

Resolución de la Prueba Parcial ABD

Abril de 2017

Pregunta 1

Enunciado

Indica de qué factores depende el tiempo que se tarda en re-organizar un fichero ASI con zona de desbordamiento (no es necesario poner la fórmula, si no la recuerdas).

Resolución

Un archivo secuencial indexado tiene dos componentes:

- el **fichero maestro**: en el que se almacenan los registros, y
- el **fichero de índice**: en el que se almacena la estructura que permite localizar más eficientemente la posición de un registro en el fichero maestro, en base a un valor de clave y dado que está ordenado por valor de la misma.

A su vez, el fichero maestro está dividido en una zona ordenada con una organización ASL y otra zona con organización ASF (no ordenada) conocida como *zona de desbordamiento*.

Para re-organizar un fichero con organización ASI, es necesario:

- Ordenar mediante algún algoritmo de ordenación la *zona de desbordamiento*, lo cual empleará un tiempo T_O .
- Una vez ordenada, se pueden fusionar las dos zonas ordenadas mediante un algoritmo de tipo *merge*, que tardará un tiempo T_M .
- Finalmente, podemos reconstruir un índice nuevo a partir del total del fichero maestro ordenado, recorriéndolo y anotando clave y posición de cada registro (en el caso de un índice denso) o clave y posición de página (en el caso de un índice no denso), lo cual tardará un tiempo T_I .

En resumen, el tiempo para re-organizar un fichero ASI con zona de desbordamiento tendría una expresión:

$$T_O + T_M + T_I$$

o una expresión similar.

En concreto, en las transparencias de clase aparece la fórmula:

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

donde:

- T_O consiste en ordenar la zona de desbordamiento (con un número de registros O),
- T_M consiste en leer todos los registros de la zona ordenada (con un número de registros n) y de desbordamiento ya ordenada (con un número de registros O), realizar el *merge* de ambas pero sólo escribir los no borrados (con un número d de registros borrados), y
- T_I consiste en recorrer el nuevo fichero maestro (con los registros que se quedan $n+O-d$) y escribir los pares clave-dirección en el fichero de índice.

Pregunta 2

Enunciado

Indica que son *segmentos* y *extensiones* en Oracle®. Justifica la utilidad de esta organización de los datos.

Resolución

Una *extensión* es un conjunto de bloques consecutivos y un *segmento* es un conjunto de extensiones consecutivos. Mediante estas dos estructuras, el SGBDR Oracle® garantiza un espacio consecutivo para el almacenamiento de bloques para el almacenamiento de los datos de la misma estructura relacional (tabla, índice o *cluster*).

Pregunta 3

Enunciado

Sean las relaciones S y T con los siguientes parámetros:

S(<u>a</u>,b,c)	T(<u>a</u>,d,e)	U(<u>a</u>,f,g)
N(S) = 1000	N(T)=500	N(U)=300
Size(a)=20	Size(a)=20	Size(a)=20
Size(b)=50		
Size(c)=15		
	Size(d)=4	
	Size(e)=20	
		Size(f)=14
		Size(g)=10
V(S,a) = ?	V(T,a) = ?	V(U,a) = ?
V(S,c)=20		

donde $S.a$, $T.a$ y $U.a$ son claves, y donde el atributo $T.a$ es llave externa a $S.a$ y $U.a$ es llave externa a $T.a$

- Determina las variabilidades del campo a , es decir, $V(S,a)$, $V(T,a)$ y $V(U,a)$.
- Teniendo en cuenta que, 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 de cada relación o resultado de operación intermedia, 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:

$$\Pi_{a,e}(\sigma_{(d=d_3)}(S \text{ JOIN } T))$$

Resolución apartado a

Cuando un atributo es clave, su valor no puede repetirse en dos registros distintos, ni puede estar vacío, por lo que habrá tantos valores distintos para el atributo como registros haya.

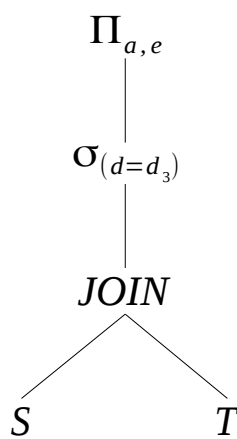
Dado que los tres atributos mencionados son claves, podemos concluir que:

- $V(S,a) = N(S) = 1000$
- $V(T,a) = N(T) = 500$
- $V(U,a) = N(U) = 300$

es decir, que la variabilidad de cada atributo es el número de registros de la correspondiente relación.

Resolución apartado 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 S para poder realizar la operación de reunión natural:

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

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

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

de modo que:

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

$$B(S) = \text{hacia arriba} \left(\frac{1000}{47} \right) = 22$$

Del mismo modo, tenemos que calcular el número de bloques de T para poder realizar la operación de reunión natural:

$$B(T) = \text{hacia arriba} \left(\frac{N(T)}{Bfr(T)} \right) = \text{hacia arriba} \left(\frac{500}{Bfr(T)} \right)$$

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

$$L(T) = 20 + 4 + 20 = 44$$

de modo que:

$$Bfr(T) = \text{parte entera} \left(\frac{4096-20}{44} \right) = 92 \quad y$$

$$B(T) = \text{hacia arriba} \left(\frac{500}{92} \right) = 6$$

A fin de poder realizar la operación de reunión natural de la forma más eficiente posible, ambos ficheros tienen que estar ordenados por el atributo de reunión, es decir, *a*. Pero, dado que *a* es la clave en ambas relaciones, ya existen índices para ambos atributos en ambas relaciones, por lo que sólo tenemos que leer dichos índices para obtener los registros de forma ordenada en ambas relaciones. En resumen: no es necesario ordenar los registros de *T* porque ya están ordenados por el atributo de reunión.

La operación de reunión natural requiere leer y mezclar todos los registros de la relación S con los de la relación T sin leer cada bloque más de una vez, por lo que será necesario:

Lectura: $B(S) + B(T) = 22 + 6 = 28$

y el resultado de la reunión natural necesita un número de bloques suficiente para

almacenar los registros resultantes de esta operación, o sea:

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

$$N(JOIN) = \frac{N(S) \cdot N(T)}{\max\{V(S, a), V(T, a)\}} = \frac{1000 \cdot 500}{1000} = 500$$

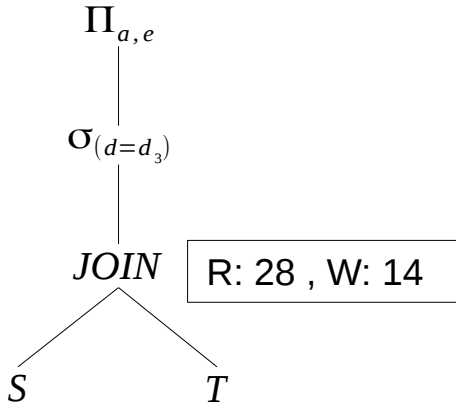
$$Bfr(JOIN) = \text{parte entera} \left(\frac{B-C}{L(JOIN)} \right) = \text{parte entera} \left(\frac{4096-20}{L(JOIN)} \right)$$

$$L(JOIN) = L(S) + L(T) - \text{Size}(a) = 85 + 44 - 20 = 109$$

de modo que:

$$Bfr(JOIN) = \text{parte entera} \left(\frac{4096-20}{L(JOIN)} \right) = \text{parte entera} \left(\frac{4096-20}{109} \right) = 37 \text{ y}$$

$$\text{Escritura: } B(JOIN) = \text{hacia arriba} \left(\frac{N(JOIN)}{Bfr(JOIN)} \right) = \text{hacia arriba} \left(\frac{500}{37} \right) = 14$$



La 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:

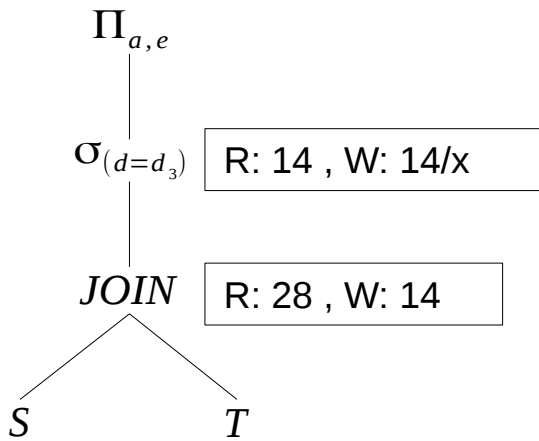
$$\begin{aligned} B(\sigma_{d=d_3}) &= \text{hacia arriba} \left(\frac{N(\sigma_{d=d_3})}{Bfr(\sigma_{d=d_3})} \right) = \\ &= \text{hacia arriba} \left(\frac{N(\sigma_{d=d_3})}{Bfr(JOIN)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{d=d_3})}{37} \right) \end{aligned}$$

Asumiendo que, al realizar la reunión natural, no afectamos a la variabilidad del atributo d procedente de la relación S , el número de tuplas resultantes de la selección sería:

$$N(\sigma_{d=d_3}) = \frac{N(JOIN)}{V(JOIN, d)} = \frac{500}{x}$$

de modo que:

$$\text{Escritura: } B(\sigma_{t=t_3}) = \text{hacia arriba} \left(\frac{N(\sigma_{d=d_3})}{37} \right) = \text{hacia arriba} \left(\frac{500}{37x} \right) \approx \frac{14}{x}$$



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:

$$\begin{aligned}
 B(\Pi_{a,e}) &= \text{hacia arriba} \left(\frac{N(\Pi_{a,e})}{Bfr(\Pi_{a,e})} \right) = \\
 &= \text{hacia arriba} \left(\frac{N(\sigma_{d=d_3})}{Bfr(\Pi_{a,e})} \right) = \\
 &= \text{hacia arriba} \left(\frac{500}{x \cdot Bfr(\Pi_{a,e})} \right)
 \end{aligned}$$

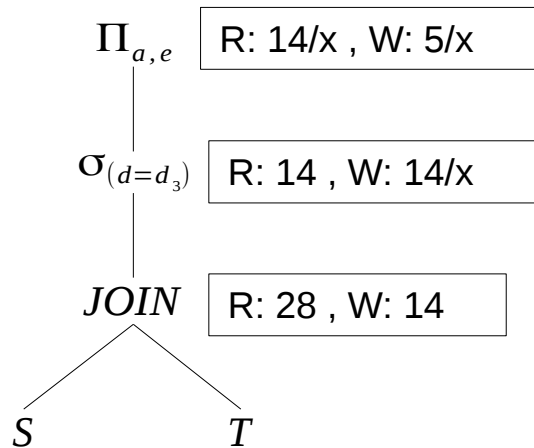
$$Bfr(\Pi_{a,e}) = \text{parte entera} \left(\frac{B-C}{L(\Pi_{a,e})} \right) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{a,e})} \right)$$

$$L(\Pi_{a,e}) = 20 + 20 = 40$$

de modo que:

$$Bfr(\Pi_{a,e}) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{a,e})} \right) = \text{parte entera} \left(\frac{4096-20}{40} \right) = 101 \quad y$$

Escritura: $B(\Pi_{a,e}) = \text{hacia arriba} \left(\frac{500}{x \cdot Bfr(\Pi_{a,e})} \right) = \text{hacia arriba} \left(\frac{500}{101x} \right) \approx \frac{5}{x}$



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:

$$(28+14) + \left(14 + \frac{14}{x}\right) + \left(\frac{14}{x} + \frac{5}{x}\right) = 62 + \frac{33}{x} \text{ bloques}$$

Pregunta 4

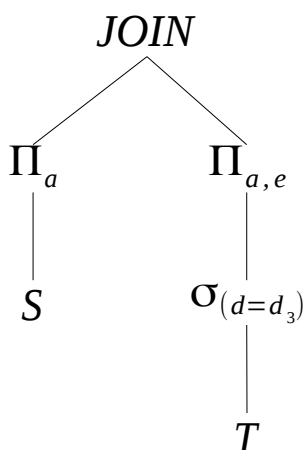
Enunciado

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 cuánto se acerque tu propuesta al óptimo.

Resolución

Un plan posible consistiría en realizar las proyecciones lo antes posible para eliminar atributos que no son necesarios en la consulta. Esta mejora es útil únicamente sobre la relación S dado que sirve de nada aplicarlo sobre la relación T (puesto que sería necesario mantener los atributos a -para la reunión natural y el resultado-, d -para la selección posterior- y e -para el resultado-).

El plan propuesto sería el que figura a continuación:



Es necesario notar que, de realizar las proyecciones en la relación S (para eliminar atributos no útiles para el resultado ni las operaciones posteriores) y T (para eliminar atributos no útiles para la reunión), no es necesario realizar una proyección final para obtener los atributos del resultado puesto que la operación de reunión ya devuelve exactamente los atributos necesarios para el resultado.

No será necesario realizar los cálculos referentes a las relaciones S y T , puesto que ya han sido realizados en el apartado b del ejercicio 3.

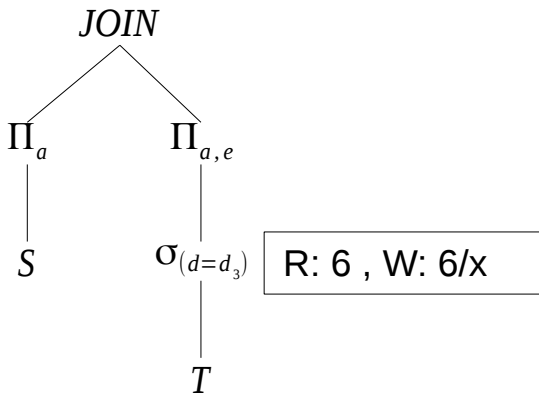
A fin de aplicar la selección sobre los registros de la relación T , es necesario leer todos los registros, almacenados en 6 bloques, y escribir en los bloques del resultado aquellos que cumplan la condición indicada. El número de bloques necesarios para escribir el resultado se calcula a continuación.

$$B(\sigma_{d=d_3}) = \text{hacia arriba} \left(\frac{N(\sigma_{d=d_3})}{Bfr(\sigma_{d=d_3})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{d=d_3})}{Bfr(T)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{d=d_3})}{92} \right)$$

$$N(\sigma_{d=d_3}) = \frac{N(T)}{V(T, d)} = \frac{500}{x}$$

de modo que:

$$B(\sigma_{d=d_3}) = \text{hacia arriba} \left(\frac{N(\sigma_{d=d_3})}{92 \cdot x} \right) = \text{hacia arriba} \left(\frac{500}{92 \cdot x} \right) \approx \frac{6}{x}$$



La operación de proyección sobre el resultado de la selección deberá leer todos los bloques del resultado de dicha selección, y escribir el resultado de la misma (atributos seleccionados) en los bloques de resultado de la proyección.

El número de bloques resultante se calcula a continuación.

$$B(\Pi_{a,e}) = \text{hacia arriba} \left(\frac{N(\Pi_{a,e})}{Bfr(\Pi_{a,e})} \right) =$$

$$= \text{hacia arriba} \left(\frac{N(\sigma_{d=d_3})}{Bfr(\Pi_{a,e})} \right) =$$

$$= \text{hacia arriba} \left(\frac{500}{x \cdot Bfr(\Pi_{a,e})} \right)$$

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

$$L(\Pi_{a,e}) = 20 + 20 = 40$$

por lo que:

$$Bfr(\Pi_{a,e}) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{a,e})} \right) = \text{parte entera} \left(\frac{4096-20}{40} \right) = 101$$

$$\text{Escritura: } B(\Pi_{a,e}) = \text{hacia arriba} \left(\frac{500}{x \cdot Bfr(\Pi_{a,e})} \right) = \text{hacia arriba} \left(\frac{500}{101 \cdot x} \right) \approx \frac{5}{x}$$

Para realizar la operación de proyección correspondiente sobre la relación S es necesario leer los 22 bloques de la relación e ir escribiendo, únicamente, el atributo a en el bloque de resultado de la operación intermedia. El número de bloques del resultado de dicha operación se calcula a continuación.

$$B(\Pi_a) = \text{hacia arriba} \left(\frac{N(\Pi_a)}{Bfr(\Pi_a)} \right) = \text{hacia arriba} \left(\frac{N(S)}{Bfr(\Pi_a)} \right) = \text{hacia arriba} \left(\frac{1000}{Bfr(\Pi_a)} \right)$$

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

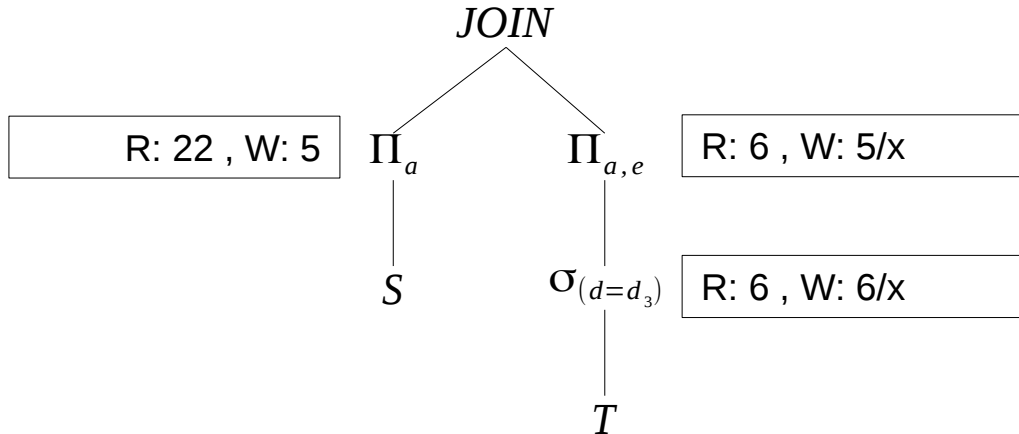
$$L(\Pi_a)=20$$

por lo que:

$$Bfr(\Pi_a)=\text{parte entera}\left(\frac{4096-20}{L(\Pi_a)}\right)=\text{parte entera}\left(\frac{4096-20}{20}\right)=203$$

Escritura: $B(\Pi_a)=\text{hacia arriba}\left(\frac{1000}{Bfr(\Pi_a)}\right)=\text{hacia arriba}\left(\frac{1000}{203}\right)=5$
--

En este punto, ya estamos preparados para realizar la última operación, y hemos realizado el número de operaciones de lectura y escritura que se puede ver a continuación:



Llegados a este punto, la operación de reunión natural requiere que las relaciones estén ordenadas por el atributo de reunión. Sin embargo, siendo a claves en las relaciones originales S y T , al haber realizado operaciones sobre ellas no se puede garantizar que sigan ordenadas, por lo que será necesario ordenarlas.

Para ordenar la relación de la izquierda, será necesario realizar un número de lecturas y escrituras de bloques para ordenarlos que se puede aproximar por:

$$B(\Pi_a) \cdot \log_2(B(\Pi_a)) = 5 \cdot \log_2(5) \approx 12$$

Para ordenar la relación de la derecha, será necesario realizar un número de lecturas y escrituras de bloques para ordenarlos que se puede aproximar por:

$$B(\Pi_{a,e}) \cdot \log_2(B(\Pi_{a,e})) = \frac{5}{x} \cdot \log_2\left(\frac{5}{x}\right) = \frac{5 \cdot \log_2(5) - 5 \cdot \log_2(x)}{x} = \frac{12 - 5 \cdot \log_2(x)}{x}$$

Posteriormente, será necesario leer los bloques ordenados de cada una de las dos relaciones (izquierda y derecha), para realizar la mezcla de ambos. El resultado se escribirá en un número de bloques que se calcula a continuación.

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

$$\begin{aligned} N(JOIN) &= \frac{N(\Pi_a) \cdot N(\Pi_{a,e})}{\max\{V(\Pi_a, a), V(\Pi_{a,e}, a)\}} = \frac{N(S) \cdot N(\sigma_{d=d_3})}{\max\{V(S, a), V(\sigma_{d=d_3}, a)\}} = \frac{1000 \cdot 500}{x \cdot \max\{1000, V(T, a)\}} = \\ &= \frac{1000 \cdot 500}{x \cdot \max\{1000, V(T, a)\}} = \frac{1000 \cdot 500}{x \cdot \max\{1000, 500\}} = \frac{500}{x} \end{aligned}$$

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

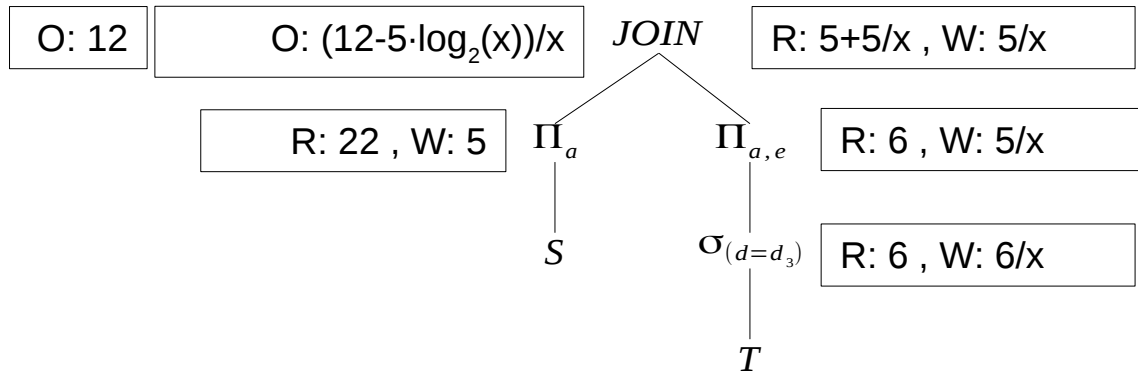
$$L(JOIN) = L(\Pi_a) + L(\Pi_{a,e}) - \text{Size}(a) = 20 + 40 - 20 = 40$$

por lo que:

$$Bfr(JOIN) = \text{parte entera}\left(\frac{B-C}{L(JOIN)}\right) = \text{parte entera}\left(\frac{4096-20}{40}\right) = 101 \text{ y}$$

$$B(JOIN) = \text{hacia arriba}\left(\frac{N(JOIN)}{Bfr(JOIN)}\right) = \text{hacia arriba}\left(\frac{500}{101x}\right) \approx \frac{5}{x}$$

Por tanto, el número total de lecturas y escrituras de bloques para este plan físico sería:



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:

$$\left(6 + \frac{6}{x}\right) + \left(6 + \frac{5}{x}\right) + (22+5) + 12 + \frac{12-5\log_2(x)}{x} + \left(5 + \frac{5}{x} + \frac{5}{x}\right) = 55 + \frac{33-5\log_2(x)}{x} \text{ bloques}$$

El número de bloques leídos y escritos en este plan físico es el mismo que en el original por lo que no podemos decir que este plan consiga una optimización evidente, siendo en su contenido mucho más eficiente.

Este plan es más eficiente que el propuesto en el ejercicio 3, siempre que se cumpla:

$$55 + \frac{33-5\log_2(x)}{x} < 62 + \frac{33}{x} \Rightarrow \frac{33}{x} - \frac{5\log_2(x)}{x} < 7 + \frac{33}{x} \Rightarrow -\frac{5\log_2(x)}{x} < 7 \Rightarrow -\frac{\log_2(x)}{x} < 1.4$$

Dado que x es un número entero igual o mayor que 1, que a partir de $x=1$ el logaritmo es positivo, que ambas funciones (logaritmo e identidad) son positivas para el dominio de x , podemos decir que esa desigualdad se cumple siempre.

Pregunta 5

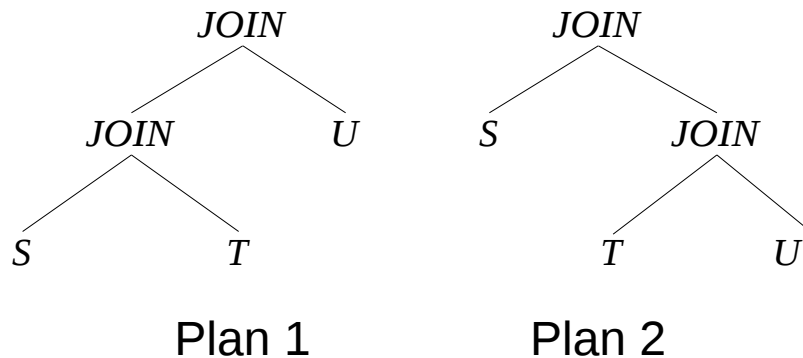
Enunciado

Considerando la propiedad asociativa de la reunión natural (*JOIN*), indica de forma numérica y razonada en qué orden se realizarán las operaciones de esta consulta:

$$S \text{ JOIN } T \text{ JOIN } U$$

Resolución

Las dos posibilidades para resolver esta consulta son:



En el caso del plan 1, sería necesario ordenar el resultado del JOIN de la izquierda, y en el caso del plan 2, sería necesario ordenar el resultado del JOIN de la derecha (ya que los resultados de ambas operaciones no tienen por qué estar ordenados).

El número de bloques del resultado del JOIN de la izquierda será de 14 bloques (resultado del ejercicio 3).

El número de bloques del JOIN de la derecha en el plan 2 se calcula a continuación:

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

$$N(\text{JOIN}) = \frac{N(T) \cdot N(U)}{\max\{V(T, a), V(U, a)\}} = \frac{500 \cdot 300}{\max\{500, 300\}} = 300$$

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

$$L(\text{JOIN}) = L(T) + L(U) - \text{Size}(a) = 44 + 44 - 20 = 68$$

por lo que:

$$Bfr(\text{JOIN}) = \text{parte entera} \left(\frac{4096 - 20}{68} \right) = 59 \quad \text{y}$$

$$B(\text{JOIN}) = \text{hacia arriba} \left(\frac{300}{59} \right) = 6$$

Por tanto, la operación de ordenación asociada al plan 2 requerirá un menor número de bloques para transferir que la asociada al plan 1.