

FUNDAMENTOS DE BASES DE DATOS

Examen Marzo 2003

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.

Parte 1 – Modelo Entidad-Relación (25 puntos)

Ejercicio 1.

Una empresa le solicita a Ud. que diseñe una base de datos para llevar encuestas sobre la programación radial. Lo que sigue es el resultado del relevamiento realizado por los analistas de la empresa.

Existe un conjunto de radios, de las que se conoce su característica (que las identifica), un nombre, la frecuencia en la que transmite y el tipo de transmisión (AM– FM).

Cada radio, emite programas que están identificados por un nombre y de los que se conoce un género (periodístico, deportivo, musical, cultural, variedades). Cada programa puede ser emitido por diferentes radios. Es importante destacar que cada radio puede realizar más de una emisión de cada programa, registrando para cada emisión la fecha, la hora de comienzo, la duración y si es una repetición o no.

A veces, para una emisión de un programa, se realiza una encuesta en la que se registran los siguientes datos: la cantidad total de encuestados, la cantidad de aprobaciones, la cantidad de rechazos y la cantidad de indiferencias. Obviamente, la empresa espera que estas cantidades cierren correctamente.

Los programas, además, pueden ser programas de “resumen” en cuyo caso interesa saber de qué otros programas toma las notas. Cada programa es producido por un único consorcio compuesto por una de las radios que lo emite y una compañía productora. De las compañías productoras se conoce su nro. del Registro Único de Contribuyentes (RUC), un nombre y sus teléfonos. No se admite que dos productoras puedan tener igual nombre.

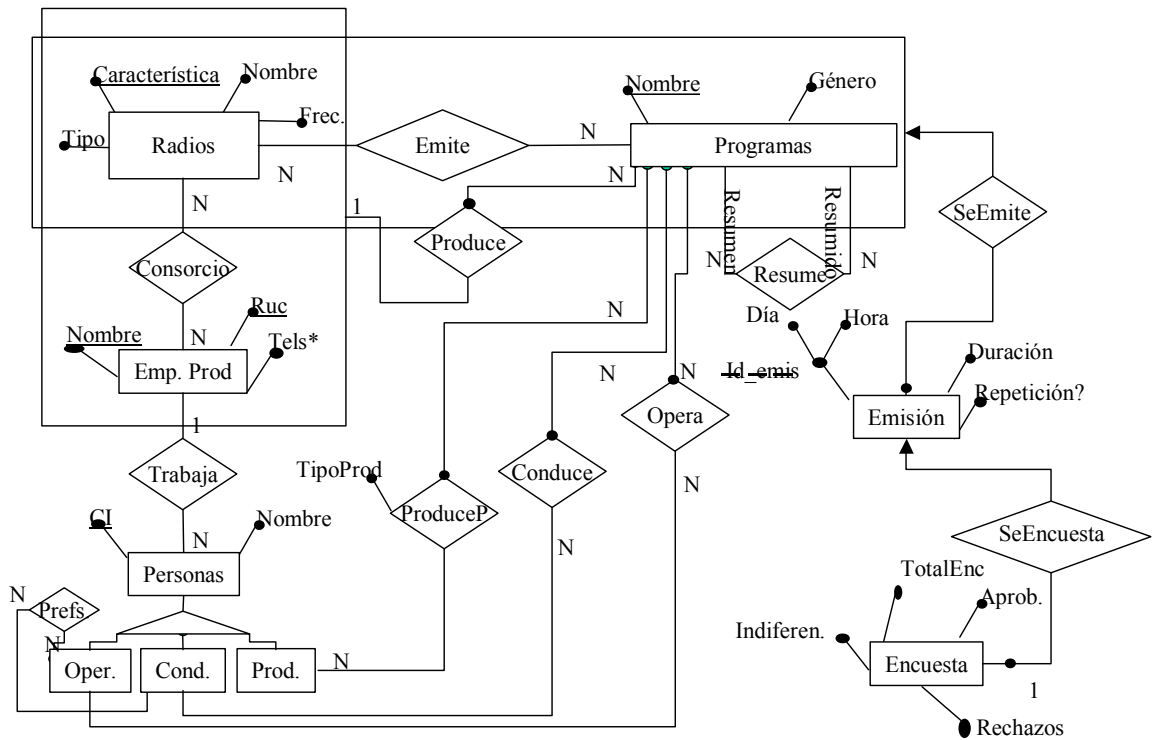
En cada productora, trabajan personas relacionadas directamente con los programas, de las que se conoce su cédula de identidad y su nombre. Estas personas pueden ser Conductores, Operadores o Productores. Las productoras no permiten que una persona trabaje en más de una de ellas. Interesa registrar para cada programa, quiénes son los conductores, los operadores y los productores. Los productores, en cada programa en el que trabajan, pueden realizar producción periodística o comercial. Por otra parte, hay algunos conductores que tienen preferencia por trabajar con algunos operadores. Es claro que todos los programas tienen al menos un conductor, un productor y un operador, aunque podrían tener varias personas para cada función.

SE PIDE: Modelo Entidad Relación completo, incluyendo restricciones de integridad.

Solucion

RNE:

- La suma de las cantidades de rechazos, aprobaciones e indiferencias debe ser igual a la cantidad de encuestados.
- La radio que emite cierto programa debe ser parte del consorcio que lo produce.
- Todos los empleados que trabajan en un programa deben ser empleados de la compañía productora que forma parte del consorcio que produce dicho programa.
- Un programa no es resumen de si mismo
- Si un conductor prefiere a un operador, ambos trabajan para la misma compañía productora



Parte 2 – Diseño Relacional (25 puntos)

Ejercicio 2 (15 puntos)

De una realidad se conocen las siguientes dependencias funcionales:

A → B, C, D
B → A
D → C
E → D
F → GH
I → A

Alguien diseñó el siguiente esquema para esta realidad.

Tabla 1

A	B	C	D	E
---	---	---	---	---

Tabla 2

F	G	H	I	C	D	E
---	---	---	---	---	---	---

Pregunta 1 -

Dar todas las claves candidatas de cada tabla. Justificar.

Pregunta 2 -

Dar la forma normal del esquema. Justificar.

Pregunta 3 -

Pasar el esquema a BCNF (usando el algoritmo visto en el curso) y dar todas las claves candidatas de cada una de las tablas del esquema en BCNF. Justificar.

Pregunta 4 -

Determinar si se perdieron dependencias funcionales y en caso de que se pierdan indicar cuales.

Solución

Pregunta 1

Las claves candidatas de la tabla 1 son AE y BE.
Las únicas clave candidata de la tabla 2 es FIE.

Pregunta 2

En la tabla 1 la dependencia funcional $A \rightarrow C$ viola 2NF.
En la tabla 2 la dependencia funcional $F \rightarrow G$ viola 2NF.

El esquema está en 1NF ya que ambas tablas tienen atributos solo con valores atómicos e individuales.

Pregunta 3

Hay varias formas posibles de descomponer este esquema. A continuación se da una de ellas.

Tabla 1

R(ABCDE)

Se proyectan las siguientes dependencias:

A → BCD
B → ACD
D → C
E → DC

A → BCD viola BCNF entonces divido en
R1 (ABCD) R2 (AE)

En R1 se proyectan las siguientes dependencias:

A → BCD

B → ACD

D → C

Y las claves candidatas son A y B

D → C viola BCNF entonces divido en:

R11 (DC) R12 (ABD) R2 (AE)

En R11 se proyecta la siguiente dependencia:

D → C

La clave de esta tabla es D y está en BCNF ya que no hay dependencia que viole esta forma normal.

En R12 se proyectan las siguientes dependencias:

A → BD

B → AD

Las claves candidatas son A y B. La tabla está en BCNF por razones análogas a la tabla anterior.

En R2 no se proyecta ninguna dependencia.

La clave del esquema es AE y está en BCNF.

Tabla 2

R (FGHICDE)

Se proyectan las siguientes dependencias:

D → C

E → DC

F → GH

I → DC

F → GH viola BCNF entonces divido en

R1 (FGH) R2(FICDE)

En R2 se proyectan las siguientes dependencias funcionales:

D → C

E → DC

I → DC

I → DC viola BCNF entonces descompongo.

R1 (FGH) R21(IDC) R22(FIE)

En R21 se proyectan las siguientes dependencias:

D → C

I → DC

I es clave de la tabla R21 y D → C viola BCNF entonces divido.

R1 (FGH) R211 (DC) R212 (ID) R22(FIE)

R1 tiene una única dependencia F → GH. Entonces F es clave y está en BCNF.

R2 tiene una única dependencia D → C. Entonces D es clave y está en BCNF.

R212 tiene una única dependencia I → D. Entonces I es clave y está en BCNF.

R22 no tiene dependencias. Entonces FIE es la clave y está en BCNF.

Pregunta 4

Se pierde $E \rightarrow DC$ y $I \rightarrow A$

Ejercicio 3 (10 puntos)

Pregunta 1

¿Está el siguiente esquema en 4NF? Justificar.

$A \rightarrow BCD$

Tabla

A	B	C	D
---	---	---	---

Solucion

A es clave del esquema.

La dependencia $A \rightarrow BCD$ tiene del lado izquierdo a A que es clave entonces no viola 4NF.

El esquema está en 4NF

Pregunta 2

¿Está el siguiente esquema en 4NF? Justificar.

$A \rightarrow BCD$

$AB \twoheadrightarrow CD$

Tabla

A	B	C	D
---	---	---	---

Solucion

El esquema está en 4NF.

A es clave y $AB \twoheadrightarrow CD$ es una dependencia trivial, ninguna de las dos dependencias viola 4NF.

Pregunta 3

Determinar en que forma normal se encuentra el siguiente esquema.

$A \rightarrow BCD$

$C \twoheadrightarrow D$

Tabla

A	B	C	D
---	---	---	---

Solucion

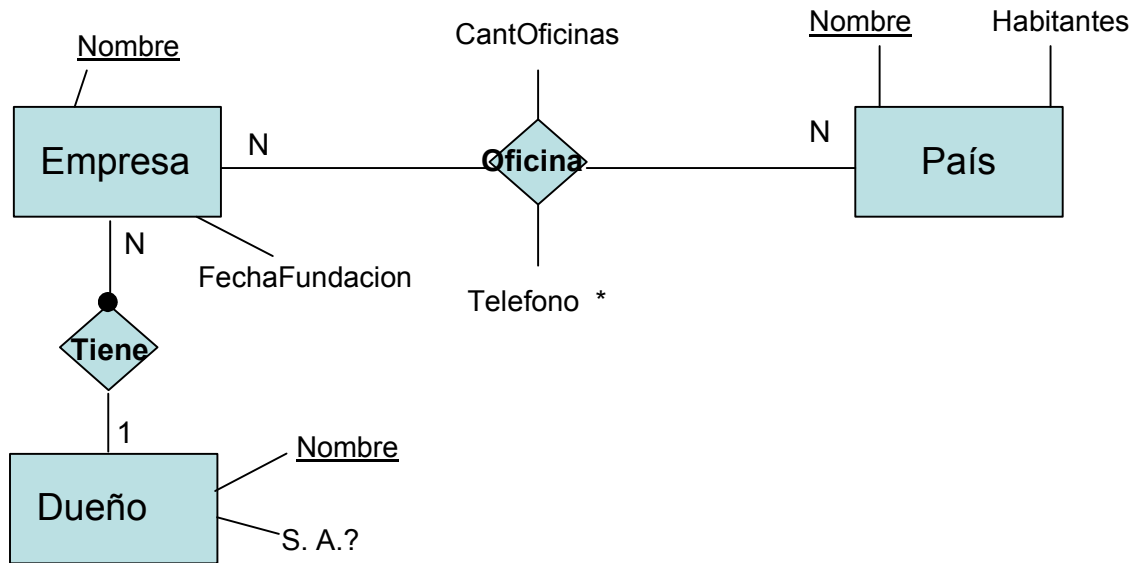
Está en BCNF.

Esto se debe a que la única dependencia funcional ($A \rightarrow BCD$) tiene del lado izquierdo a A que es clave.

No está en 4NF ya que $C \twoheadrightarrow D$ tiene a C del lado izquierdo y C no es superclave del esquema.

Pregunta 4

A partir del siguiente Mer, obtener un diseño relacional incluyendo (si existieran) dependencias funcionales, multivaluadas y de inclusión.



Solucion

Quedan las siguientes tablas.

Empresa (Nombre, FechaFundacion, NombreDueño, SA)

Dep: Nombre → FechaFundacion, NombreDueño, SA

Clave: Nombre

País (Nombre, Habitantes)

Dep: Nombre → Habitantes

Clave: Nombre

Oficinas (NombreEmp, NombrePais, cantOficinas)

Dep: NombreEmp, NombrePais → cantOficinas

Clave: (NombreEmp, NombrePais)

FK: Oficinas.NombreEmp **FK** Empresa.Nombre
Oficinas.NombrePais **FK** País.Nombre

Telefono (NombreEmp, NombrePais, Telefono)

Dep: Ninguna

Clave: NombreEmp, NombrePais, Telefono

FK: Oficinas.NombreEmp **FK** Empresa.Nombre
Oficinas.NombrePais **FK** País.Nombre

Parte 3 – Consultas (25 puntos)

Se tiene un esquema relacional correspondiente a información de una empresa proveedora de servicios informáticos a través de Internet. Estos servicios están organizados por áreas y son ofrecidos a través de páginas Web. Además, la empresa cuenta con un conjunto de páginas de índices que contienen links a otras páginas de la empresa.

Se cuenta con información acerca de los clientes registrados, de las sesiones realizadas por estos y de sus accesos a las páginas de la empresa en cada sesión.

- **Áreas(nomArea, descripcion)**

Contiene información las áreas de la empresa.

- **Servicios(idServ, nomArea, tipo)**

Contiene información de los servicios que ofrece cada área.

- **Paginas (URL, titulo, descripcion)**

Contiene información sobre las páginas de la empresa

- **PaginasIndice(URL)**

Contiene información sobre las paginas índice, las cuales no contienen servicios sino links a otras páginas.

- **PaginasServicios(URL, idServ, nomArea)**

Contiene información sobre las paginas de servicios, indicando que servicio ofrecen.

- **Links(URLOrigen, URLDestino)**

Contiene información sobre los links de las paginas índice. Las URL destino pueden ser paginas de servicios u otras páginas índice. Las URL de origen siempre son páginas de índices.

- **Usuarios(usuario, nombre, ciudad, pais, profesión, sexo, edad, nomAreaFavorita)**

Contiene información sobre los usuarios registrados. Se indica el área favorita de cada usuario

- **Sesiones(fecha, hora, usuario, duracion)**

Contiene información sobre los logins al sitio de la empresa. La duración de la sesión es expresada en minutos.

- **Accesos(fecha, hora, usuario, URL)**

Contiene información sobre las paginas de la empresa que son accedidas por el usuario durante una sesión. Se registran accesos a páginas de índices o de servicios.

- **Solicitudes(fecha, hora, usuario, idServ, nomArea, prioridad)**

Contiene información sobre los servicios solicitados por el usuario durante una sesión. Se indica la prioridad que tiene este pedido para el usuario: Alta, Media o Baja. Esta información se registra en base a la información de la tabla Accesos que corresponde a páginas de servicios.

NOTA:

- Los atributos que aparecen subrayados en una tabla, son claves de la misma.
- Subrayar todos los atributos o no subrayar ninguno son notaciones equivalentes que indican que todos los atributos son clave.

Parte 1. Resolver las siguientes consultas en Álgebra Relacional:

1. Devolver los nombres de aquellos usuarios menores de 25 años, que han solicitado servicios de todas las áreas de la empresa.

Solucion consulta 1

Solicitudes de usuarios menores de 25 años

Solic = $\Pi_{\text{usuario, nombre, nomArea}} (\Pi_{\text{nombre, usuario}} (\sigma_{\text{edad} < 25} \text{Usuarios}) \bowtie \text{Solicitudes})$

Solucion = $\Pi_{\text{nombre}} (\text{Solic} \% \Pi_{\text{nomArea}} (\text{Areas}))$

2. Devolver usuario y nombre de aquellos usuarios que cada vez que realizan solicitudes de servicio lo hacen con prioridad Alta.

Solucion consulta 2

Usuarios que solicitan servicios con prioridad Media o Baja

PrioMB = $\Pi_{\text{usuario}} (\sigma_{\text{prioridad} < \text{'Alta'}} \text{Solicitudes})$

Usuarios que sólo solicitan servicios con prioridad Alta

PrioA = $\Pi_{\text{usuario}} (\text{Solicitudes}) - \text{PrioMB}$

Res = $\Pi_{\$1, \$2} (\text{Usuarios} \bowtie \text{PrioA})$

Parte 2. Resolver la siguiente consulta en Cálculo Relacional:

3. Devolver usuario de aquellos usuarios que nunca solicitaron servicios del área de nombre "Chat"

Solucion consulta 3

Sol 1

$\{u.\text{usuario} / \text{Usuarios}(u) \wedge \neg(\exists s)(\text{Solicitudes}(s) \wedge s.\text{usuario} = u.\text{usuario} \wedge s.\text{nomArea} = \text{'Chat'})\}$

Sol 2

$\{u.\text{usuario} / \text{Usuarios}(u) \wedge (\forall s)(\text{Solicitudes}(s) \wedge s.\text{usuario} = u.\text{usuario} \rightarrow s.\text{nomArea} \neq \text{'Chat'})\}$

Sol 3

$\{u.\text{usuario} / \text{Usuarios}(u) \wedge (\exists a)(\text{Areas}(a) \wedge a.\text{nomArea} = \text{'Chat'} \wedge \neg(\exists s)(\text{Solicitudes}(s) \wedge s.\text{usuario} = u.\text{usuario} \wedge s.\text{nomArea} = a.\text{nomArea}))\}$

Sol 4

$\{u.\text{usuario} / \text{Usuarios}(u) \wedge (\exists a)(\text{Areas}(a) \wedge a.\text{nomArea} = \text{'Chat'} \wedge (\forall s)(\text{Solicitudes}(s) \wedge s.\text{usuario} = u.\text{usuario} \rightarrow s.\text{nomArea} \neq a.\text{nomArea}))\}$

Parte 3. Resolver la siguiente consulta en SQL:

4. Devolver la cantidad de sesiones de usuarios de la empresa y la duración promedio de las mismas en el día '11/09/2001', discriminando por sexo y descartando las sesiones de duración menor que 3 minutos.

NOTA: Si un usuario se conecto mas de una vez ese día, aparecerá mas de una vez en la consulta.

Solucion consulta 4

```
Select u.sexo, count(*) as cantSesiones, avg(duracion)
From usuarios u, sesiones s
Where u.usuario = s.usuario and s.fecha = '11/09/2001' and s.duracion > 3
Group by u.sexo
```


Parte 4 – Optimización, Concurrencia y Recuperación (25 puntos)

Ejercicio 5.

Dado el siguiente esquema relacional e información del Catalogo:

PROGRAMAS (nomProg, genero, actorPpal)

PROGRAMACIÓN (nomProg, dia, horaComienzo, horaFin)

EMPLEADOS (nomEmp, ciEmp, dirEmp, telEmp, genero, cargo, horaEntrada, horaSalida)

	PROGRAMAS	PROGRAMACION	EMPLEADOS
Cantidad tuplas	300	800	100
Distribución de valores	Existen 50 géneros diferentes.	Existen 300 programas diferentes.	Existe un supervisor diferente por cada genero. Todos los géneros tienen un supervisor.

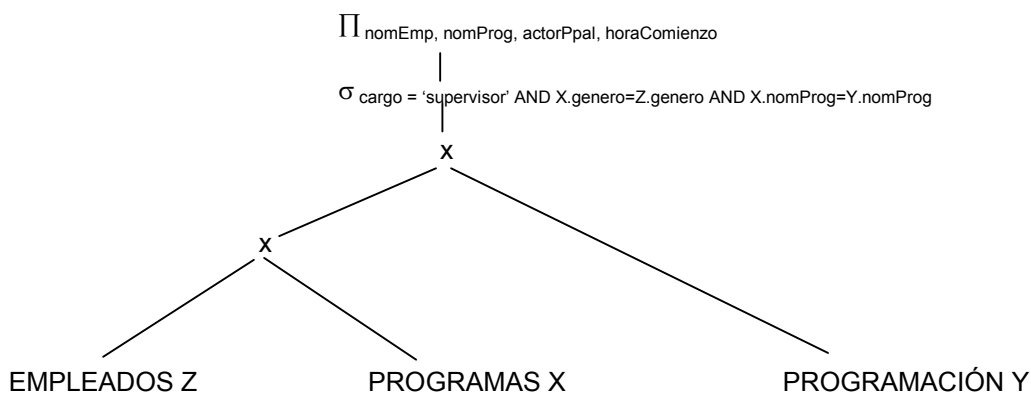
y la siguiente consulta:

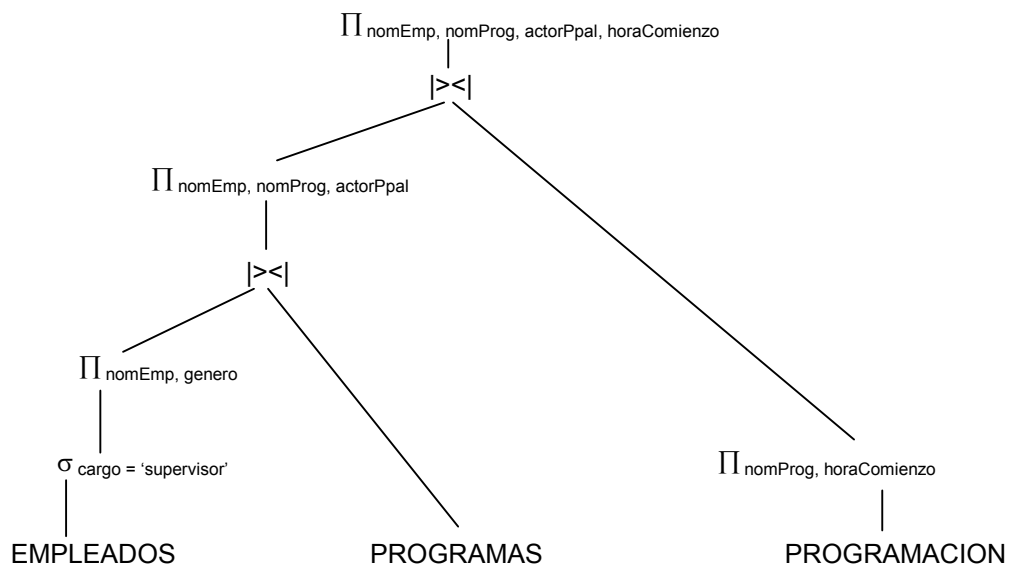
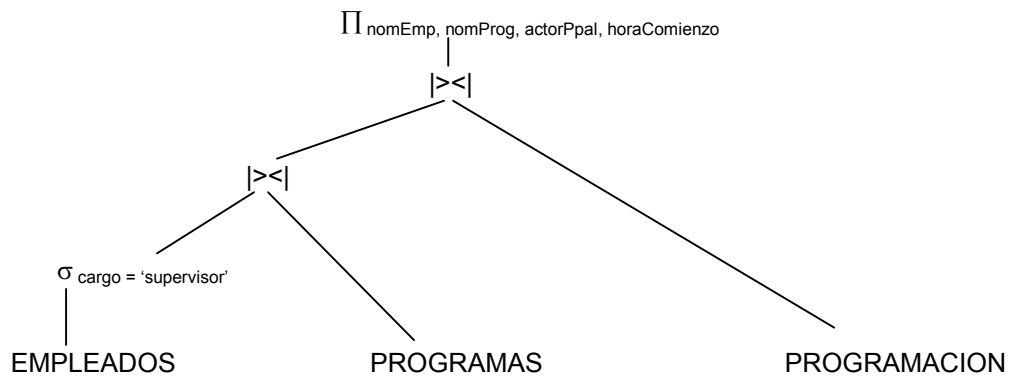
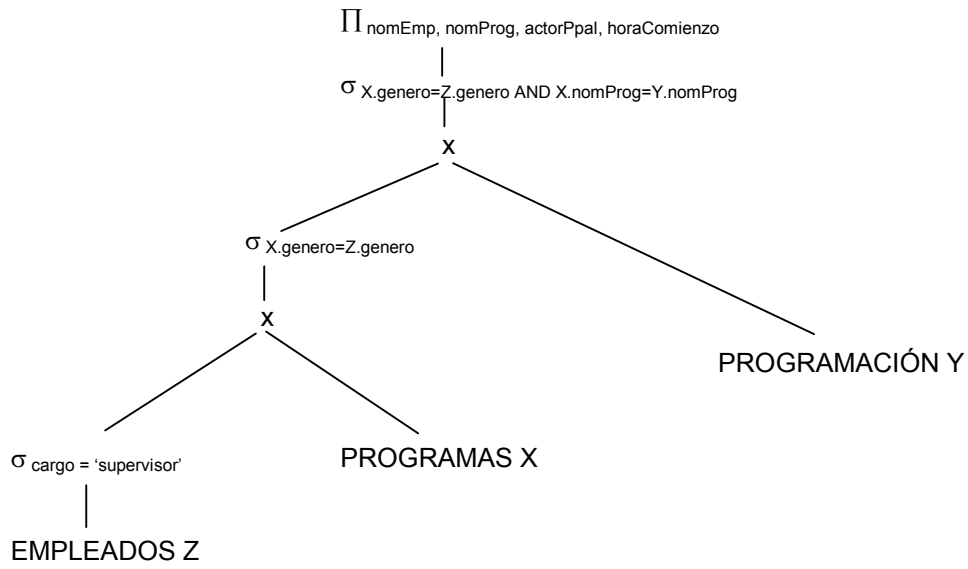
```
SELECT nomEmp, nomProg, actorPpal, horaComienzo
FROM Programas X, Programación Y, Empleados Z
WHERE X.genero = Z.genero AND X.nomProg = Y.nomProg AND Z.cargo = 'supervisor'
```

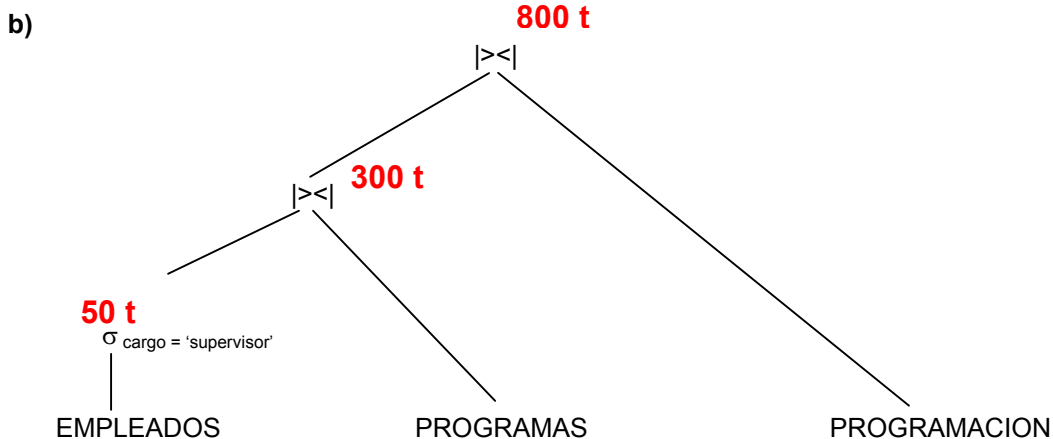
- Dar un plan lógico para la consulta, aplicando las heurísticas para obtenerlo. (Dar el árbol inicial y justificar cada paso de la transformación).
- Calcular los tamaños de los resultados de todas las operaciones que se realizan. Cambiar el orden de las hojas del árbol, si es necesario, a la luz de los resultados obtenidos.
(Tamaño $(R \bowtie S) = |R| * |S| / \text{Max}(V(A,R), V(A,S))$)
- Discutir posibles planes físicos para cada operación (excepto proyecciones) considerando la existencia de índices o no para cada caso.

SOLUCION

a)







Tamaño $(R \bowtie_A S) = |R| \cdot |S| / \text{Max}(V(A,R), V(A,S))$

Tamaño (1er. join) = $50 \cdot 300 / 50 = 300$

Tamaño (2°. Join) = $300 \cdot 800 / 300 = 800$

Calculamos tamaños sin considerar las proyecciones.

Observamos que en todos los joins queda la tabla de menor cantidad de tuplas del lado izquierdo. Por lo tanto no se debe intercambiar de lugar ninguna hoja del árbol.

c) Se debe presentar las diferentes implementaciones para el select y el join que están descritas en el libro del curso.

Ejercicio 6.

Sea un sistema ventas de pasajes de avión, existiendo distintos puntos de venta.

Procedure Venta

```

cant := input(c);
if lugares_disp > 0 then
    lugares_disp := read(X);
    write (X, lugares_disp - c);
    ganancia := read(Y);
    write(Y, ganancia + valor_pasaje * c);
end;
end;
```

a) Escribir historias H1 donde se ejecuten 2 transacciones de Venta, que cumplan:

I. H1 entrelazada, que evita abortos en cascada.

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

II. H1 no serializable y estricta.

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

III. H1 serial.

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

IV. H1 entrelazada, serializable y no recuperable.

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

- b) Escribir una historia H2 donde se ejecuten 2 transacciones de Venta y éstas sigan el protocolo 2PL básico. Decir si H2 es serializable y si es recuperable, justificando.

**H2: r1(X) r1(X) w1(X) w1(X) r1(Y) r1(Y) w1(Y) u1(X) r2(X) r2(X) w2(X) w2(X) u2(X) w1(Y)
u1(Y) c1 r2(Y) r2(Y) w2(Y) w2(Y) u2(Y) c2**

H2 es serializable, ya que las transacciones cumplen 2PL.

H2 es recuperable, ya que T2 lee de T1 y T1 hace commit antes que T2.