



Instituto Tecnológico de Costa Rica.

CAMPUS TECNOLÓGICO CENTRAL CARTAGO

ESCUELA DE INGENIERÍA EN COMPUTACIÓN.

CURSO: BASES DE DATOS II IC-4302

I SEMESTRE 2023.

Prueba Corta 5 y 6

FECHA DE ENTREGA: 02/05/2023

Profesor:

Nereo Campos Araya.

Estudiante:

Diana Sanabria Calvo 2021436548

Índice

1. Pregunta 1	3
1.1. 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.	3
2. Pregunta 2	4
2.1. 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?	4
3. Pregunta 3	4

Bases de Datos II

Prueba Corta 5 y 6

Elaborado por Diana Sanabria Calvo.

1. Pregunta 1

- 1.1. 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 índice clustered es un tipo de índice en el que los datos de la tabla se ordenan físicamente en el mismo orden que las claves del índice. Los registros de la tabla se organizan en la misma estructura de árbol que el índice clustered. El índice clustered determina el orden físico de los datos en la tabla.

Un índice non-clustered es un índice que no afecta la organización física de los datos en la tabla. Este crea una estructura de índice separada que contiene las claves del índice y un puntero a la ubicación física de los datos en la tabla.

Podemos agregar una columna adicional a la estructura de un índice non-clustered gracias al INCLUDE, esto nos permite que las consultas utilicen esas columnas sin buscar primero en la tabla principal. Al tener los datos adicionales en el índice en lugar de tener que buscarlos en la tabla principal, aumentamos la eficiencia de las consultas.

2. Pregunta 2

2.1. 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?

El memory footprint se refiere al tamaño total de la memoria utilizada por un programa o proceso en un sistema. El memory footprint puede ser crítico en sistemas con limitaciones de memoria porque un alto consumo de memoria puede hacer al sistema lento o incluso provocar errores de memoria. La creación de índices puede afectar el memory footprint de la base de datos, porque cada índice creado consume una cantidad de memoria adicional y esto aumenta el consumo de memoria total de la base de datos. Un memory footprint alto también puede tener un impacto en la paginación a disco. Cuando un sistema tiene un alto memory footprint la cantidad de memoria disponible para la paginación a disco se reduce, lo que puede llevar a un aumento en los tiempos de acceso a disco y a una disminución en el rendimiento general del sistema.

3. Pregunta 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?

Yo usaría un índice para cada una de las columnas en este caso country, city y date para mejorar los resultados de búsqueda. Para el índice de country, se puede utilizar un índice hash, que permita la búsqueda directa de registros a través del código de country. Para el índice de city, se podría utilizar

un índice de árbol B, que es un tipo de índice que se utiliza para buscar valores en un rango y para el índice de date, se podría utilizar un índice de árbol B+ que permite una búsqueda eficiente de registros en un rango de fechas.

Yo le recomendaría a TooSlow utilizar una base de datos orientada a columnas como Cassandra o HBase para almacenar sus datos. Porque este tipo de base de datos es altamente escalable y eficiente a la hora de manejar grandes cantidades de datos. También son capaces de realizar búsquedas rápidas y eficientes en grandes conjuntos de datos y esto lo hace adecuado para el caso de TooSlow con 40 billones de registros.