

Examen Bases de Datos

2do Semestre 2009

Prof. Claudio Gutiérrez, Aux. Violeta Chang

Pregunta 1 Un desarrollador encontró estas tablas y especificaciones en la documentación de su cliente. Para darse una idea de qué se trata necesita hacer ingeniería reversa. Se pide reconstruir el modelo entidad relación que generó estas relaciones y restricciones.

$T_1(A, B, C)$, llave prim. $\{A\}$, llave ext. C ref. T_2

$T_2(N, D, L)$, llave prim. $\{N\}$

$T_3(Id, Na)$, llave prim. $\{Id, Na\}$, llave ext. Id ref. T_1

$T_4(Nt, Nm, Dc, Tf)$, llave prim. $\{Nt\}$.

$T_5(Id, Ns, Di)$, llave prim. $\{Id\}$

$T_6(Id, Ir, Nt, Fs, Fd)$, llave prim. $\{Id, Ir, Nt\}$, llaves ext.: Id ref. T_1 , Ir ref. T_5 , Nt ref. T_4 .

Respuesta 1 diagrama aqui.

Pregunta 2 Demuestre que cualquier esquema de relación con dos atributos está en FNBC (Forma Normal de Boyce-Codd). ¿Es cierto esto para una relación con tres atributos? Demuestre o dé un contraejemplo.

Respuesta 2 Sea $R(A, B)$ el esquema. Las únicas df no triviales son $A \rightarrow B$ y $B \rightarrow A$, o ambas. Si es $A \rightarrow B$, entonces $\{A\}$ es llave, y luego esta en FNBC. Idem para $B \rightarrow A$. Si son ambas son df, el argumento es similar.

Para tres atributos es falso. $R(A, B, C)$ y df. $B \rightarrow C$. Entonces la llave es $\{A, B\}$ y no esta en FNBC.

Pregunta 3 La asociación de Fútbol cuenta con la siguiente información sobre jugadores, equipos y torneo. Cada tupla de la tabla **Equipo** guarda el nombre del equipo, el jugador y las fechas de ingreso y fin de su permanencia del jugador en el Equipo. **Torneo** indica los partidos y sus fechas. Todas las fechas indican años.

```
Jugador(idj, nombre, nacionalidad)
Equipo(NombreEq, idj, fechaIn, fechaFin)
Torneo(nombreEq1, nombreEq2, fecha)
```

(1) Se quiere averiguar cuáles son los equipos que violaron durante 2009 la condición de que en cada partido no se puede incluir más de 4 extranjeros.

(2) Todos los jugadores que han pertenecido al menos a tres equipos distintos. Puede suponer que no hay NULLs y no debe incluir duplicados en su resultado.

Respuesta 3 (1)

```
SELECT NombreEq
FROM Jugador AS J, Equipo AS E, Torneo as T
WHERE      J.idj = E.idj
          AND NombreEq = NombreEq1
          AND fecha = 2009
          AND fechaIn =< 2009
          AND fechaFin >= 2009
          AND NOT (nacionalidad = Chile)
GROUP BY nombreEq1, nombreEq2, fecha
HAVING COUNT(*) >=4;
```

UNION

```
SELECT NombreEq
FROM Jugador AS J, Equipo AS E, Torneo AS T
WHERE      J.idj = E.idj
```

```

        AND NombreEq = NombreEq2
        AND fecha = 2009
        AND fechaIn =< 2009
        AND fechaFin >= 2009
        AND NOT (nacionalidad = Chile)
GROUP BY nombreEq1, nombreEq2, fecha
HAVING COUNT(DISTINCT NombreEq) >=4;

```

(2)

```

SELECT DISTINCT nombre
FROM Jugador as J, Equipo as E
WHERE      J.idj = E.idj
GROUP BY idj
HAVING COUNT(*) >= 3

```

(Puede cambiar nombre por idj en el SELECT)

Pregunta 4 Considere la relación $R(A, B, C, D, E)$ que contiene 5.000.000 de registros, donde cada página de la relación tiene 10 registros. Suponga que $R.A$ es llave para R , con valores entre 0 y 4.999.999 uniformemente distribuidos, y que el archivo R se almacena ordenado sobre $R.A$. Para cada una de las siguientes consultas relacionales, indique cual de los siguientes tres enfoques es el más barato (y argumente su respuesta):

1. Accesar el archivo ordenado de R directamente.
2. Usar un índice B+ agrupado sobre el atributo $R.A$.
3. Usar un índice de hash sobre el atributo $R.A$.

Consultas:

- (a) $\sigma_{a < 50.000}(R)$
- (b) $\sigma_{a \neq 50.000}(R)$

Respuesta 4 (a) Es una consulta de rango, y recupera menos del 10 % de las tuplas. Luego (2) índice B+ agrupado sobre A.

COSTO (en accesos): 4 + 500

Alternativa usando el hecho que la consulta necesita los primeros 50.000 y la tabla está ordenada por A y A es clave: *Scan* desde el comienzo de la tabla hasta 50.000. Costo: 500.

(b) Selecciona casi toda la tabla. Claramente (1) acceder el archivo ordenado R directamente.

(NOTA: *No* sirve el indice de hash, pues se esta preguntando por los elementos diferentes de 50.000.

COSTO: 500.000.

Pregunta 5 Considere el siguiente plan:

$R_2(Z); R_2(Y); R_1(X); R_1(Y); W_1(X); R_3(Y); W_3(Y); R_2(X); W_2(Z);$

(a) Determine si este plan es o no serializable por conflicto. Si lo es, indique el plan serial equivalente. Si no lo es, indique porqué.

(b) Suponga que después de la última operación del plan, agregamos $C_3; C_2; C_1;$. Indique si este plan completo es estricto o no. Argumente su respuesta.

Respuesta 5 (a) Si lo es. El grafo de serializabilidad es acíclico.

El plan serial equivalente es:

$R_1(X); R_1(Y); W_1(X); R_2(Z); R_2(Y); R_2(X); W_2(Z); R_3(Y); W_3(Y);$

(b) (Un plan P es estricto si toda transacción sólo lee o escribe elementos X tal que la última transacción que escribió X ha sido confirmada o abortada.)

El plan no es estricto. La tx 2 lee X ($R_x(X)$) de lo que escribió la tx. 1 ($W_1(X)$), y en ese momento la tx. 1 no estaba confirmada.