

PROBLEMAS DEL ESQUEMA RELACIONAL

Una vez obtenido el esquema relacional resultantes del modelo entidad relación que representaba la base de datos, normalmente tendremos una buena base de datos. Pero otras veces, debido a fallos en el diseño o a problemas indetectables en esta fase del diseño, tendremos un esquema que puede producir una base de datos que incorpore estos problemas:

- **Redundancia.** Se llama así a los datos que se repiten continua e innecesariamente por las tablas de las bases de datos.
- **Ambigüedades.** Datos que no clarifican suficientemente el registro al que representan.
- **Pérdida de restricciones de integridad.**
- **Anomalías en operaciones de modificación de datos.** El hecho de que al insertar un solo elemento haya que repetir tuplas en una tabla para variar unos pocos datos. O que eliminar un elemento suponga eliminar varias tuplas. El principio fundamental reside en que las tablas deben referirse a objetos o situaciones muy concretas. Lo que ocurre es que conceptualmente es difícil obtener ese problema. La solución suele ser dividir la tabla con problemas en otras tablas más adecuadas.

FORMAS NORMALES

Las formas normales se corresponde a una teoría de normalización iniciada por el propio Codd y continuada por otros autores (entre los que destacan Boyce y Fagin). Codd definió en 1970 la primera forma normal, desde ese momento aparecieron la segunda, tercera, la Boyce-Codd, la cuarta y la quinta forma normal.

Una tabla puede encontrarse en primera forma normal y no en segunda forma normal, pero no al contrario. Es decir los números altos de formas normales son más restrictivos (la quinta forma normal cumple todas las anteriores).

La teoría de formas normales es una teoría absolutamente matemática, pero en el presente manual se describen de forma intuitiva.

PRIMERA FORMA NORMAL (1FN)

Una tabla se encuentra en primera forma normal si impide que un atributo de una tupla pueda tomar más de un valor. La tabla:

TRABAJADOR		
DNI	Nombre	Departamento
12121212A	Andrés	Mantenimiento
12345345G	Andrea	Dirección Gestión

Visualmente es un tabla, pero no una tabla relacional (lo que en terminología de bases de datos relacionales se llama **relación**). No cumple la primera forma normal. Lo cumpliría si:

TRABAJADOR		
DNI	Nombre	Departamento
12121212A	Andrés	Mantenimiento
12345345G	Andrea	Dirección
12345345G	Andrea	Gestión

Esa tabla sí esta en primera forma normal.

DEPENDENCIAS FUNCIONALES

Se dice que un conjunto de atributos (Y) depende funcionalmente de otro conjunto de atributos (X) si para cada valor de X hay un único valor posible para Y. Simbólicamente se denota por $X \rightarrow Y$.

Por ejemplo el nombre de una persona depende funcionalmente del DNI, para un DNI concreto sólo hay un nombre posible. En la tabla ejemplo anterior, el departamento no tiene dependencia funcional, ya que para un mismo DNO puede haber más de un departamento posible.

Al conjunto X del que depende funcionalmente el conjunto Y se le llama **determinante**. Al conjunto Y se le llama **implicado**.

DEPENDENCIA FUNCIONAL COMPLETA

Un conjunto de atributos (Y) tiene una dependencia funcional completa sobre otro conjunto de atributos (X) si Y tiene dependencia funcional de X y además no se puede obtener de X un conjunto de atributos más pequeño que consiga una dependencia funcional de Y.

Por ejemplo en una tabla de clientes, el conjunto de atributos formado por el **nombre** y el **dni** producen una dependencia funcional sobre el atributo **apellidos**. Pero no es plena ya que el **dni** sólo también produce una dependencia funcional sobre **apellidos**. El dni sí produce una dependencia funcional completa sobre el campo apellidos.

Una dependencia funcional completa se denota como $X \Rightarrow Y$

DEPENDENCIA FUNCIONAL ELEMENTAL

Se produce cuando X e Y forman una dependencia funcional completa y además Y es un único atributo.

DEPENDENCIA FUNCIONAL TRANSITIVA

Es más compleja de explicar, pero tiene también utilidad. Se produce cuando tenemos tres conjuntos de atributos X, Y y Z. Y depende funcionalmente de X ($X \rightarrow Y$), Z depende funcionalmente de Y ($Y \rightarrow Z$). Además X no depende funcionalmente de Y. Entonces ocurre que X produce una dependencia funcional transitiva sobre Z. Esto se denota como: $(X \rightarrow Y, Y \rightarrow Z) \Rightarrow X \rightarrow Z$

Por ejemplo si X es el atributo **Número de Clase** de un instituto, e Y es el atributo **Código Tutor**. Entonces $X \rightarrow Y$ (el tutor depende funcionalmente del número de clase). Si Z representa el **Código del departamento**, entonces $Y \rightarrow Z$ (el código del departamento depende funcionalmente del código tutor, cada tutor sólo puede estar en un departamento). Como no ocurre que $Y \rightarrow X$ (el código de la clase no depende

funcionalmente del código tutor, un código tutor se puede corresponder con varios códigos de clase). Entonces $X \rightarrow Z$ (el código del departamento depende transitivamente del código de la clase).

SEGUNDA FORMA NORMAL (2FN)

Ocurre si una tabla está en primera forma normal y además cada atributo que no sea clave, depende de forma funcional completa respecto de cualquiera de las claves. Toda la clave principal debe hacer dependientes al resto de atributos, si hay atributos que depende sólo de parte de la clave, entonces esa parte de la clave y esos atributos formarán otra tabla.

Ejemplo:

ALUMNOS				
<u>DNI</u>	<u>Cod Curso</u>	Nombre	Apellido1	Nota
12121219A	34	Pedro	Valiente	9
12121219A	25	Pedro	Valiente	8
3457775G	34	Ana	Fernández	6
5674378J	25	Sara	Crespo	7
5674378J	34	Sara	Crespo	6

Suponiendo que el DNI y el número de curso formen una clave principal para esta tabla, sólo la nota tiene dependencia funcional completa. El nombre y los apellidos dependen de forma completa del DNI. La tabla no es 2FN, para arreglarlo:

ALUMNOS		
<u>DNI</u>	Nombre	Apellido1
12121219A	Pedro	Valiente
3457775G	Ana	Fernández
5674378J	Sara	Crespo

ASISTENCIA		
<u>DNI</u>	<u>Cod Curso</u>	Nota
12121219A	34	9
12121219A	25	8
3457775G	34	6
5674378J	25	7
5674378J	34	6

TERCERA FORMA NORMAL (3FN)

Ocurre cuando una tabla está en 2FN y además ningún atributo que no sea clave depende transitivamente de las claves de la tabla. Es decir no ocurre cuando algún atributo depende funcionalmente de atributos que no son clave.

Ejemplo:

ALUMNOS

ALUMNOS				
<u>DNI</u>	Nombre	Apellido1	Cod Provincia	Provincia
12121349A	Salvador	Velasco	34	Palencia
12121219A	Pedro	Valiente	34	Palencia
3457775G	Ana	Fernández	47	Valladolid
5674378J	Sara	Crespo	47	Valladolid
3456858S	Marina	Serrat	08	Barcelona

La Provincia depende funcionalmente del código de provincia, lo que hace que no esté en 3FN. El arreglo sería:

ALUMNOS			
<u>DNI</u>	Nombre	Apellido1	Cod Provincia
12121349A	Salvador	Velasco	34
12121219A	Pedro	Valiente	34
3457775G	Ana	Fernández	47
5674378J	Sara	Crespo	47
3456858S	Marina	Serrat	08

PROVINCIA	
Cod Provincia	Provincia
34	Palencia
47	Valladolid
08	Barcelona

FORMA NORMAL DE BOYCE-CODD (FNBC O BCFN)

Ocurre si una tabla está en tercera forma normal y además odo determinante es una clave candidata. Ejemplo:

TUTORÍAS		
<u>DNI</u>	<u>Asignatura</u>	Tutor
12121219A	Lenguaje	Eva
12121219A	Matemáticas	Andrés
3457775G	Lenguaje	Eva
5674378J	Matemáticas	Guillermo
5674378J	Lenguaje	Julia
5634823H	Matemáticas	Guillermo

Esa tabla está en tercera forma normal (no hay dependencias transitivas), pero no en forma de Boyce - Codd, ya que (DNI, Asignatura) → Tutor y Tutor → Asignatura. En este caso la redundancia ocurre por mala selección de clave. La

redundancia de la asignatura es completamente evitable. La solución sería:

TUTORÍAS	
<u>DNI</u>	<u>Tutor</u>
12121219A	Eva
12121219A	Andrés
3457775G	Eva
5674378J	Guillermo
5674378J	Julia
5634823H	Guillermo

ASIGNATURASTUTOR	
<u>Asignatura</u>	<u>Tutor</u>
Lenguaje	Eva
Matemáticas	Andrés
Matemáticas	Guillermo
Lenguaje	Julia

En las formas de Boyce-Codd hay que tener cuidado al descomponer ya que se podría perder información por una mala descomposición.

DEPENDENCIA MULTIVALUADA

Para el resto de formas normales (las diseñadas por Fagin, mucho más complejas), es importante definir este tipo de dependencia, que es distinta de las funcionales. Si las funcionales eran la base de la segunda y tercera forma normal (y de la de Boyce-Codd), éstas son la base de la cuarta forma normal.

Una dependencia multivaluada de una tabla con atributos X, Y, Z de X sobre Z (es decir $X \twoheadrightarrow Z$) ocurre cuando los posibles valores de Y sobre cualquier par de valores X y Z dependen sólo del valor de X y son independientes de Z.

Ejemplo:

<u>Nº Curso</u>	<u>Profesor</u>	<u>Material</u>
17	Eva	1
17	Eva	2
17	Julia	1
17	Julia	2
25	Eva	1
25	Eva	2
25	Eva	3

La tabla cursos, profesores y materiales del curso. La tabla está en FNBC ya que no hay dependencias transitivas y todos los atributos son clave sin dependencia funcional hacia ellos. Sin embargo hay redundancia. Los materiales se van a repetir para cualquier profesor dando cualquier curso, ya que los profesores van a utilizar todos los materiales del curso (de no ser así no habría ninguna redundancia).

Los materiales del curso dependen del curso y no del profesor en una dependencia multivaluada. Para el par N° de curso y profesora podemos saber los materiales, pero por el curso y no por el profesor.

CUARTA FORMA NORMAL (4FN)

Ocurre esta forma normal cuando una tabla está en forma normal de Boyce Codd y toda dependencia multivaluada es una dependencia funcional. Para la tabla anterior la solución serían dos tablas:

<u>Nº Curso</u>	<u>Material</u>
17	1
17	2
25	1
25	2
25	3

<u>Nº Curso</u>	<u>Profesor</u>
17	Eva
17	Julia
25	Eva

Un teorema de Fagin indica cuando hay tres pares de conjuntos de atributos X, Y y Z si ocurre $X \twoheadrightarrow Y|Z$ (Y y Z tienen dependencia multivaluada sobre X), entonces las tablas X,Y y X,Z reproducen sin perder información lo que poseía la tabla original. Este teorema marca la forma de dividir las tablas hacia una 4FN.

QUINTA FORMA NORMAL (5FN)

Es la más compleja y polémica de todas. Polémica pues no está claro en muchas ocasiones que sea una solución mejor que el no llegar a este nivel de normalización. Fue definida también por Fagin. Es raro encontrarse este tipo de problemas cuando la normalización llega a 4FN. Se deben a restricciones muy concretas.

Ejemplo:

<u>Proveedor</u>	<u>Material</u>	<u>Proyecto</u>
1	1	2
1	2	1
2	1	1
1	1	1

Indican códigos de material suministrado por un proveedor y utilizado en un determinado proyecto.

Si ocurre una restricción especial como por ejemplo: Cuando un proveedor nos ha suministrado alguna vez un determinado material, si ese material aparece en otro proyecto, haremos que el proveedor nos suministre también ese material para ese proyecto.

Eso ocurre en los datos como el proveedor número 1 nos suministró el material número 1 para el proyecto 2 y en el proyecto 1 utilizamos el material 1, aparecerá la tupla proveedor 1, material 1 y proyecto 1.

La dependencia que produce esta restricción es lejana y se la llama de reunión. Para esa restricción esta división en tablas sería válida:

<u>Proveedor</u>	<u>Material</u>
1	1
1	2
2	1

<u>Material</u>	<u>Proyecto</u>
1	2
2	1
1	1

Esa descomposición no pierde valores en este caso, sabiendo que si el proveedor nos suministra un material podremos relacionarle con todos los proyectos que utilizan ese material.

Resumiendo, una tabla no está en quinta forma normal si hay una descomposición de esa tabla que muestre la misma información que la original.