
FUNDAMENTOS DE BASES DE DATOS

Examen Julio 2013

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 ()

Actualmente diferentes países están construyendo catálogos de datos públicos y accesibles a cualquiera que los necesite (datos abiertos). Estos catálogos, contienen información sobre el contenido, la generación y la utilización de los conjuntos de datos que publican diferentes organismos públicos. En este contexto, cada organismo público es responsable de realizar y mantener la publicación de uno o más archivos, conocidos también como **Datasets**. Sobre cada dataset publicado, diferentes empresas privadas y organismos construyen aplicaciones que son usadas por quien lo necesite. Los organismos, las empresas privadas así como los ciudadanos usan las aplicaciones. Observar que no hay organismos público/privados mixtos, o bien son organismos públicos o bien son empresas privadas. El sistema está destinado a ser usado en forma interna en el país, por lo que no hay más usuarios que los organismos, las empresas y los ciudadanos.

De las empresas se conoce su RUT, su nombre y dirección y de los ciudadanos su CI. Cada organismo además de su número que lo identifica y su nombre, tiene una URL para acceder a su sitio web.

De un Dataset se conoce una URL (ej: sites/default/files/2011/ 3305/calendario_2013.pdf) para acceder a ese archivo, una descripción textual de su contenido, la fecha de la última publicación, el período de actualización y el formato en que está (csv, odp etc.). La URL que tiene cada dataset, es relativa al organismo responsable de ese dataset.

Una parte del catálogo está destinada a los conceptos que se manejan en el estado. Los conceptos solamente pueden ser clases (que representan conjuntos de cosas ej: Personas, Vehículos) o propiedades (que representan relaciones binarias entre los conjuntos de cosas, Ej. PropietarioDe). De las clases, se conoce una descripción textual y un conjunto de tipos de datos posibles con los que se puede representar. De las propiedades se sabe qué clase funciona como dominio y qué clase como codominio. Una propiedad puede tener varias clases como dominio o codominio, pero debe tener al menos una. Cada concepto tiene una URL (absoluta) (ej: http://conceptos.gub.uy/Vehiculo) que lo identifica. Tenga en cuenta que una propiedad nunca es una clase y una clase nunca es una propiedad.

Cada Dataset contiene un conjunto de campos de los que se conoce un nombre y un tipo. Cada campo está asociado a un único dataset y se sabe que en un mismo dataset los nombres de los campos no se repiten.

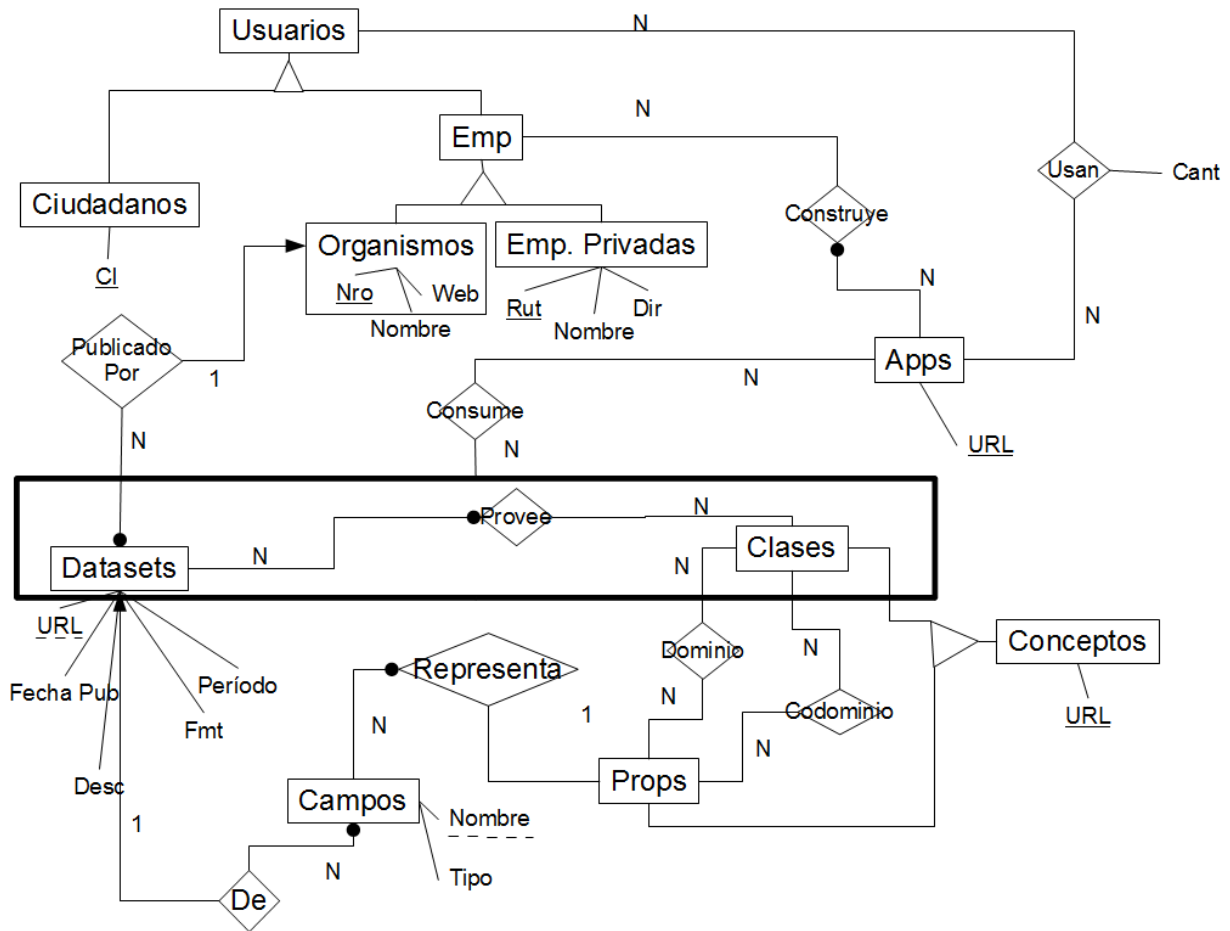
Cada campo representa una propiedad determinada y cada dataset provee una o más clases.

Tanto las empresas como los organismos construyen aplicaciones que son usadas por cualquier usuario (ya sea organismo, empresa o ciudadano). Algunas veces, las empresas u organismos se juntan entre sí para construir una misma aplicación. Las aplicaciones están identificadas por una URL. Una aplicación consume determinadas clases de determinados datasets. Se asume que no hay aplicaciones registradas que no consuman alguna clase de algún dataset. Interesa registrar para cada aplicación, la cantidad de veces que la accede cada usuario.

Observar que diferentes datasets pueden proveer una misma clase de la misma forma que diferentes una propiedad puede estar representada por varios campos. Además si un dataset tiene un campo, entonces debe proveer al menos una clase que sirve como dominio de la relación asociada a ese campo.

Se pide: Modelo Entidad Relación completo.

SOLUCION:



RNES

- La categorización de ciudadanos y empresas es total y disjunta.

$$\begin{aligned}
 \text{Usuarios} &\equiv \text{Ciudadanos} \cup \text{Emp} \\
 \text{Ciudadanos} \cap \text{Emp} &\equiv \emptyset
 \end{aligned}$$

- La categorización de Organismos y Empresas Privadas es total y disjunta.

$$\begin{aligned}
 \text{Emp} &\equiv \text{Organismos} \cup \text{Emp. Privadas} \\
 \text{Organismos} \cap \text{Emp. Privadas} &\equiv \emptyset
 \end{aligned}$$

- La categorización de Clases y Propiedades es total y disjunta.

$Conceptos \equiv Clases \cup props$

$Clases \cap props \equiv \emptyset$

• Si un dataset tiene un determinado campo, debe proveer el dominio de la propiedad asociada a ese campo.

$\forall c \in Campos. ($
 $\exists p \in Props. ((c, p) \in Representa \wedge$
 $\exists d \in Datasets. ((c, d) \in De \wedge \exists c_1. ((d, c_1) \in Provee \wedge (c_1, p) \in Dominio)))))$

Ejercicio 2 ()

Sea el esquema relación $R(A,B,C,D,E,G)$ y el conjunto de dependencias funcionales:

$$F = \{A \rightarrow BC, C \rightarrow DG, BD \rightarrow E, AB \rightarrow D, BC \rightarrow G\}$$

a) Determine si los siguientes conjuntos de dependencias funcionales son equivalentes a F , justifique su respuesta:

1. $F_1 = \{A \rightarrow B, A \rightarrow G, BG \rightarrow D, B \rightarrow C, C \rightarrow E, C \rightarrow G\}$
2. $F_2 = \{A \rightarrow BC, C \rightarrow DG, BD \rightarrow E\}$

SOLUCION

$$1. F_1 = \{A \rightarrow B, A \rightarrow G, BG \rightarrow D, B \rightarrow C, C \rightarrow E, C \rightarrow G\}$$

$$A^+_{F_1} = \{A, B, C, D, G, E\}$$

$$A^+_{F_1} = \{A, B, G, D, C, E\}$$

Coinciden

$$C^+_{F_1} = \{C, D, G\}$$

$$C^+_{F_1} = \{C, E, G\}$$

$$C \rightarrow D \in F^+ \text{ pero } C \rightarrow D \notin F_1^+.$$

Por lo tanto F_1 y F NO son equivalentes.

$$2. F_2 = \{A \rightarrow BC, C \rightarrow DG, BD \rightarrow E\}$$

$$F_2 \subset F \text{ por lo tanto } F_2^+ \subseteq F^+$$

Analizamos las dependencias de F que no pertenecen a F_2 , $AB \rightarrow D$ y $BC \rightarrow G$:

$$(AB)^+_{F_2} = \{A, B, C, D, G, E\}, \text{ por lo tanto } AB \rightarrow D \in F_2^+.$$

$$(BC)^+_{F_2} = \{B, C, D, G, E\}, \text{ por lo tanto } BC \rightarrow G \in F_2^+.$$

Por lo tanto F_2 y F son equivalentes. \checkmark

a) Dar una descomposición de R que cumpla las siguientes condiciones:

- La descomposición tiene 2 subesquemas: R1, R2.
- Cada uno de los subesquemas tiene por lo menos 2 atributos
- Ninguno de los subesquemas está incluido en el otro.
- La descomposición preserva las dependencias de F.
- La descomposición tiene join sin pérdida con respecto a F.

Justifique su respuesta.

SOLUCION

Sea la siguiente descomposición de R : R1(A,B,C,D,E), R2(A,C,D,G).
 Demostraremos que se cumplen todas las condiciones pedidas.

- $R1 \cup R2 = R \Rightarrow$ es una descomposición correcta. \checkmark
- Cada uno de los subesquemas tiene más de 2 atributos. \checkmark
 - o $R1 - R2 = \{B,E\} \Rightarrow R1 \not\subseteq R2. \checkmark$
 - o $R2 - R1 = \{G\} \Rightarrow R2 \not\subseteq R1. \checkmark$
- Consideramos las dependencias de F_2 que ya fue demostrado que es equivalente a F.
 - o $A \rightarrow BC$ se proyecta en forma directa en R1
 - o $C \rightarrow DG$ se proyecta en forma directa en R2
 - o $BD \rightarrow E$ se proyecta en forma directa en R1

Por lo tanto: $F_2 \subseteq (\Pi_{R1} F_2 \cup \Pi_{R2} F_2)$, entonces sus clausuras coinciden y la descomposición preserva las dependencias funcionales. \checkmark

- Aplicaremos el teorema visto en el curso que permite determinar si una descomposición de un esquema en dos subesquemas tiene JSP.
 - o $R1 \cap R2 = \{A,C,D\}$
 - o $R1 - R2 = \{B,E\}$ (calculado anteriormente)
 - o $A \rightarrow BCDEG \in F^+$ por lo calculado en la parte a)

Por aplicación de aumento y descomposición se puede afirmar que $ACD \rightarrow BE \in F^+$

Entonces $R1 \cap R2 \rightarrow R1 - R2 \in F^+$, aplicando el teorema antes mencionado se cumple que la descomposición tiene JSP.

b) Sea $H = F \cup \{CE \rightarrow A\}$ y la siguiente descomposición de R:
 $\rho = (S1, S2, S3)$ con : $S1=(A,D,E)$, $S2=(A,C,G,E)$ y $S3=(C,E,B,D)$

1. Calcule la proyección del conjunto de dependencias de H en cada uno de los subesquemas S1, S2 y S3.

SOLUCION

$H = \{A \rightarrow BC, C \rightarrow DG, BD \rightarrow E, CE \rightarrow A\}$

S1 (A,D,E)

$$A^+ = \{A,B,C,D,G,E\}$$

$$D^+ = \{D\}$$

$$E^+ = \{E\}$$

$$(DE)^+ = \{D,E\}$$

$$\Pi_{S1} H = \{A \rightarrow DE\}$$

S2 (A,C,G,E)

$$C^+ = \{C,D,G\}$$

$$G^+ = \{G\}$$

$$(CE)^+ = \{C,E,A,B,D,G\}$$

$$(EG)^+ = \{E,G\}$$

$$\Pi_{S2} H = \{A \rightarrow CGE, CE \rightarrow A, C \rightarrow G\}$$

S3 (C,E,B,D)

$$C^+ = \{C,D,G\}$$

$$E^+ = \{E\}$$

$$B^+ = \{B\}$$

$$D^+ = \{D\}$$

$$(CE)^+ = \{C,E,A,B,D,G\}$$

$$(CB)^+ = \{C,B,D,G,E,A\}$$

$$(EB)^+ = \{E,B\}$$

$$(ED)^+ = \{E,B\}$$

$$(BD)^+ = \{B,D,E\}$$

$$\Pi_{S3} H = \{C \rightarrow D, CE \rightarrow BD, CB \rightarrow ED, BD \rightarrow E\}$$

2. Determinar cuál es la máxima forma normal en que se encuentran cada uno de los esquemas S1, S2 y S3. Justifique la respuesta.

SOLUCION

$$S1 (A,D,E) \quad \Pi_{S1} H = \{A \rightarrow DE\}$$

A es clave de S1, por lo tanto S1 está en BCNF y como no hay dependencias multivaluadas no funcionales se puede afirmar que S1 está en 4NF.

$$S2 (A,C,G,E) \quad \Pi_{S2} H = \{A \rightarrow CGE, CE \rightarrow A, C \rightarrow G\}$$

A es clave de S2

CE es clave de S2

G no pertenece a ninguna clave ya que no pertenece a los lados izquierdos de las dfs.

En $C \rightarrow G$ parte de una clave determina un atributo no primo por lo tanto $CE \rightarrow G$ es una dependencia parcial.

S2 se encuentra en 1NF.

S3 (C,E,B,D) $\Pi_{S3} H = \{C \rightarrow D, CE \rightarrow BD, CB \rightarrow ED, BD \rightarrow E\}$

CE es clave de S3

CB es clave de S3

En $C \rightarrow D$ parte de una clave determina el atributo no primo D por lo tanto $C \rightarrow D$ es una dependencia parcial.

S3 se encuentra en 1NF.

Ejercicio 3 (25 puntos).

El siguiente esquema de base de datos representa parte de un sistema de liquidación de sueldos de un organismo del estado:

PERSONAS(ci, nombre, direccion, sexo, fechaNac, estadoCivil)

Contiene información acerca de las personas que trabajan en el organismo. De cada persona se conoce su cédula de identidad, nombre completo, dirección, sexo, fecha de nacimiento y estado civil.

CARGOS(nro, denominacion, estado)

Contiene información de los cargos del organismo. De cada cargo se conoce su número, su denominación (ejemplo: secretaria) y su estado ("vacante" u "ocupado")

PERSONAS_CARGOS(ciPersona, nroCargo, fechaDesde, fechaHasta)

Representa el vínculo entre personas y cargos. Los atributos fechaDesde y fechaHasta indican el comienzo y fin del vínculo. En esta tabla, no hay datos con fechaDesde posteriores al día de hoy y además, se conoce la fecha de fin de todos los cargos.

CONCEPTOS(id, descripción, tipoConcepto)

Contiene información acerca de los conceptos que se componen las liquidaciones. De cada concepto se conoce un identificador numérico, su descripción y el tipo de concepto ("haber" o "descuento"). Un ejemplo de concepto de tipo "haber" es "sueldo nominal", mientras que "aporte IRPF" es un tipo de concepto de tipo "descuento".

LIQUIDACIONES(ciPersona, nroCargo, idConcepto, mesLiquidacion, monto)

Contiene información acerca de las liquidaciones mensuales que se realizan en el organismo. Cada tupla en la tabla liquidación representa un renglón de la liquidación, el cual corresponde a cierto concepto para cierta persona que ocupa cierto cargo. El atributo monto indica el monto en pesos uruguayos correspondiente a ese concepto. Si el concepto es de tipo "descuento" el monto es negativo.

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

$$\Pi_{ciPersona}(PERSONAS_CARGOS) \subseteq \Pi_{ci}(PERSONAS)$$

$$\Pi_{nroCargo}(PERSONAS_CARGOS) \subseteq \Pi_{nro}(CARGOS)$$

$$\Pi_{ciPersona, nroCargo}(LIQUIDACIONES) \subseteq \Pi_{ciPersona, nroCargo}(PERSONAS_CARGOS)$$

$$\Pi_{idConcepto}(LIQUIDACIONES) \subseteq \Pi_{id}(CONCEPTOS)$$

Resolver las siguientes consultas en Álgebra Relacional:

1. Nombre de aquellas personas que durante todo el tiempo figuran en el organismo ocuparon un único cargo en el organismo.

Solución.

Personas con al menos un cargo: $A = \Pi_{ciPersona}(PERSONAS_CARGOS)$

Personas con más de un cargo: $B = \Pi_{\$1}(\text{PERSONAS_CARGOS } \bowtie_{\$1=\$5 \wedge \$2 < \$6} \text{PERSONAS_CARGOS})$

Resultado = $\Pi_{nombre}(PERSONAS * (A-B))$

2. Cédula de identidad, nombre de las personas y mes de liquidación tales que en dicho mes, en las liquidaciones de esa persona figuran todos los conceptos de tipo "haber".

Solución.

Conceptos de tipo "haber": $A = \prod_{id} (\sigma_{\text{tipoConcepto}="haber"}(\text{CONCEPTOS}))$

Resultado = $\prod_{ci, \text{nombre}, \text{mesLiquidacion}} (\text{PERSONAS} \bowtie_{ci = ciPersona} (\prod_{ciPersona, \text{mesLiquidacion}, \text{idConcepto}} (\text{LIQUIDACIONES}) \% A))$

Resolver las siguientes consultas en Cálculo Relacional:

Solución.

3. Número y denominación de aquellos cargos que están actualmente "vacante" y nunca fueron ocupados por personas de sexo femenino.

$\{ \langle c.nro, c.denominacion \rangle / \text{CARGOS}(c) \wedge c.estado="vacante" \wedge$
 $\neg((\exists pc)(\text{PERSONAS_CARGOS}(pc) \wedge pc.nroCargo=c.nroCargo \wedge$
 $(\exists p)(\text{PERSONAS}(p) \wedge pc.ciPersona=p.ci \wedge p.sexo="Femenino"))))$
 $\}$

4. Detectar inconsistencias entre el estado de un cargo y los vínculos entre personas y cargos que se registran en la tabla PERSONAS_CARGOS. Para esto, devolver los números de cargo de cargos con estado "ocupado" para los cuales no hay un vínculo activo. Un vínculo se considera activo, si hay una persona registrada que ocupa un cargo con una fechaHasta mayor o igual que el día de hoy.

Solución.

$\{ \langle c.nro \rangle / \text{CARGOS}(c) \wedge c.estado="ocupado" \wedge$
 $\neg((\exists pc)(\text{PERSONAS_CARGOS}(pc) \wedge pc.nroCargo=c.nroCargo)) \wedge \neg(\exists pc)$
 $(\text{PERSONAS_CARGOS}(pc) \wedge pc.nroCargo=c.nroCargo \wedge pc.fechaHasta < hoy)$
 $\}$

Resolver las siguientes consultas en SQL, sin utilizar vistas, ni subconsultas dentro del FROM:

5. El monto líquido en cierto mes se calcula como la suma de los conceptos. Devolver el monto líquido para aquellas personas que figuran vinculadas a más de tres cargos en la liquidación de mayo de 2013 (201305).

Solución.

```
SELECT L.ciPersona, sum(monto)
FROM LIQUIDACIONES L
WHERE L.mesLiquidacion="201305"
GROUP BY L.ciPersona
HAVING count(DISTINCT(nroCargo) > 3
```

Ejercicio 4 (25 pts).

Dadas las siguientes transacciones:

T1: w1(X), r1(Y), r1(Z), w1(Z), c1

T2: r2(X), r2(Y), w2(Y), c2

1) Si es posible, dar dos historias entrelazadas con T1 y T2 tales son equivalentes pero una evita abortos en cascada, la otra no. Justifique.

SOLUCION

La noción de equivalencia (por conflictos) observa el orden de los conflictos pero la noción de evitar abortos en cascada, sólo observa el orden de los commits y los read. Dado que tienen relativamente poco que ver una con otra, si las transacciones lo admiten, debería ser posible construir las dos historias con las condiciones que piden.

Observando las transacciones se puede ver que T1 graba en la primer instrucción el mismo ítem que lee T2 también en la primer instrucción y luego T1 lee un ítem que T2 graba en su última instrucción antes del commit. Viendo esto, es posible entrelazar las transacciones de forma que T1 lea de T2. Así se obtienen las dos historias siguientes:

H1: r2(X),w1(X),r2(Y),w2(Y),r1(Y),r1(Z),w1(Z),c1,c2

H2: r2(X),w1(X),r2(Y),w2(Y),c2,r1(Y),r1(Z),w1(Z),c1

En H1 entonces, T1 lee de T2 el ítem Y y T1 confirma antes que T2. Esta historia entonces, no es recuperable y por lo tanto, no evita abortos en cascada.

En H2 las lecturas y conflictos son exactamente los mismos que en H1 por lo tanto, H2 es una historia equivalente a H1. Sin embargo, la historia evita abortos en cascada ya que la T2 confirma antes que T1 haga una lectura de un ítem que T2 grabó.

2) Escribir T1 y T2 siguiendo el protocolo **2PL básico** (utilizando locks de lectura y de escritura). Justifique.

SOLUCION

El protocolo 2PL Básico sólo exige que se identifiquen las dos fases: crecimiento y decrecimiento de locks.

De esa forma, las siguientes transacciones son versiones de T1 y T2 pero siguiendo 2PL básico.

T1: w1(X),w1(X),rl1(Y),r1(X),rl1(Z),r1(Z),w1(Z),ul1(X),ul1(Y),w1(Z),ul1(Z),c1

T2: rl2(X),r2(X),rl2(Y),r2(Y),wl2(Y),ul2(X),w(Y),ul2(Y),c2

En estas transacciones, nunca se crece en cantidad o calidad una vez que se liberó el primer lock.

Por eso, siguen 2PL básico.

1) Dar una historia entrelazada de las T1 y T2 de la parte anterior. Decir si la historia es serializable, justificando.

SOLUCION

w1(X),w1(X),rl1(Y),r1(Y),rl1(Z),r1(Z),w1(Z),ul1(X),ul1(Y),rl2(X),r2(X),rl2(Y),r2(Y),wl2(Y),ul2(X),w(Y),ul2(Y),c2,w1(Z),ul1(Z),c1

La historia es serializable, dado que sus operaciones en conflicto siempre están en el mismo orden: primero

la de T1 y luego la de T2. De esta forma, el grafo de seriabilidad tendrá un sólo arco.
Además, cuando una historia es de transacciones que siguen 2PL, dicha historia es serializable.