

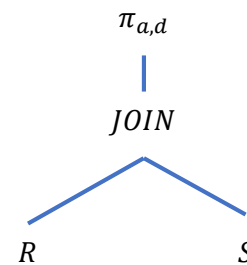
Relación de problemas II: “Optimización de consultas”

Ejercicio 1

Solución

Datos: $B = 2K, C = 20 B$, y la consulta es $\pi_{a,d}(R \times S)$, y la tabla con los datos de cada relación.

Como además indican que en la memoria sólo cabe un bloque, significa que **cada operación de E/S supone un bloque de disco**. Por tanto lo que nos piden en definitiva cuántos bloques totales se leen y se escriben para la consulta.



Vamos a hacerlo por partes y de forma ordenada, siguiendo la estrategia “abajo-arriba” del árbol algebraico (figura de la derecha). El resumen de las operaciones es el siguiente:

- i. Cargar las tablas R y S, necesarias para la reunión natural
- ii. Escribir el resultado de la reunión natural
- iii. Cargar el resultado del natural join, necesario para la proyección
- iv. Almacenar el resultado de la proyección.

Paso 1. Cargar la relación R

- $N(R) = 5000 B$
- $L(R) = 100 B$
- $Bfr = \left\lfloor \frac{B-C}{L(R)} \right\rfloor = \left\lfloor \frac{2048-20 [B]}{100 [B]} \right\rfloor = 20 \frac{\text{registros}}{\text{bloque}}$
- $B(R) = \text{entero superior} \left(\frac{5000}{20} \right) = 250 \text{ bloques hay que leer}$

Pero para que la reunión natural sea eficiente, ordenamos la relación R siguiendo el atributo común de ambas relaciones, **c**, para lo cual hay que **leer/cargar** adicionalmente $B(R)' = 250 \cdot \log_2 250 = 1992 \text{ bloques de R}$.

Paso 2. Cargar la relación S

- $N(S) = 200$
- $Bfr = \left\lfloor \frac{2048-20}{60} \right\rfloor = 33 \text{ registros/bloque}$
- $B(S) = \text{entero superior} \left(\frac{200}{30} \right) = 7 \text{ bloques para cargar S}$

Nota: suponemos que la clave primaria de la tabla S es c, que es el atributo común, por lo tanto no hará falta ordenar la tabla por dicho atributo. Si esta suposición no se realizase, bastaría con añadir al total de operaciones de E/S [o bloques] totales, el coste de ordenar S, que sería $B(S) \cdot \log_2 B(S)$.

Paso 3. Para la reunión natural, $R \bowtie S$

- Ya tenemos R y S cargados en memoria.
- $N(X) = \frac{N(R) \cdot N(S)}{\max(V(R,c), V(S,c))} = \frac{5000 \cdot 200}{\max 5,5} = 200\,000$ registros
- $L(X) = L(R) + L(S) - \text{size}(c) = 100 - 60 - 20 = 140$ B
- $Bfr = \left\lfloor \frac{2028}{140} \right\rfloor = 14$ registros/bloque
- $B(X) = \text{entero superior} \left(\frac{N(X)}{Bfr(X)} \right) = 14286$ bloques a escribir

Paso 4. Para la proyección, π

- Hay que leer **14286 bloques**, resultado de la reunión natural anterior, de disco.
- $N(X') = N(X) = 200000$ registros
- $L(X') = \text{size}(a) + \text{size}(d) = 20B + 40B = 60$ B
- $Bfr = \left\lfloor \frac{2028}{60} \right\rfloor = 33$ registros/bloque
- $B(X') = \text{entero superior} \left(\frac{200\,000}{33} \right) = 6061$ bloques [a escribir]

Paso final: cálculo del número total de bloques

Recapitulando, podemos calcular el total de operaciones de E/S como la suma de:

$$\text{operaciones E/S} = 250 + 1992 + 7 + 14286 \cdot 2 + 6061 = \mathbf{36882 \text{ operaciones E/S}}$$

Ejercicio 2

Solución

Teniendo en cuenta lo explicado, al crear una tabla se creará un segmento, y dentro de él, una extensión, y dentro de ésta, un bloque. La primera extensión ocupará 40 K.

Además, como sabemos, al crear una tabla, se le crea automáticamente un índice sobre su clave primaria, creando para él un segmento de índice, con una extensión (también de 40 K), y con un bloque.

Numéricamente:

$$\text{tamaño inicial} = 40\,K + 40\,K = \mathbf{80\,K}$$



Ejercicio 5

Solución

- a) **Falso.** Depende realmente de: (i) tiempo necesario para ordenar el fichero de desbordamiento; (ii) tiempo en leer todos los registros más el de escribir sólo los válidos; (iii) tiempo necesario para recrear el índice.
- b) **Verdadero.** Esto sucede porque el acceso al siguiente registro es inminente estando posicionado en el actual cuando se accede por el mismo índice.