

Ejercicio 1:

Dadas las tablas:

persona (nombre, DNI, edad)

trabajaEn(DNI, empresa, antigüedad)

Considerar la consulta: $\text{persona} \bowtie (\vee (\Pi_{\text{DNI}} \text{trabajaEn}))$

Asumir que trabajaEn tiene 100.000 tuplas y ocupa 2000 bloques. Asumir que persona tiene 200.000 tuplas. Asumir que $V(\text{trabajaEn}, \text{DNI}) = 50000$. Asumir bloques de 4096 bytes. Asumir que solo entran 12 bloques en memoria.

Pueden asumir que la proyección se guarda en una organización de archivo secuencial.

Obtener el costo de algoritmo para operador eliminación de duplicados (peor caso) y también calcular el tamaño en número de bloques de resultado intermedio luego de computar ese operador.

Consulta:

$\text{persona} \bowtie (\vee (\pi_{\text{DNI}} \text{trabajaEn}))$

Datos:

trabajaEn: 100.000 registros , 2000 bloques

Persona: 200.000 registros

$V(\text{trabajaEn}, \text{DNI}) = 50.000$

bloques de 4096 bytes

12 bloques en memoria

$\pi_{\text{DNI}} \text{trabajaEn}$ trae 100.000 registros

Proyección:

Por algoritmo de proyección, sabemos que vamos a tener que recorrer todos los registros y realizar proyección en cada uno

Costo proyeccion = br transferencias de bloques + 1 acceso a bloque

br denota el número de bloques conteniendo registros de la tabla r

Costo proyección = 2001 (2000 transferencias de bloques + 1 acceso a bloque)

Eliminación de duplicados:

La operación de eliminación de duplicados pide que ordenemos la tabla usando external merge sort.

$M = 12$ (12 bloques en memoria por consigna)

$$\begin{aligned}
 \text{costo del ordenamiento} &= br * (2 * \lceil \log_{M-1} (br / M) \rceil + 1) \\
 &= 2000 * (2 * \lceil \log_{12-1} (2000 / 12) \rceil + 1) \\
 &= 2000 * (2 * \lceil 2,13 \rceil + 1) \\
 &= 2000 * (2 * 3 + 1) \\
 &= \mathbf{14.000 \text{ transferencias de bloque}}
 \end{aligned}$$

Costo total de la eliminación de duplicados = $14.000 + 2001 = 16.001$

transferencias de bloque.

Sabemos que la proyección nos devuelve 100.000 registros, ya que DNI es clave primaria de trabajaEn y tenemos 100.000 registros.

También sabemos por el dato de $V(\text{DNI}, \text{trabajaEn}) = 50.000$ que hay 50.000 DNIs únicos, por lo que la remoción de duplicados nos deja 50.000 registros en la tabla intermedia.

Sabemos que hay 100.000 registros y 2000 bloques, por lo que tenemos 50 registros máximo por bloque.

Tenemos 50.000 registros resultantes en la tabla intermedia, por lo que necesitamos $50.000 / 50 = 1000$ bloques.

Asumimos que el atributo DNI es un int de 4 bytes, por lo que cuando nosotros hagamos la proyección, vamos a tener 100.000 registros, lo cual multiplicado por el tamaño de DNI nos deja $100.000 * 4 = 400.000$ bytes.

Cuando hagamos la remoción de duplicados, vamos a tener $50.000 * 4 = 200.000$ bytes.

Como cada bloque es de 4096 bytes, vamos a necesitar $200.000 / 4096 = 48,82$ ósea **49 bloques**.