

¿Qué es la normalización de bases de datos?

En bases de datos relacionales, la normalización es un concepto que se aplica al momento de definir o estructurar una base de datos con el fin de:

- Reducir redundancia (duplicados) en los datos
- Aumentar la velocidad de las transacciones
- Mejorar la integridad de los datos

Cuenta con 6 reglas llamadas "Formas Normales" pero las más utilizadas y esenciales son las 3 primeras:

1NF:

La Primer Forma Normal (1st Normal Form) nos indica los siguientes parámetros:

- Cada celda debe contener un solo valor: Cada dato almacenado en una celda debe ser lo más atómico posible
- Cada fila es única: No puede haber dos filas en las que se repitan todos sus valores

La 1NF nos sugiere que, aunque se repitan algunos otros datos entre las filas, separemos (atomicemos) los datos que pudieran estar juntos en una celda.

Veamos un ejemplo de una tabla no normalizada:

Proveedores							
Doc	Nombre	Apellido	Calle	Num	Telefono	Email	Profesion
123456	Francisco	Duarte	Boyaca	515	301-304-1234	fduarte@email.com	Pintor
234567	Melisa	Gutierrez	Irigoyen	535	432-405-2345 432-505-4789	mgutierrez@email.com	Contadora
345678	Javier	Palacio	Tortugeros	127	589-506-7894	jpalacio@email.com jpalacio_personal@email.com	Pintor

Ilustración 1 Tabla no normalizada

Como se aprecia, las columnas de Teléfono y de Email contienen celdas con más de un dato (no atómicos).

Ahora veamos la tabla normalizada con la 1NF:

Proveedores							
Doc	Nombre	Apellido	Calle	Num	Telefono	Email	Profesion
123456	Francisco	Duarte	Boyaca	515	301-304-1234	fduarte@email.com	Pintor
234567	Melisa	Gutierrez	Irigoyen	535	432-405-2345	mgutierrez@email.com	Contadora
234567	Melisa	Gutierrez	Irigoyen	535	432-505-4789	mgutierrez@email.com	Contadora
345678	Javier	Palacio	Tortugeros	127	589-506-7894	jpalacio@email.com	Pintor
345678	Javier	Palacio	Tortugeros	127	589-506-7894	jpalacio_personal@email.com	Pintor

Ilustración 2 1NF aplicada

Como se ve, se duplicó la fila de cada persona que tenía más de un teléfono o email con el propósito que dichos datos estén en sus respectivas celdas.

2NF:

La Segunda Forma Normal nos indica los siguientes puntos:

- La base de datos debe estar en 1NF
- Los valores que se puedan repetir en más de una celda en una misma columna y que estén ligados al Primary Key, deben ser llevados a sus propias tablas.

La 2NF nos indica que cualquier dato que se pudiera repetir debido a la 1NF y pueda "ligarse" al Primary Key sea separado a su propia tabla.

Vamos a ver un ejemplo usando la tabla anterior:

Proveedores				Direcciones		
Doc	Nombre	Apellido	Profesion	Doc	Calle	Num
123456	Francisco	Duarte	Pintor	123456	Boyaca	515
234567	Melisa	Gutierrez	Contadora	234567	Irigoyen	535
345678	Javier	Palacio	Pintor	345678	Tortugeros	127

Ilustración 3 Tablas 1 y 2 obtenidas a partir de la 2NF

Telefonos		Emails	
Doc	Telefono	Doc	Email
123456	301-304-1234	123456	fduarte@email.com
234567	432-405-2345	234567	mgutierrez@email.com
234567	432-505-4789	345678	jpalacio@email.com
345678	589-506-7894	345678	jpalacio_personal@email.com

Ilustración 4 Tablas 3 y 4 obtenidas a partir de la 2NF

Como podemos apreciar, logramos dividir nuestra tabla en 4, uniendo la información de cada fila mediante un Foreign Key (número de documento) y el Primary Key (también el número de documento) de la tabla principal (Proveedores). Con esto cada proveedor tendrá una sola fila para sus datos primarios en una tabla, una sola fila para sus datos de direcciones en otra tabla y cuantas filas necesite para almacenar teléfonos y direcciones de email en sus respectivas tablas.

3NF:

La Tercer Forma Normal dice:

- La base de datos debe estar en 2NF
- Si los datos de una columna se repiten, pero no dependen de la clave primaria, sepárelos en su propia tabla y genere una clave para identificarlos

Esto significa que cualquier dato que no tenga una relación exclusiva con cada Primary Key pero que se puedan repetir, debemos separarlos y crear tablas dimensionales con ellos.

Continuemos con el ejemplo anterior para ver cómo aplicar la 3NF:

Proveedores			
Doc	Nombre	Apellido	Profesion_id
123456	Francisco	Duarte	10
234567	Melisa	Gutierrez	20
345678	Javier	Palacio	10

Direcciones		
Doc	Calle	Numero
123456	Boyaca	515
234567	Irigoyen	535
345678	Tortugueros	127

Ilustración 5 Aplicación de la 3NF: Separación del campo "profesión"

Telefonos	
Doc	Telefono
123456	301-304-1234
234567	432-405-2345
234567	432-505-4789
345678	589-506-7894

Emails	
Doc	Email
123456	fduarte@email.com
234567	mgutierrez@email.com
345678	jpalacio@email.com
345678	jpalacio_personal@email.com

Profesiones	
Profesion_id	Profesion
10	Pintor
20	Contadora

Ilustración 6 Aplicación de la 3NF: Tabla adicional para profesiones

Como vemos, en lugar de repetir las profesiones una y otra vez para distintos proveedores, tomamos ese "listado" de profesiones y las almacenamos una vez cada una en su propia tabla, generando en ella un Primary Key que nos permita agregarlo como Foreign Key en la tabla principal de los proveedores.

En el ejemplo utilicé un Primary Key (en la tabla Profesiones) no auto incremental para demostrar que podemos utilizar lo que nosotros deseemos, aunque se recomienda siempre que sean números enteros y por supuesto, que nunca se repitan.