

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)*.

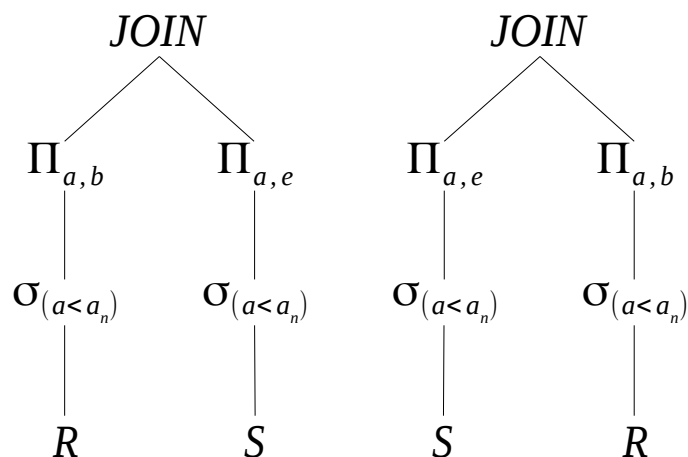
Teniendo en cuenta que, que el **tamaño de bloque** es de **4KB**, que la **cabecera** es de **40B**, que se usa **bloqueo fijo**, que los **bloques** son **homogéneos**, que en memoria sólo cabe un bloque de cada relación o resultado de operación intermedia, y suponiendo que las operaciones de **proyección 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

$$V(S,d) = N(S) = 10000$$

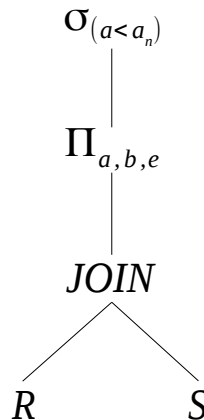
y, en general:

$$\frac{N(S)}{V(S,a)} \leq V(S,d) \leq N(S)$$

Sin embargo, puesto que sólo puede haber un valor en el catálogo para dicho parámetro, se opta por el valor superior de 10000.

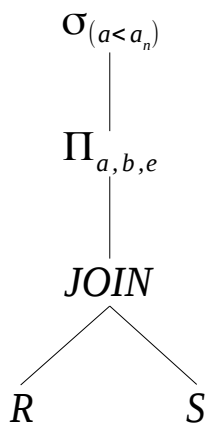
Apartado b)

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:

$$Bfr(\Pi_{a,b,e}) = \text{parte entera} \left(\frac{4096 - 40}{L(\Pi_{a,b,e})} \right) = \text{parte entera} \left(\frac{4096 - 40}{100} \right) = 40 \quad y$$

$$\text{Escritura: } B(\Pi_{a,b,e}) = \text{hacia arriba} \left(\frac{10000}{Bfr(\Pi_{a,b,e})} \right) = \text{hacia arriba} \left(\frac{10000}{40} \right) = 250$$

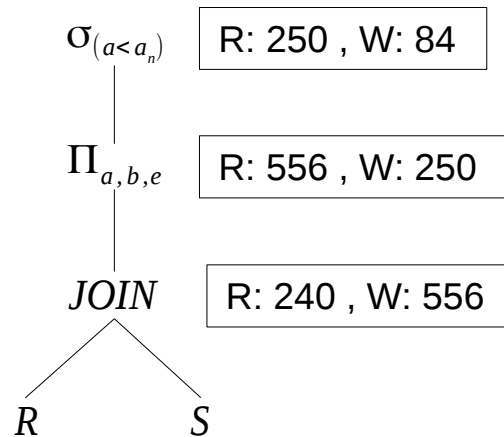
La operación de selección debe leer los bloques resultantes de la operación anterior (250) y escribir tantos bloques como sean necesarios para almacenar los registros que resultan de seleccionar aquellos que cumplan la condición, es decir:

$$B(\sigma_{a < a_n}) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n})}{Bfr(\sigma_{a < a_n})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n})}{Bfr(\Pi_{a,b,e})} \right) = \text{hacia arriba} \left(\frac{N(\sigma_{a < a_n})}{40} \right)$$

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:

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

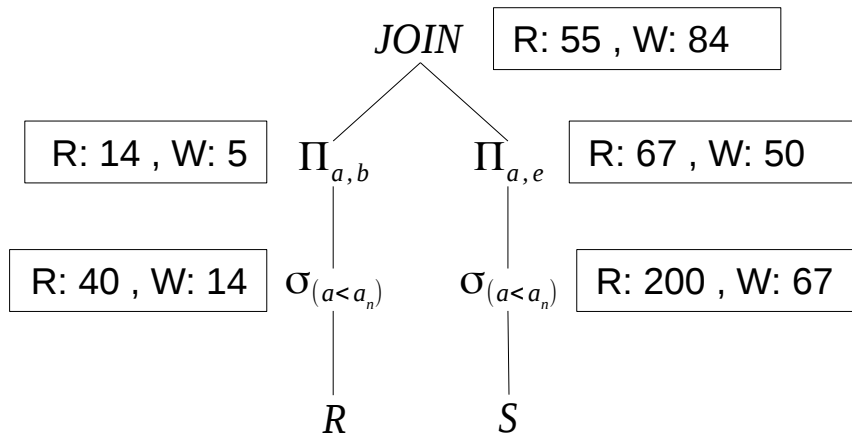
$$L(JOIN) = L(\Pi_{a,b}) + L(\Pi_{a,e}) - \text{Size}(a) = 60 + 60 - 20 = 100$$

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.