

FUNDAMENTOS DE BASES DE DATOS

Examen Febrero 2017

Solución

La duración del examen es de 3 horas.

Presentar la resolución del examen:

- 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.
- Comenzando cada ejercicio en una nueva hoja

Ejercicio 1 (25 puntos).

Se desea modelar la información más relevante de una escuela de gastronomía. De las personas involucradas con la escuela se conoce su cédula de identidad, nombre, todos sus teléfonos y dirección. Entre las personas hay docentes y alumnos, de los primeros interesa saber en qué año obtuvieron el título de chef, en qué instituto y en qué país. Éstos últimos no pueden asistir a ningún curso.

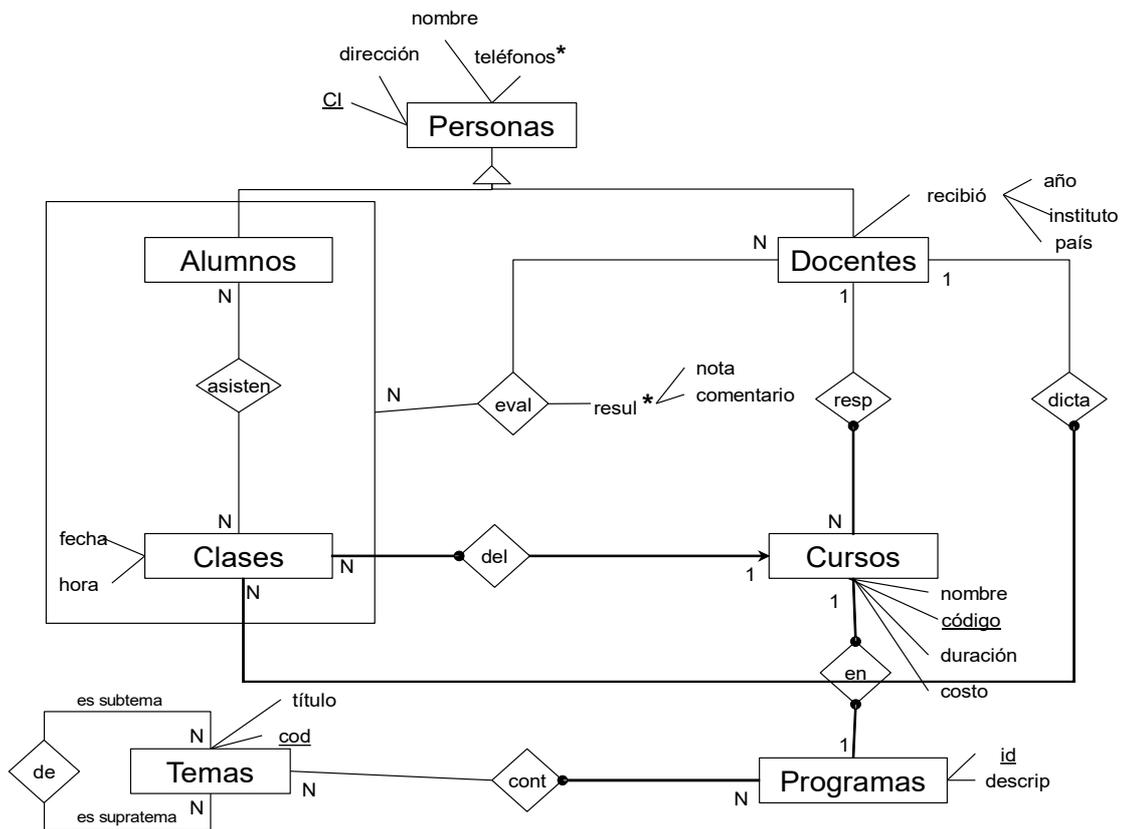
De los cursos dictados en la escuela se conoce su código, que los identifica, nombre, duración y costo. Para cada curso interesa registrar cada clase que se dicta, guardando la fecha y hora de la misma. Puede haber clases de distintos cursos el mismo día a la misma hora. Los alumnos asisten a las clases de todos los cursos que desean y podrían haber clases desiertas, es decir, sin alumnos.

Por otro lado, cada curso tiene un programa asignado. Dicho programa tiene un identificador y una descripción. Cada programa está formado por un conjunto de temas y hay temas que forman parte de más de un programa. Por ejemplo, el tema "Higiene en la cocina" está en todos los programas. Cada tema tiene un código identificador y un título e interesa saber cómo se ordenan dichos temas (independientemente del programa). Es decir, hay temas que tienen subtemas, por ejemplo "Limpieza y lavado" es subtema de "Higiene en la cocina", a su vez "Higiene en la cocina" tiene otros subtemas asociados.

Cada curso tiene un docente responsable y cada docente puede ser responsable de varios cursos. Diferentes clases de un mismo curso pueden ser dictadas por diferentes docentes, pero una clase siempre es dictada por un sólo docente. Un docente puede dictar varias clases de varios cursos.

A veces, cuando finaliza una clase, se realizan evaluaciones de forma sorpresiva. Las evaluaciones son realizadas por varios docentes, entre ellos el docente que dicta la clase. Cada docente asigna una nota y un comentario a cada alumno evaluado en cada evaluación, y toda esta información debe ser registrada. En una misma clase el alumno puede ser evaluado tantas veces como los docentes consideren necesario.

Se pide: Modelo Entidad-Relación completo.



RNE:

$\text{Alumnos} \cap \text{Docentes} = \emptyset$

El docente que dicta la clase el día de la evaluación es uno de los docentes que participa en la evaluación:
 $(\exists d \in \text{DOCENTES})(\exists a \in \text{ALUMNOS})(\exists c \in \text{CLASES})$
 $(\langle a, c \rangle \in \text{ASISTEN} * \langle d, c \rangle \in \text{DICTA} \rightarrow \langle d, \langle a, c \rangle \rangle \in \text{EVAL})$

La relación "de" es asimétrica e irreflexiva

Ejercicio 2 (25 puntos).

Sea el esquema relacional $R(A,B,C,D,E,G)$ y el conjunto de dependencias $F=\{GA \rightarrow ED, B \rightarrow G, C \rightarrow EA, D \rightarrow B\}$

1. Determinar si los siguientes conjuntos de atributos son claves de R con respecto a F . Justificar su respuesta.
 - a) CBA
 - b) CD
 - c) AG

SOLUCIÓN

Un conjunto de atributos X es clave de R según F $X^+_F = R$ (X es superclave) y no existe Y , $Y \subset X$ tal que $Y^+_F = R$ (minimal)

Por propiedad vista en el curso se sabe que si un atributo de R no pertenece a los lados derechos de las dependencias funcionales entonces pertenece a todas las claves. Entonces C tiene que pertenecer a **todas** las claves. **Por lo tanto AG no es clave.**

$C^+ = \{C,E,A\} \neq R$, por lo tanto C no es clave.

$(CD)^+ = \{C,E,A,D,B,G\} = R$, por lo tanto **CD es clave** (es superclave y no se puede quitar ningún atributo).

$(CBA)^+ = \{C,B,A,E,G,D\} = R$, por lo tanto CBA es superclave pero

$(CB)^+ = \{C,B,E,A,G,D\} = R$, entonces CBA no es minimal

CBA no es clave.

2. Obtener una cubrimiento minimal de F . Justificar.

SOLUCIÓN

Primer Paso: Descomposición de las dependencias de forma tal que los lados derechos tengan un único atributo.

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

Segundo Paso: Eliminación de atributos redundantes en los lados izquierdos de las dependencias.

$G^+_{F_1} = \{G\}$, $E \notin G^+_{F_1}$ y $D \notin G^+_{F_1}$ por lo tanto A no es redundante en $GA \rightarrow E$ ni en $GA \rightarrow D$

No hay atributos redundantes en los lados izquierdos de las dependencias funcionales.

Tercer paso: Eliminación de dependencias redundantes.

Las únicas posibles dependencias redundantes son $GA \rightarrow E$ y $C \rightarrow E$ ya que no hay otro par de dependencias que tengan el mismo lado derecho.

$(GA)^+_{F_1 - \{GA \rightarrow E\}} = \{G,A,D,B\}$, $E \notin (GA)^+_{F_1 - \{GA \rightarrow E\}}$ entonces $GA \rightarrow E$ no es redundante en F_1 .

$(C)^+_{F_1 - \{C \rightarrow E\}} = \{C,A\}$, $E \notin (C)^+_{F_1 - \{C \rightarrow E\}}$ entonces $C \rightarrow E$ no es redundante en F_1 .

No hay dependencias redundantes.

F_1 es un cubrimiento minimal de F .

3. Dar una descomposición de R en dos esquemas (R_1, R_2) tales que ninguno de los esquemas está incluido en otro y la descomposición es con Join Sin Pérdida respecto de F . Justificar la respuesta.

SOLUCIÓN

Sea la descomposición: $(R1(A,C,D,E), R2(C,D,B,G))$

Para demostrar que la descomposición tiene JSP aplicaremos el teorema visto en el curso. Para que se cumpla JSP es necesario que se cumpla alguna de estas condiciones:

$$R1 \cap R2 \rightarrow R1 - R2 \in F^+$$

$$R1 \cap R2 \rightarrow R2 - R1 \in F^+$$

para esta descomposición en particular las dependencias son:

$$CD \rightarrow AE \in F^+$$

$$CD \rightarrow BG \in F^+$$

y estas dependencias se cumplen ya que se demostró que CD es clave de R según F.

4. Obtener una descomposición de R en 3NF, con preservación de dependencias funcionales y JSP según F aplicando el algoritmo visto en el curso. Justificar todos los pasos.

SOLUCIÓN

Primer Paso: se genera un esquema relación a partir de las dependencias funcionales de un cubrimiento minimal que coinciden en el lado izquierdo.

R1(G,A,D,E)

R2(B,G)

R3(C,E,A)

R4(D,B)

Por lo calculado en la primer parte: CD y CB son claves de R según F.

Es necesario determinar si existen más claves ya que no hay ningún esquema que contenga estos conjuntos de atributos y si se confirma que no hay un esquema que contenga una clave es necesario agregarlo a la descomposición.

Si hay más claves tienen que estar contenidas en (ACEG)

$(ACEG)^+ = \{A,C,E,G,D,B\} = R$, por lo tanto hay más claves.

$(C,G)^+ = \{C,G,A,E,D,B\} = R$, por lo tanto CG es clave.

Si hay más claves debe estar contenida en (ACE)

$(ACE)^+ = \{A,C,E\} \not\subseteq R$, por lo tanto no hay más claves.

No hay un esquema que contenga una clave por lo tanto se agrega uno.

Descomposición resultante:

R1(G,A,D,E)

R2(B,G)

R3(C,E,A)

R4(D,B)

R5(C,D)

5. Dada la siguiente descomposición: R3(A,B,D,E) y R4(A,C,D,G) de R.
a) Obtener todas las claves de R3 según F. Justificar la respuesta.

SOLUCIÓN

Primero es necesario proyectar las dependencias de F en $R3 = F3$

$$A^+ = \{A\} \quad D^+ = \{D,B,G\} \quad * \quad D \overset{23}{\parallel} B \times F3$$

$$B^+ = \{B,G\} \quad E^+ = \{E\}$$

$$(AB)^+ = \{A,B,G,D,E\} \quad * \quad AB \overset{23}{\parallel} DE \times F3$$

$$(AD)^+ = \{A,D,B,G,E\} \quad * \quad AD \overset{23}{\parallel} E \times F3 \text{ (se deduce de las anteriores)}$$

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

$$F3 = \{ D \overset{23}{\parallel} B, AB \overset{23}{\parallel} DE \}$$

A debe pertenecer a todas las claves de R3 según F ya que no pertenece al lado derecho de las dependencias funcionales pero no es clave ya que $A^+ \not\subseteq R3$

- $(AB)^+ = R3$ * AB es clave de R3 según F.
 $(AD)^+ = R3$ * AD es clave de R3 según F.
 $(AE)^+ \neq R3$ * AE no es clave de R3 según F y no hay más claves.

Únicas claves: AB,AD.

- b) Determinar la máxima forma normal que cumple R3. Justificar la respuesta.

SOLUCIÓN

$R3(A,B,D,E)$, $F3 = \{ D \xrightarrow{23} B, AB \xrightarrow{23} DE \}$, claves: AB, AD

En $D \xrightarrow{23} B$: D no es superclave pero B es un atributo primo, por lo tanto cumple 3NF.

En $AB \xrightarrow{23} DE$: AB es superclave por lo tanto cumple BCNF y 3NF.

Por lo tanto la máxima forma normal que cumple R3 es 3NF.

- c) Sea $J = \star_{R3} F^{\text{sb}} \{ E \twoheadrightarrow A \}$.

Obtener una descomposición de R3 en 4NF con join sin pérdida según J aplicando el algoritmo visto en el curso.

SOLUCIÓN

$R3(A,B,D,E)$, $J = \{ D \xrightarrow{23} B, AB \xrightarrow{23} DE, E \twoheadrightarrow A \}$, claves: AB,AD

$D \xrightarrow{23} B$ viola 4NF ya que D no es superclave de R3, por lo que se debe descomponer:

$R31(D,B)$ Cumple 4NF por ser un esquema con 2 atributos.

$R32(A,D,E)$, $J32 = \{ AD \xrightarrow{23} E, E \twoheadrightarrow A \}$, clave: AD

$E \twoheadrightarrow A$ viola 4NF ya que E no es superclave de R32, por lo que se debe descomponer:

$R321(E,A)$ Cumple 4NF por ser un esquema con 2 atributos

$R322(D,E)$ Cumple 4NF por ser un esquema con 2 atributos

Descomposición de R3 en 4NF según J: **$R31(D,B)$, $R321(E,A)$, $R322(D,E)$**

Ejercicio 3 (30 puntos).

Un complejo de vacaciones cuenta con un sistema de información para la gestión de los alquileres de cabañas. Este sistema usa las siguientes tablas:

CABAÑAS (codCab, direccionCab, codCat)

Cada cabaña se identifica por un código, y se conoce su dirección y su categoría.

CATEGORIAS (codCat, descripcionCat)

Ejemplos de descripciones de categorías son 'Económica', 'Básica', 'Lujo', etc.

COMODIDADES (codCom, descripcionCom)

Ejemplos de descripciones de comodidades son 'Aire acondicionado', 'Wi-fi', 'Lavarropas', etc.

EQUIPAMIENTO (codCab, codCom)

Esta tabla representa las comodidades disponibles en cada cabaña.

TARIFAS (codCat, fechaDesde, fechaHasta, tarifaDiaria)

El complejo no tienen una lista de precios por cabaña, si no que las tarifas diarias para el alquiler de cabañas corresponden a cierta categoría y tienen un período de vigencia (comprendido entre fechaDesde y fechaHasta). A modo de ejemplo, a todas las cabañas categoría 'Lujo' se les aplica la misma tarifa diaria en cierto período.

CLIENTES (tipoDoc, nroDoc, nombre, apellido, telefono, direccionCliente, paisOrigen)

Esta tabla almacena información sobre los clientes que realizan reservas.

RESERVAS (codCab, comienzoEstadia, finEstadia, tipoDoc, nroDoc, fechaReserva, cantDias)

Esta tabla almacena las reservas realizadas. El campo cantDias se calcula y almacena a partir de las fechas comienzoEstadia y finEstadia para facilitar los cálculos de costos de la reserva.

En la base de datos no hay tablas vacías y se cumplen las siguientes dependencias de inclusión:

$$\Pi_{\text{codCab}}(\text{EQUIPAMIENTO}) \subseteq \Pi_{\text{codCab}}(\text{CABAÑAS})$$

$$\Pi_{\text{codCom}}(\text{EQUIPAMIENTO}) \subseteq \Pi_{\text{codCom}}(\text{COMODIDADES})$$

$$\Pi_{\text{codCat}}(\text{TARIFAS}) \subseteq \Pi_{\text{codCat}}(\text{CATEGORIAS})$$

$$\Pi_{\text{codCat}}(\text{CABAÑAS}) \subseteq \Pi_{\text{codCat}}(\text{CATEGORIAS})$$

$$\Pi_{\text{codCab}}(\text{RESERVAS}) \subseteq \Pi_{\text{codCab}}(\text{CABAÑAS})$$

$$\Pi_{\text{tipoDoc, nroDoc}}(\text{RESERVAS}) \subseteq \Pi_{\text{tipoDoc, nroDoc}}(\text{CLIENTES})$$

Se pide:

1. Resolver las siguientes consultas en **Álgebra Relacional**.

a) Obtener el código y dirección de las cabañas que tienen *Lavarropas* y nunca fueron reservadas.

Cabañas con lavarropas

$$A = \Pi_{\text{codCab}}(\text{EQUIPAMIENTO} * (\sigma_{\text{descripcionCom} = \text{'Lavarropas'}}(\text{COMODIDADES})))$$

$$R = \Pi_{\text{codCab, nombreCab}}((A - \Pi_{\text{codCab}}(\text{RESERVAS})) * \text{CABAÑAS})$$

b) Obtener el código y dirección de las cabañas correspondientes a la/s categoría/s con tarifa diaria más barata considerando las tarifas cuyo período de vigencia comienza el 15/02/2017 y finaliza el 20/02/2017.

Categorías y tarifas vigentes en el período indicado

$A = \Pi_{\text{codCat}, \text{tarifaDiaria}} (\sigma_{\text{fechaDesde}=15/02/2017 \text{ AND } \text{fechaHasta}=20/02/2017} \text{TARIFAS})$

Tarifa mínima en el período indicado

$B = (\Pi_{\text{tarifaDiaria}} A) \mid > < \mid_{\text{tarifaDiaria} <= \text{diaria2}} (\rho_{\text{tarifaDiaria} \rightarrow \text{diaria2}} (\Pi_{\text{tarifaDiaria}} A)) \% \rho_{\text{tarifaDiaria} \rightarrow \text{diaria2}} (\Pi_{\text{tarifaDiaria}} A)$

Categorías con tarifa mínima en el período indicado

$C = A * B$

$R = \Pi_{\text{codCab}, \text{nombreCab}} (\text{CABAÑAS} * C)$

2. Resolver las siguientes consultas en **Cálculo Relacional**.

- c) Obtener el nombre, apellido y país de origen de aquellos clientes que sólo realizan reservas para cabañas de cierta categoría.

$\{c.\text{nombre}, c.\text{apellido}, c.\text{paisOrigen} / \text{CLIENTES}(c) \wedge$
 $(\exists r1) (\text{RESERVAS}(r1) \wedge r1.\text{tipoDoc}=c.\text{tipoDoc} \wedge r1.\text{nroDoc} = c.\text{nroDoc} \wedge$
 $(\exists ca)(\text{CABAÑAS}(ca) \wedge r1.\text{codCab}=ca.\text{codCab} \wedge$
 $\neg (\exists r2) (\text{RESERVAS}(r2) \wedge r2.\text{tipoDoc}=c.\text{tipoDoc} \wedge r2.\text{nroDoc} = c.\text{nroDoc} \wedge$
 $(\exists ca1)(\text{CABAÑAS}(ca1) \wedge r2.\text{codCab}=ca1.\text{codCab} \wedge ca.\text{codCat} <> ca1.\text{codCat}))\}$

3. Resolver las siguientes consultas en **SQL** (sin utilizar vistas)

- d) Obtener nombre y apellido de los clientes que han reservado 5 o más cabañas distintas para estadias comprendidas en el período 1/1/2015 – 31/12/2016.

```
SELECT c.nombre,c.apellido
FROM CLIENTES c NATURAL JOIN RESERVAS r
WHERE r.comienzoEstadia >= 1/1/2015 AND r.finEstadia <= 31/12/2016
GROUP BY c.tipoDoc, c.nroDoc, c.nombre, c.apellido
HAVING COUNT (DISTINCT r.codCab)) >=5
```

- e) Resolver la misma consulta de la parte b).

```
SELECT c1.codCab, c1.direccionCab
FROM CABAÑAS c1 NATURAL JOIN TARIFAS t
WHERE t.fechaDesde = 15/02/2017 AND t.fechaHasta = 20/02/2017 AND
t.tarifaDiaria =
(SELECT MIN (s.tarifaDiaria)
FROM TARIFAS s
WHERE s.fechaDesde = 15/02/2017 AND s.fechaHasta = 20/02/2017)
```

Ejercicio 4 (20 puntos)

Considere las siguientes historias:

$H_1: r_1(x), r_1(y), w_1(x), r_2(x), w_2(x), r_2(y), w_1(y), w_2(y), c_1, c_2$
 $H_2: r_1(x), r_2(x), r_1(y), w_1(x), w_2(x), r_2(y), w_1(y), w_2(y), a_2, c_1$
 $H_3: r_1(x), r_2(x), r_1(y), w_1(x), w_2(x), r_2(y), w_1(y), w_2(y), c_2, c_1$

- Suponiendo que la operación $w_1(y)$, graba el valor 8 y $w_2(y)$ graba el valor 5, indique para cada una de las historias, cuál es el valor en la base para item y, posterior a la ejecución completa de esa historia. Justifique sus respuestas.
- Escriba una historia serial equivalente por conflictos para cada una de las historias H_1 y H_3 . En caso que no sea posible, indíquelo. Justifique sus respuestas.
- Escriba las transacciones T_1 y T_2 involucradas en la historia H_3 y agregue locks read/write de forma que las transacciones sigan 2PL Básico pero no Estricto. Coloque los locks lo más tarde posible para cumplir con las condiciones. Justifique su respuesta.
- Construya, si es posible, una historia entrelazada entre T_1 y T_2 teniendo en cuenta los locks que agregó en la parte c). Justifique su respuesta.

SOLUCION:

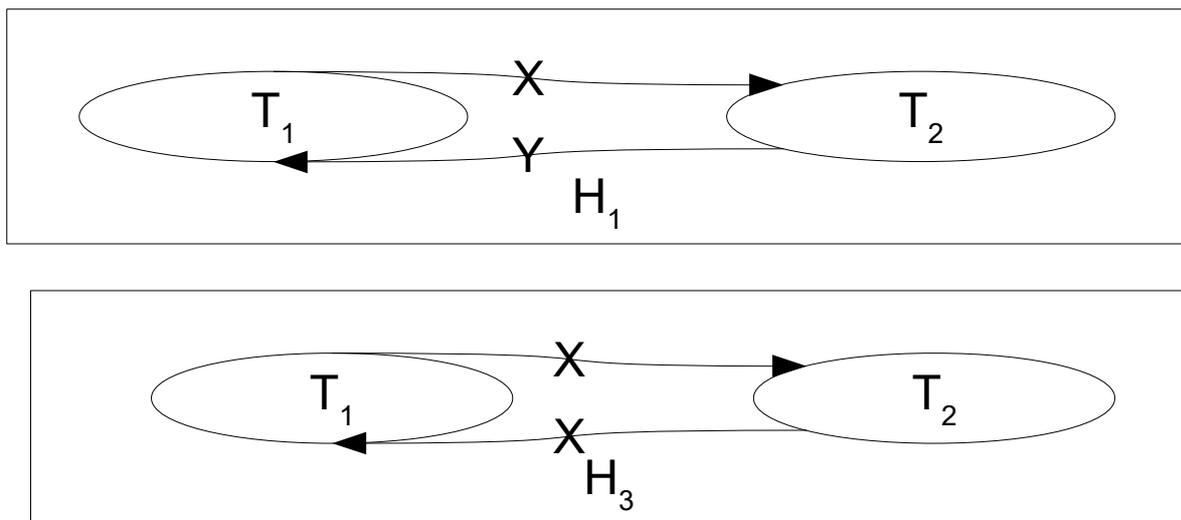
a)

En las historias H_1 y H_3 confirman las dos transacciones involucradas, por lo que el valor en la base es el que se grabó en la operación de escritura que se ejecuta último. En ambos casos, es el valor escrito por la operación $w_2(y)$ y por lo tanto es 5.

En la historia H_2 , la transacción 2 aborta, por lo que tiene que quedar en la base sólo el valor grabado por la otra transacción y por lo tanto, es 8.

b)

La forma de determinar si hay o no una historia serial equivalente e incluso el orden de las transacciones es construyendo el grafo de seriability de cada historia. En la figura 1 se ven los grafos de seriability de las 2 historias.



La historia H_1 no es serializable, dado que hay un conflicto entre las operaciones $w_1(x)$ y $r_2(x)$ y otro entre las operaciones $r_2(y)$ y $w_1(y)$ lo que genera un ciclo en el grafo de seriability.

Con la historia H_3 pasa algo similar, pero ahora con las operaciones sobre X. Los conflictos a considerar son entre $w_1(x)$ y $w_2(x)$ y el otro en $r_2(x)$ y $w_1(x)$.

c)

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

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

Para seguir 2PL Básico se deben identificar dentro de cada transacción, una zona de crecimiento de locks y una zona de disminución. Sin embargo, para no cumplir con 2PL Estricto, NO deben liberarse todos los locks al finalizar la transacción. Por eso, las transacciones pueden quedar de la siguiente forma:

$T_1: rl_1(x), r_1(x), rl_1(y), r_1(y), wl_1(x), w_1(x), wl_1(y), u_1(x), w_1(y), u_1(y), c_1$

$T_2: rl_2(x), r_2(x), wl_2(x), w_2(x), rl_2(y), r_2(y), wl_2(y), u_2(x), w_2(y), u_2(y), c_2$

La ubicación de los locks es "lo más tarde posible" dado que la operación inmediatamente posterior a cada lock es la correspondiente a ese tipo de lock e ítem. (Ej: en T_1 , inmediatamente después del read lock sobre X aparece el read de X.)

d)

Considerando la solución de la parte c), una historia podría ser la siguiente:

$rl_1(x), r_1(x), rl_1(y), r_1(y), wl_1(x), w_1(x), wl_1(y), u_1(x), rl_2(x), r_2(x), wl_2(x), w_2(x), w_1(y), u_1(y), rl_2(y), r_2(y), wl_2(y), u_2(x), w_2(y), u_2(y), c_2, c_1$