

Prueba Corta #5 y 6

Estudiante: Celina Madrigal Murillo

Carné: 2020059364

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)

En un clustered index el índice se encuentra en los datos y se arma la estructura en disco, así se evita que el puntero vaya a memoria por el dato, solo puede tener 1 índice por tabla. Los índices non-clustered viven en un heap sin ordenamiento, cuando se tienen que encontrar los datos se tienen que ir al disco, como se tienen los datos en disco y en las hojas estos se duplican.

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)

Memory footprint se refiere a la cantidad de memoria principal que utiliza un programa o a la que hace referencia mientras se ejecuta. El índice se guarda en memoria y cada vez que se crean índices mas grandes se llena cada vez mas la memoria y se hace lento, en lugar de ram empieza a usar disco.

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? (50pts)

Como se quiere optimizar las busquedas en las columnas country, city y date entonces sería beneficioso utilizar una busqueda orientada a columnas ya que se omite la busqueda de datos innecesarios y así se incrementa la velocidad de búsqueda. Algunas bases de datos relacionales las cuales tienen busqueda por columnas son MariaDB y Mysql entonces yo recomendaría una de estas.

Referencias:

Wikipedia Contributors. (2022, August 21). Memory footprint. Wikipedia; Wikimedia Foundation.
https://en.wikipedia.org/wiki/Memory_footprint