

1. - ENUNCIADO

Se tiene una tabla (*clientes*) con $5 \cdot 10^6$ filas que ocupan de media 290 B (reales). El gestor precisa que cada cubo tenga una cabecera con información de control (200 B). El tablespace en el que se almacena nuestra tabla tiene cubos de 4 bloques de espacio cada uno, siendo el tamaño de bloque de disco 2 KB (cada cubo tiene 8 KB). Este tablespace tiene definido un 10% de espacio libre distribuido para modificaciones (pctfree) y no tiene definida ocupación mínima. Se tienen tres procesos críticos:

- **SELECT** * FROM *CLIENTES*;
- **DELETE** * FROM *CLIENTES* WHERE **DNI= VALOR**;
- **SELECT** * FROM *CLIENTES* WHERE **CIUDAD= VALOR**;

Debe tenerse en cuenta que la clave DNI (cadena de caracteres no nula, tamaño fijo 9B) está marcada como **UNIQUE**, y que la consulta **SELECT DISTINCT CIUDAD FROM CLIENTES**; devuelve $2.5 \cdot 10^4$ filas (*ciudad* es también varchar2, tamaño variable 20B). Existen otros procesos (inserciones y modificaciones) que no serán tenidos en cuenta en la optimización física, pero que hacen que la cardinalidad de la tabla se mantenga constante (siempre tiene más o menos la misma cantidad de filas). Las frecuencias relativas de los tres procesos descritos son $f = \{0.3, 0.4, 0.3\}$. Se ha ensayado una función de dispersión por la clave DNI, que sobre $N=2^{18}$ produce una tasa de desbordamientos del 1% y sobre $N=2^{19}$ arroja una tasa de desbordamientos del 10^{-3} %.

Se proponen dos organizaciones físicas, a saber:

- O₁: **Direccionada sobre CD=DNI**, con el espacio de direccionamiento que se estime oportuno, y gestión de desbordamientos a elegir entre 'área independiente serial' y 'área independiente encadenada a cubo' (extensiones).
- O₂: **Serial (no consecutiva)**.

Se pide:

- Comparar el coste global (en accesos) de las organizaciones candidatas y decidir cuál de ellas es la más conveniente. Justificar la respuesta, y comentar las ventajas e inconvenientes de cada una de ellas. Como argumentos pueden utilizarse los costes, densidades, degeneración, etc.
- Se pueden añadir índices B, B⁺ y/o BITMAP para mejorar el rendimiento (el gestor usa punteros de tamaño 7 B, tanto para ptros. internos como externos. Propónganse índices que podrían ser útiles en alguna de las dos organizaciones y calcúlense sus parámetros
- Propónganse nuevos diseños físicos sobre las organizaciones base anteriores (O₁ y O₂), con la inclusión de uno o más índices de los calculados en el apartado (b) y calcúlense los nuevos costes globales.
- Discútase la solución más apropiada de entre las estudiadas y por qué, justificando también el por qué deben descartarse otras alternativas. Propónganse otras posibilidades de mejora del diseño físico si esto es posible, o justifíquese por qué no puede mejorarse si este es el caso.

a.) $T_r = 290B$

$$T_c = \frac{(8192 - 200) 0'9}{290} = 24 \text{ regs/wbo}$$

$$N = \frac{5 \cdot 10^6}{24} = 208334 \text{ wbo}$$

$$C(O_2, P_1) = N = 208334 \text{ acc. wbo} = 833\,336 \text{ acc. blq.}$$

$$C(O_2, P_2) = \frac{N+1}{2} + 1 = 416\,674 \text{ acc. blq.}$$

$$C(O_2, P_3) = N = 833\,336 \text{ acc. blq.}$$

$$C(O_2, P) = 666\,672 \text{ acc.}$$

O_1 con $N = 2^{18}$ y gest. desd. encadenamiento a cubo.

$$r' = 50.000 \Rightarrow N' = 50.000$$

$$1\% \Rightarrow 0'01$$

↳ cada reg. pertenece a un cubo.

$$C(O_1, P_1) = N + N' = 312\,144 \text{ acc. cubo} = 1\,248\,576 \text{ acc. blq.}$$

1% . \rightarrow bloque en Area desbordamiento

$$C(O_1, P_2) = 1 + 0'01 \cdot 1 + 1 = 2'01 \text{ acc. cubo} = 8 \text{ acc. blq.}$$

$$C(O_1, P_3) = C(O_1, P_2) = 1\,248\,576 \text{ acc. blq.}$$

$$C(O_1, P) = 0'6 \cdot 1\,248\,576 + 0'4 \cdot 8 = 749\,148'8 \text{ acc. blq.}$$

No se
aumentan
desbordados

$$d_0 = \frac{4\,950\,000}{262\,144 \cdot 24} = \frac{r - r'}{N \cdot T_c} = 78'7 \%$$

b.) puntero = 7B

Arbol B sobre DNI

$e = 5\,000\,000$

→ Etiqueta identificativa.

$$m \cdot 7 + k(9+7) \leq 2048; \quad m = k+1$$

$$7k + 7 + 16k \leq 2048; \quad k = 88 \quad k_{\min} = \frac{88}{2} = 44$$

$$m_{\min} = \frac{(88+1)+1}{2} = 45$$

Nivel	nodos	entradas	e. acumuladas
1	1	1	1
2	2	$2 \cdot 44 = 88$	$88 + 1 = 89$
3	$2 \cdot 45 = 90$	$2 \cdot 45 \cdot 44 = 3960$	4049
4	4050	178 200	182 249
5	182 250	8 019 000	8 102 249 > 5 000 000

$n_1 = 4$ niveles.

$$T_{\text{indice}} = \frac{e}{k_{\min}} = 113\,636 \text{ nodos ó bloques.}$$

Arbol B^+ sobre CIUDAD $c = 2.5 \cdot 10^4 \frac{\text{req}}{\text{valor}} = \frac{5000000}{2.5 \cdot 10^4} = 200 \text{ req/val}$

$$m \cdot 7 + (m-1)(1+20) \leq 2048$$

$$m = 73 \quad m_{\min} = \frac{73+1}{2} = 37$$

$$7 + k(1+20 + 200 \cdot 7) \leq 2048$$

$$k = 1 \quad k_{\min} = \frac{1+1}{2} = 1$$

$$\# \text{ nodos } (n) = \frac{2.5 \cdot 10^4}{1} = 25000$$

$$\# \text{ nodos } (n-1) = \frac{2.5 \cdot 10^4}{37} = 675$$

$$\# \text{ nodos } (n-2) = \frac{675}{37} = 18$$

$$\# \text{ nodos } (n-3) = \frac{18}{37} < 2 \rightarrow \text{raiz}$$

$$n-3 = 1 ; n = 4$$

$$n_2 = 4 \text{ niveles} \quad T_{\text{ind}} = 25000 + 675 + 18 + 1 = 25694 \text{ bloques.}$$

CD = DNI y B^+ para Ciudad.

$$C(O'_1, P_1) = N + N' = 312144 \text{ acc. wbo} = 1248576 \text{ acc. blq.}$$

$$C(O'_1, P_2) = 1 + 0'01 \cdot 1 + 1 = 2'01 \text{ acc. wbo} = 8 \text{ acc. blq.}$$

$$C(O'_1, P_3) = (n_2 - 1) + 200 \cdot 4 = 803 \text{ acc bloque.}$$

$$C(O'_1, P) = 0'6 \cdot 1248576 + 0'4 \cdot 803 = 371816'9 \text{ acc. blq.}$$

Serial e índices B sobre DNI y B⁺ sobre Ciudad.

$$C(O'_2, P_1) = N = 208\,334 \text{ acc. wbo} = 833\,336 \text{ acc. blq.}$$

$$C(O'_2, P_2) = (\underbrace{n_1}_{\text{nivel}} - 1) + \underbrace{1 \cdot 4}_{\text{el prefijo}} + \underbrace{1 \cdot 4}_{\text{L}} = 11 \text{ accesos bloques.}$$

$$C(O_2, P_3) = (n_2 - 1) + 200 \cdot 4 = 803 \text{ acc. blq.}$$

$$C(O'_2, P) = 250\,246 \cdot 4 \text{ acc. blq.}$$

c.) O₃: Secuencial por CO: Ciudad e índice B sobre DNI

$$N = 208\,334$$

$$C(O'_3, P_1) = C(O_2, P_1) = N = 208\,334 \text{ acc. blq.}$$

$$C(O'_3, P_2) = C(O'_2, P_1) = (n_1 - 1) + 1 \cdot 4 + 1 \cdot 4 = 11 \text{ acc. blq.}$$

$$\begin{aligned} C(O'_3, P_3) &= \log_2(x+1) + \left\lceil \frac{k+1}{T_c} \right\rceil = \\ &= \log_2(25001) + \left\lceil \frac{201}{24} \right\rceil = 24 \text{ acc. blq.} \end{aligned}$$

$$C(O'_3, P) = 250\,012 \cdot 4 \text{ acc blq.}$$

La diferencia entre O₃' y O₂' no merece la pena en coste global, pero el P₃ mejora 30 veces.