

Prueba Corta # 5 y #6

Iván Solís Ávila 2018209698

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. (25 pts)

Un clustered index organiza físicamente las filas de una tabla usando la clave del índice. En otras palabras, la tabla se ordena por el índice. Puede haber uno por tabla y esto ayuda a hacer las consultas más eficientes gracias a que la base de datos puede acceder directamente a los datos necesarios, en vez de buscar en toda la tabla.

La principal diferencia entre estos dos, es la forma en que se almacenan y organizan los datos de manera física, ya que el clustered index ordena las filas de acuerdo a la clave del índice, y el INCLUDE no afecta el orden de las filas en la tabla. Además, como se dijo anteriormente, solo puede haber un clustered index por tabla, mientras que un non-clustered con INCLUDE puede tener varias instancias por tabla. Otra diferencia es que el INCLUDE agrega más columnas a la estructura del índice.

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? (25 pts)

El memory footprint es la cantidad de memoria utilizada por un proceso o programa siendo ejecutado. Es importante tomar en cuenta esto a la hora de crear índices porque, si se tiene un memory footprint alto, esto puede causar que se utilice más memoria de la disponible en el sistema, lo cual hará que este comience a paginar parte de esta memoria a disco, afectando así al rendimiento de manera considerable.

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: a. country: este es un código de país b. city: está es una ciudad en un país específico. c. date: está es la fecha en que el registro fue agregado a los datos. d. 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? (50 pts)

Primero buscaría qué columnas se encuentran más relacionadas. En este caso, country y city serían las más cercanas, por lo que generaría un clustered index para estas. Para date y payload, podríamos utilizar un non-clustered index, esto por su independencia de los otros datos. Además, ya que payload es un documento JSON, se podría utilizar una base NoSQL para almacenar estos datos, especialmente bases como MongoDB, que almacenan los datos en forma de documentos JSON. Tomando esto en cuenta, también se podría utilizar un árbol B para los índices no-clusterizados, sin embargo, esta no es la única estructura

de datos que se podría utilizar, ya que una tabla hash podría proveer búsquedas más rápidas para las consultas simples.