

FUNDAMENTOS DE BASES DE DATOS

Examen Febrero 2018

Solución

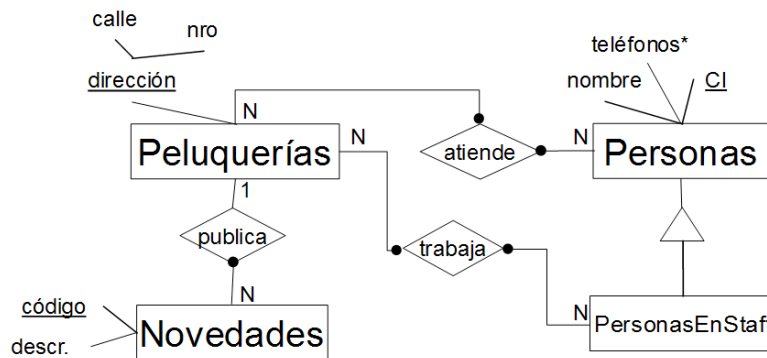
La duración del examen es de 3 horas y media.

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 (28 puntos).

Se presenta a continuación el modelo entidad-relación correspondiente a una cadena de Peluquerías.



Parte a

Indique si las siguientes afirmaciones son verdaderas o falsas. **Justifique.**

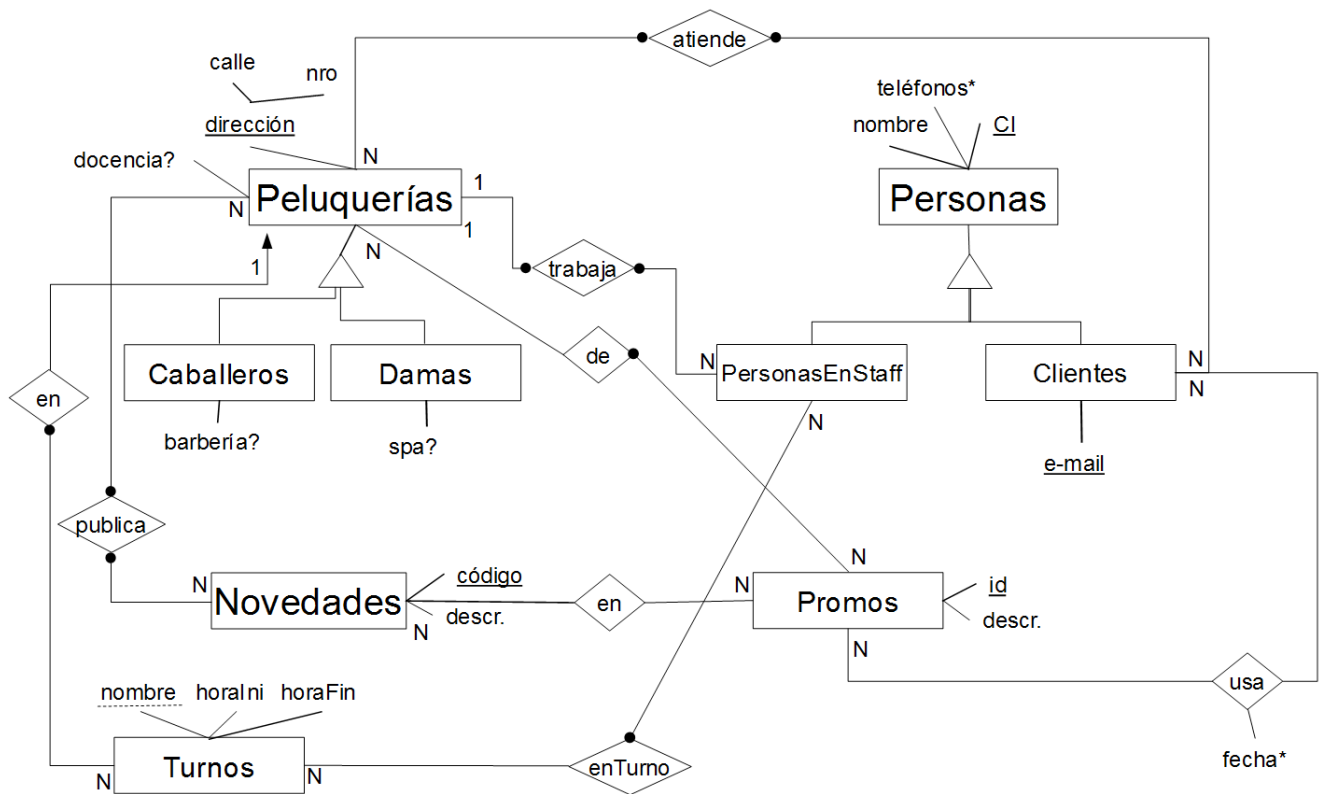
1. Hay Personas que no pertenecen a ningún staff.
V. Porque no hay ninguna RNE que diga lo contrario
2. Las Personas pueden atenderse en varias peluquerías.
V. Porque la cardinalidad de la relación *atiende*, del lado de las *Peluquerías* es N.
3. Las Personas tienen al menos un teléfono.
F. Porque el atributo *teléfonos* es multivaluado, lo que significa que la *Personas* podrían tener un conjunto de teléfonos y la cardinalidad mínima del mismo puede ser cero.
4. Toda Peluquería publica novedades.
F. Porque no hay totalidad en la relación *publica* del lado de las *Peluquerías*.

Parte b

El número de peluquerías comienza a crecer, por lo que se desea considerar los siguientes puntos:

- para las peluquerías que atienden sólo damas, interesa registrar si cuenta con servicio de Spa.
- para las peluquerías que atienden sólo caballeros, se registra si cuentan con servicio de barbería.
- interesa saber en qué peluquerías se dan clases.
- las peluquerías van a hacer promociones, por lo que interesa registrar el e-mail de las personas que son clientes. Se sabe que éste también identifica a los clientes.
- cada peluquería ofrece promociones (promos) e interesa el identificador de cada promo y su descripción.
- hay novedades que pueden no estar asociadas a las promos.
- una promo puede pertenecer a varias peluquerías y puede estar asociada a varias novedades.
- ahora las personas del staff deben trabajar en una sola peluquería.
- las personas que pertenecen a un staff pueden ser clientes de cualquier peluquería
- interesa saber si las promos asociadas a novedades son más usadas, para esto se registra la promo, el cliente y la fecha en que fue usada dicha promo por el cliente.
- un cliente puede usar una promo más de una vez.
- en las peluquerías se realizan distintos turnos horarios (matutino, vespertino y nocturno). Cada turno tiene un nombre, una hora de comienzo y una hora de fin, dichas horas pueden variar según la peluquería. Interesa saber qué turno o turnos hace cada empleado en la peluquería donde trabaja.

Realice al modelo dado en la **Parte a**, todas las modificaciones que considere necesarias.



- Clientes U Staff = Personas
- Caballeros \cap Damas = \emptyset
- La promo que usa el cliente es de la peluquería en la que se atiende
- Para todo $\langle s, t \rangle$ perteneciente a EnTurno, existe p perteneciente a Peluquería tal que $\langle s, p \rangle$ pertenece a Trabaja y $\langle t, p \rangle$ pertenece a Horarios.

Ejercicio 2 (28 puntos).

Una empresa de automóviles de alquiler posee una base de datos para la gestión del mantenimiento que se le realiza a éstos, la cual tiene el siguiente esquema:

AUTOS (mat, marca, modelo, año)

Representa los automóviles de la empresa.

MECÁNICOS (ci, nombre, teléfono, dirección)

Representa los mecánicos que trabajan en el mantenimiento de los autos.

REPUESTOS (código, descripción, precioLista)

Contiene el catálogo de repuestos con los que cuenta el taller para utilizar en los mantenimientos, por ejemplo filtro de aire, bomba de aceite, etc. De cada repuesto se conoce un código, su descripción y el precio de lista.

MANTENIMIENTOS (mat, ci, fecha, km)

Contiene el registro de los mantenimientos realizados a cada auto. De cada uno se registra la matrícula del auto, la cédula del mecánico responsable, la fecha y la cantidad de kilómetros recorridos por el auto hasta ese momento.

INSUMOS (mat, ci, fecha, codRep, cantidad)

Contiene el registro de los repuestos utilizados en los mantenimientos. De cada repuesto se conoce la cantidad utilizada (que es un entero mayor o igual que 1).

En este esquema no existen tablas vacías y se cumplen las siguientes dependencias de inclusión.

- $\Pi_{mat}(\text{MANTENIMIENTOS}) \subseteq \Pi_{mat}(\text{AUTOS})$
- $\Pi_{ci}(\text{MANTENIMIENTOS}) \subseteq \Pi_{ci}(\text{MECÁNICOS})$
- $\Pi_{mat,ci,fecha}(\text{INSUMOS}) \subseteq \Pi_{mat,ci,fecha}(\text{MANTENIMIENTOS})$
- $\Pi_{codRep}(\text{INSUMOS}) \subseteq \Pi_{código}(\text{REPUESTOS})$

Resolver en álgebra relacional las siguientes consultas:

- Obtener la cédula y el nombre de los mecánicos que utilizaron un único repuesto en todos sus mantenimientos.

Nota: Los mecánicos devueltos tienen que haber utilizado al menos un repuesto en algún mantenimiento.

Solución

$$A = \rho_{ci \rightarrow ci1, codRep \rightarrow cR} INSUMOS$$

$$B = INSUMOS \bowtie_{ci=ci1 \wedge codRep \neq cR} A$$

$$C = \Pi_{ci} INSUMOS - \Pi_{ci} B$$

$$RES = \Pi_{ci, nombre} (MECANICOS * C)$$

- Obtener la cédula de los mecánicos que mantuvieron todos los mismos autos que el mecánico con cédula 123456 y que además mantuvieron algún auto de la marca Chevrolet.

Solución

$$A = \Pi_{mat} (\sigma_{ci=123456} (\text{MANTENIMIENTOS}))$$

$$B = \Pi_{ci, mat} (\text{MANTENIMIENTOS}) \% A$$

$$C = \sigma_{marca='Chevrolet'} (\text{AUTOS})$$

$$D = \Pi_{ci} (\text{MANTENIMIENTOS} * C)$$

$$RES = B * D$$

Resolver en cálculo relacional las siguientes consultas:

3. Dar cedula y nombre de los mecánicos que han utilizado todos los repuestos existentes en un mantenimiento.

Solución

$$\{ \langle m.ci, m.nombre \rangle / Mecánicos(m) \wedge \\ (\exists ma) (Mantenimientos(ma) \wedge m.ci = ma.ci \wedge \\ (\forall r) (Repuestos(r) \rightarrow (\exists i) (Insumos(i) \wedge i.ci = ma.ci \wedge \\ i.mat = ma.mat \wedge i.fecha = ma.fecha \wedge \\ i.codRep = r.codigo))) \\ \}$$

Resolver en SQL, sin utilizar vistas ni sub-consultas en el FROM, las siguiente consulta:

4. Obtener la matricula de los autos que tuvieron algún mantenimiento en donde se hayan utilizado mas de 3 repuestos diferentes.

Solución

```
SELECT matricula
FROM AUTOS a
WHERE EXISTS (
    SELECT i.mat, i.ci, i.fecha
    FROM INSUMOS i
    WHERE i.mat = a.mat
    GROUP BY i.mat, i.ci, i.fecha
    HAVING COUNT(distinct i.codRep) >3)
```

Ejercicio 3 (29 puntos)

Una empresa que organiza espectáculos internacionales en Uruguay dispone de una base de datos para registrar los eventos que organiza. Esta base de datos tenía una única tabla:

Eventos (org, ne, dir, art, fecha, vtaE, cve)

La descripción de los atributos es la siguiente:

Org (Organizador): Es el nombre de la empresa encargada de organizar el espectáculo. Un mismo organizador puede organizar diferentes eventos, pero un evento es organizado por un único organizador.

Ne (Nombre Evento): Es el nombre del evento. Por ejemplo: Tour 2018 de Julieta Venegas. Debido a restricciones comerciales, el nombre del evento no puede repetirse.

Dir (dirección): Es la dirección donde se desarrolla el evento. Se asume que no hay dos eventos en la misma dirección y fecha, y un evento se realiza en una sola dirección independientemente de la fecha.

Art (Artista): Es el nombre del artista. En un evento solo actuará un artista y un artista no puede actuar en dos eventos en la misma fecha. Los artistas no tienen necesariamente contratos exclusivos con los organizadores, por lo tanto, un mismo artista puede encabezar eventos de diferentes organizadores.

Fecha: Fecha del evento. Un mismo evento puede efectuarse en diferentes fechas y en una misma fecha pueden efectuarse diferentes eventos.

VtaE (Canal de Venta de Entradas): Es el canal de venta de las entradas del evento, por ejemplo, TickAntel, RedUTS, Abitab. Un evento puede tener varios canales de venta asociados. Las entradas de "Tour 2018 de Julietas Venegas" puede venderlas tanto TickAntel como RedUTS al mismo tiempo. Los canales de venta, siempre venden entradas para todas las fechas en que se realiza un evento.

CVE (Comisión de Canal de Venta de Entradas): Es la comisión, representada como un porcentaje sobre el valor de las entradas, que cobra el canal de venta al organizador del evento. La comisión es independiente del evento, esto es, el canal de venta siempre cobra la misma comisión a cada organizador, sin embargo, puede cobrar diferentes comisiones a diferentes organizadores.

- Determine un cubrimiento (no necesariamente minimal) de todas las dependencias funcionales existentes teniendo en cuenta que hay cuatro lados izquierdos diferentes.
- Determine al menos una multivaluada embebida que no sea funcional o complementaria de funcional.
- En una reestructura de la base de datos, construyeron las siguientes tablas:
A (org, ne)
B (ne, dir, art, fecha)
C (ne, vtaE, cve)
Indique en qué forma normal está este nuevo esquema.
- Encuentre una descomposición en 3NF con JSP y sin pérdida de dependencias funcionales del esquema original y de acuerdo a su respuesta en la parte a). Aplique el algoritmo correspondiente visto en el curso.

Solución:

a) El esquema es **Eventos** (org, ne, dir, art, fecha vtaE, cve)

Las dependencias funcionales son las siguientes:

$ne \rightarrow org, art, dir$

- Un evento es organizado por un único organizador
- En un evento actúa solo un artista
- Un evento se puede desarrollar en una única locación

$dir, fecha \rightarrow ne$

- No hay dos eventos en una misma fecha y dirección

$art, fecha \rightarrow ne$

- Un artista no puede actuar en dos eventos en la misma fecha
- En un evento solo actúa un artista

$org, vtae \rightarrow cve$

- La comisión de venta es determinada por el canal de venta para cada organizador

b) La siguiente es una dependencia multivaluada embebida, no es funcional ni complementaria de funcional:

$ne \twoheadrightarrow fecha \mid vtae$

c) El primer paso para determinar en que forma normal está cada tabla es determinar que dependencias funcionales se cumplen sobre cada una de ellas. Por eso es necesario realizar las proyecciones del conjunto de dependencias:

A (org, ne)

B (ne, dir, art, fecha)

C (ne, vtaE, cve)

$\prod_A(F) = \{ne \rightarrow org\}$

$\prod_B(F) = \{ne \rightarrow art; ne \rightarrow dir; dir, fecha \rightarrow ne; art, fecha \rightarrow ne\}$

$\prod_C(F) = \{vtaE, ne \rightarrow cve\}$

Observe que en el caso de la proyección sobre C, la dependencia funcional $vtaE, ne \rightarrow cve$ se infiere a partir de $ne \rightarrow org$ y $org, vtae \rightarrow cve$

De esta forma:

- A está al menos en BCNF
En $ne \rightarrow org$, ne es superclave de A
- B está en 2NF
{ne, dir, fecha} es clave de B, pues $\{ne, dir, fecha\}^+ = \{ne, dir, fecha, art\}$
Para $ne \rightarrow dir$ ne no es superclave de B, por lo tanto no cumple BCNF. El atributo dir es primo, por lo tanto cumple 3NF
Para $ne \rightarrow art$ ne no es superclave de B, por lo tanto no cumple BCNF. El atributo art no es primo, por lo tanto tampoco cumple 3NF.
- C está al menos en BCNF
En $vtaE, ne \rightarrow cve$ {vtaE, ne} es superclave de C, por lo tanto cumple BCNF.

d) El conjunto de dependencias es el siguiente:

$ne \rightarrow org, art, dir$

$dir, fecha \rightarrow ne$

$art, fecha \rightarrow ne$

$org, vtae \rightarrow cve$

El primer paso indica encontrar un cubrimiento minimal de este conjunto de dependencias:

Paso 1. $G := F$

Paso 2. Separar dependencias:

$ne \rightarrow org$
 $ne \rightarrow art$
 $ne \rightarrow dir$
 $dir, fecha \rightarrow ne$
 $art, fecha \rightarrow ne$
 $org, vtae \rightarrow cve$

Paso 3. Eliminar atributos redundantes en el lado izquierdo de las dependencias.

$dir^+ = \{dir\}$
 $fecha^+ = \{fecha\}$
 $org^+ = \{org\}$
 $vtae^+ = \{vtae\}$

Por lo que no hay atributos redundantes en los lados izquierdos de las dependencias.

Paso 4. Eliminar dependencias redundantes

$ne \rightarrow org : ne^+ = \{ne, art, dir\}$
 $ne \rightarrow art : ne^+ = \{ne, org, dir\}$
 $ne \rightarrow dir : ne^+ = \{ne, org, art\}$
 $dir, fecha \rightarrow ne : \{dir, fecha\}^+ = \{dir, fecha\}$
 $art, fecha \rightarrow ne : \{art, fecha\}^+ = \{art, fecha\}$
 $org, vtae \rightarrow cve : \{org, vtae\}^+ = \{org, vtae\}$

Por lo que no hay dependencias funcionales redundantes.

Por lo tanto, los esquemas resultantes son los siguientes:

A (ne, org, art, dir)
B (dir, fecha, ne)
C (art, fecha, ne)
D (org, vtae, cve)

Ahora falta revisar si hay alguna clave de la tabla original incluida:

Por lo que se calculan las claves de la tabla original:

SoloIzq={fecha,vtae}
SoloDer={cve}
IzqDer={ne,art,dir,org}

fecha,vtae no es una clave porque $\{fecha,vtae\}^+ = \{fecha,vtae\}$

Por lo que las claves posibles deben surgir de combinar estos atributos con los que están de ambos lados.

Es así que se detectan 3 claves:

{fecha,vtae,ne}
{fecha,vtae,art}
{fecha,vtae,dir}

La restante combinación de 3 atributos no es superclave y cualquier otra combinación más grande incluye a alguna de las claves anteriores, por lo que son sólo estas tres claves.

Si observamos cada una de las tablas A,B,C,D, detectamos que ninguna contiene a alguna de las claves, por lo que hay que agregar un esquema más (E) con alguna de las claves, por ejemplo:

E(fecha,vtae,ne)

El esquema final en 3NF tiene las 5 tablas.

Ejercicio 4 (15 puntos)

Considere el siguiente esquema de la base de datos que lleva datos de compradores frecuentes de un grupo de tiendas.

Tiendas(idTienda, nomTienda, dirTienda, barrioTienda, deptoTienda): Datos de las tiendas.

Cientes(CI, nomCliente, dirCliente, barrioCliente, deptoCliente): Datos de los clientes.

CientesFrecuentes(CI, idTienda, fechaCompra, monto): Datos de las compras realizadas por cada cliente en cada tienda.

Se sabe que en cada tabla de este esquema existe un índice primario por la clave primaria correspondiente. Además, se conocen los índices y la cantidad de tuplas de cada tabla:

Tabla	Tipo	Atributo	Tuplas
Tiendas	Secundario	deptoTienda	500
Cientes	-	-	40.000
CientesFrecuentes	Secundario	fechaCompra	1.500.000

Notas:

- los clientes se distribuyen en las tiendas uniformemente.
- el 40% de las compras se hicieron el 24/12/2017.
- el 50% de las compras realizadas el 24/12/2017 fueron mayores de \$2000.
- el 20% de las tiendas están ubicadas en Montevideo.

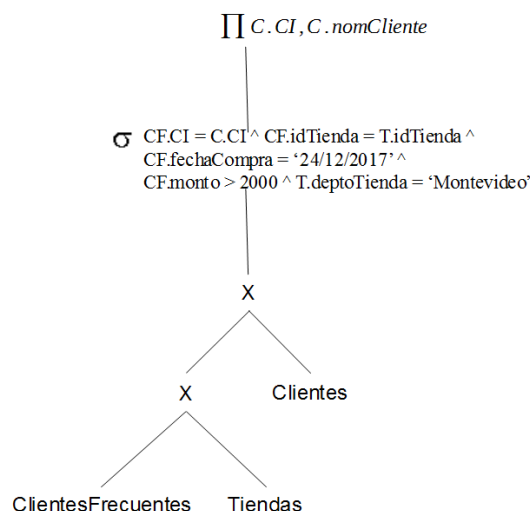
Considere la siguiente consulta sobre el esquema dado:

```

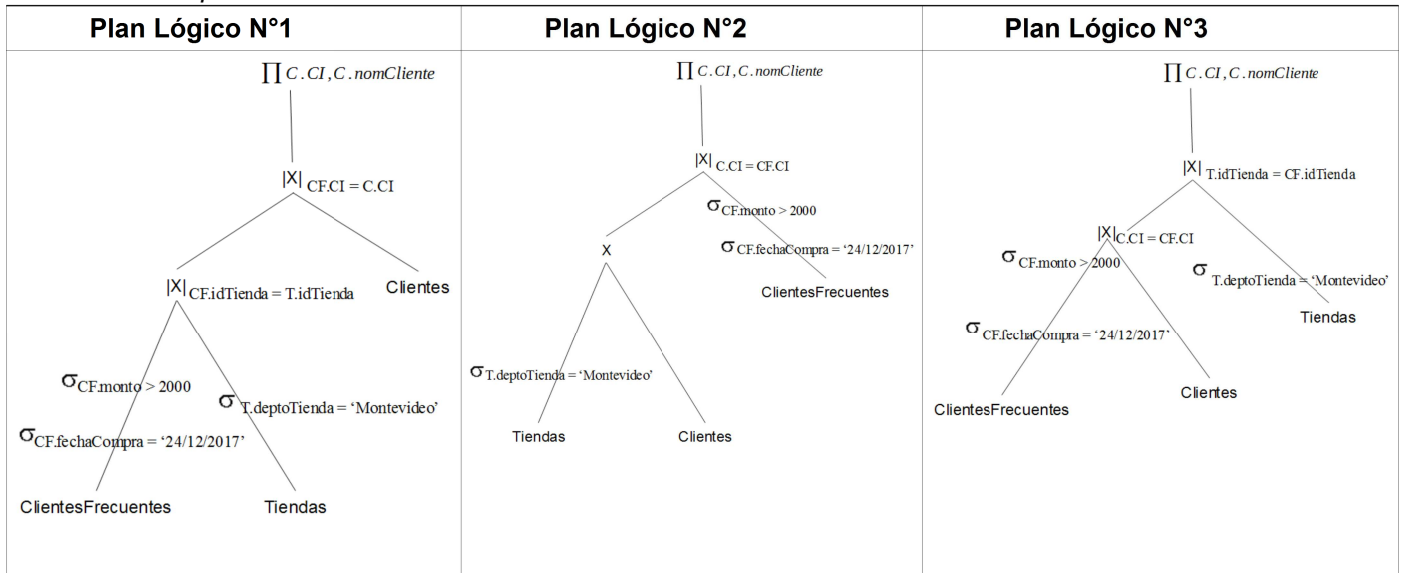
SELECT C.CI, C.nomCliente
FROM CientesFrecuentes CF, Tiendas T, Cientes C
WHERE CF.CI = C.CI AND CF.idTienda = T.idTienda AND CF.fechaCompra = '24/12/2017'
AND CF.monto > 2000 AND T.deptoTienda = 'Montevideo';
    
```

Se pide:

1. Construya el árbol canónico de la consulta dada.



2. **Justificando su elección**, indique cuál de los siguientes planes lógicos es **el más conveniente según las heurísticas**. Además, **explique brevemente** por qué no considera convenientes a los otros 2 planes lógicos.



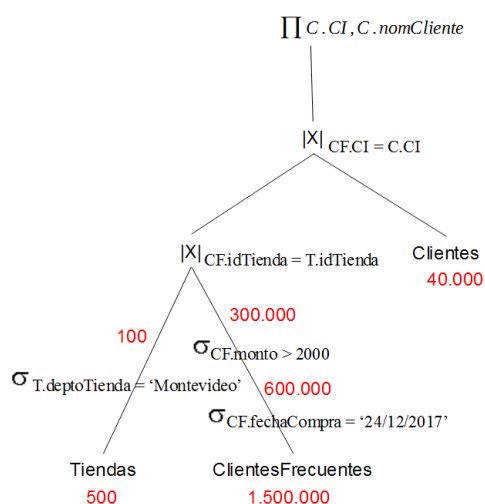
El plan lógico más conveniente es el N°1, porque está balanceado a la izquierda y es el que más respeta las heurísticas y las mismas mencionan que se debe:

- 1- cambiar las selecciones conjuntivas por una “cascada” de selecciones simples.
- 2- mover las selecciones lo más abajo que se pueda en el árbol.
- 3- Poner a la izquierda de los productos las hojas que generen menos tuplas, asegurando que el orden de las hojas no cause operaciones de producto cartesiano (que no pueden convertirse en join).
- 4- cambiar secuencias de selecciones y productos por join's.
- 5- mover las proyecciones lo más abajo posible en el árbol, agregando las proyecciones que sean necesarias.

Los planes lógicos no respetan exactamente el punto 3. En el caso de los planes N.º 1 y N.º 3 no colocan a la izquierda las hojas que generan menos tuplas, mientras que el plan N.º 2 presenta un producto cartesiano entre las tablas Tiendas y Clientes. Finalmente, el problema que presenta el plan lógico N°1 es el menos grave, por lo tanto, éste es el más conveniente según las heurísticas.

3. ¿Considera que es posible obtener un plan lógico más conveniente, según las heurísticas, que el seleccionado en la Parte 2? **Justifique su respuesta.** Si la respuesta es afirmativa, dibuje el plan lógico más conveniente.

Como se mencionó en la parte anterior, el plan lógico N°1 no respeta exactamente el paso 3 de las heurísticas, por lo tanto es posible obtener un plan lógico más conveniente, y para esto deberíamos dejar la tabla Tiendas lo más a la izquierda posible. Esto es así porque se debe tener en cuenta la cantidad de tuplas de cada tabla y las Notas dadas en la letra. Finalmente, el plan lógico más conveniente es el que se presenta a continuación:



4. De acuerdo con la respuesta de la Parte 3, construya el plan físico correspondiente al plan lógico más conveniente.

El plan físico para el plan lógico obtenido en la Parte 3 es el siguiente:

