

FUNDAMENTOS DE BASES DE DATOS

Examen Julio 2002

Presentar la resolución del examen:

- Con las hojas adicionales numeradas y escritas de un solo lado.
- Con las hojas escritas a lápiz.
- Poner cédula de identidad y nombre en cada hoja (incluidas estas).
- Escrito en forma prolija.
- Las opciones elegidas se deben marcar poniendo el identificador de la opción en un círculo claramente identificado.
- Poner la cantidad de hojas adicionales entregadas en la primer hoja.

Ejercicio 1 – Modelo Entidad-Relación (30 puntos)

Se desea modelar la información sobre las publicaciones existentes en una biblioteca perteneciente a un instituto de enseñanza e investigación

Las publicaciones se identifican por un código y además se guarda su nombre y los números de copias existentes y disponibles en la biblioteca. Las publicaciones pueden ser de tres tipos: libros, revistas y publicaciones del instituto. Sobre las revistas interesa saber la fecha de publicación. Sobre las publicaciones del instituto interesa saber si son tesis de grado, tesis de posgrado o reporte técnico. Algunos libros son compendios de artículos presentados en conferencias. Las revistas están formadas por un conjunto de artículos.

Los artículos tienen un número que los identifica dentro de la revista o del libro del que forman parte. Además tienen una descripción (el resumen del artículo) y un conjunto de palabras clave. Los artículos de conferencia tienen también el dato de la sesión de la conferencia en la que fueron presentados.

La información de los autores es muy importante para todas las publicaciones. En el caso de los libros que no son de artículos y las publicaciones del instituto interesa saber los autores que les corresponden. Para las publicaciones integradas por artículos no se guardan autores, sino que los autores se llevan con respecto a cada artículo. Sobre los autores interesa saber el nombre (que los identifica), nacionalidad y la institución a la cual pertenecen.

También se guarda información sobre los socios de la biblioteca, los préstamos y reservas de éstos. De los socios interesa la cédula de identidad, el nombre, la dirección, el teléfono, y si son estudiante o docente. Los socios pueden sacar una publicación en préstamo o reservarla. En el caso de los préstamos interesa la fecha de retiro de la publicación y la fecha de devolución. La cantidad de publicaciones prestada no puede ser mayor que la cantidad de copias existentes en la biblioteca.

Cada libro es editado por al menos una editorial. Una editorial puede editar un mismo libro en distintas fechas, siendo éstas distintas ediciones para una misma pareja libro-editorial. De cada editorial se guarda un código que la identifica, el nombre, la dirección y el teléfono. Para cada edición interesa saber en qué idioma está escrita.

Se pide: Modelo Entidad-Relación completo.

Ejercicio 2 – Diseño Relacional (20 puntos)

a)

Una empresa se dedica a brindar servicios de contralor a una parte del sistema financiero de un determinado país. De cada banco se conoce un código que lo identifica (cod_b), su nombre (nom_b), la dirección de su casa central (dcc_b) y los teléfonos de dicha casa central (tcc_b).

De cada cuenta, se conoce el banco en el cual se abre la cuenta, un identificador que asigna el banco (cta), las C.I. de los habilitados para operar con la cuenta (CI), el tipo de cuenta ($c.ahorro$, cta . Corriente, portafolio) (t_cta), la fecha y hora de la apertura ($fecha_cta$) y el saldo al instante ($saldo_cta$).

De cada cuenta, además, se conocen todos los movimientos realizados desde que fue abierta. De las extracciones de una cuenta se conoce la fecha y hora ($fecha_e$), el monto ($monto_e$) y la cédula de quien hizo el movimiento. Lo mismo sucede con los depósitos ($fecha_d$, $monto_d$). De las transferencias, se conoce la fecha y hora ($fecha_t$), la cuenta destino ($ctadest_t$), el código de banco de destino ($bancodest_t$), el monto ($monto_t$) y la cédula de quien realizó la transferencia.

De los clientes se conoce su cédula de identidad (CI), su nombre (nom_cli), su dirección (dir_cli), sus teléfonos (tel_cli) (que pueden ser celulares).

OBSERVACIONES:

Note que no hay un tipo de movimiento sino que se consideran todos por separado.

No hay dos bancos con el mismo nombre, ni en la misma dirección, ni comparten teléfonos.

Una persona puede tener abiertas todas la cuentas que quiera en cualquier banco que quiera.

Se asume que no puede haber dos movimientos sobre la misma cuenta a la misma hora del mismo día.

SE PIDE:

1. De la siguiente lista, seleccione las dependencias que Ud. considera que se cumplen en la realidad anterior:

a. $tel_cli \rightarrow CI$

b. $cod_b \rightarrow nomb_b, dcc_b$

c. $tcc_b \rightarrow cod_b$

d. $cta \rightarrow cod_b$

e. $CI \rightarrow nom_cli, dir_cli$

f. $cta, cod_b \rightarrow CI$

g. $nom_b \rightarrow cod_b$

h. $dcc_b \rightarrow cod_b$

i. $cta \rightarrow t_cta, fecha_cta, saldo_cta$

j. $cta, cod_b \rightarrow monto_d, CI$

k. $cta, cod_b \rightarrow t_cta, fecha_cta, saldo_cta$

l. $cta, cod_b, fecha_d \rightarrow monto_d, CI$

m. $cta, cod_b, fecha_t \rightarrow monto_t, CI, ctadest_t$

n. $cta, cod_b, fecha_e \rightarrow monto_e, CI$

o. $ctadest_t, banco_dest \rightarrow cta, cod_b$

2. Anote otras dependencias funcionales que le parezca que se cumplan y no estén entre las anteriores:

3. Suponiendo que el siguiente esquema es la relación universal correspondiente a la realidad y considerando las dependencias que seleccionó en la parte anterior, indique en qué forma normal se encuentra el esquema y justifique su respuesta.

R(CI,nom_cli,tel_cli,dir_cli,cta,t_cta,fecha_cta,saldo_cta,fecha_t,monto_t,ctadest_t, fecha_e,monto_e,fecha_d,monto_d,cod_b,dcc_b,tel_b,nom_b)

- b) Sea $R(A,B,C,D,E,G)$ un esquema relacional y considere el siguiente conjunto de dependencias funcionales sobre R : $F=\{ABC \rightarrow E, BG \rightarrow E, DA \rightarrow G, ABD \rightarrow GC, E \rightarrow C\}$

1. De la siguiente lista de conjuntos de atributos, seleccione todas las claves:

- | | |
|----------|----------|
| a. AD | g. ABCD |
| b. ABC | h. ABEG |
| c. ABCDE | i. ABE |
| d. DBC | j. ABG |
| e. AB | k. ABDEG |
| f. ABD | |

2. Justifique por qué las anteriores son todas las claves.

3. Del siguiente conjunto de dependencias funcionales, seleccione las que son resultado del primer paso de aplicar el algoritmo de cubrimiento minimal al conjunto original.

- | | |
|------------------------|-----------------------|
| a. $CAB \rightarrow E$ | f. $AD \rightarrow G$ |
| b. $ABD \rightarrow C$ | g. $E \rightarrow C$ |
| c. $ABD \rightarrow G$ | h. $GA \rightarrow D$ |
| d. $BD \rightarrow G$ | i. $AB \rightarrow C$ |
| e. $BG \rightarrow E$ | j. $CD \rightarrow E$ |

4. Del siguiente conjunto de dependencias funcionales, seleccione las que son resultado del segundo paso de aplicar el algoritmo de cubrimiento minimal a su selección anterior.

- | | |
|------------------------|-----------------------|
| a. $BAC \rightarrow E$ | f. $DA \rightarrow G$ |
| b. $ABD \rightarrow G$ | g. $GA \rightarrow D$ |
| c. $DAB \rightarrow C$ | h. $E \rightarrow C$ |
| d. $BD \rightarrow G$ | i. $AB \rightarrow C$ |
| e. $BG \rightarrow E$ | j. $CD \rightarrow E$ |

5. Del siguiente conjunto de dependencias funcionales, seleccione las que son resultado del primer paso de aplicar el algoritmo de cubrimiento minimal a su selección anterior.

- | | |
|------------------------|-----------------------|
| a. $ABC \rightarrow E$ | f. $DA \rightarrow G$ |
| b. $ABD \rightarrow C$ | g. $E \rightarrow C$ |
| c. $ABD \rightarrow G$ | h. $GA \rightarrow D$ |
| d. $BD \rightarrow G$ | i. $AB \rightarrow C$ |
| e. $BG \rightarrow E$ | j. $CD \rightarrow E$ |

6. Aplique un paso del algoritmo para obtener una descomposición en 4NF con join sin pérdida, considerando el esquema original R y el cubrimiento minimal hallado en la parte anterior. Aplique ese paso para la primer dependencia que no cumpla 4NF considerando el orden de las opciones. Proyecte las dependencias funcionales sobre la descomposición resultante.

7. La dependencia ABD→G, se perdió en la descomposición? SI NO

Ejercicio 3 – Consultas (30 puntos)

La **ciudad universitaria** es un conjunto de residencias estudiantiles ubicadas en un mismo predio, que dan alojamiento a jóvenes universitarios de todo el país.

Cada residencia se construyó con fondos de una institución, que puede ser una Intendencia Departamental o una fundación particular.

En cada residencia pueden habitar estudiantes de cualquier departamento del país, que estén estudiando en cualquier facultad o escuela de la Universidad de la República.

Para alojarse en una residencia, un estudiante debe hacer un contrato que tiene validez por ese año. Para alojarse al año siguiente debe hacer un nuevo contrato.

Las instituciones otorgan algunas becas para alojar gratuitamente a estudiantes con pocos recursos económicos.

Se tiene una base de datos con las siguientes relaciones:

RESIDENCIAS (cod-res, nom-res, cod-ins, año-constr, arquitecto, num-hab)

Esta relación contiene los datos de todas las residencias de la ciudad universitaria. Sus atributos son: código de la residencia, nombre de la residencia, código de la institución que la construyó, año de construcción, arquitecto que la proyectó y número de habitaciones.

INSTITUCIONES (cod-ins, nom-ins, cant-becas)

Esta relación contiene los datos de las instituciones que construyeron residencias, ya sean intendencias departamentales o fundaciones de particulares. Sus atributos son: código de la institución, nombre de la institución y cantidad de becas que otorga.

I-INTENDENCIAS (cod-ins, departamento, programa, encargado, tel-contacto)

Esta relación contiene los datos de las instituciones que corresponden a proyectos de Intendencias Departamentales. Todas las instituciones de esta tabla están también en la tabla instituciones. Sus atributos son: código de la institución, departamento, programa o proyecto de la intendencia, encargado del proyecto y teléfono de contacto.

I-PARTICULARES (cod-ins, nom-fundador, encargado, tel-contacto)

Esta relación contiene los datos de las instituciones que corresponden a fundaciones de particulares (los que no corresponden a Intendencias Departamentales). Todas las instituciones de esta tabla están también en la tabla instituciones. Sus atributos son: código de la institución, nombre del fundador, encargado del proyecto y teléfono de contacto.

ESTUDIANTES (CI, nombre, depto-origen, edad)

Esta relación contiene los datos de todos los estudiantes que se han alojado en alguna residencia de la ciudad universitaria. Sus atributos son: CI del estudiante, nombre del estudiante, departamento del que viene, y edad.

EST-FAC (CI, facultad, año-curso, promedio)

Esta relación indica en qué facultades (o escuelas) estudia cada estudiante. Cada estudiante de la relación estudiantes debe estudiar en al menos una facultad, y cada estudiante de esta relación debe estar en la relación estudiantes. Sus atributos son: CI del estudiante, nombre de la facultad, año que está cursando y promedio de notas.

CONTRATOS (cod-res, CI, año, becado, habitacion, teléfono)

Esta relación indica en qué residencia se ha alojado cada estudiante en cada año, ya que los contratos de alojamiento son anuales. Las residencias y estudiantes están en las tablas correspondientes. Sus atributos son: código de la residencia, CI del estudiante, año del contrato, indicación de existencia de una beca (valores SI o NO), habitación que ocupa y teléfono de la habitación.

SE PIDE:

1) Escribir en Algebra Relacional las siguientes consultas:

- a) Devolver código de institución y teléfono de contacto de todas las instituciones que otorgan 10 o más becas.
- b) Devolver los códigos de las residencias construidas por intendencias, que sólo han alojado a estudiantes del mismo departamento que la intendencia.

2) Escribir en Cálculo Relacional las siguientes consultas:

- a) Devolver CI y nombre de los estudiantes que sólo se han alojado un año.
- b) Devolver los códigos de las instituciones, para las cuales todas sus residencias han alojado algún becado.

3) Escribir en SQL las siguientes consultas:

- a) Listar los códigos y nombres de las residencias construidas por la “Fundación Juan Perez” que tienen ocupadas todas sus habitaciones en el año 2002.
- b) Listar código de residencia y código de institución de las residencias con más de 100 habitaciones en las que no se ha alojado ningún estudiante mayor de 25 años.

Ejercicio 4 – Optimización (20 puntos)

Dados el esquema relacional del Ejercicio 3 y la siguiente consulta

$\Pi_{\text{nombre, nom-res}} [(\sigma_{\text{edad} > 25} \text{Estudiantes}) \bowtie \text{Contratos} \bowtie (\sigma_{\text{año-constr} = 1990} \text{Residencias})]$

- 1) Dar un plan lógico de la consulta, aplicando las heurísticas y calculando los tamaños intermedios.
- 2) Elegir una implementación para el primer join que se ejecuta según el plan lógico y calcular su costo estimado. (Aclaración: No tener en cuenta las proyecciones para el cálculo.)

DATOS:

	ESTUDIANTES	CONTRATOS	RESIDENCIAS
Cantidad tuplas	4500	6000	54
Indices primarios	ci (niveles: 2)		cod-res (niveles: 1)
Indices secundarios (B+)	edad (niveles: 3)		
Cantidad de tuplas por bloque	50	60	30
Observaciones	- Edades: entre 18 y 35. Distribución uniforme.	- Residencias distintas: 54. - Estudiantes distintos: 2300 - Distribución uniforme.	- Años distintos en que se construyeron residencias: 27. Distribución uniforme.

	ESTUDIANTES \bowtie CONTRATOS	CONTRATOS \bowtie RESIDENCIAS
Cantidad de tuplas por bloque	30	20

Fórmulas para cálculo de costo y de tamaños:

Join (R,S)	Nested Loop (ciclo anidado) sin utilizar índices	$b_R + b_R * b_S$
	Nested Loop (ciclo anidado) utilizando índice secundario para recuperar tuplas que matchean	$b_R + R * (x + s)$
	Nested Loop (ciclo anidado) utilizando índice primario para recuperar tuplas que matchean	$b_R + R * (x + 1)$
Escribir resultados en disco	Join (R,S)	$(js * R * S) / fbl_{RS}$
Selectividad	Selección (atrib. A)	$1 / V(A,R)$
	Join (atrib. A)	$1 / \text{Max}(V(A,R), V(A,S))$

Notación: **b** – cantidad de bloques

fbl – factor de bloqueo

x – cantidad de niveles del índice

|R| - cantidad de tuplas de R

s – selectividad de la selección

js – selectividad del join

V(A,R) - cantidad de valores distintos del atributo A en R

