

NORMALIZACIÓN - (GUÍA RÁPIDA)

El objetivo de la Normalización de una base de datos es evitar redundancias que desaprovechan espacio en disco, eliminar dependencias innecesarias entre atributos y solucionar problemas de elección de claves que puedan haber surgido durante la fase de realización de los esquemas relacionales.

El proceso de normalización se realiza en varias etapas secuenciales. Cada etapa está asociada a una forma normal, que establece unos requisitos a cumplir por la tabla sobre la que se aplica.

Existen varias formas normales: Primera, Segunda, Tercera, Boyce-Codd, Quinta y Dominio Clave.

El paso de una forma normal a otra es consecutivo, si no se satisface una determinada forma normal no puede pasarse al análisis de la siguiente.

Como norma general, para garantizar que no existan problemas en la actualización de datos, es recomendable aplicar el proceso de normalización hasta Tercera Forma Normal o incluso hasta Forma Normal de Boyce-Codd.

Vamos a tomar como ejemplo la siguiente tabla y restricciones:

(En esta ocasión se añaden algunos registros para más fácil comprensión).

PELÍCULAS

Título	Director	Nac_direc	Formato	Precio	Género	Duración	ID_Prod
El padrino	Francis Ford	EEUU	HD	15	Drama, suspense	175	1
El padrino	Francis Ford	EEUU	3D	24	Drama, suspense	175	1
Toy Story	John Lasseter	EEUU	3D	12	Animación, aventura	81	2

Restricciones:

1. Una película puede pertenecer a varios géneros.
2. Cada película solo está dirigida por un director.
3. Una película puede estar en formato HD o en formato 3D.

PRIMERA FORMA NORMAL

Dada una tabla relacional, decimos que está en 1ª Forma Normal si:

- No tiene atributos multivaluados.
- No tiene grupos de repetición.

Ejemplo:

*El campo “**Género**” tiene valores separados por comas (**multivaluado**).*

- Podría intentarse añadir tantas columnas “Género” como tenga la película que a más géneros pertenezca. **ERROR.**
- También podríamos repetir el registro, poniendo en cada caso uno de los géneros. **ERROR.**

El primer caso nos llevaría a tener que modificar la tabla cada vez que apareciese una película que superara en nº de géneros a la que más tenga, además, generaría gran cantidad de campos vacíos para el resto de las películas.

*En el segundo caso, llevaría a la aparición de **grupos de repetición**:*

PELÍCULAS 

Título	Director	Nac_direc	Formato	Precio	Género	Duración	ID_Prod
El padrino	Francis Ford	EEUU	HD	15	Drama	175	1
El padrino	Francis Ford	EEUU	HD	15	Suspense	175	1
El padrino	Francis Ford	EEUU	3D	24	Drama	175	1
El padrino	Francis Ford	EEUU	3D	24	Suspense	175	1
Toy Story	John Lasseter	EEUU	3D	12	Animación	81	2
Toy Story	John Lasseter	EEUU	3D	12	Aventura	81	2



La solución pasa por dividir la tabla, separando los grupos de repetición.

PELÍCULAS

Título	Director	Nac_direc	Formato	Precio	Duración	ID_Prod
El padrino	Francis Ford	EEUU	HD	15	175	1
El padrino	Francis Ford	EEUU	3D	24	175	1
Toy Story	John Lasseter	EEUU	3D	12	81	2

GENEROS

ID_Género	Género
1	Drama
2	Suspense
3	Animación
4	Aventuras

PELÍCULAS – GÉNEROS

Título	ID_Género
El padrino	1
El padrino	2
Toy Story	3
Toy Story	4

Nótese que hemos creado 2 tablas:

- La tabla **GÉNEROS**, para solucionar el problema de grupos de repetición.
- La tabla **PELÍCULAS – GÉNEROS** porque se trata de una relación **N:M** (*varias películas pueden pertenecer al mismo género y una película puede tener varios géneros [R1]*), y por tanto debe generarse una tabla intermedia.

Ahora ya, vemos que no existen campos multivaluados ni grupos de repetición y podemos decir que la tabla se encuentra en 1ª Forma Normal.

(Nota: Como el proceso de separar atributos multivaluados consiste siempre en crear tablas con cardinalidad 1:N, no es necesario crear una tabla intermedia).

SEGUNDA FORMA NORMAL

Una tabla está en 2ª Forma Normal si:

- Esta en 1ª Forma Normal.
- Todas las columnas de la tabla dependen de la misma clave primaria.

(Las tablas en 1FN cuya clave primaria es 1 solo campo, ya están en 2ª Forma Normal).

Debemos fijarnos en las tablas con clave primaria compuesta (más de una columna como clave primaria).

Se trata de separar de la tabla aquellas columnas que sólo dependan de una parte de la clave primaria.

Ejemplo:

- *La tabla **GÉNEROS** tiene clave primaria atómica, por lo que ya está en 2FN.*
- *La tabla **PELÍCULAS – GÉNEROS** no tiene más columnas además de las que pertenecen a la clave primaria, por tanto, no hay nada que modificar.*

Sin embargo:

- *La tabla **PELÍCULAS** tiene clave principal compuesta (título y formato).
Analizamos las dependencias:*

- **Título** → director, nac_direc, duración, id_prod
- **Formato** → ∅
- **Título + Formato** → precio



La solución pasa por separar la tabla según la dependencia de clave primaria:

PELÍCULAS

Título	Director	Nac_direct	Duración	ID_Prod
El padrino	Francis Ford	EEUU	175	1
Toy Story	John Lasseter	EEUU	81	2

PRECIOS

Título	Formato	Precio
El padrino	HD	15
El padrino	3D	24
Toy Story	3D	12

GÉNEROS

ID_Género	Género
1	Drama
2	Suspense
3	Animación
4	Aventuras

PELÍCULAS – GÉNEROS

Título	ID_Género
El padrino	1
El padrino	2
Toy Story	3
Toy Story	4

TERCERA FORMA NORMAL

Una tabla está en tercera forma normal si:

- Está en 2ª Forma Normal
- No existen dependencias transitivas de ninguna columna sobre otra, es decir, todas las dependencias son funcionales.

Una dependencia transitiva ocurre cuando una columna depende de otra que a su vez depende de la clave principal.

Es decir: Sean A, B y C columnas de una tabla.

$A \rightarrow B$: B depende de A	Por tanto, C depende transitivamente de A:
$B \rightarrow C$: C depende de B	$A \rightarrow B \rightarrow C$ $\quad \quad \quad \wedge$ $\quad \quad \quad $ $\quad \quad \quad $

Ejemplo:

- La tabla **PELÍCULAS** tiene una columna con dependencia transitiva:
 - Título \rightarrow Director \rightarrow Nac_direc

Debemos crear una tabla nueva, "**DIRECTORES**". Además, por la **R2**, sabemos que la cardinalidad será **N:1**, por lo que no será necesario crear una tabla intermedia.

Por tanto, separamos la tabla para que todas las columnas tengan dependencia funcional de la clave primaria:



PELÍCULAS

Título	Director	Duración	ID_Prod
El padrino	Francis Ford	175	1
Toy Story	John Lasseter	81	2

PRECIOS

Título	Formato	Precio
El padrino	HD	15
El padrino	3D	24
Toy Story	3D	12

GENEROS

ID_Género	Género
1	Drama
2	Suspense
3	Animación
4	Aventuras

DIRECTORES

Director	Nac_direct
Francis Ford	EEUU
John Lasseter	EEUU

PELÍCULAS – GÉNEROS

Título	ID_Género
El padrino	1
El padrino	2
Toy Story	3
Toy Story	4

Comprobamos que todas las columnas tienen dependencias funcionales de sus claves primarias, lo que significa que están en 3FN.

FUSIONAR TABLAS

Si a lo largo del proceso de normalización, se obtiene tablas que comparten la misma clave primaria, estas tablas son susceptibles de ser fusionadas en una.