

Capítulo 4

Procesamiento de consultas

Ejemplo de materialización

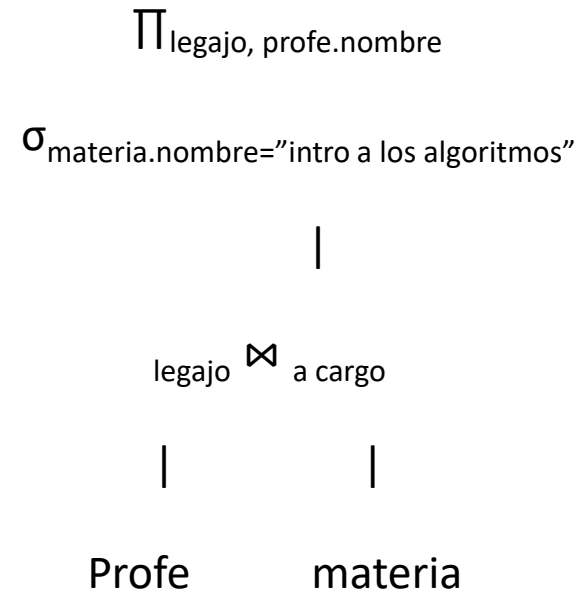
Caso de estudio

- Sea la consulta:

Π legajo, profe.nombre (σ materia.nombre = “Intro a los Algoritmos” (profe legajo \bowtie a_cargo materia))

- La tabla materia tiene 100 registros, ocupando 20 bloques.
- La tabla profe tiene 2.000 registros, ocupando 500 bloques.

Árbol binario de ejecución



Factor de selectividad

- $r = \text{profe}_{\text{legajo}} \bowtie_{\text{a cargo}} \text{materia}$
- $S = \sigma_{\text{materia.nombre} = \text{"intro a los algoritmos"}} r$
- $\text{fs}(\text{profe.legajo} = \text{materia.a_cargo}, \text{profe}, \text{materia}) = 1/|\text{profe}| = 1/2000$
- porque a_cargo clave foránea en materia referenciando a profe.
- Si B clave foránea en s referenciando r: $\text{fs}(r.A = s.B, r, s) = 1/|r|$
- $\text{fs}(A = c, r) = 1/V(A, r)$, donde $V(A, r)$ número de distintos valores que aparecen en r para A .
- $\text{fs}(\text{nombre_materia} = \text{'intro a los algoritmos'}, s) = 1/V(\text{materia.nombre}, s) = 1/100$
- Aquí asumo uniformidad

Operadores físicos

- **Reunión selectiva**

- Algoritmo de reunión merge-sort: ordenamos profe y materia según *legajo, a cargo*.
- Asumimos que todos los bloques entran en memoria.
- Los otros algoritmos usando índices son peores.

- **Selección**

- Como opera sobre resultado intermedio, no usa índices.
- Consideramos algoritmo de búsqueda lineal con selección de igualdad para atributo.

- **Proyección**

- Requiere recorrer todos los registros y realizar una proyección en cada uno.
- Se recorren todos los bloques de la tabla.

Tamaño en bloques de las tablas

- Según el enunciado materia tiene 20 bloques y profe tiene 500 bloques.

Tamaño de los resultados intermedios

- $r = \text{profe}_{\text{legajo}} \bowtie_{\text{a cargo}} \text{materia}$
- $|r| = |\text{profe}| * |\text{materia}| * \text{fs}(\text{profe.legajo} = \text{materia.a_cargo}, r) = 2000 * 100 * 1/2000 = 100$
- Hay 500 bloques de profe y hay 2000 profes: en un bloque entran: 4 materias por bloque.
- Hay 20 bloques de materia y hay 100 materias: en un bloque entran 5 materias.
- Cuantas tuplas de la reunión selectiva entran por bloque. Van a entrar 2 tuplas de r por bloque.
- $B_r = 100 / 2 = 50$
- $S = \sigma_{\text{materia.nombre} = \text{"intro a los algoritmos"}} r$
- $|s| = |r| * \text{fs}(\text{materia.nombre} = \text{'intro_algoritmos'}, r) = 100 * 1/100 = 1$
- $B_s = 1$

Costo de los operadores físicos

- **Reunión selectiva**

- Costo en cantidad de transferencia de bloques sin contar ordenación:
- $b_r + b_s = 500 + 20 = 520$
- Por lo tanto, el costo es 520 transferencias de bloques.

- **Selección**

- Costo $b_r = 50$ transferencias de bloques
- b_r denota el número de bloques conteniendo registros de la tabla r

- **Proyección**

- Estimación de costo = b_r transferencias de bloques + 1 acceso a bloque
- b_r denota el número de bloques conteniendo registros de la tabla r
- *1 transferencia de bloque*

Sumar los costos totales

- Costo total = costo de operaciones + costo de materialización
- $= (520 + 50 + 1) + (50 + 1) = 622$ transferencias de bloques