Segundo Parcial de Fundamentos de Base de Datos

Noviembre 2010

Presentar la resolución del parcial:

- 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.**

<u>Ejercicio 1</u> (25 puntos)

a) Sea un esquema R= (D, E, Q, P), y un conjunto de dependencias
F = {E→Q, PQ→D, D→E}.

1. Encontrar todas las claves candidatas, indicando el proceso realizado.
2. Encontrar un cubrimiento minimal para F.

***Solución*:**

1. P no está en ningún lado derecho)| pertenece a toda CC. Pero P+= {P} Entonces, buscamos las combinaciones de P con otro atributo.

(PQ)+= (PE)+= (PD)+ = R => las 3 son CC y son las únicas. El resto son superclaves.

2. F es un cubrimiento minimal de si mismo. Demotración: Todas las DF son de la forma X-> A. La única DF con 2 atributos en el lado izquierdo es PQ-> D. Como PQ es CC, ni P ni Q son redundantes. Todas las DF tienen lado derecho distintos, por lo cual ninguna es redundante.

b) Agregar al conjunto F definido en la parte a) la dependencia multivaluada P->>Q. Se pide:

1. Encontrar todas las claves candidatas, indicando el proceso realizado.
2. Encontrar un cubrimiento minimal para F.

Solución:

1. Por Axioma 8, P -> Q (sale por P ->> Q y E-> Q) .
Con esta nueva DF, P^+ = {P, Q, D, E} = R => La clave candidata es P y es única.

2. F_{min} = {E -> Q, P -> D, D -> E}

Demostración:

Todas las df son de la forma X -> A.
La única df con 2 atributos en el lado izquierdo es PQ-> D. En ella, Q es redundante por P->Q.
Finalmente, por P -> D, D -> E y E -> Q, P -> Q es redundante.

<u>Ejercicio 2</u> (10 puntos)

Dados:

- El esquema relacional R (A,B,C,E,G)
- Un conjunto de dependencias sobre R, F = {AC → E, G → A, B → E}
- Una descomposición de R, ρ = (R1(A,B,G,E), R2(G,B,C))

- Las siguientes consultas:

```
SELECT count(*)
 FROM R1 NATURAL JOIN R2
 WHERE   NOT EXISTS
         (SELECT *
          FROM R
          WHERE        R.A = R1.A  AND
                       R.B = R1.B  AND
                       R.C = R2.C  AND
                       R.E = R1.E  AND
                       R.G = R1.G
         )
```

- Una instancia cualquiera de R: r y sus correspondientes proyecciones en R1 y R2 : r1 = $\prod_{A,B,G,E}(r)$ y r2 = $\prod_{G,B,C}(r)$

Se pide: En caso de ser posible, determine el resultado de aplicar la consulta a las instancias r, r1 y r2. Justifique la respuesta.

Solución:

La consulta resuelve la cantidad de tuplas de (r1 * r2) – r, la cantidad de tuplas que pertenecen al join de r1 con r2 que no están en r.

Para determinar este resultado es necesario saber si la descomposición ρ es con JSP. Para esto se aplicará el teorema visto en el curso. ρ es con JSP sii $(R1 \cap R2) \rightarrow R1 - R2 \in F^+$ o $(R1 \cap R2) \rightarrow R2 - R1 \in F^+$

$(R1 \cap R2) = \{ B,G\}$
$R1 - R2 = \{A,E\}$
$R2 - R1 = \{C\}$

Para determinar si $BG \rightarrow AE \in F^+$ o $BG \rightarrow C \in F^+$ se calcula $(BG)^+_F$
$(BG)^+_F = \{ B,G,E,A\}$

Como $\{A,E\} \subseteq (BG)^+_F$ entonces es posible afirmar que $BG \rightarrow AE \in F^+$ y por aplicación del teorema se demuestra que ρ es una descomposición de R con JSP según F.

Por definición de JSP se cumple que para toda instancia r de R : r = $\prod_{R1}$ r * $\prod_{R2}$ r
Por lo tanto el resultado de la consulta es **0**.

<u>Ejercicio 3</u> (25 puntos)

Dado el esquema relación R (A,B,C,D,E,G,H) y el conjunto de dependencias F sobre R:

$F = \{ AB \rightarrow C, C \rightarrow E, GA \rightarrow H, E \rightarrow AD\}$

1) Dada la siguiente descomposición de R, ρ1 = (R1(A,D,B), R2(G,E,B,H), R3(C,D,H))
   a. Se afirma que R2 y R3 coinciden en la máxima forma normal que cumplen según F. ¿Esta afirmación es correcta? Justificar su respuesta.

   Solución:

   Para determinar la correctitud de esta afirmación se calcula la máxima forma normal en que se encuentran R2 y R3 y se comparan los resultados.

R2(G,E,B,H) , J2 = $\prod_{R2}(F)$ = { GE $\rightarrow$ H}

G^+ = {G}
E^+ = {E}
$(GE)^+$ = {G,E,A,D,**H**}
$(GBH)^+$ = {G,B,H}
$(EBH)^+$ = {E,B,H,A,D}

GEB pertenece a todas las claves de R2 según J2 ya que no pertenecen a los lados derechos de las df.

$(GBE)^+$ = {G,B,E,A,D,H}, por lo tanto GBE $\rightarrow$ GBEH

Entonces GBE es la única clave de R2 según J2. Por lo tanto los atributos primos son G,B y E.
GBE $\rightarrow$ H es una dependencia parcial de clave a un atributo no primo, ya que en GBE $\rightarrow$ H el atributo B es redundante y H no es primo, por lo tanto R2 no cumple las condiciones de 2NF.

R2 se encuentra en 1NF.

R3(C,D,H), J3 = $\prod_{R3}(F)$ = { C $\rightarrow$ D}

C^+ = {C,E,A,D}
$(DH)^+$ = {D,H}

CH pertenece a todas las claves de R3 según J3 ya que no pertenecen a los lados derechos de las df.

$(CH)^+$ = {C,H,D}, por lo tanto CH $\rightarrow$ CHD

Entonces CH es la única clave de R3 según J3. Por lo tanto los atributos primos son C y H.

CH $\rightarrow$ D es una dependencia parcial de clave a un atributo no primo, ya que en CH $\rightarrow$ D el atributo H es redundante y D no es primo, por lo tanto R3 no cumple las condiciones de 2NF.

R3 se encuentra en 1NF.

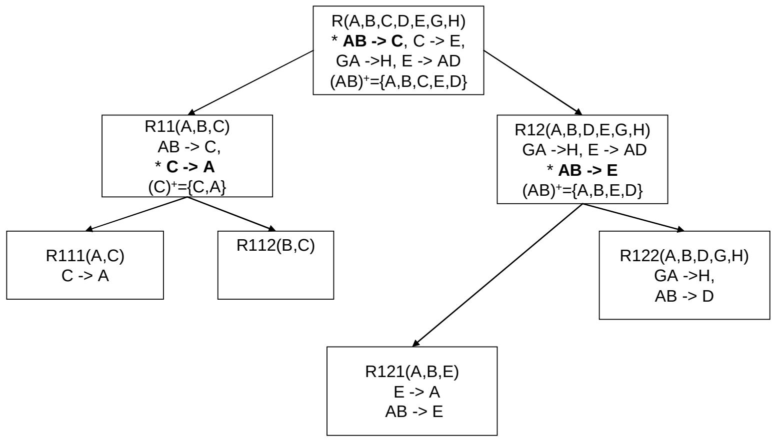
La máxima FN en que se encuentran R2 y R3 es 1NF por lo tanto **la afirmación es correcta.**

b. Determinar la máxima forma normal en que se encuentra $\rho 1$ según F. Justificar su respuesta.

Solución:

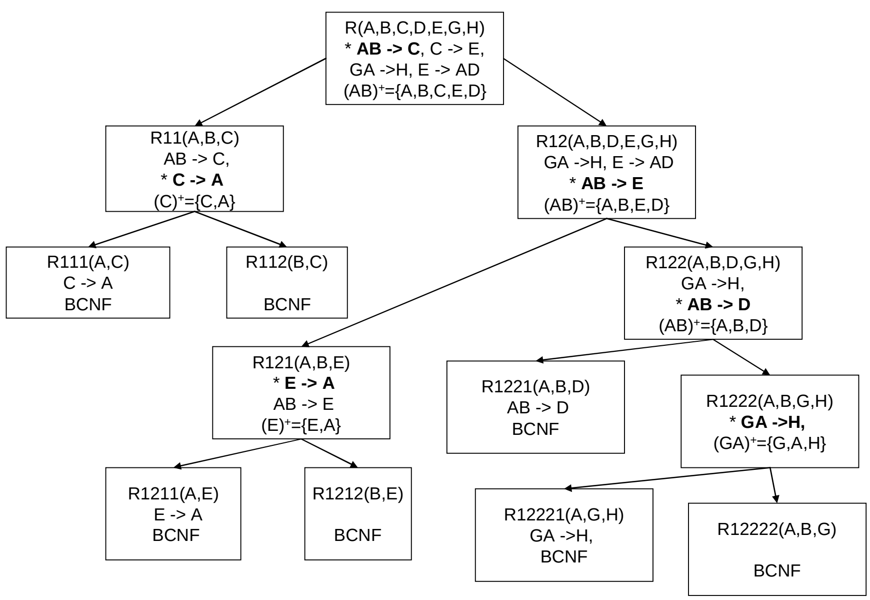
Como la máxima FN que se encuentra uno de los esquemas de $\rho 1$ es 1NF la descomposición se encuentra en 1NF (la FN de una descomposición es la mínima FN que se encuentran todos los esquemas de la descomposición).

2) Aplicando el algoritmo de descomposición en 4NF visto en el curso se obtuvo la siguiente descomposición:



Determinar si se aplicó el algoritmo en forma completa obteniéndose una descomposición en 4NF o si hace falta continuar descomponiendo. En caso de que la aplicación esté incompleta, completarla.

Solución:



Justificación:

R111(C,A) **R112(B,C)**

Todo esquema con 2 atributos está en BCNF (propiedad del práctico), por lo tanto R111 y R112 cumplen BCNF.

R111(C,A), $\prod_{R111}$ (F) = {C$\rightarrow$A}
R112(B,C), $\prod_{R112}$ (F) = {}

R121(A,B,E) **R122(A,B,D,G,H)**

R121(A,B,E) $\prod_{R121}$ (F) = {AB $\rightarrow$ E,E$\rightarrow$A}

E^+ = {E,A} $\neq$ R121
E no es superclave de R121 entonces E $\rightarrow$ A viola la condición de BCNF, se continúa con la descomposición:

R1211(E,A) **R1212(B,E)**

Todo esquema con 2 atributos está en BCNF (propiedad del práctico), por lo tanto R1211 y R1212 cumplen BCNF.

R1211(E,A) $\prod_{R1211}$ (F) = {E$\rightarrow$A}

R1212(B,E) $\prod_{R1212}$ (F) = { }

R122(A,B,D,G,H), $\prod_{R122}$ (F) = {AB $\rightarrow$ D, GA $\rightarrow$ H}

$(AB)^+$ = { A,B,E,D} $\neq$ R122

AB no es superclave de R122 entonces AB $\rightarrow$ D viola la condición de BCNF, se continúa con la descomposición.

R1221(A,B,D) R1222(A,B,G,H)

R1221(A,B,D) $\prod_{R1221}$ (F) = {AB $\rightarrow$ D}

AB es superclave de R1221 entonces R1221 esta en BCNF.

R1222(A,B,G,H), $\prod_{R1222}$ (F) = {GA $\rightarrow$ H}

$(GA)^+$ = { G,A,H} $\neq$ R1222

GA no es superclave de R1222 entonces GA $\rightarrow$ H viola la condición de BCNF, se continúa con la descomposición.

R12221(G,A,H) **R12222(A,B,G)**

R12221(G,A,H), $\prod_{R12221}$ (F) = {GA $\rightarrow$ H}
GA es superclave de R12221 entonces R12221 esta en BCNF.

R12222(A,B,G), $\prod_{R12222}$ (F) = {}
No se cumplen df no triviales por lo tanto R12222 esta en BCNF.

Por lo tanto la siguiente descomposición cumple las condiciones pedidas:

R111(C,A), $\prod_{R111}$ (F) = {C$\rightarrow$A}
R112(B,C), $\prod_{R112}$ (F) = {}
R1211(E,A), $\prod_{R1211}$ (F) = {E$\rightarrow$A}
R1212(B,E), $\prod_{R1212}$ (F) = { }
R1221(A,B,D), $\prod_{R1221}$ (F) = {AB $\rightarrow$ D}
R12221(G,A,H), $\prod_{R12221}$ (F) = {GA $\rightarrow$ H}
R12222(A,B,G), $\prod_{R12222}$ (F) = {}

Ejercicio 4 (puntos)

Sean las siguientes relaciones de la base de datos de una fábrica:

PRODUCCION(NumProducto, Modelo, Cantidad, Maquina)
PEDIDO(NumPedido, Cliente, Vendedor, ValorComision)
DETALLE-PEDIDO(NumPedido, NumProducto, Cantidad)

Con los siguientes valores:

	bf_R	n_R	Índice	Tamaño tupla (bytes)
PRODUCCION	100	200.000	NumProducto – Primario Modelo – Secundario B+	30
PEDIDO	200	50.000	NumPedido – Primario Cliente – Secundario B+ Vendedor – Secundario B+	15
DETALLE-PEDIDO	200	65.000	(NumPedido,NumProducto)	15

Tener en cuenta que:

- todos los índices primarios tienen 1 sólo nivel.
- todos los índices secundarios tienen 3 niveles.
- se dispone de 5 buffers de memoria.
- en la tabla **Producción** existen 200 modelos diferentes distribuidos uniformemente.
- en la tabla **Pedidos** existen 16 vendedores diferentes distribuidos uniformemente.
- el 85% de los pedidos de la tabla **Detalle-Pedido** corresponden a productos con modelos diferentes al modelo 'HP314'

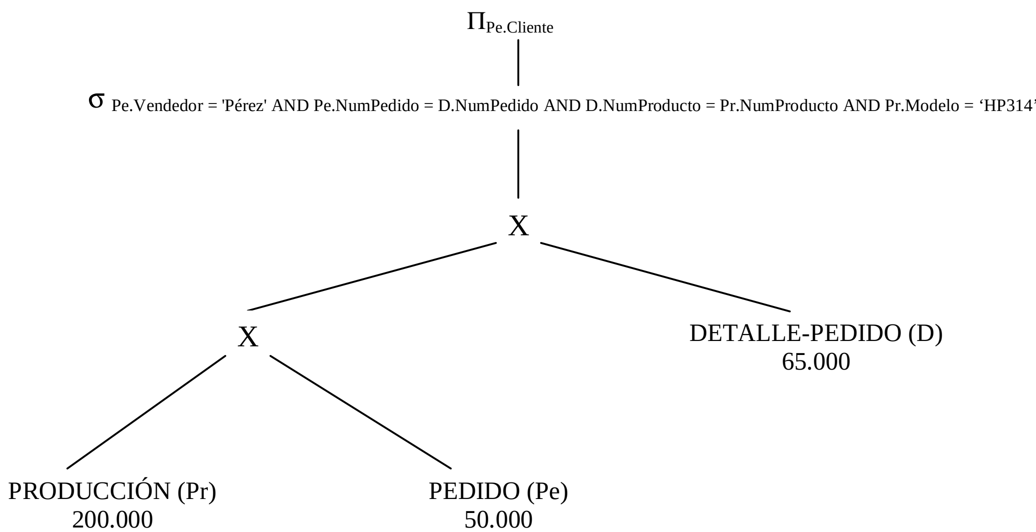
Dada la siguiente consulta:

"Encontrar los clientes que le compraron al vendedor Pérez productos del modelo HP314"

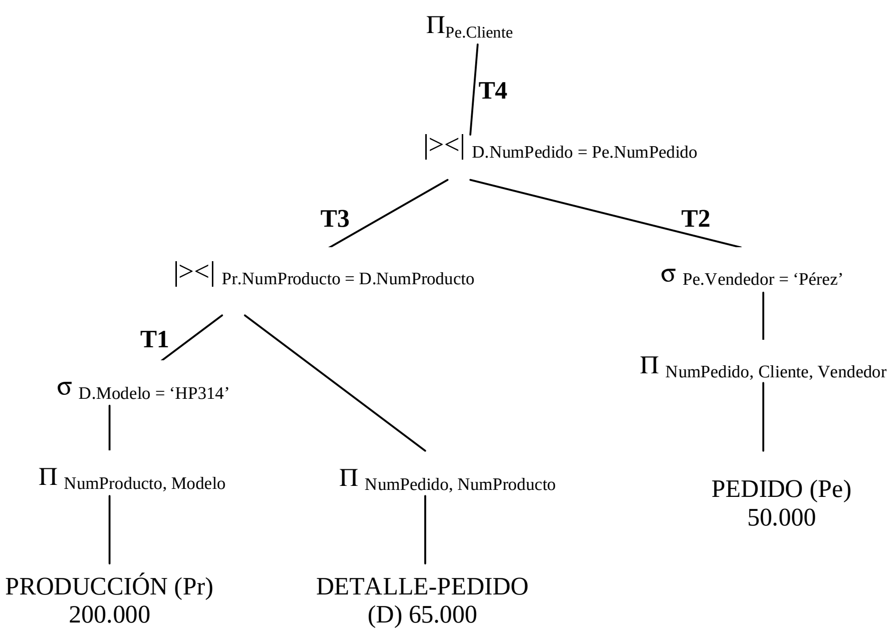
```
SELECT      Pe.Cliente
FROM        Producción Pr, Pedido Pe, Detalle-Pedido D
WHERE       Pe.Vendedor = "Pérez" AND
            Pe.NumPedido = D.NumPedido AND
            D.NumProducto = Pr.NumProducto AND
            Pr.Modelo = "HP314"
```

Se pide:

a) Construir el árbol canónico de la consulta.



b) Construir el plan lógico de la consulta, aplicando las heurísticas y calculando los tamaños intermedios.



$T1 = \sigma_{\text{Modelo} = \text{'HP314'}} Pr$
$T1 = |Pr| * (1/ V(Modelo, Pr))$ ➔ $T1 = \lceil 200.000/200 \rceil$
$V(Modelo, Pr) = 200$ **T1 = 1000**
$|Pr| = 200.000$

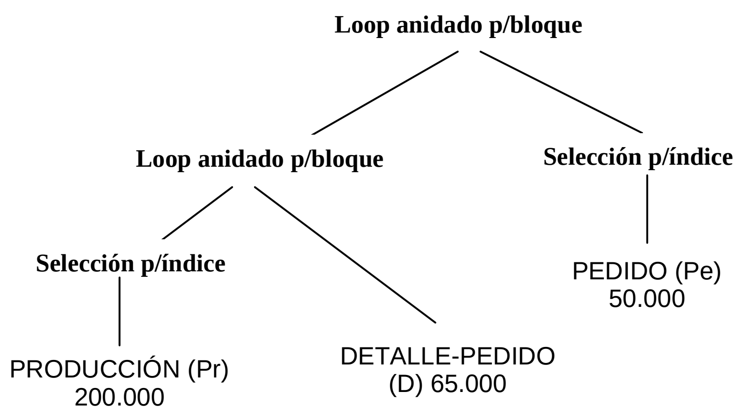
$T2 = \sigma_{\text{Vendedor} = \text{'Pérez'}} Pe$
$T2 = |Pe| * (1/ V(Vendedor, Pe))$ ➔ $T2 = \lceil 50.000/16 \rceil$
$V(Vendedor, Pe) = 16$ **T2 = 3125**
$|Pe| = 50.000$

$T3 = T1 \bowtie Pe._{NumPedido} = D._{NumPedido}$ ➔ $T3 = \lceil 65.000*0.15 \rceil$
$T3 = |D| * 0.15$ **T3 = 9750**

|D| = 65.000 y el 85% de los pedidos de la tabla Detalle-Pedido corresponden a productos con modelos diferentes al modelo ‘HP314’, por lo tanto, el 15% de los pedidos corresponden a productos con modelos ‘HP314’.

No alcanza la información dada para calcular el tamaño de T4, además, sólo se pide el cálculo de los tamaños intermedios.

c) Dar un plan físico para el plan lógico de la parte b.



Selección por índice:
- se utilizan los índices secundarios B+ tanto para la relación **Pedido** como para la relación **Producción.**

Loop anidado por bloque porque:
- para la relación **Detalle-Pedido** no se puede utilizar el índice por la clave
- para la relación **Pedido** no se puede utilizar el índice por clave porque los mismos se pierden después de aplicar la selección

<u>Ejercicio 5</u> (20 puntos)

Dadas las siguientes transacciones:

T1: w1(X), r1(Y), r1(Z), w1(Z), c1
T2: r2(X), r2(Y), w2(Y), c2

a)

a1) Dar dos historias entrelazadas de T1 y T2 tal que una evite abortos en cascada y la otra no sea recuperable. Justificar los resultados dados.

Solución:

H1: r2(X), r2(Y), w1(X), r1(Y), w2(Y), c2, r1(Z), w1(Z), c1 EAC

H2: w1(X), r1(Y), r2(X), r2(Y), r1(Z), w2(Y), c2, w1(Z), c1 NO RECUPERABLE

H1 evita abortos en cascada porque ninguna transacción lee valores escritos por transacciones que no hicieron commit. H2 no es recuperable porque T2 lee de T1 y T2 hace commit antes que T1.

a2) Dada la siguiente historia de T1 y T2:

H1: w1(X), r1(Y), r2(X), r2(Y), r1(Z), w1(Z), w2(Y), c2, c1

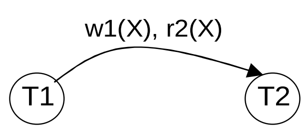
Decir si es serializable y en caso de que lo sea, dar la historia serial equivalente.

Solución:

Es serializable. Historia serial equivalente:

H: w1(X), r1(Y), r1(Z), w1(Z), c1, r2(X), r2(Y), w2(Y), c2

Grafo de seriabilidad de ambas historias:



b) Dar dos historias de T1 y T2 que sean equivalentes (pero no idénticas). Justificar explicando por qué lo son.

Solución:

Las historias de la parte a2):

H1: w1(X), r1(Y), r2(X), r2(Y), r1(Z), w1(Z), w2(Y), c2, c1

H: w1(X), r1(Y), r1(Z), w1(Z), c1, r2(X), r2(Y), w2(Y), c2

Son equivalentes porque sus operaciones en conflicto, w1(X), r2(X), se encuentran en el mismo orden en las dos historias.

c)

c1) Escribir T1 y T2 siguiendo el protocolo 2PL (utilizando locks de lectura y de escritura).

T1: wl1(X), w1(X), **rl1(Y)**, r1(Y), **rl1(Z)**, r1(Z), **wl1(Z)**, **u1(X)**, **u1(Y)**, w1(Z), **u1(Z)**, c1
T2: rl2(X), r2(X), **rl2(Y)**, r2(Y), **wl2(Y)**, **u2(X)**, w2(Y), **u2(Y)**, c2

c2) Dar una historia entrelazada de las T1 y T2 de la parte **c1)**. Decir si la historia es serializable, justificando.

H: **wl1(X)**, w1(X), **rl1(Y)**, r1(Y), **rl1(Z)**, r1(Z), **wl1(Z)**, **u1(X)**, **rl2(X)**, r2(X), **rl2(Y)**, r2(Y), **u1(Y)**, **wl2(Y)**, **u2(X)**, w2(Y), **u2(Y)**, c2, w1(Z), **u1(Z)**, c1