

# Segundo Parcial de Fundamentos de Bases de Datos

diciembre 2014

## Indicaciones generales

- La duración del parcial es de **tres (3)** horas.
- En esta prueba **no** se permite consultar material alguno.
- Puntaje: **60** puntos.
- **Toda respuesta debe estar fundamentada.**
- Numerar todas las hojas e incluir en cada una su nombre y cédula de identidad, utilizar las hojas de un solo lado, escribir con lápiz, iniciar cada ejercicio en hoja nueva y poner en la primera hoja la cantidad de hojas entregadas.

## Ejercicio 1 (6 puntos)

Considere un esquema  $R$ , un conjunto de dependencias  $F$  y la descomposición de  $R$ :  $\rho = (R_1, R_2)$ , con  $R_1 = (A, B, C, D)$  y  $R_2 = (A, B, E, G, H)$ . Suponga que se sabe que  $\rho$  tiene JSP respecto a  $F$ .

- a. Sea  $r$  una instancia válida de  $R$  según  $F$ , y sean  $r_1 = \prod_{R_1}(r)$  y  $r_2 = \prod_{R_2}(r)$  ¿Cuál es el resultado de la siguiente consulta si se realiza sobre  $r_1$  y  $r_2$ ? Justifique su respuesta.

```
SELECT R1.A, R1.B, R1.C, R1.D, R2.E, R2.G, R2.H
FROM R1 NATURAL JOIN R2
```

- b. Para cada una de las siguientes afirmaciones, indicar si es *verdadera*, *falsa*, o si *no es posible realizar dicha afirmación*. Justificar las repuestas.

- $AB \rightarrow CD \in F^+$  o  $AB \rightarrow EGH \in F^+$
- $AB \rightarrow CD \in F$  o  $AB \rightarrow EGH \in F$
- $CD \rightarrow EGH \in F^+$  o  $EGH \rightarrow CD \in F^+$

## Bosquejo de solución

- a. Por definición de JSP el resultado es  $r$ .

- b. I.  $AB \rightarrow CD \in F^+$  o  $AB \rightarrow EGH \in F^+$

**SOL:** Por el teorema del Join Si Pérdida visto en clase, una de estas dos dependencias deben estar en  $F^+$  dado que la primera es de la forma forma:

$$R_1 \cap R_2 \rightarrow R_1 - R_2$$

y la segunda de la forma:

$$R_1 \cap R_2 \rightarrow R_2 - R_1$$

II.  $AB \rightarrow CD \in F$  o  $AB \rightarrow EGH \in F$

**SOL:** Que las dependencias deban pertenecer a  $F^+$  no nos dice nada si pertenecen a  $F$ . Por lo tanto no es posible afirmar nada sobre estas dependencias.

III.  $CD \rightarrow EGH \in F^+$  o  $EGH \rightarrow CD \in F^+$  **SOL:** Tampoco es posible afirmar nada sobre estas dependencias. El teorema nos dice nada sobre otras dependencias que no sean las indicadas en la parte

## Ejercicio 2 (8 puntos)

Se cuenta con el siguiente esquema relacional donde se representan personas y proyectos en los cuales trabajan.

**Personas** (idPersona, nroProy, nomProy, idResp, correo)

De las personas se conoce un identificador y sus direcciones de correo. De cada proyecto se conoce un número de proyecto, su nombre y el identificador del responsable. Las direcciones de correo no tienen relación con los proyectos en los que trabaja cada persona. Además, en este esquema se cumple la dependencia funcional  $nroproy \rightarrow nomproy, idresp$ . Considere la siguiente instancia válida de la relación **Personas**.

	idPersona	nroProy	nomProy	idResp	correo
$t_1$	1	101	sueldo	2	uno@mail.com
$t_2$	1	102	facturacion	3	uno@mail.com
$t_3$	2	101	sueldo	2	dos@mail.com
$t_4$	3	102	facturacion	3	tres@mail.com

Teniendo en cuenta la realidad planteada y la instancia válida dada, indique para cada una de las siguientes dependencias multivaluadas si se cumple, si no se cumple o si no es posible afirmar ninguna de las dos opciones. Justifique sus respuestas.

a.  $idPersona \rightarrow nroproy$

b.  $idPersona \rightarrow nroproy, nomproy, idResp$

## Bosquejo de solución

a.  $idPersona \rightarrow nroproy$

**SOL:** Esta dependencia multivaluada no se cumple, dado que, si se cumpliera, la instancia válida debería ser:

	idPersona	nroProy	nomProy	idResp	correo
$t_1$	<b>1</b>	<b>101</b>	<b>sueldo</b>	<b>2</b>	<b>uno@mail.com</b>
$t_2$	<b>1</b>	<b>102</b>	<b>facturacion</b>	<b>3</b>	<b>uno@mail.com</b>
$t_3$	2	101	sueldo	2	dos@mail.com
$t_4$	3	102	facturacion	3	tres@mail.com
$t_5$	<b>1</b>	<b>101</b>	<b>facturación</b>	<b>3</b>	<b>uno@mail.com</b>
$t_6$	<b>1</b>	<b>102</b>	<b>sueldo</b>	<b>2</b>	<b>uno@mail.com</b>

Esta instancia es imposible si se cumple que el nombre del proyecto y el responsable son únicos para una identificación de proyecto dada.

- b.  $idPersona \rightarrow nroproy, nomproy, idResp$  **SOL:** En la descripción de la realidad se dice que: *Las direcciones de correo no tienen relación con los proyectos en los que trabaja cada persona.*

Esto indica que dada una dirección de correo asociada a una persona que trabaje en diferentes proyectos, se debe repetir para cada uno de esos proyectos. Esto hace que la instancia dada en la letra sea válida con respecto a esta dependencia (o su complementaria).

### Ejercicio 3 (12 puntos)

Sea  $R(A,B,C,D,E,G,H)$  y  $F$  un conjunto de dependencias sobre  $R$ .

- a. Sea  $F = \{A \rightarrow BCE, E \rightarrow AD, ?df\}$ , donde  $?df$  es una dependencia funcional desconocida por el momento. Indique un posible valor para  $?df$  de forma tal que  $R$  se encuentre en BCNF según  $F$ . Justifique su respuesta.
- b. Considere ahora  $F = \{A \rightarrow BCE, E \rightarrow AD, B \rightarrow GH, G \rightarrow C, B \twoheadrightarrow EA|G\}$  y  $\rho = (R_1, R_2, R_3)$  donde  $R_1(A, E, B, G)$ ,  $R_2(E, D, G, C)$  y  $R_3(E, H, B)$
- I. Calcule  $\prod_{R_1}(F)$ ,  $\prod_{R_2}(F)$  y  $\prod_{R_3}(F)$
  - II. Determine cuál es la máxima forma normal en que se encuentran cada uno de los  $R_i$
  - III. Dar una descomposición de  $R_1$  en 4NF con JSP aplicando el algoritmo visto en el curso.

### Bosquejo de solución

- a. Por definición:

*Un esquema relacional  $R$  se encuentra en BCNF si, siempre que una  $df$   $X \rightarrow Y$  se cumple en  $R$ , entonces  $X$  es una superclave de  $R$ .*

Por lo tanto si se busca que el esquema  $R$  se encuentre en BCNF todos los atributos a la izquierda de las dependencias funcionales deben ser superclave.

Para verificar si esto se cumple, se procede a calcular las clausuras de  $A$  y  $E$  con el algoritmo visto en el teórico para  $F = \{A \rightarrow BCE, E \rightarrow AD\}$ :

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

Entonces  $A \rightarrow BCE$  y  $E \rightarrow AD$  violan BCNF por no ser ni  $A$  ni  $E$  superclaves. Se puede ver que el atributo  $A$  no determina a los atributos  $GH$  ni que  $E$  determina a los atributos  $GH$  por lo que se prueba  $?df = A \rightarrow GH$  y  $?df = E \rightarrow GH$

$A \rightarrow GH$ :

El conjunto de dependencias es  $F = \{A \rightarrow BCE, E \rightarrow AD, A \rightarrow GH\}$ . Se procede a calcular las clausuras de los atributos que se encuentran a la izquierda de las dependencias según el algoritmo visto en el curso.

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

Se puede ver que tanto  $A$  como  $E$  son superclave, y por lo tanto ninguna dependencia viola  $BCNF$  por lo que el esquema relacional  $R$  se encuentra en  $BCNF$ .

Análogamente, se puede mostrar que  $?df = E \rightarrow GH$  es válida. Notar que existen más dependencias que cumplen la propiedad de que el esquema relacional  $R$  se encuentre en  $BCNF$ . La condición fundamental es que  $GH$  aparezcan del lado derecho de la dependencia sin aparecer del izquierdo.

b. I. El mecanismo de proyección puede parecer oscuro. Hay algunas formas de realizarlo correctamente. La siguiente es una de ellas. Dado un esquema de relación  $R$

- 1) Para cada dependencia  $X \rightarrow A$  tal que  $X \subseteq R$  busco encadenamientos hacia adelante vía transitiva y pseudo transitiva obteniendo dependencias  $X \rightarrow A'$ . Para cada  $A' \in R$ , agrego  $X \rightarrow A'$  en la proyección sobre  $R$ . No es necesario agregar dependencias que se deduzcan de las que ya tenemos.
- 2) Para cada dependencia  $X \rightarrow A$  tal que  $A \in R$ , busco encadenamientos **hacia atrás** vía transitiva y pseudo transitiva obteniendo dependencias  $X' \rightarrow A$ . Para cada  $X' \subseteq R$ , agrego  $X' \rightarrow A$  en la proyección sobre  $R$ . No es necesario agregar dependencias que se deduzcan de las que ya tenemos.

De esta forma, en este caso particular que tenemos,

- En  $R_1(A, E, B, G)$  se proyectan las dependencias

$$F_1 = \{A \rightarrow EBG, E \rightarrow ABG, B \rightarrow G, B \rightarrow EA\}$$

. La última por definición de dependencia multivaluada embebida.

- En  $R_2(E, D, G, C)$  se proyectan las dependencias

$$F_2 = \{E \rightarrow DGC, G \rightarrow C\}$$

- En  $R_3(E, H, B)$  se proyectan las dependencias

$$F_3 = \{E \rightarrow HB, B \rightarrow H\}$$

II.  $R_1(A, E, B, G)$

Observando las dependencias se ve que tanto  $A$  como  $E$  determinan a los atributos restantes del esquema por lo que son claves. Por otra parte, se observa que en la dependencia multivaluada, no es posible aplicar la regla de combinación por lo que no hay más dependencias funcionales que las que se deducen de las propias funcionales existentes en  $F$ . Es por esto que  $B$  sólo determina a  $G$  que a su vez no determina a nadie más (no aparece nunca a la izquierda). Por todo esto, se puede afirmar que:

- $A$  y  $E$  son las únicas claves.
- $B$  y  $G$  son no primos.

Esto induce una dependencia transitiva desde las claves sobre  $G$ . Además, dado que las claves son de un solo atributo, tenemos garantía que no existen dependencias parciales desde las claves, por lo que el esquema  $R_1$  está en **2NF**.

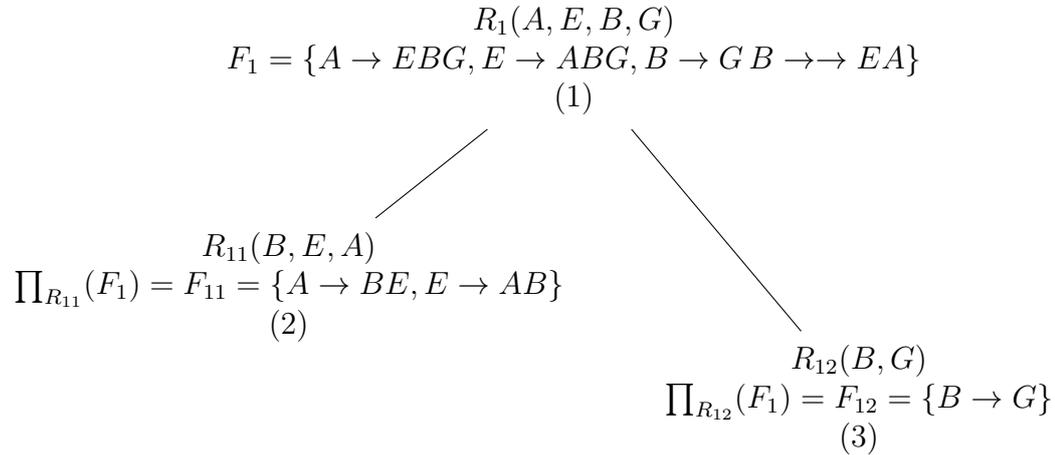
$R_2(E, D, G, C)$

Se observa que  $E$  determina a todos los atributos del esquema, por lo que es clave. Por otra parte,  $G$  determina sólo a  $C$ . En forma análoga a la parte anterior, se induce una dependencia transitiva de clave, no existiendo parciales por lo que,  $R_2$  se encuentra en **2NF**.

$$R_3(E, H, B)$$

En forma análoga a los anteriores se deduce que  $E$  es la única clave de  $R_3$ , se inducen dependencias transitivas de las claves y no hay parciales por lo que  $R_3$  se encuentra en **2NF**.

III. Se aplica el algoritmo para transformar un esquema relacional a  $4NF$  con  $JSP$ .



- (1)  $B \twoheadrightarrow EA$  viola  $4NF$  porque  $B$  no es superclave.  $(B)_{F_1}^+ = \{B, G\}$
- (2)  $R_{11}$  está en  $4NF$  dado que la dependencia multivaluada, aquí es trivial.
- (3)  $R_{12}$  está en  $4NF$ , dado que no existe ninguna dependencia que viole  $4NF$ .

### Ejercicio 4 (19 puntos)

a. Considere las siguientes transacciones:

$$T1: r_1(x), w_1(x), c_1$$

$$T2: r_2(x), w_2(x), c_2$$

$$T3: r_3(x), w_3(x), c_3$$

Considere además, el siguiente grafo de seriabilidad sobre estas transacciones:

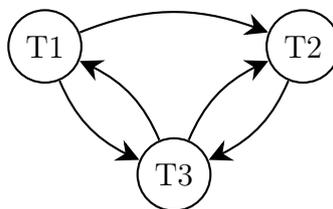


Figura 1: grafo de seriabilidad

- I. Considerando solamente las operaciones de las transacciones T1 y T2, ¿qué condiciones debe cumplir cualquier historia posible que genere el grafo de la Figura 1? . Justifique.
- II. Considerando solamente las operaciones de las transacciones T1 y T3, ¿qué condiciones debe cumplir cualquier historia posible que genere el grafo de la Figura 1? . Justifique.

- III. Considerando solamente las operaciones de las transacciones T2 y T3, ¿qué condiciones debe cumplir cualquier historia posible que genere el grafo de la Figura 1? . Justifique.
- IV. Escriba una historia completa que tenga ese grafo de seriabilidad en donde confirme cada transacción lo antes que sea posible.
- V. Clasifique la historia que escribió en la parte anterior con respecto a la recuperabilidad (no recuperable, recuperable, EAC, estricta). Justifique su respuesta.
- b. Considere el siguiente LOG, donde en las operaciones de escritura el primer número es el BFIM y el segundo el AFIM:

```

<Start,T1>
<Write,T1,X,20,30>
<checkpoint>
<Commit,T1>
<Start, T2>
<Write,T2,X,30,80>
<Commit,T2>
<Start,T3>
<Write,T3,X,80,10>
SYSTEM FAULT

```

- a) Indique qué transacciones se deben ignorar, cuáles se deben rehacer, cuáles deshacer y cuáles lanzar al final en una estrategia de **ACTUALIZACION DIFERIDA**.
- b) Indique qué transacciones se deben ignorar, cuáles se deben rehacer, cuáles deshacer y cuáles lanzar al final en las dos posibles estrategias de **ACTUALIZACION INMEDIATA**.

## Bosquejo de solución

### Parte a

- a. Considerando solamente las operaciones de las transacciones T1 y T2. Estas deben cumplir que  $w1(X)$  debe estar antes que  $r2(X)$  (siendo  $w1(X)$  y  $r2(X)$  las operaciones en conflicto). Esto basta para garantizar que sólo haya una arista de T1 a T2, ya que el orden de las operaciones de las transacciones no se debe alterar.
- b. Para el caso en el que se consideran solamente las operaciones de las transacciones T1 y T3 se debe garantizar que existan dos aristas (de T1 a T3 y viceversa). Para esto las operaciones de ambas transacciones deben estar entrelazadas.
- c. Para el caso en el que se consideran solamente las operaciones de las transacciones T2 y T3 se debe garantizar que existan dos aristas (de T2 a T3 y viceversa). Para esto las operaciones de ambas transacciones deben estar entrelazadas.
- d. La historia completa que se corresponde al grafo dado y que cada transacción confirma lo antes posible es la que se presenta a continuación:
- H:  $r1(X) r3(X) w1(X) c1 r2(X) w2(X) c2 w3(X) c3$

e. La historia es recuperable, EAC y es estricta.

### Parte b

a. En la actualización diferida los datos no se guardan en disco hasta realizar el commit. Por tanto, si hay una falla no es necesario deshacer ninguna operación.

En este caso no se deshace ninguna por ser una actualización diferida.  
No se ignora ninguna porque ninguna termina antes del checkpoint.  
T1 y T2 se rehacen y T3 se relanza.

b. En la actualización inmediata los datos se guardan en disco enseguida que se realizan las operaciones, es decir, no se espera a un commit. Por tanto, si hay una falla se deben deshacer todas las operaciones de la transacción.

En este caso se deshacen todas las operaciones por ser una actualización inmediata.  
No se ignora ninguna porque ninguna termina antes del checkpoint.

Para la estrategia **UNDO/REDO**

T1 y T2 se rehacen  
T3 se relanza

Para la estrategia **UNDO/No-REDO**

T1, T2 y T3 se relanzan

## Ejercicio 5 (15 puntos)

Considere el siguiente esquema relacional:

**Vehiculos** (mat, marca, modelo, año )

**Cientes** (ci, nombre, dir,tel)

**Alquileres** (ci, mat, fecha, duracion, costo ) donde

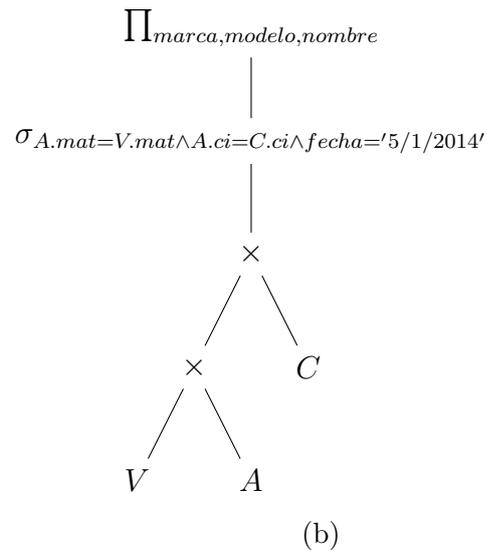
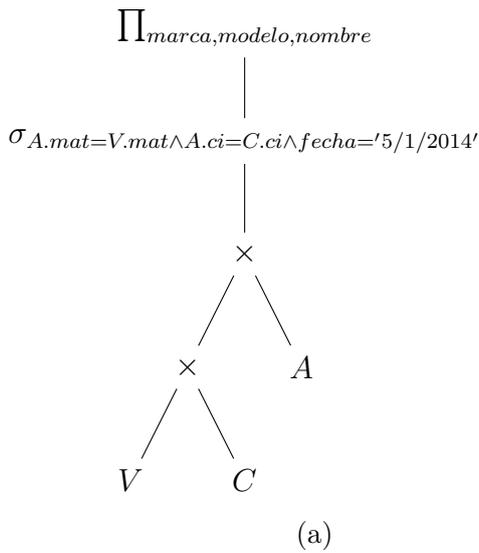
$$\prod_{ci}(Alquileres) \subseteq \prod_{ci}(Clientes)$$

$$\prod_{mat}(Alquileres) \subseteq \prod_{mat}(Vehiculos)$$

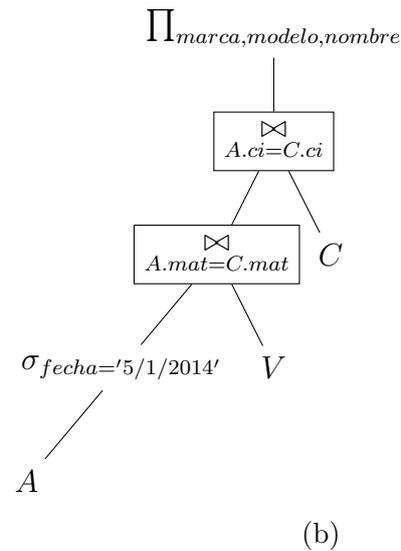
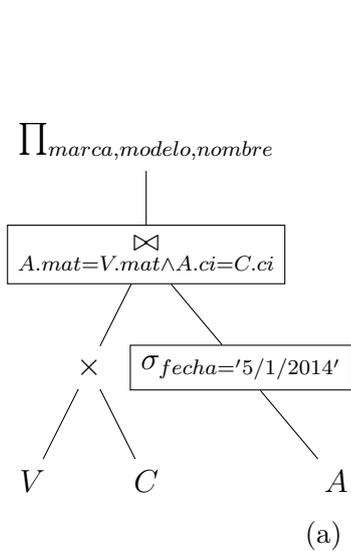
y la siguiente consulta en SQL

```
SELECT marca,modelo,nombre
FROM vehiculos V, clientes C, alquileres A
WHERE A.mat = V.mat and A.ci = C.ci and fecha = '5/1/2014'
```

a. Indique cuál de los siguientes árboles corresponde al árbol canónico. Justifique su respuesta.



b. Para los nodos recuadrados en los siguientes planes lógicos, indique qué algoritmos **NO** se pueden usar para construir planes físicos, considerando que todas las tablas tienen un índice primario por la clave. Justifique su respuesta.



### Bosquejo de solución