

# WUOLAH



irenychuchu

[www.wuolah.com/student/irenychuchu](http://www.wuolah.com/student/irenychuchu)



14603

## examen2019.pdf

*Parcial*



**3º Administración de Bases de Datos**



**Grado en Ingeniería Informática**



**Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación  
Universidad de Granada**

**CUNEF**

**POSTGRADO EN  
DATA SCIENCE**

Excelencia, futuro, éxito.

 **Santander**

*Programa Financiación a la  
Excelencia CUNEF-Banco  
Santander e incorporación  
al banco finalizado el máster.*

# Resolución Prueba Parcial Voluntaria 2019

## Ejercicio 1

### Enunciado

Sean las relaciones R y S con los siguientes parámetros:

R (a,b,c)	S (d,e,b)	R (a,b,c)	S (d,e,b)
N(R) = 1000	N(S)=5000	...	
Size(a)=20		V(R,a)=1000	
Size(b)=30	Size(b)=30	V(R,b)=200	V(S,b)=500
Size(c)=100		V(R,c)=20	
	Size(d)=20		V(S,d)=5000
	Size(e)=40		V(S,e)=40

donde *a* es llave primaria de *R* y *d* es llave primaria de *S*, y donde **no** existe una relación de llave externa entre las relaciones *R* y *S*, aunque ambas tengan un atributo común en nombre y dominio (con valores comunes) *b*.

Teniendo en cuenta 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 únicamente cabe un bloque de cada relación o resultado de operación intermedia, y considerando que las operaciones de **proyección y selección “no respetan” los índices** (es decir, si las relaciones sobre las que se aplica la operación tienen un índice, el resultado de la misma no está indexado),

- a) **(1 punto)** construye el plan lógico que se generaría para la consulta:

$$\sigma_{e=e_k}(\Pi_{a,e}(R JOIN S))$$

- b) **(3 puntos)** determina el número de operaciones de E/S (plan físico) para el plan lógico del apartado a,
- c) **(3 puntos)** propón un plan lógico cuyo plan físico mejore el del apartado b, justificando numéricamente la mejora.



**CUNEF**

**POSTGRADO EN  
DATA SCIENCE**

*Haz como  
Marcos y  
convierte tu  
talento en  
oportunidades  
profesionales.*

“ El Máster en Data Science de CUNEF me ha permitido tanto ampliar mis conocimientos de programación y matemáticas como conseguir el trabajo que quería. Era importante para mí encontrar un máster con conocimientos no sólo teóricos, sino también enfocado en las aplicaciones prácticas que tiene la ciencia de datos para resolver problemas de negocio.”

MARCOS BARERRA - Data Scientist

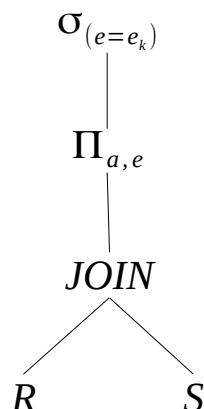


**Más de 1.600 acuerdos con empresas.**

Excelencia,  
futuro, **éxito.**

## Resolución

### Apartado a)



### 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  $R$  para poder realizar la operación de reunión natural:

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

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

$$L(R) = 20 + 30 + 100 = 150$$

de modo que:

$$Bfr(R) = \text{parte entera} \left( \frac{4096-40}{150} \right) = 27 \quad y$$

$$B(R) = \text{hacia arriba} \left( \frac{1000}{27} \right) = 38$$

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

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

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

$$L(S) = 30 + 20 + 40 = 90$$

de modo que:

$$Bfr(S) = \text{parte entera} \left( \frac{4096 - 40}{90} \right) = 45 \quad y$$

$$B(S) = \text{hacia arriba} \left( \frac{5000}{45} \right) = 112$$

Dado que ni  $R$  ni  $S$  están ordenadas a través del atributo de reunión  $b$ , es necesario realizar la ordenación de ambas relaciones para poder realizar la reunión natural mediante la operación de mezcla (*merge*).

Se estima que el número de operaciones de lectura y escritura de bloques para ordenar  $R$  es de:

$$B(R) \times \log_2(B(R)) = 38 \times \log_2(38) \approx 200$$

y que el número de operaciones de lectura y escritura de bloques para ordenar  $S$  es de:

$$B(S) \times \log_2(B(S)) = 112 \times \log_2(112) \approx 763$$

En ambos casos, se han redondeado los cálculos hacia arriba.

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 leer:

$$B(R) + B(S) = 38 + 112 = 150$$

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

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

$$N(JOIN) = \frac{N(R) \cdot N(S)}{\max\{V(R, b), V(S, b)\}} = \frac{1000 \cdot 5000}{500} = 10000$$

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

$$L(JOIN) = L(R) + L(S) - \text{Size}(b) = 150 + 90 - 30 = 210$$

de modo que:

$$Bfr(JOIN) = \text{parte entera} \left( \frac{4096 - 40}{L(JOIN)} \right) = \text{parte entera} \left( \frac{4096 - 40}{210} \right) = 19 \quad y$$

$$B(JOIN) = \text{hacia arriba} \left( \frac{N(JOIN)}{Bfr(JOIN)} \right) = \text{hacia arriba} \left( \frac{10000}{19} \right) = 527$$

La operación de proyección debe leer los bloques resultantes de la operación anterior y escribir tantos bloques como sean necesarios para almacenar el resultado de dicha operación, que se calculan a continuación:

$$B(\Pi_{a,e}) = \text{hacia arriba} \left( \frac{N(\Pi_{a,e})}{Bfr(\Pi_{a,e})} \right) = \text{hacia arriba} \left( \frac{N(JOIN)}{Bfr(\Pi_{a,e})} \right) = \text{hacia arriba} \left( \frac{10000}{Bfr(\Pi_{a,e})} \right)$$

Más de 1.600  
acuerdos con  
empresas

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

$$L(\Pi_{a,e}) = 20 + 40 = 60$$

de modo que:

$$Bfr(\Pi_{a,e}) = \text{parte entera}\left(\frac{4096-40}{L(\Pi_{a,e})}\right) = \text{parte entera}\left(\frac{4096-40}{60}\right) = 67 \quad y$$

$$B(\Pi_{a,e}) = \text{hacia arriba}\left(\frac{10000}{Bfr(\Pi_{a,e})}\right) = \text{hacia arriba}\left(\frac{10000}{67}\right) = 150$$

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:

$$B(\sigma_{e=e_k}) = \text{hacia arriba}\left(\frac{N(\sigma_{e=e_k})}{Bfr(\sigma_{e=e_k})}\right) = \text{hacia arriba}\left(\frac{N(\sigma_{e=e_k})}{Bfr(\Pi_{a,e})}\right) = \text{hacia arriba}\left(\frac{N(\sigma_{e=e_k})}{67}\right)$$

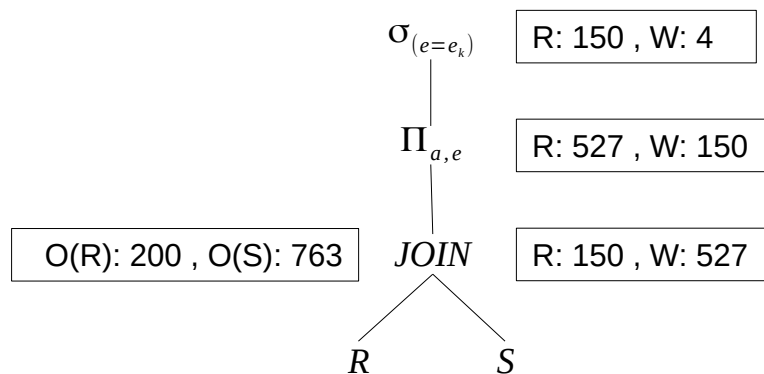
Considerando que, al realizar la reunión natural, no afectamos a la variabilidad del atributo  $e$  procedente de la relación  $S$ , y a que el número de registros resultantes de la proyección (10000) permiten la variabilidad del atributo  $e$  (al ser mucho mayor), el número estimado de tuplas resultantes de la selección sería:

$$N(\sigma_{e=e_k}) = \frac{1}{V(S,e)} \cdot N(\Pi_{a,e}) = \frac{1}{V(S,e)} \cdot N(JOIN) = \frac{1}{40} \cdot 10000 = 250$$

de modo que:

$$B(\sigma_{e=e_k}) = \text{hacia arriba}\left(\frac{N(\sigma_{e=e_k})}{67}\right) = \text{hacia arriba}\left(\frac{250}{67}\right) = 4$$

**El plan físico asociado a la consulta planteada (si se supone una única resolución para cada operador) involucraría un número de lecturas y escrituras de bloques de 2471 operaciones.**



amazon

McKinsey & Company

KPMG

accenture

pwc

Morgan Stanley

CUNEF

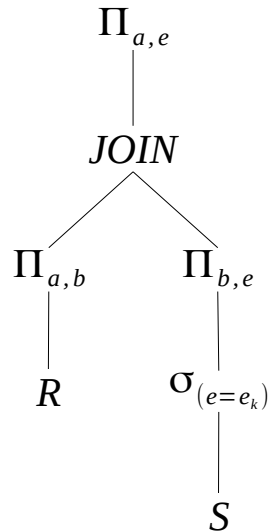
Excelencia,  
futuro, éxito.

WUOLAH



## Apartado c)

Un plan lógico más eficiente al propuesto en el apartado a) se obtiene realizando las proyecciones y las selecciones lo antes posible, con el fin de reducir el número de bloques involucrados en las operaciones más costosas (como la reunión natural), como se ven en el plan lógico que sigue:



Aprovecharemos los cálculos realizados en el apartado b) para el número de bloques de las relaciones  $R$  y  $S$ .

La operación de proyección sobre  $R$  debe leer los bloques de dicha relación (38) y escribir tantos bloques como sean necesarios para almacenar el resultado de la proyección, que se calculan a continuación:

$$B(\Pi_{a,b}) = \text{hacia arriba} \left( \frac{N(\Pi_{a,b})}{Bfr(\Pi_{a,b})} \right) = \text{hacia arriba} \left( \frac{N(R)}{Bfr(\Pi_{a,b})} \right) = \text{hacia arriba} \left( \frac{1000}{Bfr(\Pi_{a,b})} \right)$$

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

$$L(\Pi_{a,b}) = 20 + 30 = 50$$

de modo que:

$$Bfr(\Pi_{a,b}) = \text{parte entera} \left( \frac{4096-40}{L(\Pi_{a,b})} \right) = \text{parte entera} \left( \frac{4096-40}{50} \right) = 81 \quad \text{y}$$

$$B(\Pi_{a,b}) = \text{hacia arriba} \left( \frac{1000}{Bfr(\Pi_{a,b})} \right) = \text{hacia arriba} \left( \frac{1000}{81} \right) = 13$$

La operación de selección sobre  $S$  debe leer los bloques de dicha relación (112) y escribir tantos bloques como sean necesarios para almacenar el resultado de la selección, que se calculan a continuación:

$$B(\sigma_{e=e_k}) = \text{hacia arriba} \left( \frac{N(\sigma_{e=e_k})}{Bfr(\sigma_{e=e_k})} \right) = \text{hacia arriba} \left( \frac{N(\sigma_{e=e_k})}{Bfr(S)} \right) = \text{hacia arriba} \left( \frac{N(\sigma_{e=e_k})}{45} \right)$$

$$N(\sigma_{e=e_k}) = \frac{1}{V(S, e)} \cdot N(S) = \frac{1}{40} \cdot 5000 = 125$$

por lo que:

$$B(\sigma_{e=e_k}) = \text{hacia arriba} \left( \frac{N(\sigma_{e=e_k})}{45} \right) = \text{hacia arriba} \left( \frac{125}{45} \right) = 3$$

La operación de proyección sobre lo seleccionado sobre  $S$  debe leer los bloques resultantes de la selección y escribir tantos bloques como sean necesarios para almacenar el resultado de la proyección, que se calculan a continuación:

$$B(\Pi_{b,e}) = \text{hacia arriba} \left( \frac{N(\Pi_{b,e})}{Bfr(\Pi_{b,e})} \right) = \text{hacia arriba} \left( \frac{N(\sigma_{e=e_k})}{Bfr(\Pi_{b,e})} \right) = \text{hacia arriba} \left( \frac{125}{Bfr(\Pi_{b,e})} \right)$$

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

$$L(\Pi_{b,e}) = 30 + 40 = 70$$

de modo que:

$$Bfr(\Pi_{b,e}) = \text{parte entera} \left( \frac{4096-40}{L(\Pi_{b,e})} \right) = \text{parte entera} \left( \frac{4096-40}{70} \right) = 57 \quad y$$

$$B(\Pi_{b,e}) = \text{hacia arriba} \left( \frac{125}{Bfr(\Pi_{b,e})} \right) = \text{hacia arriba} \left( \frac{125}{57} \right) = 3$$

Dado que ninguna de las relaciones resultantes está ordenadas a través del atributo de reunión  $b$ , es necesario realizar la ordenación de ambas relaciones para poder realizar la reunión natural mediante la operación de mezcla (*merge*).

Se estima que el número de operaciones de lectura y escritura de bloques para ordenar el resultado de la proyección sobre  $R$  es de:

$$B(\Pi_{a,b}) \times \log_2(B(\Pi_{a,b})) = 13 \times \log_2(13) \approx 49$$

y que el número de operaciones de lectura y escritura de bloques para ordenar  $S$  es de:

$$B(\Pi_{b,e}) \times \log_2(B(\Pi_{b,e})) = 3 \times \log_2(3) \approx 5$$

En ambos casos, se han redondeado los cálculos hacia arriba.

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

$$B(\Pi_{a,b}) + B(\Pi_{b,e}) = 13 + 3 = 16$$

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

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



$$N(JOIN) = \frac{N(\Pi_{a,b}) \cdot N(\Pi_{b,e})}{\max\{V(\Pi_{a,b}, b), V(\Pi_{b,e}, b)\}} = \frac{1000 \cdot 125}{500} = 250$$

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

$$L(JOIN) = L(\Pi_{a,b}) + L(\Pi_{b,e}) - \text{Size}(e) = 50 + 70 - 30 = 90$$

de modo que:

$$Bfr(JOIN) = \text{parte entera}\left(\frac{4096-40}{L(JOIN)}\right) = \text{parte entera}\left(\frac{4096-40}{90}\right) = 45 \quad y$$

$$B(JOIN) = \text{hacia arriba}\left(\frac{N(JOIN)}{Bfr(JOIN)}\right) = \text{hacia arriba}\left(\frac{250}{45}\right) = 6$$

La operación de proyección debe leer los bloques resultantes de la operación anterior y escribir tantos bloques como sean necesarios para almacenar el resultado de dicha operación, que se calculan a continuación:

$$B(\Pi_{a,e}) = \text{hacia arriba}\left(\frac{N(\Pi_{a,e})}{Bfr(\Pi_{a,e})}\right) = \text{hacia arriba}\left(\frac{N(JOIN)}{Bfr(\Pi_{a,e})}\right) = \text{hacia arriba}\left(\frac{250}{Bfr(\Pi_{a,e})}\right)$$

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

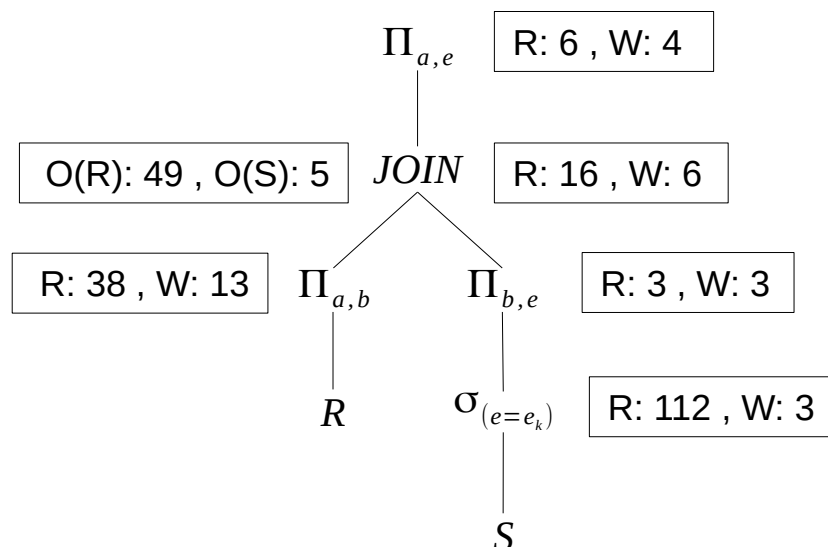
$$L(\Pi_{a,e}) = 20 + 40 = 60$$

de modo que:

$$Bfr(\Pi_{a,e}) = \text{parte entera}\left(\frac{4096-40}{L(\Pi_{a,e})}\right) = \text{parte entera}\left(\frac{4096-40}{60}\right) = 67 \quad y$$

$$B(\Pi_{a,e}) = \text{hacia arriba}\left(\frac{250}{Bfr(\Pi_{a,e})}\right) = \text{hacia arriba}\left(\frac{250}{67}\right) = 4$$

**El plan físico asociado a la consulta planteada (si se supone una única resolución para cada operador) involucraría un número de lecturas y escrituras de bloques de 258 operaciones.**



“ El Máster en Data Science de CUNEF es específico para el sector financiero y tiene como elemento diferenciador la combinación de ciencia (modelos y técnicas) y experiencia (conocimiento del negocio de las entidades financieras).”

JUAN MANUEL ZANÓN  
Director - CRM & Commercial  
Intelligence Expert

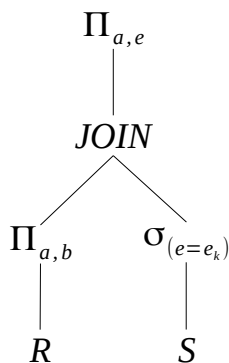
YGROUP



Convierte el desafío en  
oportunidad y especialízate  
en Data Science.

Más de 1.600  
acuerdos con  
empresas

Como puede observarse, la operación de proyección en la rama sobre  $S$  no aporta ninguna ventaja, ya que lee y escribe el mismo número de bloques. Podríamos pensar que un plan mejor entre los estudiados por el optimizador de consultas es el que sigue:



que reduce en 6 operaciones de lectura y escritura de bloques la estimación total, dejándola en un total de 252 operaciones.

POSTGRADO EN DATA SCIENCE



Excelencia,  
futuro, éxito.

WUOLAH

## Ejercicio 2

### Enunciado

Supongamos que las relaciones  $R$  y  $S$  planteadas en el *ejercicio 1*, fueron creadas dentro de un clúster, indicando que el atributo  $b$  representa a cada una de ellas en dicho clúster.

- (2 puntos)** Indica el espacio necesario (en número de bloques) para almacenar ambas relaciones, y cualquier particularidad que consideres respecto de los bloques necesarios para almacenar los contenidos de ambas relaciones.
- (1 punto)** ¿Qué mejora numérica supondría la existencia de dicho clúster en el plan físico asociado a la consulta del *apartado 1*?

### Resolución

#### Apartado a)

Según lo estudiado en el tema 3, la estructura de clúster almacena conjuntamente los registros de todas las tablas que forman parte del mismo. Para hacerlo, cada valor del atributo común  $b$  se almacena una sola vez, seguido de registros de las relaciones  $R$  y  $S$  a los que se les elimina el atributo común.

La única configuración de bloques capaz de almacenar la información de esta manera es la de bloques heterogéneos (que contienen registros de varias relaciones).

Podemos estimar el número de bloques necesarios para una estructura así considerando el número de bloques necesarios para almacenar los siguientes registros:

- 500 registros con el valor común (el máximo de las variabilidades del atributo  $b$  en las relaciones  $R$  y  $S$ ), de longitud 30,
- 1000 registros con el resto de los valores de la relación  $R$ , de longitud 120, y
- 5000 registros con el resto de los valores de la relación  $S$ , de longitud 60.

Si estimamos por separado los factores de bloqueo para los tres tipos de registro, tendremos:

- un factor de bloqueo de 135 para los valores comunes, y una cantidad de 4 bloques para almacenarlos,
- un factor de bloqueo de 33 para los registros de  $R$  sin  $b$ , y una cantidad de 31 bloques para almacenarlos, y
- un factor de bloqueo de 67 para los registros de  $S$  sin  $b$ , y una cantidad de 75 bloques para almacenarlo.

La estructura completa se podría almacenar en un total de 110 bloques (sumando el número de bloques necesarios para cada tipo de registro aunque los mismos no sean consecutivos). Este número de bloques mejora el número de bloques necesarios para almacenar  $R$  y  $S$  por separado (150 bloques, calculados en el apartado b del ejercicio 1). Por supuesto, sería necesario estimar el número de bloques necesarios para almacenar el índice del clúster (con un valor de  $b$  y una referencia a un registro en cada entrada), pero el número de bloques para almacenar 500 registros de este tipo no debe ser elevado.

Supongamos un tamaño de 4 B por cada referencia, lo que nos daría una longitud de 34 B para cada registro del índice. Para esta longitud, el factor de bloqueo del índice sería de 119 registros por bloque. Sólo serían necesarios 5 bloques para almacenar el índice.

En total, el clúster ocuparía 115 bloques en vez de 150 bloques necesario para almacenar las dos relaciones por separado.

### **Apartado b)**

En el caso de la existencia de dicho clúster, no sería necesario realizar ninguna ordenación por el atributo de reunión al estar los registros ya ordenados al almacenarlos. Por ello, podemos estimar que el plan físico calculado en el apartado b) del ejercicio 1 (para el plan lógico representando en el apartado a) del ejercicio 1) se reduce en 963 operaciones de lectura y escritura de bloques, necesarias para ordenar las dos relaciones, y en 35 operaciones de lectura ya que la operación de JOIN no lee 150 bloques de las dos relaciones sino 115 bloques correspondientes al clúster. Sin embargo, el resultado del JOIN requiere la escritura de los mismos 527 bloques.