

Problema de Ficheros – Estructuras Auxiliares

El almacenaje de sistema domótico tiene bloques de 2 KB y dispone de memoria intermedia para usos básicos (no se permite bloquear ninguna página en memoria intermedia). Se desea desarrollar una aplicación que requerirá un archivo volátil que va a recoger 15000 registros de media, organizados de forma serial no consecutiva en cubos de 8 KB. La ocupación útil de estos registros es 392.4 caracteres, y su volumen medio es 407.5 bytes. El proceso más probable es la selección de todos los valores de cierta clave a partir de una condición sobre otra clave (P_4), que ocurre el 40% de las veces. El resto de los procesos a los que se somete son equiprobables (20% cada uno), y tienen la siguiente descripción:

- P1. Selección de un registro por clave de identificación (K_1 , de tamaño fijo 9 bytes).
- P2. Selección de todos los registros por clave de búsqueda no identificativa (clave K_2 , de tamaño variable con 10 bytes de tamaño medio, y definida sobre un dominio que presenta 250 valores distintos repartidos uniformemente en el archivo).
- P3. Listado de todos los registros.
- P4. Obtener todos los valores de K_1 dado un valor de K_2

Para cada clave de búsqueda (K_1 y K_2) se requiere comparar el rendimiento de un índice **secuencial** no consecutivo con el 30% de espacio libre distribuido, con un índice en **árbol B⁺**. Especifíquese también el tamaño auxiliar empleado en cada caso.

Nota: el impacto de los procesos de reorganización local deberá comentarse, pero no se incluirá en el cálculo de los costes.

SOLUCIÓN

Org. Base:

$$T_c = 2048 / 309.5 = 6$$

$$12000/6 = 2000 \text{ cubos} \rightarrow 2 \text{ bytes parte alta puntero} + 2 \text{ bytes parte baja} = 4 \text{ B}$$

Entradas:

Para K_1 : $T_{entrada} = 11 + 4 = 15 \text{ B}$

Para K_2 : $12000/250 = 48 \text{ reg/valor} \rightarrow T_{entrada} = 1 + 10 + 1 + 4 * 48 = 204 \text{ B}$

Índices Secuenciales no consec. con ELD=40% $\rightarrow 2048 * 0.6 = 1228$ B/cubo

Para K_1 :

$$T_c = 1228 / 15 = 81 \text{ ent/cubo}$$

$$T_{ind} = 12000 / 81 = 149 \text{ cubos}$$

Para K_2 : $12000/250 = 48 \text{ reg/valor}$

$$T_c = 1228 / 204 = 6 \text{ ent/cubo}$$

$$T_{ind} = 250 / 6 = 42 \text{ cubos}$$

Índices en árbol B+ -- puntero interno 1 byte

Para K_1 :

$$\text{Hojas: } k * 15 \text{ B} + 1 \leq 2048 \rightarrow k = 136 \text{ y } k_{min} = 68$$

$$\text{No hoja: } m * 1 + (m-1) * 11 \leq 2048 \rightarrow m = 171 \text{ y } m_{min} = 86$$

$$\text{nodos}(n) = 12000 / 68 = 176 \text{ hojas}$$

$$\text{nodos}(n-1) = 176 / 86 = 2 \text{ nodos}$$

$$\text{nodos}(n-2) = 2 / 86 = 1 \text{ nodo (raíz)} \rightarrow n = 3 \text{ niveles}$$

Para K2:

Hojas: $k \cdot 204B+1 \leq 2048 \rightarrow k=10$ y $k_{min}=5$

No hoja: $m \cdot 1 + (m-1) \cdot (10+1) \leq 2048 \rightarrow m=171$ y $m_{min}=86$

nodos(n)=250/5=50 hojas

nodos(n-1)= 50/86 = 1 nodo (raíz) $\rightarrow n=2$ niveles

Comparativa de Costes (en número de accesos):

	proceso 1	proceso 2	P3	proceso 4
K1sec & K2sec	$\log_2(149+1)+1=9$	$42+1/2+48=69.5$	-	$21.5+149=170.5$
K1 B+ & K2sec	$3+1=4$	69.5	-	$21.5+176=197$
K1sec & K2 B+	9	$2+48=50$	-	$2+149=151$
K1 B+ & K2 B+	4	50	-	$2+176=178$

	COSTE GLOBAL
K1sec & K2sec	$(9 \cdot 20 + 69.5 \cdot 20 + 170.5 \cdot 40) / 80 = \mathbf{104,875}$ accesos
K1 B+ & K2sec	$(4 \cdot 20 + 69.5 \cdot 20 + 197 \cdot 40) / 80 = \mathbf{116,875}$ accesos
K1sec & K2 B+	$(9 \cdot 20 + 50 \cdot 20 + 151 \cdot 40) / 80 = \mathbf{90.25}$ accesos
K1 B+ & K2 B+	$(4 \cdot 20 + 50 \cdot 20 + 178 \cdot 40) / 80 = \mathbf{102.5}$ accesos

Resultado: K1 secuencial y K2 en B+

Su tamaño es para **K1 149 nodos** (298 KB) y para **K2 50+1=51 nodos** (102 KB).

Observación:

Si no se hace acceso invertido se ahorra en este caso, porque el número de resultados es inferior al número de bloques del segundo índice (K1). El proceso P4 tendría entonces el mismo coste que el proceso P2, y el coste global sería:

	COSTE GLOBAL
K1sec & K2sec	$(76 \cdot 20 + 69.5 \cdot 60) / 80 = \mathbf{71,125}$ accesos
K1 B+ & K2sec	$(4 \cdot 20 + 69.5 \cdot 60) / 80 = \mathbf{53,125}$ accesos
K1sec & K2 B+	$(76 \cdot 20 + 50 \cdot 60) / 80 = \mathbf{56.5}$ accesos
K1 B+ & K2 B+	$(4 \cdot 20 + 50 \cdot 60) / 80 = \mathbf{38.5}$ accesos

Las estructuras elegidas serían las mismas, pero el coste baja notablemente

2 - Problema de Ficheros – Estructuras Auxiliares – Curso 2006/07

El almacenaje de sistema domótico tiene bloques de 2 KB y dispone de memoria intermedia para usos básicos (no se permite bloquear ninguna página en memoria intermedia). Se desea desarrollar una aplicación que requerirá un archivo volátil que va a recoger 25000 registros de media, organizados de forma serial no consecutiva en cubos de 4 KB. La ocupación útil de estos registros es 223.8 caracteres, y su volumen medio es 247.5 bytes. El proceso más probable es la selección de todos los valores de cierta clave a partir de una condición sobre otra clave (P_4), que ocurre el 40% de las veces. El resto de los procesos a los que se somete son equiprobables (20% cada uno), y tienen la siguiente descripción:

- P1. Selección de un registro por clave de identificación (K_1 , de tamaño fijo 9 bytes).
- P2. Selección de todos los registros por clave de búsqueda no identificativa (clave K_2 , de tamaño variable con 10 bytes de tamaño medio, y definida sobre un dominio que presenta 250 valores distintos repartidos uniformemente en el archivo).
- P3. Listado de todos los registros.
- P4. Obtener todos los valores de K_1 dado un valor de K_2

Para cada clave de búsqueda (K_1 y K_2) se requiere comparar el rendimiento de un índice **secuencial** no consecutivo con el 25% de espacio libre distribuido, con un índice en **árbol B^+** . Especifíquese también el tamaño auxiliar empleado en cada caso.

Nota: el impacto de los procesos de reorganización local deberá comentarse, pero no se incluirá en el cálculo de los costes.

3 - Problema de Ficheros – Estructuras Auxiliares – Curso 2006/07

El almacenaje de sistema domótico tiene bloques de 2 KB y dispone de memoria intermedia para usos básicos (no se permite bloquear ninguna página en memoria intermedia). Se desea desarrollar una aplicación que requerirá un archivo volátil que va a recoger 12000 registros de media, organizados de forma serial no consecutiva en cubos de 8 KB. La ocupación útil de estos registros es 293.2 caracteres, y su volumen medio es 309.5 bytes. El proceso más probable es la selección de todos los valores de cierta clave a partir de una condición sobre otra clave (P_4), que ocurre el 40% de las veces. El resto de los procesos a los que se somete son equiprobables (20% cada uno), y tienen la siguiente descripción:

- P1. Selección de un registro por clave de identificación (K_1 , de tamaño fijo 11 bytes).
- P2. Selección de todos los registros por clave de búsqueda no identificativa (clave K_2 , de tamaño variable con 10 bytes de tamaño medio, y definida sobre un dominio que presenta 250 valores distintos repartidos uniformemente en el archivo).
- P3. Listado de todos los registros.
- P4. Obtener todos los valores de K_1 dado un valor de K_2

Para cada clave de búsqueda (K_1 y K_2) se requiere comparar el rendimiento de un índice **secuencial** no consecutivo con el 40% de espacio libre distribuido, con un índice en **árbol B^+** . Especifíquese también el tamaño auxiliar empleado en cada caso.

Nota: el impacto de los procesos de reorganización local deberá comentarse, pero no se incluirá en el cálculo de los costes.