

CC20A-1  
Lucía Hernández  
30-10-2008

## Auxiliar 10

### Ejemplo

Supongamos que alguien implanta esta relación

Depto

NombreD	<u>NumeroD</u>	RutGte	LugaresD

no está en 1FN (múltiples lugaresD)

Transformación a 1FN

<u>NDepto</u>	<u>Lugar</u>	RutGte	NombreD
5	Ñuñoa	555	Administración
5	Santiago	555	Administración
6	La Reina	284	Investigación
10	La Reina	425	Márketing

### Ejemplo

Factura

<u>CodigoCliente</u>	<u>CodigoArticulo</u>	Cliente	Articulo

CodigoCliente → Cliente

CodigoArticulo → Articulo

Para el ejemplo de la factura se dividirá en

Factura

<u>CodigoCliente</u>	<u>CodigoArticulo</u>
----------------------	-----------------------

Producto

<u>CodigoArticulo</u>	Articulo
-----------------------	----------

Cliente

<u>CodigoCliente</u>	Cliente
----------------------	---------

## Ejemplo

Lotes

<u>Id_propiedad</u>	Nombre_munic	Num_lote	Area	Precio	TasaFiscal
df <sub>1</sub>	↑	↑	↑	↑	↑
df <sub>2</sub> ↑			↑	↑	↑
	df <sub>3</sub>				↑
			df <sub>4</sub>	↑	

¿Está en 1FN? Sí

¿Está en 2FN? Por definición, no

Normalizamos

Lotes1

<u>Id_propiedad</u>	Nombre_munic	Num_lote	Area	Precio
df <sub>1</sub>	↑	↑	↑	↑
df <sub>2</sub> ↑			↑	↑
			df <sub>4</sub>	↑

Lotes2

<u>Nombre_munic</u>	TasaFiscal
df <sub>3</sub>	↑

Estas dos relaciones están en 2FN, sin embargo, Lotes1 tiene un problema: la  $df_4$

Esto ocurre porque precio puede depender de Id\_propiedad “a través de” Area (una transitividad).

Mejoremos Lotes1

Lotes 1a

<u>Id_propiedad</u>	Nombre_munic	Num_lote	Area
$df_1$	↑	↑	↑
$df_2$	↑	↑	↑

Lotes 1b

<u>Area</u>	Precio
$df_4$	↑

Ambas relaciones están en 3FN

## Dependencias multivaloradas

**Ejemplo:** EMP

<u>Nom_emp</u>	<u>Nom_proy</u>	<u>Nom_carga</u>
Pérez	X	Juan
Pérez	Y	Juan
Pérez	X	Ana
Pérez	Y	Ana

$Nom\_emp \twoheadrightarrow Nom\_proy$

$Nom\_emp \twoheadrightarrow Nom\_carga$

**Ejemplo:** Veamos si EMP está en 4FN

Habíamos visto que  $Nom\_emp \twoheadrightarrow Nom\_proy$  y  $Nom\_emp \twoheadrightarrow Nom\_carga$  y eran no triviales.

$Nom\_emp$  debería ser superclave de EMP, ¡pero no lo es!. Por lo tanto EMP no está en 4FN.

Para dejar EMP en 4FN, descomponemos

EMP1

<u>Nom_emp</u>	Nom_proy
----------------	----------

EMP2

<u>Nom_emp</u>	Nom_carga
----------------	-----------

Las que están en 4FN

Se puede observar que, en general, los problemas que se ocasionan por la 4FN ocurren en relaciones en las cuales todos los atributos son parte de la clave primaria.

Pregunta 1:

Un entusiasta del cine ha construido una base de datos relacional con la siguiente información de la colección de videos que posee:

Película (título, año, duración, nombreEstudio)

Actor (nombre, dirección, género, año Nac)

ActúaEn (nombreActor, título, año)

Estudio (nombreEstudio, dirección)

(los atributos de "dirección" se refieren a dirección postal)

- a) Dé una expresión SQL para mostrar el título y año de todas las películas no hechas en Hollywood.

Solución:

SELECT título, año

FROM Película

WHERE nombreEstudio IN (SELECT nombreEstudio

FROM Estudio

WHERE direccion NOT LIKE '%Hollywood%')

- b) Dé una expresión SQL para encontrar el año más antiguo en que apareció una película de cada actor (independiente del género) que haya actuado en al menos 3 películas. Los resultados deben mostrar nombre de actor y año solicitado. Por ejemplo, en el listado debe aparecer Michelle Pfeiffer 1983m porque el coleccionista tiene los videlos de The Witches if Eastwick (1987), Scarface (1983), Batman Returns (1992), The Fabulous Baker Boys (1989), y I am Sam (2001), películas todas en que ella actúa.

Solución:

```
SELECT nombreActor, MIN(año)
```

```
FROM ActúaEn
```

```
GROUP BY nombreActor
```

```
HAVING COUNT (*) >= 3
```

- c) Dé una expresión SQL para encontrar los títulos de películas en que ninguna persona que actúa allí haya nacido antes de 1985.

Solución:

```
SELECT título
```

```
FROM Película
```

```
WHERE 1985 < = ALL (SELECT añoNac
```

```
FROM ActúaEn, Actor
```

```
WHERE nombre = nombreActor
```

```
AND ActúaEn.título = Película.título)
```

Pregunta 2:

- a) Normalize esta tabla hasta 4FN:

<b>Figura</b>	<b>Color</b>	<b>Tamaño</b>
Cuadrado	Rojo	Grande
Cuadrado	Azul	Grande
Cuadrado	Azul	Mediano
Círculo	Blanco	Mediano
Círculo	Azul	Pequeño
Círculo	Azul	Mediano

Solución:

Se puede ver que esta tabla está en 1FN porque no hay valores no atómicos. Está en 2FN sin saber la llave, pero a simple vista cada atributo es llave, y no hay atributos que no dependan de otro que no sea llave. Está en 3FN porque no hay dependencias transitivas. Está en FNBC por la super llave que posee. Sin embargo, no está en 4FN.

Existen las dmv  $\text{Figura} \twoheadrightarrow \text{Color}$ ,  $\text{Figura} \twoheadrightarrow \text{Tamaño}$  (que no son triviales), pero no existen las df  $\text{Figura} \rightarrow \text{Color}$ ,  $\text{Figura} \rightarrow \text{Tamaño}$ , por lo tanto,  $\text{Figura}$  no es llave candidata. Por lo tanto hay que separar en dos tablas,  $R_1$  ( $\text{Figura}, \text{Color}$ ) y  $R_2$  ( $\text{Figura}, \text{Tamaño}$ )

Figura	Color
Cuadrado	Azul
Cuadrado	Rojo
Círculo	Azul
Círculo	Blanco

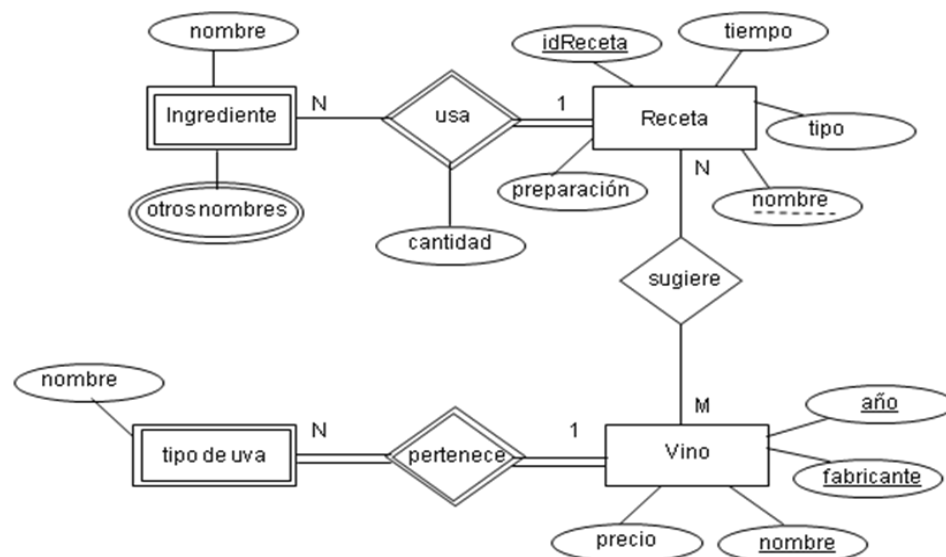
Figura	Tamaño
Cuadrado	Grande
Cuadrado	Mediano
Círculo	Pequeño
Círculo	Mediano

- b) Considere la relación  $R(A, B, C, D)$ . Suponga que las únicas dependencias funcionales que hay para  $R$  (excepto las inferibles) son:  $A \rightarrow B$ ,  $B \rightarrow C$  y  $C \rightarrow D$ . Identifique todas las llaves candidatas para  $R$ . Justifique.

Solución: Por transitividad, podemos ver que  $A \rightarrow C$  y que  $A \rightarrow D$ , por lo tanto, como  $A$  determina a todos los atributos, es llave candidata. Como  $B$  no determina a  $A$ , no es llave candidata. Análogamente  $C$  no es llave candidata, ya que sólo determina a  $D$ . Por ende la única llave candidata es  $A$ .

Pregunta 3:

Dado el siguiente diagrama E-R, construya el correspondiente esquema de base de datos, normalizado hasta 3FN.



Pregunta 4:

Sea  $R$  la relación con los atributos  $A, B, C, D, X, Y, K$ . Se verifican las dependencias:

$B \rightarrow XK$

$C \rightarrow A$

$D \rightarrow AY$

$Y \rightarrow D$

Solución:

Hay dos posibilidades\_ BCD o BDY

Pregunta 5:

Sea el esquema de relación R con los atributos A, B, C, D, E, F, G, H, I, J. Se verifican las dependencias funcionales:

$AB \rightarrow C$

$A \rightarrow DE$

$B \rightarrow F$

$D \rightarrow IJ$

$F \rightarrow GH$

¿Cuál es la clave de R?

Solución: AB

Pregunta 6:

Locovich se complicó con las llaves foráneas. De hecho, nunca las entendió muy bien. Se le ocurrió, entonces, una idea para **juntar** relaciones. Por ejemplo, este par de relaciones:

CLIENTE

<u>Rut</u>	Nombre	Teléfono	Dirección	Num-cuenta
------------	--------	----------	-----------	------------

CUENTA

<u>Num-cuenta</u>	Tipo-cuenta	Saldo	Ejecutivo-cuenta
-------------------	-------------	-------	------------------

sugiere reemplazarlas por esta única relación:

CLIENTE-CUENTA

<u>Rut</u>	Nombre	Teléfono	Dirección	Num-cuenta	Tipo-cuenta	Saldo	Ejecutivo-cuenta
------------	--------	----------	-----------	------------	-------------	-------	------------------

¿será buena idea? Justifique.

Pregunta 7:

Profesor	Curso	Texto
Silva	Est. de datos	Bartram
Silva	Gest. de datos	Al-Nour
Heras	Compiladores	Hoffman
Bravo	Est. de datos	Augenthaler
Ruiz	Gest. de datos	Augenthaler

[Tupla en azul agregada posteriormente para ejemplificar la respuesta negativa de la siguiente pregunta]

¿Podemos “deducir” que Texto  $\rightarrow$  Curso?

Queda claro que      Curso  $\nrightarrow$  Texto  
Profesor  $\nrightarrow$  Curso

Pregunta 4:

Tico y Tuco se graduaron de ingenieros hace poco y entraron a trabajar en Aceites Oil Ltda. La compañía es una productora mediana de aceites de oliva. El primer tema que les pidieron trabajar es la base de datos relacional de Aceites Oil.

- a) A Tico le explicaron que un aceite comercial se obtiene como una mezcla de las producciones de varias cepas, a fin de compensar los eventuales defectos que cada una de estas cepas pudiese tener, además de acomodar los gustos de los compradores. El gerente de producción le explica además que se usan hasta cinco cepas en cada producto terminado. Tico entonces inventa esta relación:

ACOMERCIAL (c1, p1, c2, p2, c3, p3, c4, p4, c5, p5, vol)

Donde cada c es una cepa, cada p es el porcentaje, y vol es el volumen de venta anual comprometido hasta el momento. Por ejemplo, esta tupla es una fila de la instancia de ACOMERCIAL:

< Aromatic27, 45, Rosebud3, 55, nulo, nulo, nulo, nulo, nulo, nulo, 765 >

Indicando que esta mezcla comercial se fabrica con 45% de aceite de la cepa Aromatic27 y 55% de Rosebud3; hasta el momento se han comprometido ventas por 765 mil litros.

Tuco, mirando lo que su amigo está haciendo, le dice que eso no está bien. Si Tuco tiene razón, diga por qué y cuál es la solución correcta. Si por el contrario, lo hecho por Tico está correcto (aunque haya otras soluciones) justifíquelo.



Solución:

No está bien, el ejemplo mismo muestra que hay valores nulos para partes de la clave primaria. Una solución es dar nombre a cada mezcla:

ACOM(nombre, vol)

FORMULA(nombre, c1, p1, c2, p2, c3, p3, c4, p4, c5, p5)

- b) Tuco, por su parte, está modelando la información referente a las diversas plantaciones que tiene la empresa: hay predios en diversas localidades del país (Catapilco, Rosaio, etc). En cada uno de estos predios se pueden producir varias variedades (cepas). La propuesta de Tuco es la siguiente:

PREDIO (numpredio, hectáreas, localidad)

CEPA (nombrecepa, numpredio1, hect1, ..., numpredion, hectn)

(La relación CEPA indica en qué predios se tiene plantada esa cepa y cuantas hectáreas, respectivamente)

¿Estará bien? Si lo está, dé tres tuplas ejemplo de cómo podrían ser estas relaciones. Si no lo está, diga por qué y dé un modelamiento alternativo.

Solución:

No está bien. En CEPA no puede haber un número variable de atributos. La solución propuesta es:

PREDIO(numpredio, hectareas, localidad)

CEPA\_PREDIO(nombrecepa, numepredio, hectareas)

- c) Finalmente, Tico tiene un problema con la relación VENTAS. Específicamente, no sabe cuál puede ser la clave primaria. Consulta a su amigo y Tuco le da una solución: poblar la relación y buscar aquellos atributos que no tienen valores repetidos (también hacer esto para combinaciones de atributos). Tuco argumenta que tiene que haber una clave en la relación. Además, afortunadamente, están los datos de ventas de los últimos cincuenta años. Así es que le sugiere, para mayor seguridad, chequear también las tablas correspondientes a estos cincuenta años. Las claves que se encuentren, Tuco dice que se llaman claves candidatas, y que de ellas se elige arbitrariamente una (no necesariamente la más breve en términos de número de atributos) como clave primaria.

Tico dice que Tuco no está 100% correcto, pero es incapaz de decir en qué está errado- Diga Usted quien tiene razón y justifique.

Solución:

Tuco está en error en el tema de determinación de la clave: no se puede obtener de una instancia (o de 51 instancias). En lo que está correcto es que si hay varias claves candidatas, se puede elegir cualquiera de ellas, independiente del número de atributos. Para encontrar claves candidatas, se puede ver qué conjunto de atributos determina al resto de los atributos. Tuco tiene razón en que hay al menos una clave: la de todos los atributos.