

Bases de Datos Avanzadas Solemne 1 Abril/2022

1. Dado el siguiente modelo de datos,

LIBROS: Codigo, Nombre, Autor, Editorial, Valor_unitario BIBLIOTECAS: Cod_biblio, Cod_libro, Copias, Ciudad PRESTAMOS: Rut_lector, Cod_libro, Cod_biblio, Fecha_prestamo, Fecha_devol LECTORES: Rut, Nombre, Ciudad
--

Supuestos:

- Una ciudad puede tener más de una biblioteca
- Un lector no puede pedir más de un libro diario
- En promedio, un libro es solicitado por 5 lectores
- Para el 1% de los libros hay una sola copia disponible
- Los libros están disponibles en 3 bibliotecas en promedio

calcular el número de accesos requeridos para responder a la siguiente consulta

```
SELECT Lectores.ciudad, Libros.nombre
FROM Lectores, Prestamos, Bibliotecas, Libros
WHERE Bibliotecas.copias = 1
  AND Bibliotecas.cod_biblio = Prestamos.cod_biblio
  AND Prestamos.rut_lector = Lectores.rut
  AND Bibliotecas.cod_libro = Libros.codigo
```

* Indices: LIBROS.codigo, BIBLIOTECAS.(cod_biblio, cod_libro),
 PRESTAMOS.(cod_biblio, cod_libro, rut_lector), LECTORES.rut
* 200.000 Libros, 2.000.000 Préstamos, 20.000 Lectores
* Orden del árbol: 10 Filas por bloque: 50

1. Buscar copias = 1 en Bibliotecas y Obtener cod_biblio y cod_libro
--> 1% de 200.000 = 2.000 libros
2. Para cada par (cod_biblio, cod_libro) de (1)
 Buscar (cod_biblio, cod_libro) en Prestamos y Obtener rut_lector
 --> 2.000 libros * 5 lectores por libro = 10.000 rut_lector
3. Para cada rut_lector de (2), Buscar rut en Lectores y Obtener ciudad
 --> 10.000 * 1 = 10.000 ciudades
4. Para cada cod_libro de (1), Buscar codigo en Libros y Obtener nombre
 --> 2.000 * 1 = 2.000 nombres

$N1 = (200.000 * 3 \text{ bibliotecas por libro}) / 50 = 12.000$

$N2 = 2.000 * (\log_{10} 2.000.000 + 4 / 10 + 0) = 16.000$

$N3 = 10.000 * (\log_{10} 20.000 + 1) = 60.000$

$N4 = 2.000 * (\log_{10} 200.000 + 1) = 14.000$

$NT = 12.000 + 16.000 + 60.000 + 14.000 = 102.000 \text{ accesos}$

Plan de ejecución: 4 pasos * (2 + 1) puntos c/paso = 12 puntos

Numero de accesos: 4 cálculos * 2 puntos c/calculo = 8 puntos

Total pregunta: 20 puntos

2. Explique el siguiente plan de ejecución asociado a Personas2

```
explain delete from personas2 using ruts where personas2.rut = ruts.rut;
```

QUERY PLAN

```
-----
Delete on public.personas2 (cost=0.56..84816.00 rows=10000 width=12)
-> Nested Loop (cost=0.56..84816.00 rows=10000 width=12)
    Inner Unique: true
    -> Seq Scan on pg_temp_5.ruts (cost=0.00..155.00 rows=10000 width=22)
    -> Index Scan using personas2_pkey on public.personas2 (cost=0.56..8.47
        rows=1 width=14)
        Index Cond: (personas2.rut = ruts.rut)
```

1. El proceso comienza en el acceso secuencial por bloque (Seq Scan) en la tabla ruts, lo que tiene un costo de 155 unidades para traer las 10.000 filas de la tabla.
2. Por cada (Nested Loop) rut recuperado en (1), hace una búsqueda por índice (Index Scan) de ese rut en Personas2, lo que le toma 8.47 unidades de costo por cada búsqueda.
3. Finalmente, teniendo la dirección de los rut's en la tabla personas2 procede a eliminar las filas correspondientes, con un costo total de 84.816 unidades de costo.

La descripción correcta de cada uno de los 3 pasos obtiene 7 puntos.

Total pregunta: 20 puntos

3. Comente el siguiente párrafo justificando sus comentarios.

“Para facilitar la asociación entre las tablas de una base de datos, se recomienda definir un índice por cada llave foránea que tengan estas tablas. Al hacerlo así, todas las búsquedas van a ser efectuadas a escala logarítmica al trabajar solamente con los índices y no habrá necesidad de acceder a las tablas de datos para hacer este proceso. El beneficio que se obtiene es un menor tiempo de respuesta, dado que se está minimizando el número de accesos que se hace al dispositivo de almacenamiento.”

- No se debe definir índices en las llaves foráneas, pues dado un valor de una foránea, lo que se quiere es obtener algún dato asociado a ella en la tabla en la que es primaria.
- Al hacer búsquedas por índice, solo se va a obtener los valores del índice y, lo mas probable, es que se necesite ir a buscar algún dato de la tabla.
- Al indexar por la llave foránea, se van a generar valores repetidos, lo que implicará mas accesos para resolverlos. Esto va a impactar negativamente al tiempo de respuesta de la consulta.

Cada análisis válido (máximo tres) obtiene 7 puntos.

Total pregunta: 20 puntos