

NORMALIZACIÓN

Servando Campillay

INTRODUCCIÓN

- Los seres humanos tenemos la tendencia de simplificar las cosas al máximo. Lo hacemos con casi todo, desde los animales hasta con los automóviles. Vemos una imagen de gran tamaño y la hacemos más simple agrupando cosas similares juntas. Las guías que la normalización provee crean el marco de referencia para simplificar una estructura de datos compleja.



¿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.
- También se puede entender la normalización como una serie de reglas que sirven para ayudar a los diseñadores de bases de datos a desarrollar un esquema que minimice los problemas de lógica.



GRADOS DE NORMALIZACIÓN

- Existen básicamente tres niveles de normalización:
 - Primera Forma Normal (1NF), Segunda Forma Normal (2NF) y Tercera Forma Normal (3NF).
 - Cada una de estas formas tiene sus propias reglas.
 - Cuando una base de datos se conforma a un nivel, se considera normalizada a esa forma de normalización.



GRADO DE NORMALIZACIÓN

- No siempre es una buena idea tener una base de datos conformada en el nivel más alto de normalización, puede llevar a un nivel de complejidad que pudiera ser evitado si estuviera en un nivel más bajo de normalización.



GRADO DE NORMALIZACIÓN

- En la tabla siguiente se describe brevemente en que consiste cada una de las reglas, y posteriormente se explican con más detalle.

| Regla | Descripción |
|----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Primera Forma Normal (1FN) | Incluye la eliminación de todos los grupos repetidos. |
| Segunda Forma Normal (2FN) | Asegura que todas las columnas que no son llave sean completamente dependientes de la llave primaria (PK). |
| Tercera Forma Normal (3FN) | 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. |



1. Objetivo de Normalización: Obtener los datos agrupados en tablas distintas, mediante la aplicación de una serie de reglas, para conseguir que los datos capturados tengan una estructura óptima para su implementación, gestión y explotación desde distintas aplicaciones futuras.
2. Una de las ventajas principales que se obtiene al realizar la normalización es que la información no estará duplicada innecesariamente dentro de las estructuras: habrá mínima redundancia.

Modelo de datos

evitar

- Confusión ... redundancia de la información
- Duplicación de la información ... incapacidad de representar información
- Mal funcionamiento al explorar información
- Registro de información no identificable



PRIMERA FORMA NORMAL

- Se dice que una tabla está en primera forma normal si todos los valores que componen a sus tuplas son atómicos: un atributo no puede tener más de un valor.

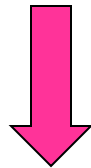
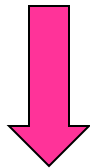
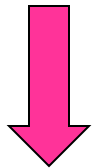


Para normalizar una tabla que no esté en 1FN han de seguirse los siguientes pasos:

1. Se localizan los atributos correspondientes a la clave principal
2. Se realiza una proyección sobre la tabla y así se descompone en varias, de manera que se hace la proyección de la clave con los atributos que tengan los valores únicos.



POR EJEMPLO, LA SIGUIENTE TABLA NO ESTÁ EN 1FN:



| IdPedido | IdCliente | IdVendedor | Fecha_Pedido | Fecha_Entrega | Productos |
|----------|-----------|------------|--------------|---------------|-------------------------------|
| 10548 | Wilman | 5 | 25/02/02 | 26/02/02 | 3 chorizos, 2 ajos, 1 cebolla |
| 10549 | Toldar | 6 | 01/04/02 | 05/04/02 | 4 botes, 2 fresas |
| 10550 | Canair | 3 | 05/06/02 | 06/06/02 | 1 revista, 2 zumos, 1 pan |
| 10551 | Suprem | 5 | 07/07/02 | 07/07/02 | 2 habas, 1 pan, 2 quesos |
| 10552 | Wilman | 1 | 08/07/02 | 09/07/02 | 1 bote, 2 panes |



CONCEPTO DE DEPENDENCIA FUNCIONAL

- Se dice que un atributo B depende funcionalmente de A ($A \rightarrow B$) si cada valor de A se corresponde con un único valor de B o, visto de otra manera, si dado A puedo obtener B . Ejemplo: $RUT \rightarrow Nombre$, pues dado un RUT puedo obtener el nombre de la persona con ese RUT .



SEGUNDA FORMA NORMAL

- Esta forma normal se considerará únicamente cuando la clave principal sea compuesta, si no (la clave principal está formada por un único atributo) la tabla estaría en segunda forma normal.



Decimos que una tabla está en segunda forma normal si se cumplen las siguientes condiciones:

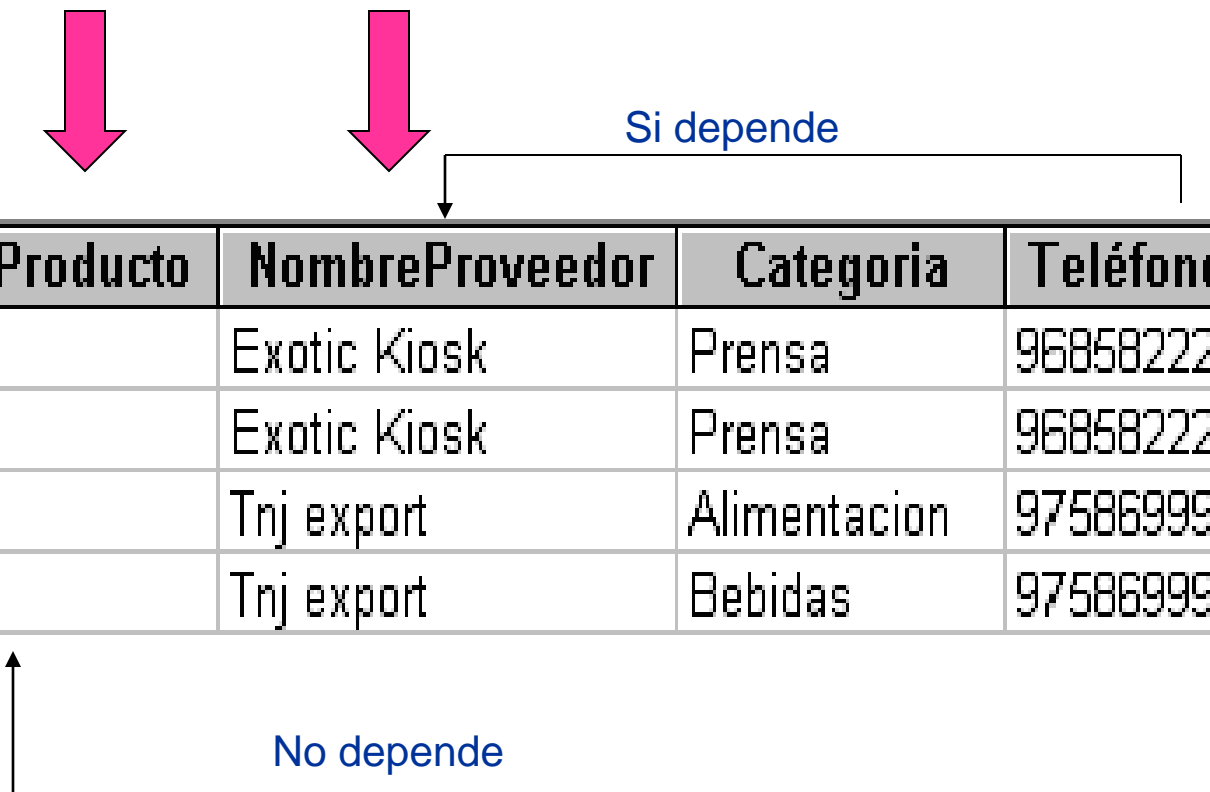
1. Está en 1FN
2. Todo atributo secundario (los que no pertenecen a la clave principal) tiene una dependencia funcional total de la clave completa y no de una parte de ella.



- Para convertir una tabla que no esté en 2FN a 2FN se creará una tabla con la clave y todas sus dependencias funcionales totales y otra tabla con la parte de la clave que tiene dependencias con los atributos secundarios.



POR EJEMPLO, LA SIGUIENTE TABLA NO ESTÁ EN 2FN:



| NombreProducto | NombreProveedor | Categoria | TeléfonoProveedor |
|----------------|-----------------|--------------|-------------------|
| Diarios | Exotic Kiosk | Prensa | 968582222 |
| Revistas | Exotic Kiosk | Prensa | 968582222 |
| Habas | Tnj export | Alimentacion | 975869999 |
| Botes | Tnj export | Bebidas | 975869999 |



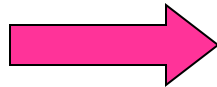
- el campo "TelefonoProveedor" no es dependiente de la clave candidata {"NombreProducto, NombreProveedor"} sino únicamente de NombreProveedor".

Se trata de no representar dos entidades distintas en una sola tabla.



EN ESTE EJEMPLO, REORGANIZARÍAMOS LOS DATOS DE LA SIGUIENTE MANERA

Tabla Producto



| IdProducto | NombreProduct | Categoría | IdProveedor |
|------------|---------------|--------------|-------------|
| 1 | Diarios | Prensa | 1 |
| 2 | Revistas | Prensa | 1 |
| 3 | Habas | Alimentación | 2 |
| 4 | Botes | Alimentación | 2 |

Tabla Proveedor



| IdProveedor | NombreProveedor | TeléfonoProveedor |
|-------------|-----------------|-------------------|
| 1 | Exotic Kiosk | 968582222 |
| 2 | Tnj Export | 975869999 |



TERCERA FORMA NORMAL

Una tabla está en 3FN si:

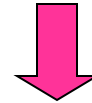
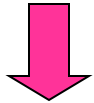
1. Está en 2FN
2. No existen atributos no primarios (no pertenecen a la clave) que son transitivamente dependientes de cada posible clave de la tabla, o lo que es lo mismo, un atributo secundario sólo puede ser conocido a través de la clave principal o claves secundarias de la tabla y no por medio de otro atributo no primario.



- Para convertir una tabla que no esté en 3FN a 3FN se realizará una proyección de la clave a los elementos que no tengan dependencia funcional transitiva y otra tabla con una nueva clave a los elementos que anteriormente tenían esta dependencia.



POR EJEMPLO, LA SIGUIENTE TABLA NO ESTÁ EN 3FN:



| Licencia | Nombre | Club | Edad | Categoría |
|----------|--------|---------|------|-----------|
| 189585 | Pepe | Alcoy | 12 | Junior |
| 205888 | Tomas | Jaca | 10 | Cadete |
| 748523 | Andres | Almansa | 9 | Alevín |

ya que, dado un número de licencia, podemos obtener la edad del inscrito, y dada la edad del inscrito, podemos averiguar la categoría a la que pertenece: tenemos una dependencia funcional transitiva. Evidentemente, dado el número de licencia podemos averiguar la categoría pero lo importante aquí es que la categoría depende de un atributo que no forma parte de la clave.



PARA NORMALIZAR, DESCOMPONDREMOS LA TABLA EN LAS SIGUIENTES:

Tabla Atleta

| Licencia | Nombre | Club | Edad |
|----------|--------|---------|------|
| 189585 | Pepe | Alcoy | 12 |
| 205888 | Tomas | Jaca | 10 |
| 748523 | Andres | Almansa | 9 |

Tabla Categoría

| Edad | Categoría |
|------|-----------|
| 9 | Alevín |
| 10 | Cadete |
| 12 | Junior |



CASO

1. Una ISAPRE mantiene la siguiente tabla con sus afiliados: AFILIADO (RUT, NOM, DIR, RUT-EMPLEADOR, CODIGO-PLAN). Dada cada una de las siguientes semánticas, indicar clave primaria de la tabla y en qué FN está:
 - Un Afiliado tiene un solo empleador y un solo plan
 - Un Afiliado tiene varios empleadores y un solo plan
 - Un Afiliado tiene varios empleadores y varios planes
 - ¿Cómo sería la tabla de Afiliados en 3FN, si el usuario fuera la superintendencia de ISAPRES?



CASO

2. Un ejecutivo de Telefónica, que dice saber de modelamiento de datos, considera que la siguiente tabla no tiene problemas de anomalías de modificación:
 - LLAMADAS (FONO-LLAMADA, FONO-DONDE-LLAMÓ, FECHA-HORA-LLAMADA, NOMBRE-CLIENTE, DIR-CLIENTE, DURACION-LLAMADA, TIPO-HORARIO-LLAMADA, PRECIO-TIPO-HORARIO) ¿Qué opinas UD?



○ Ejercicio 1

1.- Las líneas aéreas, manejan los datos de todos sus vuelos, entre estos están el número del vuelo, origen y destino. El pasaje tiene un código de identificación y el valor, que también son almacenados. Además se guardan los datos de los pasajeros, considerando su nacionalidad y datos personales. Por otro lado se lleva un control estricto del equipaje de cada pasajero, ya que es importante que nada se extravíe durante el viaje.



