



**Universidad  
Europea de Madrid**

**LAUREATE** INTERNATIONAL UNIVERSITIES

**NORMALIZACIÓN**

**EJERCICIO RESUELTO**

**Índice**

Presentación.....	3
Descripción del problema.....	4
Estudio de la 1FN.....	5
Extracción de las DF.....	6
Estudio de la 2FN.....	7
Estudio de la 3FN.....	9
Estudio de la FNBC.....	10
Extracción de las DMV y análisis de la 4FN.....	11
Esquema resultado.....	12
Resumen.....	14

### Presentación

A lo largo de esta unidad de aprendizaje hemos descrito el proceso a seguir para normalizar una relación. Este tema lo vamos a dedicar a utilizar un ejemplo que ilustre el proceso desde el principio. Anteriormente vimos cómo obtener el esquema relacional inicial aplicando las reglas para pasar a tablas el diagrama E/R. El proceso de normalización puede aplicarse:

1. Sobre este esquema inicial.
2. Directamente sobre la información extraída de la especificación del sistema resultado del proceso de análisis.

Normalmente actuaremos según la primera opción: construiremos el **DER**, a partir de éste el **esquema relacional** inicial y aplicaremos el **proceso de normalización** sobre estas tablas. El esquema relacional inicial suele tener un buen comportamiento, es decir, muchas de las tablas que lo forman ya están en 5FN. Exponer aquí un ejemplo siguiendo esta estrategia no ilustraría lo suficientemente el proceso de normalización. Por ello, en este tema utilizaremos la opción 2.



En este tema aplicarás, mediante un ejemplo, el proceso de normalización de una BD, siguiendo los siguientes pasos:

- Describir el problema y obtener el esquema en una sola tabla.
- Realizar el estudio con la 1FN.
- Extraer a partir del enunciado las DF que se dan entre los atributos del problema para poner el esquema hasta FNBC.
- Estudiar la existencia de DMV.
- Normalizar el esquema hasta 4FN.

### Descripción del problema

Eres el administrador de base 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**, cada uno de ellos identificado por un **nombre** y un **código**.



A cada grupo de investigación pertenecen varios **profesores**, cada uno de ellos identificado por el **número de empleado**, el **nombre**, el **departamento** al que pertenece, las **palabras clave** que identifican sus intereses investigadores, la **categoría profesional** y el **salario**.

Cada categoría profesional (director, agregado, adjunto, asociado, ayudante, colaborador y auxiliar) tiene asignado según el plan de carrera docente un salario fijo. Un profesor puede pertenecer a distintos grupos de investigación. Un profesor participa obligatoriamente en **todos** los proyectos del grupo de investigación al que pertenece.

Cada grupo de investigación tiene varios **proyectos** en marcha. Para cada proyecto se almacena el **nombre**, la **línea de investigación** a la que pertenece, la **financiación** conseguida y la **organización** de la que recibe dicha financiación.

La financiación de un proyecto depende de la línea de investigación a la que pertenezca y de la organización que subvencione el proyecto. Cada proyecto tiene un **investigador principal** que estará a cargo del mismo. En cada proyecto pueden participar varios grupos, cada uno de los cuales cubrirá las tareas derivadas de su línea de investigación.

### Estudio de la 1FN

El esquema que tenemos que normalizar según la descripción anterior es el siguiente:

```
Investigación_UEM (#grupo, nombre_gr, n_empleado, nombre_prof,
dpto_prof, palabras_clave, categoría_prof, salario, proyecto,
línea, financiación, organismo, investigador_pral)
```

Este esquema almacena información acerca de los grupos de investigación, los profesores que lo componen y los proyectos que tienen en marcha.

Esta relación está en **1FN** si cada atributo tiene un valor único. Fijémonos en el atributo **palabras\_clave** que define los intereses investigadores de un profesor. Este atributo incumple esta condición puesto que puede contener varios valores por **profesor**. Hay que tener en cuenta la siguiente observación: el atributo palabras\_clave sería un atributo **multivaluado** en el DER.

Para transformar esta tabla a 1FN tenemos dos opciones:

1. **Aplanar la tabla** haciendo que cada fila incluya una de las palabras clave del profesor por lo que aparecería la información del **profesor**, de los **grupos** a los que pertenece y de los **proyectos** de investigación en los que participa repetidos varias veces.
2. **Descomponer la relación en dos**: una que contenga el **grupo repetitivo** (en este caso palabra\_clave) y otra que contenga el **resto de atributos no repetitivos más la clave**.

La opción dos es mejor puesto que evita redundancias. Seguiremos, por tanto, esta estrategia, resultando las siguientes relaciones:

```
Investigación_UEM (#grupo, nombre_gr, n_empleado, nombre_prof, dpto_prof, categoría_prof, salario,
proyecto, línea, financiación, organismo, investigador_pral)
```

```
Palabras_clave (n_empleado, palabra_clave)
```

### Extracción de las DF

El siguiente paso es definir el conjunto de DF que se dan entre los atributos de la tabla.

Si fijamos la atención en la tabla **Investigación\_UEM** comprobamos que realmente se están mezclando los atributos propios de **grupos de investigación** con los de **profesores** y con los de **proyectos**. Esto implica que hay DF entre **grupos** y sus **atributos**. Lo mismo puede decirse de los **profesores** y de los **proyectos**. Las DF que aparecen son:

#grupo  $\rightarrow$  nombre\_gr

n\_empleado  $\rightarrow$  nombre\_prof, dpto\_prof, categoría\_prof, salario

proyecto  $\rightarrow$  línea, financiación, organismo, investigador\_pral

Además hay que modelizar las restricciones que explícitamente aparecen en el enunciado del problema:

- “[...] Cada categoría profesional (director, agregado, adjunto, asociado, ayudante, colaborador y auxiliar) tiene asignado según el plan de carrera docente un salario fijo [...]”. Esto implica que sabiendo la categoría del profesor podemos saber unívocamente el salario que le corresponde y, por tanto, existe una DF entre estos atributos.

Categoría\_prof  $\rightarrow$  salario

- “[...] La financiación de un proyecto depende de la línea de investigación a la que pertenezca y de la organización que subvencione el proyecto [...]”. Esto implica que, sabiendo la línea de investigación y la organización que subvenciona el proyecto, conocemos la financiación obtenida por el proyecto. La DF que modeliza esta restricción es:

Línea, organismo  $\rightarrow$  financiación

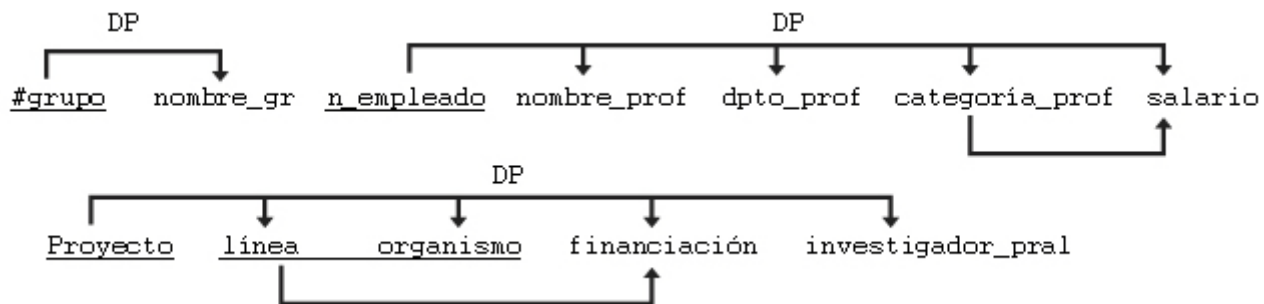
### Estudio de la 2FN

El primer paso es comprobar cuáles son las **claves** de la relación. Debemos encontrar un conjunto de atributos cuyo cierre contenga todos los atributos de la relación investigación\_UEM. Aplicando las DF establecidas anteriormente:

```
{#grupo, n_empleado, proyecto}+= {nombre_gr, nombre_prof,
dpto_prof, categoría_prof, salario, línea, financiación,
organismo, investigador_pral}
```

La clave primaria es, por tanto, **#grupo + n\_empleado + proyecto**. Además, no existe ningún otro conjunto de atributos que cumpla la condición anterior por lo que no existen claves candidatas.

Para detectar las DP buscamos las flechas que partan de un subconjunto de la clave primaria y desemboquen en un atributo no primo. La figura muestra las DP:



Para obtener un esquema equivalente en 2FN creamos una tabla para cada DP y otra con el resto de atributos más los determinantes de las DP.

Grupos (#grupo, nombre\_gr)

Profesores (n\_empleado, nombre\_prof, dpto\_prof, categoría\_prof, salario)

Proyectos (proyecto, línea, financiación, organismo, investigador\_pral)

Investigación\_UEM (#grupo, n\_empleado, proyecto)

La otra **tabla** del esquema en 1FN, está en 2FN puesto que no se establecen DF entre los atributos de ésta.

**Claves**

Para comprobar si la relación investigación\_UEM está en 2FN tenemos que detectar la existencia de dependencias parciales en el esquema. Una dependencia parcial se establece entre un subconjunto de la clave primaria ó de la clave candidata -atributo determinante-, y un atributo no primo -atributo determinado-. Por tanto, antes de esto tenemos que saber cuáles son la clave primaria y las claves candidatas de la relación.

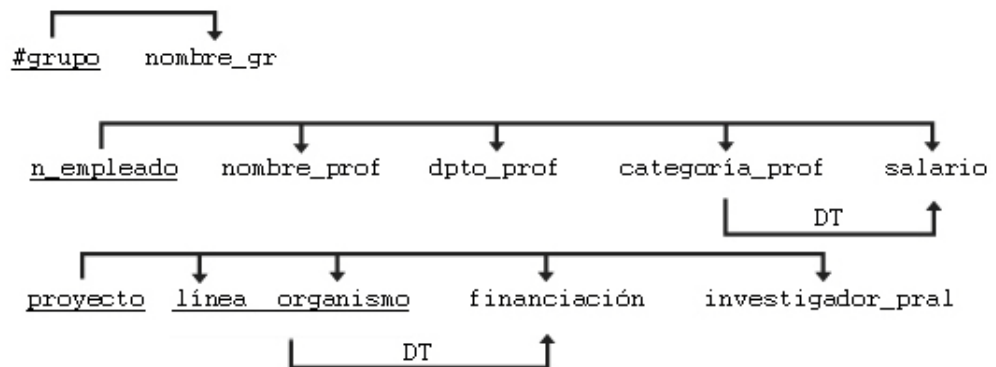
**Tabla**

Palabras\_clave (n\_empleado, palabra\_clave)



### Estudio de la 3FN

Una relación está en 3FN si todo atributo no primo depende de forma no transitiva de la clave primaria ó de las claves candidatas. Tenemos que buscar DT en nuestras tablas. Las DT son aquellas que se establecen entre atributos no primos. La siguiente figura muestra las tablas donde se identifican las DT. No se muestran las tablas **investigación\_UEM** ni **Palabras\_clave** puesto que no existen DF entre sus atributos.



Para eliminar las DT del esquema hay que descomponer las tablas de **profesores** y **proyectos**. Las descomponemos en dos: una tabla que tenga la **DT** y otra con el **resto de los atributos más el determinante de la relación**. Quedaría:

Salarios (categoría\_prof, salario)

Financiación (línea, organismo, financiación).

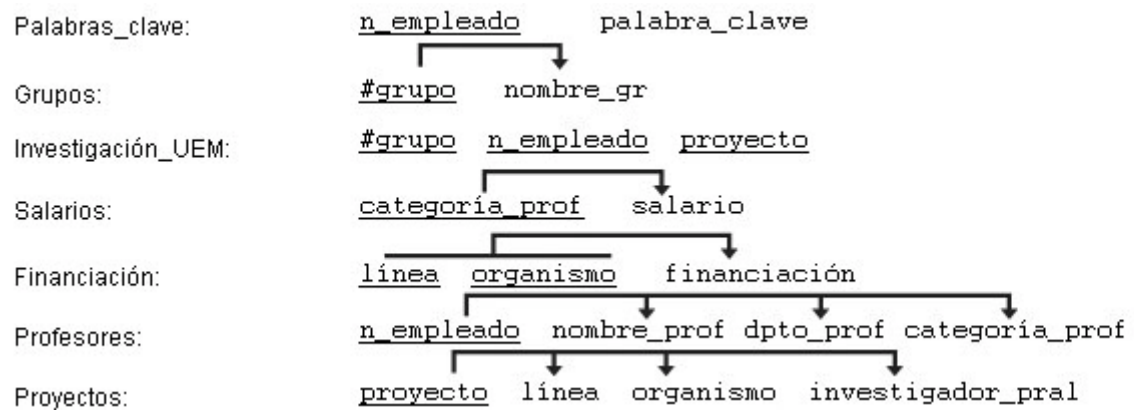
Profesores (n\_empledo, nombre\_prof, dpto\_prof, categoría\_prof)

Proyectos (proyecto, línea, organismo, investigador\_pral)

La tabla de **grupos de investigación** queda como está porque no tiene dependencias transitivas. Esta descomposición es sin pérdidas y además conserva todas las dependencias que se daban en la relación original.

### Estudio de la FNBC

Una relación está en FNBC si todo determinante es clave primaria ó clave candidata. El siguiente diagrama muestra la situación actual de las DF.



Como se puede comprobar toda flecha parte de las claves primarias de la relación por lo que el esquema ya se encuentra en FNBC.

### Alternativa

El estudio de la 2FN y la 3FN nos lo podíamos haber ahorrado aplicando directamente la FNBC sobre el esquema inicial en 1FN. La siguiente figura muestra esta forma de operar. En la parte superior aparece la relación investigación\_UEM inicial y en la parte inferior aparecen las tablas en las que la hemos descompuesto para solucionar todas las DF perjudiciales: aquellas que parten de un atributo que no es la clave primaria y desembocan en un atributo que es no primo.



Esquema de la FNBC

### Extracción de las DMV y análisis de la 4FN

Pasemos ahora al análisis de las DMV. Existe una DMV entre dos atributos cuando conocido el **determinante** podemos conocer el conjunto de valores que tomará el **determinado**. En este ejemplo se dan las siguientes:

#grupo  $\twoheadrightarrow$  n\_empleado  
 #grupo  $\twoheadrightarrow$  proyecto  
 n\_empleado  $\twoheadrightarrow$  pal\_clave

Las **DMV no triviales** son las que son perjudiciales en las relaciones. Para que una DMV sea no trivial tiene que ocurrir que: se den dos DMV dentro de la misma relación y que los determinantes sean independientes entre sí.

La última DMV de las mostradas antes se encuentra aislada en la tabla **Palabras\_clave** por lo que no tiene problema. Las otras dos se encuentran en la tabla **investigación\_UEM**. ¿Son independientes los atributos **n\_empleado** y **proyecto**? Si volvemos al enunciado del problema encontramos la restricción: “[...] Un profesor participa obligatoriamente en **todos** los proyectos del grupo de investigación al que pertenece [...].”

Esto implica que los proyectos en los que participa un profesor solo dependen del grupo de investigación al que pertenece y, por tanto, n\_empleado y proyecto son atributos independientes entre sí. Resumiendo, en la relación investigación\_UEM se da la siguiente DMV no trivial:

#grupo  $\twoheadrightarrow$  n\_empleado | proyecto

Descomponemos la relación anterior en dos:

Grupo\_profesor(#grupo, n\_empleado)

Grupo\_proyecto (#grupo, proyecto)



**Esquema resultado**

El esquema resultado obtenido es:

Palabras\_clave (n\_empleado, palabra\_clave)

Grupos (#grupo, nombre\_gr)

Salarios (categoría\_prof, salario)

Financiación (línea, organismo, financiación).

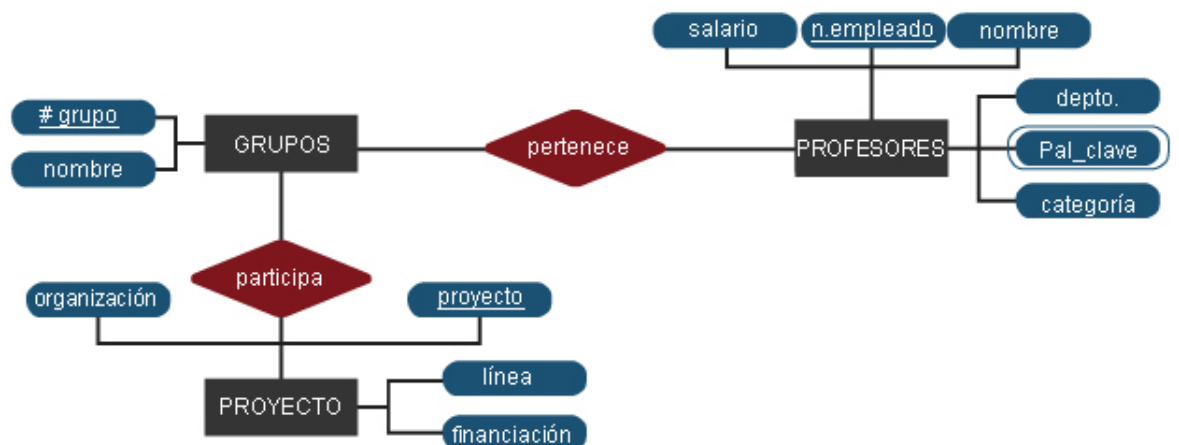
Profesores (n\_empleado, nombre\_prof, dpto\_prof, categoría\_prof)

Proyectos (proyecto, línea, organismo, investigador\_pral)

Grupo\_profesor(#grupo, n\_empleado)

Grupo\_proyecto (#grupo, proyecto)

¿Sería este el esquema que se extraería del DER?



La imagen muestra el DER que se obtiene del enunciado del problema. A partir de éste se obtienen otras **tablas**

**Tablas**

Las otras tablas que se obtendrían son:

Grupo\_profesor (#grupo, n\_empleado)

Grupo\_proyecto (#grupo, proyecto).

Palabras\_clave(n\_empleado, palabra\_clave)

Profesores (n\_empleado, nombre\_prof, dpto\_prof, categoría\_prof, salario)

Proyectos (proyecto, línea, financiación, organismo, investigador\_pral)

Grupos (#grupo, nombre\_gr)

Estas tablas están en 4FN salvo por las DT estudiadas. Sólo habría que descomponer las tablas **empleados** y **proyectos** para aislar estas DT y, así, obtendríamos un esquema igual que el anterior.

## Resumen

A lo largo de este tema hemos seguido el proceso de normalización completo para un ejemplo. La estrategia de diseño que seguiremos a partir de ahora consistirá en obtener el **diagrama ER**, pasarlo a **tablas** para conseguir el **esquema inicial** y aplicar a las tablas de éste las **formas normales** descritas. En este capítulo hemos seguido el proceso de normalización desde el principio para que ilustre mejor el proceso.

El proceso de normalización tiene las siguientes fases:

- Estudiar el estado de las tablas del esquema en 1FN comprobando si tiene atributos con varios valores para una fila (**grupo repetitivo**). En ese caso aislaremos el grupo repetitivo en una tabla que contendrá además la clave de la relación original.
- 
- Extraer el conjunto de DF que se dan en nuestro problema. En este paso tenemos dos opciones:
  - Aplicar 2FN (aislando dependencias parciales en relaciones aparte), 3FN (aislando dependencias transitivas) y FNBC (aislando las DF hacia atributos no primos que no partan de claves primarias ó candidatas) ó
  - Aplicar directamente la **FNBC**.
- Estudiar la existencia de **DMV perjudiciales** del tipo  $A \twoheadrightarrow B|C$  y obtener un esquema equivalente en 4FN.

El resultado de todo este proceso será un esquema relacional que tiene un buen comportamiento frente a los problemas de actualización ya conocidos.