

# BASES DE DATOS II

## Solución Actividad Ejercicios Diseño Físico

Dados los siguientes esquemas: EMP (eid, enombre, sal, edad, did, ... (y 50 atributos más))  
DEPT (did, planta, presupuesto, mgr\_eid, ... (y 50 atributos más))

donde mgr\_eid es el identificador del empleado que es jefe del departamento. Los salarios oscilan entre 10.000 y 100.000, las edades entre 20 y 80, cada departamento tiene una media de 5 empleados, hay 10 plantas y los presupuestos varían entre 10.000 y 1.000.000. Puede asumirse que hay una distribución uniforme de valores en todos los casos.

Para cada una de las siguientes consultas:

- 1) escribir la sentencia SQL correspondiente
- 2) ¿cuál de las opciones indicadas elegirías para mejorar su rendimiento? Explica brevemente el funcionamiento del SGBD con esta elección
- 3) Si el SGBD soporta planes *solo-índice* ¿cómo cambiaría la respuesta? Explica brevemente el funcionamiento del SGBD con esta elección

CONSULTA 1) *Nombre, edad y salario de todos los empleados*

- a. Índice hash agrupado sobre los campos (*enombre, edad, sal*) de EMP
- b. Índice hash no agrupado sobre los campos (*enombre, edad, sal*) de EMP
- c. Índice B+ agrupado sobre los campos (*enombre, edad, sal*) de EMP
- d. Índice hash no agrupado sobre los campos (*enombre, edad*) de EMP
- e. No crearía índice

1) *select enombre, edad, sal from emp*

2) (e) ya que, en este caso, es necesario acceder a todos los registros de EMP, por lo que mantener un índice no ayuda en nada.

3) (b) porque no interesa un índice agrupado (no se accede a la tabla), y el índice debe ser por los tres campos de salida

CONSULTA 2) *Identificadores (did) de los departamentos que se encuentran en la planta 10 y tienen un presupuesto inferior a 15.000*

- a. Índice hash agrupado sobre el campo (*planta*) de DEPT
- b. Índice hash no agrupado sobre el campo (*planta*) de DEPT
- c. Índice B+ agrupado sobre los campos (*planta, presupuesto*) de DEPT
- d. Índice B+ agrupado sobre el campo (*presupuesto*) de DEPT
- e. No crearía índice

1) *select did from dept where (planta = 10) and (presupuesto < 15000)*

2) (c) porque los registros deben ordenarse por planta y presupuesto. Así, cuando se ejecuta la consulta, se recupera el primer registro con planta = 10 y se leen los registros consecutivos del mismo bloque.

3) (c) por el mismo motivo anterior. La opción solo-índice no se aplica en este caso porque no hay ningún índice que incluya did.

# BASES DE DATOS II

CONSULTA 3) *Nombres de los empleados que dirigen algún departamento (no necesariamente al que pertenecen) y tienen un salario mayor de 12.000*

- a. Índice B+ agrupado sobre el campo (*sal*) de EMP
- b. Índice hash agrupado en el campo (*mgr\_eid*) de DEPT
- c. Índice hash no agrupado en el campo (*mgr\_eid*) de DEPT
- d. Índice hash no agrupado en el campo (*eid*) de EMP
- e. Índice B+ agrupado en el campo (*sal*) de EMP y un índice hash agrupado en el campo (*mgr\_eid*) de DEPT

1) *select distinct enombre from emp, dept where (emp.eid=dept.mgr\_eid) and (sal > 12000)*

2) (d) porque así se puede hacer un recorrido secuencial de la tabla DEPT y para cada tupla aplicar la función hash sobre el atributo mgr\_eid, traer la tupla correspondiente de EMP (utilizando el índice creado) y comprobar si el salario de ese eid es mayor de 12000.

No tiene sentido hacer un índice sobre sal, ya que es muy poco restrictivo (la gran mayoría de los empleados tendrán un salario superior a 12000, al considerar una distribución uniforme de valores).

3) (d) dado que la opción solo-índice no se aplica en este caso porque no hay ningún índice que incluya enombre.

CONSULTA 4) *Salario medio de cada departamento*

- a. Índice B+ agrupado disperso sobre el campo (*did*) de EMP
- b. Índice B+ agrupado denso sobre el campo (*did*) de EMP
- c. Índice B+ agrupado sobre los campos (*did, sal*) de EMP
- d. Índice hash no agrupado sobre los campos (*did, sal*) de EMP
- e. No crearía índice

1) *select avg(sal) from emp group by did*

2) (a) porque así el índice será tan eficiente como un índice sobre (*did, sal*), ya que en ambos casos se accede a los registros de datos en el orden de *did*. Sin embargo, eligiendo esta opción, como el índice solo tiene un atributo su mantenimiento es menos costoso.

Es mejor opción que la (b) porque un índice disperso es más pequeño que uno denso.

3) (c) porque de este modo basta con acceder al índice (un índice no agrupado sería suficiente, pero no está incluido en la lista de opciones que se presenta)