

FUNDAMENTOS DE BASES DE DATOS

Examen Julio 2020

SOLUCIONES

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 (28 puntos).

En la época de la “nueva normalidad”, se desea modelar la realidad de peluquerías distribuidas en todo el país.

De las peluquerías se conoce un código identificador y su nombre. Además, se registra el conjunto de sucursales de cada peluquería. A cada sucursal se le asigna un número que la identifica entre las sucursales de la peluquería a la que pertenece. Además, de cada sucursal se conoce su teléfono, la dirección (calle, número y departamento) en que se encuentra.

Con las sucursales de las peluquerías se vinculan dos tipos de personas: staff y clientes. De las personas se conoce su cédula de identidad, su nombre y sus teléfonos. Para los clientes, además se considera su e-mail, que también lo identifica. Por otro lado, el staff es exclusivo de cada sucursal,. Además, cada sucursal de cada peluquería tiene un staff formado por varias personas.

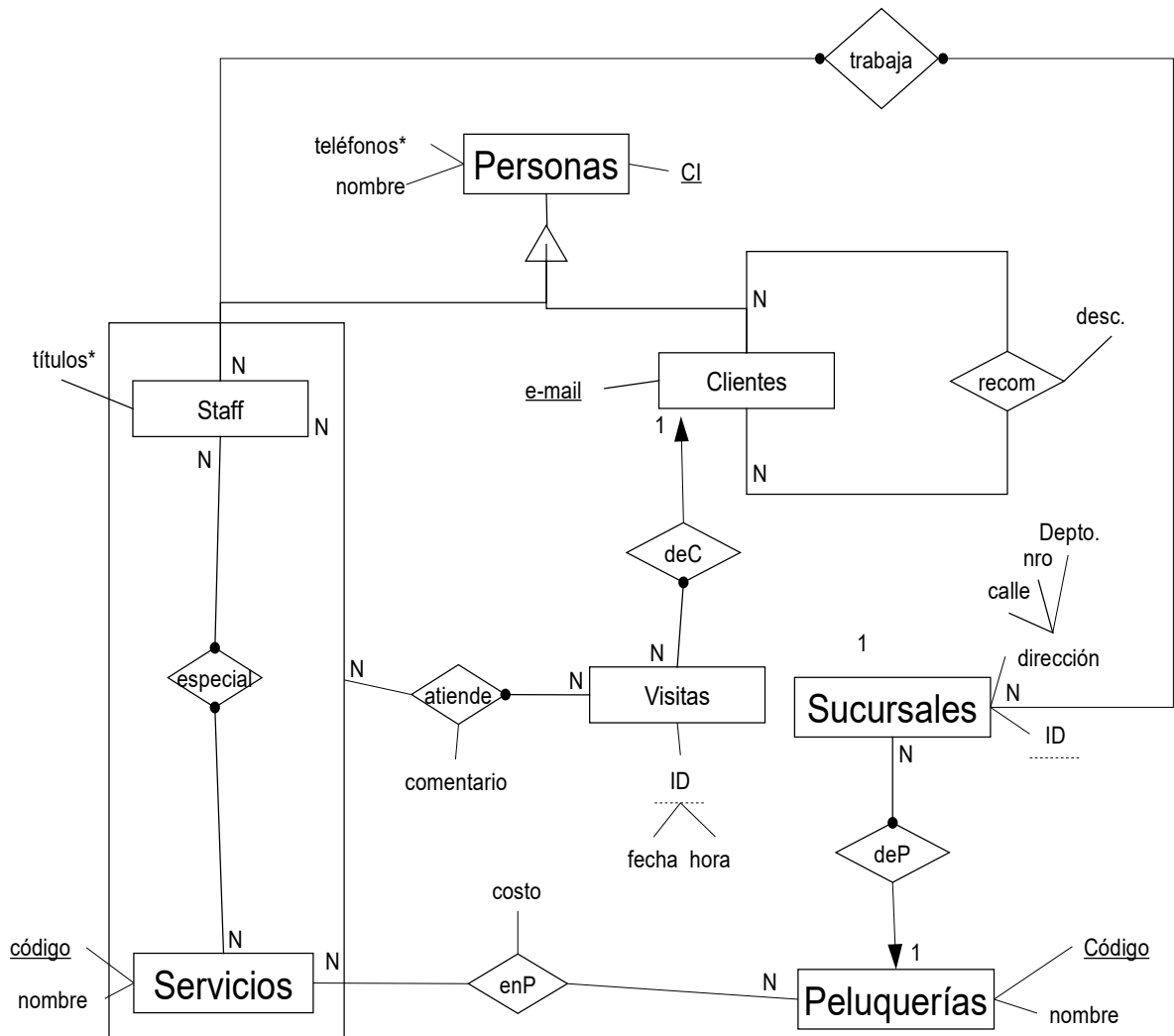
Las peluquerías dan distintos servicios (*brushing, color, corte*, etc.) y el costo de cada servicio puede variar entre las peluquerías. Sin embargo, el costo de los servicios es el mismo para todas las sucursales de una misma peluquería. De los servicios se conoce el código que los identifica y el nombre. Los servicios que da una peluquería están disponibles en todas sus sucursales.

Cada persona del staff de una sucursal, está especializado en uno o más servicios. Es claro que cada persona del staff sólo da servicios en los que está especializado.

Los clientes realizan visitas en las que personas del staff les brindan varios servicios. Cada visita de un cliente, está identificadas por la fecha y hora. Tenga en cuenta, que un mismo cliente no puede hacer dos visitas distintas en la misma fecha y hora, pero diferentes clientes pueden asistir en la misma fecha y hora la misma sucursal. Para cada servicio que da una persona del staff en una visita de un cliente, se registra un comentario con detalles del mismo (porcentaje de cada color usado, tipo de corte realizado, etc).

Finalmente, el grupo de peluquerías tienen un sistema de promociones, cuando un cliente recomienda a otro, sin importar la peluquería, ambos clientes reciben un descuento. Por lo tanto, interesa registrar cuando un cliente recomienda a otro y el descuento que ambos recibieron.

Se pide: Modelo Entidad Relación completo.



RNE:

- Clientes U Staff = Personas
- “Coinciden” es irreflexiva y asimétrica
- Los miembros del staff atienden servicios que son brindados en las peluquerías que trabajan dichos miembros de staff.

$(\forall st \in \text{Staff}) (\forall se \in \text{Servicios}) (\forall v \in \text{Visitas}) \langle \langle st, se \rangle, v \rangle \in \text{atiende} \rightarrow$

$(\exists p \in \text{Peluquerias}) (\exists su \in \text{Sucursales}) \langle se, p \rangle \in \text{enP} \wedge \langle p, su \rangle \in \text{deP} \wedge \langle st, su \rangle \in \text{trabaja}$

Ejercicio 2 (28 puntos)

Para poder tomar pruebas presenciales en Facultad durante la pandemia, hay que asignar a cada estudiante de cada prueba, además del salón que le toca, una hora de entrada por una puerta determinada. Además, en la entrada el sistema registra la hora en que entró, la puerta por la que entró y la temperatura registrada.

La base de datos que se presenta, es el resultado acumulado de todos los días desde el inicio del período hasta un día cualquiera.

Los datos representan un sólo período de parciales que se toman en dos turnos (matutino y vespertino). Se asume que sólo hay una prueba por asignatura. Las horas de entrada se asignan cada 5 minutos, es decir, hay una tanda de estudiantes que se les asigna las 9:30 en una puerta y en la misma puerta la siguiente tanda es 9:35. En diferentes puertas, pueden coincidir las horas.

A continuación se describen las tablas.

Salones (salon, aforo, puerta)

- Una tupla representa un salón y tiene un identificador de salón (salon), la máxima cantidad de estudiantes que pueden entrar en el salón y la identificación de la puerta del edificio por la que deben ingresar los estudiantes para ese salón.

Puertas (puerta, edificio)

- Una tupla representa una puerta de entrada a un edificio. Tiene el nombre de la puerta (Ej: 'Aulario Massera 1') y el nombre del edificio en donde está esa puerta.

Prueba (nomAsig, turno, fecha, hora, cantEst)

- Una tupla representa un parcial de una asignatura y tiene el nombre de la asignatura, el turno ('matutino', 'vespertino') en que se tomará el parcial, la fecha, la hora de inicio prevista y la cantidad de estudiantes anotados.

Ingreso (Ci, nomAsig, puertaAsig, horaAsig, salon, horaReal, puertaReal, temp)

- Una tupla representa el ingreso planificado y el ingreso real de un estudiante a un edificio para presentarse a un parcial. Además del documento del estudiante y la asignatura del parcial, el ingreso planificado está dado por la puerta asignada (puertaAsig), la hora de entrada asignada (horaAsig) y el salón. El ingreso real está dado por la hora real (horaReal) en la que el estudiante digitó su documento, la puerta real (puertaReal) en que lo hizo y la temperatura¹ (temp) que presentó en la toma.

Se cumplen además, las siguientes dependencias de inclusión:

$$\begin{aligned}\Pi_{puerta}(Salones) &\subseteq \Pi_{puerta}(Puertas) \\ \Pi_{nomAsig}(Ingreso) &\subseteq \Pi_{nomAsig}(Prueba) \\ \Pi_{puertaAsig, salon}(Ingreso) &\subseteq \Pi_{puerta, salon}(Salones) \\ \Pi_{puertaReal}(Ingreso) &\subseteq \Pi_{puerta}(Puertas)\end{aligned}$$

Se asume que no hay tablas vacías.

1 En la versión real del sistema, la temperatura no queda registrada.

Se Pide:

a) Resolver las siguientes consultas en Álgebra Relacional.

I. Devolver los salones y asignaturas en los que todos los estudiantes de esa prueba en ese salón, tuvieron más que 37,6° de temperatura.

$$\Pi_{salon, nomAsig}(Ingreso) - (\Pi_{salon, nomAsig}(\sigma_{temp \leq 37,6}(Ingreso)))$$

II. Devolver las fechas tal que todos los salones del edificio "Aulario Massera" están asignados en el turno "matutino".

$$\Pi_{fecha, salon}(Ingreso * \sigma_{turno='matutino'}(Prueba)) \% \Pi_{salon}(\sigma_{edificio='Aulario Massera'}(Salones * Puertas))$$

b) Resolver las siguientes consultas en Cálculo Relacional.

I. Devolver, para cada asignatura la hora planificada del primer y del último ingreso.

$$\left\{ \left(\min.nomAsig, \min.hora, \max.hora \right) / Ingresos(\min) \wedge Ingresos(\max) \wedge \right. \\ \left. \max.nomAsig = \min.nomAsig \wedge \right. \\ \left. \forall x_1 (Ingresos(x_1) \wedge x_1.nomAsig = \min.nomAsig \rightarrow \right. \\ \left. (x_1.horaAsig \geq \min.horaAsig \wedge x_1.horaAsig \leq \max.horaAsig) \right) \\ \left. \right\}$$

II. Devolver nombre de asignatura y salón, tal que en ese salón todos los estudiantes llegaron antes de la hora de comienzo.

$$\left\{ \left(t.nomAsig, t.salon \right) / Ingreso(t) \wedge \right. \\ \left. \exists p (prueba(p) \wedge t.nomAsig = p.nomAsig \wedge \right. \\ \left. \neg \exists e (Ingreso(e) \wedge t.nomAsig = e.nomAsig \wedge t.salon = e.salon \wedge p.hora < e.horaReal) \right) \\ \left. \right\}$$

c) Resolver la siguiente consulta en SQL

I. Devolver las fechas, horas y puertas en las que esa puerta tiene asignados más de 20 estudiantes en un horario determinado.

```
select puertaAsig, fecha, hora
from Ingreso natural join prueba
group by puertaAsig, fecha, hora
having count(*) > 20
```

Ejercicio 3 (24 puntos)

Considere el siguiente esquema relacional $R(A,B,C,D,E)$ y el siguiente conjunto de dependencias funcionales:

$$F = \{ AB \rightarrow C, C \rightarrow A, DE \rightarrow B \}$$

a) **Determine todas las claves. Justifique su respuesta.**

Para determinar todas las claves vamos a calcular 3 conjuntos de atributos: los que nunca están a la derecha (ND), los que sólo están a la derecha (SD) y los que están de ambos lados de las dependencias (ID).

	Atributos	Observaciones
ND	D, E	Los atributos que nunca están a la derecha deberán estar en todas las claves.
SD		Los que solamente están a la derecha, no pueden pertenecer a ninguna clave.
ID	A, B, C	Los que están a ambos lados de las dependencias podrían pertenecer o no a alguna/s clave/s.

Si ND es una superclave (todos los atributos dependen de ella) entonces es la única. Al hacer la clausura, vemos que esto no se cumple: $ND^+ = \{D, E\}^+ = \{D, E, B\}$

Vayamos probando qué pasa cuando agregamos de a un elemento de ID.

Si agregamos A: $\{D, E, A\}^+ = \{D, E, A, B, C\}$ vemos que $\{D, E, A\}$ es una superclave.

Ahora, si agregamos B: $\{D, E, B\}^+ = \{D, E, B\}$ vemos que $\{D, E, B\}$ no es una superclave.

Por último, si agregamos C: $\{D, E, C\}^+ = \{D, E, C, B, A\}$ vemos que $\{D, E, C\}$ sí es superclave.

Como el sacarle un atributo le quita la propiedad de ser superclave, decimos que $\{D, E, A\}$ y $\{D, E, C\}$ son las dos claves que surgen del conjunto F de dependencias funcionales.

b) **Determine en qué forma normal se encuentra el esquema. Justifique su respuesta.**

Dado que la dependencia $DE \rightarrow B$ viola 2NF, porque B depende parcialmente de una clave y B no es un atributo primo, el esquema relacional R se encuentra en **1NF**.

c) **Obtenga un esquema en BCNF aplicando el algoritmo visto en el curso. Justifique su respuesta.**

Para que una base de datos relacional esté en BCNF, todas las dependencias funcionales deben tener una superclave del lado izquierdo. Para eso, descompondremos el esquema R hasta obtener una descomposición que cumpla BCNF.

Las claves para el esquema R son DEA y DEC. Por tanto, tomemos $DE \rightarrow B$ que provoca que R viole BCNF; descomponemos en dos relaciones:

$R_1(D, E, B)$ $\Pi_{R_1}(F) = \{DE \rightarrow B\}$

Claves: DE

R_1 está en BCNF, dado que la única dependencia funcional en su proyección tiene a una superclave del lado izquierdo.

 $R_2(A, C, D, E)$ $\Pi_{R_2}(F) = \{DEA \rightarrow C, C \rightarrow A\}$

Claves: DEC, DEA

R_2 está en 1NF ya que $C \rightarrow A$ viola 2NF

En R_2 tenemos dos superclaves DEA y DEC. Ahora vamos a descomponer R_2 según $C \rightarrow A$, que es la que genera que R_2 viole BCNF:

 $R_{21}(C, A)$ $\Pi_{R_{21}}(F) = \{C \rightarrow A\}$ R_{21} está en BCNF

Claves: C

 $R_{22}(C, D, E)$ $\Pi_{R_{22}}(F) = \{\}$ R_{22} está en BCNF

Claves: DEC

Por tanto, esquema final está formado por $R_1(D, E, B)$, $R_{21}(C, A)$ y $R_{22}(C, D, E)$

Nota: Esta no es la única descomposición posible, dado que dependiendo qué dependencias funcionales tenemos para generar la descomposición podremos obtener diferentes esquemas.

d) Considere nuevamente el esquema R y el conjunto F. Agregue una dependencia funcional de forma que se mantengan las mismas claves. Justifique su respuesta.

Si tomamos cualquier dependencia $X \rightarrow Y \in F$, al agregar $XY \rightarrow Y$ estaremos agregando redundancia. Sea F' el nuevo conjunto de dependencias funcionales, se cumple que: $F^+ = F'^+$, por lo tanto las claves en F son las mismas que en F' .

Podemos agregar entonces la dependencia funcional $CA \rightarrow A$.

e) Considerando el esquema R y el conjunto de dfs F, determine si es posible que se cumpla la mvd: $DE \twoheadrightarrow AB$. Justifique su respuesta.

Para determinar si es posible que se cumpla la multivaluada $DE \twoheadrightarrow AB$ evaluaremos cómo se comporta la definición.

	D	E	A	B	C
t1	d1	e1	a1	b1	c1
t2	d1	e1	a2	b2	c2
t3	d1	e1	a1	b1	c2
t4	d1	e1	a2	b2	c1

Sin embargo, al planterar la definición para dicha multivaluada, se viola la dependencia funcional $C \rightarrow A$. Por tanto, **la dependencia multivaluada no se cumple.**

Ejercicio 4 (20 puntos).

Considere el siguiente esquema de bases de datos:

Peluquerías(codPelú, nombre, departamento)

Peluquerías de todo el país.

Sucursales(codPelú, nroSuc, capacidad)

Sucursales de cada peluquería. La capacidad es la cantidad de clientes que puede atender a la vez.

Cientes(CI, codPelú, nroSuc, nombre, edad, sexo)

Cientes que se atienden en cada sucursal. Se registra el nombre, la edad y el sexo del cliente.

Se sabe que en cada tabla existe un índice primario por la clave primaria correspondiente y, además, existen los siguientes índices:

Tabla	Tipo	Atributos	Nombre
Peluquerías	Índice Secundario	departamento	IndDepartamento
Cientes	Índice Secundario	CodPelú, nroSuc	IndPelúSuc

También se conoce la siguiente información:

Tabla	Cantidad de Tuplas
Peluquerías	4000
Sucursales	4500
Cientes	65000

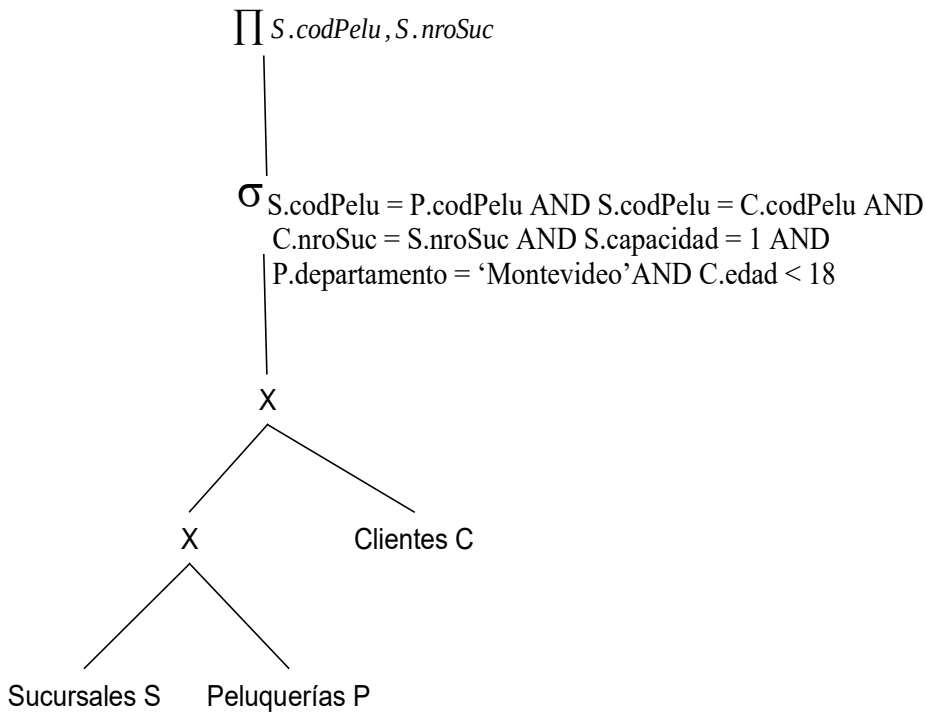
- El 10% de las sucursales tienen capacidad = 1.
- El 50% de las peluquerías están en Montevideo.
- El 1% de los clientes son menores de edad.

Considere la siguiente consulta sobre el esquema dado:

```
SELECT C.CI, S.codPelú, S.nroSuc
FROM Sucursales S, Peluquerías P, Cientes C
WHERE S.codPelú = P.codPelú AND S.codPelú = C.codPelú AND
      C.nroSuc = S.nroSuc AND S.capacidad = 1 AND P.departamento = 'Montevideo'
AND C.edad < 18;
```

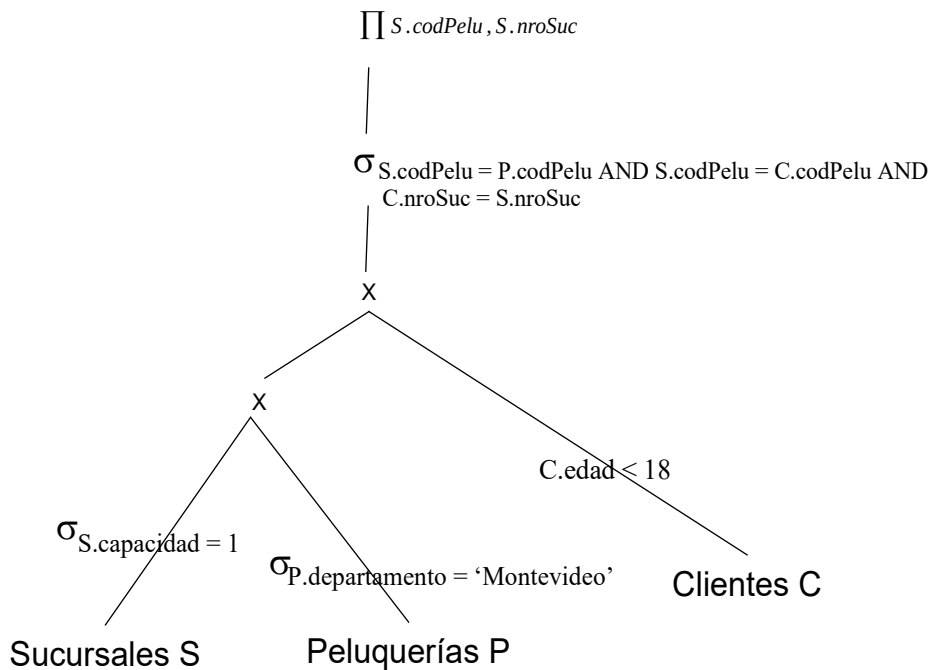
Se pide:

a) Dar el árbol canónico de la consulta planteada.

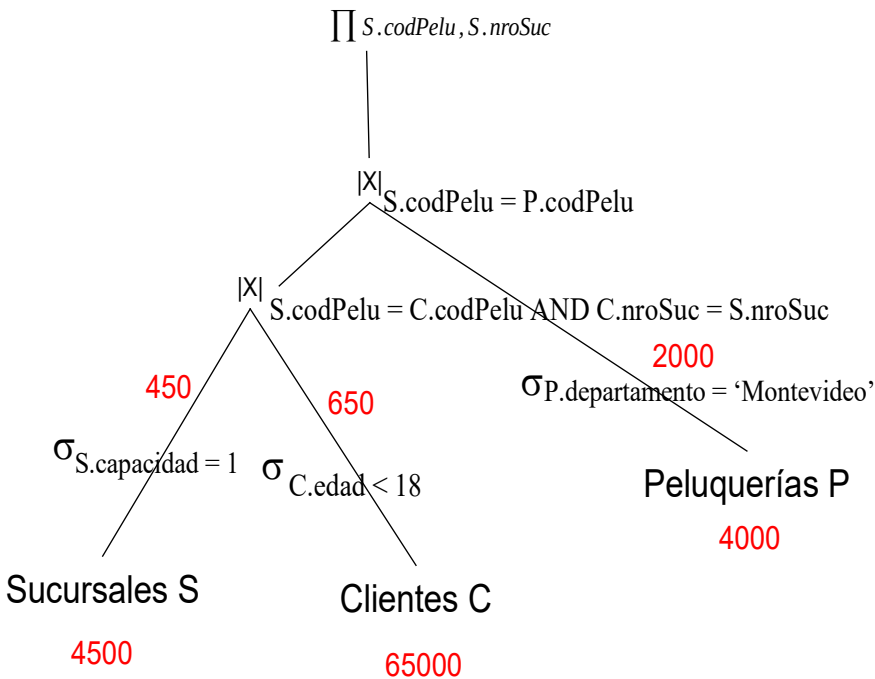
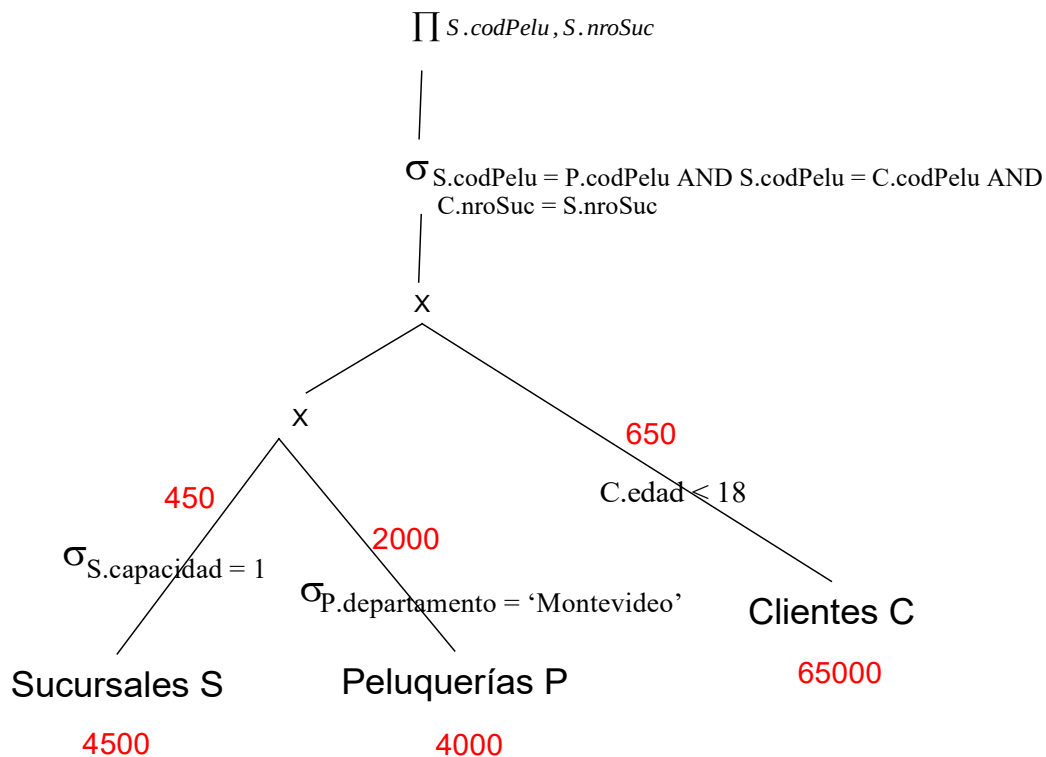


b) Aplicando las heurísticas, y calculando los tamaños intermedios cuando sea necesario, dar el mejor plan lógico para la consulta anterior.

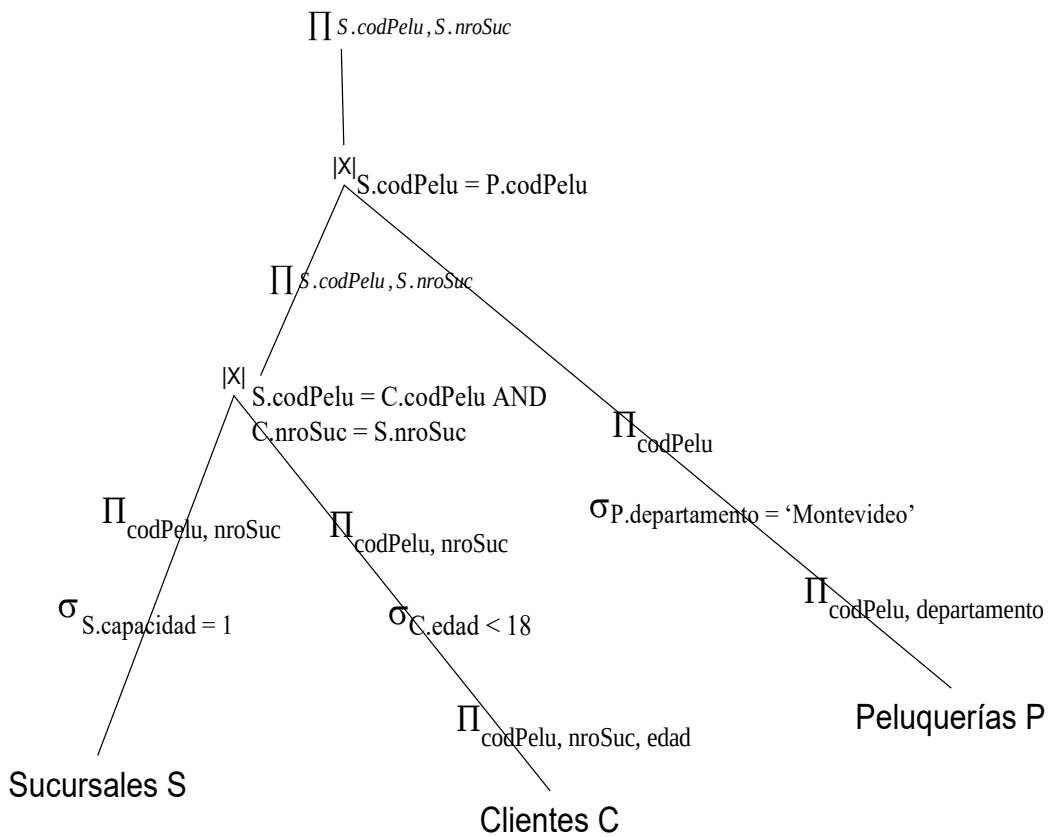
Heurísticas 1 y 2:



Heurísticas 3 y 4:



Heurística 5:



c) Para el plan lógico de la parte b), dar un plan físico considerando la mayor cantidad de índices que sea posible.

