

FUNDAMENTOS DE BASES DE DATOS

Examen Diciembre 2012 - SOLUCION

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 realizar un relevamiento de las guarderías del departamento de Montevideo.

De cada guardería se conoce un código identificador asignado por el Ministerio de Educación y Cultura, su nombre, dirección, y el conjunto de salones de la guardería. De cada salón interesa su capacidad y el nombre que lo identifica dentro de la guardería, por ejemplo "Maternal", "2 años", "3 años", etc. Los nombres de los salones pueden repetirse en guarderías distintas.

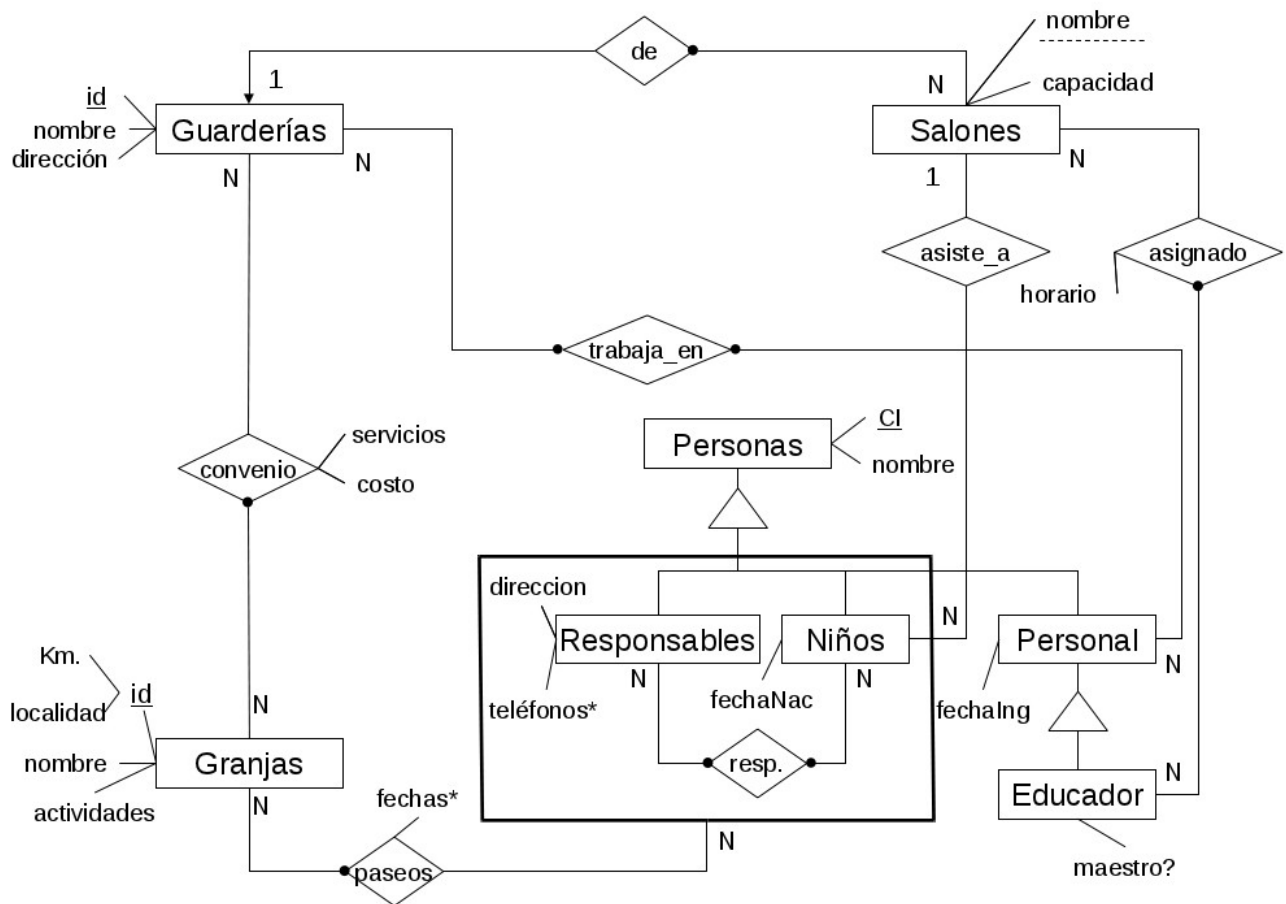
Interesa modelar a las personas que se relacionan con la guardería, entre las cuales están tanto el personal que trabaja en las guarderías, como los niños que asisten a las mismas y los adultos responsables de éstos. El personal que trabaja en las guarderías puede ser personal administrativo o educador. Tanto los administrativos como los educadores pueden trabajar en más de una guardería, y de los educadores interesa saber si tienen título de maestro y a qué salones están asignados dentro de la guarderías en que trabajan. Del personal se conoce su CI y su nombre completo, y de los educadores se conoce el horario en que trabajan en cada salón. De los adultos responsables se conoce su CI, su nombre completo, su dirección y una lista de teléfonos de contacto. De cada niño se conoce al menos un adulto responsable, el cual puede trabajar en una guardería. Además se conoce su CI, su nombre completo, su fecha de nacimiento y a que salón está asignado. Cada niño es asignado a un solo salón.

Algunas guarderías tienen convenios con granjas, las cuales son visitadas en paseos. Cada granja hace convenios particulares con cada guardería, y de éstos interesa el tipo de servicio brindado (ej: "paseo", "paseo con almuerzo", etc.) y el costo asociado por niño. Una guardería puede tener convenio con varias granjas e interesa registrar únicamente las granjas que tienen convenio con al menos una guardería. De las granjas se conoce su nombre, su identificador (el km. y localidad en la cual se encuentra) y una lista de actividades que pueden realizarse en ella (ej: "paseos a caballo", "huerta", etc).

Para que los niños puedan asistir a paseos a las granjas las guarderías solicitan que uno de los responsables del niño firme un permiso. Interesa saber, para cada paseo, la lista de niños que asistieron, la fecha de realización y que responsable firmó el permiso de cada niño.

Se pide: Modelo Entidad Relación completo.

Solución



RNE:

- $\text{Personal} \cap \text{Niños} = \emptyset$
- $\text{Responsables} \cap \text{Niños} = \emptyset$
- $\text{Personal} \cup \text{Niños} \cup \text{Responsables} = \text{Personas}$
- Los educadores son asignados a salones de las guarderías en que trabajan
- Los niños van a paseos a granjas con las cuales tienen convenios las guarderías a las que asisten

Ejercicio 2 (30 puntos).

Un organismo no gubernamental que se dedica a la ayuda social y divulgación de sus creencias filosóficas maneja toda su información en una base de datos con los siguientes esquemas relación:

PERSONAS(ci, nombre, medioConocimiento, fechaContacto)

Contiene la información sobre las personas que se ponen en contacto con la organización ya sea en busca de información o por asistir a una conferencia.

De cada una de estas personas se conoce su número de cédula de identidad que los identifica, su nombre, el medio por el cuál se enteraron de la existencia de la organización y la fecha en que se realiza el primer contacto con la persona.

MIEMBROS(ci, fechaIncorporacion, cireferente)

Contiene la información de las personas que se incorporan a la organización, o sea, pasan ser miembros de la misma.

De cada uno de estos se conoce su número de cédula de identidad, la fecha de incorporación a la organización y el número de cédula de identidad del miembro que pasa a ser su referente.

GRUPOS(codgrupo, nombre, areaTrabajo, lugarFisico)

Contiene la información de los diferentes grupos de trabajo que existen dentro de la organización. De cada uno de ellos se conoce un código que los identifica, su nombre, su área de trabajo y una identificación del lugar físico que ocupa dentro del edificio de la organización.

DEDICACION(ci, codgrupo, fecha, horas)

Contiene la información sobre las horas de trabajo dedicadas por los miembros de la organización en cada uno de los grupos que trabajó en las diferentes fechas en que lo hizo.

CONFERENCIAS(codconf, tema, fecha, codgrupo, ciConferencista)

Contiene la información de las diferentes conferencias que realiza la organización tanto como parte de la formación de sus miembros como de divulgación de sus actividades.

De cada conferencia se conoce un código que la identifica, el tema de la misma, la fecha en que se realiza, el código del grupo responsable de su realización y el número de cédula del miembro de la organización que es el conferencista.

ASISTENCIA(codconf, ci)

Contiene la información de la asistencia de las diferentes personas a las distintas conferencias.

En este esquema no existen tablas vacías y se cumplen las siguientes relaciones de inclusión:

- $\prod_{ci} (\text{MIEMBROS}) \subseteq \prod_{ci} (\text{PERSONAS})$
- $\prod_{ci} (\text{DEDICACION}) \subseteq \prod_{ci} (\text{MIEMBROS})$
- $\prod_{codgrupo} (\text{DEDICACION}) \subseteq \prod_{codgrupo} (\text{GRUPOS})$
- $\prod_{codgrupo} (\text{CONFERENCIAS}) \subseteq \prod_{codgrupo} (\text{GRUPOS})$
- $\prod_{ciConferencista} (\text{CONFERENCIAS}) \subseteq \prod_{ci} (\text{MIEMBROS})$
- $\prod_{codconf} (\text{ASISTENCIA}) \subseteq \prod_{codconf} (\text{CONFERENCIAS})$
- $\prod_{ci} (\text{ASISTENCIA}) \subseteq \prod_{ci} (\text{PERSONAS})$

Resolver en álgebra relacional las siguientes consultas:

- 1) Tema y código del grupo responsable de las conferencias a las que no asistió ningún miembro de la organización.

$$A = \prod_{codconf} (\text{ASISTENCIA} * \text{MIEMBROS})$$

$$B = \prod_{codconf} (\text{ASISTENCIA}) - A$$

$$\text{SOL} = \prod_{tema, codgrupo} (B * \text{CONFERENCIAS})$$

- 2) Nombre y fecha de contacto con la organización de las personas que han dedicado horas en todos y cada uno de los grupos que tienen por área de trabajo "Ayuda Social" y nunca asistieron a conferencias organizadas por esos grupos.

$$A = \prod_{\text{codgrupo}} (\sigma_{\text{areaTrabajo} = \text{Ayuda Social}} \text{GRUPOS})$$

$$B = (\prod_{\text{ci, codgrupo}} \text{DEDICACION}) \% A$$

$$C = \prod_{\text{ci}} (\text{CONFERENCIAS} * A * \text{ASISTENCIA})$$

$$\text{SOL} = \prod_{\text{nombre, fechaContacto}} ((B - C) * \text{PERSONAS})$$

Resolver en cálculo relacional las siguientes consultas:

- 3) Nombre de las personas que se incorporaron a la organización el mismo día que asistieron a una conferencia y a partir de ese momento asistieron a todas las conferencias del tema de esa conferencia inicial.

$$\{ p.\text{nombre} / \text{PERSONAS}(p) \wedge (\exists m)(\text{MIEMBROS}(m) \wedge$$

$$(\exists c)(\text{CONFERENCIAS}(c) \wedge (\exists a)(\text{ASISTENCIA}(a) \wedge a.\text{ci} = p.\text{ci} \wedge a.\text{codconf} = c.\text{codconf}$$

$$\wedge c.\text{fecha} = m.\text{fechaIncorporacion} \wedge$$

$$(\forall c1)(\text{CONFERENCIAS}(c1) \wedge c1.\text{tema} = c.\text{tema} \wedge c1.\text{fecha} > c.\text{fecha}$$

$$\rightarrow (\exists a1)(\text{ASISTENCIA}(a1) \wedge a1.\text{codconf} = c1.\text{codconf} \wedge a1.\text{ci} = p.\text{ci}$$

$$))$$

$$)$$

$$)$$

$$\}$$

- 4) Nombre de los miembros de la organización que desde su incorporación sólo van a conferencias del mismo tema.

$$\{ p.\text{nombre} / \text{PERSONAS}(p) \wedge (\exists m)(\text{MIEMBROS}(m) \wedge m.\text{ci} = p.\text{ci} \wedge$$

$$(\exists a)(\text{ASISTENCIA}(a) \wedge$$

$$(\exists c)(\text{CONFERENCIAS}(c) \wedge a.\text{ci} = m.\text{ci} \wedge a.\text{codconf} = c.\text{codconf} \wedge$$

$$(\forall a2)(\text{ASISTENCIA}(a2) \wedge a2.\text{ci} = a.\text{ci} \wedge$$

$$(\exists c2)(\text{CONFERENCIAS}(c2) \wedge a2.\text{codconf} = c2.\text{codconf} \wedge$$

$$c2.\text{fecha} \geq m.\text{fechaIncorporacion}$$

$$\rightarrow c2.\text{tema} = c.\text{tema}$$

$$))$$

$$)$$

$$)$$

$$\}$$

Resolver en SQL, sin utilizar vistas ni sub-consultas en el FROM, la siguientes consulta:

- 5) Nombre y área de trabajo de los grupos que fueron los responsables de la mayor cantidad de conferencias en el 2012.

```
select G.nombre, G.areaTrabajo
from Grupos G
where G.codgrupo IN
(select C.codgrupo
 from Conferencias C
 where C.fecha >= 1/1/12 and C.fecha <= 31/12/12
 group by C.codgrupo
 having count(*) >= ALL
 (select count(*)
  from Conferencias C2
  where C2.fecha >= 1/1/12 and C2.fecha <= 31/12/12
  group by C2.codgrupo
 )
 )
```

Ejercicio 3 (25 puntos).

Una productora de música lleva información de los músicos que participan de sus producciones.

De los músicos se conoce su documento (mDoc) y su nombre (mNom), los nombres de las bandas (bNom) en las que participa y los instrumentos (I) que toca. De cada banda se conocen las fechas (bFpres) en las que se presentó además de los géneros (bGen) que toca en cada presentación (no tiene por qué tocar los mismos géneros en todas las presentaciones). En cada banda en la que participa (que pueden ser muchas para cada músico), cada músico toca sólo un instrumento. Además, un músico se presenta en una sola banda en una fecha dada. Se sabe que cada banda usa todos sus instrumentos en todas las presentaciones.

Considere la tabla $R(mDoc, mNom, bNom, I, bFpres, bGen)$.

- a) Indique cuáles de las siguientes dependencias funcionales se cumplen en R y cuáles no. Justifique su respuesta para cada dependencia:
- $bFpres, bNom \rightarrow mDoc$
 - $bNom, mDoc \rightarrow I$
 - $mDoc, I \rightarrow bNom$
 - $bNom, mDoc \rightarrow Fpres$
 - $mDoc, bFpres \rightarrow bNom$
 - $mDoc \rightarrow mNom$
- b) Determine todas las claves de R considerando las dependencias que marcó que se cumplen en la parte a). Justifique su respuesta.
- c) Indique cuáles de las siguientes dependencias multivaluadas se cumplen en R y cuáles no. Justifique su respuesta.
- $mDoc \twoheadrightarrow I$
 - $mDoc \twoheadrightarrow bNom$
 - $mDoc \twoheadrightarrow bFpres$
- d) Indique cuáles de las siguientes dependencias multivaluadas embebidas se cumplen en R y cuáles no. Justifique su respuesta.
- $mDoc \twoheadrightarrow I \mid bNom$
 - $bNom \twoheadrightarrow I \mid bFpres$
 - $bNom \twoheadrightarrow bGen \mid bFpres$
- e) Indique en qué forma normal está cada una de las siguientes tablas que conforman una descomposición de R .
- $R_1(bNom, I, bFpres)$, $R_2(mDoc, mNom, bNom, bGen)$
- f) Encuentre una descomposición de R en BCNF aplicando el algoritmo visto en el curso.

SOLUCION

- a) .
- $bFpres, bNom \rightarrow mDoc$: No se cumple, ya que en una presentación de una banda, puede haber varios músicos.
 - $bNom, mDoc \rightarrow I$: Si se cumple, ya que en una banda, un músico toca un sólo instrumento.
 - $mDoc, I \rightarrow bNom$: no se cumple, ya que un músico, puede tocar el mismo instrumento en varias bandas.
 - $bNom, mDoc \rightarrow Fpres$: no se cumple, ya que un mismo músico, puede presentarse con la misma banda en varias noches.
 - $mDoc, bFpres \rightarrow bNom$: Si se cumple, dado que un músico, en una fecha dada, sólo se presenta con una sola banda.
 - $mDoc \rightarrow mNom$: Si se cumple, dado el documento identifica al músico.
- b) .
- $mDoc \rightarrow I$: No se cumple, dado que el instrumento que toca un músico, depende de la banda y un músico puede tocar en varias bandas y varios instrumentos. Viola que $bNom, mDoc \rightarrow I$. Si hubiera dos tuplas t_1 y t_2 tales que $t_1[mDoc]=t_2[mDoc]$ y $t_1[I] <> t_2[I]$ y $t_1[bNom] <> t_2[bNom]$, no pueden existir t_3 y t_4 con los cruces necesarios.
 - $mDoc \rightarrow bGen$: No se cumple, dado que un músico puede tocar varios géneros en varias bandas, pero eso no obliga a que deba tocar los mismos géneros en todas las bandas en las que toca.
- c) .
- $mDoc \rightarrow I | bNom$: No se cumple, dado que un músico, puede tocar en varias bandas pero en cada banda sólo puede tocar un instrumento. Si el músico toca varios instrumentos y además toca en varias bandas, esta dependencia obligaría a que tocara todos los instrumentos en todas las bandas en las que toca.
 - $bNom \rightarrow bFpres | I$: Si se cumple, dado que una banda utiliza todos sus instrumentos en todas las presentaciones.

- d) Sobre el esquema $R(mDoc, mNom, bNom, I, bFpres, bGen)$, las dependencias funcionales que se cumplen son:
 $F = \{mDoc \rightarrow mNom; bNom, mDoc \rightarrow I; mDoc, bFpres \rightarrow bNom; bNom \rightarrow bFpres | I\}$

Para determinar las claves se separan los atributos de R en tres conjuntos.

SóloALaDerecha = $\{mNom, I\}$
 NuncaALaDerecha = $\{bGen, mDoc, bFpres\}$
 DeLosDosLados = $\{bNom\}$

Los que nunca aparecen a la derecha, deben estar en todas las claves, mientras que los que aparecen sólo a la derecha nunca aparecen en una clave. Los que aparecen de los dos lados, pueden aparecer en algunas de las claves o en ninguna.

Ahora chequeamos si los que nunca aparecen a la derecha conforman una superclave:

$(bGen, mDoc, bFpres)^+ = \{bGen, mDoc, bFpres, mNom, bNom, I\}$

Efectivamente ese conjunto es superclave por lo que es la única clave posible, dado que todos los atributos tienen que estar en todas las claves y si se agrega otro atributo a ese conjunto quedaría una superclave no minimal.

- e) El primer paso es proyectar las dependencias sobre cada uno de los siguientes esquemas:

$R_1(bNom, I, bFpres), R_2(mDoc, mNom, bNom, bGen)$

$\Pi_{R_1}(F) = \{bNom \rightarrow bFpres\}$ (porque la embebida se vuelve "efectiva" en este esquema).

$\Pi_{R_2}(F) = \{mDoc \rightarrow mNom\}$

De esta forma, la dependencia multivaluada que se cumple en R_1 viola 4NF pero como no hay ninguna dependencia funcional que viole BCNF, entonces R_1 está en BCNF.

Con respecto a R_2 , $mNom$ es un atributo claramente no primo (sólo aparece a la derecha) y $mDoc$, no es clave (basta ver que $mDoc^+ = \{mDoc, mNom\}$) pero no aparece a la derecha de una dependencia funcional tiene que ser parte de cualquier clave, por lo que es un atributo primo. De esta forma, R_2 está en 1NF.

- f) El esquema a normalizar es $R(mDoc, mNom, bNom, l, bFpres, bGen)$ y las dependencias son:
- $bNom, mDoc \rightarrow l$
 - $mDoc, bFpres \rightarrow bNom$
 - $mDoc \rightarrow mNom$

La dependencia c, viola BCNF. Por lo tanto partimos R en dos tablas con las respectivas proyecciones :

$R_{11}(mDoc, mNom) \quad \{ mDoc \rightarrow mNom \}$

$R_{12}(mDoc, bNom, l, bFpres, bGen) \quad \{ bNom, mDoc \rightarrow l; mDoc, bFpres \rightarrow bNom \}$

R_{11} ya quedó en BCNF. Sin embargo R_{12} no está BCNF, dado que $(bNom, mDoc) \neq \{bNom, mDoc, l\}$ y no alcanza a bFpres o bGen por lo que no es superclave.

De esta forma, partimos nuevamente en dos tablas:

$R_{121}(bNom, mDoc, l) \quad \{ bNom, mDoc \rightarrow l \}$

$R_{122}(mDoc, bNom, bFpres, bGen) \quad \{ mDoc, bFpres \rightarrow bNom \}$

R_{121} esta en BCNF pero no lo está R_{122} dado que $(mDoc, bFpres) \neq \{mDoc, bFpres, bNom\}$ y no alcanza a bGen por lo que no es superclave.

Partimos nuevamente en dos tablas:

$R_{1221}(mDoc, bFpres, bNom) \quad \{ mDoc, bFpres \rightarrow bNom \}$

$R_{1222}(mDoc, bFpres, bGen) \quad \{ \}$

Las dos últimas tablas ya quedaron en BCNF, por lo que la descomposición es la siguiente:

$R_{11}(mDoc, mNom) \quad R_{121}(bNom, mDoc, l) \quad R_{1221}(mDoc, bFpres, bNom) \quad R_{1222}(mDoc, bFpres, bGen)$

Ejercicio 4 (20 puntos).

Considere las siguientes transacciones y las siguientes historias con esas transacciones:

$T_1: r_1(X), r_1(X), w_1(X), w_1(X), w_1(Y), w_1(Y), c_1, u_1(X), u_1(Y)$

$T_2: r_2(Z), r_2(Z), u_2(Z), w_2(Z), w_2(Z), w_2(V), u_2(Z), w_2(V), u_2(V), c_2$

$T_3: r_3(X), r_3(X), u_3(X), w_3(X), w_3(X), w_3(V), u_3(X), w_3(V), u_3(V), c_3$

$H_1: r_1(X), r_1(X), r_2(Z), w_1(X), r_2(Z), u_2(Z), w_1(X), w_1(Y), w_2(Z), w_2(Z), w_2(V), w_1(Y), u_2(Z), w_2(V), u_2(V), c_2, c_1, u_1(X), u_1(Y)$

$H_2: r_1(X), r_1(X), r_3(X), w_1(X), w_1(X), w_1(Y), r_3(X), u_3(X), w_3(X), w_3(X), w_3(V), w_1(Y), u_3(X), w_3(V), u_3(V), c_3, c_1, u_1(X), u_1(Y)$

$H_3: r_1(X), r_1(X), r_3(X), r_3(X), u_3(X), w_3(X), w_3(X), w_3(V), u_3(X), w_1(X), w_1(X), w_1(Y), w_1(Y), w_3(V), u_3(V), c_3, c_1, u_1(X), u_1(Y)$

- Indique si las transacciones anteriores siguen el protocolo 2PL, y en caso afirmativo indique de qué tipo. Justifique su respuesta.
- Construya los grafos de seriabilidad de las cada una de las historias anteriores. Justifique su respuesta.
- Indique si alguna de las historias es imposible. Justifique su respuesta.
- Indique cuáles (de las que identificó como posibles en la parte c) evitan abortos en cascada y cuáles son serializables. Justifique su respuesta.

Nota: Se corrigió error de la letra en historia H_3 .

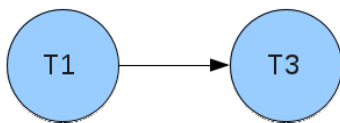
a) T1 sigue 2PL riguroso, ya que cumple con las dos fases del 2PL, y además, no se liberan locks de escritura ni lectura hasta confirmada la transacción. T2 y T3 no siguen ningún protocolo 2PL, porque no se respetan las dos fases.

b)

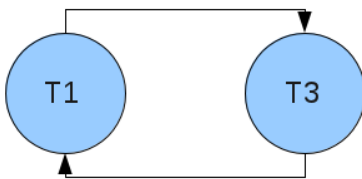
En H1 no hay operaciones en conflicto ya que las transacciones que participan (T1 y T2) operan sobre ítems distintos, resultando el siguiente grafo de seriabilidad:



En H2, T3 lee de T1, ya que en la historia aparece una lectura de X de T3 luego de una escritura del mismo ítem por T1, resultando el siguiente grafo de seriabilidad (el resto de las operaciones en conflicto no generan nuevos arcos en el grafo):



En H3, aparece una escritura de X de T1 luego de una lectura del mismo ítem por T3, resultando el siguiente grafo de seriabilidad (el resto de las operaciones en conflicto no generan nuevos arcos en el grafo):



c)

Solamente H1 es posible.

H2 y H3 no son posibles. H2 es imposible dado que T1 realiza un lock de escritura sobre X luego de un lock de lectura de T3 sobre el mismo ítem. H3 es imposible porque T3 realiza un lock de escritura sobre X cuando T1 tiene sobre X un lock de lectura.

d)

H1 evita abortos en cascada, ya que no tiene operaciones en conflicto, y es serializable dado que su grafo de seriabilidad asociado es acíclico (aplicación directa del Teorema de Seriabilidad).