

Prueba Parcial ABD (Voluntaria)
29 de Abril de 2013

Nombre:

1. **(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:

$$\Pi_{a,d}(R \text{ JOIN } S)$$

2. **(2 puntos)** Explica qué elementos del nivel interno de Oracle se crearían y cuánto ocuparía la tabla recién creada si se ejecuta la consulta:

```
create table prueba(nombre varchar2(40) primary key, DNI
varchar2(8)) tablespace users
storage (initial 40k next 20k maxextents 10);
```

3. **(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.
4. **(2 puntos)** Propón dos planes lógicos para la siguiente consulta e indica de qué dependerá que se escoja uno u otro de sus planes físicos asociados:

$$\Pi_C(\sigma_{A=a \wedge B=b}(R))$$

5. **(2 puntos)** Indica si son ciertas o falsas las siguientes afirmaciones y, brevemente, explica por qué:
- a) El tiempo que se tarda en reorganizar un fichero ASI (Archivos Secuenciales Indexados) denso depende únicamente del tiempo necesario para reordenar el fichero de desbordamiento y de reescribir ordenadamente el nuevo fichero maestro.
 - b) Para consultas por rango, los ASI (Archivos Secuenciales Indexados) son más adecuados.

Resolución de la prueba parcial de ABD en 2013

Pregunta 1

Establecemos que el número de operaciones de E/S coincide con el número de bloques que es necesario leer y escribir al disco, tanto de las relaciones originales como de las relaciones intermedias, necesarias para el cálculo del resultado.

Número de bloques en R:

$$L_R = 20 B + 60 B + 20 B = 100 B$$

$$Bfr_R = \text{parte entera} ((B - C) / L_R) = \text{parte entera} ((2048 B - 20 B) / 100 B) = 20 \text{ registros / bloque}$$

Hay que leer *redondeo hacia arriba* $(5000 / 20) = 250$ bloques para cargar R

Para que la reunión natural sea eficiente, R debería estar ordenado por el atributo común, lo que implica $250 \cdot \log_2 250 = 1992$ bloques leídos

Número de bloques en S:

$$L_S = 20 B + 40 B = 60 B$$

$$Bfr_S = \text{parte entera} ((B - C) / L_S) = \text{parte entera} ((2048 B - 20 B) / 60 B) = 33 \text{ registros / bloque}$$

Hay que leer *redondeo hacia arriba* $(200 / 33) = 7$ bloques para cargar S

Número de bloques en R JOIN S:

$$L_{JOIN} = 100 B + 60 B - 20 B = 140 B$$

$$Bfr_{JOIN} = \text{parte entera} ((B - C) / L_S) = \text{parte entera} ((2048 B - 20 B) / 140 B) = 14 \text{ registros / bloque}$$

Número de tuplas del JOIN: $N_{JOIN} = N_R \cdot N_S / \max \{V(R,c), V(S,c)\} = 5000 \cdot 200 / 5 = 200000$ registros

Hay que escribir *redondeo hacia arriba* $(200000 / 14) = 14286$ bloques del resultado del JOIN

Hay que cargar 14286 bloques del resultado del JOIN para resolver la proyección

Número de bloques de la proyección

$$L_{PROYECCION} = 20 B + 40 B = 60 B$$

$$Bfr_{PROYECCION} = \text{parte entera} ((B - C) / L_{PROYECCION}) = \text{parte entera} ((2048 - 20) / 60) = 33 \text{ registros / bloque}$$

Hay que escribir *redondeo hacia arriba* $(200000 / 33) = 6061$ bloques

En resumen, hay que leer y escribir:

- leer 1992 bloques de R para ordenar los 250 bloques,
- escribir los 250 bloques de R ordenados,
- leer 250 bloques de R ordenado,
- leer 7 bloques de S,
- escribir 14286 bloques del JOIN,
- leer 14286 bloques del JOIN, y
- escribir 6061 bloques de la proyección

o sea, **37132 bloques** en total.

Pregunta 2

Durante la creación de la tabla se crean:

- una segmento,
- una extensión y
- un bloque.

Teniendo en cuenta que la tabla tiene una clave primaria, también se crearía un índice.

La extensión inicial y el segmento ocupan un total de 40 KB.

Pregunta 3

Índice sobre A

Cada registro del índice tiene un tamaño $L(I_A) = 10\text{ B} + 10\text{ B} = 20\text{ B}$

El factor de bloqueo del índice sobre A será: *parte entera* $((B - C) / L(I_A)) = \text{parte entera} ((4096 - 10)/20) = 204$ registros/bloque

El número de registros del índice es igual al número de registros de la relación R $N(R)$

El número de bloques será *redondeo hacia arriba* $(N(I_A) / \text{Bfr}(I_A)) = \text{redondeo hacia arriba} (1000 / 204) = 5$ bloques

Índice sobre B

Dado que un índice tiene que tener valores únicos y el atributo b no los tiene, podemos montar un índice multinivel en el que, el nivel raíz contenga valores únicos y una referencia a bloques del primer nivel. Dichos bloques, contendrán pares valor del atributo b y posiciones del archivo maestro.

Cada registro del índice de primer nivel tiene un tamaño $L(I_{B,1}) = 40\text{ B} + 10\text{ B} = 50\text{ B}$

El factor de bloqueo del índice de primer nivel será *parte entera* $((B - C) / L(I_{B,1})) = \text{parte entera} ((4096 - 10)/50) = 81$ registros/bloque

El número de registros del índice de primer nivel es igual al número de registros de la relación R $N(R)$

El número de bloques del índice de primer nivel será *redondeo hacia arriba* $(N(I_A) / \text{Bfr}(I_{B,1})) = \text{redondeo hacia arriba} (1000 / 81) = 13$ bloques

Cada registro del índice raíz contendrá un valor del atributo b y una dirección de bloque del índice de primer nivel. Teniendo en cuenta que hay 200 valores distintos de b, este índice tendrá sólo 200 entradas distintas con direcciones a un bloque del índice de nivel 1 que contenga todas las entradas del mismo valor de b.

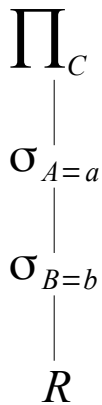
El factor de bloqueo del índice raíz será *parte entera* $((B - C) / L(I_{B,2})) = \text{parte entera} ((4096 - 10)/50) = 81$ registros/bloque

El número de bloques del índice raíz será *redondeo hacia arriba* $(N(I_{B,2}) / \text{Bfr}(I_{B,2})) = \text{redondeo hacia arriba} (200 / 81) = 3$ bloques

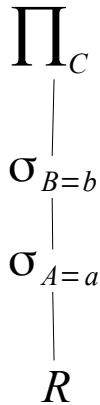
En total, para montar el índice sobre el atributo B, necesitaremos $13 + 3$ bloques (**16 bloques**).

Pregunta 4

Alternativa 1



Plan a



Plan b

Las relaciones iniciales de ambos planes (R) son las mismas (mismo tamaño, mismos bloques, mismas tuplas). Lo mismo ocurre con las segundas selecciones en cada plan y con las relaciones resultantes entre sí. La diferencia estará en la primera selección y el criterio de elección dependerá del tamaño de la relación intermedia resultante de la primera selección en cada plan.

Siendo así, para cada relación conocemos $V(R,A)$ y $V(R,B)$. El número de tuplas en ambas es de:

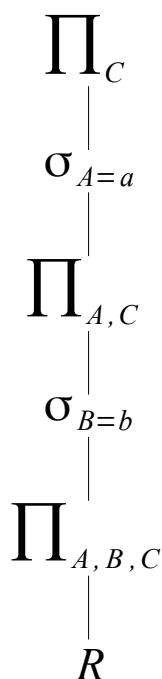
$$N_1 = \frac{N(R)}{V(R,A)} \text{ y } N_2 = \frac{N(R)}{V(R,B)}$$

Se escogerá el primer plan (a) sobre el segundo (b) si el tamaño de la primera selección en el plan a, en número de tuplas, es menor que el de la primera selección en el plan b:

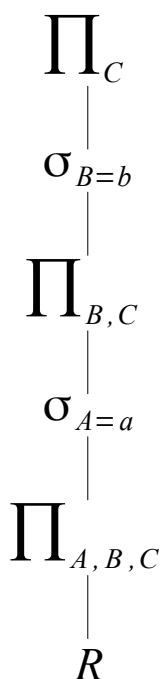
$$\frac{N(R)}{V(R,A)} < \frac{N(R)}{V(R,B)} \Rightarrow \frac{1}{V(R,A)} < \frac{1}{V(R,B)} \Rightarrow V(R,B) < V(R,A)$$

es decir, se escoge el primer plan si la variabilidad del atributo B en la relación es menor que la variabilidad del atributo A.

Alternativa 2



Plan a



Plan b

Las relaciones iniciales de ambos planes (R) son las mismas (mismo tamaño, mismos bloques, mismas tuplas). Se aplica la misma consideración a las relaciones resultantes y a la primera proyección en cada plan.

En cualquiera de los casos, se escogerá el plan que reduzca el número de bloques en mayor número antes que el otro.

El número de tuplas en la primera selección del plan (a) y del plan (b) es de:

$$N_1 = \frac{N(R)}{V(R,A)} \text{ y } N_2 = \frac{N(R)}{V(R,B)}$$

Dado que ambas relaciones intermedias tienen el mismo esquema, tendrán el mismo tamaño en las tuplas L_1 y, por tanto, el factor de bloqueo será exactamente el mismo Bfr_1 , siendo el número de bloques:

$$Bloques_1 = \frac{N(R)}{V(R,A) \cdot Bfr_1} \text{ y } Bloques_2 = \frac{N(R)}{V(R,B) \cdot Bfr_1}$$

Si comparamos, el número de bloques sólo dependerá de la variabilidad de los atributos, escogiendo el plan (a) si la variabilidad de B es menor (igual que en la alternativa

1).

La segunda proyección es asunto distinto, ya que los resultados de la relación intermedia en cada plan no tiene por qué tener el mismo tamaño en número de bloques.

$$Bfr_{a,2} = \frac{B-C}{V_A+V_C} \text{ y } Bfr_{b,2} = \frac{B-C}{V_B+V_C}$$

El tamaño en bloques de cada relación será de:

$$Bloques_{a,2} = \frac{N(R)}{V(R,A) \cdot \frac{B-C}{V_A+V_C}} \text{ y } Bloques_{b,2} = \frac{N(R)}{V(R,B) \cdot \frac{B-C}{V_B+V_C}}$$

Se escogerá el plan (a) sobre el (b) si el número de bloques de dicho plan es menor:

$$\begin{aligned} \frac{N(R)}{V(R,A) \cdot \frac{B-C}{V_A+V_C}} < \frac{N(R)}{V(R,B) \cdot \frac{B-C}{V_B+V_C}} &\Rightarrow \frac{1}{V(R,A) \cdot \frac{B-C}{V_A+V_C}} < \frac{1}{V(R,B) \cdot \frac{B-C}{V_B+V_C}} \Rightarrow \\ \frac{V_A+V_C}{V(R,A) \cdot (B-C)} < \frac{V_B+V_C}{V(R,B) \cdot (B-C)} &\Rightarrow \frac{V_A+V_C}{V(R,A)} < \frac{V_B+V_C}{V(R,B)} \end{aligned}$$

Pregunta 5

- a) **Falso**, porque depende del tiempo que tardemos en reordenar el fichero de desbordamiento, el tiempo que se tarda en reescribir ordenadamente el nuevo fichero maestro (mezclando el fichero maestro antiguo con el de desbordamiento reordenado) y el tiempo que tardamos en regenerar el índice.
- b) **Verdadero**, pero esto sólo ocurre cuando se consulta por el mismo atributo del índice. Si no, es altamente ineficiente.