

FUNDAMENTOS DE BASES DE DATOS

Solución Examen Febrero 2012

La duración del examen es de 3 horas y $\frac{1}{2}$.

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 (25 puntos).

Se desea modelar la realidad correspondiente a eventos científicos que se realizan en cierto país. Cada evento tiene un número identificador, un nombre, un costo, una fecha de inicio, una fecha de fin y la dirección donde se realiza el evento. De esta última se conoce la ciudad, la calle y el número de puerta.

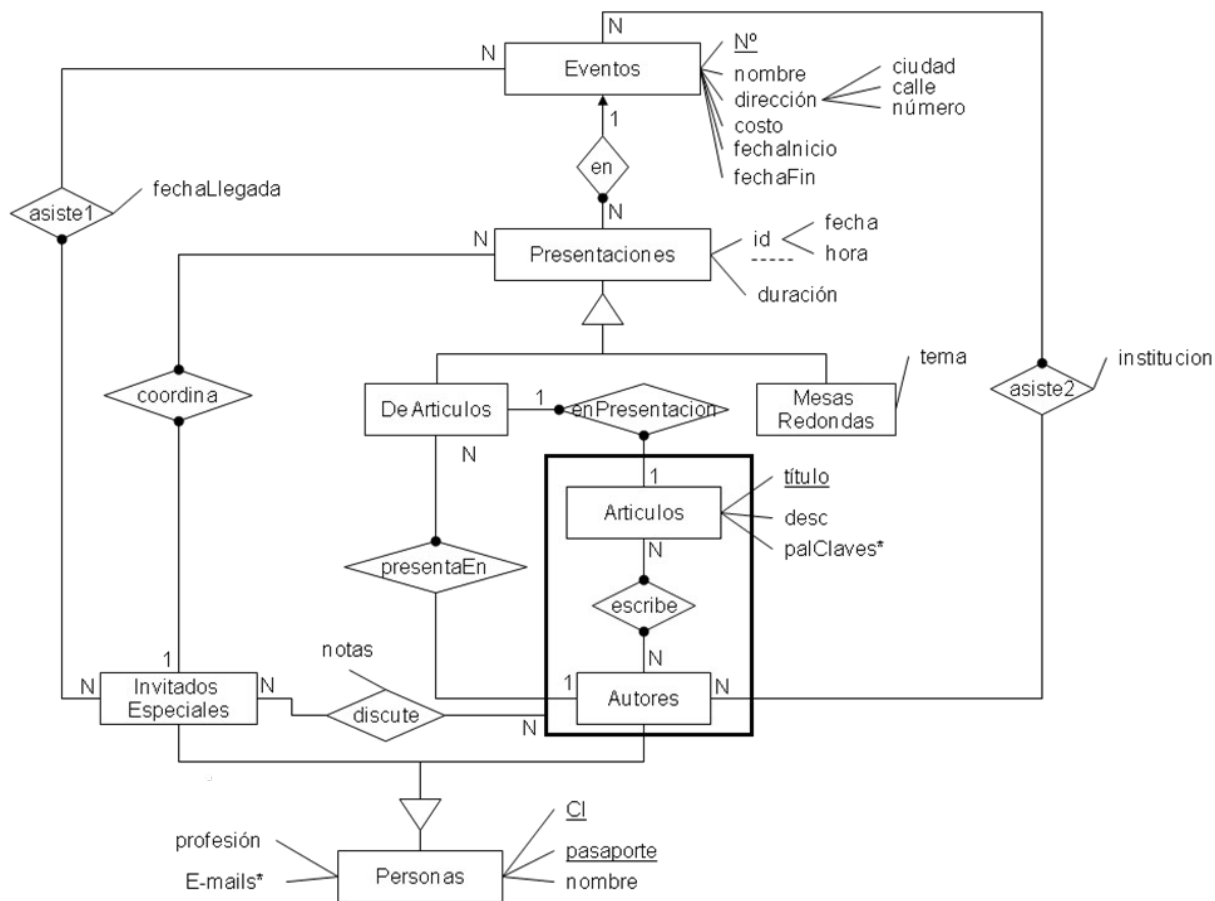
En cada evento se realizan presentaciones. Cada una de ellas se identifica por el número de evento al cuál pertenece conjuntamente con la fecha y la hora a la que se realiza. Además, se conoce la duración de la misma. Hay dos tipos de presentaciones, de artículos y Mesas Redondas. De las últimas interesa saber el tema que es presentado.

Los artículos que se presentan en los eventos se identifican por su título, tienen una descripción y un conjunto de palabras clave. Un artículo puede ser escrito por varios autores y un autor puede escribir muchos artículos. Se sabe que todos estos autores escribieron al menos un artículo que es presentado en algún evento. Cada artículo es presentado por uno solo de los autores y una sola vez en un único evento. Sólo se modelan artículos presentados en algún evento.

Hay dos tipos de personas que están relacionadas con los eventos y/o los artículos: autores e invitados especiales. Tanto de los autores como de los invitados especiales se conoce la CI, el número de pasaporte, el nombre, la profesión y todos sus e-mails. Tanto la CI como el pasaporte sirven para identificar a las personas. Para cada evento al que asiste un autor interesa registrar la institución que lo registró (tenga en cuenta que un autor puede ir a un evento por una institución y a otro por otra). De los invitados especiales, interesa registrar la fecha en la que llega al evento. Tenga en cuenta que alguien que es autor en un evento puede ser invitado especial en otro (o incluso en el mismo evento).

Cada presentación tiene un único coordinador, que es uno de los invitados especiales. Todos los invitados especiales coordinan al menos una presentación. También se sabe que cuando un invitado especial se encuentra con un autor discuten al menos uno de los artículos de dicho autor. Interesa registrar las notas que surgen de las discusiones de cada artículo, para generar posteriormente posibles nuevos artículos.

Se pide: Modelo Entidad-Relación completo del problema.



RNE:

$Presentaciones = DeArticulos \cup Mesas\ Redondas$

$De\ Articulos \cap Mesas\ Redondas = \emptyset$

$Personas = Invitados\ Especiales \cup Autores$

Un invitado especial coordina presentaciones de eventos a los que él asiste:

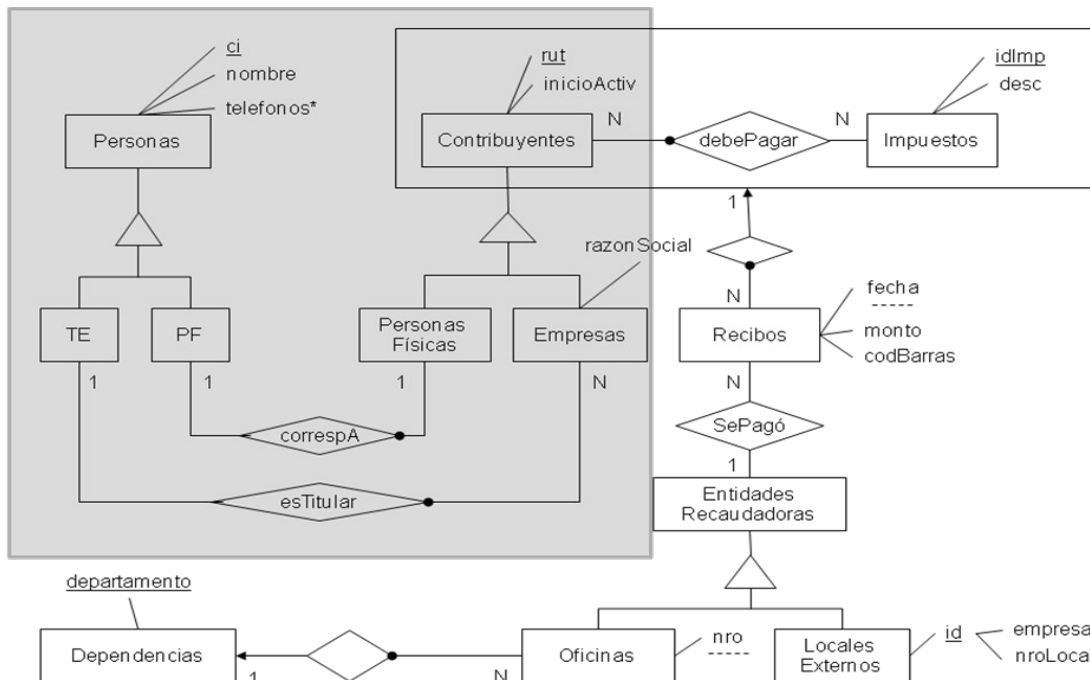
$(\forall i \in Invitado\ Especial)(\forall p \in Presentaciones)$
 $(\langle i, p \rangle \in coordina \rightarrow (\exists e \in Eventos) (\langle i, e \rangle \in asiste1 \wedge \langle p, e \rangle \in en))$

Todo artículo que es presentado, es presentado por un autor de dicho artículo que asiste al evento en el cual éste está siendo presentado

$(\forall a \in Autores)(\forall ar \in Artículos)(\forall p \in Presentaciones)$
 $(\langle a, ar \rangle \in escribe \wedge \langle a, p \rangle \in presenta \wedge \langle ar, p \rangle \in enPresentacion \rightarrow$
 $(\exists e \in Eventos) (\langle a, e \rangle \in asiste2 \wedge \langle p, e \rangle \in en))$

Ejercicio 2 (25 puntos).

Considere el Modelo Entidad Relación y las RNE que se indican a continuación.



1. $\text{LocalesExternos} \cup \text{Oficinas} = \text{EntidadesRecaudadoras}$
2. $\text{LocalesExternos} \cap \text{Oficinas} = \emptyset$
3. $\text{Empresas} \cup \text{PersonasFísicas} = \text{Contribuyentes}$
4. $\text{Empresas} \cap \text{PersonasFísicas} = \emptyset$
5. $\text{TE} \cup \text{PF} = \text{Personas}$
6. $\text{TE} \cap \text{PF} = \emptyset$
7. $\forall p \in \text{Personas}. (\exists pf \in \text{PersonasFísicas} . \langle p, pf \rangle \in \text{correspA}) \rightarrow (\neg \exists e \in \text{Empresas} . \langle p, e \rangle \in \text{esTitular})$

Parte 1)

A continuación se presentan dos posibles representaciones relacionales de las entidades *EntidadesRecaudadoras* y *Dependencias*. Indique cuál representación es correcta y cuál incorrecta. Justifique su respuesta considerando aspectos de identificación, y otras restricciones.

Representación (1)	<p>EntidadesRecaudadoras(nro,departamento,empresa,nroLocal,ERID) $ERID \rightarrow nro, departamento, empresa, nroLocal$ $\Pi_{departamento}(EntidadesRecaudadoras) \subseteq Dependencias$</p> <p>Dependencias(departamento)</p>
Representación (2)	<p>LocalesExternos(empresa,nroLocal,ERID) $ERID \rightarrow empresa, nroLocal$ $empresa, nroLocal \rightarrow ERID$</p> <p>Oficinas(nro,departamento,ERID) $ERID \rightarrow nro, departamento$ $nro, departamento \rightarrow ERID$ $\Pi_{departamento}(Oficinas) \subseteq Dependencias$</p> <p>EntidadesRecaudadoras(ERID) $EntidadesRecaudadoras = \Pi_{ERID}(LocalesExternos) \cup \Pi_{ERID}(Oficinas)$</p> <p>Dependencias(departamento)</p>

Solución parte 1)

La opción 1) fuerza a tener valores nulos en algunos atributos, por ejemplo cuando la *EntidadRecaudadora* sea un *LocalExterno* los atributos *departamento* y *nro* serán nulos. Además, permite tener el mismo valor de los atributos *nro*, *departamento*, *empresa*, *nroLocal* asociados a diferentes *ERID*. Esto induce potenciales errores de consistencia como por ejemplo, que existan dos entidades recaudadoras (distinto *ERID*) para la misma *empresa* y *nroLocal*. Por otra parte también permite que un *ERID* tenga valores en *empresa*, *nroLocal*, *nro* y *departamento* a la vez, lo cual sería incorrecto ya que por la RNE 2 las entidades recaudadoras o son Oficinas o Locales Externos.

La opción 2) representa adecuadamente la realidad ya que:

- Sólo para los *LocalesExternos* se guardan los atributos *empresa* y *nroLocal*
- No puede haber una pareja *empresa* y *nroLocal* asociada a *ERID* distintos en *LocalesExternos*.
- No puede haber una pareja *nro* y *departamento* asociada a *ERID* distintos en *Oficinas*
- Se garantiza que las únicas *EntidadesRecaudadoras* que hay son *LocalesExternos* o son *Oficinas*

Parte 2)

Considere la siguiente representación relacional de la relación *debePagar*:

DebePagar(CI,rut,inicioActiv,idImp,desclmp)

- a) Determine cuáles son las dependencias funcionales que se cumplen sobre esa tabla, suponiendo que el Modelo Entidad Relación representa fielmente la realidad.
- b) Indique en qué forma normal está esa tabla.
- c) Aplique los algoritmos vistos en el curso para construir una representación de esa tabla en 4NF.

Solución parte 2)

La tabla sobre la que hay que trabajar es: DebePagar(CI,rut,inicioActiv,IdImp,desclmp)

a)

Las dependencias que se cumplen sobre esta tabla son las siguientes:

- **rut → InicioActiv** Esta dependencia surge porque Rut es determinante en
- **idImp → Desclmp**. Esta dependencia surge porque IdImp es determinante en Impuesto.
- **rut → CI**. Esta dependencia surge porque todo Contribuyente es o bien una persona física y por lo tanto tiene asociada una única Persona, o bien es una empresa que también tiene asociada una única Persona por otra relación. Además, como las dos categorías son excluyentes, no es posible que tengan asociadas más de una Persona en distintas relaciones.

b) La única clave de la tabla es rut,idImp. Esto se puede ver calculando la clausura de los atributos y observando que ninguno es redundante. De esta forma, rut → CI induce una dependencia parcial de clave sobre un atributo no primo por lo que la tabla está en 1NF.

c) Aplicando el algoritmo para 4NF sobre rut → CI obtenemos las siguientes tablas y proyecciones de dependencias:

A_{11} (rut,CI)
 { rut → CI }
 A_{12} (rut, inicioActiv,idImp,desclmp)
 { Rut → InicioActiv; IdImp → Desclmp }

La tabla A_{12} (Rut sigue teniendo como Clave Rut,IdImp por lo que sigue estando en 1NF. Partiendo esta tabla obtenemos:

A_{121} (Rut,InicioActiv)
 { Rut → InicioActiv }
 A_{122} (Rut,IdImp,Desclmp)
 { IdImp → Desclmp }

La tabla A_{122} sigue estando en 1NF porque la clave sigue siendo Rut,IdImp por lo que la partimos nuevamente obteniendo:

A_{1221} (IdImp, Desclmp)
 { IdImp → Desclmp }
 A_{1222} (Rut,IdImp)

Ahora nuestra descomposición en 4NF es: { A_{11} (Rut,CI), A_{121} (Rut,InicioActiv), A_{1221} (IdImp,

DescImp), A₁₂₂₂ (Rut, IdImp) }

NOTA: Observar que si se estudia el MER, es posible renombrar las tablas de acuerdo al siguiente esquema:

- A₁₂₁ es la tabla correspondiente a Contribuyentes. (hay que tener en cuenta que en realidad el atributo RazonSocial no aparece)
- A₁₂₂₁ es la tabla correspondiente a Impuestos.
- A₁₂₂₂ es la tabla correspondiente a DebePagar.
- A₁₁ es la unión de las relaciones CorrespA y EsTitular.

Parte 3)

Construya un modelo relacional adecuado para la porción recuadrada en gris del MER presentado. Indique para cada tabla las dependencias funcionales, multivaluadas, de inclusión y subraye las claves primarias.

Solución parte 3)

Para representar los contribuyentes conviene representar las subcategorías y definir la supercategoría como la unión de las proyecciones adecuadas, obteniendo:

PersonasFisicas(Rut, InicioActiv, ciCorresponde)
Empresa(Rut, InicioActiv, RazonSocial, ciTitular)
Contribuyentes(Rut, InicioActiv)

Contribuyentes = $\Pi_{rut, InicioActiv}(PersonasFisicas) \cup \Pi_{rut, InicioActiv}(Empresas)$

Por las totalidades y cardinalidades de las relaciones *correspA* y *esTitular* sabemos que: todo contribuyente que es una PersonaFisica tiene una Persona asociada, a la cual corresponde y que todo contribuyente que es una empresa tiene una Persona asociada, la cual esTitular de la Empresa. De ahí surgen los atributos ciCorresponde y ciTitular respectivamente.

$\Pi_{ciCorresponde}(PersonasFisicas) \subseteq \Pi_{ci}(PF)$
 $\Pi_{ciTitular}(Empresas) \subseteq \Pi_{ci}(TE)$

Para la entidad personas hay que usar dos tablas: una para la entidad propiamente dicha y otra para los teléfonos. Por otra parte, la entidad personas está categorizada en forma exclusiva y total, lo que permite especificar mejor ciertas restricciones que al pasar a relacional se pueden expresar como inclusiones. Por este motivo, parece razonable representar la categorización de personas usando una única tabla y las subcategorías como vistas. De las consideraciones anteriores se obtiene:

Personas(Cl, nombre, TipoPersona)
TE(Cl)
TE = $\Pi_{Cl}(\sigma_{TipoPersona='TE'}(Personas))$
PF(Cl)
PF = $\Pi_{Cl}(\sigma_{TipoPersona='PF'}(Personas))$
TelPer(Ci, Tel)
 $\Pi_{Ci}(TelPer) \subseteq \Pi_{Ci}(Personas)$

El atributo tipoPersona permite indicar si es TE o PF

Ejercicio 3 (25 puntos).

Resolver en álgebra relacional las siguientes consultas:

- a) Obtener el nombre y la dirección de aquellos escenarios que, en alguna fecha, tienen programada la presentación de los primeros premios.

$$A = \prod_{\text{nombre}} (\sigma_{\text{puesto}=1} \text{PREMIOS})$$

$$B = \prod_{\text{codigoescenario}, \text{fecha}, \text{nombregrupo}} (\text{PROGRAMACIÓN}) \% A$$

$$\text{SOL} = \prod_{\text{nombre}, \text{direccion}} (B \triangleright \triangleleft \text{ESCENARIOS})_{\text{codigoescenario}=\text{codigo}}$$

- b) Obtener el nombre de los grupos que solo realizan presentaciones con grupos de su misma categoría.

$$A = \prod_{\text{codigoescenario}, \text{fecha}, \text{nombre}, \text{categoria}} (\text{PROGRAMACIÓN} \triangleright \triangleleft \text{GRUPOS})_{\text{nombregrupo}=\text{nombre}}$$

$$B = \prod_{\$2} (A \triangleright_{\$1=\$5 \wedge \$2=\$6 \wedge \$4 \neq \$8} \triangleleft A)$$

$$\text{SOL} = \epsilon \prod_{\text{nombregrupo}} (\text{PROGRAMACIÓN}) \rightarrow B$$

Resolver en cálculo relacional las siguientes consultas:

- c) Obtener el nombre y la dirección de los escenarios en los cuales todas las actuaciones en primer lugar corresponden a grupos con más de 15 integrantes y que obtuvieron algún premio en el concurso anterior.

```
{ e.nombre, e.direccion / ESCENARIOS(e) ^
  (∀a)(PROGRAMACIÓN(a) ^ a.codigoescenario = e.codigo ^ a.lugar = 1
    →
      (∃g)(GRUPOS(g) ^ g.nombre = a.nombregrupo ^ g.cantidadIntegrantes >= 15
        ^ (∃p)(PREMIOS(p) ^ p.nombre = g.nombre
          )
        )
      )
}
```

- d) Obtener el nombre y la categoría de los grupos, premiados en el concurso anterior, que solo realizan presentaciones en escenarios de un único barrio.

```
{ g.nombre, g.categoria / GRUPOS(g) ^
  (∃p)(PREMIOS(p) ^ p.nombre = g.nombre ^
    (∃a)(PROGRAMACIÓN(a) ^ a.nombregrupo = g.nombre ^
      (∃e)(ESCENARIOS(e) ^ e.codigo = a.codigoescenario ^
        ¬(
          (∃a2)(PROGRAMACIÓN(a2) ^ a2.nombregrupo = g.nombre ^
            (∃e2)(ESCENARIOS(e2) ^ e2.codigo = a2.codigoescenario ^
              e2.barrio ≠ e.barrio
            )
          )
        )
      )
    )
}
```

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

- e) Obtener el nombre y el barrio de aquellos escenarios que tienen plaza de comida y tienen por lo menos 10 fechas distintas donde se incluye alguna presentación completa.

```
select E.nombre,E.barrio
from escenarios E
where E.plazacomida='SI'
and E.codigo IN
      (select A.codigoescenario
       from PROGRAMACIÓN A
       where A.completa='SI'
       group by A.codigoescenario
       having count(distinct fecha) >= 10
      )
```

- f) Obtener nombre de los grupos y su categoría, tales que la mayoría de las presentaciones que realizan son en escenarios con calidad mayor a 3.

```
select G.nombre,G.categoria
from grupos G, PROGRAMACIÓN A, escenarios E
where E.codigo = A.codigoescenario
and A.nombregrupo = G.nombre
and E.calidad > 3
group by G.nombre,G.categoria
having count(*) >
      (select count(*)
       from grupos G2, PROGRAMACIÓN A2, escenarios E2
       where E2.codigo = A2.codigoescenario
       and A2.nombregrupo = G2.nombre
       and E2.calidad <= 3
       and G2.nombre = G.nombre
      )
```

Ejercicio 4 (25 puntos).

La siguiente realidad corresponde a una clínica odontológica en la cual se registra la atención de sus pacientes desde el momento en que abrió por primera vez. Las tablas relevantes de la base son 3:

Pacientes (ciPac, nombre, nacionalidad, teléfono) : Contiene la información de los pacientes de la clínica.

Dentistas (ciDent, nombre, teléfono, especialidad): Contiene la información de todos los dentistas que trabajaron en algún momento en la clínica.

Atencion (ciDent, ciPac, fechaAtención): Contiene la información de cada vez que un dentista atendió a un paciente.

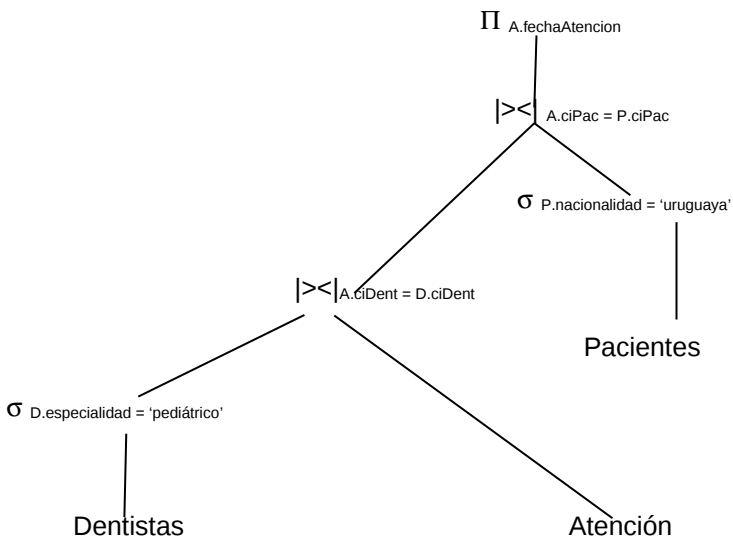
Sea la siguiente consulta SQL:

```
SELECT A.fechaAtencion
FROM Pacientes P, Dentistas D, Atencion A
WHERE A.ciDent = D.ciDent and A.ciPac = P.ciPac and
      P.nacionalidad = "uruguayana" and D.especialidad = "pediátrico";
```

Se conoce la siguiente información, y se dispone de 5 buffers de memoria:

Relación R	bf _R	n _R	Atributos	Índices
Dentistas D	100	100	V(nombre, D) = 25 V(especialidad, D) = 5 (distribución uniforme)	- Índice Prim. de 1 nivel sobre ciDent - Índice Sec. árbol b+ de 3 niveles sobre telefono - Índice Sec. árbol b+ de 3 niveles sobre especialidad
Pacientes P	250	2000	V(nacionalidad, P) = 25 (distribución uniforme)	- Índice Prim. de un nivel sobre ciPac - Índice Sec. árbol b+ de 3 niveles sobre nacionalidad - Índice Sec. árbol b+ de 3 niveles sobre nombre.
Atención A	1000	50000	El 60% de las tuplas corresponden a la atención por parte de dentistas de especialidad 'pediátrico', y el 30% de dichas atenciones fueron hechas a pacientes uruguayos.	Índice Prim. de 1 nivel sobre ciDent, ciPac, fechaAtención.
Dentistas x Atención	90			

Dado el siguiente plan lógico para la consulta:

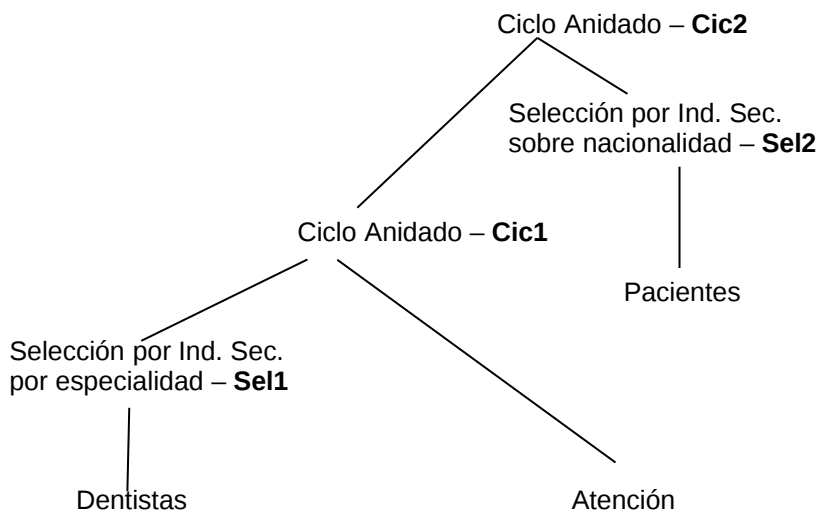


Se pide:

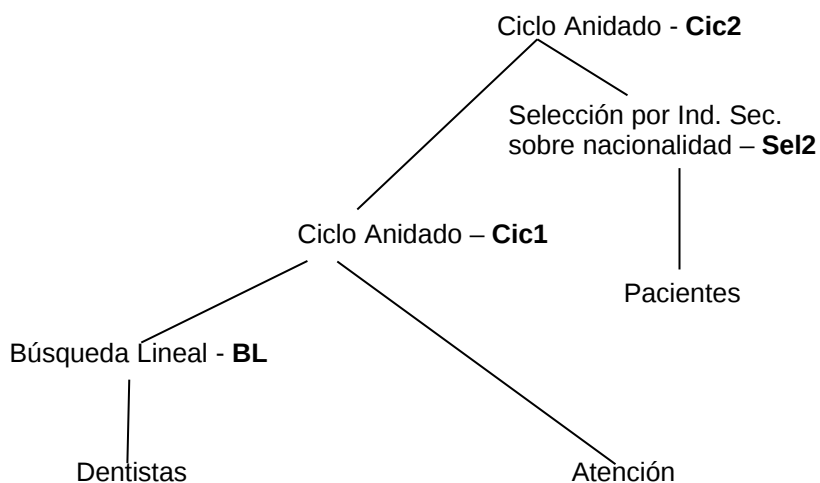
1. Dar 2 planes físicos posibles.
2. Calcular los tamaños intermedios.
3. Calcular los costos para cada plan físico construido en la parte 1), incluyendo los costos de las grabaciones intermedias.

Solución:

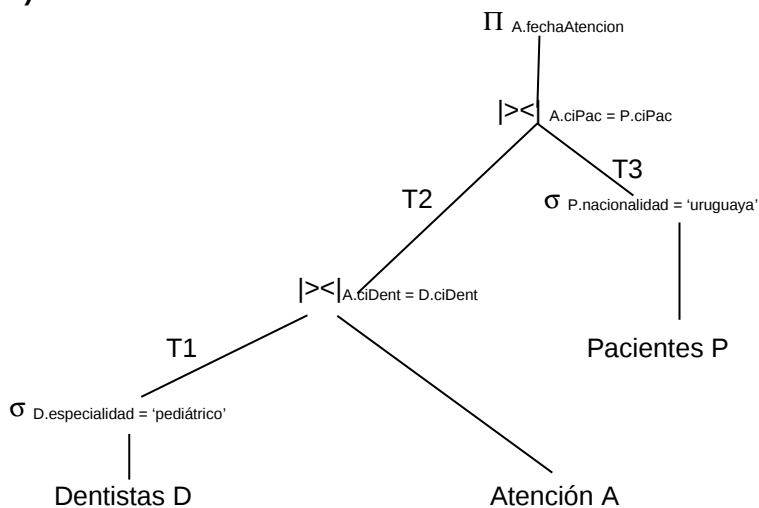
1) **PLAN 1**



PLAN 2



2)



Tamaños:

$$|T1| = |D| * 1/V(\text{especialidad, D}) = 100 / 5 = 20$$

$$|T2| = |A| * 0,6 = 50000 * 0,6 = 30000$$

$$|T3| = |P| * 1/ V(\text{nacionalidad, P}) = 2000 / 25 = 80$$

3)

Costo PLAN 1:

$$\text{Costo Sel1} = x + s = 3 + 20 = 23$$

$$\text{Costo grabación} = \lceil s / bf_D \rceil = \lceil 20 / 100 \rceil = 1$$

$$s = |T1|$$

$$\text{Costo Cic1} = b_R + \lceil b_R / M-2 \rceil * b_S = 1 + \lceil 1/3 \rceil * 50000/1000 = 51$$

$$\text{Costo grabación} = \lceil |T2| / bf_{Rxs} \rceil = \lceil 30000 / 90 \rceil = 334$$

$$\text{Costo Sel2} = x + s = 3 + 80 = 83$$

$$\text{Costo grabación} = \lceil s / bf_P \rceil = \lceil 80 / 250 \rceil = 1$$

$$s = |T3|$$

$$\text{Costo Cic2} = b_R + \lceil b_R / M-2 \rceil * b_S = 334 + \lceil 334 / 3 \rceil * 1 = 334 + 112 = 446$$

$$\text{COSTO TOTAL PLAN 1} = 23 + 1 + 51 + 334 + 83 + 1 + 446 = \mathbf{939}$$

Costo PLAN 2:

$$\text{Costo BL} = 100 / 100 = 1$$

El resto del plan, idem Plan 1.

$$\text{COSTO TOTAL PLAN 2} = 1 + 1 + 51 + 334 + 83 + 1 + 446 = \mathbf{917}$$

Observamos que en este caso no es beneficioso utilizar el índice en la primer selección.