

Resolución de la Prueba Parcial ABD
30 de Abril de 2015

1. **(1 punto)** ¿De qué depende el tiempo necesario para reorganizar un Archivo Secuencial Lógico?
-

La reorganización de un Archivo Secuencial Lógico (ASL) sólo se produce si hay zona de desbordamiento, puesto que la zona principal ya está ordenada.

El tiempo que se tarda en esta operación depende del tiempo que se tarda en ordenar los registros de la zona de desbordamiento y en hacer un *merge* (mezcla ordenada) de los registros de la zona principal y de la zona de desbordamiento ordenada en un nuevo fichero.

$$T_Y = T_C(O) + (n+O) \cdot T + (n+O-d) \cdot T_w$$

El primer sumando es el tiempo de ordenación de la zona de desbordamiento, el segundo sumando representa la lectura de los registros de la zona principal y de la zona de desbordamiento ordenada y el tercer sumando la escritura del resultado de la mezcla de los registros de la zona principal (que no hayan sido borrados) y de la zona de desbordamiento ordenada.

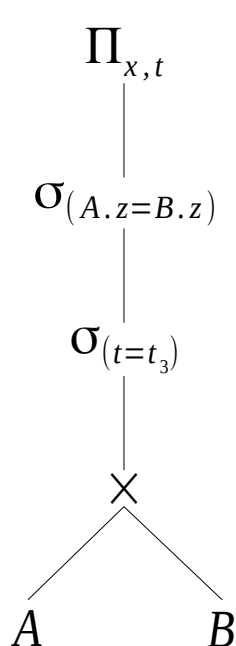
2. (2 puntos) Sean las relaciones A y B con los siguientes parámetros:

A(x,y,z)	B(z,t,u)
N(A) = 1000	N(B)=100
Size(x)=20	
Size(y)=50	
Size(z)=15	Size(z)=15
	Size(t)=4
	Size(u)=20
V(A,z) = 5	V(B,z)=3
	V(B,t)=5

Teniendo en cuenta que el tamaño de bloque es de 4KB, que la cabecera es de 20B, que se usa bloqueo fijo, que los bloques son homogéneos y que en memoria sólo cabe un bloque, determina el número de operaciones de E/S que supondría la ejecución de la consulta si el plan lógico fuera tal y como plantea la consulta:

$$\Pi_{x,t}(\sigma_{(A.z=B.z) \wedge (t=t_3)}(A \times B))$$

Para calcular de forma aproximada el número de bloques que es necesario transferir (lectura y escritura) durante la resolución de esta consulta, suponiendo que el plan lógico coincide exactamente con la misma, es necesario estimar el número de bloques que es necesario transferir para cada una de sus operaciones.



Empecemos por calcular el número de bloques que tiene la relación A para poder realizar la operación de producto cartesiano:

$$B(A) = \text{hacia arriba}(\frac{N(A)}{Bfr(A)}) = \text{hacia arriba}(\frac{1000}{Bfr(A)})$$

$$Bfr(A) = \text{parte entera}(\frac{B-C}{L(A)}) = \text{parte entera}(\frac{4096-20}{L(A)})$$

$$L(A) = 20 + 50 + 15 = 85$$

de modo que:

$$Bfr(A) = \text{parte entera}(\frac{4096-20}{85}) = 47 \quad \text{y}$$

$$B(A) = \text{hacia arriba}(\frac{1000}{47}) = 22$$

Del mismo modo, tenemos que calcular el número de bloques de B para poder realizar la operación de producto cartesiano:

$$B(B) = \text{hacia arriba} \left(\frac{N(B)}{Bfr(B)} \right) = \text{hacia arriba} \left(\frac{100}{Bfr(B)} \right)$$

$$Bfr(B) = \text{parte entera} \left(\frac{B-C}{L(B)} \right) = \text{parte entera} \left(\frac{4096-20}{L(B)} \right)$$

$$L(B) = 15 + 4 + 20 = 39$$

de modo que:

$$Bfr(B) = \text{parte entera} \left(\frac{4096-20}{39} \right) = 104 \quad y$$

$$B(B) = \text{hacia arriba} \left(\frac{100}{104} \right) = 1$$

La operación de producto cartesiano combina cada registro de A con todos los registros de B (cada registro de cada bloque de A con todos los registros de todos los bloques de B). Para cada bloque de A tiene que leer todos los bloques de B, por lo que el número de lecturas de bloques para la operación es de:

$$\text{Lectura: } B(A) \cdot B(B) = 1 \cdot 22 = 22$$

y el resultado del producto cartesiano necesita un número de bloques suficiente para almacenar los registros resultantes de esta operación, o sea:

$$B(\times) = \text{hacia arriba} \left(\frac{N(\times)}{Bfr(\times)} \right) = \text{hacia arriba} \left(\frac{1000 \cdot 100}{Bfr(\times)} \right)$$

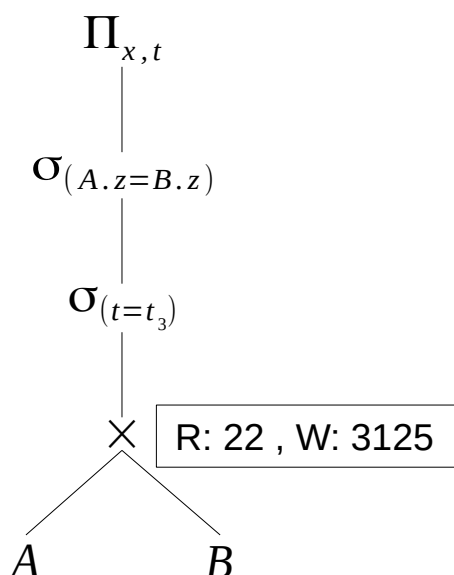
$$Bfr(\times) = \text{parte entera} \left(\frac{B-C}{L(\times)} \right) = \text{parte entera} \left(\frac{4096-20}{L(\times)} \right)$$

$$L(\times) = L(A) + L(B) = 85 + 39 = 124$$

de modo que:

$$Bfr(\times) = \text{parte entera} \left(\frac{4096-20}{124} \right) = \text{parte entera} \left(\frac{4096-20}{124} \right) = 32 \quad y$$

$$\text{Escritura: } B(\times) = \text{hacia arriba} \left(\frac{1000 \cdot 100}{Bfr(\times)} \right) = \text{hacia arriba} \left(\frac{1000 \cdot 100}{32} \right) = 3125$$



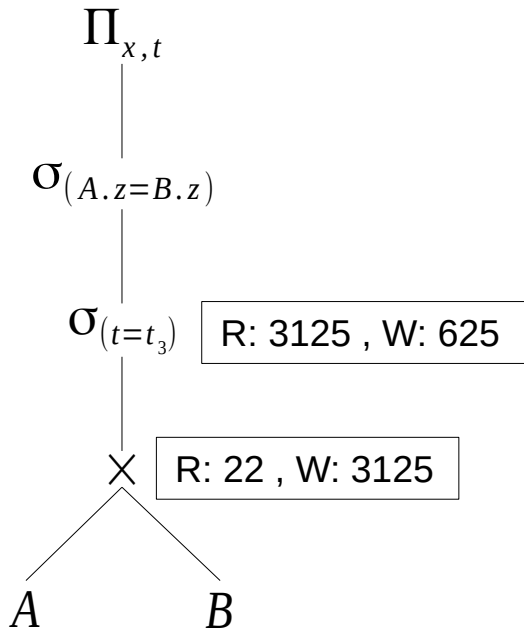
La primera operación de selección debe leer los bloques resultantes de la operación anterior y escribir tantos bloques como sean necesarios para almacenar los registros que resultan de seleccionar aquellos que cumplan la condición, es decir:

$$\begin{aligned} B(\sigma_{t=t_3}) &= \text{hacia arriba} \left(\frac{N(\sigma_{t=t_3})}{Bfr(\sigma_{t=t_3})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{t=t_3})}{Bfr(\times)} \right) = \\ &= \text{hacia arriba} \left(\frac{N(\sigma_{t=t_3})}{32} \right) \end{aligned}$$

$$N(\sigma_{t=t_3}) = \frac{N(\times)}{V(B,t)} = \frac{100 \cdot 1000}{5} = 20000$$

de modo que:

$$\text{Escritura: } B(\sigma_{t=t_3}) = \text{hacia arriba}(\frac{N(\sigma_{t=t_3})}{32}) = \text{hacia arriba}(\frac{2000}{32}) = 625$$



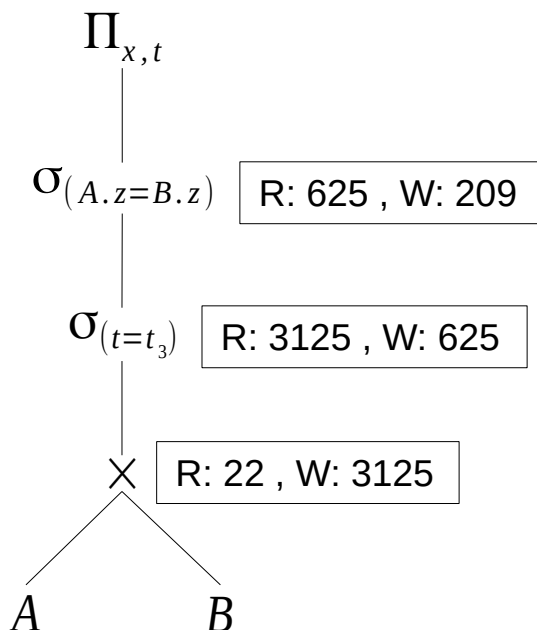
La segunda operación de selección debe leer los bloques resultantes de la operación anterior y escribir tantos bloques como sean necesarios para almacenar los registros que resultan de seleccionar aquellos que cumplan la condición, es decir:

$$\begin{aligned} B(\sigma_{A.z=B.z}) &= \text{hacia arriba}(\frac{N(\sigma_{A.z=B.z})}{Bfr(\sigma_{A.z=B.z})}) = \\ &= \text{hacia arriba}(\frac{N(\sigma_{A.z=B.z})}{Bfr(\sigma_{t=t_3})}) = \text{hacia arriba}(\frac{N(\sigma_{A.z=B.z})}{Bfr(\sigma_x)}) = \\ &= \text{hacia arriba}(\frac{N(\sigma_{A.z=B.z})}{32}) \\ N(\sigma_{A.z=B.z}) &= \frac{N(\sigma_{t=t_3})}{\min(V(A,z), V(B,z))} = \frac{20000}{3} \approx 6667 \end{aligned}$$

El hecho de calcular el número de registros de esta selección mediante el uso del mínimo entre las variabilidades del atributo z en ambas relaciones, se debe a que al realizar esta selección, sólo nos quedamos con aquellas parejas de registros que tienen igual los valores de los dos atributos z. Hay dos valores de z en A, pongamos z4 y z5, que no están presentes en B, por lo que no habría tuplas en el resultado que cumplieran la condición.

De este modo, se escriben:

$$B(\sigma_{A.z=B.z}) = \text{hacia arriba}(\frac{N(\sigma_{A.z=B.z})}{32}) = \text{hacia arriba}(\frac{6667}{32}) = 209$$



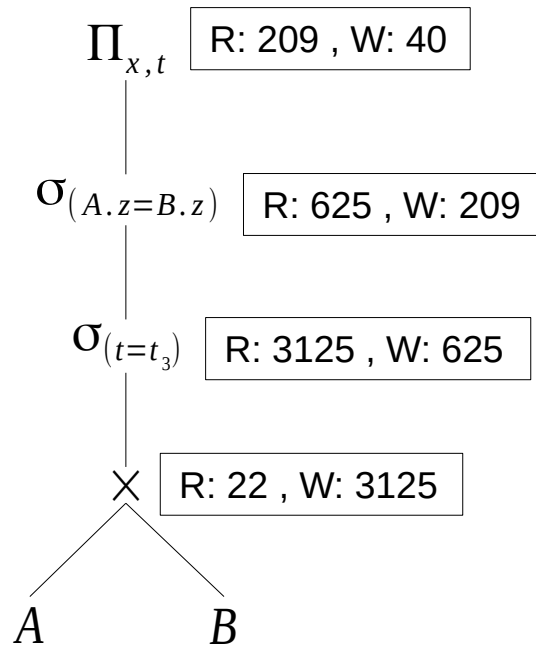
Por último, la operación de proyección debe leer los bloques resultantes de la operación anterior y escribir tantos bloques como sean necesarios para contener los registros resultantes de la operación de proyección:

$$\begin{aligned} B(\Pi_{x,t}) &= \text{hacia arriba}(\frac{N(\Pi_{x,t})}{Bfr(\Pi_{x,t})}) = \\ &= \text{hacia arriba}(\frac{N(\sigma_{A.z=B.z})}{Bfr(\Pi_{x,t})}) = \text{hacia arriba}(\frac{6667}{Bfr(\Pi_{x,t})}) \\ Bfr(\Pi_{x,t}) &= \text{parte entera}(\frac{B-C}{L(\Pi_{x,t})}) = \text{parte entera}(\frac{4096-20}{L(\Pi_{x,t})}) \\ L(\Pi_{x,t}) &= 20+4=24 \end{aligned}$$

de modo que:

$$Bfr(\Pi_{x,t}) = \text{parte entera} \left(\frac{4096 - 20}{L(\Pi_{x,t})} \right) = \text{parte entera} \left(\frac{4096 - 20}{24} \right) = 169 \text{ y}$$

$$\text{Escritura: } B(\Pi_{x,t}) = \text{hacia arriba} \left(\frac{6667}{Bfr(\Pi_{x,t})} \right) = \text{hacia arriba} \left(\frac{6667}{169} \right) = 40$$



Una vez estimados el número de bloques que cada operación lee y escribe, podemos estimar que el número total de bloques leídos y escritos para la resolución del plan físico correspondiente es de:

$$(22+3125)+(3125+625)+(625+209)+(209+40)=$$

$$=7980 \text{ bloques}$$

3. **(3 puntos)** Propón un plan lógico para la consulta del ejercicio anterior cuyo plan físico sea más eficiente que el calculado para dicho ejercicio y justifica numéricamente tu respuesta. Ten en cuenta que hay varios planes posibles por lo que la calificación de esta pregunta dependerá de cuán óptima sea tu propuesta.

Un plan posible para una mejor ejecución de la consulta, consistiría en realizar proyecciones lo antes posible, a fin de eliminar atributos que no estén involucrados en el resultado ni en las operaciones intermedias. Para ello, podemos bajar la proyección hacia las hojas del árbol de expresión algebraica. Sin embargo, al realizar la proyección sobre la relación A, nos vemos obligados a incluir el atributo z a fin de poder realizar la operación de reunión natural. Lo mismo ocurre con la proyección sobre la relación B, para poder garantizar que se realice la operación de reunión natural sobre el atributo común z.

A fin de estimar el número de bloques necesarios para la resolución de esta consulta, tendremos en cuenta tantos cálculos realizados durante el ejercicio anterior como sea posible ($B(A)$, $B(B)$, $Bfr(A)$, $Bfr(B)$, ...).

Resolveremos la estimación recorriendo el árbol mediante un recorrido en post-orden (antes de resolver un nodo, es necesario resolver su hijo a la izquierda y su hijo a la derecha). Siguiendo esta lógica, es necesario estimar, en primer lugar, el número de bloques necesario para almacenar la proyección sobre A, después de haber leído los bloques de la relación A:

$$B(\Pi_{x,z}) = \text{hacia arriba} \left(\frac{N(\Pi_{x,z})}{Bfr(\Pi_{x,z})} \right) = \text{hacia arriba} \left(\frac{N(A)}{Bfr(\Pi_{x,z})} \right) = \text{hacia arriba} \left(\frac{1000}{Bfr(\Pi_{x,z})} \right)$$

$$Bfr(\Pi_{x,z}) = \text{parte entera} \left(\frac{B-C}{L(\Pi_{x,z})} \right) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{x,z})} \right)$$

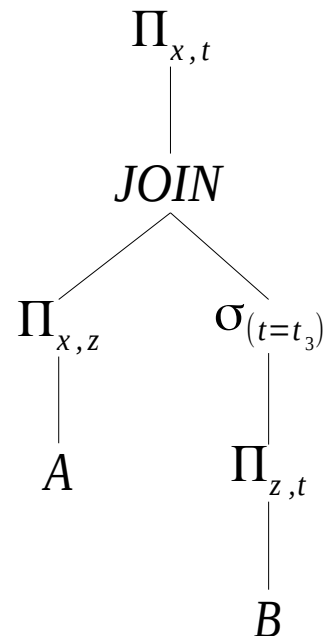
$$L(\Pi_{x,z}) = 20 + 15 = 35$$

de modo que:

$$Bfr(\Pi_{x,z}) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{x,z})} \right) = \text{parte entera} \left(\frac{4096-20}{35} \right) = 116 \quad y$$

$$\text{Escritura: } B(\Pi_{x,z}) = \text{hacia arriba} \left(\frac{1000}{Bfr(\Pi_{x,z})} \right) = \text{hacia arriba} \left(\frac{1000}{116} \right) = 9$$

Realizamos los mismos cálculos para estimar el número de bloques requeridos para realizar la operación de proyección sobre B después de haber leído los bloques de la relación B:



$$B(\Pi_{z,t}) = \text{hacia arriba} \left(\frac{N(\Pi_{z,t})}{Bfr(\Pi_{z,t})} \right) = \text{hacia arriba} \left(\frac{N(B)}{Bfr(\Pi_{z,t})} \right) = \text{hacia arriba} \left(\frac{100}{Bfr(\Pi_{z,t})} \right)$$

$$Bfr(\Pi_{z,t}) = \text{parte entera} \left(\frac{B-C}{L(\Pi_{z,t})} \right) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{z,t})} \right)$$

$$L(\Pi_{z,t}) = 15 + 4 = 19$$

de modo que:

$$Bfr(\Pi_{z,t}) = \text{parte entera} \left(\frac{4096-20}{19} \right) = \text{parte entera} \left(\frac{4096-20}{19} \right) = 214 \quad y$$

$$\text{Escritura: } B(\Pi_{z,t}) = \text{hacia arriba} \left(\frac{100}{214} \right) = \text{hacia arriba} \left(\frac{100}{214} \right) = 1$$

Antes de realizar la operación de reunión natural, es necesario aplicar una operación de selección sobre la proyección realizada sobre la relación B. Dicha operación requiere leer los bloques resultantes de la operación de proyección mencionada y la escritura de tantos bloques como sea necesario para almacenar los registros resultantes de realizar la selección de aquellos registros que cumplan la condición, es decir:

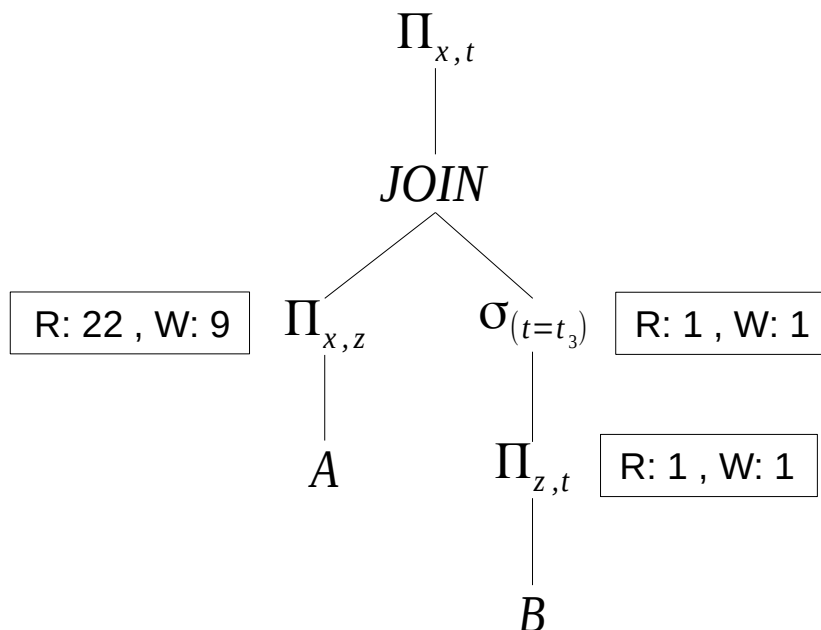
$$B(\sigma_{t=t_3}) = \text{hacia arriba} \left(\frac{N(\sigma_{t=t_3})}{Bfr(\sigma_{t=t_3})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{t=t_3})}{Bfr(\Pi_{z,t})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{t=t_3})}{214} \right)$$

$$N(\sigma_{t=t_3}) = \frac{N(\Pi_{z,t})}{V(B,t)} = \frac{N(B)}{3} = \frac{100}{3} \approx 34$$

de modo que:

$$\text{Escritura: } B(\sigma_{t=t_3}) = \text{hacia arriba} \left(\frac{N(\sigma_{t=t_3})}{214} \right) = \text{hacia arriba} \left(\frac{34}{214} \right) = 1$$

Recopilando lo estimado hasta ahora:



El algoritmo más eficiente para realizar la operación de reunión natural requiere que ambas relaciones estén ordenadas por el atributo de reunión (z en este caso). Si entendemos que el primer atributo de cada relación es la clave de la misma, y que una relación está ordenada por su clave, la relación A (y, por tanto, tampoco la proyección realizada sobre ella) no estaría ordenada por su el atributo z y sería necesario obtener una relación con los mismos registros que la proyección sobre la relación A pero con los registros ordenados por los valores del atributo z.

Como sabemos, el algoritmo de ordenación más eficiente, tiene que realizar un número de lecturas y escrituras de bloques de:

Lecturas y escrituras: $B(\Pi_{x,z}) \cdot \log_2(B(\Pi_{x,z})) = 9 \cdot \log_2(9) \approx 29$

Después de esta ordenación, la proyección natural leería los bloques de la relación de la izquierda (ya ordenada por z) y de la relación de la derecha, para escribir tantos bloques como sea necesario para almacenar los registros resultantes de la reunión natural, es decir:

$$B(JOIN) = \text{hacia arriba} \left(\frac{N(JOIN)}{Bfr(JOIN)} \right)$$

$$N(JOIN) = \frac{N(\Pi_{x,z}) \cdot N(\sigma_{t=t_3})}{\max(V(A,z), V(B,z))} = \frac{N(A) \cdot N(\sigma_{t=t_3})}{\max(V(A,z), V(B,z))} = \frac{1000 \cdot 34}{5} = 6800$$

Como podemos ver, el número de registros resultantes de la reunión natural no concuerda con el número de registros resultantes del producto cartesiano y posterior selección (para realizar el mismo efecto de la reunión natural) del ejercicio anterior. Dicha diferencia se debe al redondeo que produjimos en el número de tuplas de la selección sobre B:

$$N(\sigma_{t=t_3}) = \frac{N(\Pi_{z,t})}{V(B,t)} = \frac{N(B)}{3} = \frac{100}{3} = 33,333... \approx 34$$

Si tomamos en consideración una mayor precisión del resultado y no el aproximado:

$$N(JOIN) = \frac{1000 \cdot 33,333...}{5} = 6666,666... \approx 6667$$

descubrimos que los resultados finales concuerdan en ambos planes físicos.

Ahora, es necesario calcular el factor de bloqueo del resultado de la reunión natural:

$$Bfr(JOIN) = \text{parte entera} \left(\frac{B-C}{L(JOIN)} \right) = \text{parte entera} \left(\frac{4096-20}{L(JOIN)} \right)$$

$$L(JOIN) = 20 + 15 + 4 = 39$$

de modo que:

$$Bfr(JOIN) = \text{parte entera} \left(\frac{4096-20}{39} \right) = 104 \quad y$$

Escritura: $B(JOIN) = \text{hacia arriba} \left(\frac{N(JOIN)}{Bfr(JOIN)} \right) = \text{hacia arriba} \left(\frac{6667}{104} \right) = 64$
--

Por último, la operación de proyección final supondrá la lectura de los bloques

almacenados por la operación de reunión natural y la posterior escritura de tantos bloques como sean necesarios para almacenar los registros resultantes de dicha proyección, es decir:

$$B(\Pi_{x,t}) = \text{hacia arriba} \left(\frac{N(\Pi_{x,t})}{Bfr(\Pi_{x,t})} \right) = \text{hacia arriba} \left(\frac{N(JOIN)}{Bfr(\Pi_{x,t})} \right) = \text{hacia arriba} \left(\frac{6667}{Bfr(\Pi_{x,t})} \right)$$

$$Bfr(\Pi_{x,t}) = \text{parte entera} \left(\frac{B-C}{L(\Pi_{x,t})} \right) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{x,t})} \right)$$

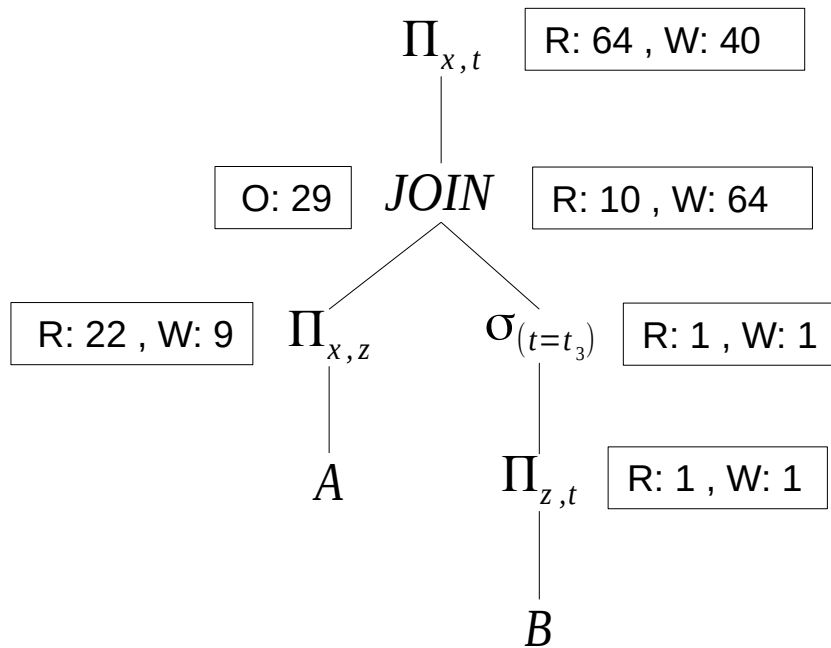
$$L(\Pi_{x,t}) = 20 + 4 = 24$$

de modo que:

$$Bfr(\Pi_{x,t}) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{x,t})} \right) = \text{parte entera} \left(\frac{4096-20}{24} \right) = 169 \quad y$$

$$\text{Escritura: } B(\Pi_{x,t}) = \text{hacia arriba} \left(\frac{6667}{Bfr(\Pi_{x,t})} \right) = \text{hacia arriba} \left(\frac{6667}{169} \right) = 40$$

Una vez estimados el número de bloques que cada operación lee y escribe,



podemos estimar que el número total de bloques leídos y escritos para la resolución del plan físico correspondiente es de:

$$(22+9)+(1+1)+(1+1)+(29+10+64)+(64+40)=242 \text{ bloques}$$

Como podemos observar, el número de bloques leídos y escritos por este plan lógico es mucho mejor que el planteado en el resultado del ejercicio 2.

4. Según la consulta del ejercicio 2 (optimizada en el ejercicio 3), responde a las siguientes preguntas de la forma más breve posible:
- (1 punto)** ¿Sería interesante tener algún índice en la tabla B que mejorara la ejecución de alguna parte de la consulta planteada? ¿o sería contraproducente?; justifica tu respuesta.
 - (2 punto)** Estima qué estructura de índice haría falta para el atributo t de la tabla B del ejercicio 2 y calcula su tamaño en bloques si $P = 10B$ y el resto de los parámetros son los mismos que los del ejercicio 2.
-

Apartado a)

En el plan planteado en el ejercicio 2, de tener un índice sobre el atributo t de la relación B ahorraríamos tiempo en la selección sobre dicha relación pero, puesto que la selección no se realiza sobre dicha relación sino sobre la relación resultante de otra operación, dicho índice no tiene utilidad.

En el plan planteado en el ejercicio 3, se aplica el mismo razonamiento sobre un posible índice sobre el atributo t de la relación B. Podría también ser útil un índice sobre el atributo z de la relación A, para ahorrarnos el ordenar la relación A por dicho atributo para la reunión natural. Pero dicha reunión no se aplica sobre la relación A sino sobre otra resultante de una proyección, así que tampoco tendría utilidad alguna.

Apartado b)

Dada la variabilidad del atributo t en la relación B (5), y que dicha relación contiene 100 registros, no podemos más que asumir que hay varios registros de B con el mismo valor de t . Si asumimos una distribución uniforme de los valores de t en los 100 registros, tendríamos 20 registros para cada uno de los 5 valores distintos de t .

Para que un índice sea útil, tiene que contener valores distintos, pero sólo hay 5 valores distintos y 100 registros.

A fin de localizar cada uno de los 5 valores, construimos un índice que contendrá dichos 5 valores y, para cada valor, una referencia a un bloque de una segunda estructura de índice, que contendrá todos los registros con dicho valor para el atributo t .

El primer índice necesita un número de bloques igual a:

$$B(I_1) = \text{hacia arriba} \left(\frac{N(I_1)}{Bfr(I_1)} \right)$$

$$N(I_1) = V(B, t) = 5$$

$$Bfr(I_1) = \text{parte entera} \left(\frac{4096 - 20}{L(I_1)} \right)$$

$$L(I_1) = 4 + 10 = 14$$

de modo que:

$$Bfr(I_1) = \text{parte entera}(\frac{4096-20}{14}) = 291 \quad y$$

$$B(I_1) = \text{hacia arriba}(\frac{N(I_1)}{Bfr(I_1)}) = \text{hacia arriba}(\frac{5}{291}) = 1$$

Para el segundo nivel de índice, al menos habrá un bloque para contener a todos los registros con igual valor de t y, dado que hay 5 valores de t , hay como mínimo 5 bloques. Veamos cuántos bloques hacen falta para almacenar todos los registros con el mismo valor de t , teniendo en cuenta que ambos índices están contruidos sobre el atributo t por lo que tienen la misma longitud de registro y el mismo factor de bloqueo:

$$\begin{aligned} B(I_2) &= v(B, t) \cdot \text{hacia arriba}(\frac{N(\text{igual valor de } t)}{Bfr(I_2)}) = 5 \cdot \text{hacia arriba}(\frac{N(\text{igual valor de } t)}{Bfr(I_1)}) = \\ &= 5 \cdot \text{hacia arriba}(\frac{N(\text{igual valor de } t)}{291}) \end{aligned}$$

$$N(\text{igual valor de } t) = \frac{N(B)}{V(B, t)} = \frac{100}{5} = 20$$

de modo que:

$$B(I_2) = 5 \cdot \text{hacia arriba}(\frac{N(\text{igual valor de } t)}{291}) = 5 \cdot \text{hacia arriba}(\frac{20}{291}) = 5$$

En total, haría falta 1 bloque para el primer nivel de índice y 5 bloques para el segundo:

$B(I) = B(I_1) + B(I_2) = 1 + 5 = 6 \text{ bloques}$
--

5. **(1 puntos)** Indica los distintos tipos de segmentos que se pueden encontrar en los datafiles.

Son de cuatro tipos:

- de datos (tablas)
- de índice (índices)
- temporales (resultados intermedios de order by, group by, union, intersect, minus, ...)
- rollback (valores antiguos de datos en update)