

Problema 1 - Temas 2, 4, 5, 6.

Se está planteando un sistema de control de tráfico en una zona de acceso restringido. El sistema de almacenamiento necesario tendrá un volumen muy elevado de accesos sobre el fichero de 'vehículos' (aprox, unos $5 \cdot 10^5$ registros de 168 bytes reales cada uno). Las claves de búsqueda (selección) en este registro son K_1 , K_2 , y K_3 , de 12, 17, 15 bytes de media cada una, respectivamente. Se tendrá un soporte direccionado cuyo tamaño de bloque es de 2KB. Se precisa tomar una decisión acerca de la organización más adecuada con el objetivo de optimizar el rendimiento en número de accesos a soporte.

En esta decisión se quiere tener en cuenta el conjunto de procesos $P = \{P_1, P_2, P_3\}$, cuyas frecuencias relativas son $f(P) = \{60\%, 20\%, 20\%\}$, y su descripción:

- P_1 : Modificación de registros: selección por clave de identificación K_1 ; las claves K_1 y K_2 no alteran su contenido (pero la clave K_3 si).
- P_2 : Consulta por clave no unívoca K_2 (de media 25 registros por valor).
- P_3 : Selección por clave no unívoca K_3 (625 valores distintos).

El rendimiento del resto de los procesos no es crítico, y por ello no serán tenidos en cuenta para la decisión final, aunque deberá señalarse como inconveniente la degeneración de la organización, en su caso. Se dispone de una transformación sobre $N=65.000$ para K_1 que arroja una tasa de desbordamientos del 0.1%. Las alternativas que se manejan son todas organizaciones no consecutivas, con $E_c=1$ y espacio libre distribuido para modificaciones (PCTFREE=10%). La tasa de modificación es $T_M=30\%$ diaria, pero con el espacio libre distribuido, todos los registros modificados caben en el cubo donde estaban. Las propuestas son:

- O_1 : serial no consecutiva.
- O_2 : direccionada con $CD = K_1$, y gestión de desbordamientos serial.
- O_3 : secuencial no consecutiva con clave de ordenación física K_3 .

Se pide:

a) Hallar el Coste Global inicial de cada organización (en número de accesos) y la densidad real de cada organización, sabiendo que la densidad ideal es del 92%.

	O_1 : serial nc.	O_2 : dir. K_1	O_3 : sec. nc. K_3
<i>densidad real</i>			
$C(O_i, P_1)$			
$C(O_i, P_2)$			
$C(O_i, P_3)$			
$C(O_i, P)$			

b) Calcular el **periodo de reorganización máximo en O_3** para que su densidad no sea peor que la de O_2 . ¿Cuál es el coste de O_3 justo antes de la reorganización? $C'(O_3, P) =$ ¿Qué organización de las tres te parece más recomendable? Justifica tu respuesta.

c) Se opta por implantar la organización O_2 mejorada con la introducción de algunos índices para los procesos menos eficientes. Compara el rendimiento de un **índice serial** (no consecutivo) con el de otro en **árbol B^+** ($T_{\text{ptro_int}}=2B$) para cada clave involucrada.
¿Qué índice/s se decide incluir? (sobre qué clave/s y qué tipo/s de índice)
Calcula el coste Global de O_4 (O_2 más este/os índice/s)

d) Se observa que el proceso P_3 tiene unas características peculiares: la condición se establece sobre K_3 , y el proceso devuelve el valor de K_2 para ese registro (el resto de campos no son relevantes). Se opta por implementar este proceso mediante un acceso invertido, para lo que se precisa contar con sendos índices invertidos sobre estas claves.
¿Modificarías alguna de las decisiones tomadas en el apartado (c)?
¿Cómo queda la organización definitiva O_5 ? (org. base + orgs. auxiliares)
¿Cuál es el coste global de la nueva organización O_5 ?