

FUNDAMENTOS DE BASES DE DATOS

Examen Febrero 2008

Solución

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 (26 puntos).

Se desea diseñar una base de datos para gestionar las actividades de una automotora (venta de automóviles).

Los vendedores se identifican por su número de vendedor, se registra su nombre y su antigüedad. Cada vendedor tiene una cartera de clientes. No se registra un cliente en esta realidad si no está asociado a un vendedor, que además debe ser único.

Los clientes se identifican por número de cliente, existen clientes con el mismo número para distintos vendedores, y se registra además su nombre, dirección y sus teléfonos.

Los automóviles pueden ser identificados tanto por su matrícula como por su número de chasis. Otros datos a registrar de los automóviles son su marca, modelo, año, kilometraje y su precio. Existen automóviles utilitarios de los cuales se registra si tuvo un único dueño. También hay automóviles deportivos, de los cuales se registra sus caballos de fuerza. Un automóvil no puede ser utilitario y deportivo a la vez. Se registran también automóviles que no son ni utilitarios ni deportivos.

Los medios publicitarios son medios por los cuales se publican avisos clasificados de los automóviles de la automotora. Un medio publicitario se identifica por su nombre y tipo (ejemplo: observador-digital, observador-escrito), además se registra una breve descripción del mismo.

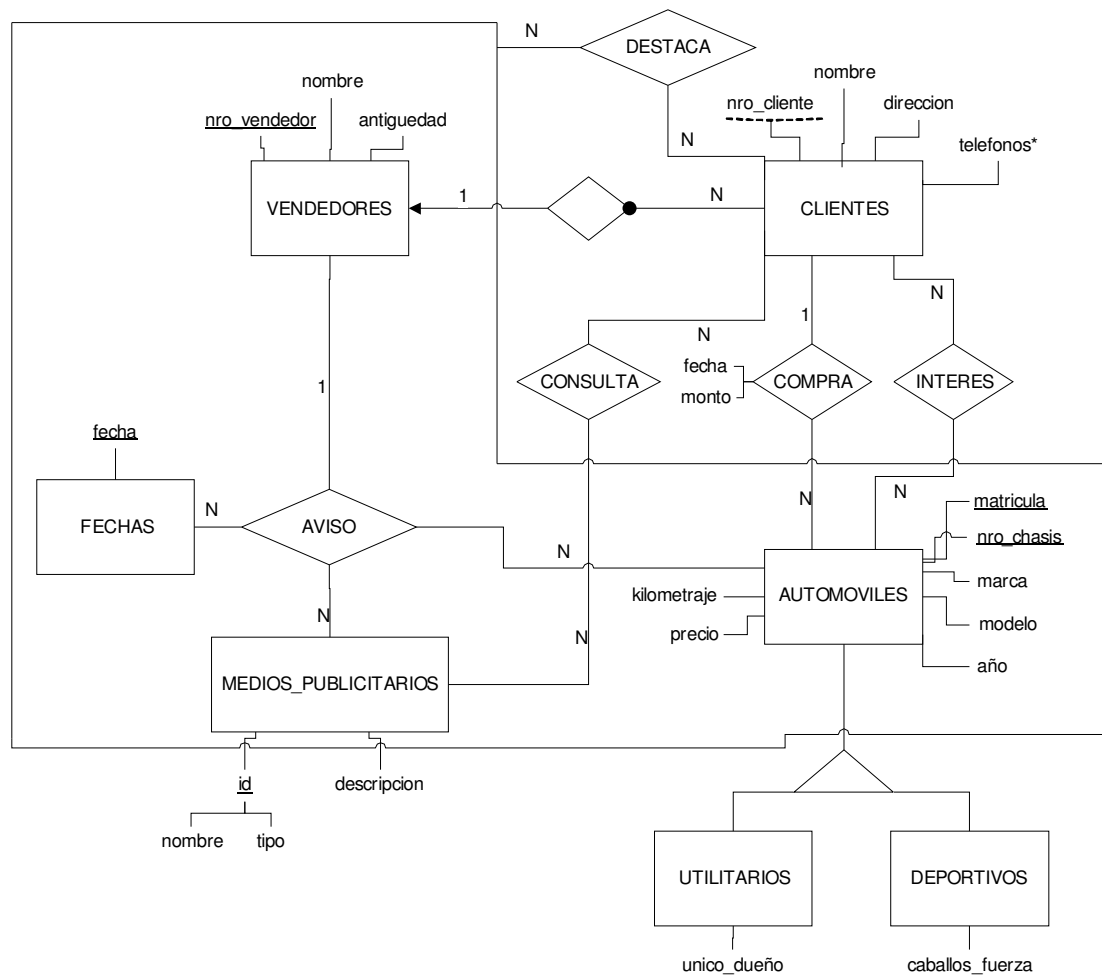
Un aviso lo publica un vendedor en determinado medio publicitario en determinada fecha y es sobre un automóvil en particular. Un vendedor puede publicar varios avisos de distintos automóviles, incluso en la misma fecha y medio publicitario. Por otro lado, no puede haber varios avisos del mismo automóvil, en el mismo medio publicitario y en la misma fecha.

Se registra información de gestión de los clientes, como ser que medios publicitarios consulta cada uno. También se registra que avisos destaca cada cliente. A su vez se registra sobre que automóviles demuestra interés, los cuales no necesariamente tengan que estar asociados a los avisos que destacó.

Se registra la compra de automóviles por parte de los clientes, un cliente puede comprar varios automóviles, pero un automóvil puede ser comprado una sola vez. De la compra se registra su fecha y el monto de la transacción, que no necesariamente tiene que ser el mismo que el precio del automóvil. Un cliente solamente puede comprar un automóvil sobre el cual había demostrado interés.

Se pide Modelo Entidad Relación completo.

Solucion:



RNE:

- $UTILITARIOS \cap DEPORTIVOS = \emptyset$
- Un cliente solamente puede comprar un automóvil sobre el cual había demostrado interés

Ejercicio 2 (4 puntos).

Dada la instancia válida r del esquema relacional R , contestar Verdadero (V) o Falso (F) en cada una de las partes justificando sus respuestas.

Notación: en los valores de la instancia, subíndices distintos significan valores distintos

$R(A, B, C, D)$

r :

A	B	C	D
a_1	b_1	c_1	d_1
a_2	b_1	c_2	d_2
a_3	b_2	c_3	d_3
a_3	b_3	c_4	d_2

- El atributo D no es una clave en el esquema R.
- Si el esquema R satisface la dependencia $AD \rightarrow B$ y $C \rightarrow D$ entonces en el esquema R satisface la dependencia $AC \rightarrow B$.

Solucion:

- VERDADERO.

Para dos valores de D iguales existen dos valores de C distintos por tanto $D \rightarrow C$ no se cumple en la instancia r , por tanto D no es clave en la instancia r , mucho menos en el esquema R.

- VERDADERO

Por propiedad si $\{X \rightarrow Y, WY \rightarrow Z\} \models WX \rightarrow Z$, como en el esquema R se cumple que $C \rightarrow D$ y $AD \rightarrow B$, entonces aplicando la propiedad la instancia r satisface $AC \rightarrow B$.

Ejercicio 3 (22 puntos).

Sea el esquema relación $R(A B C D E G)$ con el conjunto de dependencias funcionales $F = \{ ABD \rightarrow EG, A \rightarrow B, A \rightarrow D, CB \rightarrow DG, AG \rightarrow EB, G \rightarrow BA \}$.

Se pide:

- Hallar todas las claves, justificar la respuesta.
- Hallar un cubrimiento minimal de F en R . Mostrar los pasos seguidos
- Considere la descomposición $\delta = \{ R_1(ABEG), R_2(BCDG) \}$
 - Decir si δ tiene join sin pérdida, justificando.
 - Decir en qué forma normal esta la descomposición, justificando

- d) Llevar R a 3NF con JSP y preservación de dependencias.
- e) Llevar la descomposición de la parte d a BCNF con JSP, justificando.

Solución:

Parte a) calculo de claves

C pertenece a toda clave pues no se genera a partir de otros (aparece solo del lado izquierdo de dfs)

E no pertenece a ninguna clave, no aparece a la derecha de ninguna df por lo cual no genera a nadie

$C^+ = \{ C \} \Rightarrow C$ no es clave

Calculo de claves con 2 atributos

$CA^+ = \{ C, A, B, D, G, E \} \Rightarrow CA$ es clave

$CB^+ = \{ C, B, D, G, A, E \} \Rightarrow CB$ es clave

$CD^+ = \{ C, D \}$

$CG^+ = \{ C, G, B, A, D, E \} \Rightarrow CG$ es clave

No hay mas claves cualquier otra combinación de 3 o mas atributos deberia contener a algunas de claves encontradas por lo cual seria superclave no minimal.

Todas las claves son: CA, CB, y CG

Parte b) cubrimiento minimal

Paso 1: llevar todas las dfs a la forma $X \rightarrow A_i$

$F1 = \{ ABD \rightarrow E, ABD \rightarrow G, A \rightarrow B, A \rightarrow D, CB \rightarrow D, CB \rightarrow G, AG \rightarrow E, AG \rightarrow B, G \rightarrow B, G \rightarrow A \}$

Paso 2: Eliminar atributos redundantes

$AB_{+F1} = \{ A B D E G \} \Rightarrow$ a partir de AB se llega tanto a E como a G en F1
 $\Rightarrow D$ es redundante en las dfs $ABD \rightarrow E, ABD \rightarrow G$

$A_{+F1} = \{ A B D E G \} \Rightarrow$ a partir de A se llega tanto a E como a G en F1
 $\Leftrightarrow B$ también es redundante en las dfs $ABD \rightarrow E, ABD \rightarrow G$

$C_{+F1} = \{ C \}$

$$B_{+F_1} = \{ B \}$$

=> ni C ni B son redundantes en $CB \rightarrow D, CB \rightarrow G$

$$A_{+F_1} = \{ A B D E G \} \Rightarrow \text{se llega a B a partir de A en } F_1 \Rightarrow G \text{ es redundante en } AG \rightarrow B$$

=> se llega a E a partir de A en $F_1 \Rightarrow G$ es redundante en $AG \rightarrow E$

$$F_2 = \{ A \rightarrow E, A \rightarrow G, A \rightarrow B, A \rightarrow D, CB \rightarrow D, CB \rightarrow G, G \rightarrow B, G \rightarrow A \}$$

Paso 3: Eliminar dfs redundantes

$A \rightarrow E$ no es redundante pues es la única forma de llegar a E

$A_{+F_2 - \{A \rightarrow G\}} = \{A, E, B, D, \dots\} \Rightarrow$ no se puede llegar a G a partir de A sin $A \rightarrow G \Rightarrow$ esta no es redundante

$A_{+F_2 - \{A \rightarrow B\}} = \{A, E, G, D, B\} \Rightarrow$ se puede llegar a b a partir de A sin $A \rightarrow B \Rightarrow$ esta **es redundante**

$$F_2' = \{ A \rightarrow E, A \rightarrow G, A \rightarrow D, CB \rightarrow D, CB \rightarrow G, G \rightarrow B, G \rightarrow A \}$$

$A_{+F_2' - \{A \rightarrow D\}} = \{A, E, G, B, \dots\} \Rightarrow$ no se puede llegar a D a partir de A sin $A \rightarrow D \Rightarrow$ esta no es redundante

$CB_{+F_2' - \{CB \rightarrow D\}} = \{C, B, G, A, E, D\} \Rightarrow$ se puede llegar a D a partir de CB sin $CB \rightarrow D \Rightarrow$ esta **es redundante**

$$F_2'' = \{ A \rightarrow E, A \rightarrow G, A \rightarrow D, CB \rightarrow G, G \rightarrow B, G \rightarrow A \}$$

$CB_{+F_2'' - \{CB \rightarrow G\}} = \{C, B\} \Rightarrow$ no se puede llegar a G a partir de CB sin $CB \rightarrow G \Rightarrow$ esta no es redundante

Si se saca $G \rightarrow B$ no habrá forma de llegar a B => esta dfs no es redundante

Si se saca $G \rightarrow A$ no habrá forma de llegar a A => esta dfs no es redundante

$$F_{\min} = \{ A \rightarrow E, A \rightarrow G, A \rightarrow D, CB \rightarrow G, G \rightarrow B, G \rightarrow A \}$$

Otra solución a la que puede llegar: $F_{\min} = \{ A \rightarrow E, A \rightarrow G, A \rightarrow B, A \rightarrow D, CB \rightarrow G, G \rightarrow A \}$

Parte c)

$F = \{ ABD \rightarrow EG, A \rightarrow B, A \rightarrow D, CB \rightarrow DG, AG \rightarrow EB, G \rightarrow BA \}$

$\delta = \{ R_1(ABEG), R_2(BCDG) \}$

Parte c-1) δ tiene JSP ?

$R_1 \cup R_2 = R$

$R_1 \cap R_2 = \{ B G \}$ y $R_1 - R_2 = \{ A E \}$

$BG_{+F} = \{ B G A E B \dots \} \Rightarrow BG \rightarrow AE$ es decir $R_1 \cap R_2 \rightarrow R_1 - R_2 \Rightarrow$ por propiedad JSP

Parte c-2) forma normal de δ ?

$F_{R_1} = \{ A \rightarrow BEG, G \rightarrow BEA \}$ Claves A y G

Toda dfs tiene como parte izquierda una superclave (clave = superclave minimal)
 $\Rightarrow R_1$ esta en BCNF

$F_{R_2} = \{ G \rightarrow BD, CB \rightarrow DG \}$ Claves CB, CG

$CB \rightarrow DG$ tiene como parte izquierda una superclave \Rightarrow esta en BCNF

$G \rightarrow BD$ / G no es superclave \Rightarrow no esta en BCNF

D atributo no primo \Rightarrow G parte de una clave \Rightarrow existe un atributo no primo que depende parcialmente de una clave por lo tanto no esta en 2 FN

\Rightarrow 1 FN

R_1 esta en BCNF y R_2 esta en 1 NF $\Rightarrow \delta$ esta en 1 NF

Parte d) Llevar R a 3NF con JSP y preservación de dependencias.

Partiendo del conjunto minimal se crea un único esquema relación con todas las dependencias de la forma $X \rightarrow A_i$ formando $R_i (X, A_1..A_i)$

$F_{min} = \{ A \rightarrow E, A \rightarrow G, A \rightarrow D, CB \rightarrow G, G \rightarrow B, G \rightarrow A \}$

=> R1 (AEGD)
R2 (CBG)
R3 (GBA)

con una clave R2 incluye una clave de R entonces no es necesario agregar un sub esquema clave
Todos los atributos de R estan en algun/os de los subesquemas

=> resumiendo la descomposición $\rho_{3NF} = (R1, R2, R3)$

$F_{R1} = \{ A \rightarrow E, A \rightarrow G, A \rightarrow D, G \rightarrow A \}$ claves A y G => BCNF

$F_{R2} = \{ CB \rightarrow G, G \rightarrow B \}$ claves CB => 3 NF ($G \rightarrow B$ viola BCNF pues G no es superclave, pero como B es atributo primo cumple 3 NF)

$F_{R3} = \{ G \rightarrow A, G \rightarrow B, A \rightarrow G \}$ claves A y G => BCNF (lados izquierdos son claves)

Parte 3) Llevar la descomposición de la parte d a BCNF con JSP, justificando.

Habrà que descomponer es esquema que viola BCNF según algoritmo visto en clase

R1 (AEGD) => BCNF

R2 (CBG) => 3 NF

$F_{R2} = \{ CB \rightarrow G, G \rightarrow B \}$ $G \rightarrow B$ viola BCNF

=> R2 se descompone / R21 (GB) y R22 (CG)

$FR21 = \{ G \rightarrow B \}$ G clave de R21 => esta en BCNF

$FR22 = \{ \}$ => esta en BCNF

Ejercicio 4 (28 puntos).

Dada la realidad de una base de datos para el manejo de las estadísticas de ventas y gestión de clientes de una empresa:

CLIENTES (idCliente, nomCliente, rut)

Representa información de los clientes, su identificador, nombre y número de registro tributario.

PRODUCTOS (idProducto, nomProducto, categProducto)

Representa información de los productos, su identificador, nombre y su categoría.

GRUPOS (idGrupo, nomGrupo)

Representa información de los grupos de clientes, su identificador y nombre.

GRUPOSCLIENTES (idCliente, idGrupo)

Registra que clientes pertenecen a cada grupo.

INTERES (idCliente, idProducto)

Registra sobre que productos tienen interés cada cliente.

VENTAS (idCliente, idProducto, fecha, cantidad)

Registra las ventas de determinado producto a cada cliente en determinada fecha, así como la cantidad que se vendió.

Además se cumplen las siguientes dependencias de inclusión:

$$\Pi_{idCliente} (GRUPOSCLIENTES) \subseteq \Pi_{idCliente} (CLIENTES)$$

$$\Pi_{idGrupo} (GRUPOSCLIENTES) \subseteq \Pi_{idGrupo} (GRUPOS)$$

$$\Pi_{idCliente} (INTERES) \subseteq \Pi_{idCliente} (CLIENTES)$$

$$\Pi_{idProducto} (INTERES) \subseteq \Pi_{idProducto} (PRODUCTOS)$$

$$\Pi_{idCliente} (VENTAS) \subseteq \Pi_{idCliente} (CLIENTES)$$

$$\Pi_{idProducto} (VENTAS) \subseteq \Pi_{idProducto} (PRODUCTOS)$$

Resolver las siguientes consultas en Algebra Relacional:

1. Devolver nombre de los clientes que tienen interés en todos los productos que se vendieron entre 1/8/2007 y 30/8/2007, (no necesariamente se vendió ese producto al cliente determinado).
2. Devolver nombre del producto, nombre del cliente y fecha pertenecientes a las ventas cuya cantidad fue la mayor.

Resolver las siguientes consultas en Cálculo Relacional:

1. Nombre de los grupos tales que para cada producto, alguno de sus clientes miembros tiene interés en él.
2. Nombre de los grupos tales que cada uno de sus clientes cumple que: o tiene interes en todos los productos de la categoría 'HIGIENE', o se le han vendido todos los productos de la categoría nombrada.

Resolver las siguientes consultas en SQL, sin utilizar vistas:

1. Nombre de los clientes a los que se le vendió todos los productos sobre los cuales declararon interés.
2. Identificador de los clientes que pertenezcan a algún grupo a los cuales se les ha vendido a sus miembros 20 o más productos de la categoría 'HIGIENE' o tienen interés en 5 o más productos de la categoría nombrada.

Solución:

Resolver las siguientes consultas en algebra relacional:

1. Nombre de los clientes que tienen interés en todos los productos que se vendieron entre 1/8/2007 y 30/8/2007, (no necesariamente se vendió ese producto al cliente determinado).

$$A = \Pi_{\text{idProducto}} (\sigma_{\text{fecha} >= '1/8/2007' \wedge \text{fecha} <= '30/8/2007'} (\text{VENTAS}))$$
$$S = \Pi_{\text{nomCliente}} (\text{CLIENTES} * (\text{INTERES} \% A))$$

2. Nombre del producto, nombre del cliente y fecha pertenecientes a las ventas cuya cantidad fue la mayor.

$$A = (\Pi_{\$1, \$2, \$3} \text{VENTAS}) - (\Pi_{\$1, \$2, \$3} (\text{VENTAS} \bowtie_{\$4 < \$8} \text{VENTAS}))$$
$$S = \Pi_{\text{nomProducto}, \text{nomCliente}, \text{fecha}} (A * \text{PRODUCTOS} * \text{CLIENTES})$$

Resolver las siguientes consultas en cálculo relacional:

1. Nombre de los grupos tales que para cada producto, alguno de sus clientes miembros tiene interés en él.

$$\{ g.\text{nomGrupo} / \text{GRUPOS}(g) \wedge (\forall p)(\text{PRODUCTOS}(p) \rightarrow$$
$$(\exists c)(\text{GRUPOSCLIENTES}(c) \wedge c.\text{idGrupo} = g.\text{idGrupo} \wedge$$
$$(\exists i)(\text{INTERES}(i) \wedge i.\text{idCliente} = c.\text{idCliente} \wedge$$
$$i.\text{idProducto} = p.\text{idProducto})) \wedge$$
$$(\exists p1)(\text{PRODUCTOS}(p1)) \}$$

2. Nombre de los grupos tales que cada uno de sus clientes cumple que: o tiene interés en todos los productos de la categoría 'HIGIENE', o se le han vendido todos los productos de la categoría nombrada.

$$\begin{aligned}
 & \{ g.\text{nomGrupo} / \text{GRUPOS}(g) \wedge \\
 & \quad (\forall c1) (\text{GRUPOSCLIENTES}(c1) \wedge c1.\text{idGrupo} = g.\text{idGrupo} \rightarrow \\
 & \quad \quad ((\forall p1) (\text{PRODUCTOS}(p1) \wedge p1.\text{categProducto} = \text{'HIGIENE'} \rightarrow \\
 & \quad \quad \quad (\exists i)(\text{INTERES}(i) \wedge i.\text{idCliente} = c1.\text{idCliente} \wedge \\
 & \quad \quad \quad \quad i.\text{idProducto} = p1.\text{idProducto} \\
 & \quad \quad \quad \quad) \\
 & \quad \quad \quad) \\
 & \quad \quad \vee \\
 & \quad \quad (\forall p1) (\text{PRODUCTOS}(p1) \wedge p1.\text{categProducto} = \text{'HIGIENE'} \rightarrow \\
 & \quad \quad \quad (\exists v)(\text{VENTAS}(v) \wedge v.\text{idCliente} = c1.\text{idCliente} \wedge \\
 & \quad \quad \quad \quad v.\text{idProducto} = p1.\text{idProducto} \\
 & \quad \quad \quad \quad) \\
 & \quad \quad \quad) \\
 & \quad \quad) \wedge (\exists p)(\text{PRODUCTOS}(p) \wedge p.\text{categProducto} = \text{'HIGIENE'}) \\
 & \quad) \wedge (\exists c) (\text{GRUPOSCLIENTES}(c) \wedge c.\text{idGrupo} = g.\text{idGrupo}) \\
 & \}
 \end{aligned}$$

Resolver las siguientes consultas en SQL, sin utilizar vistas:

1. Nombre de los clientes a los que se le vendió todos los productos sobre los cuales declararon interés.

```

SELECT c.nomCliente
FROM CLIENTES c
WHERE NOT EXISTS (
    SELECT *
    FROM INTERES i
    WHERE i.idCliente = c.idCliente
    AND i.idProducto NOT IN (
        SELECT v.idProducto
        FROM VENTAS v
        WHERE v.idCliente=c.idCliente
    )
)

```

2. Identificador de los clientes que pertenezcan a algún grupo a los cuales se les ha vendido a sus miembros 20 o más productos de la categoría 'HIGIENE' o tienen interés en 5 o más productos de la categoría nombrada.

```
SELECT idCliente
FROM GRUPOSCLIENTES
WHERE idGrupo IN (SELECT g.idGrupo
                  FROM PRODUCTOS p, VENTAS v, GRUPOSCLIENTES g
                  WHERE p.idProducto = v.idProducto
                  AND p.categProducto = 'HIGIENE'
                  AND v.idCliente = g.idCliente
                  GROUP BY g.idGrupo
                  HAVING COUNT(*) >= 20)

UNION

SELECT i.idCliente
FROM PRODUCTOS p, INTERES i
WHERE p.idProducto = i.idProducto
AND p.categProducto = 'HIGIENE'
GROUP BY i.idCliente
HAVING COUNT(*) >= 5
```

Otra posible solución:

```
SELECT DISTINCT idCliente
FROM GRUPOSCLIENTES
WHERE idGrupo IN (SELECT g.idGrupo
                  FROM PRODUCTOS p, VENTAS v, GRUPOSCLIENTES g
                  WHERE p.idProducto = v.idProducto
                  AND p.categProducto = 'HIGIENE'
                  AND v.idCliente = g.idCliente
                  GROUP BY g.idGrupo
                  HAVING COUNT(*) >= 20)

OR idCliente IN (SELECT i.idCliente
                 FROM PRODUCTOS p, INTERES i
                 WHERE p.idProducto = i.idProducto
                 AND p.categProducto = 'HIGIENE'
                 GROUP BY i.idCliente
                 HAVING COUNT(*) >= 5)
```

Ejercicio 5 (20 puntos).

El siguiente es parte del esquema de base de datos de una sociedad médica:

Sanatorios (nombre-san, especialidad, direccion, cantidad-salas)

Salas (nombre-san, nro-sala, cant-camas, piso)

Internados (ci-paciente, nombre-san, nro-sala, diagnostico)

Además se tiene la siguiente información sobre las tablas de este esquema:

Indices:

En cada tabla existe un índice primario por la clave primaria correspondiente.

Además existen los siguientes índices:

Tabla	Tipo	Atributos	Nombre
Sanatorios	Indice Secundario	especialidad	Ind-especialidad
Internados	Indice Secundario	nombre-san, nro-sala	Ind-sanat-sala

Tamaños:

Tabla	Cantidad de tuplas
Sanatorios	7
Salas	340
Internados	970

Considere la siguiente consulta sobre el esquema dado:

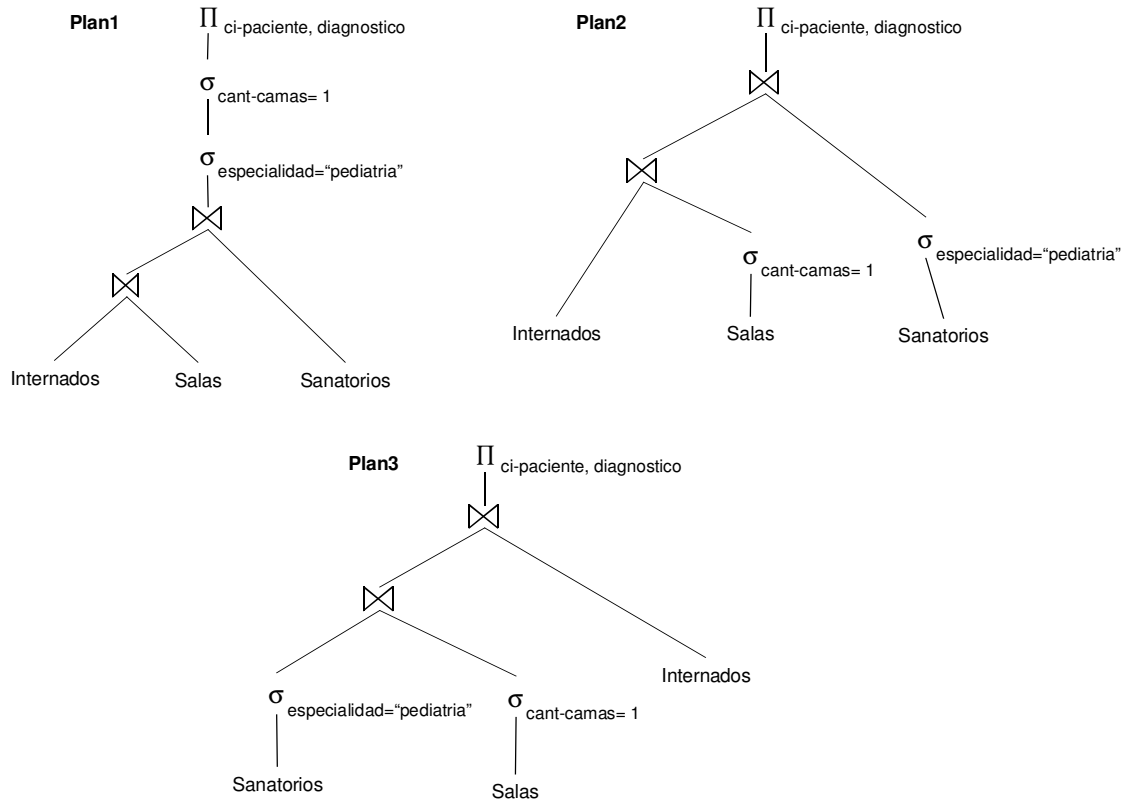
```
SELECT I.ci-paciente, I.diagnostico
FROM Internados I, Salas S1, Sanatorios S2
WHERE I.nombre-san = S1.nombre-san AND
      I.nro-sala = S1.nro-sala AND
      S2.nombre-san = S1.nombre-san AND
      S2.especialidad = 'pediatria' AND
      S1.cant-camas = 1
```

Se pide:

- Dar 3 posibles planes lógicos para la consulta anterior, y decir, justificando, cuál le parece ser el más conveniente.
- Para el plan lógico de la parte a) seleccionado como el más conveniente, dar 2 posibles planes físicos.
- Qué información necesitaría para saber el tamaño del resultado de la consulta?
 - Qué información necesitaría para calcular el costo de grabar el resultado de la consulta en disco?

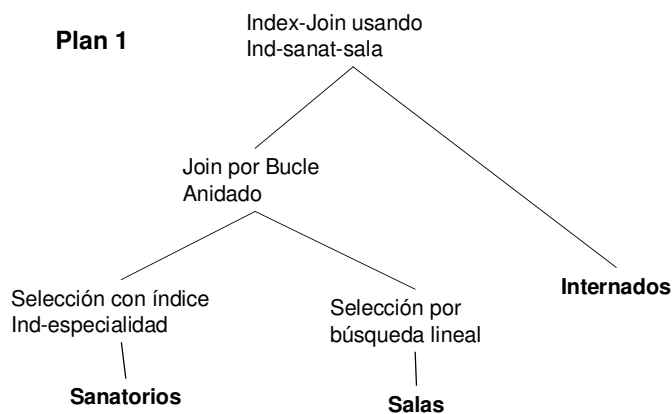
Solución:

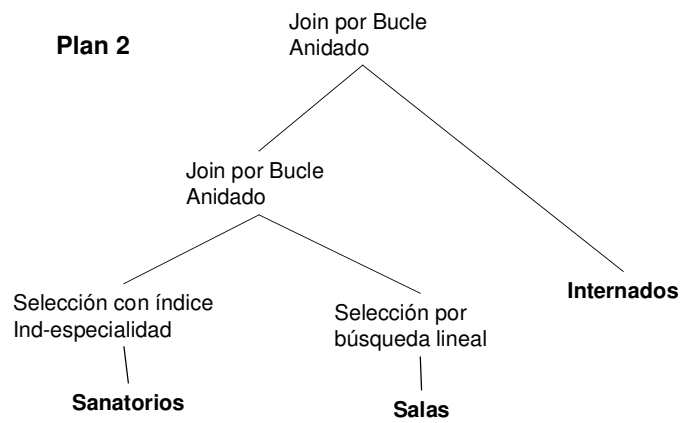
a) Planes Lógicos:



El más conveniente es el Plan 3, ya que realiza las selecciones lo antes posible y además tiene las ramas más restrictivas colocadas a la izquierda del árbol. Esto es lo que indican las heurísticas para optimización de los planes.

b) Planes Físicos:





c)

c1) La cantidad de internados en sanatorios de especialidad 'pediatria' en salas con cant-camas = 1.

c2) Idem c1 y además la cantidad de tuplas que entran en un bloque.