

FUNDAMENTOS DE BASES DE DATOS

Examen Julio 2004

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.

Parte 1. Modelo Entidad-Relación (25 puntos)

Ejercicio 1. (25 puntos)

El ministerio de Agricultura y Actividades productivas de Ruritania quiere plantear nuevas políticas de exportación en función de la PRODUCTIVIDAD de los productores. Para ello necesita llevar un registro con la siguiente información:

Se llamarán productores a las personas o empresas que deseen registrarse para adquirir los beneficios previstos en determinado plan.

Al momento de inscribirse, se le asignará un número identificatorio que deberá presentarse para cualquier trámite posterior.

Para registrarse, las personas deberán llenar un formulario con la siguiente información: Cédula de identidad, nombre completo, dirección, teléfonos, profesión (Ing. Agrónomo, Técnico Agropecuario, productor, ingeniero, etc.)

Las empresas en cambio, deberán llenar otro formulario con la siguiente información: Ruc, Nombre, dirección, teléfonos

Cada productor deberá registrar además cada uno de los establecimientos en donde cultiva, indicando su nro. de padrón, departamento, sección judicial, nombre del establecimiento y superficie. (Se asume que nunca hay dos padrones con el mismo número en la misma sección judicial del mismo departamento). Además interesa conocer en calidad de qué cultiva el productor en cada establecimiento: Propietario o por Arriendo. En caso de ser propietario, interesa conocer desde que año lo es. Y si es por arrendamiento, se desea conocer quien es el propietario y desde que fecha se está arrendando, en qué fecha termina el arrendamiento y por qué monto mensual.

De cada producto se lleva registro de la producción que tuvo en un año dado. Para esto, de cada producto que produjo cada productor, se registra el nombre del cultivo (soja, trigo, duraznos, lechugas, etc), el total (Kg.) acumulado para el año, el total de hectareas cultivadas en ese año. Estos datos son por año y para cada cultivo de cada productor (no por establecimiento).

Además se desea registrar el aporte de cada establecimiento donde cultiva cada productor, en la producción total del mismo, registrando cuantos kilos aportó cada establecimiento de cada productor para la producción de ese año y cultivo.

También hay un registro de exportadores de los que se conoce el nombre, el ruc, un número de permiso de exportación, su dirección y su teléfono.

Cada exportador, exporta parte de la producción de determinado cultivo de determinado productor. Sin embargo esta exportación, no necesariamente se realiza en el año de producción. Por este motivo, junto con el país de destino de la exportación es necesario registrar el año en que se realiza y también la cantidad exportada medida en Kg.

El ministerio desea controlar la existencia de "producción clandestina", es decir, kilos declarados de algún cultivo y que no surgen de ningún establecimiento.

También desea controlar el "Arrendamiento ilegal". Se denomina así al caso en que un productor se declara como arrendatario de alguno de los establecimientos de los cuales es propietario, con la finalidad de evadir impuestos.

SE PIDE: Modelo Entidad Relación completo, incluyendo restricciones de integridad.

Parte 2. Diseño Relacional (25 puntos)

Ejercicio 2 (13 puntos)

Se desea construir una base de datos que permita almacenar información relativa a un congreso de informática.

De cualquier participante, interesa registrar su nombre (PA_NOM), su nacionalidad (PA_NAC), la universidad de la que proviene (PA_UNIV) y su e-mail (PA_EMAIL). No hay dos participantes que tengan un mismo e-mail.

Los artículos tienen un identificador (ART_ID), un título (ART_TITULO), un conjunto de autores (ART_AUTOR) los cuales no necesariamente participan del congreso y el idioma en que está escrito (ART_IDIOMA). La organización no acepta artículos diferentes que tengan igual nombre y un autor en común. Interesa registrar qué artículos están vinculados temáticamente con qué otros.

El congreso se organiza en diferentes sesiones. Cada sesión tiene un tema (SES_TEMA) que es único entre todas las sesiones del congreso. Además se realizan en una fecha dada (SES_FECHA), en un determinado salón (SES_SALON) y en determinado turno (SES_TURNO) que puede ser mañana o tarde. Para cada sesión interesa registrar los artículos que se presentarán en ella. Un artículo se presentará en una única sesión y es presentado por un único participante. Interesa registrar a qué hora (HORA_PRES) dentro de esa sesión se presentará el artículo. Obviamente, en cada salón y en determinada fecha y hora de presentación puede haber solamente una presentación de un único artículo y por otro lado, dada la hora de presentación es posible saber el turno.

Se Pide:

1. Determinar las dependencias funcionales que se cumplen.
2. Alguien diseñó el siguiente esquema relacional. Asociar a cada tabla todas las dependencias funcionales y multivaluadas no triviales que se cumplen.

PARTICIPANTES(PA_EMAIL, PA_NOM, PA_NAC, PA_UNIV, ART_IDIOMA)

SESIONES(SES_TEMA, SES_FECHA, SES_TURNO, SES_SALON)

ARTICULOS(ART_ID, ART_TITULO, ART_AUTOR)

PROGRAMA(SES_TEMA, HORA_PRES, ART_ID, ART_TITULO, PA_EMAIL)

EXPOSITORES(PA_EMAIL, ART_AUTOR, ART_ID)

3. Indicar en qué forma normal está el esquema propuesto.
4. Indicar si el esquema preserva dependencias funcionales indicando en caso que se pierdan dependencias, dar al menos una dependencia funcional que se pierda.

Justifique todas sus respuestas.

Ejercicio 3 (12 puntos)

Dado el esquema $R(A,B,C,D,E,G,H)$ y el siguiente conjunto de dependencias funcionales:
 $F = \{C \rightarrow D, A \rightarrow C, AD \rightarrow GBH, B \rightarrow A, BD \rightarrow H\}$

1. Hallar todas las claves de F , justificando porqué son claves y porqué son todas.
2. Encontrar una descomposición del esquema R en 3NF con JSP y que preserve dependencias funcionales.
3. Encontrar una descomposición del esquema R en 4NF con JSP. En caso de perderse dependencias indicar cuales se pierden.

Parte 3. Consultas (25 puntos)

Ejercicio 4 (25 pts)

Una empresa que se dedica a ofrecer servicios para cumpleaños infantiles maneja su información en una base de datos con las siguientes tablas:

LOCALES (codloc, capacidad, dir, exterior?)

En esta tabla se almacena la información relativa a cada local que dispone la empresa para la realización de las fiestas.

De cada uno de ellos se conoce un código (codloc) que lo identifica, su capacidad (capacidad), la dirección (dir) y un campo booleano que indica si el local cuenta con espacio exterior (exterior?)

ANIMACIONES (codgrupo, tipo, fecha_inicio, cod_responsable)

En esta tabla se almacena la información relativa a cada uno de los grupos que realizan animaciones en las fiestas.

De cada grupo se conoce un código que lo identifica (codgrupo), el tipo de animación que realizan (tipo) pudiendo ser magia, titeres, payasos, etc. También se conoce la fecha en que el grupo comenzó a trabajar con la empresa (fecha_inicio) y un código de una persona responsable del grupo.

INFLABLES (codinf, superbase, altura, contact, fechaincorp)

En esta tabla se almacena la información relativa a cada inflable que tiene la empresa.

De cada uno de ellos se conoce un código que lo identifica (codinf), la superficie de la base (superbase), la altura (altura) del mismo, la cantidad de actividades que es posible realizar (ej 2 que corresponde a pelotero y tobogan) (contact) y la fecha en que este fue incorporado a la empresa (fechaincorp).

CLIENTES (codcli, nombre, mes_nac, año_nac, responsable, telefono)

En esta tabla se almacena la información relativa a cada uno de los clientes que alguna vez contrataron por lo menos unos de los servicios que ofrece la empresa para alguna fiesta.

De cada cliente se conoce un código que lo identifica (codcli), mes y año de nacimiento (mes_nac y año_nac respectivamente), datos del adulto responsable (responsable) y teléfono donde ubicarlo (telefono).

FIESTAS (codfiesta, codcli, fecha)

En esta tabla se almacena la información relativa a cada fiesta para la cual la empresa brindó por lo menos un servicio. De cada fiesta se conoce un código que la identifica (codfiesta), el cliente que la realiza (codcli) y la fecha de realización (fecha)

Es importante notar que para cada fiesta no necesariamente los clientes contratan todos los servicios, ni existen alguno obligatorio. Por ejemplo un cliente para una fiesta puede contratar local y nada más; local y cualquier cantidad de animaciones; solo inflables, cualquier combinación con la única restricción de que en caso de contratar local para la fiesta se contrata un único local.

La información relativa a los servicios que se contrataron para cada fiesta se encuentra en las siguientes tablas.

FIESTA-LOCAL (codfiesta, codloc)

Para las fiestas que se contrata local se conoce el código del mismo.

FIESTA-ANIMACION (codfiesta, codgrupo)

Para las fiestas que se contrata animaciones se conocen los códigos de los grupos contratados.

FIESTA-INFLABLE (codfiesta, codinf)

Para las fiestas que se contrata inflables se conocen los códigos de los mismos.

NOTAS:

* No existen tablas vacías.

$$* \Pi_{\text{codfiesta}}(\text{FIESTA-LOCAL}) \subseteq \Pi_{\text{codfiesta}}(\text{FIESTAS})$$

$$* \Pi_{\text{codfiesta}}(\text{FIESTA-ANIMACION}) \subseteq \Pi_{\text{codfiesta}}(\text{FIESTAS})$$

$$* \Pi_{\text{codfiesta}}(\text{FIESTA-INFLABLE}) \subseteq \Pi_{\text{codfiesta}}(\text{FIESTAS})$$

$$* \Pi_{\text{codloc}}(\text{FIESTA-LOCAL}) \subseteq \Pi_{\text{codloc}}(\text{LOCALES})$$

$$* \Pi_{\text{codgrupo}}(\text{FIESTA-ANIMACION}) \subseteq \Pi_{\text{codgrupo}}(\text{ANIMACIONES})$$

$$* \Pi_{\text{codlinf}}(\text{FIESTA-INFLABLE}) \subseteq \Pi_{\text{codinf}}(\text{INFLABLES})$$

$$* \Pi_{\text{codcli}}(\text{FIESTAS}) = \Pi_{\text{codcli}}(\text{CLIENTES})$$

SE PIDE:

a) Resolver las siguientes consultas en Álgebra Relacional:

1. Obtener el código de los inflables que solo han sido contratados para fiestas realizadas en los locales de la empresa.
2. Obtener los códigos de los clientes que han contratado todos los tipos de animación que ofrece la empresa.

b) Resolver las siguientes consultas en Cálculo Relacional:

3. Obtener el nombre de los clientes tales que todas las fiestas que han realizado solo contrataron local.
4. Obtener el nombre de los clientes que han contratado locales y siempre contrataron locales que tienen espacio exterior.

c) Resolver las siguientes consultas en SQL

5. Obtener el código y nombre de los clientes y la cantidad de fiestas para las cuales contrataron por lo menos dos tipos de animaciones distintas.
6. Para cada inflable obtener la cantidad de fiestas no realizadas en locales de la empresa para las cuales han sido contratados.

Parte 4 Optimización y Concurrencia (25 puntos)

Ejercicio 5. (15 pts.)

Dado el siguiente esquema y la siguiente consulta:

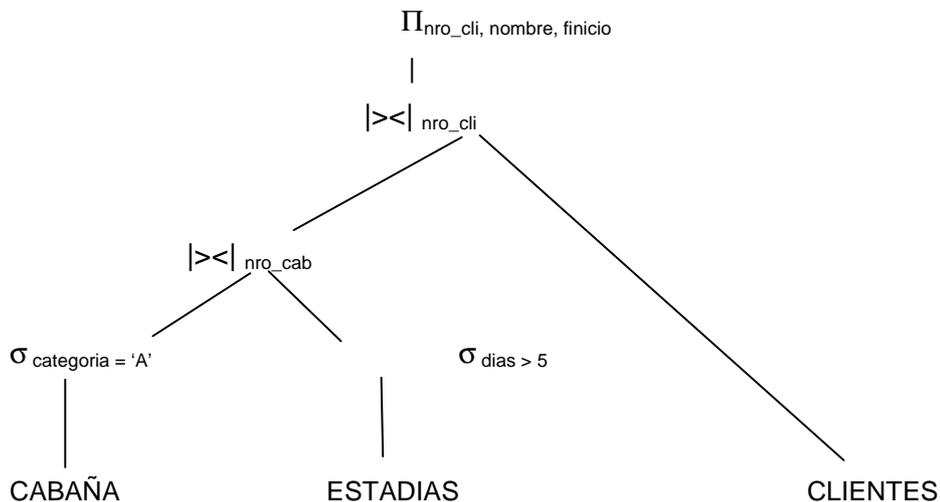
CABAÑA (nro_cab, capacidad, zona, categoría)

CLIENTES (nro_cli, nombre, fecontacto, origen)

ESTADIAS (nro_cab, nro_cli, finicio, dias)

```
SELECT C.nro_cli, C.nombre, E.finicio
FROM Clientes C, Estadias E, Cabaña A
WHERE C.nro_cli = E.nro_cli AND E.nro_cab = A.nro_cab
AND A.categoría = 'A' AND E.dias > 5
```

Considerando el siguiente plan lógico para la consulta, con los tamaños intermedios calculados:



Tamaños:

Llamamos:

t1 al resultado de aplicar $\sigma_{categoría = 'A'}$

t2 al resultado de aplicar $\sigma_{dias > 5}$

t3 al resultado de aplicar $\><|_{nro_cab}$

t4 al resultado de aplicar $\><|_{nro_cli}$

T(CABAÑA) = 15

T(ESTADIAS) = 1300

T(CLIENTES) = 600

T(t1) = 5

T(t2) = 975

T(t3) = 325

T(t4) = 325

Se pide :

Dar un plan físico adecuado para la consulta y calcular su costo total.

Nota: Considerar costo de escritura en disco para todas las operaciones, excepto la última.

DATOS:

	CLIENTES	ESTADIAS	CABANA	CABANA >< ESTADIA
Cantidad tuplas por bloque (fbl)	20	30	20	25
Indices primarios	nro-CLI (niveles: 3)			
Indices secundarios (B+)		dias (niveles: 2)		

FORMULAS:

Selección (R)	Búsqueda lineal	b
	Búsqueda binaria	$\log_2 b + \lceil s/fbl_R \rceil - 1$
	Índice primario	$x + 1$
	Índice secundario (B+)	$x + s$
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	$b_R + R * (x + s)$ $s = \text{card. de la selección en S por el atributo del join.}$
	Nested Loop (ciclo anidado) utilizando índice primario para recuperar tuplas que matchean	$b_R + R * (x + 1)$
Escribir resultados en disco	Todas la operaciones	n_{res} / fbl_{res}

Notación: b – cantidad de bloques
 fbl – factor de bloqueo
 x – cantidad de niveles del índice
 $|R|$ - cantidad de tuplas de R
 s – cardinalidad de la selección (nro. de tuplas del resultado)
 n_{res} – nro. de tuplas del resultado

Aclaración: Se considera que existe 1 buffer por relación y 1 buffer para el resultado (3 buffers).

Ejercicio 6. (10 pts.)

Dadas las siguientes transacciones T_1, T_2 :

$T_1: r_1(X), r_1(Y), w_1(Y), w_1(X), c_1$

$T_2: r_2(X), r_2(Y), w_2(X), w_2(Z), c_2$

a) Para cada historia decir si es serializable y si es recuperable, justificando.

- i) $r_1(X), r_1(Y), r_2(X), w_1(Y), w_1(X), r_2(Y), w_2(X), w_2(Z), c_1, c_2$
- ii) $r_2(X), r_2(Y), w_2(X), r_1(X), r_1(Y), w_1(Y), w_1(X), c_1, w_2(Z), c_2$
- iii) $r_1(X), r_1(Y), r_2(X), r_2(Y), w_1(Y), w_1(X), w_2(X), w_2(Z), c_2, c_1$

b) Indique como haría los bloqueos y desbloqueos en las transacciones para garantizar seriability y a la vez permitir que se entrelacen. Justifique. (Además, debe escribir las transacciones con los cambios propuestos).