

FUNDAMENTOS DE BASES DE DATOS

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.

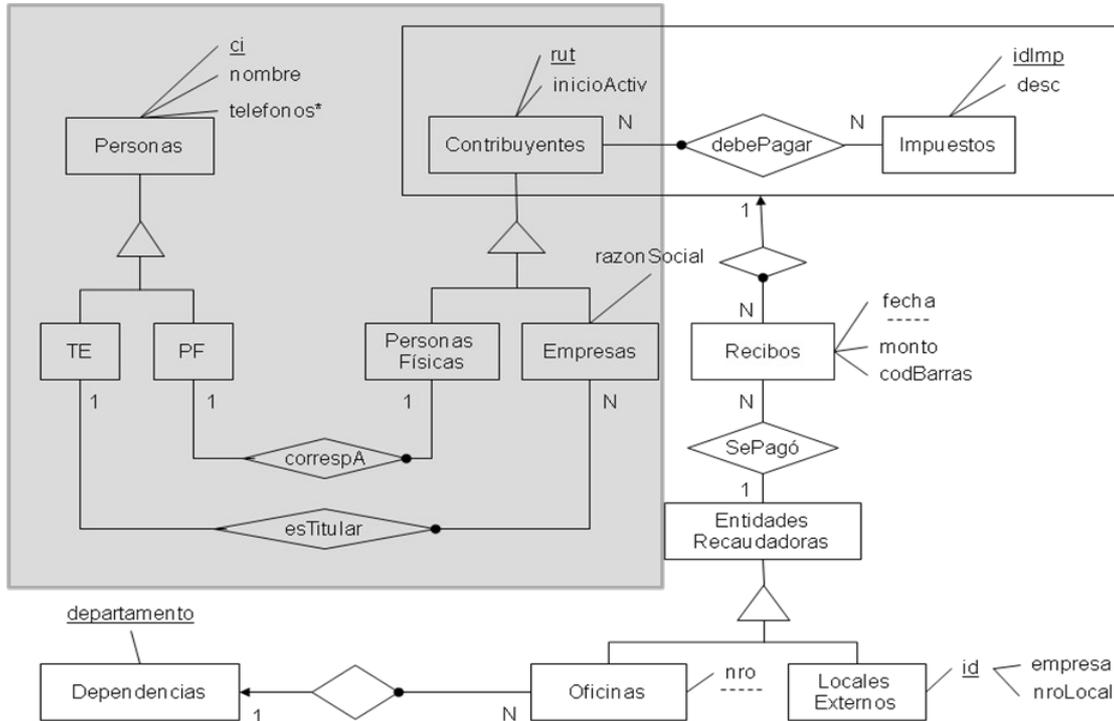
Hay dos tipos de personas que están relacionadas con los eventos y/o los artículos: autores e invitados especiales. Tanto de los autores como de los invitados especiales se conoce la CI, el número de pasaporte, el nombre, la profesión y todos sus e-mails. Tanto la CI como el pasaporte sirven para identificar 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.

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. $\forall p \in \text{Personas}. (\exists pf \in \text{PersonasFísicas} . \langle p, pf \rangle \in \text{correspA}) \rightarrow (\neg \exists e \in \text{Empresas} . \langle p, e \rangle \in \text{esTitular})$

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

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

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

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

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.

Ejercicio 3 (25 puntos).

El siguiente esquema de base de datos representa parte de la información de los distintos grupos de carnaval, los escenarios donde actúan y las presentaciones que se realizan.

ESCENARIOS (codigo, nombre, barrio, direccion, plazacomida, calidad)

Contiene información sobre los diferentes escenarios que se encuentran activos.

De cada uno de ellos se conoce un código, su nombre, el barrio donde se encuentra, la dirección exacta, un indicador de si tiene plaza de comida (valores posibles Si,No) y un indicador de la calidad del mismo con respecto al sonido. Los valores posibles de la calidad son un número entre 1 y 5.

GRUPOS (nombre, categoria, director, cantidadintegrantes)

Contiene la información de los diferentes grupos que participan del carnaval.

De cada uno de los grupos se conoce un nombre, la categoría a que corresponde, el nombre del director y la cantidad de integrantes.

PREMIOS (categoria, puesto, nombre)

Contiene el nombre de los grupos que ocuparon los tres primeros puestos de cada categoría en el concurso anterior. Como dato adicional se sabe que todos estos grupos participan actualmente del carnaval.

PROGRAMACIÓN (codigoescenario, fecha, lugar, nombregrupo, completa)

Contiene la información sobre la programación de todos los escenarios, donde se realizan las distintas presentaciones de los grupos. Para cada escenario, en cada fecha y en cada lugar (orden de aparición en esa fecha) se conoce el grupo que actúa y si la presentación es completa o no.

En este esquema no existen tablas vacías y se cumplen las siguientes relaciones de inclusión:

- $\prod_{\text{nombre,categoria}}(\text{PREMIOS}) \subseteq \prod_{\text{nombre,categoria}}(\text{GRUPOS})$
- $\prod_{\text{codigoescenario}}(\text{PROGRAMACIÓN}) \subseteq \prod_{\text{codigo}}(\text{ESCENARIOS})$
- $\prod_{\text{nombregrupo}}(\text{PROGRAMACIÓN}) \subseteq \prod_{\text{nombre}}(\text{GRUPOS})$

1) Resolver en álgebra relacional las siguientes consultas:

- Obtener el nombre y la dirección de aquellos escenarios que, en alguna fecha, tienen programada la presentación de los primeros premios.
- Obtener el nombre de los grupos que solo realizan presentaciones con grupos de su misma categoría.

2) Resolver en cálculo relacional las siguientes consultas:

- Obtener el nombre y la dirección de los escenarios en los cuales todas las presentaciones en primer lugar corresponden a grupos con más de 15 integrantes y que obtuvieron algún premio en el concurso anterior.
- Obtener el nombre y la categoría de los grupos, premiados en el concurso anterior, que sólo realizan presentaciones en escenarios de un único barrio.

3) 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.
- 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.

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.

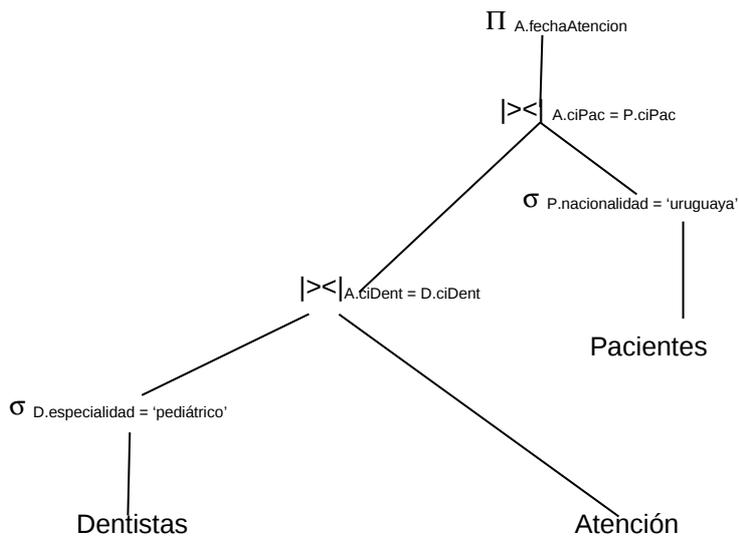
Sea la siguiente consulta SQL:

```
SELECT A.fechaAtencion
FROM Pacientes P, Dentistas D, Atencion A
WHERE A.ciDent = D.ciDent and A.ciPac = P.ciPac and
      P.nacionalidad = "uruguaya" 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). Aclaración: No se puede usar el índice por prefijos.
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.

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

1

--

Implementaciones de los Operadores.

Oper.	Algoritmo	Costo	Cond.	Organización
$\sigma_c(R)$	Índice secundario B+	$x + s$ peor caso	Cualquier Caso	Cualquiera
	Grabación Intermedia	s/bf_R	Cualquier caso	Cualquiera
$R \bowtie X \bowtie S$	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

1

Donde Z depende del tipo de índice:

- primario: $Z = x + 1$, hash = h

- secundario: $Z = x + s_S$, cluster: $Z = x + \lceil s_S/bf_S \rceil$