Titulación: GRADO INGENIERIA INFORMATICA Año Académico: 2019/2020 -- Curso: 2º Asignatura: Ficheros y bases de datos Ejercicio F4 – Ejercicio de estructuras de ficheros	uc3m Universidad Carlos III de Madrid

Se tiene un fichero que contiene 10000 registros de 1100 B cada uno, almacenados de manera serial consecutiva (O_0) en un dispositivo cuyo tamaño de bloque es de 1KB. Tras un proceso de optimización (basado exclusivamente en la introducción de marcas) se consigue reducir el tamaño del registro a 150B.

Los procesos, cuyas frecuencias son (0.4,0.299,0.3,0.001), a los que está sometido este fichero son los siguientes:

P_1: Inserción de nuevos registros 2 ind

P_2: Consulta por clave identificativa K_1 -ind k2

-ind k1

P_3: Consulta por clave no identificativa K_2 (con 710 valores distintos; 14 ocurrencias de media)

P_4: Borrado de registros antiguos (selección por rango en la 'fecha de entrada', K_3, en media afecta a 400 registros cada vez que se ejecuta)

Se pretende comparar la eficiencia de tres organizaciones base no consecutivas con cubos de Ec= 4 bloques y con espacio libre distribuido (10%). Las alternativas son las siguientes:

O_1: serial no consecutiva

O_2: secuencial no consecutiva por CO= K_1

O_3: direccionada por CD=K_2 sobre N=550 con gestión de desbordamientos en área independiente de organización serial. La organización presenta una tasa de desbordamientos del 11% del total de registros.

Se pide:

a) (5ptos) Calcula el coste global de la organización original y de cada una de las tres propuestas.
b) (4ptos) Tomando un puntero estándar de tamaño 4 B (tanto para punteros internos como externos), calcúlense índices arbóreos B sobre K1 y B+ sobre K2, y aplíquese cada uno donde proceda para obtener las organizaciones O_1', O_2' yO_3'. Calcúlese el coste global de cada uno de ellas (el tamaño de la clave K_2 es 12 B de media (variable), mientras que las claves K_1 y K_3 son de tamaño fijo, con 9 B y 8 B respectivamente).
c) (1pto) Qué organización es más conveniente. Opcionalmente, el alumno puede proponer nuevas mejoras sobre la organización elegida, u otras organizaciones diferentes que el alumno estime puedan mejorar la eficiencia del sistema descrito.

a) $n = \dfrac{1100B \cdot 10.000 \text{ regs}}{1024} = 10743$ bloques.

Insertar $C(O_0, P_1) = 1$ acc

Consulta iden. $C(O_0, P_2) = \dfrac{n+1}{2} = \dfrac{10743+1}{2} = 5372$ acc

Consulta no iden $C(O_0, P_3) = n = 10743$ acc

Borrar un rango $C(O_0, P_4) = n^{+\text{borrar}} = 10743 + 400 = 11143$ acc

$C(O_0, P) = 0'4 \cdot 1 + 0'299 \cdot 5372 + 0'30 \cdot 10743 \overset{+0'001 \cdot 11143}{=} \boxed{4840'27 \text{ acc}}$

$T_c = \dfrac{(4 \cdot 1024 - 0)\,0'9}{150} = 24$ reg/cubo $\qquad N = \dfrac{10\,000}{24} = 417$ cubos.

$C(O_1, P_1) = 1$ acc a cubo $= 4$ acc.

$C(O_1, P_2) = \dfrac{417+1}{2} = 209$ acc cubo $= 836$ acc

$C(O_1, P_3) = N = 417$ acc cubo $= 1668$ acc

$C(O_1, P_4) = N + 400 = 817$ acc cubo $= 3268$ acc

$C(O_1, P) = 0'4 \cdot 4 + 0'2999 \cdot 836 + 0'3 \cdot 1668 + 0'001 \cdot 3268 = 755'232$ acc

$CO = K_1$ con area desbordada, mismo cubo que O_1 $\quad N(O_2) = N(O_1)$

$C(O_2, P_1) = 1 \text{ acceso} = 4$ ó $\log_2 (417+1)$? $\times 1$

$C(O_2, P_2) = \log_2 (417+1) = 9$ acc cubo $= 36$ acc

$C(O_2, P_3) = C(O_1, P_3) = 1668$ acc

$C(O_2, P_4) = C(O_1, P_4) = 3268$ acc

$C(O_2, P) = 0'4 \cdot 4 + 0'2999 \cdot 36 + 0'3 \cdot 1668 + 0'001 \cdot 3268 = 516$ acc

$N' = r' / T_c = \frac{0'11 \cdot 10.000}{24} = 46$ cubos

$C(O_3, P_1) = 2$ accesos cubo $= 8$ acc ← encuentro el cubo → leo el cubo, si: lleno lo recorro secuencial / hay hueco lo meto en el cubo ⇒ 2 acc. a blq

$C(O_3, P_2) = \frac{N + N' + 1}{2} = \frac{1 + 550 + 46}{2} = 298'5$ acc cubo $= 1192$ acc

$C(O_3, P_3) = 1 + N' = 47$ acc cubo $= 188$ acc

$C(O_3, P_4) = N + N' + 400 = 996$ acc cubo $= 3984$ acc

$C(O_3, P) = \ldots = 420'59$

b.) Arbol B sobre K1 (tamaño fijo 9B)

$T_{entrada} = \underbrace{9}_{\text{clave}} + \underbrace{4}_{\text{puntero}} = 13$ B

$\underbrace{m \cdot 4}_{\text{p. hijo \cdot punt. hijo}} + \underbrace{k \cdot 13}_{\text{entrada \cdot entra. Te}} \leq T_{nodo} = 1024 = 1$ bloque ← lo + pequeño. ; $m = k+1$

$4k+4$

Lo no ha inf. d. control

$17k \leq 1020$; $k = 60 \Rightarrow k_{min} = \left\lfloor \frac{60}{2} \right\rfloor = 30$ entradas.

$m = k+1 = 61 \Rightarrow m_{min} = k_{min}+1 = 31$ hijos

Buscamos que en acumulado llegue a ser nº registros. Al ser $e = r = 10.000$

nivel	# nodos	# entradas	acumulado.
1	1	1	1
2	2	$2k_{min} = 60$	61
3	$2m_{min} = 62$	$2m_{min} k_{min} = 1860$	1921
~~4~~	~~$62 m_{min} = 1922$~~	~~$1922 \cdot 30 = 57660$~~	~~59581 > registros.~~ 10.000

Nivel imposible, demasiados ⇒ nivel anterior. } Demasiados

$\boxed{n = 3}$

$T_{indice} = \frac{e}{k_{min}} = \left\lfloor \frac{10.000}{33} \right\rfloor = 333$ kB

Aproximación

10.000 entradas y 33 entradas por bloque

333 · 1 bloque

Arbol B^+ sobre k2 (tamaño variable 12 B)

$$T_{entrada} = 1B + 12B + \overset{\text{nº ocurrencia veces}}{14} \cdot 4B + \overset{\text{long. lista punt}}{1B} = 70B$$

$e = 710$ entradas

Que tienen hijos → $m \cdot \overset{\text{punteros interno}}{4} + (m-1)\overbrace{(\underset{\text{marca}}{1}+12)}^{\text{clave}} \leq 1024$

$m = \frac{1037}{17} = 61 \Rightarrow m_{min} = \lfloor \frac{m+1}{2} \rfloor = 31$

Ejemplo de imposible: $T_{entrada} = 1B + 12B + 2B + 140 \cdot 4B = 5616B$

$k \cdot 5615B + 1 + 4B \leq 1024 \Rightarrow k < 1 \Rightarrow$ Imposible

$T_{nodo} = 6$ bloques

Hojas con punt ext. → $\underbrace{k \cdot 70B}_{\text{Hojas}} + \overset{\text{1 puntero interno}}{1 \cdot 4B} \leq 1024$; $k = 14 \Rightarrow k_{min} = \lfloor \frac{k+1}{2} \rfloor = 7$

$\#\text{hojas} = \#\text{nodos}(n) = e/k_{min} = \lceil \frac{710}{7} \rceil = 101$ hojas

$\#\text{nodos}(n-1) = \lceil \frac{101}{31} \rceil = 3$ nodos

$\#\text{nodos}(n-2) = \lfloor \frac{3}{31} \rfloor \simeq 0 \Rightarrow 1$ nodo $\Rightarrow$ Raíz

$1 = n-2 \Rightarrow \boxed{n = 3}$

$T_{ind} = \#\text{hojas} + \#\text{nodos}(n-1) + \ldots + \#\text{nodos}(1) = 101 + 3 + 1 = 105$

O_1' : Serial con $\overset{\text{Arbol}}{B}$ sobre k1 y B^+ sobre k2

actualizar los 2 índices.

$C(O_1', P_1) = 1$ acc. a cubo $+ n(k_1) + n(k_2) = 4$ acc. blq $+ 3 + 3 = 10$ acc bloque.

$C(O_1', P_2) = (n-1) + 1$ acc. cubo $= (3-1) + 4$ blq acc. $= 6$ acc. a bloque

lo de media (el peor caso 1 reg por cubo)

$C(O_1', P_3) = (n-1) + \underset{\text{cubos}}{14} \cdot 4$ acc. blq $= 58$ acc. bloque.

$C(O_1', P_4) = N + 400 + \underbrace{400 \cdot (n(k_1) + n(k_2))}_{\text{Actualizar para cada borrado.}} = 847$ acc. cubo $+ 2400$ acc blq $= 5668$ acc bloque

$$\boxed{C(O_1', P) = 0'4 \cdot 10 + 0'2996 + 0'3 \cdot 58 + 0'001 \cdot 5668 = 28'86 \text{ accesos}}$$

Al ser infrecuente el borrado, aunque es el que peor le viene, en global mejora.

O_2' : O_2 que tiene $\underbrace{CO = k1}_{\text{Por eso no se usa B sobre k1}}$ e índice B^+ sobre $k2$

$C(O_2', P_1) = 1 \text{ cubo} + n(k2) = 4 + 3 = 7$ accesos.

$C(O_2', P_2) = \log_2(417+1) \overset{+1}{=} 9 \text{ acc. cubo} = 3\ 6$ accesos.

$C(O_2', P_3) = C(O_1', P_3) = 58$ accesos.

$C(O_2', P_4) = \underbrace{C(O_2, P_4)}_{\text{Borrar en tabla + borrar en índice}} + n(k2) \cdot 400 = 4468$ accesos.

$C(O_2', P) = 0'4 \cdot 7 + 0'299 \cdot 36 + 0'3 \cdot 58 + 0'001 \cdot 4468 = \boxed{35'432}$

O_3' : O_3 con índice B sobre $k1$ y CDispersión $= k2$

$C(O_3', P_1) = \underset{\text{normal}}{\overset{2\cdot 4}{2 \text{ cubo}}} + \underset{\text{actualiz}}{\overset{3}{n(k1)}} = 11$ accesos

$C(O_3', P_2) = (n-1) + 1 \text{ acc cubo} = 3 - 1 + 4 = 6$ accesos

$C(O_3', P_3) = C(O_3, P_3) = 1 + N' = 47 \text{ acc cubo} = 188$ accesos.

$C(O_4', P_4) = N + N' + 400 + 400\, n(k1) = 996 \text{ acc. cubo} + 1200 =$

$= 5184$ accesos.

$C(O_4', P) = 0'4 \cdot 11 + 0'299 \cdot 6 + 0'3 \cdot 188 + 0'001 \cdot 5184 = \boxed{67'78}$