

# Normalización

La normalización de bases de datos es una técnica para organizar el contenido de estas últimas. Sin la normalización, los sistemas de bases de datos pueden ser inexactos, lentos e ineficientes.

La teoría de normalización iniciada por Edgar Frank Codd a principios de la década de 1970 y continuada por otros autores (entre los que destacan Boyce y Fagin). Cada "nivel" de normalización se conoce como una forma, y hay 5 formas en total. La primera forma normal es el nivel de normalización más bajo, y así hasta la quinta forma normal, que es el nivel de normalización más alto.

Objetivos de la normalización:

- **Evitar la redundancia** de datos. Eliminar la repetición continua de datos, agiliza el trabajo de los usuarios, facilita la actualización de los datos, evita incoherencias entre los datos, reduce el espacio que ocupan las BBDD y en consecuencia hace que sean más manejables.
- **Evitar la ambigüedad** de datos. Datos que no clarifican suficientemente el registro al que representan o que pueden referirse a más de un registro.
- **Garantizar la integridad referencial**. Es decir, asegurarse que las relaciones entre datos son correctas

## Primera Forma Normal (1FN)

### Definición

Una relación está en primera forma normal (1FN) si, y solo si, ningún atributo de la relación es en sí mismo una relación, ni descomponible ni con multiplicidad de valores.

Los datos, pues, deben ser únicos y atómicos, es decir mínimos e indivisibles. Además, cada conjunto de datos relacionados debe ser único.

### Acciones para aplicarla

- Atomizar los datos, descomponiendo datos que contiene varios valores y desagrupando datos compuestos.
- Separar las columnas duplicadas de una tabla, en una nueva tabla.
- Separar las repeticiones de grupos de datos relacionados de una tabla, en una nueva tabla.
- Identificar cada tabla con una clave primaria.
- Establecer relaciones entre tablas usando claves foráneas.

### Ejemplos

Descomponer datos que contiene varios valores:

Por ejemplo, la siguiente tabla contiene varios teléfonos en el mismo registro.

#### Personas

id	nombre	direccion	telefono
025	Juan Pérez	Calle Bonete, 34, Montejo de la Sierra, Madrid, 28190	555-861-025
365	Andrea Gutiérrez	c/Sarmiento, 5 -21720 Rociana del Condado, Huelva	931596782

			619542396
385	Javier Barrientos	San Andrés, 24, 37290 Boada, Salamanca	774 794 965 625879568
472	Teresa Canales	C/ Amoladera, 78, Mejorada del Campo 28840 Madrid	686 54 79 51

Debe crearse una tabla independiente de teléfonos para cada valor con clave primaria y foranea que relaciona con la persona:

#### Telefonos

id_tel	id_per	telefono
1	025	555-861-025
2	365	931596782
3	365	619542396
4	385	774 794 965
5	385	625879568
6	472	686 54 79 51

#### Personas

id_per	nombre	direccion
025	Juan Pérez	Calle Bonete, 34, Montejo de la Sierra, Madrid, 28190
365	Andrea Gutiérrez	c/Sarmiento, 5 -21720 Rociana del Condado, Huelva
385	Javier Barrientos	San Andrés, 24, 37290 Boada, Salamanca
472	Teresa Canales	C/ Amoladera, 78, Mejorada del Campo 28840 Madrid

Ejemplo de desagrupar datos compuestos.

En la tabla de personas, el nombre completo debe dividirse en Nombre y apellido. A su vez la direccion debe dividirse en calle, nºmunicipio, provincia y codigo postal

id_per	nombre	apellido	calle	municipio	cp	provincia
025	Juan	Pérez	Bonete, 34	Montejo de la Sierra	28190	Madrid
365	Andrea	Gutiérrez	Sarmiento, 5	Rociana del Condado	21720	Huelva
385	Javier	Barrientos	San Andrés, 24	Boada	37290	Salamanca
472	Teresa	Canales	Amoladera, 78	Mejorada del Campo	28840	Madrid

## Segunda Forma Normal (2FN)

### Definición

Una relación está en segunda forma normal (2FN) si, y solo si, está en primera forma normal y todo atributo que no forma parte de una clave candidata depende completamente de todas las claves candidatas de la relación.

Esto significa que las columnas en la tabla deben contener solamente información sobre la entidad definida por la clave primaria y, por tanto, las columnas en la tabla deben contener datos acerca de una sola cosa.

### Acciones

- Eliminar cualquier columna que no dependa de la clave primaria de la tabla
- Crear nuevas tablas con esas columnas y la clave primaria de la cual dependen.
- Establecer relaciones entre tablas usando claves foráneas.

### Ejemplos

La siguiente tabla esta referida a alumnos, asignaturas y notas .

dni	asignatura	nombre	apellido	nota
12121219A	Lengua	Pedro	Valiente	9
12121219A	Matemáticas	Pedro	Valiente	8
3457775G	Lengua	Ana	Fernández	6
5674378J	Matemáticas	Sara	Crespo	7
5674378J	Lengua	Sara	Crespo	6

Deberia separarse en dos para evitar repeticiones: Alumnos.

dni	nombre	apellido
12121219A	Pedro	Valiente
3457775G	Ana	Fernández
5674378J	Sara	Crespo

y calificacion

dni	asignatura	nota
12121219A	Lengua	9
12121219A	Matemáticas	8
3457775G	Lengua	6
5674378J	Matemáticas	7
5674378J	Lengua	6

## Tercera Forma Normal (3FN)

### Definición

Una relación está en tercera forma normal (3FN) si, y solo si, está en segunda forma normal y ningún atributo que no forma parte de una clave depende de un conjunto de atributos que contiene alguno que no forma parte de una clave.

Se debe eliminar cualquier columna no clave que sea dependiente de otra columna no clave. En general, siempre que el contenido de un grupo de campos pueda aplicarse a más de un único registro de la tabla, considere colocar estos campos en una tabla independiente.

### Acciones

- Crear nuevas tablas con las columnas que son dependientes de otra columna no clave.

### Ejemplos

En la tabla de persona anterior, se puede extraer una tabla de poblaciones.

id_per	nombre	apellido	calle	municipio	cp	provincia
025	Juan	Pérez	Bonete, 34	Montejo de la Sierra	28190	Madrid
365	Andrea	Gutiérrez	Sarmiento, 5	Rociana del Condado	21720	Huelva
385	Javier	Barrientos	San Andrés, 24	Boada	37290	Salamanca
472	Teresa	Canales	Amoladera, 78	Mejorada del Campo	28840	Madrid

Quedando personas con la clave foranea del cp:

id_per	nombre	apellido	calle	cp
025	Juan	Pérez	Bonete, 34	28190
365	Andrea	Gutiérrez	Sarmiento, 5	21720
385	Javier	Barrientos	San Andrés, 24	37290
472	Teresa	Canales	Amoladera, 78	28840

Y poblaciones del siguiente modo

cp	municipio	provincia
28190	Montejo de la Sierra	Madrid
21720	Rociana del Condado	Huelva
37290	Boada	Salamanca
28840	Mejorada del Campo	Madrid

## **Excepción**

Cumplir la tercera forma normal, aunque en teoría es deseable, no siempre es práctico. Si tiene una tabla Clientes y desea eliminar todas las dependencias posibles entre los campos, debe crear tablas independientes para las ciudades, códigos postales, representantes de venta, clases de clientes y cualquier otro factor que pueda estar duplicado en varios registros. En teoría, la normalización merece el trabajo que supone. Sin embargo, muchas tablas pequeñas pueden degradar el rendimiento o superar la capacidad de memoria o de archivos abiertos.

Puede ser más factible aplicar la tercera forma normal sólo a los datos que cambian con frecuencia. Si quedan algunos campos dependientes, diseñe la aplicación para que pida al usuario que compruebe todos los campos relacionados cuando cambie alguno.