|  |  |
| --- | --- |
| Titulación: GRADO INGENIERIA INFORMATICA  Año Académico: 2019/2020 -- Curso: 2º  Asignatura: Ficheros y bases de datos  Ejercicios de Repaso | UC3M. Universidad Carlos III de Madrid.  UC3M. Universidad Carlos III de Madrid. |

**Ejercicio de organizaciones de ficheros**

La información de un supermercado está almacenada en una base de datos y para obtener un mejor rendimiento, se analiza los procesos más críticos que actúan sobre la tabla PRODUCTOS. A continuación se muestran los resultados de este análisis.

|  |  |  |  |
| --- | --- | --- | --- |
| P | Descripción del proceso | filas | Frecuencia |
| P1: | SELECT price FROM products WHERE barcode=y; | 1 | 0.8 |
| P2: | SELECT barcode FROM products WHERE prod\_type=‘value’; | 21 | 0.1 |
| P3: | UPDATE products SET stock = stock-x WHERE barcode=y; | 1 | 0.1 |

El fichero contiene 16.800 filas cuya media es de 259B (de los cuales 230B son bytes útiles). El espacio de cubo es de un bloque (Ec=1), con un 10% de espacio libre distribuido (PCTFREE) y con información de control (cabecera de cubo) de 50B. El tamaño de bloque es de 1024B y se utiliza un dispositivo de soporte aleatorio. El objetivo es comparar la eficiencia de las tres organizaciones básicas no consecutivas (con el mismo espacio de cubo) que se presentan a continuación:

|  |  |
| --- | --- |
| O0: | Serial no consecutiva |
| O1: | Cluster disperso por CD=*barcode* en N=9000 con una gestión de desboramientos independiente en area serial cuya tasa de desbordamiento es de 0.1% (sobre el total de registros). |
| O2: | Cluster disperso por CD=*barcode* en N=6500 con una gestión de desbordamiento igual que la anterior pero con una tasa de desbordamiento de 2.5% (sobre el total de registros). |
|  |  |

**TO DO:**

1. (5 pnts.) Calcula los costes totales de todas las organizaciones propuestas y sus densidades (ideal, real y de ocupación). Realiza una discusión de los resultados, indica la organización más destacada y cuál de ellas se debería evitar y porqué. Propón mejoras a través de la indexación.

di= útil/real= 230/259 = 88,8%

Tc= (1024-50)(1-0,1) / 259 = 3 reg/cubo

O0: serial

N= 16800 / 3 = 5600 cubos

dr(O0) = 230 \*16800 / 5600 \* 1024 = 67,4%

C(O0,P1)= (N+1)/2= 2800,5 acc a cubo = 2800,5 acc

C(O0,P2)= N = 5600 acc

C(O0,P3)= C(O0,P1) +1 = 2801,5 acc

C(O0,P)= 0,8\*2800,5 + 0,1 \*5600+ 2801,5\*0,1= 3080,55 acc

O1: N=9000 por barcode y desb 0,1

🡪 r’ = 0,001\*16800 = 17 reg 🡪 N’= 17/3 = 6 cubos de desbordamiento

dr(O1) = 230 \*16800 / (9000+6) \* 1024 = 41,9%

do(O1) = (16800-17) / 9000 \* 3 = 62,16%

C(O1,P1)= 1 + 0,001\* (6+1)/2 = 1,035 acc = 1 acc

C(O1,P2)= N +N’ = 9006 acc

C(O1,P3)= C(O1,P1) + 1 = 1+1 acc = 2 acc

C(O1,P)= 0,8\*1 + 0,1 \*9006+ 2\*0,1= 901,6 acc

O2: N=6500 por barcode y desb 2,5%

🡪 r’ = 0,025\*16800 = 420 reg 🡪 N’= 420/3 = 140 cubos de desbordamiento

dr(O1) = 230 \*16800 / (6500+140) \* 1024 = 56,8%

do(O1) = (16800-420) / 6500 \* 3 = 84%

C(O2,P1)= 1 + 0,025\* (140+1)/2 = 2,76 acc

C(O2,P2)= N +N’ = 6640 acc

C(O2,P3)= C(O1,P1) + 1 = 3,76 acc

C(O2,P)= 0,8\*2,76 + 0,1 \*6640+ 3,76\*0,1= 666,6 acc

1. (5 pnts.) Teniendo en cuenta el análisis anterior, elija la organización que piense que es mejorable agregando índices. Teniendo en cuenta que los punteros tanto externo como interno son de 7B, calcule árboles B y B+ para las claves de indexación que estime oportunas. Y por último, realize el coste de la organización elegida con estos índices y discuta los resultados.

B para barcode

Te= 4B + 7B = 11B

m\*7 + k\* 11B <= 1024 🡪 m=k+1 ,, 18k<= 1017 🡪 k = 56 🡪 kmin=k/2=28 🡪 m=57 🡪 mmin=29

e=r= 16800

nivel nodos entradas acumulado

1 1 1 1

2 2 2\*28=56 57

3 2mmin=58 58\*28= 1624 1681

~~4 1682 47096 >e~~

**n=3**

B+ para prod\_type

Te= 1B(m\_longitud) +9,5B(media ocupación) + 1(long\_lista) + 21\*7 =158,5B

e = 800 entradas

m\*7 + (m-1) (1+9,5) <= 1024 🡪 17,5 m <= 1024 + 10,5 🡪 m= 59 🡪 mmin=(m+1)/2= 30

k\*158,5B + 7 <=1024 🡪 k = 6 🡪 kmin=(k+1)/2 = 3

#hojas= #nodos(n) = 800 /3 = 266 hojas

#nodos(n-1) = 266 hijos / mmin = 266/30= 8 nodos

#nodos(n-2) = 8 hijos / 30 <2 🡪 nodo raíz (nivel 1) 🡪 n-2=1 🡪 **n2=3**

O0’: serial con sendos índices

C(O0’,P1)= (n1 -1) +1 \*1 = 3 acc

C(O0’,P2)= (n2-1) + 21 \*1 = 23 acc

C(O0’,P3)= C(O0’,P1) + 1 = 4 acc

C(O0’, P)= 0,8\*3 + 0,1 \*23+ 4\*0,1= 5,1 acc

O1’: O1 con índice en prod\_type

C(O1’,P1)= C(O1,P1)= 1 acc

C(O1’,P2)= C(O0’,P1)= 23 acc

C(O1’,P3)= C(O1,P3) = 2 acc

C(O1’, P)= 0,8\*1 + 0,1 \*23 + 2\*0,1= 3,3 acc

O2’: O2 con índice en prod\_type

C(O2’,P1)= C(O2,P1)= 2,76 acc

C(O2’,P2)= C(O0’,P1)= 23 acc

C(O2’,P3)= C(O2,P3) = 3,76 acc

C(O2’, P)= 0,8\*2,76 + 0,1 \*23 + 3,76\*0,1= 4,9 acc