

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 50 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 ¿cuál de las opciones indicadas elegirías para mejorar su rendimiento sabiendo que el SGBD soporta planes solo-índice? (Solo se puede elegir 1 opción).

Justifica por qué se descartan cada una de las restantes opciones planteadas.

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

SELECT enombre, edad, sal FROM EMP

- a. <EMP.enombre> B+ agrupado
- b. <EMP.enombre, EMP.edad, EMP.sal> B+ no agrupado
- c. <EMP.enombre, EMP.edad, EMP.sal> B+ agrupado
- d. No crearía índice

RESPUESTA:

(b) El índice debe ser por los tres campos de salida para disponer de toda la información y que no sea necesario acceder a la tabla.

Se descarta la opción (a) porque no incluye todos los atributos de la sentencia SELECT y, por lo tanto, supone la necesidad de acceder a la tabla. No ofrece ninguna ventaja.

Se descarta la opción (c) porque no tiene sentido agrupar, dado que con el índice (b) no es necesario acceder a la tabla.

Se descarta la opción (d) porque el índice (b) evita la necesidad de traer la tabla a memoria, lo que acelera el proceso.

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

SELECT did FROM DEPT WHERE (planta = 10) AND (presupuesto < 15000)

- a. <DEPT.planta> B+ agrupado
- b. <DEPT.presupuesto> B+ agrupado
- c. <DEPT.planta, DEPT.presupuesto> B+ agrupado
- d. <DEPT.presupuesto, DEPT.planta> B+ agrupado

RESPUESTA:

(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 el menor presupuesto y se leen consecutivamente las páginas de datos hasta acceder a un departamento con presupuesto igual o superior a 15000. La opción solo-índice no se aplica en este caso porque no hay ningún índice que incluya did.

Se descarta la opción (a) porque no incluye todos los atributos de la parte WHERE. Las entradas de datos solo incluyen información de la planta, por lo que resulta más complejo localizar los departamentos con presupuesto inferior a 15000.

Se descarta la opción (b) porque no incluye todos los atributos de la parte WHERE. Las entradas de datos solo incluyen información del presupuesto, por lo que resulta más complejo localizar los departamentos que se encuentran en la planta 10.

Se descarta la opción (d) porque de este modo las entradas de datos en el índice están ordenadas, en primer lugar, por presupuesto y, dentro de cada presupuesto, por planta. Esto dificulta el filtrado de los departamentos que se encuentran en una planta concreta.

Solución Actividad Ejercicios Diseño Físico

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

SELECT DISTINCT enombre FROM EMP, DEPT WHERE (emp.eid=dept.mgr_eid) AND (sal > 12000)

- a. <EMP.sal> B+ agrupado
- b. <DEPT.mgr_eid> B+ agrupado
- c. <DEPT.mgr_eid> B+ no agrupado
- d. <EMP.eid> B+ no agrupado
- e. <EMP.sal> B+ agrupado y <DEPT.mgr_eid> B+ agrupado

RESPUESTA:

(d) porque así se puede hacer un recorrido secuencial de la tabla DEPT y seguir los siguientes pasos:

- para cada tupla de DEPT se obtiene el valor del atributo *mgr_eid*
 - se localiza el valor *eid* en el índice creado para EMP
 - se recupera el registro de EMP correspondiente al valor *eid* y se comprueba si su salario es mayor de 12000.
- La opción solo-índice no se aplica en este caso porque el índice no incluye *enombre*.

Se descartan las opciones (a) y (e) porque no tiene sentido hacer un índice en EMP 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).

Se descartan las opciones (b) y (c) porque la tabla DEPT es mucho más pequeña que EMP (cada departamento tiene una media de 50 empleados), por lo que es mejor elegir una opción que ayude en el recorrido de la tabla EMP.

CONSULTA 4) *Salario medio de cada departamento*

SELECT AVG(sal) FROM EMP GROUP BY did

- a. <EMP.did> B+ agrupado
- b. <EMP.sal> B+ no agrupado
- c. <EMP.did, EMP.sal> B+ agrupado
- d. <EMP.did, EMP.sal> B+ no agrupado
- e. No crearía índice

RESPUESTA:

(d) porque de este modo basta con acceder al índice, recorrer todas las entradas de datos, sumar los salarios que se encuentran en ellas y dividir esta suma por el número de entradas.

Se descarta la opción (c) porque no tiene sentido agrupar la tabla, dado que no se va a acceder a ella, pues con la opción (d) se dispone de toda la información necesaria en el índice.

Se descarta la opción (a) porque, aunque este índice permite ordenar los empleados por departamento, no se dispone de información del salario, por lo que hace necesario acceder al total de la tabla para hacer el cálculo.

Se descarta la opción (b) porque este índice no separa los empleados por departamento, por lo que sería necesario acceder igualmente al total de la tabla, ordenar los registros de datos por código de departamento y calcular las medias posteriormente. Se trata de un proceso totalmente ineficiente, incluso más caro que traer directamente la tabla a memoria sin empleo de índice.

Se descarta la opción (e) porque se propone un índice (opción (d)) que evita tener que traer la tabla de datos a memoria, por lo que resulta mucho más eficiente que el recorrido de una tabla aleatoria.