropiado usar una estructura de datos de tipo B+ Tree para organizar los datos. En ellos, el tiempo de búsqueda está ligado a la profundidad del árbol, y de esta forma, se tendría una respuesta rápida a las búsquedas. Los datos no parecen necesitar modificaciones constantes, si no más bien lo importante son las consultas. Por esto, recomendaría una base de datos NoSql como Mongo. [3] Además de esto, se podría realizar un análisis para verificar cuales registros son consultados con mayor frecuencia. Si se nota una prioridad a los registros recientes, se puede tener los registros más recientes en un warm tier y los demás en un cold tier, ya que no tienen tanto tránsito.

*Referencias*

1. Difference between clustered and non-clustered index (2022) GeeksforGeeks. Available at: <https://www.geeksforgeeks.org/difference-between-clustered-and-non-clustered-index/> (Accessed: October 22, 2022).
2. Memory footprint (no date) Definition and Examples (2022). Available at: <https://www.contractscounsel.com/g/557/us/memory-footprint> (Accessed: October 22, 2022).
3. ¿Qué es la paginación del disco duro? Available at: <https://www.centrobanamex.com.mx/que-es-la-paginacion-del-disco-duro> (Accessed: October 22, 2022).
4. Tobin, D. (2019, July 27). Which Modern Database Is Right for Your Use Case? Integrate.io. Retrieved October 22, 2022, from <https://www.integrate.io/blog/which-database/>