

Corrección Examen de contenidos teóricos (Junio de 2014)

Primera parte

1. (3 puntos) Sean las relaciones R y S con los siguientes parámetros:

R(a,b,c)	S(c,d)
N(R) = 5000	N(S)=200
V(R,a)=5000	
V(R,b)=300	
V(R,c)=5	V(S,c)=5
	V(S,d)=40
Size(a)=20	
Size(b)=60	
Size(c)=20	Size(c)=20
	Size(d)=40

Propón el plan lógico más eficiente para la siguiente consulta y justifica numéricamente por qué es más eficiente en base al número de bloques que emplea cada uno y teniendo en cuenta que el tamaño de bloque es de 2KB, que la cabecera es de 20B y que en memoria sólo cabe un bloque:

$$\Pi_{a,d}(\sigma_{(b < b_1)}(R JOIN S))$$

El plan más eficiente para la consulta planteada sería el generado por el árbol de expresión correspondiente a la consulta:

$$\Pi_{a,d}(\Pi_{a,c}(\sigma_{(b < b_1)}(R)) JOIN S)$$

Teniendo en cuenta que cada operación de E/S representa la transferencia de un bloque de datos entre el disco y la memoria, o viceversa, tendremos que calcular el número de bloques transferidos en total.

A fin de justificar que el plan que propongo es más eficiente que el planteado en el enunciado, tendré que calcular el número de bloques de E/S necesarios para uno y otro plan.

Comenzando por el planteado en el enunciado:

El número de bloques que almacenan la relación R es de:

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

Para ello, hay que calcular el factor de bloqueo de R como:

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

que dependerá de la longitud del registro de R:

$$L(R) = 20B + 60B + 20B = 100B$$

De modo que:

$$Bfr(R) = \text{parte entera} \left(\frac{2048-20}{100} \right) = 20$$

$$B(R) = \text{redondeo hacia arriba} \left(\frac{5000}{20} \right) = 250 \text{ bloques}$$

Si aplicamos los mismos cálculos para el número de bloques de la relación S:

$$L(S) = 20B + 40B = 60B$$

$$Bfr(S) = \text{parte entera} \left(\frac{2048 - 20}{60} \right) = 33$$

$$B(S) = \text{redondeo hacia arriba} \left(\frac{200}{33} \right) = 7 \text{ bloques}$$

Como ya se sabe, hay varios algoritmos para resolver la operación de reunión natural, pero la más eficiente, consiste en ordenar por el atributo de reunión aquella relación que no lo está (sólo una de las dos). Entendemos que R no lo está por lo que tenemos que aplicar un algoritmo de ordenación que requiere realizar un número de lecturas de:

$$B(R) \cdot \log_2(B(R)) = 250 \cdot \log_2(250) = 1991,45 \approx 1992 \text{ bloques}$$

Ahora que las dos relaciones están ordenadas por el atributo c, podemos realizar la operación de reunión natural entre las dos. Habrá que leer:

$$250 + 7 = 257 \text{ bloques}$$

y escribir los bloques resultantes de la operación de reunión:

$$L(\text{JOIN}) = 20B + 60B + 20B + 40B = 140B$$

$$Bfr(\text{JOIN}) = \text{parte entera} \left(\frac{B - C}{L(\text{JOIN})} \right) = \text{parte entera} \left(\frac{2048 - 20}{140} \right) = 14$$

$$B(\text{JOIN}) = \text{redondeo hacia arriba} \left(\frac{N(\text{JOIN})}{Bfr(\text{JOIN})} \right) = \text{redondeo hacia arriba} \left(\frac{N(\text{JOIN})}{14} \right)$$

El número de registros de la operación de reunión natural viene dado por:

$$N(\text{JOIN}) = \frac{N(R) \cdot N(S)}{\max\{V(R, c), V(S, c)\}} = \frac{5000 \cdot 200}{\max\{5, 5\}} = 200000$$

por lo que el número de bloques resultantes que tendremos que escribir en disco será de:

$$B(\text{JOIN}) = \text{redondeo hacia arriba} \left(\frac{200000}{14} \right) = 14286 \text{ bloques}$$

A fin de aplicar la operación de selección sobre el atributo d, tendremos que leer de nuevo:

$$14286 \text{ bloques}$$

y el número de bloques que se transferirán al disco como resultado de la selección será de:

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

La estimación del número de tuplas que cumplen una condición de desigualdad de mayor o menor es de un tercio del total de las tuplas de la relación sobre la que se aplica la operación de selección.

$$N(\text{SELEC}) = \frac{N(\text{JOIN})}{3} = \frac{200000}{3} \approx 66667$$

por lo que:

$$B(\text{SELEC}) = \text{redondeo hacia arriba} \left(\frac{66667}{14} \right) = 4762 \text{ bloques}$$

Por último, para realizar la operación de proyección, tendremos que leer:

4762 bloques

y escribir el número de bloques resultantes de dicha operación de proyección:

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

$$Bfr(\text{PROY}) = \text{parte entera} \left(\frac{B-C}{L(\text{PROY})} \right) = \text{parte entera} \left(\frac{2048-20}{L(\text{PROY})} \right)$$

$$L(\text{PROY}) = 20B + 40B = 60B$$

Con lo que:

$$Bfr(\text{PROY}) = \text{parte entera} \left(\frac{2048-20}{L(\text{PROY})} \right) = \text{parte entera} \left(\frac{2048-20}{60} \right) = 33$$

$$B(\text{PROY}) = \text{redondeo hacia arriba} \left(\frac{N(\text{SELEC})}{Bfr(\text{PROY})} \right) = \text{redondeo hacia arriba} \left(\frac{66667}{33} \right) = 2021 \text{ bloques}$$

De modo que el número total de lecturas y escrituras de bloques necesarios para resolver este plan físico es de:

$$1992 + 250 + 7 + 14286 \cdot 2 + 4762 \cdot 2 + 2021 = 42366 \text{ bloques}$$

El número de bloques involucrado en el plan físico de la consulta que propongo:

$$\Pi_{a,d}(\Pi_{a,c}(\sigma_{(b < b_1)}(R)) \text{ JOIN } S)$$

sería el que se muestra a continuación.

Como se ha mostrado en el plan físico de la consulta anterior, el número de bloques de la relación R es de 250 bloques, que habrá que leer para aplicar la operación de selección sobre la misma. El número de bloques resultante de dicha operación viene dado por los siguientes cálculos:

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

$$N(\text{SELEC}) = \frac{N(R)}{3} = \frac{5000}{3} \approx 1667$$

con lo que:

$$B(\text{SELEC}) = \text{redondeo hacia arriba} \left(\frac{N(\text{SELEC})}{Bfr(R)} \right) = \text{redondeo hacia arriba} \left(\frac{1667}{20} \right) = 84 \text{ bloques}$$

Sobre el resultado de la operación de selección, es necesario aplicar la primera operación de proyección sobre los atributos *a* y *c*, que arroja un número de bloques resultantes como sigue:

$$B(\text{PROY1}) = \text{redondeo hacia arriba} \left(\frac{N(\text{PROY1})}{Bfr(\text{PROY1})} \right) = \text{redondeo hacia arriba} \left(\frac{N(\text{SELEC})}{Bfr(\text{PROY1})} \right)$$

$$Bfr(\text{PROY1}) = \text{parte entera} \left(\frac{B-C}{L(\text{PROY1})} \right) = \text{parte entera} \left(\frac{2048-20}{L(\text{PROY1})} \right)$$

$$L(\text{PROY1}) = 20B + 20B = 40B$$

Con lo que:

$$Bfr(\text{PROY1}) = \text{parte entera} \left(\frac{2048-20}{L(\text{PROY1})} \right) = \text{parte entera} \left(\frac{2048-20}{40} \right) = 50$$

$$B(\text{PROY1}) = \text{redondeo hacia arriba} \left(\frac{N(\text{SELEC})}{Bfr(\text{PROY1})} \right) = \text{redondeo hacia arriba} \left(\frac{1667}{50} \right) = 34 \text{ bloques}$$

El primer paso de la reunión natural, consiste en re-ordenar la relación R por el atributo de reunión c , lo cual tiene un coste en bloques de:

$$B(\text{PROY1}) \cdot \log_2(\text{PROY1}) = 34 \cdot \log_2(34) \approx 173 \text{ bloques}$$

Con respecto a la relación S , está almacenada en 7 bloques (como se ha visto en el plan físico de la consulta planteada por el enunciado) que habrá que leer, junto con los 34 bloques de la proyección calculada. La reunión natural arrojará un resultado en bloques de:

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

$$N(\text{JOIN}) = \frac{N(\text{PROY1}) \cdot N(S)}{\max\{V(R, c), V(S, c)\}} = \frac{N(\text{PROY1}) \cdot 200}{\max\{5, 5\}} = \frac{N(\text{SELEC}) \cdot 200}{5} = \frac{1667 \cdot 200}{5} = 66680$$

$$Bfr(\text{JOIN}) = \text{parte entera} \left(\frac{B - C}{L(\text{JOIN})} \right) = \text{parte entera} \left(\frac{2048 - 20}{L(\text{JOIN})} \right)$$

$$L(\text{JOIN}) = 20B + 20B + 40B = 80B$$

con lo que queda:

$$Bfr(\text{JOIN}) = \text{parte entera} \left(\frac{2028}{80} \right) = 25$$

$$B(\text{JOIN}) = \text{redondeo hacia arriba} \left(\frac{66680}{25} \right) = 2668 \text{ bloques}$$

Por último, y sobre el resultado de la reunión natural, será necesario leer los 2668 bloques para aplicar la segunda operación de proyección para eliminar el atributo c , lo que nos deja con un número de bloques como sigue:

$$B(\text{PROY2}) = \text{redondeo hacia arriba} \left(\frac{N(\text{PROY2})}{Bfr(\text{PROY2})} \right) = \text{redondeo hacia arriba} \left(\frac{N(\text{JOIN})}{Bfr(\text{PROY2})} \right)$$

$$Bfr(\text{PROY2}) = \text{parte entera} \left(\frac{2048 - 20}{L(\text{PROY2})} \right)$$

$$L(\text{PROY2}) = 20B + 40B = 60B$$

con lo que queda:

$$Bfr(\text{PROY2}) = \text{parte entera} \left(\frac{2048 - 20}{60} \right) = 33$$

$$B(\text{PROY2}) = \text{redondeo hacia arriba} \left(\frac{66680}{33} \right) = 2021 \text{ bloques}$$

De modo que el número total de lecturas y escrituras de bloques necesarios para resolver este plan físico es de:

$$250 + 84 \cdot 2 + 34 + 173 + 34 + 7 + 2668 \cdot 2 + 2021 = 8023 \text{ bloques}$$

opción que supone menos de una quinta parte de las lecturas y escrituras de bloques implicadas por el plan físico correspondiente a la consulta planteada por el enunciado.

2. Enuncia brevemente usando tus propias palabras (o con la fórmula si la recuerdas y lo prefieres):
- a) **(1 punto)** ¿De qué depende el tiempo de búsqueda de un registro por su clave en un fichero ASI (archivo secuencial indexado) multinivel con zona de desbordamiento, si la búsqueda se realiza por el valor de clave?.
 - b) **(1 punto)** Enumera los elementos de los que se compone el nivel interno de Oracle Database®.
 - c) **(1 punto)** Los índices son útiles para búsquedas por valor de clave pero son costosos de mantener. Enumera las cuatro situaciones en las que es conveniente eliminarlos.
-

Apartado (a)

Depende del tiempo necesario para encontrar la clave en el índice raíz, el tiempo necesario para encontrar la clave en cada uno de los índices intermedios y el tiempo para acceder a un registro del fichero maestro por posición, además del tiempo para encontrar el registro en la cadena de desbordamiento, en caso de que no estuviera presente en la posición indicada del fichero maestro.

$$T_F = T_M + (m - 1) \cdot T_{F_i} + T'_F + T_{F \text{ Cadena}}$$

Apartado (b)

La base de datos se compone de varios *tablespaces*, que se componen, a su vez, de varios segmentos, que se componen de varias extensiones, que se componen de varios bloques.

Apartado (c)

Los índices deben eliminarse cuando concurra una de las siguientes situaciones:

- Ya no sirven
- No mejoran la eficiencia
- Hay que cambiar los campos que se indexan
- Hay que rehacerlo

Corrección de la Prueba Parcial ABD (Voluntaria)

24 de Abril de 2014

1. **(2 puntos)** Se dispone de una relación $R(a, b)$ donde a es la clave de valores únicos por la que se mantiene ordenado el archivo y b es un atributo con valores duplicados, además se tiene $B = 4096B$, $C=10B$, $P= 10B$, $N(R) = 1000$, $V(R, b) = 200$, $size(a) = 10B$, $size(b) = 40B$. Se montan dos índices I_A e I_B , uno por cada atributo. Calcula el tamaño en bloques de cada índice.
-

En primer lugar, calcularemos el tamaño en bloques del índice construido sobre el atributo a .

El tamaño de cada entrada del índice sería:

$$L(I_A) = 10B + 10B = 20B$$

Para ver cuántos registros entran en un bloque del índice, calculamos el **factor de bloqueo**:

$$Bfr(I_A) = \text{parte entera}\left(\frac{B-C}{L(I_A)}\right) = \text{parte entera}\left(\frac{4096B - 10B}{20B}\right) = 204$$

El número de bloques necesario para almacenar las 1000 entradas de la relación en el índice se calcula:

$$B(I_A) = \text{redondeo hacia arriba}\left(\frac{N(R)}{Bfr(I_A)}\right) = \text{redondeo hacia arriba}\left(\frac{1000}{204}\right) = 5$$

El tamaño del índice I_A es de 5 bloques.

Para calcular el tamaño del índice construido sobre el atributo b , necesitamos hacer algunos cálculos adicionales.

Dado que un índice no puede tener valores repetidos, puesto que la búsqueda binaria no funcionaría, hemos de construir otro tipo de índice. En concreto, podríamos usar una estructura de índice multinivel en el que, el primer nivel (raíz) no tuviera valores repetidos, y el segundo nivel tuviera valores repetidos.

Siendo así, cada valor distinto del atributo b sólo podría aparecer una vez en el primer nivel. Para calcular cuántos bloques hacen falta para el primer nivel, tenemos que tener en cuenta el número de entradas:

$$B(I_B^1) = \text{redondeo hacia arriba}\left(\frac{N(I_B)}{Bfr(I_B)}\right) = \text{redondeo hacia arriba}\left(\frac{V(R, b)}{Bfr(I_B)}\right)$$

$$Bfr(I_B) = \text{parte entera}\left(\frac{B-C}{L(I_B)}\right) = \text{parte entera}\left(\frac{4096B - 10B}{L(I_B)}\right)$$

$$L(I_B) = 40B + 10B = 50B$$

Por lo que:

$$Bfr(I_B) = \text{parte entera}\left(\frac{4086B}{50B}\right) = 81$$

$$B(I_B^1) = \text{redondeo hacia arriba} \left(\frac{V(R, b)}{Bfr(I_B)} \right) = \text{redondeo hacia arriba} \left(\frac{200}{81} \right) = 3 \text{ bloques}$$

Véase que el número de entradas del índice de primer nivel sería el número de valores distintos para el atributo b en la relación R, y que la estructura de todos los índices es exactamente igual, sean de un sólo nivel o multinivel, por lo que el cálculo del factor de bloqueo para el índice I_B es igual que el calculado para el índice I_A .

Con respecto al segundo nivel del índice, cada bloque contendrá entradas con el mismo valor de clave que se recorrerán secuencialmente. En ese caso, habría que calcular cuántos bloques se estiman necesarios para contener las 1000 entradas distintas de la tabla agrupadas en tantos grupos como valores distintos haya del atributo b.

El número de bloques estimado para contener los registros con valores iguales de b se calcula:

$$\begin{aligned} \text{redondeo hacia arriba} \left(\frac{N(I_B^2)}{Bfr(I_B)} \right) &= \text{redondeo hacia arriba} \left(\frac{\frac{L(R)}{V(R, b)}}{Bfr(I_B)} \right) = \\ &= \text{redondeo hacia arriba} \left(\frac{\frac{1000}{200}}{81} \right) = \text{redondeo hacia arriba} \left(\frac{1000}{200 \cdot 81} \right) = 1 \end{aligned}$$

De este modo, para cada uno de los 200 valores de b, se estiman 5 registros que caben en un bloque. De ese modo, cada valor distinto tendría 1 bloque en el segundo nivel del índice, lo que hacen:

$$B(I_B^2) = 1 \text{ bloque / valor} \cdot 200 \text{ valores} = 200 \text{ bloques}$$

De este modo, el cálculo total de bloques para el índice I_B es:

$$B(I_B) = B(I_B^1) + B(I_B^2) = 3 + 200 = 203 \text{ bloques}$$

El tamaño del índice I_B es de 203 bloques.

2. **(1 puntos)** ¿De qué depende el tiempo necesario para reorganizar un Archivo de Acceso Directo?

El tiempo necesario para reorganizar un archivo de acceso directo depende del tiempo necesario en recorrerlo, es decir, del tiempo que se tarda en leer las posiciones del fichero más las colisiones, y el tiempo necesario en generar un nuevo archivo de acceso directo, insertando de nuevo cada registro leído en un nuevo archivo de acceso directo (que dependerá del tiempo que se tarda en calcular la posición del registro y el tiempo que se tarda en insertarlo en una posición o en la zona de desbordamiento).

$$T_Y = T_X + T_{Carga}$$

$$T_X = (m + c) \cdot T$$

$$T_{Carga} = \sum_{i=1}^n T_I(i)$$

3. **(2 puntos)** Sean las relaciones R y S con los siguientes parámetros:

R(a,b,c)	S(c,d)
N(R) = 5000	N(S)=200
V(R,a)=5000	
V(R,b)=3000	
V(R,c)=5	V(S,c)=5
	V(S,d)=40
Size(a)=20	
Size(b)=60	
Size(c)=20	Size(c)=20
	Size(d)=40

Teniendo en cuenta que el tamaño de bloque es de 2KB, que la cabecera es de 20B y que en memoria sólo cabe un bloque, determina el número de operaciones de E/S que supondría la ejecución de la consulta:

$$\Pi_{c,d}(\sigma_{d=d_1}(R \text{ JOIN } S))$$

Teniendo en cuenta que cada operación de E/S representa la transferencia de un bloque de datos entre el disco y la memoria, o viceversa, tendremos que calcular el número de bloques transferidos en total.

El número de bloques que almacenan la relación R es de:

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

Para ello, hay que calcular el factor de bloqueo de R como:

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

que dependerá de la longitud del registro de R:

$$L(R) = 20B + 60B + 20B = 100B$$

De modo que:

$$Bfr(R) = \text{parte entera} \left(\frac{2048 - 20}{100} \right) = 20$$

$$B(R) = \text{redondeo hacia arriba} \left(\frac{5000}{20} \right) = 250 \text{ bloques}$$

Si aplicamos los mismos cálculos para el número de bloques de la relación S:

$$L(S) = 20B + 40B = 60B$$

$$Bfr(S) = \text{parte entera} \left(\frac{2048 - 20}{60} \right) = 33$$

$$B(S) = \text{redondeo hacia arriba} \left(\frac{200}{33} \right) = 7 \text{ bloques}$$

Como ya se sabe, hay varios algoritmos para resolver la operación de reunión natural, pero la más eficiente, consiste en ordenar por el atributo de reunión aquella relación que no lo está (sólo una de las dos). Entendemos que R no lo está por lo que tenemos que aplicar un algoritmo de ordenación que requiere realizar un número de lecturas de:

$$B(R) \cdot \log_2(B(R)) = 250 \cdot \log_2(250) = 1991,45 \approx 1992 \text{ bloques}$$

Ahora que las dos relaciones están ordenadas por el atributo c, podemos realizar la operación de reunión natural entre las dos. Habrá que leer:

$$250 + 7 = 257 \text{ bloques}$$

y escribir los bloques resultantes de la operación de reunión:

$$L(\text{JOIN}) = 20B + 60B + 20B + 40B = 140B$$

$$Bfr(\text{JOIN}) = \text{parte entera} \left(\frac{B-C}{L(\text{JOIN})} \right) = \text{parte entera} \left(\frac{2048-20}{140} \right) = 14$$

$$B(\text{JOIN}) = \text{redondeo hacia arriba} \left(\frac{N(\text{JOIN})}{Bfr(\text{JOIN})} \right) = \text{redondeo hacia arriba} \left(\frac{N(\text{JOIN})}{14} \right)$$

El número de registros de la operación de reunión natural viene dado por:

$$N(\text{JOIN}) = \frac{N(R) \cdot N(S)}{\max\{V(R, c), V(S, c)\}} = \frac{5000 \cdot 200}{\max\{5, 5\}} = 200000$$

por lo que el número de bloques resultantes que tendremos que escribir en disco será de:

$$B(\text{JOIN}) = \text{redondeo hacia arriba} \left(\frac{200000}{14} \right) = 14286 \text{ bloques}$$

A fin de aplicar la operación de selección sobre el atributo d, tendremos que leer de nuevo:

$$14286 \text{ bloques}$$

y el número de bloques que se transferirán al disco como resultado de la selección será de:

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

$$N(\text{SELEC}) = \frac{N(\text{JOIN})}{V(S, d)} = \frac{200000}{40} = 5000$$

por lo que:

$$B(\text{SELEC}) = \text{redondeo hacia arriba} \left(\frac{5000}{14} \right) = 358 \text{ bloques}$$

Por último, para realizar la operación de proyección, tendremos que leer:

$$358 \text{ bloques}$$

y escribir el número de bloques resultantes de dicha operación de proyección:

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

$$Bfr(\text{PROY}) = \text{parte entera} \left(\frac{B-C}{L(\text{PROY})} \right) = \text{parte entera} \left(\frac{2048-20}{L(\text{PROY})} \right)$$

$$L(\text{PROY}) = 20B + 40B = 60B$$

Con lo que:

$$Bfr(\text{PROY}) = \text{parte entera} \left(\frac{2048-20}{L(\text{PROY})} \right) = \text{parte entera} \left(\frac{2048-20}{60} \right) = 33$$

$$B(\text{PROY}) = \text{redondeo hacia arriba} \left(\frac{N(\text{SELEC})}{Bfr(\text{PROY})} \right) = \text{redondeo hacia arriba} \left(\frac{5000}{33} \right) = 152 \text{ bloques}$$

De modo que el número total de lecturas y escrituras de bloques necesarios para resolver este plan físico es de:

$$1992 + 250 + 7 + 14286 \cdot 2 + 358 \cdot 2 + 152 = 31689 \text{ bloques}$$

4. **(2 puntos)** Propón un plan lógico para la consulta del ejercicio anterior cuyo plan físico sea más eficiente que el calculado para dicho ejercicio. Justifica numéricamente tu respuesta.
-

En primer lugar, llevaremos la selección lo más cerca posible de las hojas del árbol de consulta:

$$\Pi_{c,d}(R JOIN \sigma_{d=d_1}(S))$$

A continuación, hundiremos la proyección lo más posible hasta las hojas para eliminar atributos que no deben ser considerados para la resolución de la consulta:

$$\Pi_c(R) JOIN \sigma_{d=d_1}(\Pi_{c,d}(S))$$

Sin embargo y teniendo en cuenta que el esquema de la relación S es, precisamente, c y d, la proyección sobre S no tiene sentido, por lo que el plan lógico consulta sería:

$$\Pi_c(R) JOIN \sigma_{d=d_1}(S)$$

Para calcular si el número de bloques considerados para la ejecución de esta consulta (plan físico) es mejor que el calculado para el ejercicio 3, calculamos dichos bloques.

En primer lugar, el número de bloques para la selección sobre S es:

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

$$N(SELEC) = \frac{N(S)}{V(S,d)} = \frac{200}{40} = 5$$

Por lo que:

$$B(SELEC) = \text{redondeo hacia arriba} \left(\frac{N(SELEC)}{Bfr(S)} \right) = \text{redondeo hacia arriba} \left(\frac{5}{33} \right) = 1$$

Para poder realizar la selección es necesario leer los bloques de S y escribir el bloque resultante:

$$7 + 1 = 8 \text{ bloques}$$

Para poder realizar la proyección sobre R es necesario leer los 250 bloques y escribir el número de bloques necesarios para almacenar el resultado:

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

$$Bfr(PROY) = \text{parte entera} \left(\frac{B-C}{L(PROY)} \right) = \text{parte entera} \left(\frac{2048-20}{L(PROY)} \right)$$

$$L(PROY) = 20 B$$

Con lo que:

$$Bfr(PROY) = \text{parte entera} \left(\frac{2048-20}{L(PROY)} \right) = \text{parte entera} \left(\frac{2048-20}{20} \right) = 101$$

$$B(\text{PROY}) = \text{redondeo hacia arriba} \left(\frac{N(R)}{Bfr(\text{PROY})} \right) = \text{redondeo hacia arriba} \left(\frac{5000}{101} \right) = 46 \text{ bloques}$$

Es decir, para calcular la proyección sobre R es necesario leer los bloques R y escribir los bloques del resultado. En total:

$$250 + 46 = 296 \text{ bloques}$$

Para resolver la reunión natural, será necesario ordenar la primera relación (el resultado de la proyección sobre R) por el atributo c para lo cual habrá que leer:

$$46 \cdot \log_2(46) = 254.1 \approx 255 \text{ bloques leídos y escritos para ordenar}$$

y escribir de nuevo los 46 bloques ordenados. En total:

$$255 + 46 = 301 \text{ bloques}$$

Por último, leemos los bloques ordenados de la proyección sobre R junto con los bloques de la selección sobre S, y los mezclamos. El número de bloques del resultado de la reunión natural sería:

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

Hay que ver que el esquema resultante de la reunión natural es exactamente igual que el esquema de la relación S, por lo que nos ahorramos algunos cálculos ya realizados.

$$N(\text{JOIN}) = \frac{N(\text{PROY}) \cdot N(\text{SELEC})}{\max\{V(R, c), V(S, c)\}} = \frac{5000 \cdot 5}{\max\{5, 5\}} = 5000$$

con lo que:

$$B(\text{JOIN}) = \text{redondeo hacia arriba} \left(\frac{N(\text{JOIN})}{Bfr(S)} \right) = \text{redondeo hacia arriba} \left(\frac{5000}{33} \right) = 152 \text{ bloques}$$

En resumen, el número de bloques necesarios para ejecutar este plan físico sería:

$$7 + 1 + 250 + 46 + 255 + 46 + 46 + 1 + 152 = 804 \text{ bloques}$$

Efectivamente, el número de bloques empleados en esta aproximación es mucho menor de los 31939 bloques del plan físico del ejercicio anterior.

5. **(2 puntos)** Explica qué elementos del nivel interno de Oracle se crearían o modificarían y en qué *tablespaces* para obtener un *cluster* completamente funcional (en el que se puede insertar y consultar) que incluye dos tablas R(a,b,c) y S(c,d).
-

Por una parte, para almacenar las estructuras del clúster, se crearía un segmento, en el que se crearía una extensión, en el que crearía un bloque para contener los registros de las dos tablas del clúster. Estas estructuras se crearían en el tablespace del usuario (*USERS*, en su caso).

Además, se modificaría la tabla o tablas del tablespace *SYSTEM* que contienen la parte del catálogo referente a los clústers, tablas, restricciones sobre columnas y tablas, e índices, consultables por las vistas de catálogo:

- *DBA_OBJECTS*
- *DBA_CLUSTERS*
- *DBA_TABLES*
- *DBA_TAB_COLUMNS*
- *DBA_CONSTRAINTS*
- *DBA_INDEXES*
- ...

6. **(1 puntos)** Propón un tipo de bloque y una estructura de bloque similar a la provista por Oracle para almacenar los registros del clúster del ejercicio anterior en el *tablespace* de datos.
-

El bloque necesario para almacenar esta información sería un bloque con bloqueo fijo y estructura heterogénea y contendría, al igual que un bloque de Oracle:

- cabecera
- directorio de registros
- espacio libre
- grupos de registros formados por:
 - un registro que contienen valores de los atributos comunes del cluster
 - registros de la primera tabla con dichos valores en los atributos comunes
 - registros de la segunda tabla con dichos valores en los atributos comunes

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.

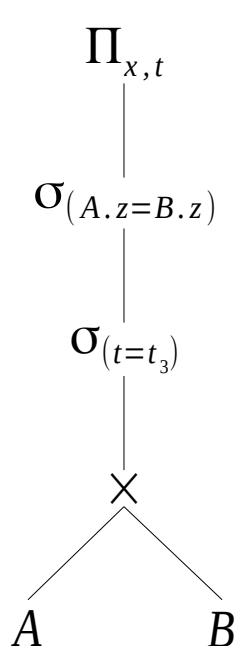
2. (2 puntos) Sean las relaciones A y B con los siguientes parámetros:

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:

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

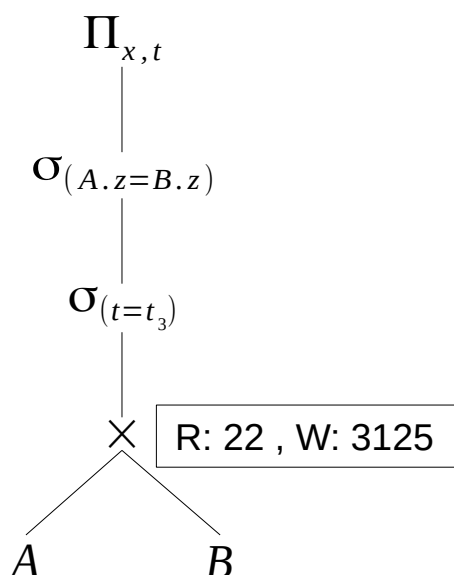
$$Bfr(\times) = \text{parte entera} \left(\frac{B-C}{L(\times)} \right) = \text{parte entera} \left(\frac{4096-20}{L(\times)} \right)$$

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

de modo que:

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

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



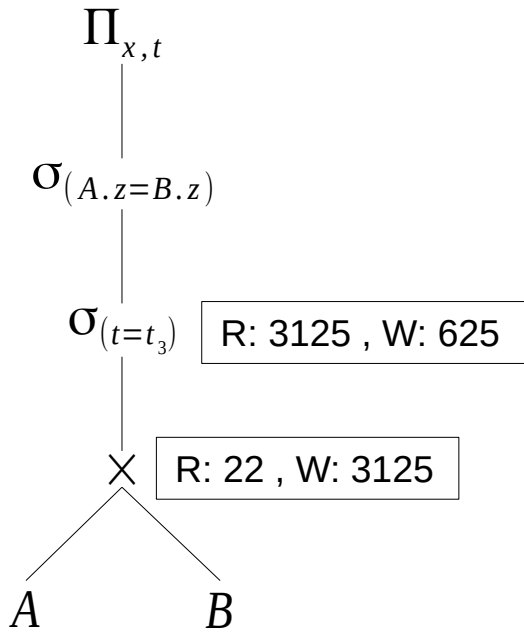
La primera operación de selección debe leer los bloques resultantes de la operación anterior y escribir tantos bloques como sean necesarios para almacenar los registros que resultan de seleccionar aquellos que cumplan la condición, es decir:

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

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

de modo que:

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



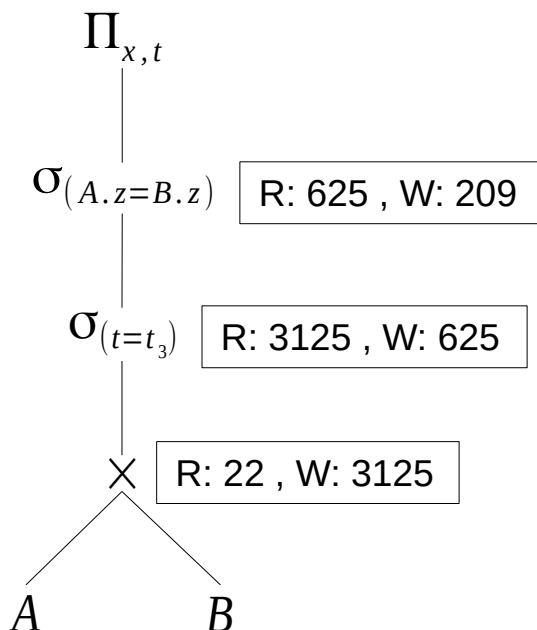
La segunda operación de selección debe leer los bloques resultantes de la operación anterior y escribir tantos bloques como sean necesarios para almacenar los registros que resultan de seleccionar aquellos que cumplan la condición, es decir:

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

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:

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



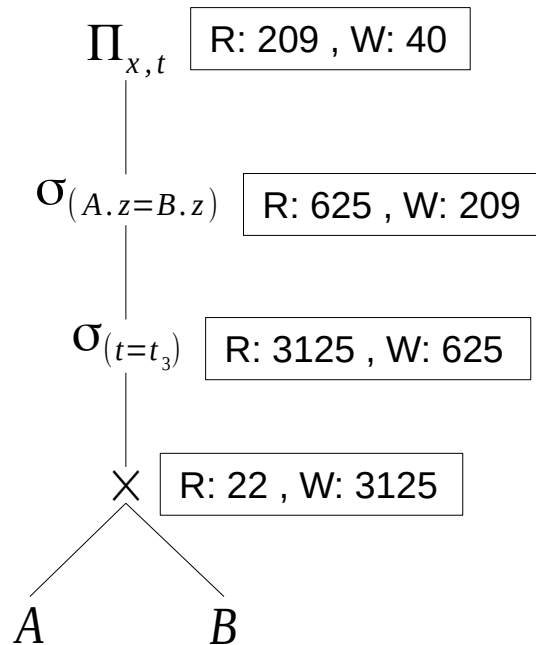
Por último, la operación de proyección debe leer los bloques resultantes de la operación anterior y escribir tantos bloques como sean necesarios para contener los registros resultantes de la operación de proyección:

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

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:

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

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

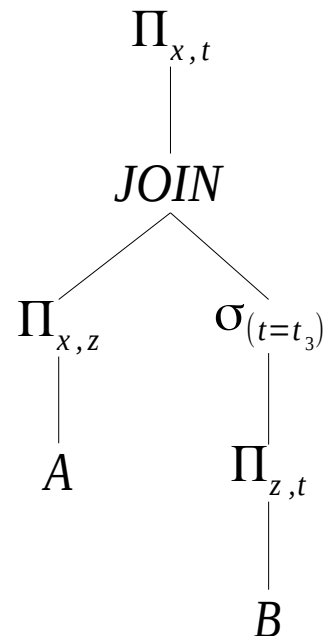
$$L(\Pi_{x,z}) = 20 + 15 = 35$$

de modo que:

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

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

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:

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

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

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:

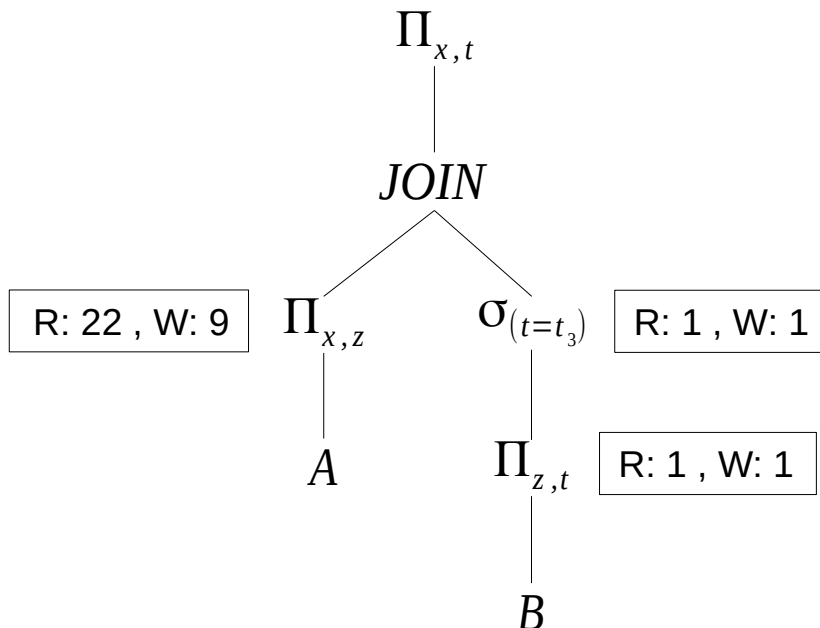
$$B(\sigma_{t=t_3}) = \text{hacia arriba} \left(\frac{N(\sigma_{t=t_3})}{Bfr(\sigma_{t=t_3})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{t=t_3})}{Bfr(\Pi_{z,t})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{t=t_3})}{214} \right)$$

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

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:

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

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

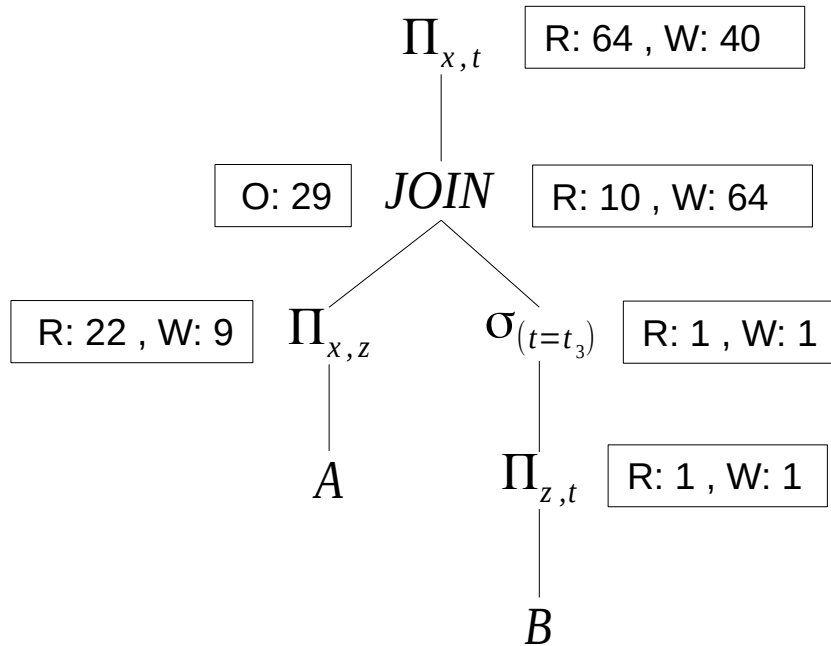
$$L(\Pi_{x,t}) = 20 + 4 = 24$$

de modo que:

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

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)

Grado en Ingeniería Informática
Administración de Bases de Datos
Resolución del Examen de contenidos teóricos – Convocatoria de Septiembre de 2015

1. (3 puntos) Sean las relaciones R y S con los siguientes parámetros:

R(a,b,c)	S(c,d)
N(R) = 5000	N(S)=200
V(R,a)=5000	
V(R,b)=300	
V(R,c)=5	V(S,c)=5
	V(S,d)=40
Size(a)=20	
Size(b)=60	
Size(c)=20	Size(c)=20
	Size(d)=40

Propón el plan lógico más eficiente para la siguiente consulta y justifica numéricamente por qué es más eficiente en base al número de bloques que emplea cada uno y teniendo en cuenta que el tamaño de bloque es de 2KB, que la cabecera es de 20B y que en memoria sólo cabe un bloque:

$$\Pi_{b,d}(\sigma_{(c=c_3 \wedge d \neq d_{20})}(R \times S))$$

Estimación del número de bloques para el plan de la consulta del ejercicio

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.

```

graph BT
    R --> X
    S --> X
    X --> S1["σc=c3"]
    S1 --> S2["σd≠d20"]
    S2 --> P["Πb,d"]
        
```

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

$$B(R) = \text{hacia arriba} \left(\frac{N(R)}{Bfr(R)} \right) = \text{hacia arriba} \left(\frac{5000}{Bfr(R)} \right)$$

$$Bfr(R) = \text{parte entera} \left(\frac{B-C}{L(R)} \right) = \text{parte entera} \left(\frac{2048-20}{L(R)} \right)$$

$$L(R) = 20 + 60 + 20 = 100$$

de modo que:

$$Bfr(R) = \text{parte entera} \left(\frac{2048-20}{100} \right) = 20 \quad \text{y}$$

$$B(R) = \text{hacia arriba} \left(\frac{5000}{20} \right) = 250$$

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

$$B(S) = \text{hacia arriba} \left(\frac{N(S)}{Bfr(S)} \right) = \text{hacia arriba} \left(\frac{200}{Bfr(S)} \right)$$

$$Bfr(S) = \text{parte entera} \left(\frac{B-C}{L(S)} \right) = \text{parte entera} \left(\frac{2048-20}{L(S)} \right)$$

$$L(S) = 20 + 40 = 60$$

de modo que:

$$Bfr(S) = \text{parte entera} \left(\frac{2048-20}{60} \right) = 33 \quad \text{y}$$

$$B(S) = \text{hacia arriba} \left(\frac{200}{33} \right) = 7$$

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

Lectura: $B(R) \cdot B(S) = 250 \cdot 7 = 1750$

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

$$B(\times) = \text{hacia arriba} \left(\frac{N(\times)}{Bfr(\times)} \right) = \text{hacia arriba} \left(\frac{5000 \cdot 200}{Bfr(\times)} \right)$$

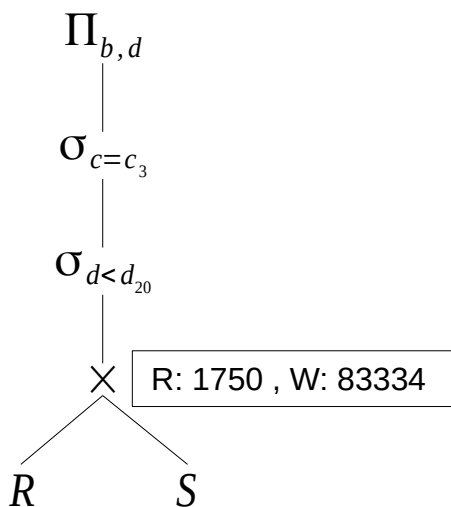
$$Bfr(\times) = \text{parte entera} \left(\frac{B-C}{L(\times)} \right) = \text{parte entera} \left(\frac{2048-20}{L(\times)} \right)$$

$$L(\times) = L(R) + L(S) = 100 + 60 = 160$$

de modo que:

$$Bfr(\times) = \text{parte entera} \left(\frac{2048-20}{L(\times)} \right) = \text{parte entera} \left(\frac{2048-20}{160} \right) = 12 \quad \text{y}$$

Escritura: $B(\times) = \text{hacia arriba} \left(\frac{5000 \cdot 200}{Bfr(\times)} \right) = \text{hacia arriba} \left(\frac{5000 \cdot 200}{12} \right) = 83334$



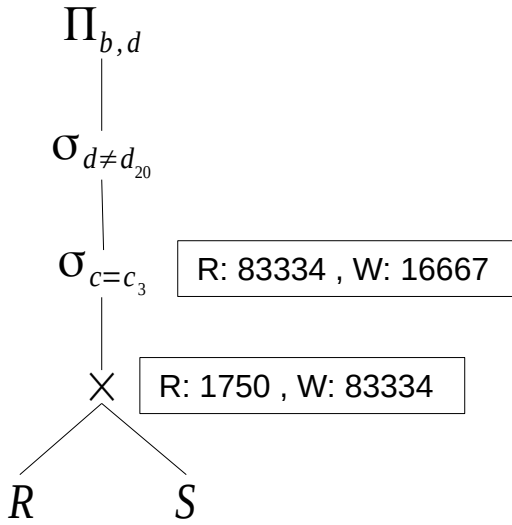
La primera operación de selección debe leer los bloques resultantes de la operación anterior y escribir tantos bloques como sean necesarios para almacenar los registros que resultan de seleccionar aquellos que cumplan la condición, es decir:

$$B(\sigma_{c=c_3}) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{Bfr(\sigma_{c=c_3})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{Bfr(\times)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{12} \right)$$

$$N(\sigma_{c=c_3}) = \frac{N(\times)}{V(R, c)} = \frac{5000 \cdot 200}{5} = 200000$$

de modo que:

Escritura: $B(\sigma_{c=c_3}) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{12} \right) = \text{hacia arriba} \left(\frac{200000}{12} \right) = 16667$



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:

$$B(\sigma_{d \neq d_{20}}) = \text{hacia arriba} \left(\frac{N(\sigma_{d \neq d_{20}})}{Bfr(\sigma_{d \neq d_{20}})} \right) =$$

$$= \text{hacia arriba} \left(\frac{N(\sigma_{d \neq d_{20}})}{Bfr(\sigma_{c=c_3})} \right) =$$

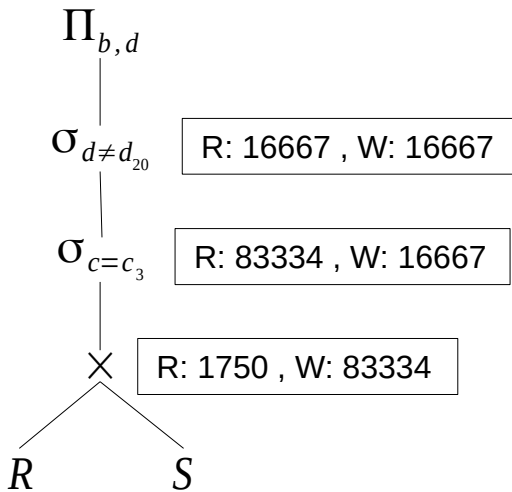
$$= \text{hacia arriba} \left(\frac{N(\sigma_{d \neq d_{20}})}{Bfr(\times)} \right) =$$

$$= \text{hacia arriba} \left(\frac{N(\sigma_{d \neq d_{20}})}{12} \right)$$

$$N(\sigma_{c=c_3}) = N(\sigma_{d \neq d_{20}}) \cdot 1 = 200000 \cdot 1 = 200000$$

De este modo, se escriben:

Escritura: $B(\sigma_{d \neq d_{20}}) = \text{hacia arriba} \left(\frac{N(\sigma_{d \neq d_{20}})}{12} \right) = \text{hacia arriba} \left(\frac{200000}{12} \right) = 16667$



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:

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

$$= \text{hacia arriba} \left(\frac{N(\sigma_{d \neq d_{20}})}{Bfr(\Pi_{b,d})} \right) = \text{hacia arriba} \left(\frac{200000}{Bfr(\Pi_{b,d})} \right)$$

$$Bfr(\Pi_{b,d}) = \text{parte entera} \left(\frac{B-C}{L(\Pi_{b,d})} \right) =$$

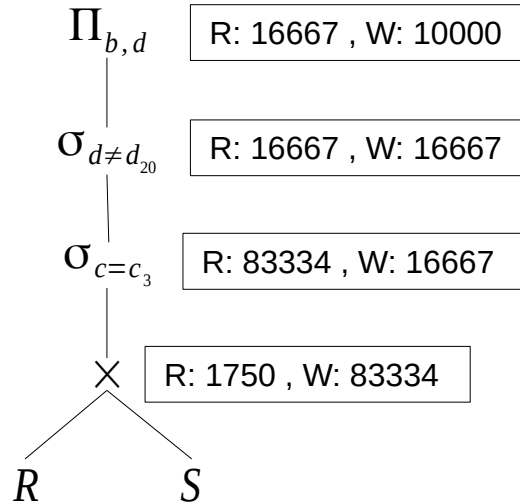
$$= \text{parte entera} \left(\frac{2048-20}{L(\Pi_{b,d})} \right)$$

$$L(\Pi_{b,d}) = 60 + 40 = 100$$

de modo que:

$$Bfr(\Pi_{b,d}) = \text{parte entera} \left(\frac{2048-20}{L(\Pi_{b,d})} \right) = \text{parte entera} \left(\frac{2048-20}{100} \right) = 20 \text{ y}$$

$$\text{Escritura: } B(\Pi_{b,d}) = \text{hacia arriba} \left(\frac{200000}{Bfr(\Pi_{b,d})} \right) = \text{hacia arriba} \left(\frac{200000}{20} \right) = 10000$$

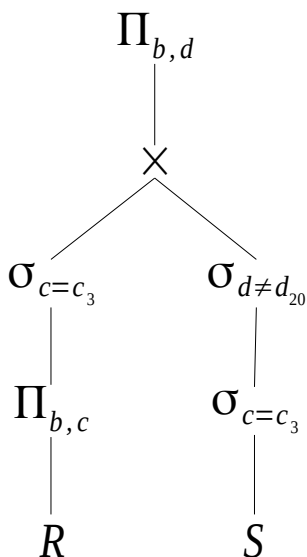


Una vez estimados el número de bloques que cada operación lee y escribe, podemos estimar que el número total de bloques leídos y escritos para la resolución del plan físico correspondiente es de:

$$(1750+83334) + (83334+16667) + (16667+16667) + (16667+10000) = 245086 \text{ bloques}$$

Estimación del número de bloques para el plan optimizado para la consulta del ejercicio

El plan lógico propuesto para la optimización del plan lógico planteado por la consulta sería el que se ve a continuación:



Para calcular el número de bloques estimados para resolver el plan físico, reutilizaremos los cálculos realizados en la parte anterior cuando sea posible.

$$B(\Pi_{b,c}) = \text{hacia arriba} \left(\frac{N(R)}{Bfr(\Pi_{b,c})} \right) = \text{hacia arriba} \left(\frac{5000}{Bfr(\Pi_{b,c})} \right)$$

$$Bfr(\Pi_{b,c}) = \text{parte entera} \left(\frac{B-C}{L(\Pi_{b,c})} \right) = \text{parte entera} \left(\frac{2048-20}{L(\Pi_{b,c})} \right)$$

$$L(\Pi_{b,c}) = 60 + 20 = 80$$

de donde podemos calcular que:

$$Bfr(\Pi_{b,c}) = \text{parte entera} \left(\frac{2048-20}{80} \right) = 25$$

$$B(\Pi_{b,c}) = \text{hacia arriba} \left(\frac{5000}{25} \right) = 200$$

Después, procedemos para calcular el número de bloques necesarios para escribir la selección de la rama izquierda:

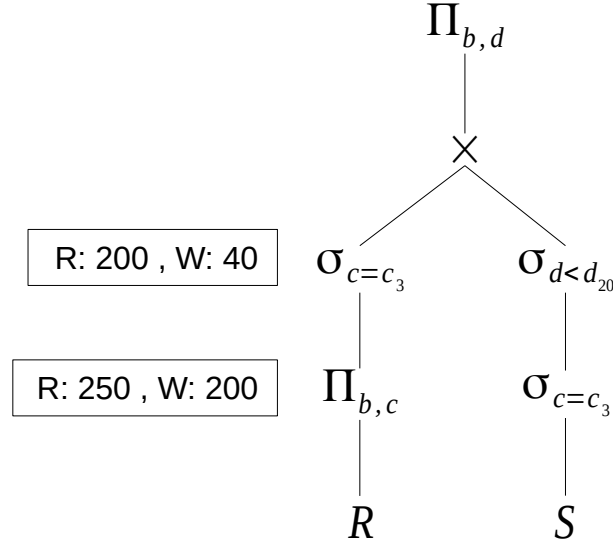
$$B(\sigma_{c=c_3}) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{Bfr(\sigma_{c=c_3})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{Bfr(\Pi_{b,c})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{25} \right)$$

$$N(\sigma_{c=c_3}) = \frac{N(\Pi_{b,c})}{V(R,c)} = \frac{N(R)}{V(R,c)} = \frac{5000}{5} = 1000$$

de donde podemos calcular que:

$$B(\sigma_{c=c_3}) = \text{hacia arriba} \left(\frac{1000}{25} \right) = 40$$

Con esto, podemos concluir el número de bloques necesario para calcular la relación izquierda del producto cartesiano:



Procedemos a realizar los cálculos para la estimación de la relación derecha del producto cartesiano:

$$B(\sigma_{c=c_3}^D) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}^D)}{Bfr(\sigma_{c=c_3}^D)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}^D)}{Bfr(S)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}^D)}{33} \right)$$

$$N(\sigma_{c=c_3}^D) = \frac{N(S)}{V(S, c)} = \frac{200}{5} = 40$$

de donde podemos calcular que:

$$B(\sigma_{c=c_3}^D) = \text{hacia arriba} \left(\frac{40}{33} \right) = 2$$

A continuación, podemos aplicar la segunda selección y estimar los bloques necesarios:

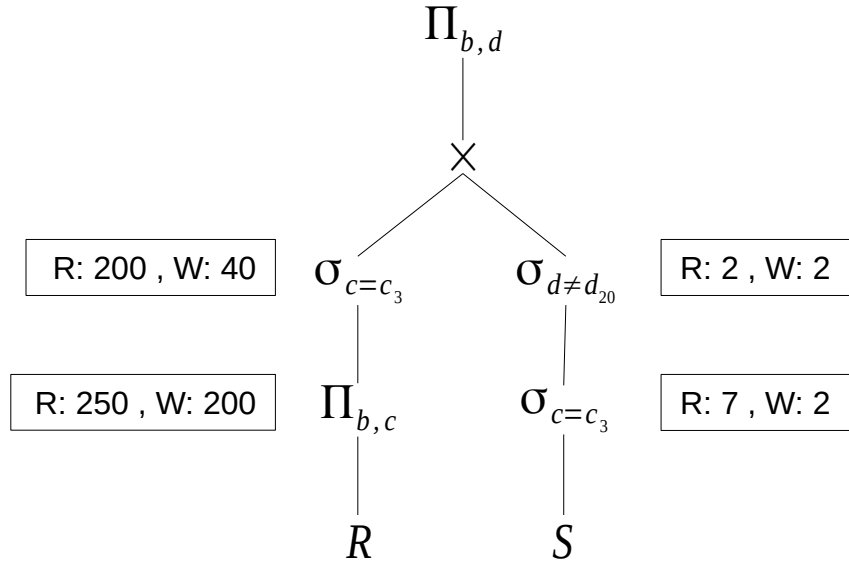
$$B(\sigma_{d \neq d_{20}}) = \text{hacia arriba} \left(\frac{N(\sigma_{d \neq d_{20}})}{Bfr(\sigma_{d \neq d_{20}})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{d \neq d_{20}})}{Bfr(\sigma_{c=c_3}^D)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{d \neq d_{20}})}{Bfr(S)} \right)$$

$$N(\sigma_{d \neq d_{20}}) = N(\sigma_{c=c_3}^D) \cdot 1 = 40 \cdot 1 = 40$$

de donde podemos calcular que:

$$B(\sigma_{d \neq d_{20}}) = \text{hacia arriba} \left(\frac{N(\sigma_{d \neq d_{20}})}{Bfr(S)} \right) = \text{hacia arriba} \left(\frac{40}{33} \right) = 2$$

Con esto, tenemos el cálculo de los bloques necesarios para calcular la relación derecha del producto cartesiano:



Ya podemos calcular el número de bloques necesarios para el cálculo del producto cartesiano. En concreto, habrá que multiplicar los registros de cada bloque de la relación izquierda por los de cada bloque de la relación derecha, lo cual requerirá leer el bloque de la relación derecha por cada bloque de la relación izquierda. En total, habrá que hacer:

$$40 \times 2 = 80 \text{ lecturas}$$

y escribir los bloques necesarios para almacenar el resultado:

$$B(\times) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}) \cdot N(\sigma_{d \neq d_{20}})}{Bfr(\times)} \right) = \text{hacia arriba} \left(\frac{1000 \cdot 40}{Bfr(\times)} \right)$$

$$Bfr(\times) = \text{parte entera} \left(\frac{B-C}{L(\times)} \right) = \text{parte entera} \left(\frac{2048-20}{L(\times)} \right)$$

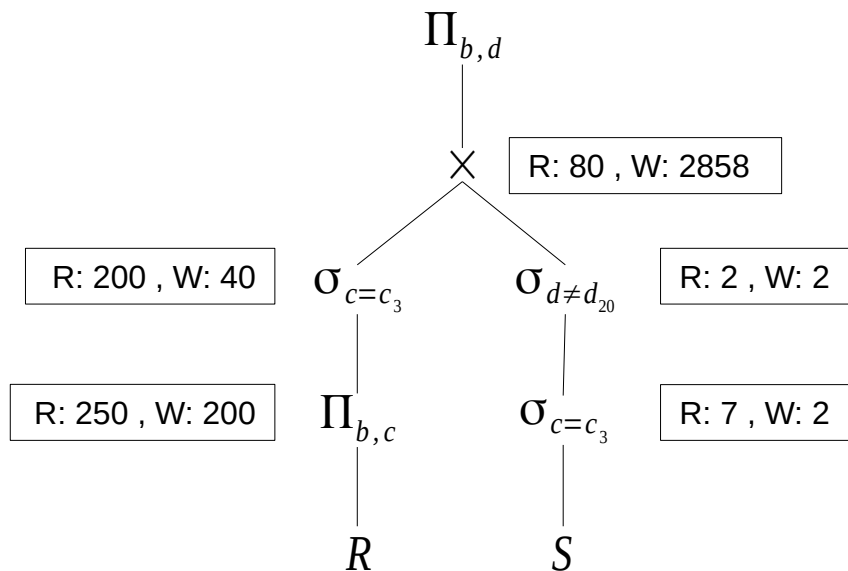
$$L(\times) = L(\sigma_{c=c_3}) + L(\sigma_{d \neq d_{20}}) = L(\Pi_{b,c}) + L(S) = 80 + 60 = 140$$

de lo que podemos calcular que:

$$Bfr(\times) = \text{parte entera} \left(\frac{2048-20}{140} \right) = 14$$

$$B(\times) = \text{hacia arriba} \left(\frac{1000 \cdot 40}{14} \right) = 2858$$

Con esto, sabemos el número de bloques necesarios para la lectura y escritura del producto cartesiano:



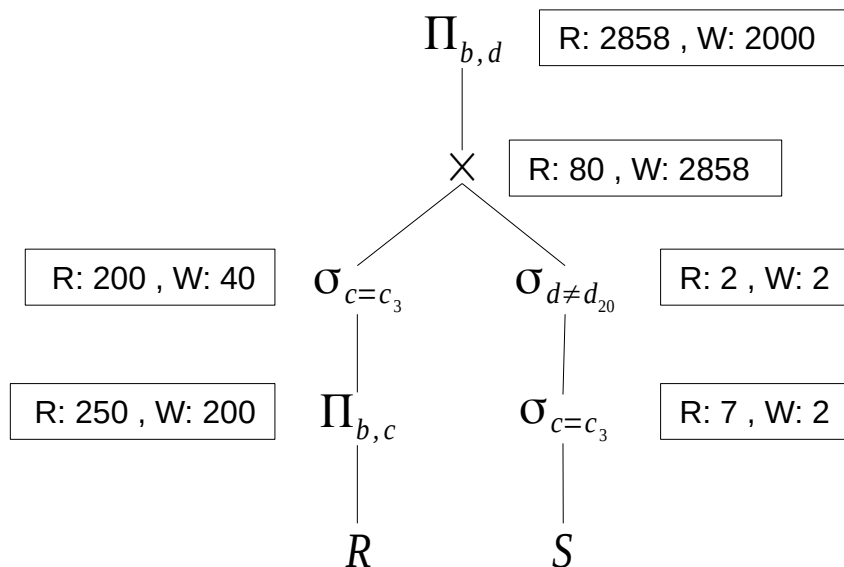
Únicamente nos resta el cálculo de los bloques necesarios para la proyección:

$$B(\Pi_{b,d}) = \text{hacia arriba} \left(\frac{N(\Pi_{b,d})}{Bfr(\Pi_{b,d})} \right) = \text{hacia arriba} \left(\frac{N(\times)}{Bfr(\Pi_{b,d})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}) \cdot N(\sigma_{d \neq d_{20}})}{Bfr(\Pi_{b,d})} \right)$$

de donde podemos calcular que:

$$B(\Pi_{b,d}) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}) \cdot N(\sigma_{d \neq d_{20}})}{Bfr(\Pi_{b,d})} \right) = \text{hacia arriba} \left(\frac{1000 \cdot 40}{20} \right) = 2000$$

Con esto, completamos los cálculos necesarios:



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:

$$(250+200)+(200+40)+(7+2)+(2+2)+(80+2858)+(2858+2000)=8499 \text{ bloques}$$

Obviamente, el presente plan mejora el plan propuesto por el enunciado, reduciendo el número de transferencias de bloques necesarias en un 96.5 %

2. **(2 punto)** Estima qué estructura de índice haría falta para el atributo c de la tabla R del ejercicio 1 y calcula su tamaño en bloques si $P=10B$ y el resto de los parámetros son los mismos que los del ejercicio 1.

Dada la variabilidad del atributo c en la relación R (5), y que dicha relación contiene 5000 registros, no podemos más que asumir que hay varios registros de R con el mismo valor de c . Si asumimos una distribución uniforme de los valores de c en los 5000 registros, tendríamos 1000 registros para cada uno de los 5 valores distintos de c .

Para que un índice sea útil, tiene que contener valores distintos, pero sólo hay 5 valores distintos y 5000 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 c .

El primer índice necesita un número de bloques igual a:

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

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

$$Bfr(I_1) = \text{parte entera} \left(\frac{2048 - 20}{L(I_1)} \right)$$

$$L(I_1) = 20 + 10 = 30$$

de modo que:

$$Bfr(I_1) = \text{parte entera} \left(\frac{2048 - 20}{30} \right) = 67 \quad y$$

$$B(I_1) = \text{hacia arriba} \left(\frac{N(I_1)}{Bfr(I_1)} \right) = \text{hacia arriba} \left(\frac{5}{67} \right) = 1$$

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

$$\begin{aligned} B(I_2) &= v(B, t) \cdot \text{hacia arriba} \left(\frac{N(\text{igual valor de } t)}{Bfr(I_2)} \right) = 5 \cdot \text{hacia arriba} \left(\frac{N(\text{igual valor de } t)}{Bfr(I_1)} \right) = \\ &= 5 \cdot \text{hacia arriba} \left(\frac{N(\text{igual valor de } t)}{67} \right) \end{aligned}$$

$$N(\text{igual valor de } t) = \frac{N(R)}{V(R, c)} = \frac{5000}{5} = 1000$$

de modo que:

$$B(I_2) = 5 \cdot \text{hacia arriba} \left(\frac{N(\text{igual valor de } t)}{67} \right) = 5 \cdot \text{hacia arriba} \left(\frac{1000}{67} \right) = 75$$

En total, haría falta 1 bloque para el primer nivel de índice y 75 bloques para el segundo:

$$B(I) = B(I_1) + B(I_2) = 1 + 75 = 76 \text{ bloques}$$

3. **(1 punto)** 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)

Corrección Examen de contenidos teóricos (Septiembre de 2014)

Primera parte

1. Sean las relaciones R y S con los siguientes parámetros:

R(a,b,c)	S(c,d)
N(R) = 5000	N(S)=200
V(R,a)=5000	
V(R,b)=300	
V(R,c)=5	V(S,c)=5
	V(S,d)=40
Size(a)=20	
Size(b)=60	
Size(c)=20	Size(c)=20
	Size(d)=40

Propón el plan lógico más eficiente para la siguiente consulta y justifica numéricamente por qué es más eficiente en base al número de bloques que emplea cada uno y teniendo en cuenta que el tamaño de bloque es de 2KB, que la cabecera es de 20B y que en memoria sólo cabe un bloque:

$$\Pi_{a,d}(\sigma_{(b \neq b_1)}(R JOIN S))$$

El plan más eficiente para la consulta planteada sería el generado por el árbol de expresión correspondiente a la consulta:

$$\Pi_{a,d}(\Pi_{a,c}(\sigma_{(b \neq b_1)}(R)) JOIN S)$$

Teniendo en cuenta que cada operación de E/S representa la transferencia de un bloque de datos entre el disco y la memoria, o viceversa, tendremos que calcular el número de bloques transferidos en total.

A fin de justificar que el plan que propongo es más eficiente que el planteado en el enunciado, tendré que calcular el número de bloques de E/S necesarios para uno y otro plan.

Comenzando por el planteado en el enunciado:

El número de bloques que almacenan la relación R es de:

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

Para ello, hay que calcular el factor de bloqueo de R como:

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

que dependerá de la longitud del registro de R:

$$L(R) = 20B + 60B + 20B = 100B$$

De modo que:

$$Bfr(R) = \text{parte entera} \left(\frac{2048-20}{100} \right) = 20$$

$$B(R) = \text{redondeo hacia arriba} \left(\frac{5000}{20} \right) = 250 \text{ bloques}$$

Si aplicamos los mismos cálculos para el número de bloques de la relación S:

$$L(S) = 20B + 40B = 60B$$

$$Bfr(S) = \text{parte entera} \left(\frac{2048 - 20}{60} \right) = 33$$

$$B(S) = \text{redondeo hacia arriba} \left(\frac{200}{33} \right) = 7 \text{ bloques}$$

Como ya se sabe, hay varios algoritmos para resolver la operación de reunión natural, pero la más eficiente, consiste en ordenar por el atributo de reunión aquella relación que no lo está (sólo una de las dos). Entendemos que R no lo está por lo que tenemos que aplicar un algoritmo de ordenación que requiere realizar un número de lecturas y escrituras de bloque de:

$$B(R) \cdot \log_2(B(R)) = 250 \cdot \log_2(250) = 1991,45 \approx 1992 \text{ bloques}$$

Ahora que las dos relaciones están ordenadas por el atributo c, podemos realizar la operación de reunión natural entre las dos. Habrá que leer:

$$250 + 7 = 257 \text{ bloques}$$

y escribir los bloques resultantes de la operación de reunión:

$$L(\text{JOIN}) = 20B + 60B + 20B + 40B = 140B$$

$$Bfr(\text{JOIN}) = \text{parte entera} \left(\frac{B - C}{L(\text{JOIN})} \right) = \text{parte entera} \left(\frac{2048 - 20}{140} \right) = 14$$

$$B(\text{JOIN}) = \text{redondeo hacia arriba} \left(\frac{N(\text{JOIN})}{Bfr(\text{JOIN})} \right) = \text{redondeo hacia arriba} \left(\frac{N(\text{JOIN})}{14} \right)$$

El número de registros de la operación de reunión natural viene dado por:

$$N(\text{JOIN}) = \frac{N(R) \cdot N(S)}{\max\{V(R, c), V(S, c)\}} = \frac{5000 \cdot 200}{\max\{5, 5\}} = 200000$$

por lo que el número de bloques resultantes que tendremos que escribir en disco será de:

$$B(\text{JOIN}) = \text{redondeo hacia arriba} \left(\frac{200000}{14} \right) = 14286 \text{ bloques}$$

A fin de aplicar la operación de selección sobre el atributo d, tendremos que leer de nuevo:

$$14286 \text{ bloques}$$

y el número de bloques que se transferirán al disco como resultado de la selección será de:

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

La estimación del número de tuplas que cumplen una condición de igualdad es de:

$$N(\text{SELEC}) = N(\text{JOIN}) = 200000$$

por lo que:

$$B(\text{SELEC}) = \text{redondeo hacia arriba} \left(\frac{200000}{14} \right) = 14286 \text{ bloques}$$

Por último, para realizar la operación de proyección, tendremos que leer:

14286 bloques

y escribir el número de bloques resultantes de dicha operación de proyección:

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

$$Bfr(\text{PROY}) = \text{parte entera} \left(\frac{B-C}{L(\text{PROY})} \right) = \text{parte entera} \left(\frac{2048-20}{L(\text{PROY})} \right)$$

$$L(\text{PROY}) = 20B + 40B = 60B$$

Con lo que:

$$Bfr(\text{PROY}) = \text{parte entera} \left(\frac{2048-20}{L(\text{PROY})} \right) = \text{parte entera} \left(\frac{2048-20}{60} \right) = 33$$

$$B(\text{PROY}) = \text{redondeo hacia arriba} \left(\frac{N(\text{SELEC})}{Bfr(\text{PROY})} \right) = \text{redondeo hacia arriba} \left(\frac{200000}{33} \right) = 6061 \text{ bloques}$$

De modo que el número total de lecturas y escrituras de bloques necesarios para resolver este plan físico es de:

$$1992 + 250 + 7 + 14286 + 14286 + 14286 + 14286 + 6061 = 65454 \text{ bloques}$$

El número de bloques involucrado en el plan físico de la consulta que propongo:

$$\Pi_{a,d}(\Pi_{a,c}(\sigma_{(b \neq b_1)}(R)) \text{ JOIN } S)$$

sería el que se muestra a continuación.

Como se ha mostrado en el plan físico de la consulta anterior, el número de bloques de la relación R es de 250 bloques, que habrá que leer para aplicar la operación de selección sobre la misma. El número de bloques resultante de dicha operación viene dado por los siguientes cálculos:

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

$$N(\text{SELEC}) = N(R) = 5000$$

con lo que:

$$B(\text{SELEC}) = \text{redondeo hacia arriba} \left(\frac{N(\text{SELEC})}{Bfr(R)} \right) = \text{redondeo hacia arriba} \left(\frac{5000}{20} \right) = 250 \text{ bloques}$$

Sobre el resultado de la operación de selección, es necesario aplicar la primera operación de proyección sobre los atributos a y c, que arroja un número de bloques resultantes como sigue:

$$B(\text{PROY1}) = \text{redondeo hacia arriba} \left(\frac{N(\text{PROY1})}{Bfr(\text{PROY1})} \right) = \text{redondeo hacia arriba} \left(\frac{N(\text{SELEC})}{Bfr(\text{PROY1})} \right)$$

$$Bfr(\text{PROY1}) = \text{parte entera} \left(\frac{B-C}{L(\text{PROY1})} \right) = \text{parte entera} \left(\frac{2048-20}{L(\text{PROY1})} \right)$$

$$L(\text{PROY1}) = 20B + 20B = 40B$$

Con lo que:

$$Bfr(\text{PROY1}) = \text{parte entera} \left(\frac{2048-20}{L(\text{PROY1})} \right) = \text{parte entera} \left(\frac{2048-20}{40} \right) = 50$$

$$B(\text{PROY1}) = \text{redondeo hacia arriba} \left(\frac{N(\text{SELEC})}{Bfr(\text{PROY1})} \right) = \text{redondeo hacia arriba} \left(\frac{5000}{50} \right) = 100 \text{ bloques}$$

El primer paso de la reunión natural, consiste en re-ordenar la relación R por el atributo de reunión c , lo cual tiene un coste en bloques de:

$$B(\text{PROY1}) \cdot \log_2(B(\text{PROY1})) = 100 \cdot \log_2(100) \approx 665 \text{ bloques}$$

Sería necesaria la lectura y escritura de 665 bloques para ordenar la relación R con respecto a c .

Con respecto a la relación S , está almacenada en 7 bloques (como se ha visto en el plan físico de la consulta planteada por el enunciado) que habrá que leer, junto con los bloques de la proyección calculada. La reunión natural arrojará un resultado en bloques de:

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

$$N(\text{JOIN}) = \frac{N(\text{PROY1}) \cdot N(S)}{\max\{V(R, c), V(S, c)\}} = \frac{N(\text{PROY1}) \cdot 200}{\max\{5, 5\}} = \frac{N(\text{SELEC}) \cdot 200}{5} = \frac{5000 \cdot 200}{5} = 200000$$

$$Bfr(\text{JOIN}) = \text{parte entera} \left(\frac{B - C}{L(\text{JOIN})} \right) = \text{parte entera} \left(\frac{2048 - 20}{L(\text{JOIN})} \right)$$

$$L(\text{JOIN}) = 20B + 20B + 40B = 80B$$

con lo que queda:

$$Bfr(\text{JOIN}) = \text{parte entera} \left(\frac{2028}{80} \right) = 25$$

$$B(\text{JOIN}) = \text{redondeo hacia arriba} \left(\frac{200000}{25} \right) = 8000 \text{ bloques}$$

Por último, y sobre el resultado de la reunión natural, será necesario leer los 8000 bloques para aplicar la segunda operación de proyección para eliminar el atributo c , lo que nos deja con un número de bloques como sigue:

$$B(\text{PROY2}) = \text{redondeo hacia arriba} \left(\frac{N(\text{PROY2})}{Bfr(\text{PROY2})} \right) = \text{redondeo hacia arriba} \left(\frac{N(\text{JOIN})}{Bfr(\text{PROY2})} \right)$$

$$Bfr(\text{PROY2}) = \text{parte entera} \left(\frac{2048 - 20}{L(\text{PROY2})} \right)$$

$$L(\text{PROY2}) = 20B + 40B = 60B$$

con lo que queda:

$$Bfr(\text{PROY2}) = \text{parte entera} \left(\frac{2048 - 20}{60} \right) = 33$$

$$B(\text{PROY2}) = \text{redondeo hacia arriba} \left(\frac{200000}{33} \right) \approx 6061 \text{ bloques}$$

De modo que el número total de lecturas y escrituras de bloques necesarios para resolver este plan físico es de:

$$250 + 250 + 250 + 100 + 665 + 7 + 100 + 8000 + 8000 + 6061 = 23683 \text{ bloques}$$

opción que supone menos de la mitad de las lecturas y escrituras de bloques implicadas por el plan físico correspondiente a la consulta planteada por el enunciado.

2. Enuncia brevemente usando tus propias palabras (o con la fórmula si la recuerdas y lo prefieres):
- a) **(1 punto)** ¿De qué depende el tiempo de búsqueda de un registro por su clave en un fichero ASI (archivo secuencial indexado) multinivel con zona de desbordamiento, si la búsqueda se realiza por el valor de clave?.
 - b) **(1 punto)** Enumera los elementos de los que se compone el nivel interno de Oracle Database®.
 - c) **(1 punto)** Los índices son útiles para búsquedas por valor de clave pero son costosos de mantener. Enumera las cuatro situaciones en las que es conveniente eliminarlos.
-

Apartado (a)

Depende del tiempo necesario para encontrar la clave en el índice raíz, el tiempo necesario para encontrar la clave en cada uno de los índices intermedios y el tiempo para acceder a un registro del fichero maestro por posición, además del tiempo para encontrar el registro en la cadena de desbordamiento, en caso de que no estuviera presente en la posición indicada del fichero maestro.

$$T_F = T_M + (m - 1) \cdot T_{F_i} + T'_F + T_{F \text{ Cadena}}$$

Apartado (b)

La base de datos se compone de varios *tablespaces*, que se componen, a su vez, de varios segmentos, que se componen de varias extensiones, que se componen de varios bloques.

Apartado (c)

Los índices deben eliminarse cuando concurra una de las siguientes situaciones:

- Ya no sirven
- No mejoran la eficiencia
- Hay que cambiar los campos que se indexan
- Hay que rehacerlo

Relación de problemas resuelta 8

1. (3 puntos) Sean las relaciones R y S con los siguientes parámetros:

R(a,b,c)	S(c,d)
N(R) = 5000	N(S)=200
V(R,a)=5000	
V(R,b)=300	
V(R,c)=5	V(S,c)=5
	V(S,d)=40
Size(a)=20	
Size(b)=60	
Size(c)=20	Size(c)=20
	Size(d)=40

Propón el plan lógico más eficiente para la siguiente consulta y justifica numéricamente por qué es más eficiente en base al número de bloques que emplea cada uno y teniendo en cuenta que el tamaño de bloque es de 2KB, que la cabecera es de 20B y que en memoria sólo cabe un bloque:

$$\Pi_{b,d}(\sigma_{(c=c_3 \wedge d < d_{20})}(R \times S))$$

Estimación del número de bloques para el plan de la consulta del ejercicio

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.

$$\begin{array}{c}
 \Pi_{b,d} \\
 | \\
 \sigma_{c=c_3} \\
 | \\
 \sigma_{d < d_{20}} \\
 | \\
 \times \\
 / \quad \backslash \\
 R \quad \quad S
 \end{array}$$

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

$$B(R) = \text{hacia arriba} \left(\frac{N(R)}{Bfr(R)} \right) = \text{hacia arriba} \left(\frac{5000}{Bfr(R)} \right)$$

$$Bfr(R) = \text{parte entera} \left(\frac{B-C}{L(R)} \right) = \text{parte entera} \left(\frac{2048-20}{L(R)} \right)$$

$$L(R) = 20 + 60 + 20 = 100$$

de modo que:

$$Bfr(R) = \text{parte entera} \left(\frac{2048-20}{100} \right) = 20 \quad \text{y}$$

$$B(R) = \text{hacia arriba} \left(\frac{5000}{20} \right) = 250$$

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

$$B(S) = \text{hacia arriba} \left(\frac{N(S)}{Bfr(S)} \right) = \text{hacia arriba} \left(\frac{200}{Bfr(S)} \right)$$

$$Bfr(S) = \text{parte entera} \left(\frac{B-C}{L(S)} \right) = \text{parte entera} \left(\frac{2048-20}{L(S)} \right)$$

$$L(S) = 20 + 40 = 60$$

de modo que:

$$Bfr(S) = \text{parte entera} \left(\frac{2048 - 20}{60} \right) = 33 \quad \text{y}$$

$$B(S) = \text{hacia arriba} \left(\frac{200}{33} \right) = 7$$

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

$$\text{Lectura: } B(R) \cdot B(S) = 250 \cdot 7 = 1750$$

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

$$B(\times) = \text{hacia arriba} \left(\frac{N(\times)}{Bfr(\times)} \right) = \text{hacia arriba} \left(\frac{5000 \cdot 200}{Bfr(\times)} \right)$$

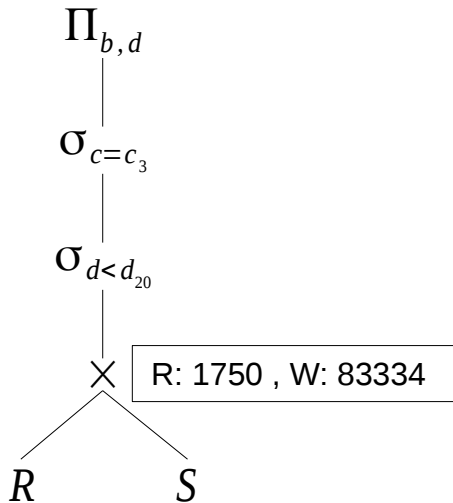
$$Bfr(\times) = \text{parte entera} \left(\frac{B-C}{L(\times)} \right) = \text{parte entera} \left(\frac{2048-20}{L(\times)} \right)$$

$$L(\times) = L(R) + L(S) = 100 + 60 = 160$$

de modo que:

$$Bfr(\times) = \text{parte entera} \left(\frac{2048-20}{L(\times)} \right) = \text{parte entera} \left(\frac{2048-20}{160} \right) = 12 \quad \text{y}$$

$$\text{Escritura: } B(\times) = \text{hacia arriba} \left(\frac{5000 \cdot 200}{Bfr(\times)} \right) = \text{hacia arriba} \left(\frac{5000 \cdot 200}{12} \right) = 83334$$

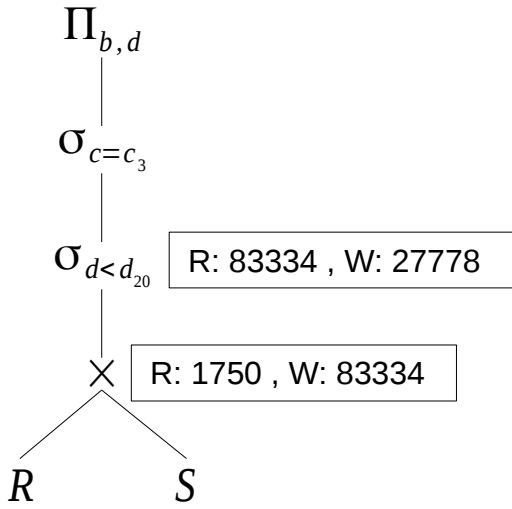


La primera operación de selección debe leer los bloques resultantes de la operación anterior y escribir tantos bloques como sean necesarios para almacenar los registros que resultan de seleccionar aquellos que cumplan la condición, es decir:

$$\begin{aligned}
 B(\sigma_{d < d_{20}}) &= \text{hacia arriba} \left(\frac{N(\sigma_{d < d_{20}})}{Bfr(\sigma_{d < d_{20}})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{d < d_{20}})}{Bfr(\times)} \right) = \\
 &= \text{hacia arriba} \left(\frac{N(\sigma_{d < d_{20}})}{12} \right) \\
 N(\sigma_{d < d_{20}}) &= \frac{N(\times)}{3} = \frac{5000 \cdot 200}{3} = 333334
 \end{aligned}$$

de modo que:

Escritura: $B(\sigma_{d < d_{20}}) = \text{hacia arriba} \left(\frac{N(\sigma_{d < d_{20}})}{12} \right) = \text{hacia arriba} \left(\frac{333334}{12} \right) = 27778$

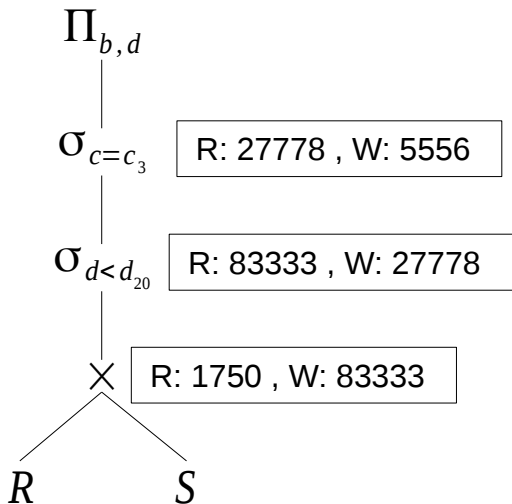


La segunda operación de selección debe leer los bloques resultantes de la operación anterior y escribir tantos bloques como sean necesarios para almacenar los registros que resultan de seleccionar aquellos que cumplan la condición, es decir:

$$\begin{aligned}
 B(\sigma_{c=c_3}) &= \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{Bfr(\sigma_{c=c_3})} \right) = \\
 &= \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{Bfr(\times)} \right) = \\
 &= \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{12} \right) \\
 N(\sigma_{c=c_3}) &= \frac{N(\sigma_{d < d_{20}})}{V(R, c)} = \frac{333334}{5} \approx 66667
 \end{aligned}$$

De este modo, se escriben:

$$B(\sigma_{c=c_3}) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{12} \right) = \text{hacia arriba} \left(\frac{66667}{12} \right) = 5556$$



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:

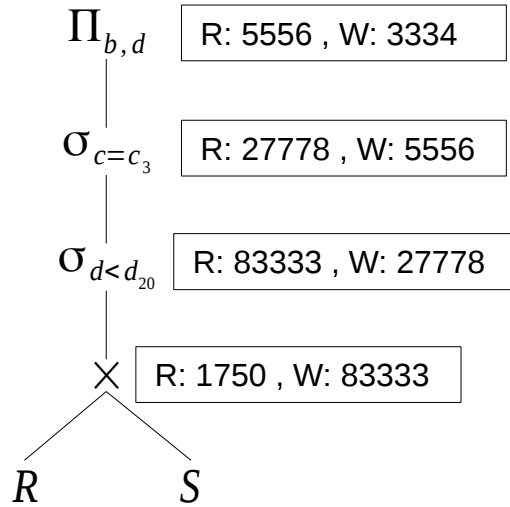
$$\begin{aligned}
 B(\Pi_{b,d}) &= \text{hacia arriba} \left(\frac{N(\Pi_{b,d})}{Bfr(\Pi_{b,d})} \right) = \\
 &= \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{Bfr(\Pi_{b,d})} \right) = \text{hacia arriba} \left(\frac{66667}{Bfr(\Pi_{b,d})} \right)
 \end{aligned}$$

$$L(\Pi_{b,d}) = 60 + 40 = 100$$

de modo que:

$$Bfr(\Pi_{b,d}) = \text{parte entera} \left(\frac{2048-20}{L(\Pi_{b,d})} \right) = \text{parte entera} \left(\frac{2048-20}{100} \right) = 20 \quad y$$

$$\text{Escritura: } B(\Pi_{b,d}) = \text{hacia arriba} \left(\frac{66667}{Bfr(\Pi_{b,d})} \right) = \text{hacia arriba} \left(\frac{66667}{20} \right) = 3334$$

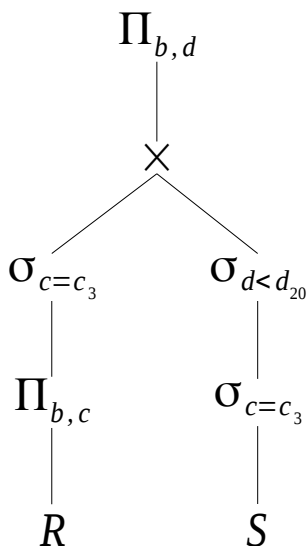


Una vez estimados el número de bloques que cada operación lee y escribe, podemos estimar que el número total de bloques leídos y escritos para la resolución del plan físico correspondiente es de:

$$(1750+83334)+(83334+27778)+(27778+5556)+(5556+3334)=238420 \text{ bloques}$$

Estimación del número de bloques para el plan optimizado para la consulta del ejercicio

El plan lógico propuesto para la optimización del plan lógico planteado por la consulta sería el que se ve a continuación:



Para calcular el número de bloques estimados para resolver el plan físico, reutilizaremos los cálculos realizados en la parte anterior cuando sea posible.

$$B(\Pi_{b,c}) = \text{hacia arriba} \left(\frac{N(R)}{Bfr(\Pi_{b,c})} \right) = \text{hacia arriba} \left(\frac{5000}{Bfr(\Pi_{b,c})} \right)$$

$$Bfr(\Pi_{b,c}) = \text{parte entera} \left(\frac{B-C}{L(\Pi_{b,c})} \right) = \text{parte entera} \left(\frac{2048-20}{L(\Pi_{b,c})} \right)$$

$$L(\Pi_{b,c}) = 60 + 20 = 80$$

de donde podemos calcular que:

$$Bfr(\Pi_{b,c}) = \text{parte entera} \left(\frac{2048-20}{80} \right) = 25$$

$$B(\Pi_{b,c}) = \text{hacia arriba} \left(\frac{5000}{25} \right) = 200$$

Después, procedemos para calcular el número de bloques necesarios para escribir la selección de la rama izquierda:

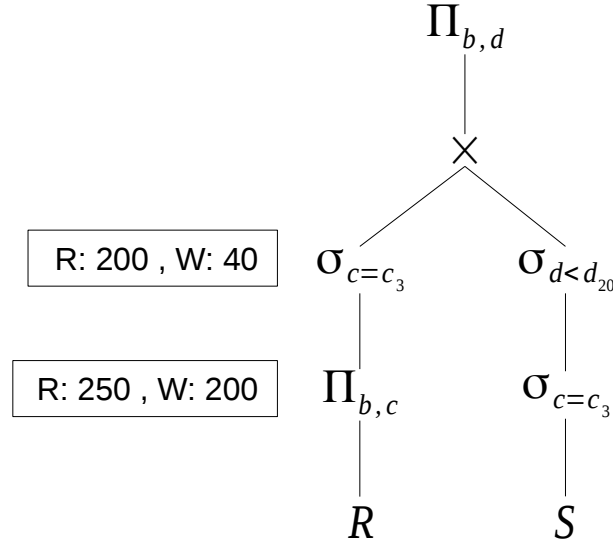
$$B(\sigma_{c=c_3}) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{Bfr(\sigma_{c=c_3})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{Bfr(\Pi_{b,c})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{25} \right)$$

$$N(\sigma_{c=c_3}) = \frac{N(\Pi_{b,c})}{V(R,c)} = \frac{N(R)}{V(R,c)} = \frac{5000}{5} = 1000$$

de donde podemos calcular que:

$$B(\sigma_{c=c_3}) = \text{hacia arriba} \left(\frac{1000}{25} \right) = 40$$

Con esto, podemos concluir el número de bloques necesario para calcular la relación izquierda del producto cartesiano:



Procedemos a realizar los cálculos para la estimación de la relación derecha del producto cartesiano:

$$B(\sigma_{c=c_3}^D) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}^D)}{Bfr(\sigma_{c=c_3}^D)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}^D)}{Bfr(S)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}^D)}{33} \right)$$

$$N(\sigma_{c=c_3}^D) = \frac{N(S)}{V(S,c)} = \frac{200}{5} = 40$$

de donde podemos calcular que:

$$B(\sigma_{c=c_3}^D) = \text{hacia arriba} \left(\frac{40}{33} \right) = 2$$

A continuación, podemos aplicar la segunda selección y estimar los bloques necesarios:

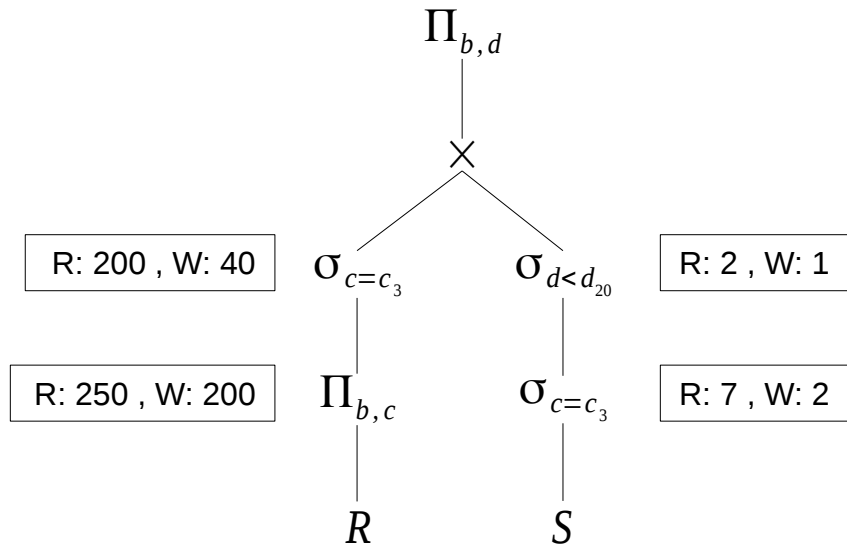
$$B(\sigma_{d < d_{20}}) = \text{hacia arriba} \left(\frac{N(\sigma_{d < d_{20}})}{Bfr(\sigma_{d < d_{20}})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{d < d_{20}})}{Bfr(\sigma_{c=c_3}^D)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{d < d_{20}})}{Bfr(S)} \right)$$

$$N(\sigma_{d < d_{20}}) = \frac{N(\sigma_{c=c_3}^D)}{3} = \frac{40}{3} \approx 14$$

de donde podemos calcular que:

$$B(\sigma_{d < d_{20}}) = \text{hacia arriba} \left(\frac{14}{33} \right) = 1$$

Con esto, tenemos el cálculo de los bloques necesarios para calcular la relación derecha del producto cartesiano:



Ya podemos calcular el número de bloques necesarios para el cálculo del producto cartesiano. En concreto, habrá que multiplicar los registros de cada bloque de la relación izquierda por los de cada bloque de la relación derecha, lo cual requerirá leer el bloque de la relación derecha por cada bloque de la relación izquierda. En total, habrá que hacer:

$$40 \times 1 = 40 \text{ lecturas}$$

y escribir los bloques necesarios para almacenar el resultado:

$$B(\times) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}) \cdot N(\sigma_{d<d_{20}})}{Bfr(\times)} \right) = \text{hacia arriba} \left(\frac{1000 \cdot 14}{Bfr(\times)} \right)$$

$$Bfr(\times) = \text{parte entera} \left(\frac{B-C}{L(\times)} \right) = \text{parte entera} \left(\frac{2048-20}{L(\times)} \right)$$

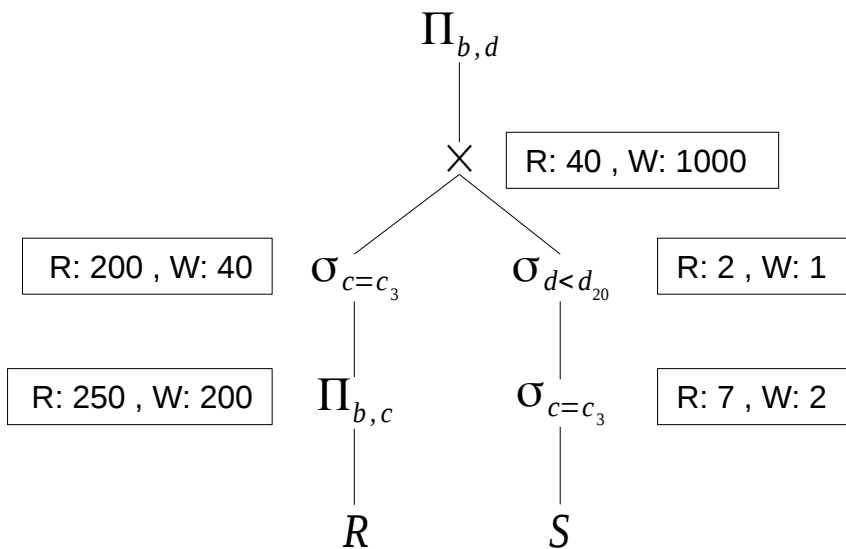
$$L(\times) = L(\sigma_{c=c_3}) + L(\sigma_{d<d_{20}}) = L(\Pi_{b,c}) + L(S) = 80 + 60 = 140$$

de lo que podemos calcular que:

$$Bfr(\times) = \text{parte entera} \left(\frac{2048-20}{140} \right) = 14$$

$$B(\times) = \text{hacia arriba} \left(\frac{1000 \cdot 14}{14} \right) = 1000$$

Con esto, sabemos el número de bloques necesarios para la lectura y escritura del producto cartesiano:



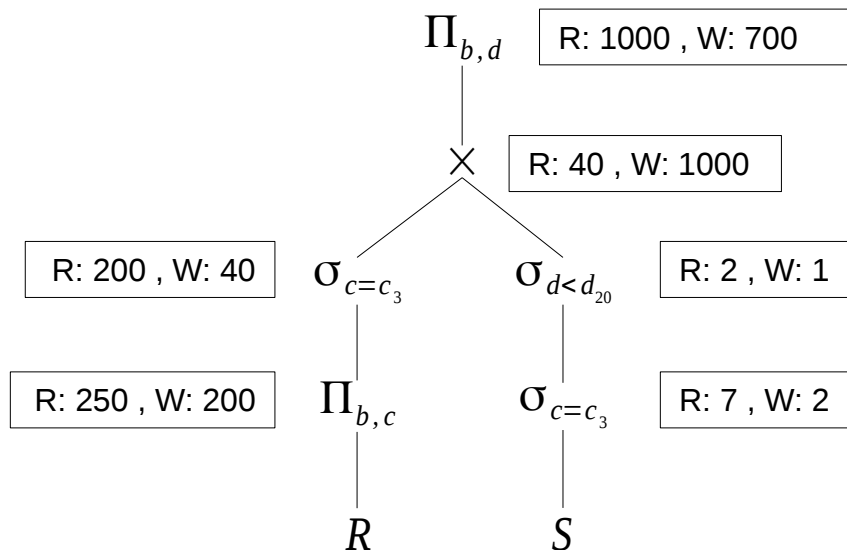
Únicamente nos resta el cálculo de los bloques necesarios para la proyección:

$$B(\Pi_{b,d}) = \text{hacia arriba} \left(\frac{N(\Pi_{b,d})}{Bfr(\Pi_{b,d})} \right) = \text{hacia arriba} \left(\frac{N(\times)}{Bfr(\Pi_{b,d})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}) \cdot N(\sigma_{d<d_{20}})}{Bfr(\Pi_{b,d})} \right)$$

de donde podemos calcular que:

$$B(\Pi_{b,d}) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}) \cdot N(\sigma_{d<d_{20}})}{Bfr(\Pi_{b,d})} \right) = \text{hacia arriba} \left(\frac{1000 \cdot 14}{20} \right) = 700$$

Con esto, completamos los cálculos necesarios:



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:

$$(200+250)+(200+40)+(7+2)+(2+1)+(40+1000)+(1000+700)=3442 \text{ bloques}$$

2. **(2 punto)** Estima qué estructura de índice haría falta para el atributo c de la tabla R del ejercicio 1 y calcula su tamaño en bloques si $P=10B$ y el resto de los parámetros son los mismos que los del ejercicio 1.

Dada la variabilidad del atributo c en la relación R (5), y que dicha relación contiene 5000 registros, no podemos más que asumir que hay varios registros de R con el mismo valor de c . Si asumimos una distribución uniforme de los valores de c en los 5000 registros, tendríamos 1000 registros para cada uno de los 5 valores distintos de c .

Para que un índice sea útil, tiene que contener valores distintos, pero sólo hay 5 valores distintos y 5000 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 c .

El primer índice necesita un número de bloques igual a:

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

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

$$Bfr(I_1) = \text{parte entera} \left(\frac{2048 - 20}{L(I_1)} \right)$$

$$L(I_1) = 20 + 10 = 30$$

de modo que:

$$Bfr(I_1) = \text{parte entera} \left(\frac{2048 - 20}{30} \right) = 67 \quad y$$

$$B(I_1) = \text{hacia arriba} \left(\frac{N(I_1)}{Bfr(I_1)} \right) = \text{hacia arriba} \left(\frac{5}{67} \right) = 1$$

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

$$\begin{aligned} B(I_2) &= v(B, t) \cdot \text{hacia arriba} \left(\frac{N(\text{igual valor de } t)}{Bfr(I_2)} \right) = 5 \cdot \text{hacia arriba} \left(\frac{N(\text{igual valor de } t)}{Bfr(I_1)} \right) = \\ &= 5 \cdot \text{hacia arriba} \left(\frac{N(\text{igual valor de } t)}{67} \right) \end{aligned}$$

$$N(\text{igual valor de } t) = \frac{N(R)}{V(R, c)} = \frac{5000}{5} = 1000$$

de modo que:

$$B(I_2) = 5 \cdot \text{hacia arriba} \left(\frac{N(\text{igual valor de } t)}{67} \right) = 5 \cdot \text{hacia arriba} \left(\frac{1000}{67} \right) = 75$$

En total, haría falta 1 bloque para el primer nivel de índice y 75 bloques para el segundo:

$$B(I) = B(I_1) + B(I_2) = 1 + 75 = 76 \text{ bloques}$$

3. **(1 punto)** 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)

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.

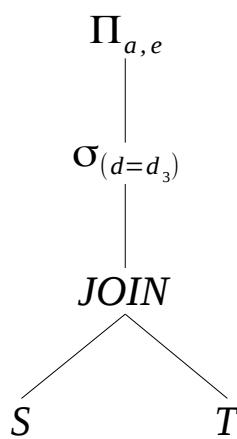
Dado que los tres atributos mencionados son claves, podemos concluir que:

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

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:

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

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

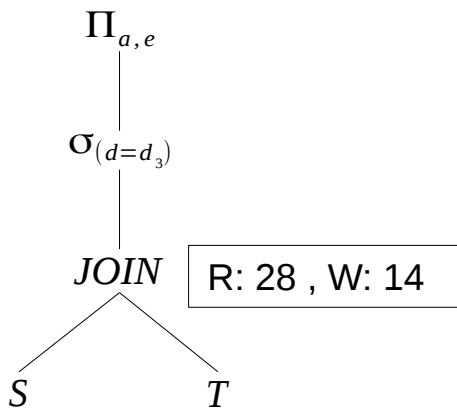
$$Bfr(JOIN) = \text{parte entera} \left(\frac{B-C}{L(JOIN)} \right) = \text{parte entera} \left(\frac{4096-20}{L(JOIN)} \right)$$

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

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:

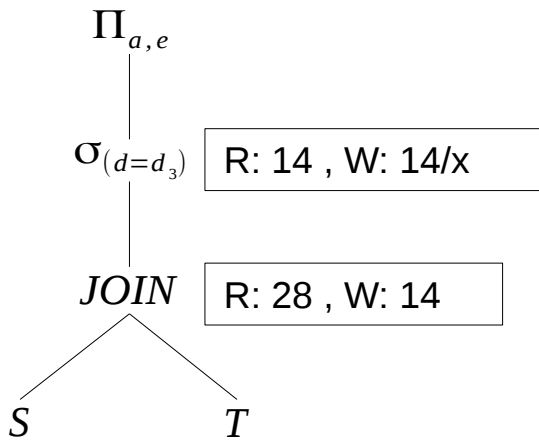
$$\begin{aligned} B(\sigma_{d=d_3}) &= \text{hacia arriba} \left(\frac{N(\sigma_{d=d_3})}{Bfr(\sigma_{d=d_3})} \right) = \\ &= \text{hacia arriba} \left(\frac{N(\sigma_{d=d_3})}{Bfr(JOIN)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{d=d_3})}{37} \right) \end{aligned}$$

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:

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

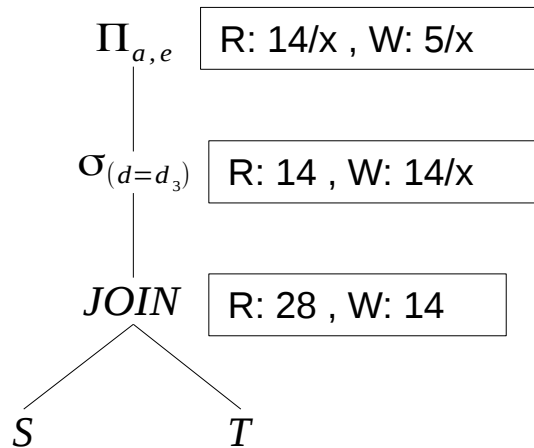
$$Bfr(\Pi_{a,e}) = \text{parte entera} \left(\frac{B-C}{L(\Pi_{a,e})} \right) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{a,e})} \right)$$

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

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

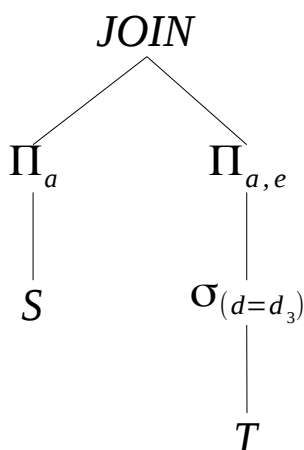
Enunciado

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.

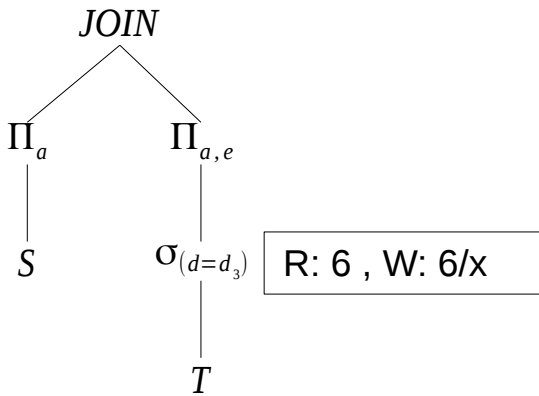
A fin de aplicar la selección sobre los registros de la relación T , es necesario leer todos los registros, almacenados en 6 bloques, y escribir en los bloques del resultado aquellos que cumplan la condición indicada. El número de bloques necesarios para escribir el resultado se calcula a continuación.

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

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

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.

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

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

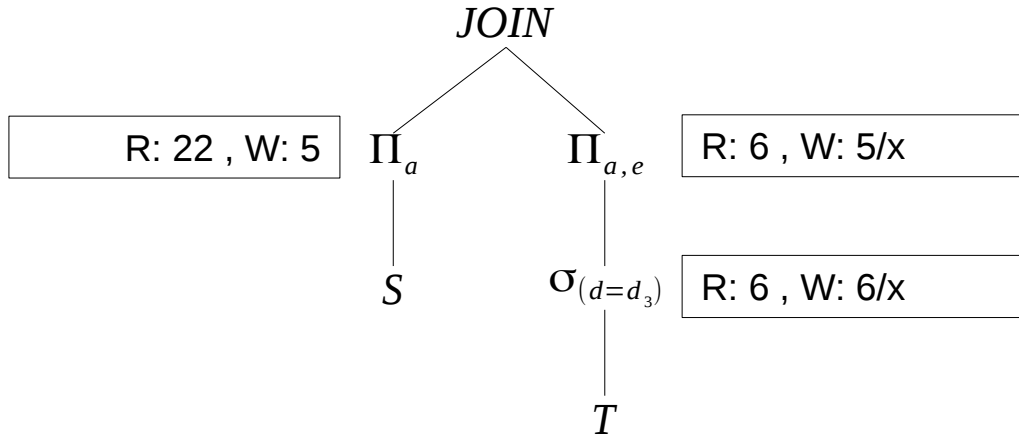
$$L(\Pi_a)=20$$

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.

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

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

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

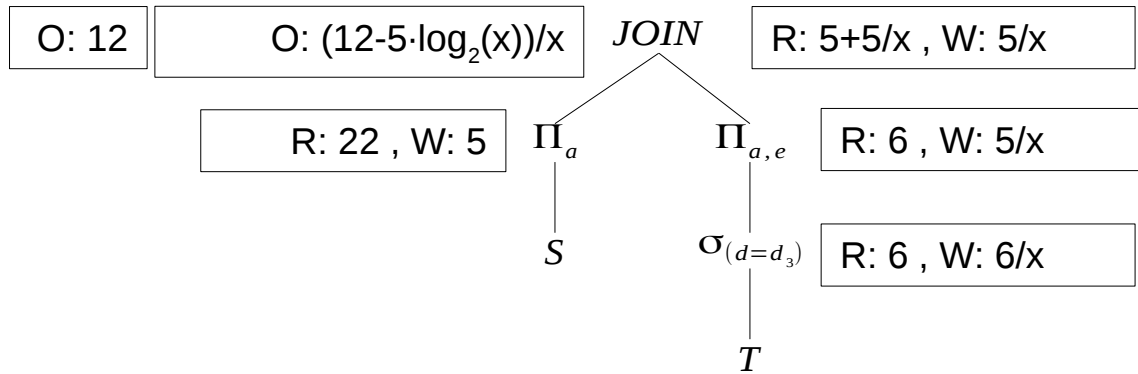
$$L(JOIN) = L(\Pi_a) + L(\Pi_{a,e}) - \text{Size}(a) = 20 + 40 - 20 = 40$$

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

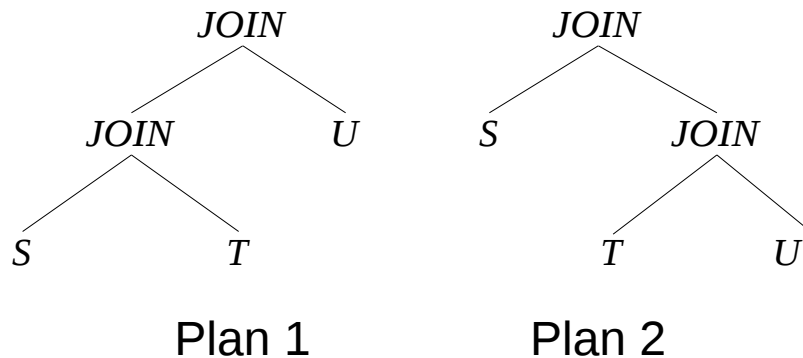
Enunciado

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:

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

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

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



Grado en Ingeniería Informática

Examen teórico – Convocatoria de Junio de 2013

Nombre: _____

Primera parte

1. (3 puntos) Sean las relaciones R y S con los siguientes parámetros:

R(a,b,c)	S(c,d)
N(R) = 5000	N(S)=200
V(R,a)=5000	
V(R,b)=3000	
V(R,c)=5	V(S,c)=5
	V(S,d)=40
Size(a)=20	
Size(b)=60	
Size(c)=20	Size(c)=20
	Size(d)=40

Propón dos planes lógicos para la siguiente consulta e indica qué plan físico se elegirá en base al número de bloques que se utilizan, teniendo en cuenta que el tamaño de bloque es de 2KB, que la cabecera es de 20B y que en memoria sólo cabe un bloque:

$$\Pi_{a,d}(\sigma_{b=b_1}(R \text{ JOIN } S))$$

Pista: uno de los planes es igual que el de la consulta; el otro se obtiene moviendo la selección.

2. Enuncia brevemente usando tus propias palabras (o con la fórmula si la recuerdas y lo prefieres):
- (1 puntos) ¿De qué depende el tiempo de búsqueda de un registro por su clave en un fichero ASI (archivo secuencial indexado) con índice denso y zona de desbordamiento, si la búsqueda se realiza por el valor de clave?
 - (1 puntos) ¿Cómo se calcula la inserción en un fichero AAD (archivo de acceso directo) con encadenamiento abierto?
 - (1 puntos) ¿Qué información se puede consultar a través de las vistas de catálogo con prefijo USER_? ¿Y con las vistas con prefijo ALL_? ¿Y con las vistas con prefijo DBA_?

Segunda parte

3. Considera el siguiente par de transacciones entrelazadas (y que no se consideran bloqueos)
- (0,5 puntos) Rellena la tabla de modificaciones en los espacios vacíos de la derecha considerando que los valores iniciales son A=2 y B=3
 - (1 puntos) Si el sistema sufre un fallo justo después de la sentencia "Escribe (C,t)", ¿qué hará el gestor de recuperaciones con la transacción 1? ¿Y con la transacción 2? ¿Lo que hace con la transacción 2 depende de algún otro factor que no aparece en la transacción pero puede estar escrito en la tabla de modificaciones en el momento del fallo?



T1	T2	T _i	Estado	Operación	Dato	V antiguo	V nuevo
Lee (A, x)							
x := x + 1							
Escribe (A, x)	Lee (A, y)						
	Lee (B, z)						
	t := y + z						
	Escribe (B, t)						
	Escribe (C, t)						

4. Considera las siguientes transacciones y responde a las siguientes cuestiones:

T1	T2	T3
Lee (A, x)		
x := x + 3		
	Lee (A, t)	Lee (C, u)
	t := t * 2	Lee (A, v)
	Escribe (A, t)	w := u + v
Escribe (A, x)		Escribe (C, w)
Lee (C, z)		
z := z + x		
Escribe (C, z)		

- (0,5 puntos)** dibuja el grafo de precedencia indicando qué transacción espera, a qué transacción espera, por culpa de qué átomo tiene que esperar y a causa de qué modo de acceso al átomo.
- (1 puntos)** ¿en qué orden se ejecutan las transacciones según el **algoritmo de ordenación parcial**?
- (1 puntos)** ¿en qué orden se ejecutan las transacciones según el **método de bloqueo en dos fases** suponiendo que: el bloqueo del átomo se realiza justo antes de la primera lectura de dicho átomo, que los desbloques se

producen después de la última sentencia de la transacción y considerando que las *lecturas* son *no protegidas* (M1) y las *escrituras protegidas* (M4)? ¿Existe situación de interbloqueo (*deadlock*)? ¿En caso de existir, cómo se resuelve?

Pista: M1 es compatible con M4, M1 es compatible con M1 pero M4 no es compatible con M4

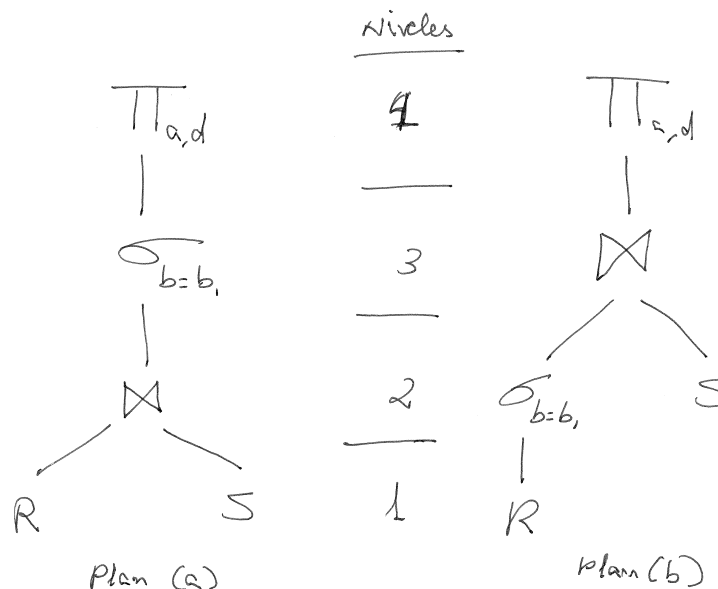
Notas:

- Duración del examen:** 1 hora y media
- Todos los alumnos y todas las alumnas tienen que realizar los ejercicios de la segunda parte.
- Aquellos alumnos y aquellas alumnas que hayan superado la prueba parcial, no están obligados a realizar los ejercicios de la primera parte. En ese caso, la calificación obtenida en dicha prueba se escalará sobre 6.

Resolución

Pregunta 1

Atendiendo al enunciado, los dos planes lógicos propuestos son:



Teniendo en cuenta los dos planes propuestos, calculemos ahora el número de bloques empleados en cada uno de los cuatro niveles para ver qué plan físico sería el escogido para ejecutar.

Plan (a)

En el plan (a), tenemos dos relaciones R y S que tienen que ser leídas:

$$L(R) = 20 \text{ B} + 60 \text{ B} + 20 \text{ B} = 100 \text{ B}$$

$$\text{Bfr}(R) = \text{parte entera}((B - C) / L(R)) = \text{parte entera}((2048 \text{ B} - 20 \text{ B}) / 100 \text{ B}) = \text{parte entera}(20,28) = 20$$

$$\text{Nbloques}(R) = \text{redondeo hacia arriba}(N(R) / \text{Bfr}(R)) = \text{redondeo hacia arriba}(5000 / 20) = 250 \text{ bloques}$$

Teniendo en cuenta que la clave de R sólo puede ser a puesto que es el único atributo que tiene tantos valores como tuplas, y que queremos reunir con la relación S que debe estar ordenada por c (que es el atributo de reunión), es necesario reordenar la relación R para poder reunirla con la relación S .

La reordenación de R consiste en reordenar los bloques de R internamente y entre sí, por lo que el algoritmo más eficiente de ordenación que conocemos tarda un tiempo aproximado de $250 \cdot \log_2(250) = 250 \cdot 7,965784285 = 1991,446071166$, es decir, aproximadamente requiere la transferencia de 1992 bloques para reordenar los registros de R .

Antes de seguir, necesitamos saber el número de bloques de S :

$$L(S) = 20 \text{ B} + 40 \text{ B} = 60 \text{ B}$$

$$\text{Bfr}(S) = \text{parte entera}((B - C) / L(S)) = \text{parte entera}((2048 \text{ B} - 20 \text{ B}) / 60 \text{ B}) = \text{parte entera}(33,8) = 33$$

$N_{\text{bloques}}(S) = \text{redondeo hacia arriba } (N(S) / B_{\text{fr}}(S)) = \text{redondeo hacia arriba } (200 / 33) = \text{redondeo hacia arriba } (6,060606061) = 7 \text{ bloques}$

Para conseguir la reunión natural, sólo necesitamos recorrer las dos relaciones ordenadas por c :

250 bloques + 7 bloques

y escribir la relación resultante:

$L(\text{JOIN}) = 20 B + 60 B + 20 B + 40 B = 140 B$

$B_{\text{fr}}(\text{JOIN}) = \text{parte entera } ((B - C) / L(\text{JOIN})) = \text{parte entera } ((2048 B - 20 B) / 140 B) = \text{parte entera } (14,485714286) = 14$

$N(\text{JOIN}) = N(R) \cdot N(S) / \max\{V(R,c), V(S,b)\} = 5000 \cdot 200 / 5 = 200000$

$N_{\text{bloques}}(\text{JOIN}) = \text{redondeo hacia arriba } (N(S) / B_{\text{fr}}(\text{JOIN})) = \text{redondeo hacia arriba } (200000 / 14) = \text{redondeo hacia arriba } (14285,714285714) = 14286 \text{ bloques}$

Cuando ya tenemos la reunión, es necesario aplicar la selección, es decir, leer los bloques de la reunión, aplicarla y escribir el resultado de la selección:

$N(\text{SELECCIÓN}) = 200000 / 3000 = 66,666666667$, o sea, aproximadamente 67 tuplas

$N_{\text{bloques}}(\text{SELECCIÓN}) = \text{redondeo hacia arriba } (N(\text{SELECCIÓN}) / B_{\text{fr}}(\text{JOIN})) = \text{redondeo hacia arriba } (67 / 14) = \text{redondeo hacia arriba } (4,785714286) = 5 \text{ bloques}$

Por último, habría que aplicar la proyección a estos cinco bloques:

$L(\text{PROYECCIÓN}) = 20 B + 40 B = 60 B$

$B_{\text{fr}}(\text{PROYECCIÓN}) = \text{parte entera } ((B - C) / L(\text{PROYECCIÓN})) = \text{parte entera } ((2048 B - 20B) / 60B) = \text{parte entera } (33,8) = 33$

$N_{\text{bloques}}(\text{PROYECCIÓN}) = \text{redondeo hacia arriba } (N(\text{PROYECCIÓN}) / B_{\text{fr}}(\text{PROYECCIÓN})) = \text{redondeo hacia arriba } (67 / 33) = \text{redondeo hacia arriba } (2,03030303) = 3 \text{ bloques}$

Teniendo en cuenta que el resultado de cada operación intermedia tiene que ser escrito a disco y leído después, podemos resumir que el número de bloques leídos y escritos en este plan físico es de:

- Reordenar R requiere transferir 1992 bloques
- Mezclar R reordenado y S (resolver el JOIN) requiere leer 250 + 7 bloques y escribir 14286 bloques
- Aplicar la selección requiere leer 14286 bloques y escribir 5 bloques
- Aplicar la proyección requiere leer 5 bloques y escribir 3 bloques

En total, para resolver este plan es necesario transferir 30834 bloques

Plan (b)

En este plan, podemos aprovecharnos de algunos cálculos realizados para el plan anterior.

Para aplicar la selección sobre R , necesitamos leer los bloques de dicha relación:

$N_{\text{bloques}}(R) = 250 \text{ bloques}$

Ahora, aplicamos la selección y calculamos el número de bloques:

$N(\text{SELECCIÓN}) = N(R) / V(R,b) = 5000 / 3000 = 1,666666667$, es decir, aproximadamente 2 registros

$N_{\text{bloques}}(\text{SELECCIÓN}) = \text{redondeo hacia arriba } (N(\text{SELECCIÓN}) / B_{\text{fr}}(R)) = \text{redondeo hacia arriba } (2 / 20) = \text{redondeo hacia arriba } (0,1) = 1 \text{ bloque}$

Para realizar la reunión, es preciso que el resultado de la selección esté ordenado, pero para ordenar

un bloque sólo se requiere el propio bloque, luego la reordenación requiere una transferencia de bloque.

Ahora que la selección de R está ordenada por c , ya podemos mezclar la selección de R con S , lo que requiere leer 1 bloque y 7 bloques más, y el resultado de la reunión tendrá:

$$N(\text{JOIN}) = N(\text{SELECCIÓN}) \cdot N(S) / \max\{V(R,c), V(S,b)\} = 2 \cdot 200 / 5 = 80 \text{ registros}$$

$$\text{Nbloques}(\text{JOIN}) = \text{redondeo hacia arriba} (N(\text{JOIN}) / \text{Bfr}(\text{JOIN})) = \text{redondeo hacia arriba} (80 / 14) = \text{redondeo hacia arriba} (5,714285714) = 6 \text{ bloques}$$

Sobre el resultado de la reunión, es necesario aplicar la proyección:

$$\text{Nbloques}(\text{PROYECCIÓN}) = \text{redondeo hacia arriba} (N(\text{JOIN}) / \text{Bfr}(\text{PROYECCIÓN})) = \text{redondeo hacia arriba} (80 / 33) = \text{redondeo hacia arriba} (2,424242424) = 3 \text{ bloques}$$

Teniendo en cuenta que el resultado de cada operación intermedia tiene que ser escrito a disco y leído después, podemos resumir que el número de bloques leídos y escritos en este plan físico es de:

- para aplicar la selección sobre R , necesitamos leer 250 bloques y escribir 1 bloque
- para reordenar el resultado de la selección necesitamos leer 1 bloque y escribir 1 bloque
- para aplicar la reunión natural, necesitamos leer $7 + 1$ bloques y escribir 6 bloques
- para aplicar la proyección, necesitamos leer 6 bloques y escribir 3 bloques

En total, para resolver este plan es necesario transferir 276 bloques

Como conclusión, entendemos que el plan físico seleccionado como mejor es el **plan (b)** puesto que requiere transferir 276 bloques, contra los 30834 bloques que requiere el plan (a).

Ejercicio 2

- a) Dependerá del tiempo que se tarda en aplicar la búsqueda binaria sobre el índice más el tiempo que se tarda en recorrer la cadena de desbordamiento si no se encuentra en el fichero maestro.
- b) El tiempo que se tarda en calcular la posición que le corresponde más el tiempo de escribir el registro o, si hay colisión, de escribir el registro en la zona de desbordamiento y modificar el registro que produce la colisión para apuntar a este nuevo.
- c) Las vistas de catálogo USER_ contienen información sobre todos los objetos creados por el usuario que la consulta, las vistas de catálogo ALL_ contienen información sobre todos los objetos a los que el usuario tiene acceso y las vistas de catálogo DBA_ contienen información sobre todos los objetos del sistema.

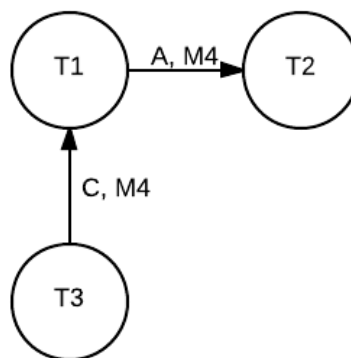
Ejercicio 3

T1	T2	T _i	Estado	Operación	Dato	V antiguo	V nuevo
Lee (A, x)		1	activa				
x := x + 1							
Escribe (A, x)	Lee (A, y)	2	activa				
	Lee (B, z)	1		update	A	2	3
	t := y + z	1	commit				
	Escribe (B, t)	2		update	B	3	5
	Escribe (C, t)	2		insert	C		5
		2	commit				

- b) Cuando se produce el fallo, la transacción 1 hace tiempo que terminó por lo que el sistema de recuperación ejecutaría una orden *REDO* sobre la transacción 1. Sin embargo, con la transacción 2 pueden ocurrir dos cosas:
1. si el fallo se produce después de haber escrito el registro *COMMIT*, el sistema de recuperación ejecutaría una orden *REDO* sobre la transacción 2, o
 2. si el fallo se produce antes de haber escrito el registro *COMMIT*, el sistema de recuperación ejecutaría una orden *UNDO* sobre la transacción 2.

Ejercicio 4

a)



- b) Las operaciones que se ven afectadas según el algoritmo de operación parcial y en el orden en el que intentan ejecutarse serían:
- Lee (T1, A), Lee (T2, A), Lee (T3, C), Lee (T3, A), Escribe (T2, A), Escribe (T1, A), Escribe (T3, C), Lee (T1, C), Escribe (T1, C)

Cuando se ejecuta *Lee (T2, A)*, no ha escrito ninguna transacción, luego se puede ejecutar sin problemas. *Lee (T3, C)* se puede ejecutar sin problemas, dado que es la primera lectura sobre C. *Lee (T3, A)* no ha escrito ninguna transacción, luego se puede ejecutar sin problemas. La operación *Escribe (T2, A)* falla puesto que la última lectura ha sido realizada por una transacción más joven (T3) y es necesario abortar la transacción T2.

Lee (T1, A), Lee (T3, C), Lee (T3, A), Escribe (T1, A), Escribe (T3,C), Lee (T1, C), Escribe (T1,C), Lee (T4, A), Escribe (T4, A)

La operación *Escribe (T1, A)* falla puesto que la última lectura ha sido realizada por una transacción más joven (T3) y es necesario abortar la transacción T1.

Lee (T3, C), Lee (T3, A), Escribe (T3,C), Lee (T4, A), Escribe (T4, A), Lee (T5, A), Escribe (T5, A), Lee (T5, C), Escribe (T5,C)

Escribe (T3, C) funciona puesto que el último que ha leído es T3 y nadie ha escrito antes.

Lee (T4, A) funciona puesto que el último que ha escrito es T3.

Escribe (T4, A) funciona puesto que el último que ha leído es T4 y el último que ha escrito es T3.

Lee (T5, A) funciona puesto que el último que ha escrito es T4.

Escribe (T5, A) funciona puesto que el último que ha leído es T5 y el último que escribió es T4.

Lee (T5, C) funciona puesto que el último que escribió fue T3.

Escribe (T5, C) funciona puesto que el último que leyó fue T5 y el último que escribió fue T3.

En resumen, las transacciones se ejecutan en el siguiente orden: **T3, T2, T1**

- c) Tomando en consideración el momento de los bloqueos y los desbloqueos, las transacciones deberían reescribirse como se ve en la tabla a continuación.

Cuando T2 intenta bloquear A en el modo M4, T1 ya lo ha hecho, y el modo es incompatible consigo mismo, de modo que T2 tendrá que esperar a que T1 termine con A.

Cuando T3 llega a bloquear C en modo M4, es la primera, por lo que consigue bloquearlo. Después de eso, intenta bloquear A en modo M1, que es compatible con el modo M4 (en el que T1 ha bloqueado, y por el que T2 está esperando en la lista), de modo que lo bloquea si sigue adelante. Hasta que completa todas sus operaciones, incluidos los desbloqueos de C y A.

Cuando T1 necesita bloquear C, T3 ya ha desbloqueado, por lo que puede bloquear sin problemas. Y continúa su ejecución hasta el desbloqueo de A y C, momento en el cual, T2 puede reanudarse, bloquear A y realizar su ejecución.

T1	T2	T3
<i>Lock (A, M4)</i>		
Lee (A, x)		
$x := x + 3$		
	<i>Lock (A, M4)</i>	<i>Lock (C, M4)</i>
	Lee (A, t)	Lee (C, u)
		<i>Lock (A, M1)</i>
	$t := t * 2$	Lee (A, v)
	Escribe (A, t)	$w := u + v$
	<i>Unlock (A)</i>	
Escribe (A, x)		Escribe (C, w)
		<i>Unlock (C)</i>
		<i>Unlock (A)</i>
<i>Lock (C, M4)</i>		
Lee (C, z)		
$z := z + x$		
Escribe (C, z)		
<i>Unlock (A)</i>		
<i>Unlock (C)</i>		

De este modo, el orden de ejecución (por el momento en que terminan) de las transacciones será **T3, T1, T2**.

Relación de problemas resuelta 8

1. (3 puntos) Sean las relaciones R y S con los siguientes parámetros:

R(a,b,c)	S(c,d)
N(R) = 5000	N(S)=200
V(R,a)=5000	
V(R,b)=300	
V(R,c)=5	V(S,c)=5
	V(S,d)=40
Size(a)=20	
Size(b)=60	
Size(c)=20	Size(c)=20
	Size(d)=40

Propón el plan lógico más eficiente para la siguiente consulta y justifica numéricamente por qué es más eficiente en base al número de bloques que emplea cada uno y teniendo en cuenta que el tamaño de bloque es de 2KB, que la cabecera es de 20B y que en memoria sólo cabe un bloque:

$$\Pi_{b,d}(\sigma_{(c=c_3 \wedge d < d_{20})}(R \times S))$$

Estimación del número de bloques para el plan de la consulta del ejercicio

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.

$$\begin{array}{c}
 \Pi_{b,d} \\
 | \\
 \sigma_{c=c_3} \\
 | \\
 \sigma_{d < d_{20}} \\
 | \\
 \times \\
 / \quad \backslash \\
 R \quad \quad S
 \end{array}$$

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

$$B(R) = \text{hacia arriba} \left(\frac{N(R)}{Bfr(R)} \right) = \text{hacia arriba} \left(\frac{5000}{Bfr(R)} \right)$$

$$Bfr(R) = \text{parte entera} \left(\frac{B-C}{L(R)} \right) = \text{parte entera} \left(\frac{2048-20}{L(R)} \right)$$

$$L(R) = 20 + 60 + 20 = 100$$

de modo que:

$$Bfr(R) = \text{parte entera} \left(\frac{2048-20}{100} \right) = 20 \quad \text{y}$$

$$B(R) = \text{hacia arriba} \left(\frac{5000}{20} \right) = 250$$

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

$$B(S) = \text{hacia arriba} \left(\frac{N(S)}{Bfr(S)} \right) = \text{hacia arriba} \left(\frac{200}{Bfr(S)} \right)$$

$$Bfr(S) = \text{parte entera} \left(\frac{B-C}{L(S)} \right) = \text{parte entera} \left(\frac{2048-20}{L(S)} \right)$$

$$L(S) = 20 + 40 = 60$$

de modo que:

$$Bfr(S) = \text{parte entera} \left(\frac{2048 - 20}{60} \right) = 33 \quad \text{y}$$

$$B(S) = \text{hacia arriba} \left(\frac{200}{33} \right) = 7$$

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

$$\text{Lectura: } B(R) \cdot B(S) = 250 \cdot 7 = 1750$$

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

$$B(\times) = \text{hacia arriba} \left(\frac{N(\times)}{Bfr(\times)} \right) = \text{hacia arriba} \left(\frac{5000 \cdot 200}{Bfr(\times)} \right)$$

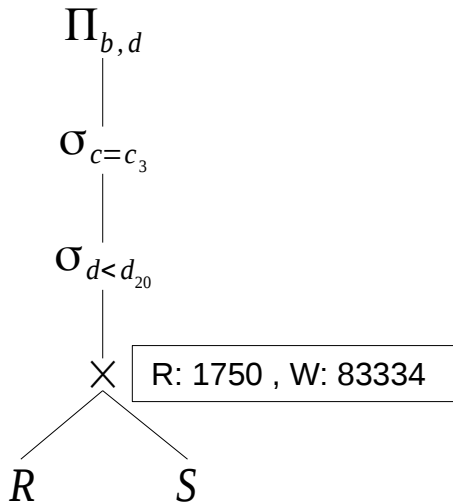
$$Bfr(\times) = \text{parte entera} \left(\frac{B-C}{L(\times)} \right) = \text{parte entera} \left(\frac{2048 - 20}{L(\times)} \right)$$

$$L(\times) = L(R) + L(S) = 100 + 60 = 160$$

de modo que:

$$Bfr(\times) = \text{parte entera} \left(\frac{2048 - 20}{L(\times)} \right) = \text{parte entera} \left(\frac{2048 - 20}{160} \right) = 12 \quad \text{y}$$

$$\text{Escritura: } B(\times) = \text{hacia arriba} \left(\frac{5000 \cdot 200}{Bfr(\times)} \right) = \text{hacia arriba} \left(\frac{5000 \cdot 200}{12} \right) = 83334$$

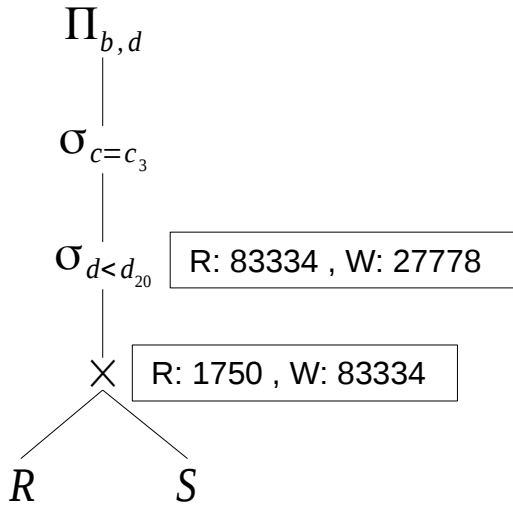


La primera operación de selección debe leer los bloques resultantes de la operación anterior y escribir tantos bloques como sean necesarios para almacenar los registros que resultan de seleccionar aquellos que cumplan la condición, es decir:

$$\begin{aligned}
 B(\sigma_{d < d_{20}}) &= \text{hacia arriba} \left(\frac{N(\sigma_{d < d_{20}})}{Bfr(\sigma_{d < d_{20}})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{d < d_{20}})}{Bfr(\times)} \right) = \\
 &= \text{hacia arriba} \left(\frac{N(\sigma_{d < d_{20}})}{12} \right) \\
 N(\sigma_{d < d_{20}}) &= \frac{N(\times)}{3} = \frac{5000 \cdot 200}{3} = 333334
 \end{aligned}$$

de modo que:

Escritura: $B(\sigma_{d < d_{20}}) = \text{hacia arriba} \left(\frac{N(\sigma_{d < d_{20}})}{12} \right) = \text{hacia arriba} \left(\frac{333334}{12} \right) = 27778$



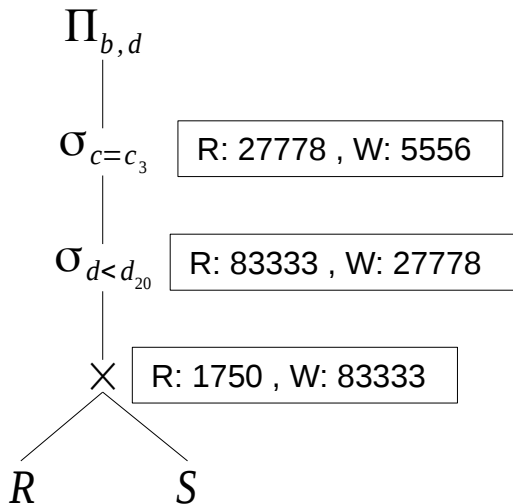
La segunda operación de selección debe leer los bloques resultantes de la operación anterior y escribir tantos bloques como sean necesarios para almacenar los registros que resultan de seleccionar aquellos que cumplan la condición, es decir:

$$\begin{aligned}
 B(\sigma_{c=c_3}) &= \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{Bfr(\sigma_{c=c_3})} \right) = \\
 &= \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{Bfr(\times)} \right) = \\
 &= \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{12} \right)
 \end{aligned}$$

$$N(\sigma_{c=c_3}) = \frac{N(\sigma_{d < d_{20}})}{V(R, c)} = \frac{333334}{5} \approx 66667$$

De este modo, se escriben:

$$B(\sigma_{c=c_3}) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{12} \right) = \text{hacia arriba} \left(\frac{66667}{12} \right) = 5556$$



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:

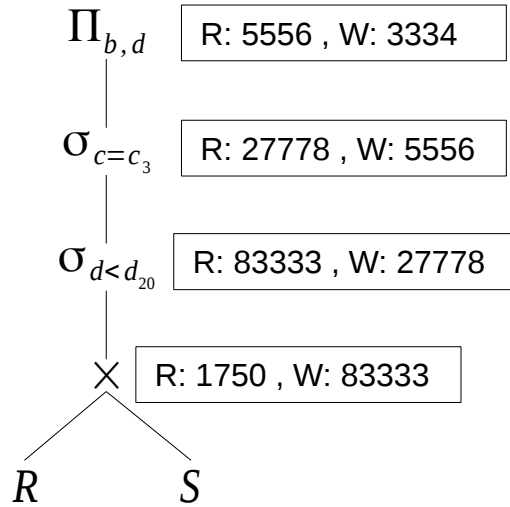
$$\begin{aligned}
 B(\Pi_{b,d}) &= \text{hacia arriba} \left(\frac{N(\Pi_{b,d})}{Bfr(\Pi_{b,d})} \right) = \\
 &= \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{Bfr(\Pi_{b,d})} \right) = \text{hacia arriba} \left(\frac{66667}{Bfr(\Pi_{b,d})} \right)
 \end{aligned}$$

$$L(\Pi_{b,d}) = 60 + 40 = 100$$

de modo que:

$$Bfr(\Pi_{b,d}) = \text{parte entera} \left(\frac{2048-20}{L(\Pi_{b,d})} \right) = \text{parte entera} \left(\frac{2048-20}{100} \right) = 20 \quad y$$

$$\text{Escritura: } B(\Pi_{b,d}) = \text{hacia arriba} \left(\frac{66667}{Bfr(\Pi_{b,d})} \right) = \text{hacia arriba} \left(\frac{66667}{20} \right) = 3334$$

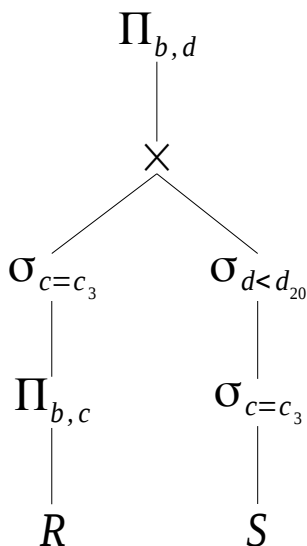


Una vez estimados el número de bloques que cada operación lee y escribe, podemos estimar que el número total de bloques leídos y escritos para la resolución del plan físico correspondiente es de:

$$(1750+83334)+(83334+27778)+(27778+5556)+(5556+3334)=238420 \text{ bloques}$$

Estimación del número de bloques para el plan optimizado para la consulta del ejercicio

El plan lógico propuesto para la optimización del plan lógico planteado por la consulta sería el que se ve a continuación:



Para calcular el número de bloques estimados para resolver el plan físico, reutilizaremos los cálculos realizados en la parte anterior cuando sea posible.

$$B(\Pi_{b,c}) = \text{hacia arriba} \left(\frac{N(R)}{Bfr(\Pi_{b,c})} \right) = \text{hacia arriba} \left(\frac{5000}{Bfr(\Pi_{b,c})} \right)$$

$$Bfr(\Pi_{b,c}) = \text{parte entera} \left(\frac{B-C}{L(\Pi_{b,c})} \right) = \text{parte entera} \left(\frac{2048-20}{L(\Pi_{b,c})} \right)$$

$$L(\Pi_{b,c}) = 60 + 20 = 80$$

de donde podemos calcular que:

$$Bfr(\Pi_{b,c}) = \text{parte entera} \left(\frac{2048-20}{80} \right) = 25$$

$$B(\Pi_{b,c}) = \text{hacia arriba} \left(\frac{5000}{25} \right) = 200$$

Después, procedemos para calcular el número de bloques necesarios para escribir la selección de la rama izquierda:

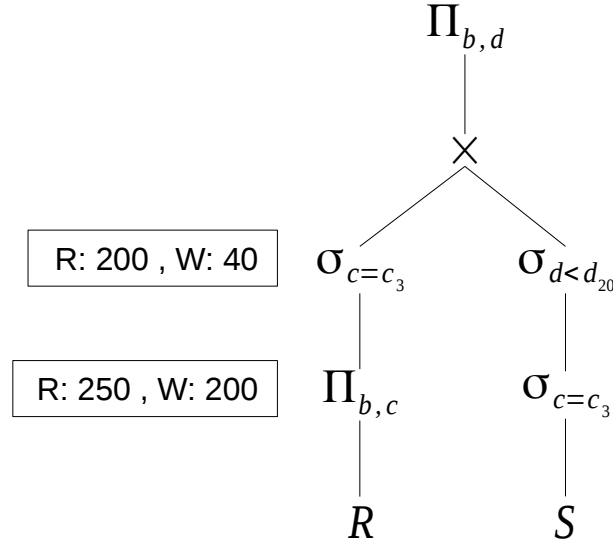
$$B(\sigma_{c=c_3}) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{Bfr(\sigma_{c=c_3})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{Bfr(\Pi_{b,c})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3})}{25} \right)$$

$$N(\sigma_{c=c_3}) = \frac{N(\Pi_{b,c})}{V(R,c)} = \frac{N(R)}{V(R,c)} = \frac{5000}{5} = 1000$$

de donde podemos calcular que:

$$B(\sigma_{c=c_3}) = \text{hacia arriba} \left(\frac{1000}{25} \right) = 40$$

Con esto, podemos concluir el número de bloques necesario para calcular la relación izquierda del producto cartesiano:



Procedemos a realizar los cálculos para la estimación de la relación derecha del producto cartesiano:

$$B(\sigma_{c=c_3}^D) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}^D)}{Bfr(\sigma_{c=c_3}^D)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}^D)}{Bfr(S)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}^D)}{33} \right)$$

$$N(\sigma_{c=c_3}^D) = \frac{N(S)}{V(S,c)} = \frac{200}{5} = 40$$

de donde podemos calcular que:

$$B(\sigma_{c=c_3}^D) = \text{hacia arriba} \left(\frac{40}{33} \right) = 2$$

A continuación, podemos aplicar la segunda selección y estimar los bloques necesarios:

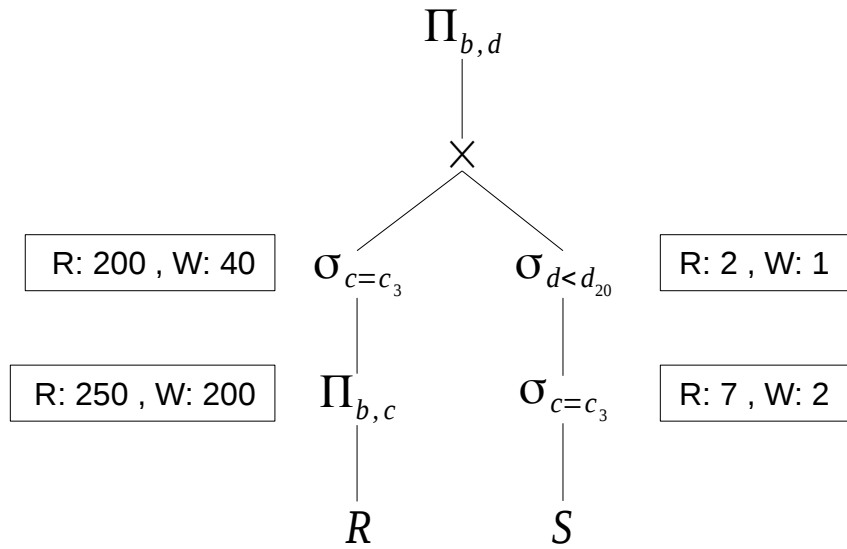
$$B(\sigma_{d < d_{20}}) = \text{hacia arriba} \left(\frac{N(\sigma_{d < d_{20}})}{Bfr(\sigma_{d < d_{20}})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{d < d_{20}})}{Bfr(\sigma_{c=c_3}^D)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{d < d_{20}})}{Bfr(S)} \right)$$

$$N(\sigma_{d < d_{20}}) = \frac{N(\sigma_{c=c_3}^D)}{3} = \frac{40}{3} \approx 14$$

de donde podemos calcular que:

$$B(\sigma_{d < d_{20}}) = \text{hacia arriba} \left(\frac{14}{33} \right) = 1$$

Con esto, tenemos el cálculo de los bloques necesarios para calcular la relación derecha del producto cartesiano:



Ya podemos calcular el número de bloques necesarios para el cálculo del producto cartesiano. En concreto, habrá que multiplicar los registros de cada bloque de la relación izquierda por los de cada bloque de la relación derecha, lo cual requerirá leer el bloque de la relación derecha por cada bloque de la relación izquierda. En total, habrá que hacer:

$$40 \times 1 = 40 \text{ lecturas}$$

y escribir los bloques necesarios para almacenar el resultado:

$$B(\times) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}) \cdot N(\sigma_{d<d_{20}})}{Bfr(\times)} \right) = \text{hacia arriba} \left(\frac{1000 \cdot 14}{Bfr(\times)} \right)$$

$$Bfr(\times) = \text{parte entera} \left(\frac{B-C}{L(\times)} \right) = \text{parte entera} \left(\frac{2048-20}{L(\times)} \right)$$

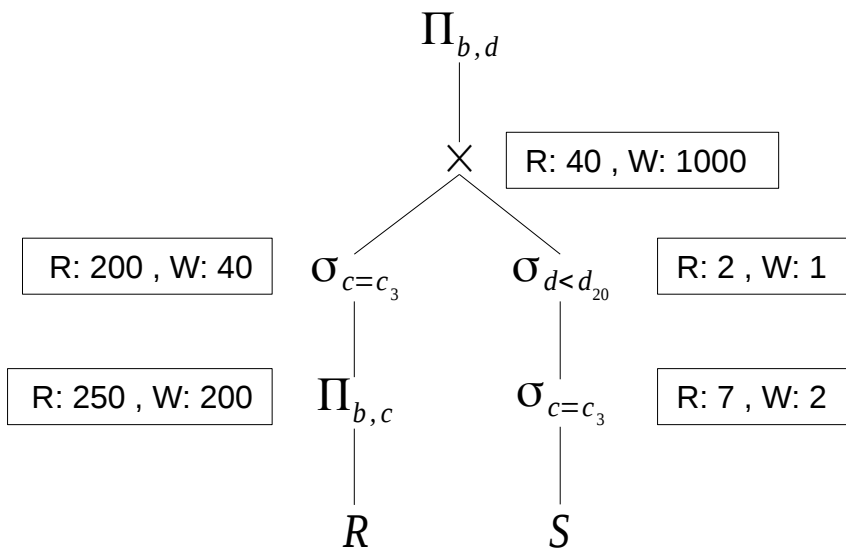
$$L(\times) = L(\sigma_{c=c_3}) + L(\sigma_{d<d_{20}}) = L(\Pi_{b,c}) + L(S) = 80 + 60 = 140$$

de lo que podemos calcular que:

$$Bfr(\times) = \text{parte entera} \left(\frac{2048-20}{140} \right) = 14$$

$$B(\times) = \text{hacia arriba} \left(\frac{1000 \cdot 14}{14} \right) = 1000$$

Con esto, sabemos el número de bloques necesarios para la lectura y escritura del producto cartesiano:



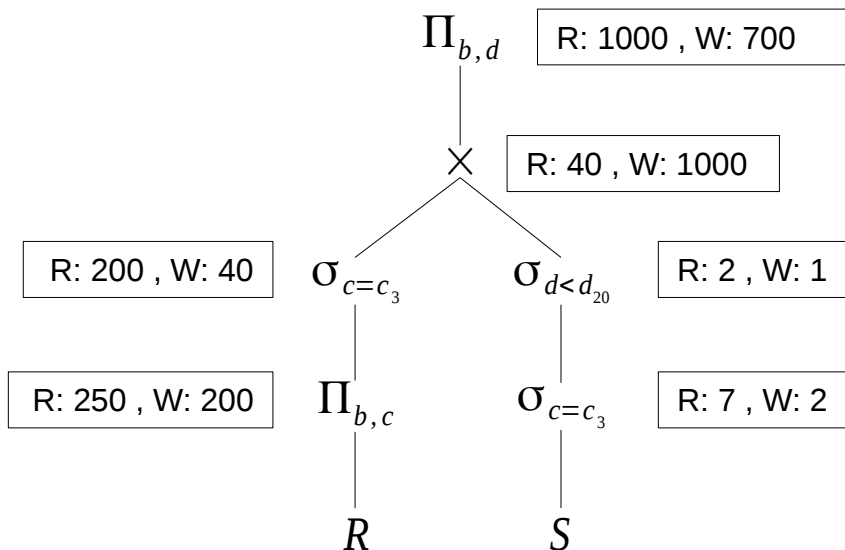
Únicamente nos resta el cálculo de los bloques necesarios para la proyección:

$$B(\Pi_{b,d}) = \text{hacia arriba} \left(\frac{N(\Pi_{b,d})}{Bfr(\Pi_{b,d})} \right) = \text{hacia arriba} \left(\frac{N(\times)}{Bfr(\Pi_{b,d})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}) \cdot N(\sigma_{d<d_{20}})}{Bfr(\Pi_{b,d})} \right)$$

de donde podemos calcular que:

$$B(\Pi_{b,d}) = \text{hacia arriba} \left(\frac{N(\sigma_{c=c_3}) \cdot N(\sigma_{d<d_{20}})}{Bfr(\Pi_{b,d})} \right) = \text{hacia arriba} \left(\frac{1000 \cdot 14}{20} \right) = 700$$

Con esto, completamos los cálculos necesarios:



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:

$$(200+250)+(200+40)+(7+2)+(2+1)+(40+1000)+(1000+700)=3442 \text{ bloques}$$

2. **(2 punto)** Estima qué estructura de índice haría falta para el atributo c de la tabla R del ejercicio 1 y calcula su tamaño en bloques si $P=10B$ y el resto de los parámetros son los mismos que los del ejercicio 1.

Dada la variabilidad del atributo c en la relación R (5), y que dicha relación contiene 5000 registros, no podemos más que asumir que hay varios registros de R con el mismo valor de c . Si asumimos una distribución uniforme de los valores de c en los 5000 registros, tendríamos 1000 registros para cada uno de los 5 valores distintos de c .

Para que un índice sea útil, tiene que contener valores distintos, pero sólo hay 5 valores distintos y 5000 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 c .

El primer índice necesita un número de bloques igual a:

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

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

$$Bfr(I_1) = \text{parte entera} \left(\frac{2048 - 20}{L(I_1)} \right)$$

$$L(I_1) = 20 + 10 = 30$$

de modo que:

$$Bfr(I_1) = \text{parte entera} \left(\frac{2048 - 20}{30} \right) = 67 \quad y$$

$$B(I_1) = \text{hacia arriba} \left(\frac{N(I_1)}{Bfr(I_1)} \right) = \text{hacia arriba} \left(\frac{5}{67} \right) = 1$$

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

$$\begin{aligned} B(I_2) &= v(B, t) \cdot \text{hacia arriba} \left(\frac{N(\text{igual valor de } t)}{Bfr(I_2)} \right) = 5 \cdot \text{hacia arriba} \left(\frac{N(\text{igual valor de } t)}{Bfr(I_1)} \right) = \\ &= 5 \cdot \text{hacia arriba} \left(\frac{N(\text{igual valor de } t)}{67} \right) \end{aligned}$$

$$N(\text{igual valor de } t) = \frac{N(R)}{V(R, c)} = \frac{5000}{5} = 1000$$

de modo que:

$$B(I_2) = 5 \cdot \text{hacia arriba} \left(\frac{N(\text{igual valor de } t)}{67} \right) = 5 \cdot \text{hacia arriba} \left(\frac{1000}{67} \right) = 75$$

En total, haría falta 1 bloque para el primer nivel de índice y 75 bloques para el segundo:

$$B(I) = B(I_1) + B(I_2) = 1 + 75 = 76 \text{ bloques}$$

3. **(1 punto)** 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)

Administración de Bases de Datos

Resolución del examen de contenidos teóricos – Junio de 2016

1. Sean las relaciones S y T con los siguientes parámetros:

S(<u>a</u> ,b,c)	T(<u>b</u> ,d,e)
N(S)=2000	N(T)=200
Size(a)=15	
Size(b)=20	Size(b)=20
Size(c)=50	
	Size(d)=20
	Size(e)=10
V(S,b)=25	V(T,b)=200

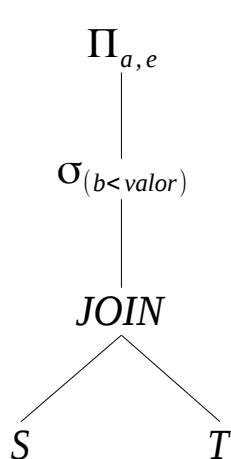
donde S.a y T.b son claves, y donde el atributo T.b es llave externa a S.b

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:

$$\Pi_{a,e}(\sigma_{(b < valor)}(S \text{ JOIN } T))$$

Y la solución propuesta es la siguiente:

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:

$$B(S) = \text{hacia arriba} \left(\frac{N(S)}{Bfr(S)} \right) = \text{hacia arriba} \left(\frac{2000}{Bfr(S)} \right)$$

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

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

de modo que:

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

$$B(S) = \text{hacia arriba} \left(\frac{2000}{47} \right) = 43$$

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

$$B(T) = \text{hacia arriba} \left(\frac{N(T)}{Bfr(T)} \right) = \text{hacia arriba} \left(\frac{200}{Bfr(T)} \right)$$

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

$$L(T) = 20 + 20 + 10 = 50$$

de modo que:

$$Bfr(T) = \text{parte entera} \left(\frac{4096-20}{50} \right) = 81 \text{ y}$$

$$B(T) = \text{hacia arriba} \left(\frac{200}{81} \right) = 3$$

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, b . Dado que b es clave en T , existe un índice asociado por lo que no es necesario ordenarlo pero, dado que no se indica la existencia de un índice para b en S será necesario ordenar los registros de dicha relación.

El coste de esta operación se estima en $43 \cdot \log_2(43) = 233,329384452 \approx 234$ operaciones de bloque

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: $B(S) + B(T) = 43 + 3 = 46$

y el resultado de la reunión natural necesita un número de bloques suficiente para almacenar los registros resultantes de esta operación, o sea:

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

$$N(JOIN) = \frac{N(S) \cdot N(T)}{\max\{V(S, b), V(T, b)\}} = \frac{2000 \cdot 200}{200} = 2000$$

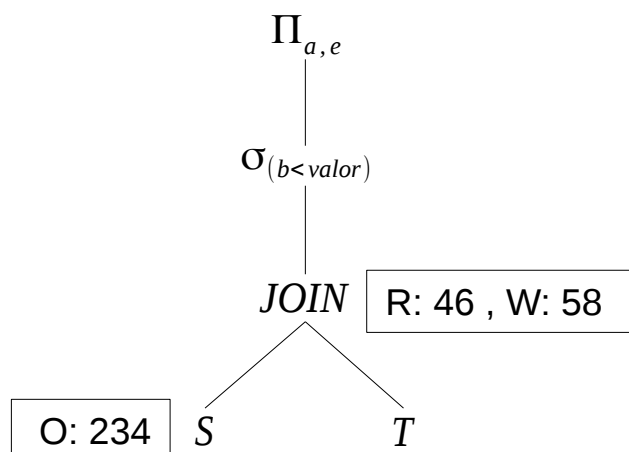
$$Bfr(JOIN) = \text{parte entera} \left(\frac{B-C}{L(JOIN)} \right) = \text{parte entera} \left(\frac{4096-20}{L(JOIN)} \right)$$

$$L(JOIN) = L(S) + L(T) - \text{Size}(b) = 85 + 50 - 20 = 115$$

de modo que:

$$Bfr(JOIN) = \text{parte entera} \left(\frac{4096-20}{115} \right) = \text{parte entera} \left(\frac{4096-20}{115} \right) = 35 \text{ y}$$

Escritura: $B(JOIN) = \text{hacia arriba} \left(\frac{N(JOIN)}{Bfr(JOIN)} \right) = \text{hacia arriba} \left(\frac{2000}{35} \right) = 58$



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:

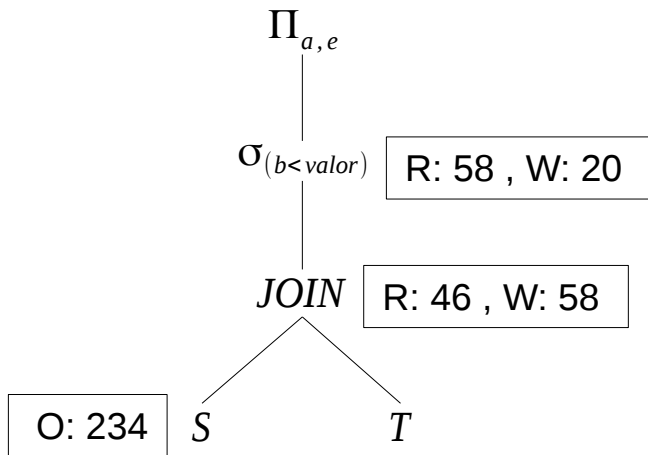
$$B(\sigma_{b < valor}) = \text{hacia arriba} \left(\frac{N(\sigma_{b < valor})}{Bfr(\sigma_{b < valor})} \right) =$$

$$= \text{hacia arriba} \left(\frac{N(\sigma_{b < \text{valor}})}{Bfr(JOIN)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{b < \text{valor}})}{35} \right)$$

$$N(\sigma_{b < \text{valor}}) = N(JOIN) \cdot \frac{1}{3} = \frac{2000}{3} \approx 667$$

de modo que:

Escritura: $B(\sigma_{b < \text{valor}}) = \text{hacia arriba} \left(\frac{N(\sigma_{b < \text{valor}})}{35} \right) = \text{hacia arriba} \left(\frac{667}{35} \right) = 20$



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:

$$\begin{aligned}
 B(\Pi_{a,e}) &= \text{hacia arriba} \left(\frac{N(\Pi_{a,e})}{Bfr(\Pi_{a,e})} \right) = \\
 &= \text{hacia arriba} \left(\frac{N(\sigma_{b < \text{valor}})}{Bfr(\Pi_{a,e})} \right) = \\
 &= \text{hacia arriba} \left(\frac{667}{Bfr(\Pi_{a,e})} \right)
 \end{aligned}$$

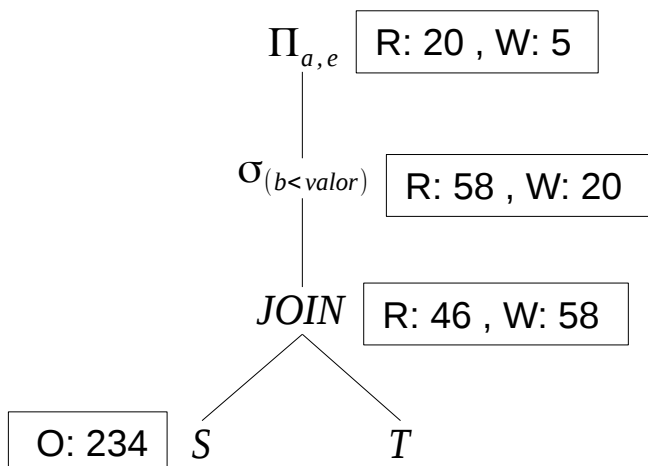
$$Bfr(\Pi_{a,e}) = \text{parte entera} \left(\frac{B-C}{L(\Pi_{a,e})} \right) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{a,e})} \right)$$

$$L(\Pi_{a,e}) = 15 + 10 = 25$$

de modo que:

$$Bfr(\Pi_{a,e}) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{a,e})} \right) = \text{parte entera} \left(\frac{4096-20}{25} \right) = 163 \quad y$$

Escritura: $B(\Pi_{a,e}) = \text{hacia arriba} \left(\frac{667}{Bfr(\Pi_{a,e})} \right) = \text{hacia arriba} \left(\frac{667}{163} \right) = 5$



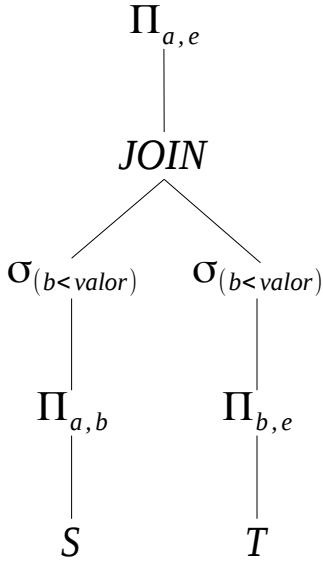
Una vez estimados el número de bloques que cada operación lee y escribe, podemos estimar que el número total de bloques leídos y escritos para la resolución del plan físico correspondiente es de:

$$234 + (43 + 3 + 58) + (58 + 20) + (20 + 5) = 441 \text{ bloques}$$

2. Plantea una versión más eficiente de la consulta del ejercicio 1 e ilustra numéricamente por qué la consideras más eficiente.

Y la solución propuesta es la siguiente:

la mejora de la eficiencia se conseguiría reduciendo el número de bloques implicados al realizar las proyecciones y selecciones lo antes posible, para eliminar el mayor número de atributos no implicados, así como reducir el número de tuplas involucradas en las selecciones.



El plan propuesto para la mejora es el que se muestra en el árbol de expresión, aunque queda por determinar si el incremento en el número de operaciones intermedias mejora realmente el número de bloques involucrados.

Reutilizaremos tantos cálculos realizados para el ejercicio 1 como sea posible.

Empezaremos por estimar el número de bloques necesarios para las operaciones de la rama izquierda del árbol. Para ello, comenzaremos con la proyección que se realiza sobre S para eliminar el atributo c, de tamaño mucho mayor que los demás.

$$B(\Pi_{a,b}) = \text{hacia arriba} \left(\frac{N(\Pi_{a,b})}{Bfr(\Pi_{a,b})} \right) = \text{hacia arriba} \left(\frac{N(S)}{Bfr(\Pi_{a,b})} \right) = \\ = \text{hacia arriba} \left(\frac{2000}{Bfr(\Pi_{a,b})} \right)$$

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

$$L(\Pi_{a,b}) = 15 + 20 = 35$$

de modo que:

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

Escritura: $B(\Pi_{a,b}) = \text{hacia arriba} \left(\frac{2000}{Bfr(\Pi_{a,b})} \right) = \text{hacia arriba} \left(\frac{2000}{116} \right) = 18$

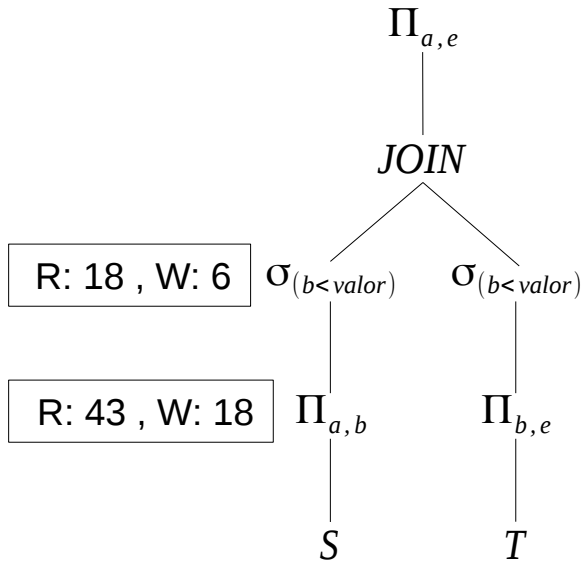
La operación de selección debe leer los bloques resultantes de la operación anterior y escribir tantos bloques como sean necesarios para almacenar los registros que resultan de seleccionar aquellos que cumplan la condición, es decir:

$$B(\sigma_{b<valor, izq}) = \text{hacia arriba} \left(\frac{N(\sigma_{b<valor, izq})}{Bfr(\sigma_{b<valor, izq})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{b<valor, izq})}{Bfr(\Pi_{a,b})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{b<valor, izq})}{116} \right)$$

$$N(\sigma_{b<valor, izq}) = N(\Pi_{a,b}) \cdot \frac{1}{3} = N(S) \cdot \frac{1}{3} = \frac{2000}{3} \approx 667$$

de modo que:

Escritura: $B(\sigma_{b<valor, izq}) = \text{hacia arriba} \left(\frac{N(\sigma_{b<valor, izq})}{116} \right) = \text{hacia arriba} \left(\frac{667}{116} \right) = 6$



Finalizada la estimación del número de bloques de la rama izquierda del árbol, podemos proceder con la rama derecha, de forma análoga.

Comenzaremos con la proyección que se realiza sobre S para eliminar el atributo c, de tamaño mucho mayor que los demás.

$$\begin{aligned}
 B(\Pi_{b,e}) &= \text{hacia arriba} \left(\frac{N(\Pi_{b,e})}{Bfr(\Pi_{b,e})} \right) = \\
 &= \text{hacia arriba} \left(\frac{N(T)}{Bfr(\Pi_{b,e})} \right) = \\
 &= \text{hacia arriba} \left(\frac{200}{Bfr(\Pi_{b,e})} \right)
 \end{aligned}$$

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

$$L(\Pi_{b,e}) = 20 + 10 = 30$$

de modo que:

$$Bfr(\Pi_{b,e}) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{b,e})} \right) = \text{parte entera} \left(\frac{4096-20}{30} \right) = 135 \quad y$$

$$\text{Escritura: } B(\Pi_{b,e}) = \text{hacia arriba} \left(\frac{200}{Bfr(\Pi_{b,e})} \right) = \text{hacia arriba} \left(\frac{200}{135} \right) = 2$$

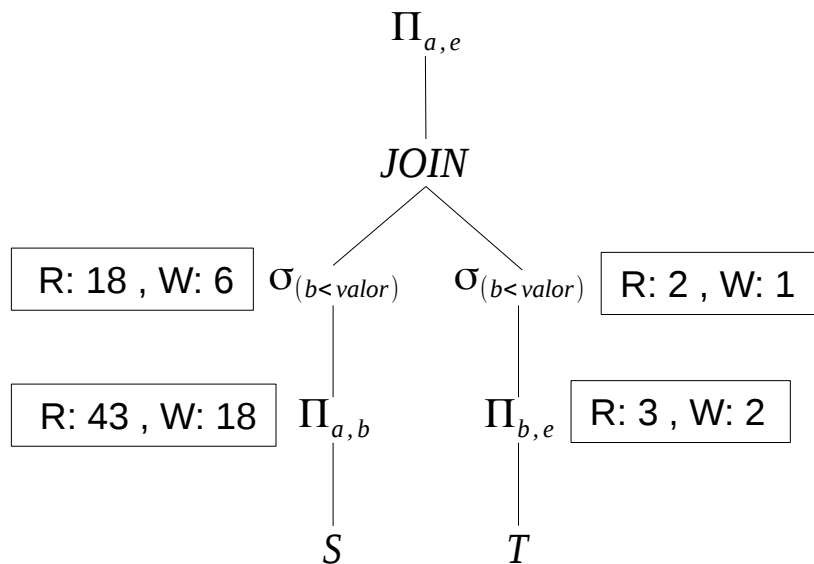
La operación de selección debe leer los bloques resultantes de la operación anterior y escribir tantos bloques como sean necesarios para almacenar los registros que resultan de seleccionar aquellos que cumplan la condición, es decir:

$$B(\sigma_{b<valor,der}) = \text{hacia arriba} \left(\frac{N(\sigma_{b<valor,der})}{Bfr(\sigma_{b<valor,der})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{b<valor,der})}{Bfr(\Pi_{b,e})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{b<valor,der})}{135} \right)$$

$$N(\sigma_{b<valor,der}) = N(\Pi_{b,e}) \cdot \frac{1}{3} = N(T) \cdot \frac{1}{3} = \frac{200}{3} \approx 67$$

de modo que:

$$\text{Escritura: } B(\sigma_{b<valor,der}) = \text{hacia arriba} \left(\frac{N(\sigma_{b<valor,der})}{135} \right) = \text{hacia arriba} \left(\frac{67}{135} \right) = 1$$



Estimado el número de bloques para realizar las operaciones de las ramas izquierda y derecha del árbol, podemos proceder al cálculo de bloques necesario para la operación de reunión natural.

Siendo b el atributo de reunión, y siendo b clave en T , no sería necesaria la ordenación de la relación T mediante ese atributo. Sin embargo, se realiza una operación de proyección sobre T , cuyo resultado no tiene por qué estar ordenado y, desde luego, no tiene un índice sobre b . De este modo, hará falta ordenar por el atributo b tanto la relación de la izquierda, para lo que se estima que harán falta las siguientes operaciones de entrada salida sobre bloques:

$$6 \cdot \log_2(6) = 15,51 \approx 16 \text{ operaciones de bloque}$$

como para la relación de la derecha, para lo que se estima que harán falta las siguientes 2 operaciones sobre bloques, una para leer el bloque, que se ordenará en memoria, y otra para escribir el bloque ordenado.

Dado que el número de tuplas de cada una de las relaciones a la izquierda y a la derecha de la reunión natural se han visto reducidas por la aplicación de la operación de selección, podemos asumir que las variabilidades se han visto reducidas en la misma proporción, es decir, un tercio de la variabilidad original de los atributos en las tablas de origen.

La operación de reunión natural requiere leer y mezclar todos los registros de las relaciones a la izquierda y a la derecha sin leer cada bloque más de una vez, por lo que será necesario:

$$\text{Lectura: } B(\sigma_{a,b}) + B(\sigma_{b,e}) = 6 + 1 = 7$$

y el resultado de la reunión natural necesita un número de bloques suficiente para almacenar los registros resultantes de esta operación, o sea:

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

$$N(\text{JOIN}) = \frac{N(\sigma_{b < valor, izq}) \cdot N(\sigma_{b < valor, der})}{\max\{V(\sigma_{b < valor, izq}, b), V(\sigma_{b < valor, der}, b)\}} = \frac{667 \cdot 67}{67} = 667$$

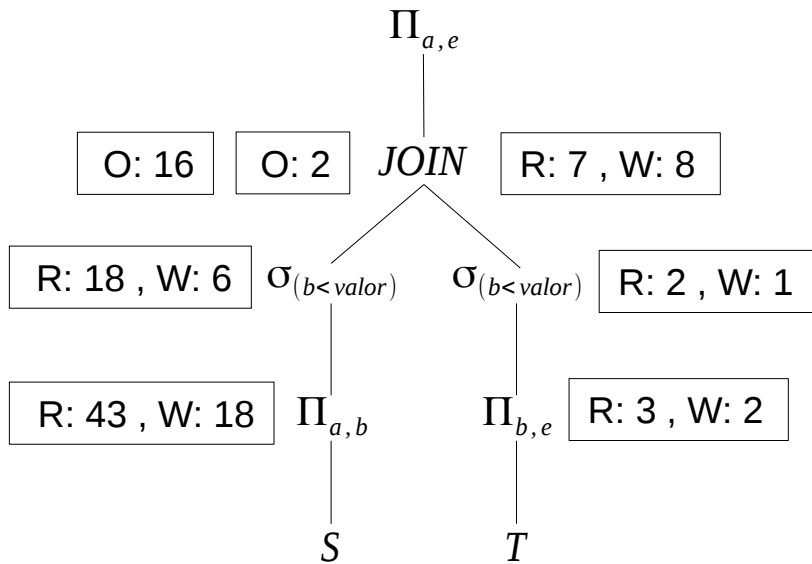
$$Bfr(\text{JOIN}) = \text{parte entera} \left(\frac{B - C}{L(\text{JOIN})} \right) = \text{parte entera} \left(\frac{4096 - 20}{L(\text{JOIN})} \right)$$

$$L(\text{JOIN}) = L(\sigma_{b < valor, izq}) + L(\sigma_{b < valor, der}) - \text{Size}(b) = L(\Pi_{a,b}) + L(\Pi_{b,e}) - \text{Size}(b) = 35 + 30 - 20 = 45$$

de modo que:

$$Bfr(\text{JOIN}) = \text{parte entera} \left(\frac{4096 - 20}{L(\text{JOIN})} \right) = \text{parte entera} \left(\frac{4096 - 20}{45} \right) = 90 \text{ y}$$

$$\text{Escritura: } B(\text{JOIN}) = \text{hacia arriba} \left(\frac{N(\text{JOIN})}{Bfr(\text{JOIN})} \right) = \text{hacia arriba} \left(\frac{667}{90} \right) = 8$$



Por último, calculamos el número de operaciones de bloques necesario para la proyección final.

$$\begin{aligned}
 B(\Pi_{a,e}) &= \text{hacia arriba} \left(\frac{N(\Pi_{a,e})}{Bfr(\Pi_{a,e})} \right) = \\
 &= \text{hacia arriba} \left(\frac{N(JOIN)}{Bfr(\Pi_{a,e})} \right) = \\
 &= \text{hacia arriba} \left(\frac{667}{Bfr(\Pi_{a,e})} \right)
 \end{aligned}$$

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

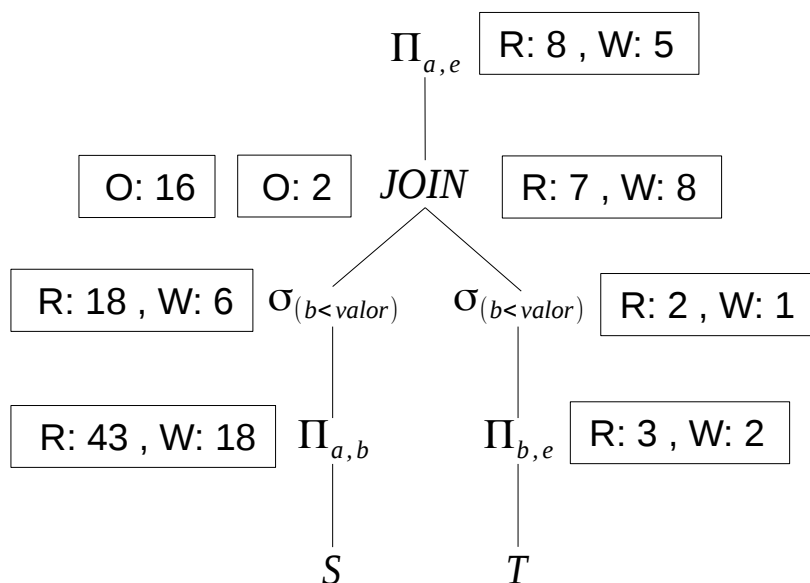
$$L(\Pi_{a,e}) = 15 + 10 = 25$$

de modo que:

$$Bfr(\Pi_{a,e}) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{a,e})} \right) = \text{parte entera} \left(\frac{4096-20}{25} \right) = 163 \text{ y}$$

$$\text{Escritura: } B(\Pi_{a,e}) = \text{hacia arriba} \left(\frac{667}{Bfr(\Pi_{a,e})} \right) = \text{hacia arriba} \left(\frac{667}{163} \right) = 5$$

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:



resultado que mejora, sustancialmente, el número de bloques necesario para el ejercicio 1.

$$(43+18)+(18+6)+(3+2)+(2+1)+16+2+(7+8)+(8+5)=139 \text{ bloques}$$

3. Enuncia qué condiciones deben cumplirse para que una vista sea actualizable.

Y la respuesta es la siguiente:

- No pueden incluir agrupadores o agregaciones
- No puede incluir la cláusula DISTINCT
- No puede incluir la reunión ni operadores de conjuntos
- Todos los atributos con restricción NOT NULL (incluida la clave) deben estar en la vista



ugr

Universidad de Granada
Departamento de Ciencias de la Computación
e Inteligencia Artificial



Grado en Ingeniería Informática
Examen de contenidos teóricos - Convocatoria de Junio de 2017
Resolución

Pregunta 1

Enunciado

Sean las relaciones S y T con los siguientes parámetros:

$S(\underline{a}, b, c)$	$T(\underline{b}, d, e)$
$N(S)=2000$	$N(T)=200$
$Size(a)=15$	
$Size(b)=20$	$Size(b)=20$
$Size(c)=50$	
	$Size(d)=20$
	$Size(e)=10$
$V(S, b)=25$	$V(T, b)=200$

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

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

$$\Pi_{a, e}(\sigma_{(b \neq valor)}(S JOIN T))$$

que supondría la ejecución de la consulta si el plan lógico fuera tal y como plantea la consulta.



ugr

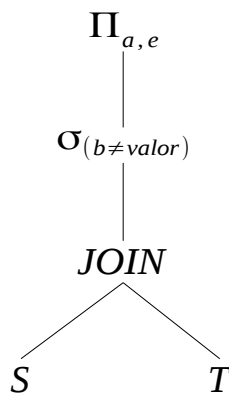
Universidad de Granada

Departamento de Ciencias de la Computación
e Inteligencia Artificial



Solución

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:

$$B(S) = \text{hacia arriba} \left(\frac{N(S)}{Bfr(S)} \right) = \text{hacia arriba} \left(\frac{2000}{Bfr(S)} \right)$$

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

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

de modo que:

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

$$B(S) = \text{hacia arriba} \left(\frac{2000}{47} \right) = 43$$

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

$$B(T) = \text{hacia arriba} \left(\frac{N(T)}{Bfr(T)} \right) = \text{hacia arriba} \left(\frac{200}{Bfr(T)} \right)$$

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

$$L(T) = 20 + 20 + 10 = 50$$

de modo que:

$$Bfr(T) = \text{parte entera} \left(\frac{4096-20}{50} \right) = 81 \quad y$$

$$B(T) = \text{hacia arriba} \left(\frac{200}{81} \right) = 3$$

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, b . Dado que b es clave en T, existe un índice asociado por lo que no es necesario ordenarlo pero, dado que no se indica la existencia de un índice para b en S será necesario ordenar los registros de dicha relación.

El coste de esta operación se estima en $43 \cdot \log_2(43) = 233,329384452 \approx 234$ operaciones de bloque

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

Lectura: $B(S) + B(T) = 43 + 3 = 46$

y el resultado de la reunión natural necesita un número de bloques suficiente para almacenar los registros resultantes de esta operación, o sea:



ugr

Universidad de Granada

Departamento de Ciencias de la Computación
e Inteligencia Artificial



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

$$N(JOIN) = \frac{N(S) \cdot N(T)}{\max\{V(S, b), V(T, b)\}} = \frac{2000 \cdot 200}{200} = 2000$$

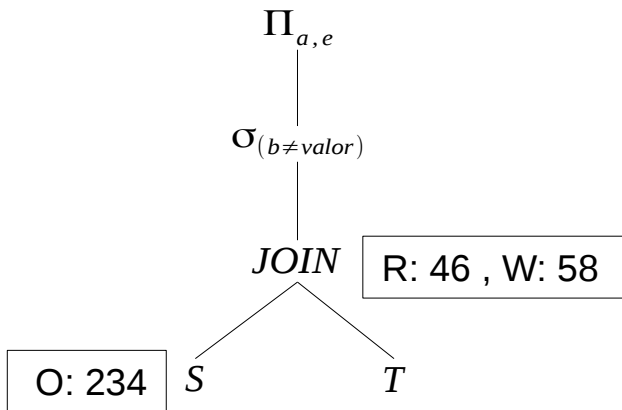
$$Bfr(JOIN) = \text{parte entera} \left(\frac{B - C}{L(JOIN)} \right) = \text{parte entera} \left(\frac{4096 - 20}{L(JOIN)} \right)$$

$$L(JOIN) = L(S) + L(T) - \text{Size}(b) = 85 + 50 - 20 = 115$$

de modo que:

$$Bfr(JOIN) = \text{parte entera} \left(\frac{4096 - 20}{L(JOIN)} \right) = \text{parte entera} \left(\frac{4096 - 20}{115} \right) = 35 \text{ y}$$

Escritura: $B(JOIN) = \text{hacia arriba} \left(\frac{N(JOIN)}{Bfr(JOIN)} \right) = \text{hacia arriba} \left(\frac{2000}{35} \right) = 58$



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:

$$B(\sigma_{b \neq \text{valor}}) = \text{hacia arriba} \left(\frac{N(\sigma_{b \neq \text{valor}})}{Bfr(\sigma_{b \neq \text{valor}})} \right) =$$

$$= \text{hacia arriba} \left(\frac{N(\sigma_{b \neq \text{valor}})}{Bfr(JOIN)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{b \neq \text{valor}})}{35} \right)$$

$$N(\sigma_{b \neq \text{valor}}) = N(JOIN) \cdot 1 = 2000 \cdot 1 = 2000$$

de modo que:

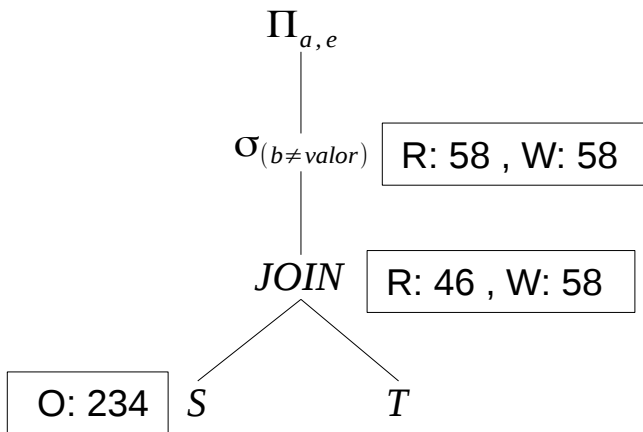
Escritura: $B(\sigma_{b \neq \text{valor}}) = \text{hacia arriba} \left(\frac{N(\sigma_{b \neq \text{valor}})}{35} \right) = \text{hacia arriba} \left(\frac{2000}{35} \right) = 58$



ugr

Universidad de Granada

Departamento de Ciencias de la Computación
e Inteligencia Artificial



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:

$$\begin{aligned}
 B(\Pi_{a,e}) &= \text{hacia arriba} \left(\frac{N(\Pi_{a,e})}{Bfr(\Pi_{a,e})} \right) = \\
 &= \text{hacia arriba} \left(\frac{N(\sigma_{b \neq \text{valor}})}{Bfr(\Pi_{a,e})} \right) = \\
 &= \text{hacia arriba} \left(\frac{2000}{Bfr(\Pi_{a,e})} \right)
 \end{aligned}$$

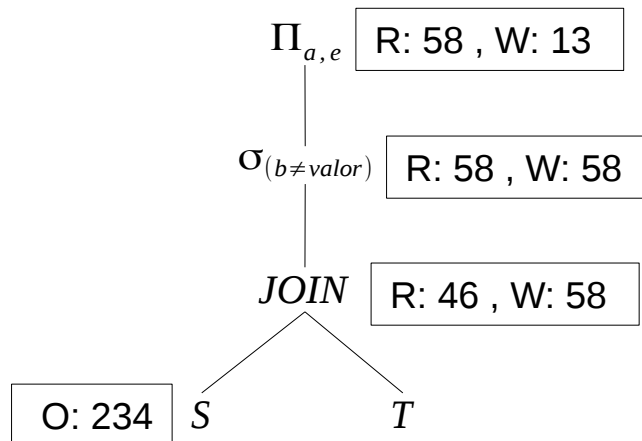
$$Bfr(\Pi_{a,e}) = \text{parte entera} \left(\frac{B-C}{L(\Pi_{a,e})} \right) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{a,e})} \right)$$

$$L(\Pi_{a,e}) = 15 + 10 = 25$$

de modo que:

$$Bfr(\Pi_{a,e}) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{a,e})} \right) = \text{parte entera} \left(\frac{4096-20}{25} \right) = 163 \quad \text{y}$$

$$\text{Escritura: } B(\Pi_{a,e}) = \text{hacia arriba} \left(\frac{2000}{Bfr(\Pi_{a,e})} \right) = \text{hacia arriba} \left(\frac{2000}{163} \right) = 13$$



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:

$$234 + (43 + 3 + 58) + (58 + 58) + (58 + 13) = 525 \text{ bloques}$$



ugr

Universidad de Granada

Departamento de Ciencias de la Computación
e Inteligencia Artificial



Pregunta 2

Enunciado

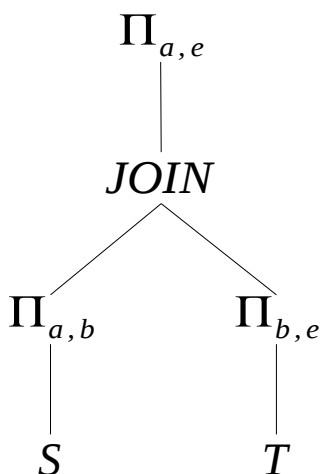
Plantea una versión más eficiente de la consulta del ejercicio 1 e ilustra numéricamente por qué la consideras más eficiente.

Solución propuesta

La mejora de la eficiencia se conseguiría reduciendo el número de bloques implicados al realizar las proyecciones y selecciones lo antes posible, para eliminar el mayor número de atributos no implicados, así como reducir el número de tuplas involucradas en las selecciones.

Dado que el número estimado de tuplas que resultan de la selección de aquellas tuplas con valores distintos a uno dado es exactamente igual al de las tuplas de la relación que se selecciona, dicha estimación no tendría efecto ni se consideraría para optimización. Por ello, al estimar el número de bloques del plan propuesto, es igual hacer la selección como no hacerla.

Puesto que el objetivo es encontrar un plan mejor, la selección se realizaría en primer lugar (junto a las relaciones S y T), en todos los planes. Por ello, podemos asumir que no realizaremos la selección en ninguno de los planes.



El plan propuesto para la mejora es el que se muestra en el árbol de expresión, aunque queda por determinar si el incremento en el número de operaciones intermedias mejora realmente el número de bloques involucrados.

Reutilizaremos tantos cálculos realizados para el ejercicio 1 como sea posible.

Empezaremos por estimar el número de bloques necesarios para las operaciones de la rama izquierda del árbol. Para ello, comenzaremos con la proyección que se realiza sobre S para eliminar el atributo c, de tamaño mucho mayor que los demás.

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

$$= \text{hacia arriba} \left(\frac{2000}{Bfr(\Pi_{a,b})} \right)$$

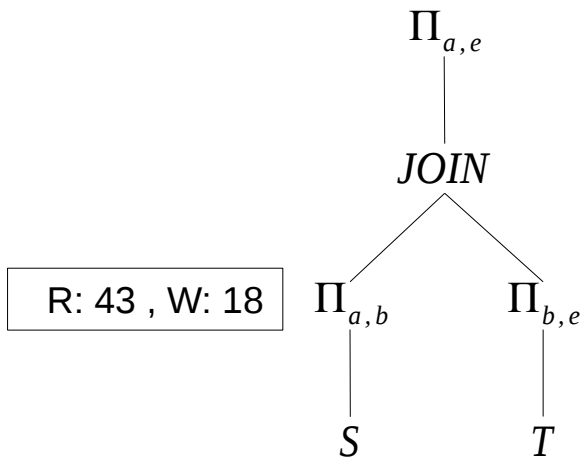
$$Bfr(\Pi_{a,e}) = \text{parte entera} \left(\frac{B-C}{L(\Pi_{a,b})} \right) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{a,b})} \right)$$

$$L(\Pi_{a,b}) = 15 + 20 = 35$$

de modo que:

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

$$\text{Escritura: } B(\Pi_{a,b}) = \text{hacia arriba} \left(\frac{2000}{Bfr(\Pi_{a,b})} \right) = \text{hacia arriba} \left(\frac{2000}{116} \right) = 18$$



Finalizada la estimación del número de bloques de la rama izquierda del árbol, podemos proceder con la rama derecha, de forma análoga.

Comenzaremos con la proyección que se realiza sobre S para eliminar el atributo c , de tamaño mucho mayor que los demás.

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

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

$$= \text{hacia arriba} \left(\frac{200}{Bfr(\Pi_{b,e})} \right)$$

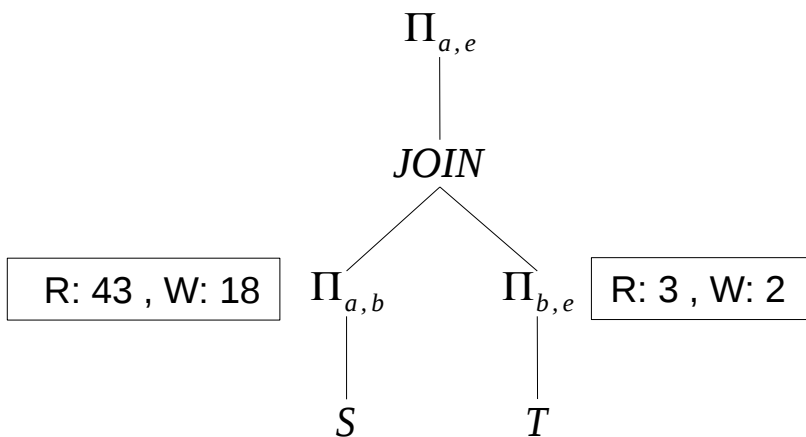
$$Bfr(\Pi_{b,e}) = \text{parte entera} \left(\frac{B-C}{L(\Pi_{b,e})} \right) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{b,e})} \right)$$

$$L(\Pi_{b,e}) = 20 + 10 = 30$$

de modo que:

$$Bfr(\Pi_{b,e}) = \text{parte entera} \left(\frac{4096-20}{L(\Pi_{b,e})} \right) = \text{parte entera} \left(\frac{4096-20}{30} \right) = 135 \quad y$$

Escritura: $B(\Pi_{b,e}) = \text{hacia arriba} \left(\frac{200}{Bfr(\Pi_{b,e})} \right) = \text{hacia arriba} \left(\frac{200}{135} \right) = 2$



Estimado el número de bloques para realizar las operaciones de las ramas izquierda y derecha del árbol, podemos proceder al cálculo de bloques necesario para la operación de reunión natural.

Siendo b el atributo de reunión, y siendo b clave en T, no sería necesaria la ordenación de la relación T mediante ese atributo. Sin embargo, se realiza una operación de proyección sobre T, cuyo resultado no tiene por

qué estar ordenador y, desde luego, no tiene un índice sobre b . De este modo, hará falta ordenar por el atributo b tanto la relación de la izquierda, para lo que se estima que harán falta las siguientes operaciones de entrada salida sobre bloques:

Para ordenar un bloque es necesario leerlo y después escribirlo. De modo que se puede hacer con dos operaciones. Pero es necesario ordenar ambos lados de la operación de reunión, por lo que habrá que realizar dos operaciones de lectura y escritura para ordenar los resultados de cada una de las dos selecciones.



ugr

Universidad de Granada

Departamento de Ciencias de la Computación
e Inteligencia Artificial



El coste de esta operación se estima en $18 \cdot \log_2(18) = 75.058650026 \approx 76$ operaciones de bloque para la proyección sobre S y en $2 \cdot \log_2(2) = 2$ operaciones de bloque para la proyección sobre T.

La operación de reunión natural requiere leer y mezclar todos los registros de las relaciones a la izquierda y a la derecha sin leer cada bloque más de una vez, por lo que será necesario:

$$\text{Lectura: } B(\Pi_{a,b}) + B(\Pi_{b,e}) = 18 + 2 = 20$$

y el resultado de la reunión natural necesita un número de bloques suficiente para almacenar los registros resultantes de esta operación, o sea:

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

$$N(\text{JOIN}) = \frac{N(\Pi_{a,b}) \cdot N(\Pi_{b,e})}{\max\{V(\Pi_{a,b}, b), V(\Pi_{b,e}, b)\}} = \frac{N(S) \cdot N(T)}{\max\{V(S, b), V(T, b)\}} = \frac{2000 \cdot 200}{200} = 2000$$

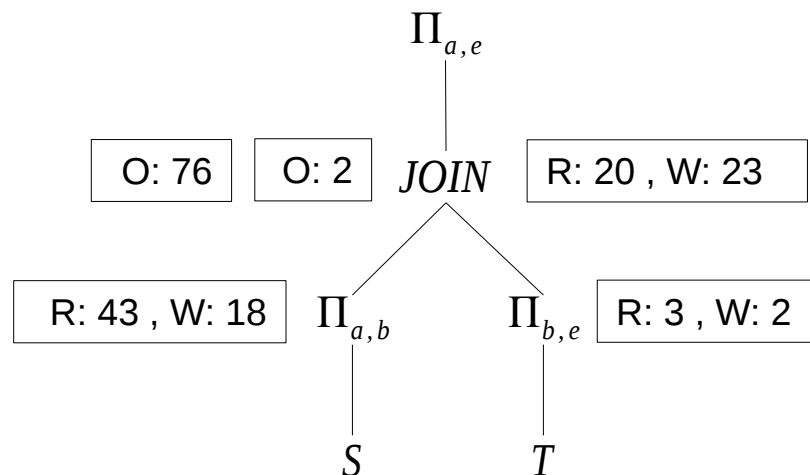
$$Bfr(\text{JOIN}) = \text{parte entera} \left(\frac{B - C}{L(\text{JOIN})} \right) = \text{parte entera} \left(\frac{4096 - 20}{L(\text{JOIN})} \right)$$

$$L(\text{JOIN}) = L(\Pi_{a,b}) + L(\Pi_{b,e}) - \text{Size}(b) = 35 + 30 - 20 = 45$$

de modo que:

$$Bfr(\text{JOIN}) = \text{parte entera} \left(\frac{4096 - 20}{L(\text{JOIN})} \right) = \text{parte entera} \left(\frac{4096 - 20}{45} \right) = 90 \quad \text{y}$$

$$\text{Escritura: } B(\text{JOIN}) = \text{hacia arriba} \left(\frac{N(\text{JOIN})}{Bfr(\text{JOIN})} \right) = \text{hacia arriba} \left(\frac{2000}{90} \right) = 23$$



Por último, calculamos el número de operaciones de bloques necesario para la proyección final.

$$B(\Pi_{a,e}) = \text{hacia arriba} \left(\frac{N(\Pi_{a,e})}{Bfr(\Pi_{a,e})} \right) = \text{hacia arriba} \left(\frac{N(\text{JOIN})}{Bfr(\Pi_{a,e})} \right) = \text{hacia arriba} \left(\frac{2000}{Bfr(\Pi_{a,e})} \right)$$

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

$$L(\Pi_{a,e}) = 15 + 10 = 25$$



ugr

Universidad de Granada

Departamento de Ciencias de la Computación
e Inteligencia Artificial

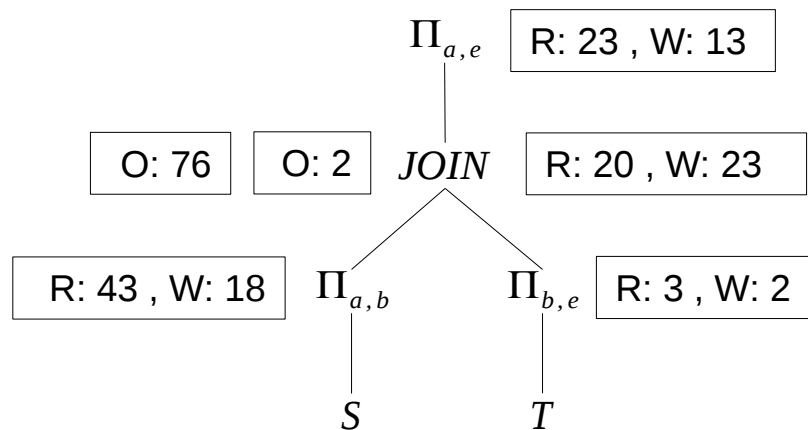


de modo que:

$$Bfr(\Pi_{a,e}) = \text{parte entera}\left(\frac{4096-20}{L(\Pi_{a,e})}\right) = \text{parte entera}\left(\frac{4096-20}{25}\right) = 163 \text{ y}$$

$$\text{Escritura: } B(\Pi_{a,e}) = \text{hacia arriba}\left(\frac{2000}{Bfr(\Pi_{a,e})}\right) = \text{hacia arriba}\left(\frac{2000}{163}\right) = 13$$

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:



resultado que mejora, sustancialmente, el número de bloques necesario para el ejercicio 1.

$$(43+18)+(3+2)+76+2+(20+23)+(23+13)=223 \text{ bloques}$$

De haber realizado las selecciones después de las proyecciones, el montante total de bloques de la operación propuesta se incrementaría en la lectura y escritura de 18 bloques para la selección sobre S y en la lectura y escritura de 2 bloques para la selección sobre T. Es decir, en un total de 263 operaciones con bloques, lo que mejora a las 525 operaciones con bloques del ejercicio 1.

Por lo tanto, la consulta seleccionada para ejecución según su plan físico sería:

$$\Pi_{a,e}(\sigma_{(b \neq valor)}(\Pi_{a,b}(S)) \text{ JOIN } \sigma_{(b \neq valor)}(\Pi_{b,e}(T)))$$



ugr

Universidad de Granada

Departamento de Ciencias de la Computación
e Inteligencia Artificial



Pregunta 3

Enunciado

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

Respuesta

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*).

Administración de Bases de Datos

Grado en Ingeniería Informática

Resolución de la Prueba Parcial ABD (Voluntaria), Abril de 2018

Ejercicio 1

Enunciado

1. Se dispone de un archivo secuencial indexado con un factor de bloqueo de 4 registros para almacenar registros de longitud fija, con la siguiente estructura en el fichero maestro (de datos):

--	--	--	--

- a) **(1 punto)** rellena sobre el bloque del enunciado el resultado de insertar los registros con valores de clave 7, 2, 5 y 3,
- b) **(1 punto)** indica qué ocurre cuando se añaden los registros con valores de clave 6, 4, 8, 9 y 1, tanto en el fichero maestro como en el fichero de índice.

Resolución

Apartado a)

El resultado de completar el bloque con los valores especificados resulta en que, puesto que el bloque está almacenado en memoria, se aprovecha la oportunidad para insertar los valores ordenados hasta que se complete el bloque.

	2	3	5	7
--	---	---	---	---

Apartado b)

A la llegada del valor 6, ya no queda espacio en el bloque, lo cual fuerza a la creación de un nuevo bloque para el fichero maestro. Este procedimiento no puede realizarse si, previamente, no se construye el índice asociado al fichero ASL:

Índice: 2, 3, 5, 7

Posteriormente, se crea un segundo bloque en el fichero maestro para seguir insertando datos:

	2	3	5	7		6		
--	---	---	---	---	--	---	--	--

y se añade el valor nuevo de clave al índice:

Índice: 2, 3, 5, 6, 7

Aún queda espacio para los tres siguientes registros (4, 8 y 9), quedando los bloques de datos como sigue:

	2	3	5	7		6	4	8	9
--	---	---	---	---	--	---	---	---	---

y el índice como sigue:

Índice: 2, 3, 4, 5, 6, 7, 8, 9

Por último, cuando se inserta el último registro (1), se crea un nuevo bloque en el fichero maestro y se añade el citado registro:

	2	3	5	7		6	4	8	9		1			
--	---	---	---	---	--	---	---	---	---	--	---	--	--	--

y el índice como sigue:

Índice: 1, 2, 3, 4, 5, 6, 7, 8, 9

Ejercicio 2

Enunciado

2. Sean las relaciones R y S con los siguientes parámetros:

R (a,b,c)	S (a,d,e)	R (a,b,c)	S (a,d,e)
N(R) = 1000	N(S)=10000	V(R,a)=1000	V(S,a)=1000
Size(a)=20	Size(a)=20	V(R,b)=200	
Size(b)=40		V(R,c)=20	
Size(c)=100			V(S,d)=???
	Size(d)=20		V(S,e)=40
	Size(e)=40		

donde *a* es llave primaria de *R* y (*a,d*) es llave primaria de *S*, y donde el atributo *S.a* es llave externa a *R.a*

a) **(1 punto)** Considerando la estructura de llaves (primarias y externa) de *R* y *S*, deduce la variabilidad $V(S,d)$.

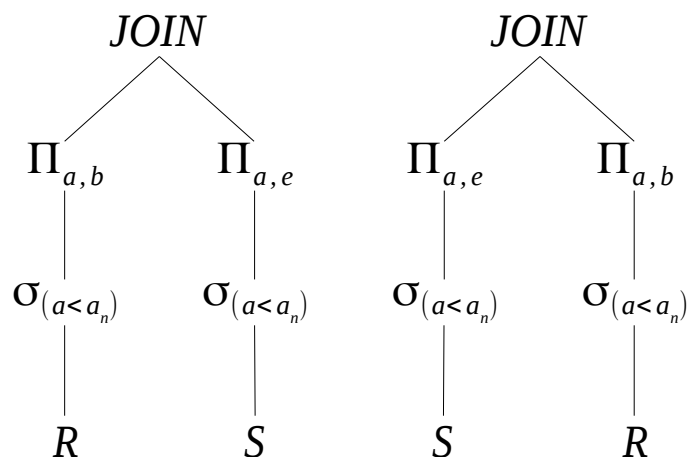
Teniendo en cuenta que, que el **tamaño de bloque** es de **4KB**, que la **cabecera** es de **40B**, que se usa **bloqueo fijo**, que los **bloques** son **homogéneos**, que en memoria sólo cabe un bloque de cada relación o resultado de operación intermedia, y suponiendo que las operaciones de **proyección y selección “respetan” los índices** (es decir, si la relación sobre la que se aplica la operación tiene un índice, el resultado también lo tendrá),

b) **(1 punto)** construye el plan lógico que se generaría para la consulta:

$$\sigma_{a < a_n}(\Pi_{a,b,e}(R JOIN S))$$

c) **(3 puntos)** determina el número de operaciones de E/S (plan físico) para el plan lógico del apartado b,

- d) **(2 puntos)** determina el número de operaciones E/S (planes físicos) que supondría la ejecución de los siguientes planes lógicos como alternativas al planteado por la consulta



- e) **(1 punto)** indica cuál de los planes físicos (el del apartado c o uno de los planes del apartado d) será seleccionado para ejecutar y por qué.

Resolución

Apartado a)

Alternativa 1

Según el enunciado, la relación R tiene 1000 registros, que $R.a$ es la clave y que la variabilidad de a en R , $V(R,a)$ es de 1000 (un valor distinto de a por cada registro de R).

Por su parte, la relación S tiene 10000 registros, una clave formada por $S.a$ y $S.d$ y una clave externa de $S.a$ a $R.a$. De acuerdo a las variabilidades de a en R y S , todos los valores de a presentes en R también lo están en S , o sea 1000.

Siendo la combinación de a y d la clave de S , podemos asumir que, puesto que hay 1000 valores de a , los 10000 registros pueden distribuirse uniformemente entre esos 1000 valores, dejándonos un total de:

$$\frac{10000 \text{ registros}}{1000 \text{ valores}} = 10 \text{ registros/valor}$$

De modo que tendríamos un valor de d por cada 1000 valores de a , con un total de 10 valores de d .

Como conclusión:

$$V(S,d) = 10$$

Alternativa 2

Se puede suponer que hay un valor distinto de d por cada valor de a en S , por lo que

$$V(S,d) = N(S) = 10000$$

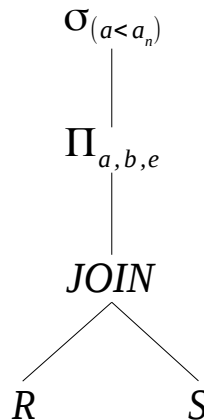
y, en general:

$$\frac{N(S)}{V(S,a)} \leq V(S,d) \leq N(S)$$

Sin embargo, puesto que sólo puede haber un valor en el catálogo para dicho parámetro, se opta por el valor superior de 10000.

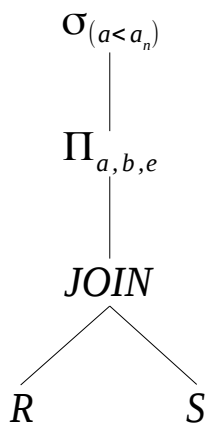
Apartado b)

El plan lógico para la consulta planteada, sin ningún tipo de modificación o heurística aplicada, es el siguiente:



Apartado c)

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 R para poder realizar la operación de reunión natural:

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

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

$$L(R) = 20 + 40 + 100 = 160$$

de modo que:

$$Bfr(R) = \text{parte entera} \left(\frac{4096-40}{160} \right) = 25 \quad y$$

$$B(R) = \text{hacia arriba} \left(\frac{1000}{25} \right) = 40$$

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

$$B(S) = \text{hacia arriba} \left(\frac{N(S)}{Bfr(S)} \right) = \text{hacia arriba} \left(\frac{10000}{Bfr(S)} \right)$$

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

$$L(S) = 20 + 20 + 40 = 80$$

de modo que:

$$Bfr(S) = \text{parte entera} \left(\frac{4096-40}{80} \right) = 50 \quad y$$

$$B(S) = \text{hacia arriba} \left(\frac{10000}{50} \right) = 200$$

Dado que R y S están ordenadas ambas a través de a , aunque en un índice multi-atributo en S , ambas relaciones están ordenadas por el atributo de reunión, no siendo necesaria ninguna modificación (ordenación) para poder realizar la reunión natural mediante la operación de mezcla (*merge*).

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

Lectura: $B(R) + B(S) = 40 + 200 = 240$

y el resultado de la reunión natural necesita un número de bloques suficiente para almacenar los registros resultantes de esta operación, o sea:

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

$$N(JOIN) = \frac{N(R) \cdot N(S)}{\max\{V(R, a), V(S, a)\}} = \frac{1000 \cdot 10000}{1000} = 10000$$

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

$$L(JOIN) = L(R) + L(S) - \text{Size}(a) = 160 + 80 - 20 = 220$$

de modo que:

$$Bfr(JOIN) = \text{parte entera} \left(\frac{4096-40}{220} \right) = \text{parte entera} \left(\frac{4096-40}{220} \right) = 18 \quad y$$

Escritura: $B(JOIN) = \text{hacia arriba} \left(\frac{N(JOIN)}{Bfr(JOIN)} \right) = \text{hacia arriba} \left(\frac{10000}{18} \right) = 556$

La operación de proyección debe leer los bloques resultantes de la operación anterior (556) y escribir tantos bloques como sean necesarios para almacenar el resultado de dicha operación, que se calculan a continuación:

$$B(\Pi_{a,b,e}) = \text{hacia arriba} \left(\frac{N(\Pi_{a,b,e})}{Bfr(\Pi_{a,b,e})} \right) = \text{hacia arriba} \left(\frac{N(JOIN)}{Bfr(\Pi_{a,b,e})} \right) = \text{hacia arriba} \left(\frac{10000}{Bfr(\Pi_{a,b,e})} \right)$$

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

$$L(\Pi_{a,b,e}) = 20 + 40 + 40 = 100$$

de modo que:

$$Bfr(\Pi_{a,b,e}) = \text{parte entera} \left(\frac{4096 - 40}{L(\Pi_{a,b,e})} \right) = \text{parte entera} \left(\frac{4096 - 40}{100} \right) = 40 \quad y$$

$$\text{Escritura: } B(\Pi_{a,b,e}) = \text{hacia arriba} \left(\frac{10000}{Bfr(\Pi_{a,b,e})} \right) = \text{hacia arriba} \left(\frac{10000}{40} \right) = 250$$

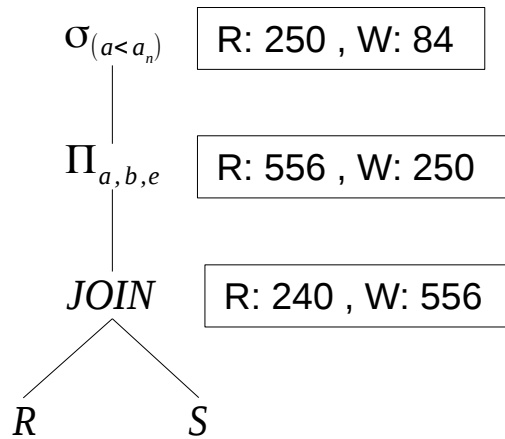
La operación de selección debe leer los bloques resultantes de la operación anterior (250) y escribir tantos bloques como sean necesarios para almacenar los registros que resultan de seleccionar aquellos que cumplan la condición, es decir:

$$B(\sigma_{a < a_n}) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n})}{Bfr(\sigma_{a < a_n})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n})}{Bfr(\Pi_{a,b,e})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n})}{40} \right)$$

Considerando que, al realizar la reunión natural, no afectamos a la variabilidad del atributo a procedente de las relaciones R y S , el número de tuplas resultantes de la selección sería:

$$N(\sigma_{a < a_n}) = \frac{1}{3} \cdot N(\Pi_{a,b,e}) = \frac{1}{3} \cdot N(JOIN) = \frac{10000}{3} = 3334$$

de modo que:



$$\text{Escritura: } B(\sigma_{a < a_n}) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n})}{40} \right) = \text{hacia arriba} \left(\frac{3334}{40} \right) = 84$$

El plan físico asociado al plan lógico planteado, suponiendo una sola resolución para cada operador (con un único coste por operador) sería:

Lo que arroja un total de 1936 operaciones de lectura y escritura de bloques.

Apartado d)

La única diferencia entre los dos planes propuestos es el orden en el que se realiza la reunión natural. En cualquiera de los casos, tanto considerando como ignorando la hipótesis planteada en el ejercicio propuesto (“que las operaciones de proyección y selección respetan los índices”), ambos planes son equivalentes en coste:

- si no se respetan los índices, ambos operandos (relaciones resultantes de la proyección) tendrían con ordenarse en ambos planes y con los mismos costes, y
- si se respetan los índices, ningún operando (relaciones resultantes de la proyección) tendría que ordenarse puesto que ya estarían ordenados por el criterio de reunión (atributo a).

En resumen, basta con calcular el coste de uno de los planes. Eso sí, respetando la hipótesis planteada en el enunciado (“que las operaciones de proyección y selección respetan los índices”) por lo que no es necesario ordenar ningún operando (relación) antes de realizar la operación de reunión.

Aprovecharemos los cálculos realizados en el Apartado c) para las relaciones R y S .

Para realizar la operación de selección sobre R , debemos calcular el número de bloques necesarios para almacenar el resultado de la misma:

$$B(\sigma_{a < a_n}(R)) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n}(R))}{Bfr(\sigma_{a < a_n}(R))} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n}(R))}{Bfr(R)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n}(R))}{25} \right)$$

por lo que:

$$N(\sigma_{a < a_n}(R)) = \frac{1}{3} \cdot N(R) = \frac{1000}{3} = 334 \quad y$$

$$B(\sigma_{a < a_n}(R)) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n}(R))}{25} \right) = \text{hacia arriba} \left(\frac{334}{25} \right) = 14$$

Para realizar la operación de selección sobre S , debemos calcular el número de bloques necesarios para almacenar el resultado de la misma:

$$B(\sigma_{a < a_n}(S)) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n}(S))}{Bfr(\sigma_{a < a_n}(S))} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n}(S))}{Bfr(S)} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n}(S))}{50} \right)$$

por lo que:

$$N(\sigma_{a < a_n}(S)) = \frac{1}{3} \cdot N(S) = \frac{10000}{3} = 3334 \quad y$$

$$B(\sigma_{a < a_n}(S)) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n}(S))}{50} \right) = \text{hacia arriba} \left(\frac{3334}{50} \right) = 67$$

Para realizar la operación de proyección en la rama de R , debemos calcular el número de bloques necesarios para almacenar el resultado de la misma:

$$B(\Pi_{a,b}) = \text{hacia arriba} \left(\frac{N(\Pi_{a,b})}{Bfr(\Pi_{a,b})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n}(R))}{Bfr(\Pi_{a,b})} \right) = \text{hacia arriba} \left(\frac{334}{Bfr(\Pi_{a,b})} \right)$$

$$Bfr(\Pi_{a,b}) = \text{parte entera} \left(\frac{B - C}{L(\Pi_{a,b})} \right) = \text{parte entera} \left(\frac{4096 - 40}{L(\Pi_{a,b})} \right)$$

$$L(\Pi_{a,b}) = 20 + 40 = 60$$

por lo que:

$$Bfr(\Pi_{a,b}) = \text{parte entera} \left(\frac{4096 - 40}{L(\Pi_{a,b})} \right) = \text{parte entera} \left(\frac{4096 - 40}{60} \right) = 67$$

$$\text{Escritura: } B(\Pi_{a,b}) = \text{hacia arriba} \left(\frac{334}{Bfr(\Pi_{a,b})} \right) = \text{hacia arriba} \left(\frac{334}{67} \right) = 5$$

Para realizar la operación de proyección en la rama de S , debemos calcular el número de bloques necesarios para almacenar el resultado de la misma:

$$B(\Pi_{a,e}) = \text{hacia arriba} \left(\frac{N(\Pi_{a,e})}{Bfr(\Pi_{a,e})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n}(S))}{Bfr(\Pi_{a,e})} \right) = \text{hacia arriba} \left(\frac{3334}{Bfr(\Pi_{a,e})} \right)$$

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

$$L(\Pi_{a,e}) = 20 + 40 = 60$$

por lo que:

$$Bfr(\Pi_{a,e}) = \text{parte entera} \left(\frac{4096 - 40}{L(\Pi_{a,e})} \right) = \text{parte entera} \left(\frac{4096 - 40}{60} \right) = 67$$

$$\text{Escritura: } B(\Pi_{a,e}) = \text{hacia arriba} \left(\frac{3334}{Bfr(\Pi_{a,e})} \right) = \text{hacia arriba} \left(\frac{3334}{67} \right) = 50$$

Puesto que, por hipótesis, ambas ramas conservan los índices que poseían R y S sobre el atributo a en cada una de ellas, no es necesaria la ordenación de los bloques, y bastará con la lectura de los 5 bloques de R y los 50 de S para proceder a realizar la reunión natural.

Para almacenar el resultado de dicha operación, será necesario un número de bloques como sigue:

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

$$\begin{aligned} N(JOIN) &= \frac{N(\Pi_{a,b}) \cdot N(\Pi_{a,e})}{\max\{V(\Pi_{a,b}, a), V(\Pi_{a,e}, a)\}} = \frac{N(\sigma_{a < a_n}(R)) \cdot N(\sigma_{a < a_n}(S))}{\max\{V(\sigma_{a < a_n}(R), a), V(\sigma_{a < a_n}(S), a)\}} = \\ &= \frac{334 \cdot 3334}{\max\{V(\sigma_{a < a_n}(R), a), V(\sigma_{a < a_n}(S), a)\}} \end{aligned}$$

El principal escollo de este cálculo consiste en calcular como las selecciones afectan a la variabilidad del atributo a . Al tratarse de una clave, había un valor distinto por cada registro (como se establece en el Apartado a) y, en concreto, 1000 valores para el atributo. Puesto que las operaciones de selección han reducido el conjunto de tuplas en un tercio, podemos deducir que el conjunto de valores de a se ha reducido en igual medida, por lo que

$$V(\sigma_{a < a_n}(R), a) = V(\sigma_{a < a_n}(S), a) = \frac{1000}{3} \approx 334$$

quedando:

$$N(JOIN) = \frac{334 \cdot 3334}{\max\{V(\sigma_{a < a_n}(R), a), V(\sigma_{a < a_n}(S), a)\}} = \frac{334 \cdot 3334}{334} = 3334$$

Además:

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

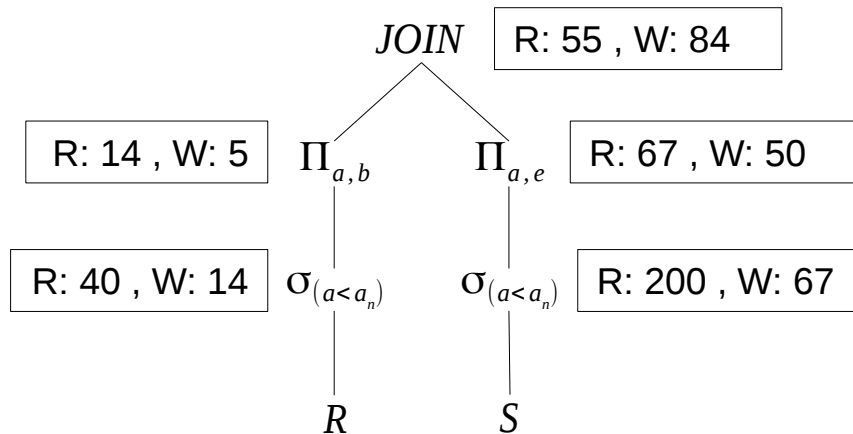
$$L(JOIN) = L(\Pi_{a,b}) + L(\Pi_{a,e}) - \text{Size}(a) = 60 + 60 - 20 = 100$$

por lo que:

$$Bfr(JOIN) = \text{parte entera}\left(\frac{B-C}{L(JOIN)}\right) = \text{parte entera}\left(\frac{4096-40}{100}\right) = 40 \quad y$$

Escritura: $B(JOIN) = \text{hacia arriba}\left(\frac{N(JOIN)}{Bfr(JOIN)}\right) = \text{hacia arriba}\left(\frac{3334}{40}\right) = 84$

El plan físico planteado (cualquiera de los dos tiene el mismo coste por lo indicado al principio del apartado) tiene un coste de:



Lo que arroja un total de 596 operaciones de entrada y salida de bloques.

Apartado e)

Evidentemente, se seleccionará cualquiera de los dos planes indicados en el Apartado d) por su evidente ahorro en el número de operaciones de entrada y salida.