

FUNDAMENTOS DE BASES DE DATOS

Examen Julio 2012 - SOLUCION

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).

El comité olímpico internacional desea registrar todos los datos correspondientes a la organización de los juegos olímpicos. Para ésto, se mantiene información de todas las personas que participan en los mismos, por lo que se conoce de ellos nombre, fecha y país de nacimiento, país donde se emitió el pasaporte y número de pasaporte. Los dos últimos datos son necesarios para identificar a cada participante. Algunos de los participantes son médicos, entrenadores o deportistas.

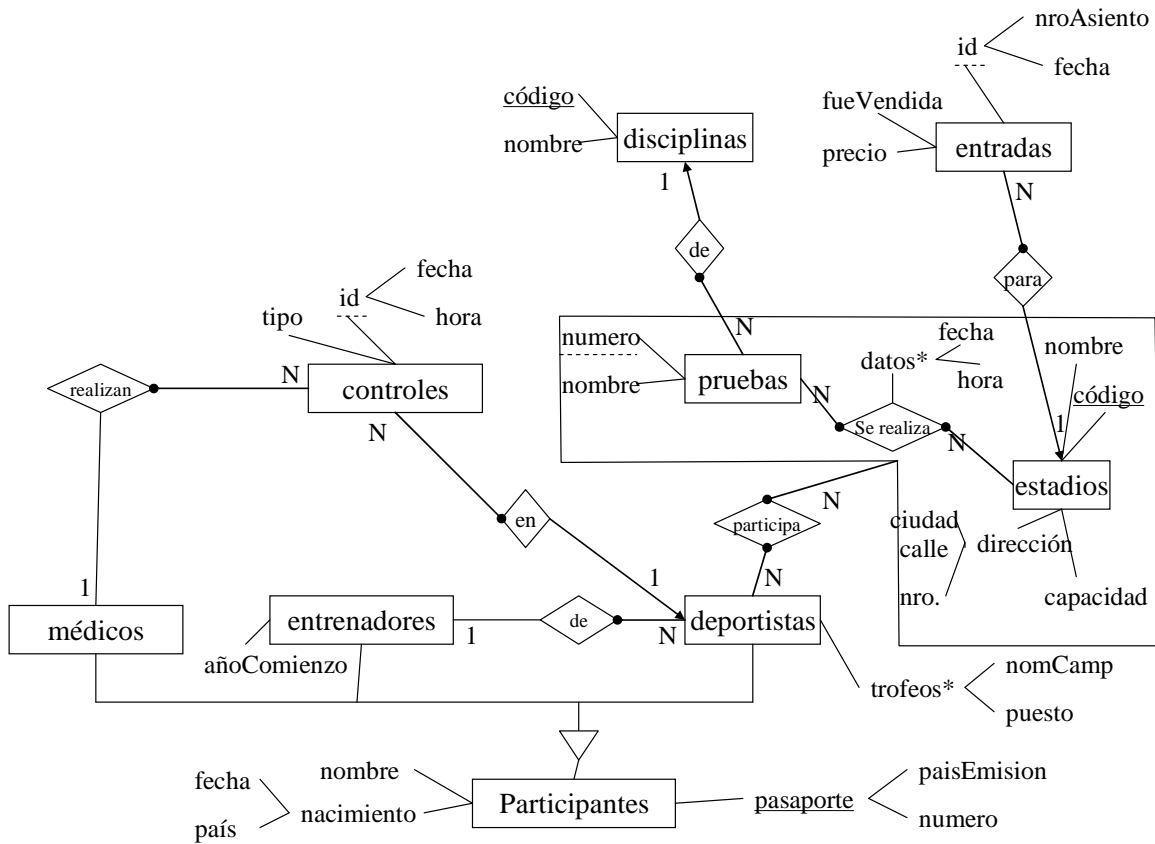
Cada deportista es entrenado por un único entrenador y se sabe que ningún deportista se entrena a sí mismo. Además, los entrenadores están dedicados únicamente a entrenar y están habilitados a entrenar a varios deportistas. De los entrenadores interesa saber el año en el cual comenzaron a ejercer como tal, mientras que de los deportistas interesa tener el registro de todos los trofeos que han ganado, de cada uno de los trofeos interesa el nombre del campeonato en el cual fue obtenido (por ejemplo, juegos olímpicos 2008, campeonato nacional del 2010, etc.) y el puesto obtenido (por ejemplo primer, segundo o tercer puesto). Los médicos realizan controles antidoping a los deportistas. De cada control antidoping se conoce el deportista al que se le realiza el análisis, la fecha y la hora de la toma, además del tipo de toma (si es sangre u orina). Cada control antidoping es realizado únicamente por un médico, pero cada médico realiza varios controles. A todos los deportistas se les realizan al menos un control antidoping y se sabe que un médico no puede realizar más de un control a un mismo deportista.

De cada disciplina que realiza un deportista se conoce un código que la identifica y su nombre, pero además se conocen las distintas pruebas de cada una de ellas. Por ejemplo, las pruebas de 100 metros llanos masculino y 100 metros con vallas femenino corresponden a la disciplina atletismo, mientras que las pruebas de gimnasia artística y las de gimnasia rítmica corresponden a la disciplina gimnasia. De cada prueba se conoce el número que la identifica dentro de cada disciplina y su nombre. Interesa modelar en qué pruebas participa cada deportista. Cada deportista no solo puede realizar varias pruebas de una misma disciplina, sino que también puede realizar varias disciplinas. Todos los deportistas que se registran en las olimpiadas realizan al menos una disciplina y todas las disciplinas registradas son realizadas en el evento.

Cada prueba se lleva a cabo en un único estadios olímpicos identificados por un código, además se conoce su nombre, dirección (ciudad, calle y número) y capacidad. De cada entrada interesa saber si fue vendida, su precio, el número de asiento y fecha. Cada prueba se realiza en una fecha y hora dada en el estadio designado, pero hay que tener en cuenta que cada prueba se puede realizar varias veces en cada estadio en diferentes horas y fechas (clasificaciones, mal tiempo, etc). Para los deportistas que compiten en cierta prueba, se registra el resultado (posición en el ranking) final del deportista en cada participación e interesa saber en qué estadios participó en cada prueba.

Se pide: Modelo Entidad Relación completo.

SOLUCION:



RNE:

1. $deportistas \cap entrenadores = \{\}$
2. $médicos \cap entrenadores = \{\}$
3. $médicos \cap deportistas = \{\}$
4. Un médico no realiza más de un control antidoping de un mismo deportista.
 $(\forall m \in médicos)(\forall d_1 \in deportistas)(realizan(m, d_1) \rightarrow \neg(\exists d_2 \in deportistas)(realizan(m, d_2) \wedge d_1 \neq d_2))$

Ejercicio 2 (25 puntos).

Un ente público maneja datos del mantenimiento y utilización de los edificios públicos. En cada edificio hay espacios físicos ocupados por oficinas de diferentes organismos. Para ello tenían una base con una única tabla:

EDIFICIOS (sec, nomEdif, calle, nroPuerta, piso, nroOf, jefeM, telM)

La descripción de los atributos es la siguiente:

sec: Es el nombre de la sección de una oficina pública que es subdivisión de un organismo. Por ej: Gerencia de la Asesoría Financiera de UTE que ocupa una oficina física. Una misma sección puede ocupar diferentes oficinas físicas en diferentes edificios (o en el mismo) pero cada oficina física es exclusiva de una sección.

nomEdif: Es el nombre del edificio.

calle, nroPuerta: Es la calle y el número. de puerta del edificio en que se encuentra la oficina. Se asume que no hay dos edificios con el mismo nombre en direcciones diferentes y que en cada dirección puede haber sólo un edificio.

piso: es el piso en donde se encuentra la oficina.

nroOf: es el número de la oficina dentro del edificio. Las primeras cifras indican el piso en que está la oficina.

jefeM: Es el nombre del jefe de mantenimiento del edificio. Hay un solo jefe de mantenimiento por edificio.

telM: Es el teléfono para llamar por problemas de mantenimiento. Puede haber varios teléfonos en un mismo edificio, pero siempre son comunes a todas las oficinas del edificio, es decir que cualquier persona puede llamar a cualquiera de los teléfonos asociados a su oficina por un problema determinado (Ej: no funciona el enchufe X). Puede suceder que los teléfonos sean de servicios de terceros por lo que puede ocurrir que haya distintos edificios con el mismo teléfono de mantenimiento asociado.

- Determine las dependencias funcionales y al menos una multivaluada embebida no trivial (y no funcional) que se cumpla sobre esta tabla.
- En una reestructura de la base construyeron las siguientes tablas:

A(Sec,nomEdif,telM), B(sec,nomEdif,nroOf,jefeM), C(sec,calle,nroPuerta,piso,nroOf)

Indique en qué forma normal está este nuevo esquema.

- Considera la tabla D que se construye en función de la siguiente consulta:

```
select sec,nomEdif,calle,nroPuerta,piso,nroOf,jefeM  
from Edificios;
```

Indique si el esquema relacional formado por las relaciones B y C de la parte b) tiene join sin pérdida o no con respecto a la relación D. Justifique su respuesta.

- Encuentre una descomposición en 3NF con JSP y sin pérdida de dependencias funcionales del esquema original y de acuerdo a su respuesta en la parte A. Aplique el algoritmo correspondiente visto en el curso.

SOLUCION:

- El esquema es:

EDIFICIOS(sec, nomEdif, calle,nroPuerta,piso,nroOf, jefeM,telM)

Las dependencias funcionales son las siguientes:

nroOf → piso, dado que las primeras cifras del nro de oficina son las del piso.

nomEdif, nroOf → sec, dado que una oficina nunca es compartida por varias secciones.

Es necesario incluir el nombre del edificio, porque una sección puede estar ocupando diferentes oficinas en diferentes edificios.

nomEdif → calle,nroPuerta, jefeM, porque el jefe de mantenimiento es único para todo el edificio y la

dirección también es única para todo el edificio.

calle,nroPuerta→nomEdif, dado que una dirección sólo puede haber un edificio.

Multivaluadas embebidas que se cumple en esta tabla son, al menos:

NomEdif -->> telM | nroOf , dado que, para todas las oficinas de un mismo edificio, se comparten todos los teléfonos.

NomEdif-->> telM | Sec, dado que, todas las secciones que estén en un mismo edificio también comparten todos los teléfonos.

- b) El primer paso para determinar en qué forma normal está cada tabla es determinar que dependencias funcionales se cumplen sobre cada una de ellas. Por eso es necesario realizar las proyecciones del conjunto de dependencias:

$$\Pi_A(F) = \{ \}$$

$$\Pi_B(F) = \{ \text{nomEdif, nroOf} \rightarrow \text{sec}; \text{nomEdif} \rightarrow \text{jefeM} \}$$

$\Pi_C(F) = \{ \text{nroOf} \rightarrow \text{piso}; \text{calle, nroPuerta, nroOf} \rightarrow \text{sec} \}$ Observar que en este caso, la última dependencia surge de una pseudotransitividad entre **calle,nroPuerta→nomEdif** y **nomEdif, nroOf→sec**.

De esta forma:

- A está al menos en BCNF dado que no hay dependencias funcionales sobre ella.
- B está en 1NF dado que la clave de esa tabla debiera ser **nomEdif** y **nroOf** (porque es la única forma de determinar a **sec**) y **nomEdif→jefeM** induciría una dependencia parcial desde la clave, por lo que viola 2NF.
- C está en 1NF porque la clave debiera ser **calle, nroPuerta,nroOf** (dado que es la forma de determinar **sec**) y **nroOf→piso**, induciría una dependencia parcial.

- c) Aplicando el teorema del JSP para deps. funcionales hay ver si se cumple una de las siguientes dependencias:

$$C \cap B \rightarrow B - C \text{ o } C \cap B \rightarrow C - B$$

Entonces:

$$C \cap B = \{ \text{sec, nroOf} \} .$$

La clausura de estos atributos con respecto a las dependencias funcionales originales es la siguiente:

$$\{ \text{sec, nroOf} \}^+ = \{ \text{sec, nroOf, piso} \}$$

y

$$C - B = \{ \text{calle, nroPuerta, nroOf} \} \text{ y } B - C = \{ \text{nomEdif, jefeM} \} .$$

Como ninguno de estos conjuntos están incluidos en $\{ \text{sec, nroOf} \}^+$ se puede afirmar que no se cumplen ninguna de las dos dependencias buscadas, por lo que se puede afirmar que la descomposición de D en B y C no tiene JSP.

- d) El conjunto de dependencias es el siguiente:

nroOf → piso
nomEdif, nroOf → sec
nomEdif → calle, nroPuerta, jefeM
calle, nroPuerta → nomEdif

El primer paso indica encontrar un cubrimiento minimal de este conjunto de dependencias:

PASO 1) Separar dependencias. De esta forma, el conjunto de dependencias nos queda:

nroOf → piso
nomEdif, nroOf → sec
nomEdif → calle
nomEdif → nroPuerta
nomEdif → jefeM
calle, nroPuerta → nomEdif

PASO 2) Eliminar atributos redundantes en el lado izquierdo de las dependencias.

Las únicas dependencias a tratar son **nomEdif, nroOf → sec** y **calle, nroPuerta → nomEdif**.

$\text{nomEdif}^+ = \{ \text{nomEdif, calle, nroPuerta, jefeM} \}$

$\text{nroOf}^+ = \{ \text{nroOf, piso} \}$

$\text{calle}^+ = \{ \text{calle} \}$

$\text{nroPuerta}^+ = \{ \text{nroPuerta} \}$

Por lo que no hay atributos redundantes en los lados izquierdos de las dependencias.

PASO 3) Eliminar dependencias redundantes.

Dado que no hay ningún atributo que aparezca del lado derecho en más de una dependencia, no puede haber dependencias redundantes.

Conclusión:

nroOf → piso
nomEdif, nroOf → sec
nomEdif → calle
nomEdif → nroPuerta
nomEdif → jefeM
calle, nroPuerta → nomEdif

Es un cubrimiento minimal.

De esta forma, las tablas iniciales que hay que construir corresponden a una tabla por cada lado izquierdo de las dependencias con todos los atributos que aparecen en cada lado derecho.

De esta forma quedan:

nroOf → piso
nomEdif, nroOf → sec
nomEdif → calle
nomEdif → nroPuerta
nomEdif → jefeM
calle, nroPuerta → nomEdif

A(nroOf,piso), B(nomEdif,nroOf,sec), C(nomEdif,calle,nroPuerta,jefeM)

La tabla que surge de la última dependencia no se agrega porque queda incluida en C.

Ahora bien, para garantizar que esto tiene JSP hay que ver si quedó incluida alguna de las claves de la tabla inicial con lo que hay que determinar todas las claves de la tabla inicial con respecto a las dependencias.

- Los atributos que nunca aparecen en las claves son los que sólo aparecen del lado derecho de las dependencias por lo tanto son:
Piso, sec y jefeM
- Los atributos que aparecen en todas las claves son los que nunca aparecen del lado derecho. Los atributos que cumplen con esta característica son:
nroOf, telM.
- Por lo anterior, los atributos candidatos a participar en las claves aunque no necesariamente en todos son:
nomEdif, calle, nroPuerta.

Observando las dependencias, se puede ver que nomEdif y la pareja calle, nroPuerta se determinan mutuamente por lo que no se deberían poner juntos en las claves y es lo mismo poner un conjunto que el otro. De esta forma, se debe probar si nomEdif, nroOf, telM es clave o si calle, nroPuerta, nroOf, telM es clave.

Para esto calculamos sus clausuras:

$\{nomEdif, nroOf, telM\}^+ = \{nomEdif, nroOf, telM, piso, calle, nroPuerta, jefeM, sec\}$
 $\{calle, nroPuerta, nroOf, telM\}^+ = \{calle, nroPuerta, nomEdif, nroOf, telM, piso, jefeM, sec\}$

Estas son todas las claves que posibles, dado que nroPuerta por sí solo ni combinado con nroOf y/o telM determinan a ningún otro atributo (sólo piso cuando se usa nroOf) y cualquier otra combinación a una de estas claves.

Dado que ninguna de estas claves quedó incluida en las tablas que se estaban manejando, hay que agregar una tabla que contenga a una de estas claves por lo que se agrega la primera:

D(nomEdif, nroOf, telM).

Así se puede garantizar que el siguiente esquema es el resultado del algoritmo y por lo tanto está en 3NF con JSP y sin perder dependencias.

A(nroOf,piso), B(nomEdif,nroOf,sec), C(nomEdif,calle,nroPuerta,jefeM)
D(nomEdif,nroOf,telM).

Ejercicio 3 (25 puntos).

Un organismo departamental se encarga del control de los espacios públicos en especial de plazas y parques. Para esto maneja una base de datos con los siguientes esquemas relación:

EspPub(codigo, nombre, juegos?, zona)

Contiene información sobre las plazas y parques. De cada uno se conoce un código que la identifica, su nombre, si tiene juegos infantiles o no y la zona del departamento donde se encuentra.

Intervenciones (codigo, fecha, nroContrato, tipo)

Contiene información sobre las intervenciones realizadas sobre cada espacio público. Las intervenciones consideradas son por ejemplo: limpieza, pintura, poda, plantación de árboles, etc. De cada una de las intervenciones se conoce el código del espacio público donde se realiza, el tipo de la intervención, la fecha de realización y el número de contrato con la empresa encargada de realización. Se sabe que cada intervención se realiza en un único día.

Empresas (nroEmp, nombre, encargado, especialidad)

Contiene la información relativa a las empresas con las que trabaja el organismo para realizar las diferentes intervenciones en los espacios públicos.

De cada empresa se conoce un nroEmp que la identifica, el nombre del encargado y su especialidad.

Contratos (nroContrato, nroEmp, descripcion, fechafirma, costo)

Contiene la información de los contratos realizados por el organismo. De cada uno de ellos se conoce número del contrato que lo identifica, el número de la empresa con la cual se realiza este contrato, una descripción de las tareas a realizar por parte de la empresa, la fecha de firma del contrato y el costo total de este trabajo.

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

- $\Pi_{\text{codigo}}(\text{INTERVENCIONES}) \subseteq \Pi_{\text{codigo}}(\text{ESPPUB})$
- $\Pi_{\text{nroContrato}}(\text{INTERVENCIONES}) \subseteq \Pi_{\text{nroContrato}}(\text{CONTRATOS})$
- $\Pi_{\text{nroEmp}}(\text{CONTRATOS}) \subseteq \Pi_{\text{nroEmp}}(\text{EMPRESAS})$

Resolver en álgebra relacional las siguientes consultas:

- a) Nombre y especialidad de las empresas que solo realizaron intervenciones en espacios públicos de una única zona.

$$A = \Pi_{\text{codigo, zona, nroEmp}}((\text{EspPub} * \text{Intervenciones}) * \text{Contratos})$$

$$B = \Pi_3(A \mid \rightarrow < \mid \rightarrow 3 = 6 \text{ and } 2 < \rightarrow 4)$$

$$\text{SOL} = \Pi_{\text{nombre, especialidad}}((\Pi_{\text{nroEmp}}(\text{Contratos}) - B) * \text{Empresas})$$

- b) Nombre de los encargados de las empresas que realizaron todas las intervenciones de tipo "PODA" en los espacios públicos que no tienen intervenciones realizadas antes del 1/1/2010.

$$A = \Pi_{\text{codigo}}(\text{EspPub}) - \Pi_{\text{codigo}}(\sigma_{\text{fecha} <= 1/1/2010}(\text{Intervenciones}))$$

Espacios públicos que no tuvieron intervenciones antes del 1/1/2010

$$B = \Pi_{\text{nroContrato}}(\sigma_{\text{tipo} = \text{PODA}}(\text{Intervenciones} * A))$$

Contratos de tipo Poda sobre los espacios públicos de interés.

$$\Pi_{\text{encargado}}(\Pi_{\text{nroEmp, nroContrato}}(\text{Contratos} \% B) * \text{Empresas})$$

Resolver en cálculo relacional las siguientes consultas:

- c) Nombre de las empresas que **en cada zona** donde realizaron intervenciones solo realizaron intervenciones de un único tipo.

```
{e.nombre / Empresas (e) ∧
  (∀c)(Contratos(c) ∧ c.nroEmp = e.nroEmp ∧
    (∃i)(∃ep) (Intervenciones(i) ∧ EspPub(ep) ∧ i.nroContrato=c.nroContrato ∧ i.codigo = ep.codigo
      →
        ¬ ( (∃c2)(∃i2)(∃ep2)(Intervenciones(i2) ∧ Contratos(c2) ∧ EspPub(ep2) ∧
          c2.nroContrato = i2.nroContrato ∧ c2.nroEmp = e.nroEmp ∧ i2.codigo = ep2.codigo ∧
          ep2.zona = ep.zona ∧ i.tipo ≠ i2.tipo
        )
      )
    )
  )
}
```

Resolver en SQL, sin utilizar vistas ni sub-consultas en el FROM, las siguientes consultas:

- d) Obtener el nombre de los espacios públicos junto con la cantidad de intervenciones realizadas en el mismo considerando únicamente los espacios que son los únicos que tienen juegos infantiles en su zona.

```
SELECT E3.nombre, count(*)
FROM Intervenciones I Natural Join EspPub E3
WHERE I.codigo IN
  (SELECT E.codigo
   FROM EspPub E
   WHERE E.juegos? and
     NOT EXISTS
       (SELECT *
        FROM EspPub E2
        WHERE E2.juegos?
          And E2.zona = E.zona
          And E2.codigo <> E.codigo
        )
  )
GROUP BY E3.codigo, E3.nombre
```

- e) Para cada espacio público donde se realizaron más de 20 intervenciones, obtener su nombre, tipo de intervención que se le realizó y el valor total invertido en ese tipo de intervención.

```
SELECT EP.nombre, I.tipo, sum(C.costos)
FROM (Intervenciones I Natural join Contrato C ) Natural join EspPub Ep
WHERE I.codigo IN
  (SELECT codigo
   FROM Intervenciones
   GROUP BY codigo
   HAVING count(*) > 20
  )
```

Ejercicio 4 (25 puntos).

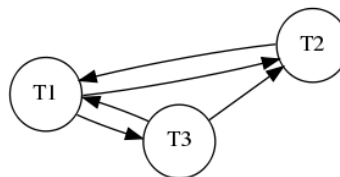
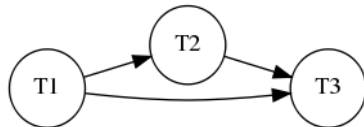
Considere las siguientes transacciones:

T1 : r1(x), r1(y), w1(x), w1(y), c1

T2 : r2(x), r2(y), w2(y), w2(x), c2

T3 : r3(x), w3(x), c3

1. Dar dos historias entrelazadas H1 y H2 que se correspondan con los siguientes grafos de seriabilidad.



2. Indique si las historias dadas son serializables. Justifique.
3. Reescribir T1 y T3 de forma que incluyan locks de lectura/escritura, siguiendo el protocolo 2PL básico.
4. Dar una historia H3 entrelazada de las transacciones de la parte anterior.
5. Indique si H3 es serializable. Justifique.
6. Indique si H3 es recuperable. Justifique.
7. Indique si H3 evita abortos en cascada. Justifique.

SOLUCIÓN

1.

H1: r1(x), r1(y), w1(x), r2(x), w1(y), r2(y), w2(y), w2(x), r3(x), w3(x), c1, c2, c3.

H2: r1(x), r1(y), r3(x), w3(x), c3, r2(x), r2(y), w2(y), w2(x), c2, w1(x), w1(y), c1.

2.

H1 es serializable dado que su grafo de seriabilidad es acíclico, mientras que H2 no lo es debido a la existencia de ciclos en su grafo.

3.

T1': r1(x), r1(x), r1(y), r1(y), wl1(x), w1(x), wl1(y), u1(x), w1(y), u1(y), c1.

T3': r13(x), r3(x), wl3(x), w3(x), u3(x), c3.

4.

H3: r11(x), r1(x), r1(y), r1(y), wl1(x), w1(x), wl1(y), u1(x), r13(x), r3(x), wl3(x), w3(x), u3(x), w1(y), u1(y), c1, c3.

5.

H3 es serializable dado que las transacciones involucradas siguen el protocolo 2PL.

7.

H3 no evita abortos en cascada porque T3 lee de T1 antes del commit de T1.