

Segundo Parcial de Fundamentos de Base de Datos

Diciembre 2015

Presentar la resolución del parcial:

- Duración: 3 horas
- Con las hojas numeradas y escritas de un solo lado.
- Con la cantidad de hojas entregadas en la primer hoja.
- Con cédula de identidad y nombre en cada hoja.
- **Escrita a lápiz y en forma prolija.**

Ejercicio 1 (8 puntos)

Sea el esquema relación $R(A,B,C,D,E,G)$ y el conjunto de dependencias $F = \{ A \rightarrow B, CD \rightarrow E, CD \rightarrow G, AG \rightarrow D, G \rightarrow E, BEA \rightarrow C \}$

Para cada una de las siguientes afirmaciones indique si es verdadera o falsa, **JUSTIFICANDO** la respuesta.

1. R tiene una única clave con respecto a F. **FALSO**

A no pertenece al lado derecho de las df por lo tanto pertenece a todas las claves.

Verifico si A es clave:

$A^+ = \{ A, B \} \neq R$, por lo tanto A no es superclave ni clave.

Busco claves con dos atributos:

$(AB)^+ = \{ A, B \} \neq R$

$(AC)^+ = \{ A, C, B \} \neq R$

$(AD)^+ = \{ A, D, B \} \neq R$

$(AE)^+ = \{ A, E, B, C \} \neq R$

$(AG)^+ = \{ A, G, B, E, C, D \} = R$

AG es clave

Si hay más claves entonces $R - \{ G \}$ tiene que ser superclave

$(ABCDE)^+ = \{ A, B, C, D, E, G \} = R$

ABCDE es superclave, por lo tanto contiene una clave.

Por lo tanto AG no es la única clave.

2. Todas las claves de R según F son de 3 o más atributos. **FALSO**

AG es clave (demostrado en la parte anterior) y AG no tiene 3 atr

3. G es un atributo primo. **VERDADERO**

G pertenece a la clave AG por lo que es un atributo primo.

4. Aplicando el algoritmo visto en el curso para obtener una descomposición de R en 3NF con JSP y preservación de dependencias según F se obtiene la siguiente descomposición:

$R_1(C,D,E,G), R_2(A,G,D), R_3(G,E), R_4(B,E,A,C)$. **FALSO**

Esta descomposición sería el resultado de aplicar el algoritmo visto en el curso si F es minimal pero en este caso F no es minimal ya que en la dependencia $BEA \rightarrow C$ el atributo B es redundante:

$(AE)^+ = \{ A, E, B, C \}$

Por lo tanto la df se sustituye por $EA \rightarrow C$ y el esquema R4 no se obtiene al aplicar el algoritmo visto en el curso.

Ejercicio 2 (8 puntos)

Sea el esquema relación $R(A,B,C,D,E,G,H)$ y el conjunto de dependencias

$F = \{ B \rightarrow DE, A \rightarrow GC, CD \rightarrow H, H \rightarrow A \}$ y la descomposición r de R en $R_1(H,G,C)$, $R_2(B,A,E)$, $R_3(A,C,D,G)$

Para cada una de las siguientes afirmaciones indique si es verdadera o falsa, **JUSTIFICANDO** la respuesta.

1. $C \rightarrow G \in (\prod_{R_1} F)^+$ **FALSO**

$$(X \rightarrow Y \in (\prod_{R_i} F)^+) \Leftrightarrow (X \rightarrow Y \in F^+ \wedge XY \subseteq R_i) \quad (I)$$

$C^+ = \{C\}$, $G \notin C^+$ por lo tanto $C \rightarrow G \notin F^+$. Entonces la afirmación es falsa por no cumplirse la primera parte de la definición (I).

2. $AD \rightarrow H \in (\prod_{R_3} F)^+$ **FALSO**

ADH no está incluido en R_3 , entonces no se cumple la segunda parte de la definición (I).

3. $CD \rightarrow A \notin (\prod_{R_3} F)^+$ **FALSO**

$CD^+ = \{C,D,H,A,G\}$

A pertenece a CD^+ , por lo tanto $CD \rightarrow A \in F^+$ (a)

$CDA \subseteq R_3$ (b)

De (a) y (b) se puede concluir que la afirmación es falsa.

4. R_2 está en BCNF - **FALSO**

Es necesario calcular la proyección de F en R_2 .

$B^+ = \{B,D,E\}$

$A^+ = \{A,G,C\}$

$E^+ = \{E\}$

$(BE)^+ = \{B,D,E\}$

$(AE)^+ = \{A,G,C,E\}$

$$(\prod_{R_3} F)^+ = \{B \rightarrow E\}$$

B no es superclave en R_3 ya que no determina A , por lo tanto $B \rightarrow E$ viola BCNF.

5. ρ preserva las dependencias de F . **FALSO**

R1(H,G,C),
 $H \rightarrow GC$

$C^+ = \{C\}$

$G^+ = \{G\}$

$CG^+ = \{C,G\}$

R2(B,A,E),
 $B \rightarrow E$

R3(A,C,D,G)

$CD \rightarrow A$

$A \rightarrow GC$

$A^+ = \{A,G,C\}$

$CD^+ = \{C,D,H,A,G\}$

$D^+ = \{D\}$

$DG^+ = \{D,G\}$

$$J = \cup_{R_i} (\prod_{R_i} F) = \{H \rightarrow GC, B \rightarrow E, A \rightarrow GC, CD \rightarrow A\}$$

Si la descomposición preserva las df entonces F y J deben ser equivalentes.

Sea $B \rightarrow D \in F$

$$(B^+)_J = \{B, E\}$$

$D \notin (B^+)_J$ por lo tanto $B \rightarrow D \notin J^+$, F y J no son equivalentes entonces la descomposición no preserva las df.

Ejercicio 3 (8 puntos)

Sea el esquema relación: ASIGNATURAS (codasig, nomasig, aniolectivo, cantincriptos) donde se cumplen dependencias funcionales pero no se conocen las mismas.

1. Dar un conjunto de dependencias funcionales F de forma tal que la máxima forma normal en que se encuentra ASIGNATURAS sea 1NF. Justifique su respuesta.

$$F = \{ \text{codasig} \rightarrow \text{nomasig}, \text{codasig}, \text{aniolectivo} \rightarrow \text{cantincriptos} \}$$

Clave de ASIGNATURAS según F : codasig,aniolectivo

codasig,aniolectivo no pertenecen a los lados derechos de las df por lo tanto pertenecen a todas las claves.

$$(\text{codasig}, \text{aniolectivo})^+ = \{ \text{codasig}, \text{aniolectivo}, \text{nomasig}, \text{cantincriptos} \}$$

Por lo tanto codasig,aniolectivo es la única clave.

La dependencia $\text{codasig}, \text{aniolectivo} \rightarrow \text{nomasig}$ $\hat{=}$ \hat{F}^+ y es una dependencia parcial de clave (el atributo aniolectivo es redundante), por lo tanto ASIGNATURAS viola 2NF. Entonces la máxima forma normal en que se encuentra es 1NF.

2. Mostrar las posibles problemas que se pueden presentar en ASIGNATURAS por no cumplir 2NF.

Cada año que se tenga información de inscriptos a una asignatura es necesario repetir el nombre de la misma.

No es posible tener la información del nombre de una asignatura si aún no se ha dictado o se tienen que permitir valores nulos.

Al borrar los datos de inscriptos a todas las ediciones de una asignatura se pierde la información del nombre de la misma o es necesario permitir valores nulos.

Ejercicio 4 (8 puntos)

Sea el esquema relación R(A,B,C,D,E) y el conjunto de dependencias $F = \{A \rightarrow B, E \rightarrow D\}$

Para cada una de las siguientes afirmaciones indique si es verdadera o falsa, **JUSTIFICANDO** la respuesta.

1. Todas las descomposiciones de R en dos subesquemas son con JSP (join sin pérdida) con respecto a F.
FALSO
2. Todas las descomposiciones de R en dos subesquemas que preservan las dependencias funcionales de F son con JSP respecto a F.
FALSO

Sea la siguiente descomposición de R : R1(A,B,C), R2(A,E,D)

- En R1 se proyecta la dependencia $A \rightarrow B$ y en R2 se proyecta $E \rightarrow D$. Por lo tanto es una descomposición que preserva las dependencias funcionales.
- Aplicando el teorema de JSP visto en el curso esta descomposición tiene JSP sii $A \rightarrow BC$ o $A \rightarrow ED$ se deducen de las dependencias de F.
 $A^+ = \{A, B\}$, BC no esta incluido en A^+ y ED no esta incluido en A^+ , por lo tanto las dependencias funcionales no pertenecen a F^+
Entonces la descomposición es con JOIN con pérdida.

La descomposición de R en R1 y R2 son un contraejemplo para las dos afirmaciones.

Ejercicio 5 (8 puntos)

En una universidad se tiene información de estudiantes que realizan cursos y que a su vez realizan deportes dentro de la universidad. Los cursos y deportes son independientes entre sí. También se tiene la información de los responsables de los cursos, habiendo un único responsable en cada curso.

Para representar esta realidad se usa el esquema relación

ESTUDIANTES(idEstudiante, idDeporte, idCurso, idResponsable)

donde se cumplen las siguientes dependencias:

idCurso \rightarrow idResponsable

idEstudiante \twoheadrightarrow idDeporte

Dada la instancia r de Estudiantes:

| idEstudiante | idDeporte | idCurso | idResponsable |
|----------------|----------------|----------------|----------------|
| a ₁ | d ₁ | c ₁ | r ₁ |
| a ₁ | d ₂ | c ₂ | r ₂ |
| a ₂ | d ₃ | c ₁ | r ₁ |
| a ₂ | d ₄ | c ₄ | r ₄ |

Donde los subíndices distintos en los valores de los atributos representan valores distintos.

1. Completar la instancia r de forma tal que sea una instancia válida de **ESTUDIANTES**. Justificar la respuesta.

| | idEstudiante | idDeporte | idCurso | idResponsable |
|----|----------------|----------------|----------------|----------------|
| t1 | a ₁ | d ₁ | c ₁ | r ₁ |
| t2 | a ₁ | d ₂ | c ₂ | r ₂ |
| t3 | a ₂ | d ₃ | c ₁ | r ₁ |
| t4 | a ₂ | d ₄ | c ₄ | r ₄ |
| t5 | a ₁ | d ₁ | c ₂ | r ₂ |
| t6 | a ₁ | d ₂ | c ₁ | r ₁ |
| t7 | a ₂ | d ₃ | c ₄ | r ₄ |
| t8 | a ₂ | d ₄ | c ₁ | r ₁ |

- Las tuplas t5 y t6 se obtienen al aplicar la definición de dependencia multivaluada considerando la dependencia idEstudiante \twoheadrightarrow idDeporte y las tuplas t1,t2.
- Las tuplas t7 y t8 se obtienen al aplicar la definición de dependencia multivaluada considerando la dependencia idEstudiante \twoheadrightarrow idDeporte y las tuplas t3,t4.

2. Explicar por qué **ESTUDIANTES** no cumple 4NF.
 $\text{idEstudiante}^+ = \{\text{idEstudiante}\} \neq \text{ESTUDIANTES}$, por lo tanto idEstudiante no es superclave, entonces la dependencia $\text{idEstudiante} \twoheadrightarrow \text{idDeporte}$ viola 4NF.
3. Obtener una descomposición de **ESTUDIANTES** en 4NF aplicando el algoritmo visto en el curso. Mostrar cada uno de los pasos aplicados.

Aplicando el algoritmo visto en el curso, considerando la dependencia $\text{idEstudiante} \twoheadrightarrow \text{idDeporte}$ descomponemos ESTUDIANTES en:

EST1($\text{idEstudiante}, \text{idDeporte}$)
EST2($\text{idEstudiante}, \text{idCurso}, \text{idResponsable}$)

En EST1 se cumple la dependencia $\text{idEstudiante} \twoheadrightarrow \text{idDeporte}$, el esquema está en 4NF ya que la dependencia es trivial.

En EST2 se cumple la dependencia $\text{idCurso} \rightarrow \text{idResponsable}$, idCurso no es superclave por lo tanto la dependencia viola BCNF, por lo tanto también 4NF. Por lo tanto es necesario continuar descomponiendo este esquema.

EST21($\text{idCurso}, \text{idResponsable}$)
EST22($\text{idCurso}, \text{idEstudiante}$).

En EST21 se cumple la dependencia $\text{idCurso} \rightarrow \text{idResponsable}$, idCurso es clave por lo tanto se cumple BCNF. Y como no hay dependencias multivaluadas que no sean funcionales entonces el esquema está en 4NF.

En EST22 no se cumplen dependencias no triviales por lo tanto el esquema está en 4NF.

En resumen la descomposición de ESTUDIANTES en EST1,EST21,EST22 está en 4NF.

Ejercicio 6 (10 puntos)

El siguiente es parte de un esquema de gestión de un instituto que dicta cursos a empresas:

ESTUDIANTES (idEst, nomEst, empresa)

Representa información de los estudiantes, su identificador, nombre y empresa a la que pertenece.

CURSOS (idCurso, nomCurso, carrera, nivel)

Representa información de los cursos, su identificador, su nombre, la carrera y el nivel al que corresponde en esa carrera.

REALIZADO (idEst, idCurso, fecha, nota)

Registra los cursos aprobados por los estudiantes, guardando la fecha, así como la nota de aprobación.

Se cuenta además con la siguiente información:

| | Estudiantes | Cursos | Realizado |
|--------------------------------|--|--------|---|
| Cantidad de Tuplas | 600 | 50 | 2500 |
| Distribución de Valores | Existen 25 empresas diferentes. Los estudiantes se distribuyen uniformemente en ellas. | | Las notas van del 1 al 12. Se asume que se distribuyen uniformemente en la tabla. |

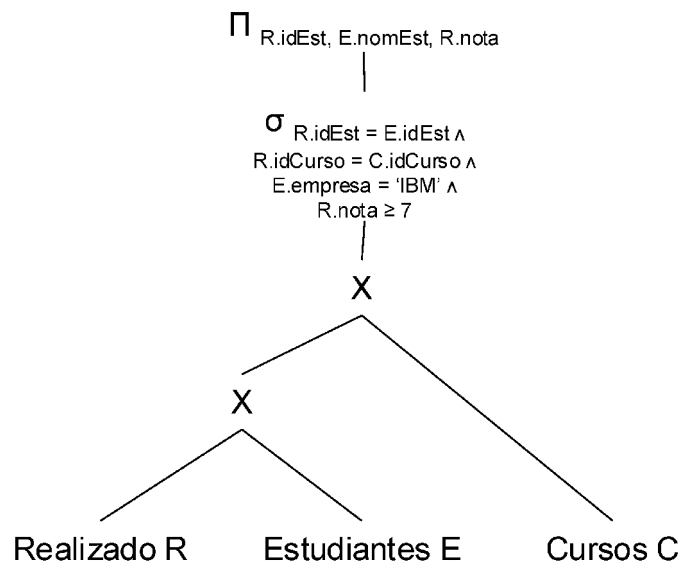
Considere la siguiente consulta sobre el esquema dado:

```

SELECT R.idEst, E.nomEst, R.nota
FROM Realizado R, Estudiantes E, Cursos C
WHERE R.idEst = E.idEst AND
      R.idCurso = C.idCurso AND
      E.empresa = 'IBM' AND
      R.nota ≥ 7
    
```

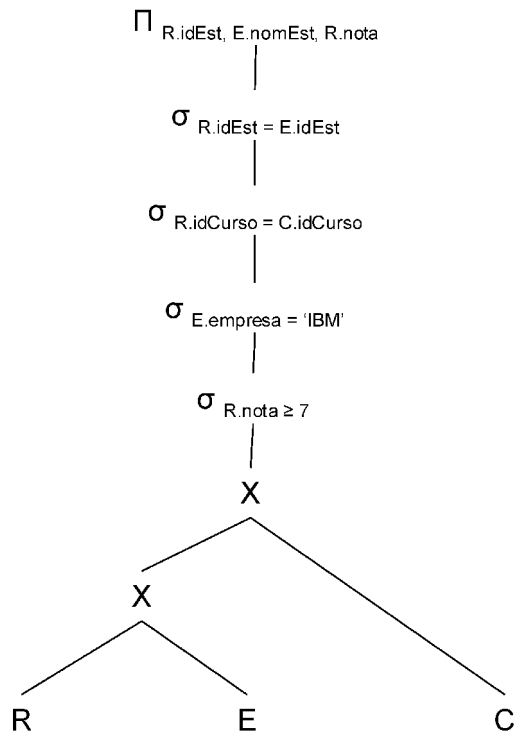
Se pide:

- a) Dar el árbol canónico del plan lógico para la consulta

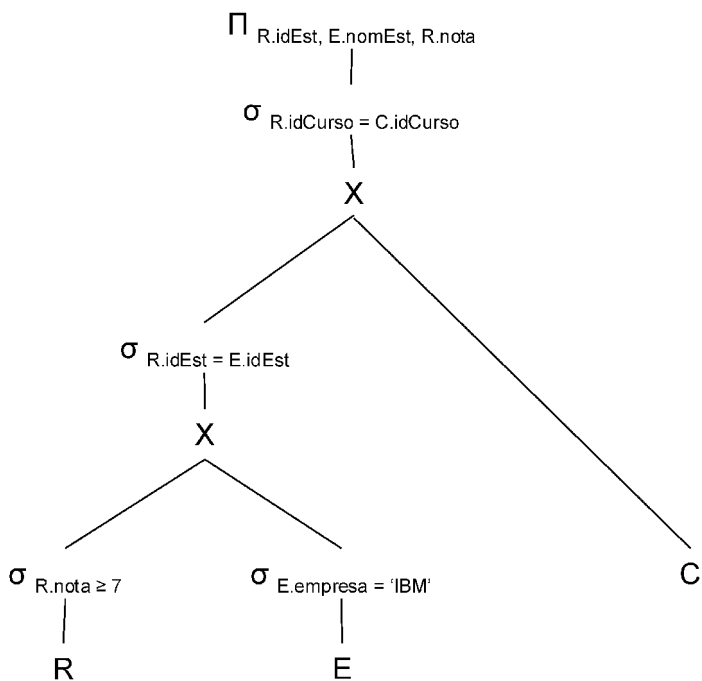


- b) Aplicar las heurísticas para optimización llegando al plan optimizado. (Explicar pasos aplicados).

Paso 1: Cascada de Selecciones



Paso 2: Bajar Selecciones lo máximo posible



Paso 3: Ordenar las hojas

Para este paso necesitamos calcular algunos tamaños:

$$|E| = 600$$

$$|\sigma_{E.empresa = 'IBM'} E| = 600 / 25 = 24$$

$$|R| = 2500$$

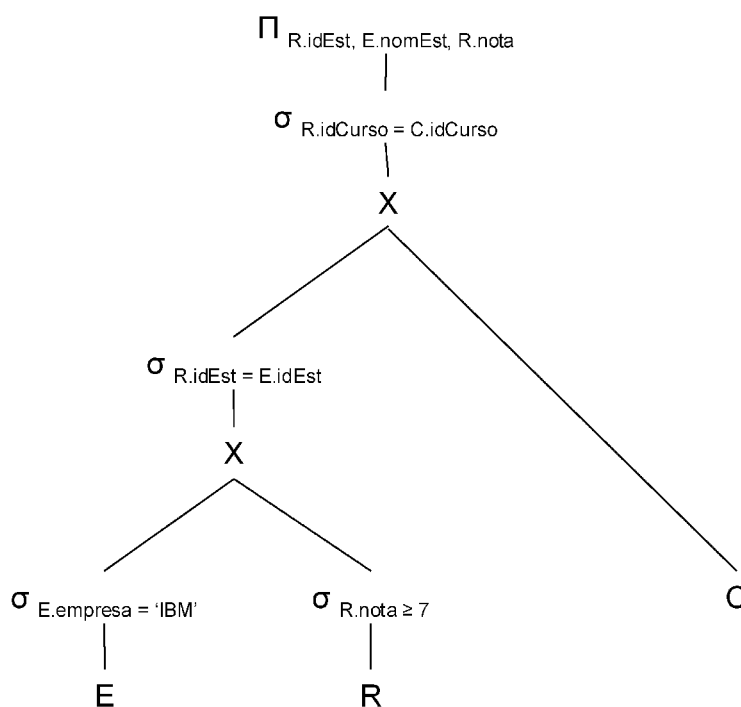
$$|\sigma_{R.nota = 7} R| = 2500 / 12$$

$$|\sigma_{R.nota \geq 7} R| = 2500 / 12 * 6 = \mathbf{1250}$$

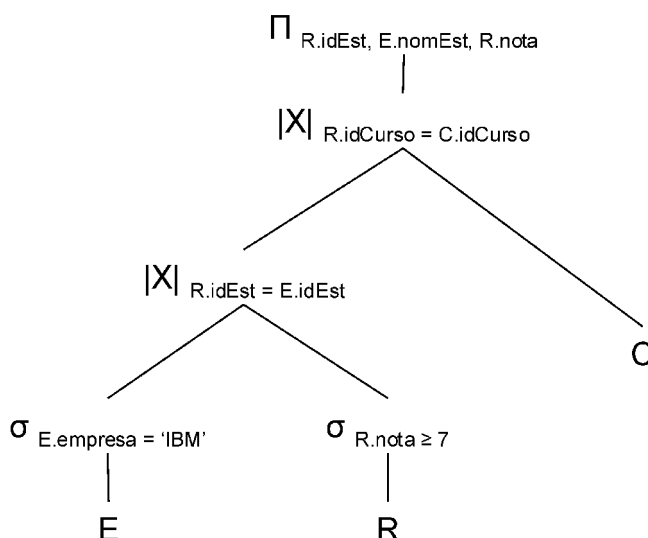
$$|C| = \mathbf{50}$$

Orden de las hojas:

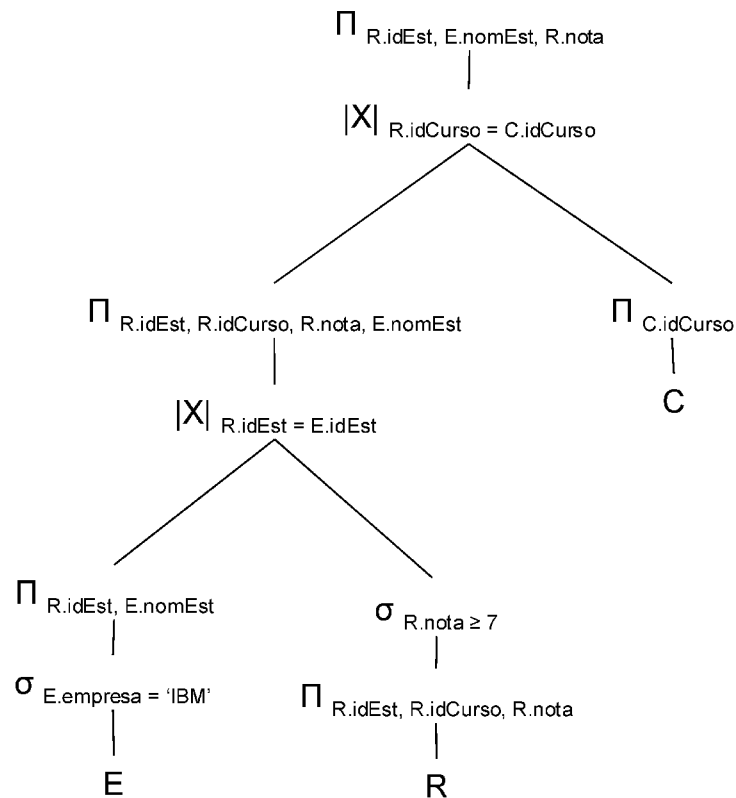
- E, C, R no es posible porque no hay condición para formar join entre E y C, y quedaría un producto cartesiano.
- E, R, C es el mejor orden, porque E es el que genera menor cantidad de tuplas y lo ponemos lo más a la izquierda posible. Luego R para que no se genere producto cartesiano, y por último C.



Paso 4: Generar los Joins



Paso 5: Bajar las Proyecciones lo máximo posible



c) Teniendo en cuenta esta consulta y que son habituales las consultas por las claves, ¿qué índices le parece útil tener para este esquema? Para cada uno, decir qué tipo de índice sería y sobre que atributos se definiría.

- Índices primarios sobre todas las claves:

Ind1 – sobre idEst en Estudiantes

Ind2 – sobre idCurso en Cursos

Ind3 – sobre idEst, idCurso, fecha en Realizado

- Índices Secundarios:

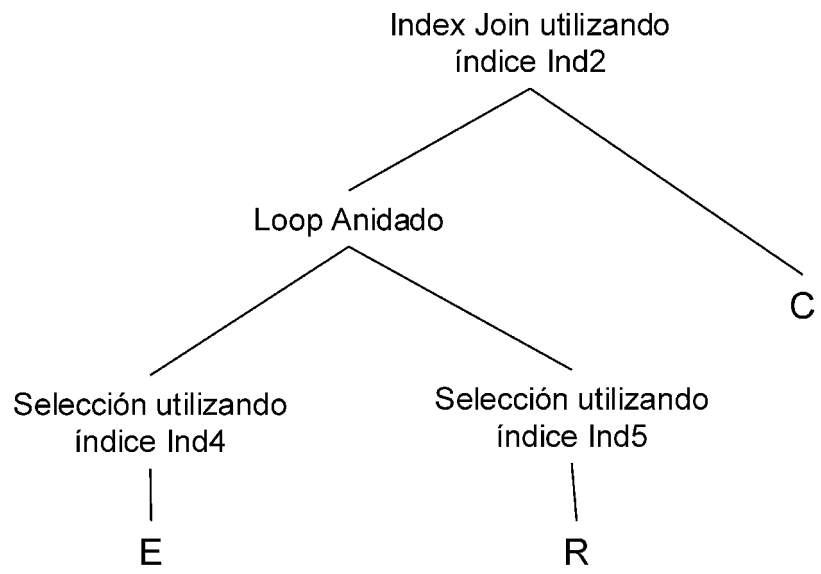
Ind4 – sobre empresa en Estudiantes (árbol B+ o Hash)

Ind5 – sobre nota en Realizado (árbol B+)

Los índices secundarios no pueden ser tipo Cluster, porque éstos necesitarían que el archivo de datos esté ordenado por ese atributo. Esto no puede ser porque deben estar ordenados por el atributo del índice primario.

El índice de nota no debería ser tipo Hash, porque es para comparación por desigualdad.

d) Considerando las respuestas de las partes a), b) y c), dar un plan físico que le parezca adecuado.



Ejercicio 7 (10 puntos)

Dadas las siguientes transacciones:

T1: r1(X) r1(Y) w1(X) w1(Y) c1

T2: w2(Y) r2(X) w2(X) c2

Parte 1)

- a) Dar una historia de T1 y T2 entrelazada, que sea recuperable, justificando.

r1(X) r1(Y) w2(Y) w1(X) r2(X) w2(X) w1(Y) c1 c2

Esta historia es recuperable porque T2 lee de T1, T1 no lee de T2, y T1 se confirma (hace commit) antes que T2.

T2 lee de T1 el elemento X (operaciones w1(X), r2(X)).

- b) ¿Qué ventaja tiene el hecho de que una historia sea recuperable?

La ventaja es que si una transacción se cancela (hace abort) en vez de confirmarse, todas las transacciones que leyeron de ella pueden cancelarse también (abortos en cascada). Esto es porque ninguna transacción se habrá confirmado antes que otra de la cual leyó.

- c) Dar una historia de T1 y T2 entrelazada, que Evite Abortos en Cascada, justificando.

w2(Y) r1(X) r2(X) w2(X) c2 r1(Y) w1(X) w1(Y) c1

Esta historia Evita Abortos en Cascada porque T1 lee de T2 luego que T2 ya hizo commit, y no hay ninguna otra lectura de una transacción hacia otra.

T1 lee de T2 el elemento Y (operaciones w2(Y), r1(Y)).

Parte 2) Ahora considere a T1 y T2 con los siguientes bloqueos y desbloqueos:

T1: r1(X) r1(X) r1(Y) r1(Y) w1(X) w1(X) w1(Y) u1(X) w1(Y) u1(Y) c1

T2: w2(Y) w2(Y) u2(Y) r2(X) r2(X) w2(X) w2(X) u2(X) c2

- a) Para cada una de las siguientes historias decir si puede suceder en el sistema o no, justificando.

H1: r1(X) r1(X) r1(Y) r1(Y) w2(Y) w2(Y) u2(Y) r2(X) r2(X) w2(X) w2(X) u2(X) c2 w1(X) w1(X) w1(Y) u1(X) w1(Y) u1(Y) c1

H1 no puede suceder en el sistema. La operación w2(Y) no puede suceder en ese lugar, ya que T1 tiene bloqueado a Y para lectura. No pueden haber a la vez un bloqueo de lectura y otro de escritura sobre el mismo objeto, de dos transacciones diferentes.

H2: w2(Y) w2(Y) u2(Y) r1(X) r1(X) r1(Y) r1(Y) w1(X) w1(X) w1(Y) u1(X) r2(X) r2(X) w2(X) w2(X) u2(X) c2 w1(Y) u1(Y) c1

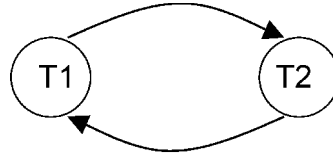
H2 puede suceder en el sistema, ya que los bloqueos y desbloqueos son respetados en la historia.

- b) Decir para cada una de las transacciones T1 y T2, si siguen el protocolo 2PL básico.
T1 sigue el protocolo 2PL básico, ya que cumple con las 2 fases requeridas: primero aumenta sus bloqueos y luego los libera.

T2 NO sigue el protocolo 2PL básico, ya que luego de la operación $u_2(Y)$ aparecen nuevas operaciones de bloqueo ($r_2(X)$ y $w_2(X)$), no respetando las 2 fases.

- c) Para cada una de las historias posibles de la parte a) decir si es serializable, justificando.

H2 es la única historia posible de la parte a). Para saber si es serializable debemos hacer el grafo de serializabilidad.



Se genera el arco T1-T2 por las operaciones en conflicto: $r_1(X), w_2(X)$, existiendo también otros pares que generan el mismo arco.

Se genera el arco T2-T1 por las operaciones en conflicto: $w_2(Y), r_1(Y)$, existiendo también otros pares que generan el mismo arco.

H2 no es serializable.