

Segundo Parcial de Fundamentos de Base de Datos

Noviembre 2012

SOLUCION

Ejercicio 1 (15 puntos)

Un sitio web desea publicar información sobre actores, las películas en las que trabajó y sus preferencias deportivas.

De los actores se conoce su nombre (anom), los deportes que le gustan ver (adep) (pueden ser más de uno), y su país de origen (apais). De las películas se conoce el nombre de la película (pnom), el nombre del director (pdir) (se asume que cada película fue dirigida por un sólo director) y el año de producción (paño). Además se conoce, para cada actor, en qué películas trabajó.

Se considera además que los deportes que le gustan a un actor son totalmente independientes de las películas en que trabajó.

El problema de diseñar la base se distribuyó entre tres personas diferentes. Cada una de ellas completó su diseño y construyó un esquema propio. Cada una de las siguientes tablas pertenece a un esquema distinto de los construidos por estas personas:

Actores(anom,pnom,adep)

ActoresPaíses(anom,pnom,adep,apais)

ActoresPelículas(anom,pnom,adep,apais,pdir,paño)

- Indique qué dependencias funcionales no triviales se cumplen en cada tabla. Justifique.
- Indique al menos 1 dependencia multivaluada (ni trivial ni funcional) que se cumpla en cada 1 de las tablas (3 en total). Justifique su respuesta.
- Indique al menos una dependencia multivaluada embebida que se cumpla en alguna de las tablas. Justifique su respuesta.
- Determine en que forma normal está el esquema ActoresPelículas. Justifique su respuesta.
- Aplique el algoritmo visto en el curso para encontrar una descomposición en 4NF de ActoresPelículas. Justifique los pasos que da en el algoritmo.

SOLUCION:

a)

Actores(anom,pnom,adep)

No tiene dependencias funcionales. Esto se debe a que para cada actor hay muchas películas y muchos deportes asociados.

ActoresPaíses(anom,pnom,adep,apais)

anom → apais. Esta dependencia se cumple porque cada actor tiene un único país de origen. Además el resto de los atributos son los mismos que los de la tabla anterior por lo que no hay dependencias entre ellos.

ActoresPelículas(anom,pnom,adep,apais,pdir,paño)

anom → apais. Los atributos son los mismos que en la tabla anterior, por lo que tiene por lo menos las mismas dependencias (esta).

$pnom \rightarrow pdir, paño$. Cada película tiene asociado un único director y un único año de producción.

Dada cualquier pareja de atributos no considerada en estas dependencias, la relación entre ellos es N:N por lo que no hay más dependencias funcionales.

b)

Actores(anom, pnom, adep)

$anom \rightarrow pnom$.

$aom \rightarrow adep$.

Estas dependencias se cumplen porque la única relación que hay entre una película y un deporte es que corresponden al mismo actor. O sea que, si al actor a_1 hizo las películas p_1 y p_2 y además le gustan los deportes d_1 y d_2 , le gustan esos deportes independientemente de las películas que hizo por lo que la tabla debe llenarse de la siguiente forma para representar correctamente esa información:

anom	Prom.	adep
a_1	p_1	d_1
A_1	p_2	d_2
A_1	p_1	d_2
A_1	p_2	d_1

ActoresPaises(anom, pnom, adep, apais)

Dado que se cumple la dependencia $anom \rightarrow apais$, también se debe cumplir (por las reglas de replicación y complemento) la dependencia $anom \rightarrow pnom, adep$.

Por otra parte $apais$, es independiente del nombre de las películas y los deportes que le gustan a ese actor.

Observar que como el valor de $apais$ es único para cada $anom$ (por la funcional), entonces se cumplen las siguientes multivaluadas:

$anom, apais \rightarrow pnom$

$anom \rightarrow pnom, apais$

$anom, apais \rightarrow adep$

$anom \rightarrow adep, apais$

y los complementos correspondientes.

Por este motivo, la tabla debe llenarse de esta forma.

anom	pnom	adep	apais
A_1	p_1	d_1	ap_1
A_1	p_2	d_2	ap_1
A_1	p_1	d_2	ap_1
A_1	p_2	d_1	ap_1

ActoresPeliculas(anom, pnom, adep, apais, pdir, paño)

Siguiendo un razonamiento análogo al de antes, se puede probar que se cumplen las siguientes multivaluadas:

$anom, apais \rightarrow pnom, pdir, paño$ (el complemento es $anom, apais \rightarrow adep$)

$anom \rightarrow pnom, pdir, paño, apais$ (el complemento es $anom \rightarrow adep$)

$anom \rightarrow adep, apais$ (el complemento es $anom \rightarrow pnom, pdir, paño$)

c)

En **ActoresPaises(anom,pnom,adep,apais)** se cumple la multivaluada embebida $anom \rightarrow pnom | adep$

dado que se cumple como multivaluada en un subesquema (ver parte a).

d)

Considerando la dependencia

$anom \rightarrow apais$

se ve que apais no aparece en ninguna otra dependencia funcional del lado izquierdo, por lo que sabemos que no está en ninguna clave (es no primo).

Si observamos las dependencias funcionales, entonces se ve que anom está nunca del lado derecho por lo que tiene que estar en todas las claves. Esto garantiza que esa dependencia induce una dependencia parcial desde cualquier clave sobre un atributo no primo, por lo que viola 2NF y por lo tanto el esquema está en 1NF.

e)

El esquema sobre el que hay que trabajar es:

ActoresPeliculas(anom,pnom,adep,apais,pdir,paño)

Las dependencias a considerar son las siguientes:

$anom \rightarrow apais$

$pnom \rightarrow pdir, paño$

$anom \rightarrow \rightarrow adep$

$anom \rightarrow pnom | adep$

La dependencia elegida en el primer paso es la detectada en la parte D y según el algoritmo se debe partir en las siguientes tablas (ya se muestran las proyecciones de las dependencias sobre los esquemas).

$AP_1(anom, apais)$ { $anom \rightarrow apais$ } en 4NF.

$AP_2(anom, pnom, adep, pdir, paño)$ { $pnom \rightarrow pdir, paño$; $anom \rightarrow \rightarrow adep$; $anom \rightarrow \rightarrow pnom | adep$ } en 1NF por $pnom \rightarrow pdir, paño$, haciendo un análisis similar al realizado en la parte d.

Por lo tanto se parte AP_2 en:

$AP_{21}(pnom, pdir, paño)$ { $pnom \rightarrow pdir, paño$ } en 4NF

$AP_{22}(anom, pnom, adep)$. Como se obtuvo el esquema de la multivaluada embebida, las dependencias a tomar en cuenta son las siguientes { $anom \rightarrow \rightarrow pnom$ }. Como anom no es clave, entonces este esquema no está en 4nf, por lo que hay que dividirlo en

$AP_{221}(anom, pnom)$ y $AP_{222}(anom, adep)$ en donde las dependencias multivaluadas consideradas son triviales.

La descomposición resultante en 4NF es la siguiente:

$AP_1(anom, apais)$ { $anom \rightarrow apais$ }

$AP_{21}(pnom, pdir, paño)$ { $pnom \rightarrow pdir, paño$ }

$AP_{221}(anom, pnom)$

$AP_{222}(anom, adep)$

Ejercicio 2 (15 puntos)

Dado el siguiente esquema relación R (A,B,C,D,E,G,H) con el siguiente conjunto de dependencias

$F = \{B \rightarrow ACD, GA \rightarrow HB, A \rightarrow D, BD \rightarrow G\}$

- Calcular todas las claves de R según F. Justificar.
- Dada la siguiente descomposición de R, $\rho=(S1,S2)$:

$S1(B,C,D,H)$

$S2(A,E,B,G,H)$

- Proyectar las dependencias funcionales sobre S1 y S2.
 - Indicar si se pierden dependencias funcionales. Justificar su respuesta.
 - Indicar si la descomposición es con join sin pérdida.
- Obtener un cubrimiento minimal de F, aplicando el algoritmo visto en el curso y mostrando cómo aplica cada paso.
 - Obtener una descomposición de R en 3NF con preservación de dependencias y join sin pérdida aplicando el algoritmo visto en el curso y mostrando cómo aplica cada paso.

Solución

- a)
E no aparece en ninguna df, por lo tanto pertenece a todas las claves.
C y H no aparecen a la izquierda en ninguna df, por lo tanto no pertenecen a ninguna clave.

$E^+ = \{E\}$

$(EA)^+ = \{E, A, D\}$

$(EB)^+ = \{E, B, A, C, D, G, H\} \Rightarrow$ **(EB) es clave**

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

$(EG)^+ = \{E, G\}$

Me fijo si hay más claves:

$\{R - (B, C, H)\} = \{E, A, D, G\}$

$(EADG)^+ = \{E, A, D, G, H, B, C\}$ es superclave

$(EAG)^+ = \{E, A, G, H, B, C, D\} \Rightarrow$ **(EAG) es clave**

$(EAD)^+ = \{E, A, D\}$

No hay más claves, ya que la única posibilidad de 4 atributos es superclave.

- b)

$S1(B,C,D,H)$

$S2(A,E,B,G,H)$

- i)

$F_{S1} = \{B \rightarrow CDH\}$

ya que:

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

No tiene más dfs porque C y H no están a la derecha de ninguna df, y D está solamente junto a B, por lo que no generan ninguna nueva df.

$$F_{S2} = \{ B \rightarrow AGH, GA \rightarrow HB \}$$

ya que:

$$B_{+F} = \{ B, A, C, D, G, H \}$$

No tiene más dfs porque E no está en ninguna df y H no está a la derecha de ninguna df, por lo que no generan ninguna nueva df. Ninguna otra combinación de atributos de S2 genera nuevas dfs.

ii)

$$K = F_{S1} \cup F_{S2}$$

$$K = \{ B \rightarrow CDH, B \rightarrow AGH, GA \rightarrow HB \}$$

$$F^+ \subseteq K^+ ?$$

Tengo que verificar que todas las dfs de F pertenecen a K+

$$F = \{ B \rightarrow ACD, GA \rightarrow HB, A \rightarrow D, BD \rightarrow G \}$$

Como

$$B_{+K} = \{ B, C, D, H, A, G \}$$

entonces $B \rightarrow AC$ y $BD \rightarrow G$ pertenecen a K+

Sin embargo,

$A_{+K} = \{A\}$ con lo cual, la dependencia $A \rightarrow D$ se perdió.

El resto de las dfs de F están en K, por lo tanto pertenecen a K+

=> Se pierde $A \rightarrow D$.

iii)

Aplicamos la propiedad de JSP vista en clase:

$D = (R1, R2)$ de R tiene JSP respecto a F sobre R sii

- la df $(R1 \cap R2) \rightarrow (R1 - R2)$ está en F+

ó - la df $(R1 \cap R2) \rightarrow (R2 - R1)$ está en F+

$$S1 \cap S2 = BH$$

$$S1 - S2 = CD$$

$$BH \rightarrow CD \in F^+ ?$$

$$(BH)^+ = \{ B, H, A, C, D, G \} \Rightarrow BH \rightarrow CD \in F^+$$

La descomposición es con JSP.

c)

Paso 1: Descomponer dfs. con más de un atributo a la derecha.

$F_1 = \{ B \rightarrow A, B \rightarrow C, B \rightarrow D, GA \rightarrow H, GA \rightarrow B, A \rightarrow D, BD \rightarrow G \}$

Paso 2: Eliminar atributos redundantes a la izquierda de las dfs.

$GA \rightarrow H$?

$G^+ = \{ G \}$

$A^+ = \{ A, D \}$

No tiene atributos redundantes.

$GA \rightarrow B$?

No tiene atributos redundantes (mirando las mismas clausuras que en la anterior).

$BD \rightarrow G$?

$B^+ = \{ B, A, C, D, \dots \} \Rightarrow D$ es redundante

$F_2 = \{ B \rightarrow A, B \rightarrow C, B \rightarrow D, GA \rightarrow H, GA \rightarrow B, A \rightarrow D, B \rightarrow G \}$

Paso 3: Eliminar dependencias funcionales redundantes.

$B \rightarrow A$ no es redundante porque A no está a la derecha de ninguna otra df

$B \rightarrow C$ ídem

$B \rightarrow D$?

$(B)^+_{F_2 - \{B \rightarrow D\}} = \{ B, A, C, D, G, H \} \Rightarrow B \rightarrow D$ es redundante

$F_3 = \{ B \rightarrow A, B \rightarrow C, GA \rightarrow H, GA \rightarrow B, A \rightarrow D, B \rightarrow G \}$

$GA \rightarrow H$ ídem

$GA \rightarrow B$ ídem

$A \rightarrow D$ ídem

$B \rightarrow G$ ídem

$F_{\text{minimal}} = \{ B \rightarrow A, B \rightarrow C, GA \rightarrow H, GA \rightarrow B, A \rightarrow D, B \rightarrow G \}$

Ejercicio 3 (15 puntos)

El siguiente esquema relacional representa la organización de los grupos de trabajo de un determinado instituto de una Facultad:

Docente (cod-doc, nom-doc, cod-grupo, horas-sem, grado)

Representa los docentes de cada grupo, su nombre, el código del grupo al cual pertenece, las horas semanales que trabaja y su grado.

Grupo (cod-grupo, nom-grupo, interno)

Representa a cada grupo, su nombre y el teléfono del interno en el instituto. El atributo nom-grupo es una clave alternativa.

Materias (cod-mat, cod-doc-resp, nom-mat, créditos)

Representa las materias dictadas, el docente responsable y los créditos de las mismas.

Sea la siguiente consulta sobre dicho esquema:

SELECT M.nom-mat, M.créditos

FROM Materias M, Docente D, Grupo G

WHERE D.cod-grupo = G.cod-grupo AND D.cod-doc = M.cod-doc-resp

AND G.nom-grupo = 'CZY' AND D.horas-sem = '40' AND D.grado = '4';

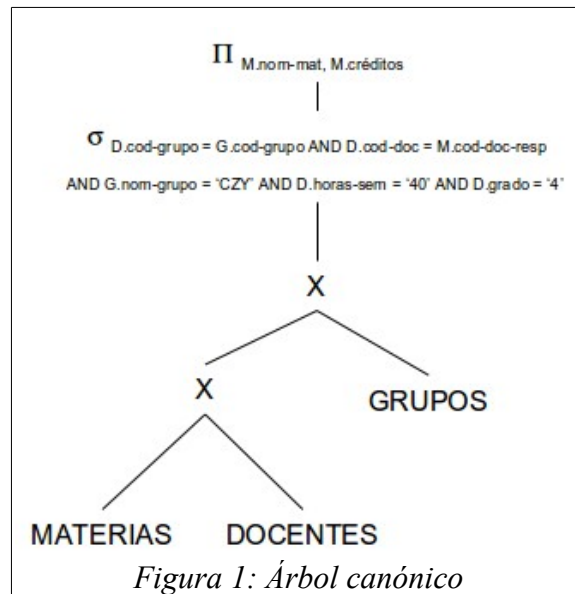
Se pide:

- Construir el árbol canónico de la consulta dada.
- Construir un plan lógico para la consulta, utilizando las heurísticas y teniendo en cuenta los tamaños.
- Construir un plan físico adecuado al plan lógico de la parte anterior.

DATOS:

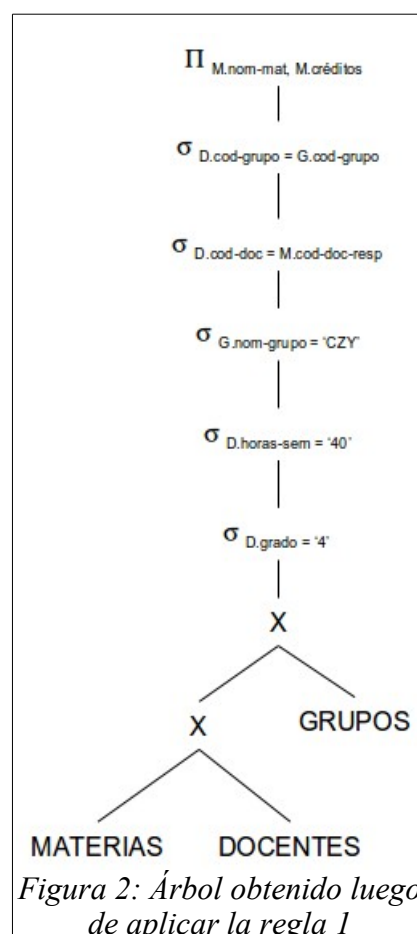
	DOCENTE	GRUPO	MATERIA	DOC >< GRUPO
Cantidad Tuplas	200	40	500	
Indices primarios (niveles: 1)		- cod-grupo	- cod-mat	
Indices secundarios (B+) (niveles: 3)	Un índice por cada uno de los siguientes atributos: - cod-grupo - horas-sem - grado	- nom-grupo	- cod-doc-resp	
Cantidad de tuplas por bloque	50	30		20
Observaciones	- grado: 5 distintos - el 90% de los docentes G4 tienen 40 horas - Se asume que hay distribución uniforme en todos los casos.			

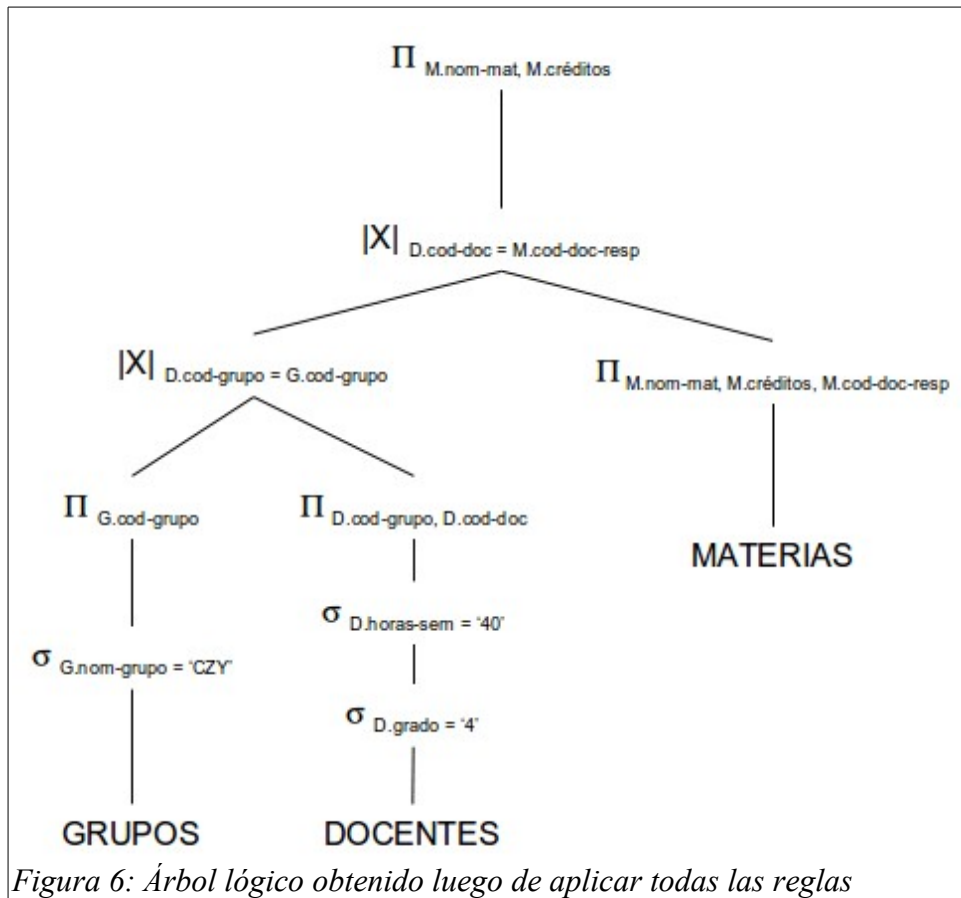
SOLUCIÓN parte a:



SOLUCIÓN parte b:

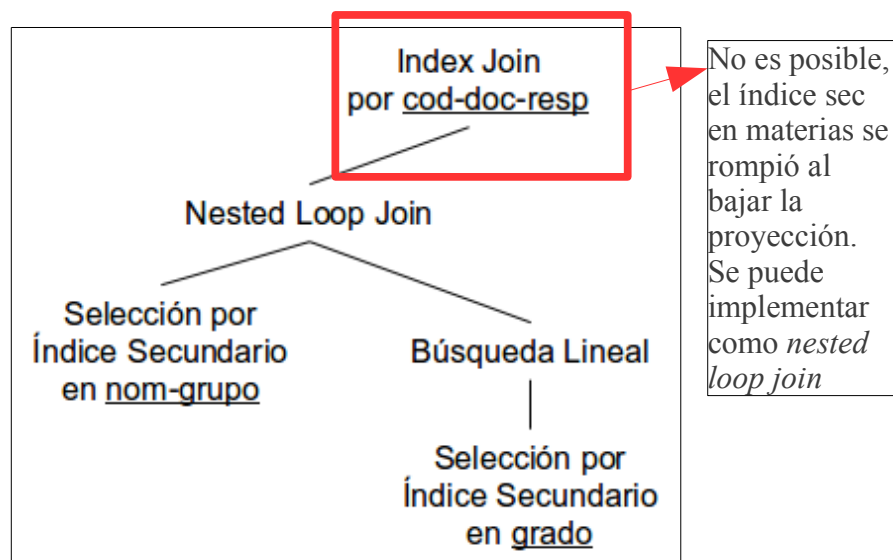
A partir del árbol canónico de la Figura 1 y aplicando la primer regla de las heurísticas de optimización vistas en el curso (transformar la selección con condición conjuntiva en una cascada de selecciones) se obtiene el árbol de la Figura 2





SOLUCIÓN parte c:

La Figura a continuación muestra un posible plan físico para el árbol de la Figura 6



Ejercicio 4 (15 puntos)

Considere las siguientes transacciones:

$T_1: r_1(y), w_1(x), w_1(y), c_1$

$T_2: r_2(x), w_2(y), w_2(x), c_2$

- Escriba una historia no serializable con estas transacciones. Justifique su respuesta.
- En esas transacciones se agregaron los bloqueos de read-write siguiendo sólo las reglas de los bloqueos pero ningún protocolo 2PL. Con esto se obtienen:

$T'_1: rl_1(y), r_1(y), u_1(y), wl_1(x), w_1(x), u_1(x), wl_1(y), w_1(y), u_1(y), c_1$

$T'_2: rl_2(x), r_2(x), u_2(x), wl_2(y), w_2(y), u_2(y), wl_2(x), w_2(x), u_2(x), c_2$

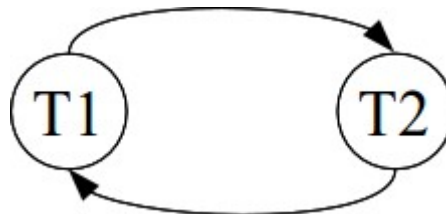
Escriba (si es posible) una historia no serializable con estas transacciones. Justifique su respuesta.

- Escriba las transacción T''_1 y T''_2 tales que son iguales con respecto a los read y write pero tienen los bloqueos y desbloqueos de forma que sigan 2PL Estricto. Justifique su respuesta.
- Escriba (si es posible), con las transacciones de la parte c), una historia no serializable. Justifique su respuesta.

SOLUCION:

- Una historia posible es la siguiente:
 $H_1: r_1(y), r_2(x), w_1(x), w_1(y), w_2(y), w_2(x), c_1, c_2$

El grafo de seriabilidad es el siguiente:



En donde el arco superior surge por el conflicto entre $r_1(y)$ y $w_2(y)$, y el arco inferior surge por el conflicto entre $r_2(x)$ y $w_1(x)$.

De esta forma, el grafo presenta un ciclo por lo que la historia no es serializable.

- Las operaciones de bloqueo y desbloqueo se colocaron siguiendo las reglas básicas, pero si seguir ninguna forma de 2PL. De esta forma, es posible considerar cada secuencia $rl(i), r(i), u(i)$ o $wl(i), w(i), u(i)$ como "operaciones unitarias". Con esta idea, se puede escribir una historia igual a la de la parte a) pero en donde se sustituye cada operación por la secuencia correspondiente. Así se obtiene:

$H'_1: rl_1(y), r_1(y), u_1(y), rl_2(x), r_2(x), u_2(x), wl_1(x), w_1(x), u_1(x), wl_1(y), w_1(y), u_1(y),$

$wl_2(y), w_2(y), u_2(y), wl_2(x), w_1(x), u_2(x), c_1, c_2$

Dado que no se cambiaron las posiciones de las operaciones en conflicto de la historia H_1 , el grafo de seriabilidad es el mismo por lo que la historia, a pesar de manejar bloqueos, no es serializable.

- c) En 2PL estricto, se respeta la idea de las dos fases (crecimiento y contracción) y se liberan los bloqueos de escritura luego de la finalización de la transacción por lo que, las transacciones quedarían de la siguiente forma:

$T''_1: r_1(y), r_1(y), w_1(x), w_1(x), w_1(y), w_1(y), c_1, u_1(x), u_1(y)$

$T''_2: r_2(x), r_2(x), w_2(y), w_2(y), w_2(x), w_2(x), c_2, u_2(y), u_2(x)$

Observar que ahora se debe trabajar con “promoción” de los bloqueos. Estas transacciones siguen 2PL estricto porque:

- Por el tipo y el orden en que se realizan los bloqueos, siempre se aumentan tanto en calidad como en cantidad (no hay r_1 ni u sobre un ítem luego de que se hizo un w_1 hasta que se liberan efectivamente los bloqueos).
 - Los bloqueos de escritura, se liberan luego de finalizada la transacción.
- d) Dado que se sigue 2PL estricto, no es posible generar historias no serializables porque el protocolo utilizado lo garantiza.