

FUNDAMENTOS DE BASES DE DATOS

Examen Diciembre 2008

La duración del examen es de 3 horas y ½.

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).

Un gobierno departamental desea implementar un sistema de información que le permita mejorar la gestión del transporte de pasajeros.

Del relevamiento surge que existen varias empresas de transporte colectivo de pasajeros, cada una identificada por su nombre. Además, de cada empresa se conoce su dirección (calle y número) y una lista de teléfonos.

Las empresas poseen ómnibus, los cuales se identifican por su matrícula. De cada ómnibus se conoce su marca, modelo, año de empadronamiento y cantidad de asientos. Los ómnibus pueden tener o no asiento para guarda.

Cada parada se identifica por sus coordenadas (latitud y longitud) y además se sabe la calle sobre la cual se ubica.

Los recorridos se identifican por un nombre y representan una secuencia de paradas, donde para cada una de ellas se sabe su posición relativa dentro del recorrido (EJ: la parada 0 es el origen, etc.). Algunos recorridos pueden estar contenidos dentro de otros e interesa modelar esto.

Cuando una empresa decide comenzar a ofrecer un nuevo recorrido se dice que crea una nueva línea, a la cual le asigna un nombre formado por letras y números. Por ejemplo, la empresa Cutcsa realiza el recorrido Ciudad Vieja – Punta Carretas con la línea 121 y el recorrido Punta Carretas – Zonamérica con la línea DM1. El nombre de la línea la identifica dentro de la empresa y cada línea realiza un único recorrido. Cada empresa asigna ómnibus de su flota a cada una de las líneas que ofrece.

En las empresas trabajan empleados, que pueden ser conductores, guardas o personal administrativo. De los empleados se conoce su CI que los identifica, su nombre, teléfono y año de ingreso a la empresa. Para los conductores se desea saber la fecha en que vence su libreta de conducir, para evitar problemas. Existen conductores que a veces trabajan de guardas. El personal administrativo no puede trabajar ni de guarda ni de chofer.

De acuerdo con las reglamentaciones laborales cada ómnibus con asiento para guarda debe tener un guarda asignado, mientras que los ómnibus sin dicho asiento no lo tienen. Dado que el trabajo de conductor es bastante estresante no se asignan conductores a ómnibus sino a ómnibus que recorren cierta línea, intentando minimizar así la necesidad de conocer varios recorridos.

Se pide Modelo Entidad Relación completo.

Ejercicio 2 (7 puntos).

Dado el esquema de relación R, R(A,B,C,D,E,G,H) y el conjunto de dependencias funcionales F que se cumplen sobre R.

$F = \{AB \rightarrow CD, E \rightarrow BG, GA \rightarrow H, C \rightarrow AE, B \rightarrow GC\}$

- a) En caso de ser posible muestre una instancia válida de R con por lo menos 4 tuplas y que si se ejecuta la siguiente consulta sobre esta instancia el resultado es 1

```
Select count(distinct R.H)
From R
```

Si no es posible justifique su respuesta.

- b) En caso de ser posible muestre una instancia válida de R con por lo menos 4 tuplas y que si se ejecuta la siguiente consulta sobre esta instancia el resultado es 1

```
Select count(distinct R.B)
From R
```

Si no es posible justifique su respuesta.

Ejercicio 3 (9 puntos).

Dados los siguientes esquemas relación con sus respectivos conjuntos de dependencias:

$R_1(A,B,C,D) \quad F_1 = \{AB \rightarrow C, D \rightarrow AC\}$

$R_2(A,C,E,G) \quad F_2 = \{E \rightarrow A, G \rightarrow CE\}$

$R_3(E,H,B,C) \quad F_3 = \{H \rightarrow BE, B \rightarrow HC, C \rightarrow EB\}$

$R_4(A,B,E) \quad F_4 = \{A \rightarrow BE, B \rightarrow A, E \twoheadrightarrow BA\}$

- a) ¿Cuál es la máxima forma normal en que se encuentra cada uno de los esquemas relación anteriores? Justificar la respuesta.
- b) ¿Cuál es la máxima forma normal en que se encuentra el esquema de base de datos (R_1, R_2, R_3, R_4) ? Justificar la respuesta.

Ejercicio 4 (12 puntos).

Dado el esquema de relación R, R(A,B,C,D,E,G,H) y el conjunto de dependencias funcionales F que se cumplen sobre R.

$F = \{AB \rightarrow EG, D \rightarrow CH, BE \rightarrow D, G \rightarrow H, AC \rightarrow DG\}$

- 1) Determinar si la descomposición $\rho_1 = (R_1, R_2)$ con

- $R_1(A,D,E,G)$
- $R_2(A,D,C,B,H)$

es una descomposición de R con join sin pérdida respecto a F. Justificar la respuesta.

- 2) Dada la siguiente descomposición de R, $\rho_2 = (R_1, R_2, R_3)$ con

- $R_1(A,B,D,H)$
- $R_2(B,E,C,H)$
- $R_3(A,D,G)$

- a) Calcular la proyección de F en cada uno de los subesquemas de ρ_2 .
- b) Determinar si las dependencias de F : $AB \rightarrow G$ y $BE \rightarrow D$ se preservan en la descomposición ρ_2 . Justificar la respuesta.
- c) ¿Es posible que en R_2 además de las dependencias proyectadas de F se cumpla la dependencia multivaluada $C \twoheadrightarrow H$? Justificar la respuesta.

Ejercicio 5 (27 puntos).

Una cadena de centros especializados en **tratamientos** estéticos desea llevar la información de su negocio en una base de datos. A continuación se describen las tablas que la componen.

CENTROS_ESTETICOS (nombre-ce, dir-ce, tel-ce, email-ce)

Contiene información de cada uno de los centros estéticos de la cadena. De cada uno se conoce su dir, teléfono de contacto y email.

ESPECIALISTAS (ci-emp, nombre-emp, dir-emp, tel-emp, especialidad, nombre-ce)

Contiene información del personal especialista que trabaja en los centros. De cada uno de ellos se conoce su CI, su nombre completo, su dirección y teléfono, el único centro de la cadena en el cual trabaja y su especialidad (EJ: *masajista, cosmólogo, nutricionista* etc.)

NO_ESPECIALISTAS (ci-emp, nombre-emp, dir-emp, tel-emp)

Contiene información de personal no especialista que trabaja en los centros. De cada uno de ellos se conoce su CI, su nombre completo, su dirección y teléfono

TRABAJAN (nombre-ce, ci-emp, tarea)

Contiene información acerca de la tarea que realiza el personal no especialista en los diferentes centros.

ACTIVIDADES (cod-act, nombre-act, desc-act)

Contiene información acerca de las actividades que se realizan en los centros estéticos. De cada actividad se conoce un código que la identifica, un nombre (EJ: *masajes*) y una descripción de la misma.

OFRECEN_ACTIVIDADES (cod-act, nombre-ce)

Representa las actividades que se ofrecen en cada centro.

TRATAMIENTOS (nombre-ce, cod-trat, nombre-trat, desc-trat)

Contiene información acerca de los tratamientos ofrecidos en cada centro. De cada tratamiento se conoce su nombre (EJ: *tratamiento para adelgazar*) y una descripción del mismo.

CONFORMAN (nombre-ce, cod-trat, cod-act)

Cada tratamiento se compone de actividades. Esta relación se representa en esta tabla.

CLIENTES (ci-cli, nombre-cli, tel-cli)

Contiene información acerca de los clientes que se atienden en la cadena de centros. De cada cliente se conoce su CI, su nombre y su teléfono

REALIZAN (ci-cli, fecha, hora, cod-act, nombre-ce, cod-trat, ci-emp, observaciones)

Contiene información acerca de las actividades realizadas por cada cliente en el marco de un tratamiento. De cada una de estas instancias se conoce la fecha y hora de comienzo, el código de la actividad, el tratamiento, el especialista que supervisó la actividad y sus observaciones.

En este esquema se cumple lo siguiente:

- No existen tablas vacías.
- $\Pi_{ci-emp}(\text{Especialistas}) \cap \Pi_{ci-emp}(\text{No_Especialistas}) = \emptyset$
- $\Pi_{nombre-ce}(\text{Especialistas}) \subseteq \Pi_{nombre-ce}(\text{Centros_Esteticos})$
- $\Pi_{ci-emp}(\text{Trabajan}) \subseteq \Pi_{ci-emp}(\text{No_Especialistas})$
- $\Pi_{nombre-ce}(\text{Trabajan}) \subseteq \Pi_{nombre-ce}(\text{Centros_Esteticos})$
- $\Pi_{cod-act}(\text{Ofrecen_Actividades}) \subseteq \Pi_{cod-act}(\text{Actividades})$
- $\Pi_{nombre-ce}(\text{Ofrecen_Actividades}) \subseteq \Pi_{nombre-ce}(\text{Centros_Esteticos})$
- $\Pi_{nombre-ce}(\text{Tratamientos}) \subseteq \Pi_{nombre-ce}(\text{Centros_Esteticos})$
- $\Pi_{cod-act}(\text{Conforman}) \subseteq \Pi_{cod-act}(\text{Actividades})$
- $\Pi_{cod-trat, nombre-ce}(\text{Conforman}) \subseteq \Pi_{cod-trat, nombre-ce}(\text{Tratamientos})$
- $\Pi_{cod-act, cod-trat, nombre-ce}(\text{Realizan}) \subseteq \Pi_{cod-act, cod-trat, nombre-ce}(\text{Conforman})$
- $\Pi_{ci-emp, nombre-ce}(\text{Realizan}) \subseteq \Pi_{ci-emp, nombre-ce}(\text{Especialistas})$
- $\Pi_{ci-cli}(\text{Realizan}) \subseteq \Pi_{ci-cli}(\text{Clientes})$

Resolver en álgebra relacional las siguientes consultas:

- i. Devolver la cédula de los clientes que realizaron todas las actividades que conforman al tratamiento de código "T1" del centro de nombre "Casa central", sin importar en que centro y en el marco de que tratamiento las hayan realizado.
- ii. Devolver el nombre de aquellos especialistas que siempre supervisan la misma actividad.

Resolver en cálculo relacional las siguientes consultas:

- iii. Devolver el nombre de aquellas actividades realizadas por los clientes que nunca han sido supervisadas por masajistas.
- iv. Devolver el nombre de aquellos clientes que han realizado actividades que forman parte de más de un tratamiento.

Resolver en SQL, sin utilizar vistas ni subconsultas en el from, las siguientes consultas:

- v. Devolver el nombre de aquellos empleados no especialistas que realizan tareas de "limpieza" en centros en los que no trabaja ningún "nutricionista"

Ejercicio 6 (20 puntos).

Las siguientes relaciones forman parte del esquema de una inmobiliaria en el contexto de una realidad reducida donde se manejan únicamente alquileres de casas.

Casas (idcasa, dir, m2, cantdor, m2patio, precioalquiler)

Esta tabla contiene los datos de las casas disponibles para alquilar en la inmobiliaria.

Interesados (ci, nombre, telefono, cantpersonas, tipodegarantia, cantdor)

Esta tabla contiene los datos de las personas interesadas en alquilar alguna casa, la cantidad de dormitorios que buscan, y el tipo de garantía que tienen. Los tipos de garantía posibles son: "CGN", "ANDA", "Particular" y "No tiene".

Observación: la inmobiliaria registra en **Interesados** las personas que hicieron ofertas sobre alguna de las casas disponibles así como aquellas personas que están buscando y aún no ofertaron.

Ofertas (ci, idcasa, precioofertado, observaciones)

En esta tabla se registran las ofertas concretas que realizan los interesados sobre las casas disponibles para alquilar, y si corresponde alguna observación al respecto.

En este esquema se cumplen las siguientes dependencias de inclusión:

vi. $\Pi_{idcasa}(\text{Ofertas}) \subseteq \Pi_{idcasa}(\text{Casas})$

vii. $\Pi_{ci}(\text{Ofertas}) \subseteq \Pi_{ci}(\text{Interesados})$

Considere la siguiente información:

1. El 15% de las casas tienen 1 dormitorio, el 50% de las casas tienen 2 dormitorios, y el resto tienen más de 2 dormitorios.
2. Los tipos de garantía en la tabla Interesados tienen distribución uniforme.
3. El 80% de las ofertas se realizan sobre casas con 2 dormitorios, y de estas ofertas, el 30% son de interesados con garantía de tipo CGN.

El tamaño de cada tabla es el siguiente:

- n. Casas: 500 tuplas
- o. Interesados: 1700 tuplas
- p. Ofertas: 800 tuplas

El factor de bloqueo para cada tabla es:

- q. Casas: 5
- r. Interesados: 10
- s. Ofertas: 10
- t. Casas \gg Ofertas: 3

La cantidad de buffers disponibles es 5.

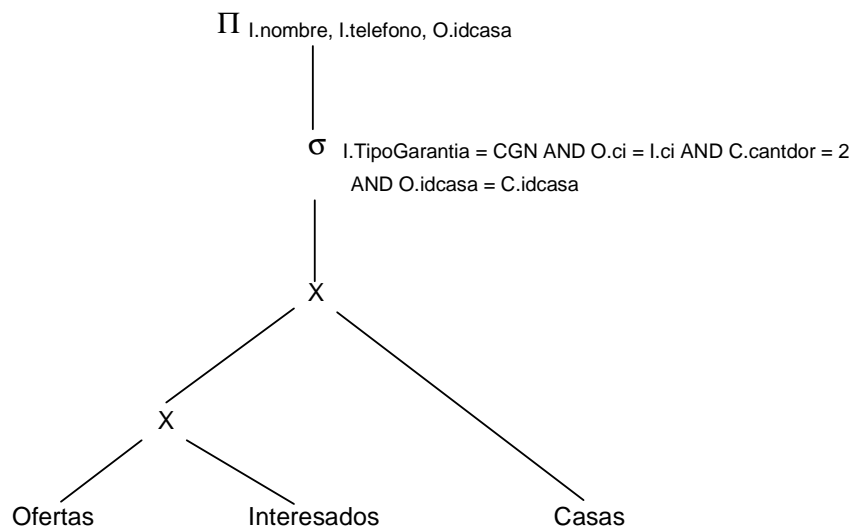
Los índices existentes son los siguientes:

Nombre Tabla	Nombre Índice	Tipo Índice	Por Atributo
Casas	IndiceCasa_idcasa	Primario	idcasa
Casas	IndiceCasa_dor	Secundario	cantdor
Casas	IndiceCasa_precio	Secundario	precioalquiler
Interesados	IndiceInt_ci	Primario	ci
Interesados	IndiceInt_dor	Secundario	cantdor
Interesados	IndiceInt_gar	Secundario	tipogarantia
Ofertas	IndiceOf_ci_idcasa	Primario	(ci, idcasa)
Ofertas	IndiceOf_ci	Secundario	ci
Ofertas	IndiceOf_idcasa	Secundario	idcasa
Ofertas	IndiceOf_precio	Secundario	precioofertado

Nota: Todos los índices nombrados son árboles B+ con cantidad de niveles: 2.

SE PIDE:

Parte a) Escribir la consulta SQL que se corresponde al siguiente árbol canónico:



- Parte b)** A partir del árbol canónico de la parte a) obtenga el plan lógico optimizado mediante las heurísticas vistas en el curso (todas). Justifique su respuesta
- Parte c)** Calcule los tamaños en tuplas (sin considerar las proyecciones) de los resultados de las selecciones y los joins.
- Parte d)** Dar un plan físico para el plan lógico de la parte b) (sin considerar las proyecciones), usando de manera eficiente los índices existentes.
- Parte e)** Calcular el costo del plan físico propuesto en la parte d). Considerar costos de grabaciones intermedias, excluyendo la del resultado final. Justifique su respuesta.

Implementaciones de los Operadores.

Oper.	Algoritmo	Costo	Condición	Organización
$\sigma_c(R)$	Búsqueda Lineal	b_R peor caso, $b_R/2$ promedio	Cualquier Caso	Cualquiera
	Búsqueda Binaria	$\log_2 b_R + \lceil s/bf_R \rceil - 1$	Cualquier caso	Registros ordenados
	Índice Primario	$x + 1$	Por igualdad a un valor	Registros Ordenados
	Hash	1 o 2 según el tipo	Por igualdad a un valor	Cualquiera
	Índices Primario	$x + (b/2)$ (promedio)	Por relación de orden.	Índice ordenado
	Índice Cluster	$x + \lceil s/bf_R \rceil$	Cualquier Caso	Registros Ordenados

CSI-INCO

Fundamentos de Bases de Datos

45

--

Implementaciones de los Operadores.

Oper.	Algoritmo	Costo	Cond.	Organización
$\sigma_c(R)$	Índice secundario B+	$x + s$ peor caso	Cualquier Caso	Cualquiera
	Grabacion Intermedia	s/bf_R	Cualquier caso	Cualquiera
$R X _cS$	Loop Anidado (registros)	$b_R + (n_R * b_s)$	Cualquier caso	Cualquiera
	Loop Anidado (bloque)	$b_R + \lceil b_R/(M-2) \rceil * b_s$	Cualquier caso	Cualquiera
	Sort Merge	$b_R + b_s + \text{costo ords.}$	Cualquier caso	índice en disco
	Index join	$b_R + n_R * Z$	Cualquier caso	índice en disco

CSI-INCO

Fundamentos de Bases de Datos

46

donde Z depende del tipo de índice:

secundario: $Z = x + sS$, cluster: $Z = x + \lceil sS/bfS \rceil$, primario: $Z = x+1$, hash = h