

## Bases de Datos

### Práctico 3: Operadores sobre Tablas y Ejecución de Planes de Evaluación

**Ejercicio 1:** definir una implementación para el operador producto cartesiano. Evaluar su costo en términos de transferencias de bloques y accesos de bloques.

**Ejercicio 2:** Para el operador físico de ordenación externa: escribir un algoritmo en pseudocódigo para la implementación de la combinación de corridas cuando  $N > M$ .

**Ejercicio 3:** Definir un operador que produce la unión de dos tablas sin repetición de tuplas en el resultado. Se pide:

- Hacer una definición recursiva del operador.
- ¿Como se puede definir el operador usando operaciones enseñadas?
- Sugerir algún operador físico para el operador y estimar su costo.

**Ejercicio 4:** Dadas las tablas  $r1(A, B, C)$  y  $r2(C, D, E)$  con las siguientes propiedades:  $r1$  tiene 20.000 tuplas,  $r2$  tiene 45.000 tuplas, 25 tuplas de  $r1$  caben en un bloque y 30 tuplas de  $r2$  que caben en un bloque. Estímese el número de transferencias de bloques y de accesos a bloques requeridos utilizando las siguientes estrategias para la reunión  $r1 \bowtie r2$ :

- a) Reunión en bucle anidado.
- b) Reunión en bucle anidado por bloques.
- c) Reunión por mezcla.

¿Y cuántos bloques se escriben? ¿Cambia este número la estrategia utilizada?

**Ejercicio 5:** Sea la siguiente tabla:

cliente(cid, cnombre, teléfono, dirección, edad)

Se tiene la siguiente información de cliente:

- cliente tiene 1400 tuplas y
- En cliente se tienen 70 valores distintos para la edad.
- La edad más grande es 87 y la más baja es de 18.
- La edad se asume equi-distribuida.
- Asumir que cliente tiene un índice primario en edad.

Además, asumir que un bloque tiene tamaño de 4096 bytes.

Sea la consulta:  $\sigma_{edad > 30}(\text{cliente})$

1. Si para el índice primario en edad se usa un árbol B+: ¿Cómo calcularías la cantidad de tuplas de ese índice que entran por bloque? ¿Cómo calcularías la cantidad de nodos en total que se necesitan para ese árbol B+?
2. Indicar el algoritmo más conveniente para implementar la selección explicando su comportamiento de manera precisa. Tiene que ser un operador eficiente.
3. Calcular el costo del algoritmo del ítem 2 en transferencias de bloques.

**Ejercicio 6:** Sea la tabla sucursal con esquema (nombre-sucursal, ciudad-sucursal, activos). Supóngase que hay un índice de árbol B+ disponible en ciudad-sucursal y que no hay más índices. ¿Cuál sería el mejor modo de manejar las siguientes selecciones con negaciones?

- a)  $\sigma \text{ not(ciudad-sucursal} < \text{"Arganzuela"}\text{)(sucursal)}$
- b)  $\sigma \text{ not(ciudad-sucursal} = \text{"Arganzuela"}\text{)(sucursal)}$
- c)  $\sigma \text{ not(ciudad-sucursal} < \text{"Arganzuela"} \parallel \text{activo} < 5000\text{)(sucursal)}$

**Ejercicio 7:** Genere los árboles sintácticos de expresiones, es decir, el plan de ejecución, para las siguientes consultas:

- a)  $\sigma \text{ precio} < 1000 (\Pi \text{ color; precio(pinturas)})$
- b)  $\Pi \text{ color, precio } (\sigma \text{ precio} < 1000(\text{pinturas}))$

Luego anote en los árboles generados el costo (en cantidad de bloques leídos y escritos) de cada operación, teniendo en cuenta los siguientes parámetros: La tabla pinturas tiene 10.000 registros, y cada bloque contiene 10 registros. Pero se pueden almacenar 100 registros conteniendo sólo color y precio. Se estima que 3/4 de las pinturas tienen un precio mayor o igual a 1000. Para ello, deberá indicar qué algoritmo utiliza. Considere las siguientes situaciones:

- a) Se tiene un índice primario en precio.
- b) Se tiene un índice secundario en precio.
- c) No se tiene un índice en precio.

Recuerde que la tabla temporal que escribe un operador no tiene índices.

**Ejercicio 8:** Realice las mismas tareas que las indicadas en el ejercicio 7 para las siguientes consultas:

- a)  $\Pi \text{ legajo, profe.nombre } (\sigma \text{ materia.nombre} = \text{"Intro a los Algoritmos"}) (\text{profe legajo} \bowtie \text{a\_cargo materia})$
- b)  $\Pi \text{ legajo, profe.nombre } (\text{profe legajo} \bowtie \text{a\_cargo } \sigma \text{ materia.nombre} = \text{"Intro a los Algoritmos"}) (\text{materia})$

La tabla materia tiene 100 registros, ocupando 20 bloques, mientras que profe tiene 2.000 registros, ocupando 500 bloques. Considere los diferentes escenarios según la existencia de índice (ninguno, primario, o secundario) en las columnas relevantes de las tablas.

**Ejercicio 9:** Calcule el costo total (de entrada y salida) de las siguientes consultas, utilizando variables para aquellos datos desconocidos. La tabla de las consultas es acceso\_dispositivo(usuario, fecha, dispositivo), y tiene un índice primario en usuario. Justifique el costo en cada parte describiendo brevemente qué haría el algoritmo en uso.

- a)  $v(\Pi \text{ dispositivo } (\sigma \text{ usuario} = \text{"raúl"}) (\text{acceso\_dispositivo}))$
- b)  $\sigma \text{ usuario} = \text{"Raúl"} (\text{acceso\_dispositivo}) ++ \sigma \text{ usuario} = \text{"Luisa"} (\text{acceso\_dispositivo})$