

NORMALIZACIÓN DE BASES DE DATOS

Jesús Reyes Carvajal

¿Qué es la normalización?

La normalización es el proceso mediante el cual se transforman datos complejos a un conjunto de estructuras de datos más pequeñas, que además de ser más simples y más estables, son más fáciles de mantener.

- La normalización hace las cosas fáciles de entender.
- Una BD normalizada ocupa menos espacio en disco.
- Elimina redundancia de datos.
- Elimina errores lógicos.

Niveles de normalización

Primera Forma Normal (1NF)

Incluye la eliminación de todos los grupos repetidos.

Segunda Forma Normal (2NF)

Asegura que todas las columnas que no son llave, sean completamente dependientes de la llave primaria (PK).

Tercera Forma Normal (3NF)

Elimina cualquier dependencia transitiva. Una dependencia transitiva es aquella en la cual las columnas que no son llave son dependientes de otras columnas que tampoco son llave.

Existen otras formas normales como la 4FN, Boyce Codd y 5FN, que no se considera explicar ya que por lo general siempre se llega hasta la 3FN.

Conceptos de dependencia

Dependencia Funcional: Es una conexión entre uno o más atributos

DNI \rightarrow Nombre y apellidos

Dependencia Funcional Reflexiva: Si “Y” está incluido en “X” entonces $X \rightarrow Y$

Si dirección y nombre están incluidos en DNI entonces con el DNI se puede recuperar la dirección y el nombre

Dependencia Funcional Transitiva:

Si $X \rightarrow Y \rightarrow Z$ entonces $X \rightarrow Z$

Fecha de nacimiento \rightarrow Edad

Edad \rightarrow Conducir

Fecha de nacimiento \rightarrow Edad \rightarrow Conducir

Ejercicio

La siguiente tabla muestra la información de una tienda de artículos deportivos.

ordenes(id_orden, fecha, id_cliente, direccion, num_articulo, n_articulo, cant, precio)

id_orden	fecha	id_cliente	n_cliente	direccion	num_articulo	n_articulo	cant	precio
2301	2019-02-23	101	Martín	Av Cali	3786	Red	3	35000
2301	2019-02-23	101	Martín	Av Cali	4011	Raqueta	6	65000
2301	2019-02-23	101	Martín	Av Cali	9132	Tenis	8	160000
2302	2019-02-25	107	Hernando	Pasadena	5794	Pelotas	4	45000
2303	2019-02-27	110	Carolina	Centro	4011	Raqueta	2	65000
2303	2019-02-27	110	Carolina	Centro	3141	Funda	2	15000

PRIMERA FORMA NORMAL 1FN

Al examinar esta tabla vemos que hay datos repetidos en **num_articulo**, **n_articulo**, **cant** y **precio**, la 1FN prohíbe grupos repetidos, por lo tanto debemos hacer cumplir las reglas de la 1FN.

- Eliminar grupos repetidos
- Crear nuevas tablas con llaves primarias de la base y del grupo repetido

La solución es crear dos tablas

ordenes(id_orden, fecha, id_cliente, n_cliente, direccion)

id_orden	fecha	id_cliente	n_cliente	direccion
2301	2019-02-23	101	Martín	Av Cali
2302	2019-02-25	107	Hernando	Pasadena
2303	2019-02-27	110	Carolina	Centro

articulo_ordenes(id_orden, num_articulo, n_articulo, cant, precio)

id_orden	num_articulo	n_articulo	cant	precio
2301	3786	Red	3	35000
2301	4011	Raqueta	6	65000
2301	9132	Tenis	8	160000
2302	5794	Pelotas	4	45000
2303	4011	Raqueta	2	65000
2303	3141	Funda	2	15000

SEGUNDA FORMA NORMAL 2FN

Una tabla está en 2FN, si está en 1FN y todos los atributos no clave dependen completamente de la llave primaria.

Los pasos a seguir son:

- Determinar las columnas que no son llave primaria y verificar si dependen completamente de la llave primaria.
- Eliminar esas columnas de tabla base
- Crear una tabla con esas columnas y traer la llave primaria de la cual dependen.

La tabla ordenes esta en 1FN y ademas todas las columnas no clave dependen completamente de la llave primaria **id_orden**.

La tabla articulos_ordenes no se encuentra en 2FN, ya que **precio** y **n_articulo** no dependen de **id_orden**, pero si dependen de **num_articulo**, por tanto lo que hacemos es eliminar esas columnas de articulos_ordenes y crear una tabla llamada articulos con dichas columnas y la llave de la cual dependen.

articulos_ordenes(id_orden,num_articulo,cant)

id_orden	num_articulo	cant
2301	3786	3
2301	4011	6
2301	9132	8
2302	5794	4
2303	4011	2
2303	3141	2

articulos(num_articulo,n_articulo,precio)

num_articulo	n_articulo	precio
3786	Red	35000
4011	Raqueta	65000
9132	Tenis	160000
5794	Pelotas	45000
3141	Funda	15000

TERCERA FORMA NORMAL 3FN

Una tabla está en 3FN, si está en 2FN y no existe ninguna dependencia funcional transitiva entre los atributos que no son clave.

Los pasos a seguir son:

Determinar las columnas que son dependientes de otras columnas no clave.

Eliminar esas columnas de tabla base

Crear una tabla con esas columnas y con la columna no clave de la cual son dependientes.

Las tablas **articulos_ordenes** y **articulos** cumplen con la 3FN, pero la tabla **ordenes** no cumple con la 3FN, ya que las columnas **n_cliente** y **direccion** dependen de **id_cliente** y el **id_cliente** depende de **id_orden**. Por tanto existe una dependencia funcional transitoria.

$\text{id_orden} \rightarrow \text{id_cliente} \rightarrow \text{n_cliente}$

$\text{id_orden} \rightarrow \text{id_cliente} \rightarrow \text{direccion}$

Para normalizar esta tabla, moveremos las columnas no clave y la columna clave de la cual dependen dentro de una nueva tabla llamada clientes.

ordenes(id_orden, fecha, id_cliente)

id_orden	fecha	id_cliente
2301	2019-02-23	101
2302	2019-02-25	107
2303	2019-02-27	110

clientes(id_cliente, n_cliente, direccion)

id_cliente	n_cliente	direccion
101	Martín	Av Cali
107	Hernando	Pasadena
110	Carolina	Centro

SOLUCIÓN FINAL

articulos_ordenes(id_orden,num_articulo,cant)

id_orden	num_articulo	cant
2301	3786	3
2301	4011	6
2301	9132	8
2302	5794	4
2303	4011	2
2303	3141	2

articulos(num_articulo,n_articulo,precio)

num_articulo	n_articulo	precio
3786	Red	35000
4011	Raqueta	65000
9132	Tenis	160000
5794	Pelotas	45000
3141	Funda	15000

ordenes(id_orden,id_cliente,fecha)

id_orden	fecha	id_cliente
2301	2019-02-23	101
2302	2019-02-25	107
2303	2019-02-27	110

clientes(id_cliente,n_cliente,direccion)

id_cliente	n_cliente	direccion
101	Martín	Av Cali
107	Hernando	Pasadena
110	Carolina	Centro

OTRO EJEMPLO

Describa el proceso de normalización de la siguiente tabla de datos.

<i>CodLibro</i>	<i>Titulo</i>	<i>Autor</i>	<i>editorial</i>	<i>NombreLector</i>	<i>echaDev</i>
1001	Variable Compleja	Murray Spiegel	McGraw Hill	Pérez Gómez, Juan	15/04/2014
1004	Visual Basic	E. Petroustsos	Anaya	Ríos Terán, Ana	17/04/2014
1005	Estadística	Murray Spiegel	McGraw Hill	Roca, René	16/04/2014
1006	Oracle University	Nancy Greenberg y Priya Nathan	Oracle Corp.	García Roque, Luis	20/04/2014
1007	Clipper 5.01	Ramalho	McGraw Hill	Pérez Gómez, Juan	18/04/2014

1FN

Esta tabla no cumple con la Primera Forma Normal (1NF) tiene campos atómicos, el nombre del lector es un campo que puede (y conviene) descomponerse en apellido paterno, apellido materno y nombres.

<i>CodLibro</i>	<i>Título</i>	<i>Autor</i>	<i>editorial</i>	<i>Paterno</i>	<i>Materno</i>	<i>Nombres</i>	<i>FechaDev</i>
1001	Variable compleja	Murray Spiegel	McGraw Hill	Pérez	Gómez	Juan	15/04/2014
1004	Visual Basic 5	E. Petroustsos	Anaya	Ríos	Terán	Ana	17/04/2014
1005	Estadística	Murray Spiegel	McGraw Hill	Roca		René	16/04/2014
1006	Oracle University	NancyGreenberg	Oracle Corp.	García	Roque	Luis	20/04/2014
1006	Oracle University	Priya Nathan	Oracle Corp.	García	Roque	Luis	20/04/2014
1007	Clipper 5.01	Ramalho	McGraw Hill	Pérez	Gómez	Juan	18/04/2014

2FN

Actualmente en tabla tenemos varias dependencias parciales si consideramos como atributo clave el código del libro. Por ejemplo, el título es completamente identificado por el código del libro, pero el nombre del lector en realidad no tiene dependencia con este código, por tanto estos datos deben ser trasladados a otra tabla.

<i>CodLibro</i>	<i>Título</i>	<i>Autor</i>	<i>Editorial</i>
1001	Variable compleja	Murray Spiegel	McGraw Hil
1004	Visual Basic 5	E. Petroustsos	Anaya
1005	Estadística	Murray Spiegel	McGraw Hill
1006	Oracle University	NancyGreenberg	Oracle Corp
1006	Oracle University	Priya Nathan	Oracle Corp.
1007	Clipper 5.01	Ramalho	McGraw Hill

<i>CodLector</i>	<i>Paterno</i>	<i>Materno</i>	<i>Nombres</i>
501	Pérez	Gómez	Juan
502	Ríos	Terán	Ana
503	Roca		René
504	García	Roque	Luis

<i>CodLibro</i>	<i>CodLector</i>	<i>FechaDev</i>
1001	501	15/04/2014
1004	502	17/04/2014
1005	503	16/04/2014
1006	504	20/04/2014
1007	501	18/04/2014

FechaDev depende de CodLibro y CodLector

3FN

La primera tabla de la 2FN conserva información acerca del libro, pero los campos autores y editoriales, contemplan dependencia transitiva, ya que estos dependen del título de libro, por tanto debemos crear nuevas tablas para satisfacer los requisitos de 3NF.

<i>CodEditorial</i>	<i>Editorial</i>
901	McGraw Hill
902	Anaya
903	Oracle Corp

<i>CodLibro</i>	<i>codEditorial</i>
1001	901
1004	902
1005	901
1006	903
1007	901

<i>CodLibro</i>	<i>Título</i>
1001	Variable compleja
1004	Visual Basic 5
1005	Estadística
1006	Oracle University

<i>CodLibro</i>	<i>codAutor</i>
1001	801
1004	802
1005	801
1006	803
1006	804
1007	806

<i>CodLector</i>	<i>Paterno</i>	<i>Materno</i>	<i>Nombres</i>
501	Pérez	Gómez	Juan
502	Ríos	Terán	Ana
503	Roca		René
504	García	Roque	Luis

<i>CodLibro</i>	<i>CodLector</i>	<i>FechaDev</i>
1001	501	15/04/2014
1004	502	17/04/2014
1005	503	16/04/2014
1006	504	20/04/2014
1007	501	18/04/2014

<i>CodAutor</i>	<i>Autor</i>
801	Murray Spiegel
802	E. Petroustsos
803	Nancy Greenberg
804	Priya Nathan
806	Ramalho