

Relación de problemas 1

1. Sea una relación con $n=10^6$ tuplas, $B=4KB$, $R=2050B$ y bloqueo fijo. Calcula el factor de bloqueo así como el desperdicio y el porcentaje de utilización de los bloques.
2. Sea una relación con $n=10^6$ tuplas, $B=4KB$, $R=120B$, $P=6B$ y $V=8B$. Calcula el número de bloques necesarios para almacenar los datos organizados mediante un archivo secuencial indexado en caso de tratarse de:
 - a) un índice denso
 - b) un índice no denso
3. Indica las ventajas e inconvenientes de tener registros de longitud variable y razona las respuestas.
4. Indica las ventajas y/o los inconvenientes de claves de búsquedas duplicadas a la hora de montar un índice, y razona las respuestas.
5. Indica cuándo crees que es más adecuado usar el bloqueo partido:
 - a) para registros de gran tamaño
 - b) para registros de tamaño pequeño
 - c) para bloques de más tamaño
 - d) para bloques de tamaño pequeño
 - e) una relación entre tamaño de registro y tamaño de bloque
6. Indica por qué mejoran las consultas mediante los índices:
 - a) el número de bloques del índice es menor
 - b) las claves están ordenadas por valor de clave en el índice
 - c) si son suficientemente pequeños están en memoria
7. Supón una tabla con nombre `_paciente`, que es un char (56) que ocupa 56B, una fecha que ocupa 10B, un peso de tipo real que ocupa 8B, un número `_intervenciones` que es un entero y ocupa 4B, un número `_hijos` que es un entero y ocupa 4B, un atributo fumador que es lógico y ocupa 1B, y un R de 83B. Calcula el factor de bloqueo y el porcentaje de utilización en caso de tratarse de bloqueo fijo en los casos de:
 - a) bloque de 2KB
 - b) bloque de 4KB
8. Se tienen registros con un nombre que es un varchar (29), una dirección que es un varchar (255), una fecha que ocupa 10B, un valor para sexo que es un lógico y ocupa 1B, y un tamaño de bloque $B=4KB$. Calcula el factor de bloqueo y el porcentaje de utilización en caso de tratarse de bloqueo fijo. Si el bloque contiene 10B de cabecera y un directorio de entradas en el bloque.

9. Se tienen registros con: char (215), integer -2B-, fecha -10B-, real -8B-, R=235B, B=4KB. Supuesta la estructura de longitud variable, una cabecera con 2 punteros -de 4B- más un carácter, calcula el factor de bloqueo para:
- bloqueo fijo
 - bloqueo encadenado
10. Supongamos una relación con 10^6 tuplas, un tamaño de bloque B=4KB, un factor de bloqueo Bfr=10, V=10B y P=8B. Calcula el tiempo de búsqueda T_F en un índice denso primario si estuviese en memoria. Y supuesto que no cabe en memoria y se monta un índice de segundo nivel, calcula el espacio adicional ocupado.
11. Supongamos Bfr=30 del fichero de datos, y Bfr_i=100 del índice. Sean n los registros de datos. Indica cuántos bloques se necesitan para:
- un índice denso
 - un índice no denso
- El anterior cálculo se realiza suponiendo una ocupación del 100% en cada bloque. Vuelve a calcular los casos a) y b) suponiendo que los bloques se ocupan inicialmente al 80%.
12. Sea Bfr=30 y Bfr_i=100, y sea n el número de registros. Montar tantos niveles de índices como sea necesario hasta llegar a un índice de un único bloque.

$$\text{Solución propuesta: } x = \frac{1}{2} \log_{10} \frac{n}{30}$$

13. Supón el siguiente esquema: dpto (d#, nombre, extension, dir) y prof (NRP, nombre, categoría, dpto)

Indique la organización que favorecería más la siguiente consulta:

SELECT depto, categoria, count(*) FROM prof GROUP BY depto, categoria;

- un índice denso sobre dpto de prof
 - un índice no denso sobre dpto de prof
 - un índice denso sobre dpto y categoría de prof
 - un hashing sobre dpto y categoría de prof
 - un índice denso sobre d# de dpto
 - ninguno
14. Suponiendo que se han definido como claves primarias d# y NRP, revisa la consulta anterior y justifica la respuesta.
15. Sea R(A, B, C, D) el esquema de una relación. A continuación, se presentan tres planes lógicos de una consulta:

- $\sigma_{A=a \wedge B=b}(R)$
- $\sigma_{A=a}(\sigma_{B=b}(R))$
- $\sigma_{B=b}(\sigma_{A=a}(R))$

Indica cuál de los tres escogería un optimizador para ejecutar y justifica tu respuesta.

16. Para los planes lógicos del ejercicio anterior y si se dispone de las siguientes estadísticas sobre los atributos A, B y C: $T(R)=1000$, $V(R,A)=75$, $V(R,B)=20$, $V(R,C)=80$, indica cuál de ellos escogería un optimizador para ejecutar la consulta. Justifica tu respuesta.

17. Indica las diferencias entre plan lógico y plan físico.

18. Suponiendo que se tiene la relación prof (NRP, nombre, categoría, dpto) y se tienen dos índices I_1 (sobre NRP) e I_2 (sobre categoría), y dada la consulta:

```
SELECT dpto, categoría, NRP FROM prof WHERE categoría='AS1';
```

Indica:

- a) algún plan lógico
- b) un plan de ejecución que podría generar el optimizador de consultas