

Relación de problemas I: Nivel interno

Ejercicio 1

Solución

Datos: $n = 10^6$ tuplas, $B = 4 \text{ KB} = 4096 \text{ B}$, $R = 2050 \text{ B}$ y suponemos bloqueo **fijo**.

Nota. Como no nos dicen el tamaño de las cabeceras, suponemos $C = 0 \text{ B}$.

Se pide:

- $Bfr = \left\lfloor \frac{B-C}{R} \right\rfloor = \left\lfloor \frac{4096-0}{2050} \right\rfloor = [1,98] = 1 \text{ registro/bloque}$
- % desperdicio = $W = \frac{(B-C)-(Bfr \cdot R)}{B-C} = \frac{4096 \text{ B} - 1 \cdot 2050 \text{ B}}{4096 \text{ B}} = 49,95 \%$
- % uso de los bloques = $100 - W = 100 - 49,95 = 50,05 \%$

Ejercicio 2

Solución

Datos:

- $n = 10^6$ tuplas; $B = 4096 \text{ B}$, $R = 120 \text{ B}$, $P = 6 \text{ B}$, $V = 8 \text{ B}$
- Suponemos $C = 0 \text{ B}$.

Se pide calcular el **número de bloques necesarios** para guardar los datos de un **ASI** (a) denso; (b) no denso.

a) Denso:

- $Bfr_{maestro} = \left\lfloor \frac{4096}{120} \right\rfloor = 34 \text{ registros/bloque}$
- $Bfr_{indice} = \left\lfloor \frac{4096}{6+8} \right\rfloor = [292,57] = 292 \text{ registros/bloque}$

Con estos datos, ya podemos calcular el número de bloques como:

- Número de bloques en el maestro = $\left\lceil \frac{n}{Bfr_{maestro}} \right\rceil = \left\lceil \frac{10^6}{34} \right\rceil = 29412 \text{ bloques}$
- Número de bloques en el índice = $\left\lceil \frac{n}{Bfr_{indice}} \right\rceil = \left\lceil \frac{10^6}{292} \right\rceil = 3425 \text{ bloques}$

En total, para el ASI denso se necesitan $29412 + 3425 = 32837$ bloques

b) No denso:

En primer lugar, los cálculos para el *fichero maestro* son los mismos, tal que se sigue cumpliendo que:

$$\text{Número de bloques en el maestro} = 29\,412 \text{ bloques}$$

Ahora hay que tener en cuenta que cada entrada de índice “apunta” a un bloque en el fichero maestro, tal que el número de entradas en el fichero índice coincide con el número de bloques en el fichero maestro: $n' = 29\,412 \text{ registros (en el fichero índice)}$

Por tanto, simplemente tenemos que calcular el número de bloques necesarios teniendo eso en cuenta:

$$\text{Número de bloques en el índice} = \left\lceil \frac{29\,412 \text{ registros de índice}}{292 \frac{\text{registros}}{\text{bloque}}} \right\rceil = 101 \text{ bloques}$$

Concluimos: número de bloques necesarios para ASI no denso: $101 + 29412 = 29\,513$ bloques.

Ejercicio 3.

Solución

Ventajas

- Es la única alternativa cuando el registro es mayor que el bloque.
- Permiten almacenar varios tipos de registro en un mismo archivo.
- Permite tener tipos de datos de longitud variable (`varchar`, entre otros).
- Permite un mejor aprovechamiento de los bloques ya que se pueden ajustar los registros mejor al tamaño del bloque e incluso dividir el último si no cabe en un bloque.

Inconvenientes

- Se ha de almacenar la estructura de los registros en las cabeceras de los bloques o en cada uno de los registros con lo que se pierde espacio de almacenamiento.
- Hay que usar marcas para separar campos entre sí.
- Algunos registros quedarán divididos en varios bloques al no poderse introducir un registro en el espacio libre del bloque donde deba insertarse.
- La búsqueda y colocación sobre un registro concreto es más compleja (por su estructura) y no inmediata.
- En línea con el ítem anterior, también cabe referir que genera problemas al reemplazar un registro por otro más grande.

Ejercicio 11.

Solución

Datos: $Bfr_{maestro} = 30$, $Bfr_{indice} = 100$, hay n registros de datos.

Se pide calcular, **para una ocupación de los bloques del 100%**, el número de los mismos que se necesitan para **guardar el índice** de un ASI:

(a) Denso:

Teniendo en cuenta que en este tipo de ASI, hay una entrada de índice por cada registro del fichero de datos, tendremos, de acuerdo a los datos del ejercicio, n entradas de índice. Con este dato, y el Bfr_{indice} , calcular el número de bloques del índice, $N(B)_{indiceDenso}$ resulta trivial:

$$N(B)_{indiceDenso} = \left\lceil \frac{n}{Bfr_{indice}} \right\rceil \text{ bloques de índice}$$

(b) No denso:

Ahora, recordemos, tenemos *una entrada de índice por cada bloque de datos*, luego el número de entradas de índice será el número de bloques el fichero maestro. Calculemos este dato, también sencillo a partir de los datos proporcionados: $N(B)_{maestro} = \left\lceil \frac{n}{Bfr_{maestro}} \right\rceil = \left\lceil \frac{n}{30} \right\rceil$. Con esto, ya podemos calcular

$$N(B)_{indiceNoDenso} = \left\lceil \frac{\left\lceil \frac{n}{30} \right\rceil}{Bfr_{indice}} \right\rceil = \left\lceil \frac{n}{Bfr_{indice} \cdot 30} \right\rceil = \left\lceil \frac{n}{100 \cdot 30} \right\rceil = \left\lceil \frac{n}{3000} \right\rceil \text{ bloques}$$

Ahora nos piden repetir los cálculos, suponiendo una ocupación del 80% de los bloques, tal que ahora cabrán sólo el 80% de los registros que cabían antes en los bloques (se supone que tanto en los datos como en el índice). Luego Bfr pasará a contarse como $0,8 \cdot Bfr$:

(a) Denso:

$$N(B)_{indiceDenso} = \left\lceil \frac{n}{0,8 \cdot Bfr_{indice}} \right\rceil = \left\lceil \frac{n}{0,8 \cdot 100} \right\rceil = \left\lceil \frac{n}{80} \right\rceil \text{ bloques}$$

(b) No Denso:

$$N(B)_{indiceNoDenso} = \left\lceil \frac{\frac{n}{Bfr_{maestro} \cdot 0,8}}{0,8 \cdot Bfr_{indice}} \right\rceil = \left\lceil \frac{n}{Bfr_{indice} \cdot Bfr_{maestro} \cdot 0,8^2} \right\rceil \text{ bloques}$$

Ejercicio 5

Solución.

La **opción a)** no sería correcta ya que si tenemos (pongamos el caso) registros grandes en bloques más pequeños, por ejemplo, ya tendríamos que fraccionar el registro, y sabemos que el bloqueo partido con registros partidos se vuelve tremendamente ineficiente.

En cuanto a la **opción b)**, no resultaría muy razonable si caben muchos registros en un bloque porque éstos son de tamaño reducido, y para no malgastar “un mínimo” de espacio al final del bloque partimos el registro, teniendo en cuenta lo ineficiente de recuperar registros partidos, como ya comentábamos antes.

Sobre la **opción c)**, se puede aplicar un razonamiento similar al anterior: si tenemos registros mucho más pequeños que el bloque, puede ser que nos sobre hueco al final del bloque, pero no lo suficiente para un registro completo, pues se partiría, con el ya comentado perjuicio en tiempo de recuperación de un *registro partido* en bloqueo partido. Quizás sería preferible dejar ese espacio desperdiciado y ganar en tiempo de acceso.

La **opción d)** sigue un planteamiento similar.

En definitiva, hablamos de que se debe de buscar un equilibrio entre el tamaño de un bloque y el tamaño del registro, teniendo un tamaño de registro lo suficientemente grande como para ocupar todo el espacio posible del bloque, y a su vez un bloque que sea lo justamente pequeño para que igualmente sobre el menor espacio posible. La idea, en definitiva es evitar tener que partir registros. Es por ello que me decanto por la **opción e)**.

Ejercicio 17

Solución

Son conceptos muy diferentes.

- El plan lógico equivale a la expresión en forma de árbol de álgebra relacional de la consulta que se ha planteado.
- El plan lógico surge tras analizar y pasar por el *parser* a la consulta original (y es posible que una posterior fase de elaboración de planes lógicos equivalentes).
- Un plan lógico llevará asociados varios planes físicos, y no a la inversa. Dicho de otra forma, de un plan físico no salen los planes lógicos, pero de un plan lógico sí salen varios planes físicos, que serán las diversas opciones a evaluar por el optimizador para ver con cuál se queda, de cara a ejecutar la consulta original [equivalente] óptimamente.
- Un plan físico indica el orden de realización de las operaciones, el plan físico indica de forma más detallada los costes del plan y los algoritmos usados para la obtención de los datos.

Ejercicio 8

Solución.

Debido a que tenemos campos de longitud variable, vamos calcular el valor de R siguiendo la expresión

$$R = a'(A + V + 1)$$

De esta forma:

$$R = 4 \cdot \left(0 + \frac{29 + 255 + 10 + 1}{4} + 1 \right) = 299 B$$

Además, como tenemos campos de longitud variable, podemos considerar que existe separador entre campos, tal que supondremos, que, por ejemplo, $M = 1 B$.

Antes de empezar a calcular el factor de bloqueo y el % uso, tenemos que pararnos en algo concreto del enunciado: dice que hay una entrada de directorio (por cada registro). Para cada entrada deberá existir, por tanto, alguna “representación” de cada registro. Cojamos, por ejemplo, el campo más “significativo de la tabla: nombre (30 B). De esta forma, tendremos que el espacio que no se puede usar será $C + 30 \cdot Bfr$.

Con esto, el factor de bloqueo se puede calcular como:

$$Bfr = \left\lfloor \frac{B - C - 30 \cdot Bfr}{R + M} \right\rfloor = \left\lfloor \frac{4096 - 10 - 30 \cdot Bfr}{299 + 1} \right\rfloor \Rightarrow \text{despejamos } Bfr \Rightarrow$$

$$Bfr = [12, 38] = 12 \text{ registros/bloque}$$

Por otra parte, para calcular el porcentaje de uso también tenemos que tener en cuenta las entradas de directorio:

$$\% uso = \frac{(R + M) \cdot Bfr}{B - C - 30 \cdot Bfr} = \frac{300 \cdot 12}{4096 - 10 - 30 \cdot 12} = 96,6 \%$$