



**Universidad
Europea de Madrid**

LAUREATE INTERNATIONAL UNIVERSITIES

NORMALIZACIÓN

NORMALIZACIÓN AVANZADA

Índice

Presentación.....	3
Problemas derivados de las claves candidatas solapadas.....	4
FNBC	6
Ejercicio resuelto	7
Situación del mundo real modelizada con DMV	9
Dependencias multivaluadas.....	11
4FN	12
Ejemplo.....	13
Otras FN.....	15
Resumen.....	16

Presentación

Durante esta unidad de aprendizaje estamos estudiando el procedimiento a seguir para obtener un conjunto de relaciones que evite, en la medida de lo posible, los **problemas de actualización** derivados de un **mal diseño**. Ya hemos dado los primeros pasos: sabemos obtener un esquema relacional que esté en 3FN, para lo cual hemos aprendido a trabajar con **dependencias funcionales**.

¿Una relación en 3FN carece de problemas? La respuesta es no. Cuando en una relación existen varias claves candidatas que tienen atributos en común pueden aparecer problemas. Además, entre los atributos de una relación se establecen otro tipo de relaciones, más difíciles de encontrar en el mundo real, pero que acarrearán problemas de actualización. Son las **dependencias multivaluadas** (DMV).

En este tema aprenderás a:

- Detectar situaciones en el mundo real de esquemas en 3FN y que siguen teniendo problemas de actualización.
- Normalizar dichos esquemas a FNBC (**Forma Normal de Boyce-Codd**).
- Detectar situaciones del mundo real que se representan con dependencias multivaluadas.
- Analizar los problemas derivados de esquemas de este tipo.
- Descomponer la relación para obtener un esquema sin pérdidas que se encuentre en 4FN.

Introduciremos la existencia de otras formas normales aunque no las trataremos en detalle pues son muy difíciles de detectar.



Problemas derivados de las claves candidatas solapadas

Una relación en 3FN puede seguir teniendo problemas de actualización y suelen aparecer cuando la relación contiene dos o más claves candidatas compuestas, o las claves candidatas tienen al menos un atributo en común, es decir, están solapadas. Veamos un ejemplo. Añadiremos una relación sobre sesiones de fisioterapia para los futbolistas.

#jugador	fecha	hora	#fisio	#habitación
j012	12-5-2006	17:30	f05	h01
j976	12-5-2006	13:00	f05	h01
j084	12-5-2006	13:00	f07	h02
j135	27-3-2006	17:30	f05	h02

Las dependencias funcionales que se cumplen en esta relación son éestas.

Esta relación está en 3FN porque no hay dependencias parciales ni transitivas con la clave principal de la relación (#jugador, fecha). De hecho no hay atributos no primos en esta relación por lo que cumple hasta la 3FN. Sin embargo, sigue teniendo problemas de actualización. Por ejemplo, modificar la habitación del fisioterapeuta F05 el 12-5-2006, supondría modificar dos filas lo que puede producir posibles incoherencias. Esto es así porque el determinante de la DF *#fisio, fecha* → *#habitación* no es clave.



Precaución

Precaución

Descomponer relaciones puede hacer que no se conserven las DF originales. En estos casos podría ser mejor no avanzar en la normalización. Habrá que analizar qué es mejor: la existencia de redundancia por dejar la relación en 3FN o perder una DF.

Ejemplo: Si solo hay una cita diaria no hace falta llegar a BCNF.

Dependencias funcionales

- A un jugador en un día determinado se le asigna una sesión de fisioterapia que es impartida por un fisioterapeuta en una habitación.

#jugador, fecha \rightarrow hora, #fisio, #habitación

- Un fisioterapeuta solo puede atender a un jugador en una hora determinada del día.

#fisio, fecha, hora \rightarrow #jugador

- Las habitaciones son individuales.
- A un jugador en un día determinado se le asigna una sesión de fisioterapia que es impartida por un fisioterapeuta en una habitación.

#habitación, fecha, hora \rightarrow #fisio, #jugador

- En un día un fisioterapeuta no cambia de habitación.

#fisio, fecha \rightarrow #habitación

FNBC

La FNBC (**Forma Normal de Boyce-Codd**) parte de la definición general de la 3FN, que elimina cualquier dependencia transitiva con respecto a cualquier clave candidata.

Una relación R está en FNBC si y sólo si todo determinante es una clave candidata.

Toda relación R que esté en FNBC está también en 3FN, pero al contrario no ocurre siempre lo mismo. Obsérvese que esta definición no hace alusión a 1FN ni a 2FN y tampoco al concepto de dependencia transitiva. Por tanto, para probar que una relación está en FNBC no hace falta probar previamente que está en 2FN y 3FN como ocurría con las FN anteriormente estudiadas.

Cualquier relación puede descomponerse sin pérdidas en un conjunto equivalente de relaciones en FNBC. Dada la DF $X \rightarrow Y$ que no cumple la FNBC, descomponemos la relación original R_0 en:

- $R_1 = R_0 - \{Y\}$.
- $R_2 = \{X\} \cup \{Y\}$.

La existencia de varias claves candidatas no es necesariamente mala. En cualquier caso, es aconsejable especificar al SGBD la existencia de claves candidatas (por ejemplo en SQL con la cláusula UNIQUE).



Ejercicio resuelto

Normalicemos la relación **tareas** utilizando directamente la FNBC. Recordemos la relación y su conjunto de DF:

```
ASIGNACION_TAREAS (dni_emp, nombre_emp, dirección_emp,  
cargo_emp, salario_emp, tarea, tipo_tarea, descripción_tarea,  
categoría_tarea, coste-hora, fecha-inic, fecha-fin, estado,  
proyecto, estimación-horas)
```

DF1 dni_emp → nombre_emp, direccion_emp, cargo_emp, salario_emp

DF2 tarea → tipo_tarea, descripcion_tarea, categoria_tarea, coste_hora, fecha_inic, fecha_fin, estado, proyecto

DF3 dni_emp, tarea → estimacion-horas

DF4 cargo_emp → tipo_tarea

DF5 categoria_tarea → coste_hora

Análisis FNBC

Una relación está en FNBC si todo determinante es clave primaria ó clave candidata. En nuestro caso las dependencias funcionales 1, 2, 4 y 5 incumplen esta regla. Hay, por tanto, que aislarlas en relaciones propias.



[Descomposición de la DF1](#)



[Descomposición de la DF2](#)



[Descomposición de la DF5](#)

El esquema resultado, que sí cumple la FNBC, es **éste**.

Descomposición de la DF1

De la DF1 obtenemos la descomposición:

- R1: empleados (dni_emp, nombre_emp, direccion_emp, cargo_emp, salario_emp)
- R0: asignación_tareas (dni_emp, tarea, tipo_tarea, descripcion_tarea, categoria_tarea, coste_hora, fecha_inicio, fecha_fin, estado, proyecto, estimacion_horas)

Descomposición de la DF2

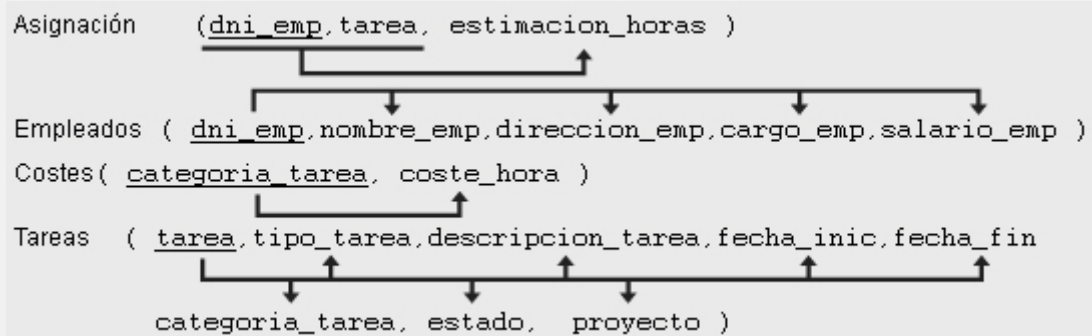
Para aislar la DF2 sustituimos R0 por:

- R2: tareas (tarea, tipo_tarea, descripcion_tarea, categoria_tarea, coste_hora, fecha_inicio, fecha_fin, estado, proyecto)
- R0: asignacion_tareas (dni_emp, tarea, estimacion_horas)

Descomposición de la DF5

La DF4 ya no está presente en nuestro esquema. Como resultado de aislar la DF5 presente ahora en R2 obtenemos la siguiente relación:

- R21: costes (categoria_tarea, coste_hora)
- R2: tareas (tarea, tipo_tarea, descripcion_tarea, categoria_tarea, fecha_inicio, fecha_fin, estado, proyecto)

Esquema resultado

Situación del mundo real modelizada con DMV

Supongamos una relación que almacena la información sobre los pedidos de materiales de oficina de los departamentos de una empresa. Se almacena información sobre el departamento, los proveedores que suministran a cada departamento y los materiales que utilizan. Esta relación tiene una serie de restricciones semánticas.

Restricciones semánticas

Estas restricciones semánticas son:

- Un **departamento** puede ser abastecido por cualquiera de los **proveedores** con los que trabaja este departamento.
- Cada **pedido** que realiza un **departamento** incluye un conjunto fijo de **materiales**.
- Un **proveedor** puede suministrar a varios **departamentos**.

La siguiente instancia de la relación **pedidos** (departamento, proveedor, material) cumple con las restricciones semánticas anteriores:

departamento	proveedor	material
sistemas informáticos	ofimat SL	torre cd
sistemas informáticos	ofimat SL	rotuladores pizarra blanca
sistemas informáticos	mediomercado	torre cd
sistemas informáticos	mediomercado	rotuladores pizarra blanca
marketing	ofimat SL	paquete 500 folios
marketing	ofimat SL	recambio impresora
marketing	ofimat SL	fundas folios taladradas

Se puede observar que en la relación **pedidos** no hay dependencias funcionales, además cumple hasta la FNBC puesto que no existen atributos no primos. Pero la relación sigue teniendo problemas de actualización. Por ejemplo, si queremos añadir un nuevo **suministrador** al departamento de Sistemas Informáticos, habría que añadir 3 tuplas a la relación **pedidos**: una por cada uno de los **materiales** que utiliza este departamento. Estos problemas vienen derivados del hecho de que suministradores y materiales son independientes entre sí. La situación mejoraría si partimos la tabla anterior en:

Proveedores (departamento, proveedor)

Materiales (departamento, material)

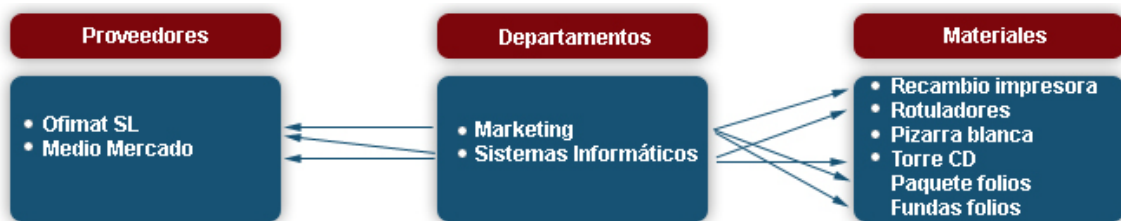
Estas dos relaciones no tienen atributos no primos por lo que están en FNBC.

Dependencias multivaluadas

En 1977 se formalizó este problema con la introducción, por parte de Fagin, del concepto de **dependencia multivaluada** (DMV). Las dependencias multivaluadas surgen como consecuencia de la restricción de la 1FN.

Se dice que X multidetermina a Y ($X \twoheadrightarrow Y$) si el conjunto de valores de Y depende de X y no depende del resto de los atributos de R. También se dice que Y es multidependiente de X.

Dicho de otro modo: $X \twoheadrightarrow Y$ si conociendo X conozco el conjunto de valores de Y. Definen relaciones de puntos a conjuntos, en el ejemplo de los pedidos, $f(\text{systemas informáticos}) = \{\text{torre cd, rotuladores pizarra blanca}\}$.



La dependencia multivaluada es una generalización del concepto de DF. Esto implica que toda DF es una DMV, es decir, que las DF son un caso particular de las DMV. Dada una DMV $X \twoheadrightarrow Y$, si el conjunto imagen es unitario obtenemos una DF. Lo contrario no es cierto. Para que una DMV sea perjudicial dentro de una relación debe darse acompañada de otra, y los atributos multidependientes tienen que ser independientes entre sí.

Dada la relación $R(A,B,C)$, la DMV $R.A \twoheadrightarrow R.B$ se cumple si, y sólo si, se cumple también $R.A \twoheadrightarrow R.C$ y $R.B$ es independiente de $R.C$. Se nota: $R.A \twoheadrightarrow R.B | R.C$.

En nuestro ejemplo de pedidos hay dos DMV: **departamento** \twoheadrightarrow **proveedor**, **departamento** \twoheadrightarrow **material**. Además, **proveedor** y **material** son independientes entre sí puesto que el conjunto de materiales suministrados por un proveedor no depende de éste sino del **departamento**.

4FN

Una DMV $X \twoheadrightarrow Y$ puede ser:

- **Trivial**, si:
 - Y es un subconjunto de X, o
 - $X \cup Y = R$.
- **No trivial**, si no se cumple ninguna de las dos restricciones anteriores.

Un DMV no trivial impondrá una restricción que afectará a la relación, por tanto son las que pueden plantear problemas.

Una relación R está en **4FN** si está en FNBC y no tiene dependencias multivaluadas (DMV) no triviales.

Una relación R de atributos A, B y C se puede **descomponer sin pérdidas** en sus proyecciones $R_1(A,B)$ y $R_2(A,C)$ si y solo si se cumple la DMV $A \twoheadrightarrow B|C$.

La relación **pedidos** (departamento, proveedor, material) cumple las DMV:

- Departamento \twoheadrightarrow proveedor: si conozco el nombre del departamento puedo saber el conjunto de proveedores que lo suministran.
- Departamento \twoheadrightarrow material: si conozco el departamento puedo saber cuál es el conjunto de materiales que solicitan en todos sus pedidos.

Además, **proveedor** y **material** son independientes entre sí. Resumiendo, se cumple la DMV:

Departamento \twoheadrightarrow proveedor | material

Esta relación la podemos descomponer sin pérdidas en las siguientes. De esta forma obtenemos un nuevo esquema que está en 4FN.

Proveedores (departamento, proveedor)

Materiales (departamento, material)

Ejemplo

Eres el administrador de bases de datos de la UEM. Se te ha pedido, desde el departamento de la OTRI (Oficina de Transferencia de Resultados de Investigación), que crees una base de datos para almacenar la información relativa a los proyectos de investigación que tiene en marcha la universidad. En la universidad existen varios grupos de investigación, a cada uno de ellos pertenecen varios profesores. Un profesor puede pertenecer a varios grupos. Cada grupo de investigación interviene en varios proyectos. Un proyecto puede ser realizado por varios grupos de investigación distintos que participan en distintas partes del proyecto. Cada profesor participa en la realización de todos los proyectos del grupo al que pertenece.

El esquema inicial está compuesto por la relación **investigación**: *investigación* (grupo, profesor, proyecto)

¿Está en FNBC?**En detalle****¿Está en FNBC?**

Analicemos en qué FN se encuentra esta relación. Empecemos por las **DF**, ¿cuáles son las DF que se cumplen en la relación **investigación**? Realmente no existe ninguna relación 1 a 1, puesto que un grupo tiene varios profesores y varios proyectos, y un profesor puede pertenecer a varios grupos y estar implicado en varios proyectos (todos los de los grupos de investigación a los que pertenece). Además, en cada proyecto pueden intervenir varios grupos de investigación y todos los profesores que pertenecen a éstos. No existe, por tanto, ninguna DF en este esquema y, por tanto, está en FNBC.

Extraigamos ahora las **DMV** que se cumplen. Son las siguientes:

Grupo → profesor

Grupo → proyecto

Grupo → profesor | proyecto

DMV

- En la universidad existen varios grupos de investigación, a cada uno de ellos pertenecen varios profesores. Esto implica que dado un grupo de investigación, sabemos qué conjunto de profesores pertenecen a él.

Grupo \rightarrow profesor

- Cada grupo de investigación interviene en varios proyectos. Dado el grupo, conocemos el conjunto de proyectos en los que interviene.

Grupo \rightarrow proyecto

- Además, en cada proyecto pueden intervenir varios grupos de investigación y todos los profesores que pertenecen a éstos. Esto implica que el proyecto en el que participa un profesor no depende de él sino del grupo al que pertenece. Profesores y proyectos son, por tanto, independientes entre sí:

Grupo \rightarrow profesor | proyecto

La relación tiene una DMV no trivial y, por tanto, no se encuentra en 4FN. La descomponemos como sigue:

Profesores (grupo, profesor)

Proyectos (grupo, proyecto)

Nos queda así un esquema que está en 4FN.

Otras FN

¿Una relación en 4FN carece de problemas de actualización? En la mayoría de los casos sí. Sin embargo, existen situaciones del mundo real que no son modelizables con DF, ni con DMV. Tales situaciones se dan muy poco y además son difíciles de detectar y difíciles de tratar. Es por esto que dejaremos el estudio de la normalización aquí aunque sí daremos un esbozo de las dependencias de reunión, aquellas que permiten situar una relación en 5FN.



Dependencias de reunión y 5FN

Supongamos el siguiente esquema SPJ($S\#,P\#,J\#$)

obtenido de (Date, 1993) cuyas tuplas significan: Un determinado proveedor S_i suministra la pieza P_i al proyecto J_i . En la relación anterior se cumple la restricción: Si **Pérez** suministra **tuercas**, el proyecto **Madrid12** usa tuercas y Pérez suministra a Madrid12, entonces Pérez es suministrador de tuercas para el proyecto Madrid12. Esta restricción hace que la relación SPJ no se puede descomponer sin pérdidas en dos proyecciones, pero sí en tres.

La relación R satisface la **dependencia de reunión** (DR) $*(R_1,R_2,...,R_n)$, donde R_i se compone de un subconjunto de atributos de R si

$$R = \text{reunión de } R_1,R_2,...,R_n$$

SPJ satisface la DR $*(SP, PJ, JS)$, es decir se descompone sin pérdidas en SP, PJ, JS.

Una relación está en 5FN si y sólo si toda DR es una consecuencia de las claves candidatas.

Resumen

Una relación está en **FNBC** si está en 1FN y todo determinante es clave. Para comprobar si una relación está en FNBC no es necesario pasar por las anteriores. Puede ocurrir que una relación esté en 3FN y no lo esté en FNBC, lo contrario no es cierto. Esto puede ocurrir cuando existan varias claves candidatas solapadas. Para obtener un **esquema equivalente** en FNBC podemos **descomponer sin pérdidas** la relación original en dos: una con los atributos de la DF y otra con el resto de los atributos de R más el determinante de la dependencia. Una relación está en **4FN** si está en **FNBC** y no contiene **DMV no triviales**. Las DMV son **relaciones de punto a conjunto** entre dos atributos de una relación. Las DMV no triviales son aquellas del tipo

$A \twoheadrightarrow B|C$, es decir, se deben dar dos DMV dentro de la relación y además los atributos determinados ser independientes entre sí.

Existen casos patológicos en los que puede ocurrir que una relación esté en 4FN y siga teniendo problemas de actualización. Estos casos son estudiados por otros tipos de dependencias como las **dependencias de reunión** relacionadas con la **5FN**.

