

# NORMALIZACIÓN DE BASES DE DATOS

## INTRODUCCION

Las formas normales definidas en la Teoría de Base de Datos Relacionales representan una guía y una orientación para el diseño de registros. Las reglas de normalización están destinadas a prevenir anomalías en las actualizaciones e inconsistencia en los datos.

Las directrices que se ofrecerán parten del supuesto de que aquellos campos que no constituyen una clave serán actualizados frecuentemente.

El propósito de la normalización es mejorar la integridad de los datos a través de la minimización de la redundancia y la inconsistencia, pero con algún posible costo en ciertas aplicaciones.

El término **normalización** se usa algunas veces en relación a una forma normal particular. Esto es, un conjunto de registros puede ser normalizado con respecto a la segunda forma normal pero no con respecto a la tercera.

## 1. PRIMERA FORMA NORMAL

La **Primera Forma Normal** (1FN) trata con el *modelo de un registro*, es decir, la forma básica o registro tipo.

"Cuando se manipulan las relaciones de un modelo de visión se les trata como conjuntos matemáticos de eneadas (tuplas). Esto requiere entre otras cosas, que los atributos dentro de cada eneada sean ordenados y completos, y que los dominios permitan sólo valores simples. Los valores simples no pueden descomponerse en múltiples valores y no pueden por sí mismos ser conjuntos o relaciones."

Según esto, una repetición anidada de un elemento (o de un grupo de elementos) no puede ser representada en un modelo como una sola relación. Ahora bien, si un grupo anidado tiene tamaño acotado y es pequeño, es posible representarlo usando un número fijo de atributos asignados a diferentes papeles sobre el mismo dominio. Debido a que los grupos pueden ser más pequeños que el número máximo fijo, el dominio debe incluir el valor nulo o indefinido. Con mayor frecuencia se traslada el grupo anidado a otra relación.

*Al proceso de eliminar grupos anidados se le denomina NORMALIZACION, y a las relaciones resultantes se les considera expresadas en la PRIMERA FORMA NORMAL*

**EJEMPLO PARA LA PRIMERA FORMA NORMAL**

Información que se desea manejar:

PROFESOR	Nombres y Apellidos
DEDICACION	Tiempo dedicado a la Universidad
MATERIAS	Nombres de las materias que dicta
CURSOS	Número de cursos por materia
DEPARTAMENTO	Nombre del Departamento al cual está adscrito el Profesor.
COMISIONES	Nombre se las comisiones a las que pertenece el Profesor.
COORDINACION	¿Coordina el Profesor la comisión?

Supongamos que tenemos los siguientes datos:

PROFESOR	José García
DEDICACION	Exclusiva
MATERIAS Y CURSOS	Métodos Numéricos (2 cursos). Estadística (1 curso). Matemática II (1 curso).
DEPARTAMENTO	Matemática.
COMISIONES	Seminario de Investigación (Coordinador) Reestructuración Curricular. Entrenamiento Docente (Coordinador).

Tratemos de construir el registro:

PROFESOR	DEDICACIÓN	MATERIAS	CURSOS	DEPART.	COMISIONES	COORDINA
José García	Exclusiva	Métodos Num., Estadística, Matemática II	2, 1, 1	Matemática	Sem. de Invest., Reestruc. Curr., Entren. Docente	Sí, No, Sí

Debe haberse notado que en los campos MATERIAS y COMISIONES existen diversos valores para el mismo individuo, además de una serie de datos relacionados tales como CURSOS POR MATERIAS y COORDINACION DE COMISIONES. Suponiendo que un Profesor no dicta más de tres materias por semestre, podríamos construir el registro de la siguiente forma

**RELACION PROFESOR**

NOMBRE	DEDIC.	MAT. 1	CUR. 1	MAT. 2	CUR. 2	MAT. 3	CUR. 3	DEPTO.
J. García	Exclus.	M. Numér.	2	Estadíst.	1	Mat. II	1	Matem.

**RELACION PARTICIPA-EN-COMISION**

<b>PROFESOR</b>	<b>COMISION</b>	<b>COORDINA</b>
José García	Seminario de Investigación	Sí
José García	Reestructuración Curricular	No
José García	Entrenamiento Docente	Sí

El procedimiento empleado ha sido el siguiente:

- El campo MATERIA para el cual se esperaban como máximo tres valores diferentes (con tres valores asociados: numero de cursos), se descompuso en tres nuevos campos, a saber: Materia 1, Materia 2 y Materia 3, denominados respectivamente como MAT1, MAT2 y MAT3. Los valores asociados (número de cursos por materia) se colocaron a su vez como tres nuevos campos: CUR1, CUR2 y CUR3, que corresponderán respectivamente a las materias numeradas del 1 al 3. Debe tenerse cuidado al manejar de esta forma los registros el no exceder el número de campos permitidos por el sistema administrador de base de datos que Ud. utiliza.
- Con respecto a las COMISIONES y a COORDINACION, se decidió emplear otra técnica que consiste en separar el campo y construir otra tabla. Tal decisión se tomó en vista de la independencia de la relación MATERIA-COMISIONES. Esta nueva tabla (la hemos llamado PARTICIPA-EN-COMISION para mayor claridad), contiene los campos PROFESOR, COMISION y COORDINA (este último de tipo lógico).

La repetición del Nombre del Profesor en el campo PROFESOR nos da el enlace a partir de las tuplas en la nueva relación PARTICIPA-EN-COMISION.

## 2. SEGUNDA Y TERCERA FORMAS NORMALES

Estas tratan con la relación entre un campo clave y uno o más campos dependientes. Bajo la Segunda y Tercera Formas Normales un campo dependiente debe proporcionar un dato acerca de la clave (relacionado con la clave), con toda la clave y con ningún otro campo. Además, debe satisfacer la primera forma normal.

Trataremos ahora solamente con datos de un solo valor. Un dato de un solo valor sería una relación *uno a muchos*, tal como el Departamento de un empleado o una relación *uno a uno* tal como la esposa de un empleado. Por lo tanto, la frase “*provee un dato acerca de X*” significa relación uno a uno o uno a muchos entre X e Y.

### 2.1. SEGUNDA FORMA NORMAL

La **Segunda Forma Normal** (2FN) es violada cuando un campo dependiente es un dato acerca de un subconjunto de la clave. Por supuesto, esto es relevante sólo si la clave es compuesta. Consideremos los siguientes ejemplos

Ejemplo 1

ARTÍCULO	DEPÓSITO	CANTIDAD	DIRECCIÓN

\_\_\_\_\_ Clave \_\_\_\_\_

Ejemplo 2

MATERIA	PROFESOR	SECCIONES	DEPARTAMENTO

\_\_\_\_\_ Clave \_\_\_\_\_

Estudiemos el primer caso:

- La clave es compuesta, consiste de los campos ARTÍCULO y DEPÓSITO.
- CANTIDAD es un campo relacionado con ARTICULO-DEPOSITO (es decir, con la clave).
- DIRECCION es un dato acerca de DEPOSITO, que es un subconjunto de la clave,

Inconvenientes de este diseño:

- La Dirección del Depósito será repetida en cada registro referido a cada artículo correspondiente a un mismo depósito.
- Si la Dirección del Depósito cambia, todos los registros referidos a un artículo de un mismo depósito deben ser actualizados.
- Debido a la redundancia, los datos pueden tornarse inconsistentes con diferentes registros exhibiendo diferentes direcciones para el mismo depósito.
- Si en algún momento no existen artículos guardados en un depósito, pudiera no existir un registro en el cual mantener la dirección del depósito.

Para satisfacer la Segunda Forma Normal la tabla deberá ser descompuesta o reemplazada por dos tablas:

ARTÍCULO	DEPÓSITO	CANTIDAD

DEPÓSITO	DIRECCIÓN

## 2.2. TERCERA FORMA NORMAL

La **Tercera Forma Normal** (1FN) es violada cuando un campo dependiente es un dato acerca de otro campo dependiente. Consideremos el siguiente ejemplo:

EMPLEADO	DEPARTAMENTO	UBICACIÓN

----- Clave -----

Estudiemos el ejemplo:

- La clave EMPLEADO es simple.
- Si cada DEPARTAMENTO está ubicado en un lugar, entonces el campo UBICACION es un dato acerca del DEPARTAMENTO, y además es un dato acerca del EMPLEADO.

Inconvenientes de este diseño:

Presenta los mismos problemas causado por la violación de la Segunda Forma Normal:

- La ubicación del DEPARTAMENTO es repetida en el registro de cada EMPLEADO asignado a ese DEPARTAMENTO.
- Si la ubicación del DEPARTAMENTO cambia, cada registro debe ser actualizado.
- Debido a la redundancia, los datos pueden tornarse inconsistente; por ejemplo, diferentes registros exhibirán diferentes DIRECCIONES para el mismo DEPARTAMENTO.
- Si un DEPARTAMENTO no tiene EMPLEADOS, puede no existir un registro en el cual mantener la UBICACIÓN de tal DEPARTAMENTO.

Para satisfacer la Tercera Forma Normal, el registro debe ser descompuesto en dos nuevos registros:

EMPLEADO	DEPARTAMENTO	DEPARTAMENTO	UBICACIÓN

En resumen, un registro está en la Segunda o Tercera Forma Normal si todos los campos son parte de la clave o proveen un dato (un valor simple) acerca de exactamente el campo clave y ningún otro campo.

### 3. CUARTA Y QUINTA FORMAS NORMALES

La Cuarta y Quinta Formas Normales tratan con datos multivaluados (múltiples valores). Un dato multivaluado puede corresponder a una relación *muchos a muchos*, como con EMPLEADOS y HABILIDADES, o a una relación *muchos a uno*, como con los HIJOS de un EMPLEADO (asumiendo que sólo un pariente es empleado). Por *muchos a muchos* entendemos que un empleado puede tener diversas habilidades y que una habilidad puede residir en diversos empleados. Note que vemos la relación muchos a uno entre hijos y padres como un dato de un solo valor acerca de un hijo pero como un dato multivaluado acerca de un padre.

En un sentido, la Cuarta y Quinta Formas Normales están alrededor de claves compuestas. Estas formas normales procuran minimizar el número de campos involucrados en una clave compuesta.

### 3.1. CUARTA FORMA NORMAL

Bajo la **Cuarta Forma Normal** (4FN) un registro tipo no deberá contener dos o más datos multivaluados independientes acerca de una entidad. Además, el registro debe satisfacer la Tercera Forma Normal.

Considere EMPLEADOS, HABILIDADES e IDIOMAS, donde un empleado puede tener diversas habilidades y conocer varios idiomas. Tenemos aquí dos relaciones *muchos a muchos*, una entre EMPLEADOS y HABILIDADES, y una entre EMPLEADOS e IDIOMAS. Bajo la Cuarta Forma Normal estas dos relaciones no podrán ser representadas en un registro simple tal como:

EMPLEADO	HABILIDAD	IDIOMA

En cambio, ellas podrán ser representadas en dos registros:

EMPLEADO	HABILIDAD
<i>Clave</i>	

EMPLEADO	IDIOMA
<i>Clave</i>	

Note que otros campos que no incluyan datos multivaluados son permitidos en el registro, como en el caso del campo CANTIDAD en el ejemplo anterior ARTICULO-DEPOSITO.

El principal problema con la violación de la Cuarta Forma Normal es que existe incertidumbre en cuanto a las políticas de mantenimiento. Es posible tener diversas políticas para mantener los datos independientes multivaluados en cada registro:

- Un formato disjunto en el cual un registro contiene o una habilidad o un idioma pero no ambos.

EMPLEADO	HABILIDAD	IDIOMA
GONZALEZ	COCINA	
GONZALEZ	MANEJA	
GONZALEZ		FRANCES
GONZALEZ		ALEMAN
GONZALEZ		GRIEGO

En este caso, no existe diferencia en cuanto a mantener dos tipos de registros separados. Por otra parte, semejante formato conduce a ambigüedades con respecto al significado del campo *vacío*. Una habilidad vacía podría significar que el empleado no tiene habilidades, que el campo no es aplicable a este empleado, que el dato es desconocido, o, como en este caso, que el dato puede estar ubicado en otro registro.

- Una mezcla al azar, con tres variaciones

- Mínimo número de registros con repeticiones

EMPLEADO	HABILIDAD	IDIOMA
GONZALEZ	COCINA	FRANCES
GONZALEZ	MANEJA	ALEMAN
GONZALEZ	MANEJA	GRIEGO

- Mínimo número de registros con valores nulos.

EMPLEADO	HABILIDAD	IDIOMA
GONZALEZ	COCINA	FRANCES
GONZALEZ	MANEJA	ALEMAN
GONZALEZ		GRIEGO

- Sin restricciones

EMPLEADO	HABILIDAD	IDIOMA
GONZALEZ	COCINA	FRANCES
GONZALEZ	MANEJA	
GONZALEZ		ALEMAN
GONZALEZ	MANEJA	GRIEGO

- El producto cartesiano, donde para cada empleado habrá un registro para todos los posibles pares de una de sus habilidades con uno de sus idiomas.

EMPLEADO	HABILIDAD	IDIOMA
GONZALEZ	COCINA	FRANCES
GONZALEZ	COCINA	ALEMAN
GONZALEZ	COCINA	GRIEGO
GONZALEZ	MANEJA	FRANCES
GONZALEZ	MANEJA	ALEMAN
GONZALEZ	MANEJA	GRIEGO

Inconvenientes de este diseño:

Presenta los mismos problemas mencionados para la Segunda y Tercera Formas Normales. Produce diferentes violaciones dependiendo de la política de mantenimiento seleccionada.



- Se presentarán repeticiones, por lo tanto la actualización deberá ejecutarse sobre registros múltiples y éstos pueden tornarse inconsistentes.
- La inserción de una nueva habilidad podría involucrar consideraciones para un registro con una habilidad vacía, insertando un nuevo registro con un posible idioma vacío, o insertando múltiples registros apareando la habilidad con alguno o todos los idiomas.
- Al borrar una habilidad, sacar la habilidad localizada en uno o más registros, o borrar uno o más registros conectados con un contenido que en la última alusión de algún idioma fue borrado también.

La Cuarta Forma Normal minimiza tales problemas de actualización.

### 3.2. QUINTA FORMA NORMAL

La **Quinta Forma Normal** (5FN) trata con casos donde la información puede ser reconstruida de muchas piezas de información las cuales pueden ser mantenidas con poca redundancia.

La Segunda, Tercera y Cuarta Formas Normales también sirven a este propósito pero la Quinta Forma Normal generaliza los casos no cubiertos por ellas. No intentaremos una exposición amplia de la Quinta Forma Normal pero ilustraremos el concepto central con un ejemplo, a saber:

- Entidades: AGENTES, COMPAÑIAS y PRODUCTOS.
- Si los AGENTES representan COMPAÑÍAS, las COMPAÑÍAS fabrican PRODUCTOS, y los AGENTES venden PRODUCTOS, entonces nosotros querríamos tener guardado un registro de cuáles agentes venden cuáles productos para cuál compañía.
- Esta información puede ser guardada en un registro con tres campos:

AGENTE	COMPAÑÍA	PRODUCTO
PARRA	FORD	AUTOMOVIL
PARRA	GENERAL MOTORS	CAMION

Esta forma es necesaria en el caso general. Ahora bien, aunque el agente PARRA vende autos hechos por FORD y camiones hechos por GENERAL MOTORS; él no vende camiones FORD ni autos GM. Esto es, necesitamos la combinación de los tres campos para saber cuáles combinaciones son válidas y cuáles no.

Ahora bien, supongamos la siguiente regla: si un agente vende cierto producto y él representa la compañía que lo fabrica, entonces él vende un producto para esa compañía.

AGENTE	COMPAÑÍA	PRODUCTO
PARRA	FORD	AUTOMOVIL
PARRA	FORD	CAMION
PARRA	GENERAL MOTORS	AUTOMÓVIL
PARRA	GENERAL MOTORS	CAMION
PEREZ	FORD	AUTOMÓVIL

En este caso, resulta que podemos reconstruir todos los datos reales de una forma normalizada consistente de tres tipos de registros separados, cada uno conteniendo dos campos

AGENTE	COMPAÑÍA	AGENTE	PRODUCTO
PARRA	FORD	PARRA	AUTOMOVIL
PARRA	GM	PARRA	CAMION
PEREZ	FORD	PEREZ	AUTOMOVIL

COMPAÑÍA	PRODUCTO
FORD	AUTOMÓVIL
FORD	CAMION
GM	AUTOMOVIL
GM	CAMION

Estos tres registros están en la Quinta Forma Normal, puesto que el correspondiente registro de tres campos presentado previamente no lo está.

Para hablar de un modo general, podemos decir que un registro está en la Quinta Forma Normal cuando la información que contiene no puede ser reconstruida por varios registros más pequeños, por ejemplo, de registros que tienen cada uno menos campos que el original. Se excluye el caso en el cual todos los registros más pequeños tienen la misma clave. Si un registro puede ser descompuesto solamente en registros pequeños, teniendo todos la misma clave, entonces el registro es considerado como en la Quinta Forma Normal, sin descomposición. Un registro en la Quinta Forma Normal está también en la Cuarta, Tercera, Segunda y Primera Formas normales.

La Quinta Forma Normal no difiere de la Cuarta Forma Normal, excepto en que allí existe una simetría compulsiva, tal como la regla acerca de los Agentes, Compañías y Productos. En ausencia de tal compulsión, un registro en la Cuarta Forma Normal está siempre en la Quinta Forma Normal.

Una ventaja de la Quinta Forma Normal es que ciertas redundancias pueden ser eliminadas. En la forma normalizada, el hecho de que PARRA venda AUTOS es registrado solamente una vez; en la forma no normalizada esto puede ser repetido muchas veces. Debemos advertir entonces que aun cuando las formas normalizadas involucran un número mayor de registros tipo comparado con el original, las ocurrencias de algunos hechos se reducen considerablemente.

Lo importante es comprender que cuanto más hechos son registrados, el tamaño del archivo normalizado crece en forma aditiva, mientras que el tamaño del archivo no normalizado crece en forma multiplicativa. Por ejemplo, si agregamos un nuevo AGENTE que vende 'X' PRODUCTOS para 'Y' COMPANIAS, donde cada una de estas compañías fabrica cada uno de estos productos, tendremos  $X + Y$  nuevos registros para la forma normalizado, pero tendremos  $X \cdot Y$  nuevos registros para la forma no normalizada

Es también importante hacer notar que todos los tres registros tipo son requeridos en la forma normalizada a fin de reconstruir la misma información. De los registros presentados anteriormente sabemos que PEREZ representa a la FORD y que FORE fabrica CAMIONES y AUTOS, Pero no podemos determinar si PEREZ vende CAMIONES o AUTOS FORD hasta que veamos el otro registro tipo.

El siguiente ejemplo ilustra un caso en el cual la regla acerca de los agentes, compañías y productos es satisfecha y el cual, evidentemente, requiere los tres registros tipo en la forma normalizada. Alguno de los registros tipos considerados anteriormente implicarán datos falsos.

AGENTE	COMPAÑÍA	PRODUCTO
PARRA	FORD	AUTOMOVIL
PARRA	FORD	CAMION
PARRA	GENERAL MOTORS	AUTOMÓVIL
PARRA	GENERAL MOTORS	CAMION
PEREZ	FORD	AUTOMÓVIL
PEREZ	FORD	CAMION
RIOS	FORD	AUTOMÓVIL
RIOS	GENERAL MOTORS	AUTOMÓVIL
RIOS	TOYOTA	AUTOMÓVIL
RIOS	TOYOTA	BUS

AGENTE	COMPAÑÍA
PARRA	FORD
PARRA	GM
PEREZ	FORD
RIOS	FORD
RIOS	GM
RIOS	TOYOTA

COMPAÑÍA	PRODUCTO
FORD	AUTOMOVIL
FORD	CAMION
GM	AUTOMOVIL
GM	CAMION
TOYOTA	AUTOMÓVIL
TOYOTA	BUS

AGENTE	PRODUCTO
PARRA	AUTOMOVIL
PARRA	CAMION
PEREZ	AUTOMOVIL
PEREZ	CAMION
RIOS	AUTOMÓVIL
RIOS	BUS

Estos tres registros están en la Quinta Forma Normal

Observe que

- PEREZ vende AUTOS y GM fabrica AUTOS, pero PEREZ no representa a GM.
- RIOS representa a la FORD y FORD fabrica CAMIONES, pero RIOS no vende CAMIONES.
- RIOS representa a la FORD y RIOS vende BUSES, pero FORD no fabrica BUSES.

## Ejemplo de Normalización de una Base de Datos

Con la finalidad de dar una visión clara del proceso de normalización, nos disponemos a desarrollar un caso donde se podrá apreciar cómo se salvan los diferentes obstáculos que se puedan presentar durante este proceso (Normalización).

Trataremos de desarrollar una base de datos que contenga la siguiente información:

- Nombre del Estudiante: NombreE
- Cédula de Identidad: CédulaE
- Edad: EdadE
- Materias que cursa: NombreM
- Semestre (de Materia): SemestreM
- Sección (de Materia): SecciónM
- Nombre del Representante: NombreR
- Dirección del Represent.: DirecciónR
- Profesor de la Materia: ProfesorM
- Notas: N1, N2, N3, N4

Veamos como se podría organizar lógicamente una relación con esta información, indicando algunos valores, para los diferentes atributos del mismo:

**Estudiante:** (NombreE, CedulaE, EdadE, NombreM, SemestreM, SeccionM, NombreR, DireccionR, ProfesorM, N1, N2, N3, N4)

NombreE	CeduE	EdadE	NombreR	DireccR	NombM	SemM	SeccM	ProfeSM	N1	N2	N3	N4
P. GARCIA	10101010	21	J. GARCIA	Haticos	Matem. I	2	A	A. PARRA	10	12	14	16
					Estadíst.	2	A	L. SOTO	09	15	16	10
					Informát.	3	C	P. PEREZ	15	15	15	15
J. ACOSTA	12121212	20	R.ACOSTA	Delicias	Matem. II	3	B	A. PARRA	05	08	07	11
					Geometr.	4	A	R. BRITO	08	08	10	12
					Informát.	3	C	P. PEREZ	14	16	18	18

Si observamos este archivo, notaríamos cómo a cada estudiante le corresponderían varias materias, secciones, semestres y profesores; es decir, cada registro o entidad tendría una serie de atributos los cuales aceptarían varios valores. Estos atributos son conocidos como atributos multivaluados.

Sabiendo que la primera forma normal (1FN) exige que los atributos deben mostrar solo valores simples o atómicos, podemos llegar a la conclusión de que la relación anterior no se encuentra en 1FN.

Ahora bien, si tratáramos de llevar esta relación a su 1FN, tendríamos que descomponerla en dos relaciones:

**Estudiante:** (NombreE, CedulaE, EdadE, NombreR, DireccionR)

**Inscripción:** (NombreE, NombreM, SeccionM, ProfesorM, N1, N2, N3, N4)

Tabulemos algunos valores para estas relaciones:

**Estudiante:**

NombreE	CedulaE	EdadE	NombreR	DireccionR
P. GARCIA	10101010	21	J. GARCIA	Haticos
J. ACOSTA	12121212	20	R.ACOSTA	Delicias

---- Clave ----

**Inscripciones:**

NombreE	NombreM	SemM	SeccM	ProfesorM	N1	N2	N3	N4
P. GARCIA	Matem. I	2	A	A. PARRA	10	12	14	16
P. GARCIA	Estadística	2	A	L. SOTO	09	15	16	10
P. GARCIA	Informática	3	C	P. PEREZ	15	15	15	15
J. ACOSTA	Matem. II	3	B	A. PARRA	05	08	07	11
J. ACOSTA	Geometría	4	A	R. BRITO	08	08	10	12
J. ACOSTA	Informática	3	C	P. PEREZ	14	16	18	18

----- Clave -----

De esta manera, tanto la relación **Estudiante** como la relación **Inscripción** están en la 1FN. Obsérvese que los valores de los atributos son todos simples, que es la restricción impuesta por la 1FN.

La segunda forma normal (2FN) exige que los atributos de la parte dependiente lo sean funcionalmente de la parte rectora completa.

Con respecto a la relación **Estudiante** podemos decir que este problema no debe existir (es decir, si se han asignado correctamente los atributos) puesto que la parte rectora de la misma es simple (la constituye un sólo atributo); no sucede lo mismo con la relación **Inscripción**, la cual posee una parte rectora o clave compuesta.

Sí analizamos el atributo **SemestreM** de la relación **Inscripción** nos daremos cuenta que este quedaría definido perfectamente por el valor del atributo **NombreM** sin importar el valor del atributo **NombreE**; dicho de otra forma, depende sólo de un elemento de la parte rectora de esa relación, por lo tanto, no se encuentra en la 2FN.

Para llevar esta relación a la 2FN, debemos construir dos nuevas relaciones:

**Materia:** (NombreM, SemestreM)

**Inscripción:** (NombreE, NombreM, SeccionM, ProfesorM, N1, N2, N3, N4)

Tabulando valores, tenemos:

**Materia:**

NombreM	SemM
Matemática I	2
Estadística	2
Informática	3
Matemática II	3
Geometría	4

**Inscripción:**

NombreE	NombreM	SeccM	ProfesorM	N1	N2	N3	N4
P. GARCIA	Matem. I	A	A. PARRA	10	12	14	16
P. GARCIA	Estadística	A	L. SOTO	09	15	16	10
P. GARCIA	Informática	C	P. PEREZ	15	15	15	15
J. ACOSTA	Matem. II	B	A. PARRA	05	08	07	11
J. ACOSTA	Geometría	A	R. BRITO	08	08	10	12
J. ACOSTA	Informática	C	P. PEREZ	14	16	18	18

Una vez hechas estas transformaciones, notamos cómo los atributos de la parte dependiente, tanto en la relación **Estudiante** como en la relación **Inscripción** lo son completamente de sus respectivas partes rectoras, condición necesaria para que estén en la 2FN.

La Tercera Forma Normal (3FN) busca eliminar las dependencias, cualesquiera que se encuentren entre los atributos de la parte dependiente. Regresando a la relación **Estudiante**, podemos notar como el atributo **DireccionR** es dependiente del atributo **NombreR**, lo que significa que esta relación no se encuentra en su 3FN.

Ahora bien, para lograr esta nueva normalización debemos construir otras dos relaciones:

**Representante:** (NombreR, DireccionR)

**Estudiante:** (NombreE, CedulaE, EdadE, NombreR)

Tabulando valores para estas relaciones tenemos:

**Representante:**

NombreR	DireccionR
J. GARCIA	Haticos
R. ACOSTA	Delicias

**Estudiante:**

NombreE	CedulaE	EdadE	NombreR
P. GARCIA	10101010	21	J. GARCIA
J. ACOSTA	12121212	20	R. ACOSTA

De esta forma las relaciones **Materia** e **Inscripción** se encuentran en su 3FN, y además se encuentran en la Forma Normal Boyce-Codd (BCFN) por cuanto no existen dependencias triviales y las dependencias no triviales lo son funcionalmente de la parte rectora completa; es decir, todos los atributos de los que depende algún otro están en una clave. No sucede lo mismo con la relación **Estudiante** donde existe una dependencia trivial entre el atributo **NombreE** y **CedulaE** ya que el primero define al segundo, pero **CedulaE** también define a **NombreE**, y además **CedulaE** no pertenece a una clave; para resolver el problema definimos la relación

**Identificación:** (CedulaE, NombreE)

De esta manera, todas las relaciones están en la Forma Normal Boyce-Codd.

Dado que la Cuarta (4FN) y Quinta (5FN) Forma Normal no son aplicables al ejemplo que hemos desarrollado, aplicaremos ésta a un nuevo ejemplo, que si bien tiene alguna relación con el anterior es independiente del mismo.



## Segundo ejemplo.

Supongamos una relación no normalizada

**Comunes:** (NombreM, Escuela, ProfesorM) (materias comunes):

Tabulemos algunos valores:

**Comunes:**

NombreM	Escuela	ProfesorM
MATEMÁTICA II	COMPUTACIÓN ELECTRICA	A. PARRA R. BRITO L. SOTO
ESTADISTICA	COMPUTACIÓN ELECTRICA QUÍMICA	F. RINCÓN J. FERRER A. PARRA

Si tratáramos de normalizar esta relación, la presentaríamos así:

**Comunes:**

NombreM	Escuela	ProfesorM
MATEMÁTICA II	COMPUTACIÓN	A. PARRA
MATEMÁTICA II	COMPUTACIÓN	R. BRITO
MATEMÁTICA II	COMPUTACIÓN	L. SOTO
MATEMÁTICA II	ELECTRICA	A. PARRA
MATEMÁTICA II	ELECTRICA	R. BRITO
MATEMÁTICA II	ELECTRICA	L. SOTO
ESTADÍSTICA	COMPUTACIÓN	F. RINCÓN
ESTADÍSTICA	COMPUTACIÓN	J. FERRER
ESTADÍSTICA	COMPUTACION	A. PARRA
ESTADÍSTICA	ELECTRICA	F. RINCÓN
ESTADÍSTICA	ELECTRICA	J. FERRER
ESTADÍSTICA	ELECTRICA	A. PARRA
ESTADÍSTICA	QUÍMICA	F. RINCÓN
ESTADÍSTICA	QUÍMICA	J. FERRER
ESTADÍSTICA	QUÍMICA	A. PARRA

Presentada de esta forma, la relación estaría normalizada en la **BCNF**, ya que todos los atributos formarían parte de la clave y no hay dependencia entre ellos. Por supuesto que una relación de esta forma tendría demasiada redundancia y problemas a la hora de una actualización.

Para resolver este problema derivaremos dos nuevas relaciones a partir de la anterior.

**Materia-Escuela:** (NombreM, Escuela)

**Materia-Profesor:** (NombreM, ProfesorM)

Obsérvese que los atributos forman parte de la clave en cada relación; así mismo se pueden adicionar a cada relación otros atributos no multivaluados.

Tabulando algunos valores para las relaciones anteriores:

**Materia-Escuela**

NombreM	Escuela
MATEMÁTICA II	COMPUTACIÓN
MATEMÁTICA II	ELECTRICA
ESTADÍSTICA	COMPUTACIÓN
ESTADÍSTICA	ELECTRICA
ESTADÍSTICA	QUÍMICA

**Materia-Profesor:**

NombreM	ProfesorM
MATEMÁTICA II	A. PARRA
MATEMÁTICA II	R. BRITO
MATEMÁTICA II	L. SOTO
ESTADÍSTICA	F. RINCÓN
ESTADÍSTICA	J. FERRER
ESTADISTICA	R. BRITO

De esta forma se resuelve el problema de la actualización, ya que para incluir un nuevo profesor de una materia determinada bastaría con agregarlo a la relación **Materia-Profesor** una sola vez; mientras que en la relación **Comunes** deberíamos agregarlo tantas veces como escuelas compartieran la misma materia.

Para ilustrar la Quinta Forma Normal (5FN), utilizaremos una relación para determinar los profesores que están dando alguna materia específica y el título o especialidad requerido para dictar la misma:

**Materia:** (ProfesorM, NombreM, Especialidad)

Nuevamente, todos los atributos forman parte de la clave y son multivaluados; tabulando algunos valores, tenemos:

NombreM	ProfesorM	Especialidad
MATEMÁTICA II	R. BRITO	LIC. MATEMÁTICA
INFORMATICA	R. BRITO	COMPUTACIÓN
MATEMÁTICA II	A. PARRA	INGENIERO
ESTADÍSTICA	R. BRITO	LIC. MATEMÁTICA
ESTADÍSTICA	J. FERRER	ADMINISTRADOR
ESTADÍSTICA	F. RINCON	INGENIERO
GEOMETRÍA	A. PARRA	INGENIERO

Como puede verse, ningún atributo define otro, por lo tanto todos están en la clave. Si nosotros descompusiéramos la relación **Materia** en dos relaciones, podría suceder que al tratar de reconstruir la relación **Materia-Estudiante**, obtuviéramos algunas eneadas sorpresivas, por ejemplo:

MATEMÁTICA II	F. RINCON	INGENIERO
---------------	-----------	-----------

Lo que no sería consistente ya que F. RINCON no dicta la materia MATEMÁTICA II. La 5FN permite descomponer una relación en estas condiciones en tres o más relaciones, así:

#### **Materia-Especialidad**

NombreM	Especialidad
MATEMÁTICA II	LIC. MATEMÁTICA
MATEMÁTICA II	INGENIERO
INFORMATICA	COMPUTACIÓN
ESTADÍSTICA	LIC. MATEMÁTICA
ESTADÍSTICA	ADMINISTRADOR
ESTADÍSTICA	INGENIERO
GEOMETRIA	INGENIERO

#### **Materia-Profesor:**

NombreM	ProfesorM
MATEMÁTICA II	R. BRITO
INFORMATICA	R. BRITO
MATEMÁTICA II	A. PARRA
ESTADÍSTICA	R. BRITO
ESTADÍSTICA	J. FERRER
ESTADÍSTICA	F. RINCÓN
GEOMETRÍA	A. PARRA

**Profesor-Especialidad:**

<b>ProfesorM</b>	<b>Especialidad</b>
R. BRITO	LIC. MATEMÁTICA
R. BRITO	COMPUTACIÓN
J. FERRER	ADMINISTRADOR
F. RINCON	INGENIERO
A. PARRA	INGENIERO

Si combináramos la tabla **Materia-Especialidad** con la tabla **Profesor-Especialidad** obtendríamos la siguiente tabla:

**Profesor-Especialidad:**

<b>NombreM</b>	<b>ProfesorM</b>	<b>Especialidad</b>	
MATEMÁTICA II	R. BRITO	LIC. MATEMÁTICA	
MATEMÁTICA II	A. PARRA	INGENIERO	
MATEMÁTICA II	F. RINCON	INGENIERO	***
INFORMATICA	R. BRITO	COMPUTACIÓN	
ESTADÍSTICA	R. BRITO	LIC. MATEMÁTICA	
ESTADÍSTICA	J. FERRER	ADMINISTRADOR	
ESTADÍSTICA	A. PARRA	INGENIERO	***
ESTADÍSTICA	F. RINCON	INGENIERO	
GEOMETRÍA	A. PARRA	INGENIERO	
GEOMETRÍA	F. RINCON	INGENIERO	***

Como puede observarse, las eneadas marcadas con asteriscos no aparecen en la relación **Materia** que vimos primeramente, lo que significa que estas combinaciones no son válidas.

Sin embargo, si se combinan las tablas **Materia-Especialidad**, **Materia-Profesor** y **Profesor-Especialidad** reproducimos la tabla **Materia** original.