



PRACTICO Nº 1

REVISIÓN, PROCESAMIENTO Y OPTIMIZACIÓN DE CONSULTAS

- 1) Indexación: Considere un disco con un tamaño de bloque de $B = 512$ bytes. Un puntero de bloque tiene una longitud de $P = 6$ bytes, y un puntero de registro tiene una longitud de $P_R = 7$ bytes. Un fichero tiene $r = 30.000$ registros EMPLEADO de longitud fija. Cada registro tiene los siguientes campos: Nombre (30 bytes), Dni (9 bytes), CodDpto (9 bytes), Direcc (40 bytes), Tlf (9 bytes), FechaNac (8 bytes), Sexo (1 byte), CodTrabajo (4 bytes), Sueldo (4 bytes, número real). Se utiliza un byte adicional como marcador de eliminación.
 - a. Calcule el tamaño de registro R en bytes.
 - b. Calcule el factor de bloqueo fb_l y el número de bloques de fichero b , asumiendo una organización no extendida.
 - c. Suponga que el fichero está ordenado por el campo clave Dni y que queremos construir un *índice principal* sobre Dni. Calcule (i) el factor de bloqueo de índice fb_l ; (ii) el número de entradas de índice de primer nivel y el número de bloques de índice de primer nivel; (iii) el número de niveles necesarios si lo convertimos en un índice multinivel; (iv) el número total de bloques que el índice multinivel necesita; y (v) el número de accesos a bloques que se necesitan para buscar y recuperar un registro del fichero (dado el valor de su Dni) utilizando el índice principal.
 - d. Suponga que el fichero no está ordenado por el campo clave Dni y queremos construir un índice secundario sobre Dni. Repita el ejercicio anterior (parte c) para el índice secundario y compárelo con el índice principal.
 - e. Suponga que el fichero no está ordenado por el campo no clave CodDpto y queremos construir un *índice secundario* sobre CodDpto, utilizando la opción 3 de la Sección 14.1.3, con un nivel extra de indirección que almacena punteros de registro. Asuma que hay 1.000 valores distintos de CodDpto y que los registros EMPLEADO están uniformemente distribuidos entre estos valores. Calcule (i) el factor de bloqueo de índice fb_l ; (ii) el número de bloques que el nivel de indirección que almacena los punteros de registro necesita; (iii) el número de entradas de índice de primer nivel y el número de bloques de índice de primer nivel; (iv) el número de niveles necesarios si lo convertimos en un índice multinivel; (v) el número total de bloques que el índice multinivel necesita y los bloques utilizados en el nivel extra de indirección; y (vi) el número aproximado de accesos a bloques necesarios para buscar y recuperar todos los registros del fichero que tienen un valor CodDpto determinado, utilizando el índice.
 - f. Suponga que el fichero está ordenado por el campo no clave CodDpto y queremos construir un *índice agrupado* sobre CodDpto que utiliza anclas de bloque (cada valor nuevo de CodDpto empieza al principio de un bloque nuevo). Asuma que hay 1.000 valores distintos de CodDpto y que los registros EMPLEADO están uniformemente distribuidos entre estos valores. Calcule (i) el factor de bloqueo de índice fb_l ; (ii) el número de entradas de índice de primer nivel y el número total de bloques de índice de primer nivel; (iii) el número de niveles necesarios si lo convertimos en un índice multinivel; (iv) el número total de bloques que el índice multinivel necesita; y (v) el número de accesos a bloques que se necesitan para buscar y recuperar todos los registros del fichero que tienen un valor concreto de CodDpto, utilizando el índice agrupado (asuma que varios bloques en un grupo son contiguos).
- 2) Porque no hay que obligar a los usuarios a que elijan explícitamente una estrategia de procesamiento de la consulta. Hay casos en que es deseable que los usuarios sepan el costo de las distintas estrategias posibles? Razone la respuesta.

Dadas las siguientes relaciones:



trabajador (**dni**, nombre, dirección, *IdLocalidad*, telefono, valorHora)
piloto (**dniPiloto**, nroLicencia, horasVuelo, fechaInicio)
avion (**nroAvion**, *tipoModelo*, año, horasVuelo)
modeloAvion (**tipoModelo**, descripcion, capacidad)
trabajadorReparacion(**dniTrabajador**, **nroAvion**, fechaInicioReparacion,
fechaFinReparacion, *tipoFallaReparada*)
fallas (**tipoFalla**, descripcion)
pilotoAvion(**dniPiloto**, **nroAvion**) /* aviones que vuela cada piloto. */
pilotoModelo(**dniPiloto**, **modelo**) /* modelos de avion que puede pilotear un piloto. */
localidad (**idLocalidad**, nombreLocalidad)

3) Para las siguientes consultas SQL:

- I. Escriba la expresión correspondiente en álgebra relacional
- II. construya el árbol canónico de la expresión lograda

- a)

```
SELECT nroAvion
FROM avion A, modeloAvion MA
WHERE MA.capacidad > 100 and A.tipoModelo=MA.tipoModelo
```
- b)

```
SELECT *
FROM trabajador T, trabajadorReparacion TR
WHERE T.valorHora > 200 and T.dni = TR.dniTrabajador
UNION ALL
SELECT *
FROM trabajador T, trabajadorReparacion TR
WHERE idLocalidad = (SELECT idLocalidad
                     FROM localidad
                     WHERE nombreLocalidad = 'Trelew')
and T.dni = TR.dniTrabajador
```
- c)

```
SELECT descripción
FROM (SELECT *
      FROM fallas F, trabajadorReparacion TR
      WHERE fechaInicioReparacion > '10/12/2011'
      and F.tipoFalla = TR.tipoFallaReparada)
```
- d)

```
SELECT T.nombre
FROM trabajador T, piloto P, pilotoAvion PA, avion A
WHERE P.dniPiloto = PA.dniPiloto and P.horasVuelo > 5000 and PA.nroAvion
= A.nroAvion and A.horasVuelo < 20000 and
P.dniPiloto = T.dni and T.valorHora < 150
```
- e)

```
SELECT *
FROM trabajadorReparacion
WHERE fechafinReparacion > '31/01/2012' and
    tipoFallaReparada IN (SELECT tipoFalla FROM fallas
                        WHERE descripción = 'Flap oscilante')
```
- f)

```
SELECT *
FROM avion A, avion B
WHERE A.horasVuelo > B.horasVuelo and B.año=2000
```

4) Mencione las estadísticas que almacenan los DBMS para cada tabla de una base de datos



5) Mencione las etapas de un algoritmo de optimización heurística y aplicarlas en los árboles del ejercicio 3.

6) Dada la relación R(A, B, C) y la consulta

```
select *  
from R  
where A > 10 and B=10 and C=5;
```

Asumir que:

nr=10.000, br=1000 (cantidad de bloques de r)

V(A, R) = 90, V(B, R) = 150, V(C, R) = 150

Suponer que existe un índice de agrupamiento para A (profundidad de 4) niveles) e índices de secundarios para B (3 niveles) y C (3 niveles).

¿Cuáles de las siguientes opciones requieren la lectura de la mínima cantidad de bloques para la consulta?

- a) leer la relación R íntegramente a memoria.
- b) realizar primero la $A > 10$.
- u) realizar primero la $B = 10$.
- d) realizar primero $C = 5$.

Razone el costo de las siguientes consultas

- e) SELECT DISTINCT B FROM R
- f) SELECT * FROM R WHERE $A = 10$

7) Estimación del tamaño del resultado de las operaciones (sección 14.2.2 y 14.2.3 del libro Fundamentos de Bases de Datos - Silberschatz y Korth - 4 edición)

Sean las siguientes tres relaciones: R(a, b, c), S(b, c, d) Y T(d, e)

Asumir que:

nr=10.000, ns=20.000, nt=30.000

V(a, R)=100, V(b, R)=200, V(c, R)=200

V(b, S)=400, V(c, S)=500, V(d, S)=200

V(d, T)=100, V(e, T)=200

Un bloque puede alojar 50 tuplas de cualquier relación

Tamaño de atributos: a=5 bytes b=2 bytes c = 3 bytes d=3 bytes e = 8 bytes

Estimar el tamaño de la relación resultante de aplicar las siguientes operaciones

- a) SELECT * FROM T WHERE $d = 800$
- b) SELECT * FROM S WHERE $c \geq 500$
- c) SELECT * FROM R WHERE $a = 10$ and $b = 58$
- d) Estimar el tamaño de la reunión natural $R \bowtie S \bowtie T$



LABORATORIO Nº 1

PL/PGSQL, PROCESAMIENTO Y OPTIMIZACIÓN DE CONSULTAS

Pautas: Se deberá entregar un informe con la resolución de los ejercicios

Sobre la base de datos tp1-aviones y para el motor de base de datos Postgres (9.0 o superior) hacer lo siguiente

- 1 Agregar tuplas en forma masiva a las tablas modeloAvion y Avion de forma tal que por cada modeloAvion insertado se agreguen 10000 aviones y que la estadística $V(\text{capacidad}, \text{modeloAvion}) \geq 1000$. Además por cada avión agregado agregar una reparación del mismo (con algún DNI de trabajador ya preexistente). Se debe presentar el script de la solución.
- 2 Hacer una selección de las reparaciones para aviones cuya capacidad es = X (tomar para X un valor válido de los que se encuentren en su base de datos)
- 3 Analizar el plan de ejecución de la consulta del punto anterior. Documentar en su informe
- 4 Podría cambiar en algo el esquema físico de la base de datos de forma tal que se consiga algún plan de ejecución que acelere la consulta
- 5 Implementar el cambio propuesto en el punto 4
- 6 Volver a analizar la consulta del punto 2, viendo si el plan de ejecución cambió. Documentar el nuevo plan logrado y comparar con el anterior explicando los cambios encontrados.