

NORMALIZACION

Cuando aparecen los problemas enumerados anteriormente entonces se los puede resolver usando reglas de normalización. Estas reglas suelen forzar la división de una tabla en dos o más tablas para arreglar ese problema.

FORMAS NORMALES

En la teoría de bases de datos relacionales, las formas normales (NF) proporcionan los criterios para determinar el grado de vulnerabilidad de una tabla a inconsistencias y anomalías lógicas. Cuánto más alta sea la forma normal aplicable a una tabla, menos vulnerable será a inconsistencias y anomalías. Cada tabla tiene una "forma normal más alta" (HNF): por definición, una tabla siempre satisface los requisitos de su HNF y de todas las formas normales más bajas que su HNF; también por definición, una tabla no puede satisfacer los requisitos de ninguna forma normal más arriba que su HNF.

Las formas normales son aplicables a tablas individuales; es decir que si una base de datos entera está en la forma normal n es decir que todas sus tablas están en la forma normal n . Los recién llegados al diseño de bases de datos a veces suponen que la normalización procede de una manera iterativa, es decir un diseño 1NF primero se normaliza a 2NF, entonces a 3NF, etcétera. Ésta no es una descripción exacta de cómo la normalización trabaja típicamente. Una tabla sensiblemente diseñada es probable que esté en 3NF en la primera tentativa; además, si está en 3NF, también es extremadamente probable que tenga una forma HNF de 5NF. Conseguir formas normales "más altas" (sobre 3NF) usualmente no requiere un gasto adicional de esfuerzo por parte del diseñador, porque las tablas 3NF usualmente no necesitan ninguna modificación para satisfacer los requisitos de estas formas normales más altas. Edgar F. Codd originalmente definió las tres primeras formas normales (1NF, 2NF, y 3NF). Estas formas normales se han resumido como requiriendo que todos los atributos no-clave sean dependientes en "la clave, la clave completa, y nada excepto la clave". Las cuarta y quinta formas normales (4NF y 5NF) se ocupan específicamente de la representación de las relaciones muchos a muchos y uno muchos entre los atributos. La sexta forma normal (6NF), en pocas palabras, se basa en el principio de que si se tiene más de dos claves candidatas en una tabla, se tendrán que crear otras tablas con estas.

Se debe tener en cuenta que muchos diseñadores opinan que basta con llegar a la forma Boyce-Codd, ya que la cuarta, quinta y sexta forma normal es polémica.

PRIMERA FORMA NORMAL (1FN)

Una tabla está en 1FN si y solo si es "isomorfa a alguna relación", lo que significa, específicamente, que satisface las siguientes cinco condiciones:

1. No hay orden de arriba-a-abajo en las filas.
2. No hay orden de izquierda-a-derecha en las columnas.
3. No hay filas duplicadas.
4. Cada intersección de fila-y-columna contiene exactamente un valor del dominio aplicable (Atomicidad).
5. Todas las columnas son regulares.

SEGUNDA FORMA NORMAL (2FN)

Una tabla que está en la primera forma normal (1NF) debe satisfacer criterios adicionales para calificar para la segunda forma normal. Específicamente: una tabla 1NF está en 2NF si y solo si, dada una clave primaria y cualquier atributo que no sea un constituyente de la clave primaria, el atributo no clave depende de toda la clave primaria en vez de solo de una parte de ella. En términos levemente más formales: una tabla 1NF está en 2NF si y solo si ninguno de sus atributos no-principales son funcionalmente dependientes en una parte (subconjunto propio) de una clave primaria (Un atributo no-principal es uno que no pertenece a ninguna clave primaria). Observe que cuando una tabla 1NF no tiene ninguna clave candidata compuesta (claves candidatas consistiendo en más de un atributo), la tabla está automáticamente en 2NF.

TERCERA FORMA NORMAL (3FN)

La definición de Codd indica que una tabla está en 3NF si y solo si las dos condiciones siguientes se cumplen:

1. La tabla está en la segunda forma normal (2NF)
2. Ningún atributo no-primario de la tabla es dependiente transitivamente de una clave primaria.
3. Es una relación que no incluye ningún atributo clave

Una dependencia transitiva es una dependencia funcional $X \rightarrow Z$ en la cual Z no es inmediatamente dependiente de X , pero si de un tercer conjunto de atributos Y , que a su vez dependen de X . Es decir, $X \rightarrow Z$ por virtud de $X \rightarrow Y$ e $Y \rightarrow Z$.

Torneo	Año	Ganador	NacimientoGanador
Indiana	2015	Juan Perez	12/05/1990
Cleveland Open	2016	Pepe Grillo	05/12/1980
Des Moines	2016	Juan Perez	12/05/1990
Indiana	2016	Rosa Adams	16/10/2000

En el ejemplo anterior la violación de la 3NF ocurre porque el atributo no primario *NacimientoGanador* es dependiente transitivamente de {Torneo, Año} vía el atributo no primario *Ganador*. El hecho de que el campo *NacimientoGanador* es funcionalmente dependiente a *Ganador* hace la tabla vulnerable a inconsistencias lógicas, pues no hay nada que impida a la misma persona ser mostrada con diferentes fechas de nacimiento en diversos registros.

FORMA NORMAL DE BOYCE-CODD

La Forma Normal de Boyce-Codd (o FNBC) es una forma normal utilizada en la normalización de bases de datos. Es una versión ligeramente más fuerte de la Tercera forma normal (3FN). La forma normal de Boyce-Codd requiere que no existan dependencias funcionales no triviales de los atributos que no sean un conjunto de la clave candidata. En una tabla en 3FN, todos los atributos dependen de una clave, de la clave completa y de ninguna otra cosa excepto de la clave (excluyendo dependencias triviales, como $A \rightarrow A$). Se dice que una tabla está en FNBC si y solo si está en 3FN y cada dependencia funcional no trivial tiene una clave candidata como determinante. En términos

menos formales, una tabla está en FNBC si está en 3FN y los únicos determinantes son claves candidatas.

NUM_EMPLEADO	DNI	NOMBRE	SEG_SOC	DPTO	PUESTO	SALARIO
10	41325B	Juan Ramos	35-1234566	COMPRAS	Gerente	2300
17	23456J	Pedro Pérez	35-1234567	NÓMINAS	Auxiliar	1200
2	34561A	María Rodríguez	35-5540114	ALMACÉN	Conserje	1530
3	222346S	Juana Cabello	35-1147895	COMPRAS	Gestión	2200

Tabla 1 Tabla que no cumple FNBC

En el ejemplo anterior vemos que existen tres claves candidatas y cada campo puede depender de cualquiera de ellas y no de la clave primaria por lo que no se encuentra en FNBC. Como solución se debería dividir los datos en dos tablas T1{num_empleado, DNI, Seg_social} y T2{num_empleado, nombre, Dpto puesto, salario}.

*Note que si la clave primaria tiene un solo atributo y está en 3FN, ya está en FNBC.

PONIENDO EN PRACTICA LA NORMALIZACION HASTA LA 3FN

Para comprender mejor cada una de las formas normales trabajaremos con la siguiente tabla como ejemplo:

idLibro	Titulo	Autor	Editorial	Lector	FechaDevolucion
1001	Variable compleja	Murray Spiegel	McGraw Hill	Pérez Gómez, Juan	15/04/2005
1004	Visual Basic 5	E. Petroustsos	Anaya	Ríos Terán, Ana	17/04/2005
1005	Estadística	Murray Spiegel	McGraw Hill	Roca, René	16/04/2005
1006	Oracle University	Nancy Greenberg y Priya Nathan	Oracle Corp.	García Roque, Luis	20/04/2005
1007	Clipper 5.01	Ramalho	McGraw Hill	Pérez Gómez, Juan	18/04/2005

1. PRIMERA FORMA NORMAL (1FN)

Se dice que una tabla se encuentra en primera forma normal si impide que un atributo de una tupla pueda tomar más de un valor.

idLibro	Titulo	Autor	Editorial	Lector	FechaDevolucion
1001	Variable compleja	Murray Spiegel	McGraw Hill	Pérez Gómez, Juan	15/04/2005
1004	Visual Basic 5	E. Petroustsos	Anaya	Ríos Terán, Ana	17/04/2005
1005	Estadística	Murray Spiegel	McGraw Hill	Roca, René	16/04/2005
1006	Oracle University	Nancy Greenberg y Priya Nathan	Oracle Corp.	García Roque, Luis	20/04/2005
1007	Clipper 5.01	Ramalho	McGraw Hill	Pérez Gómez, Juan	18/04/2005

Tabla 2. Tabla que no cumple con la 1FN

Modificando la tabla para que cumpla la 1FN tenemos:

idLibro	Título	Autor	Editorial	Paterno	Materno	Nombre	FechaDevolucion
1001	Variable compleja	Murray Spiegel	McGraw Hill	Pérez	Gómez	Juan	15/04/2005
1004	Visual Basic 5	E. Petroustsos	Anaya	Ríos	Terán	Ana	17/04/2005
1005	Estadística	Murray Spiegel	McGraw Hill	Roca		René	16/04/2005
1006	Oracle University	Nancy Greenberg	Oracle Corp.	García	Roque	Luis	20/04/2005
1006	Oracle University	Priya Nathan	Oracle Corp.	García	Roque	Luis	20/04/2005
1007	Clipper 5.01	Ramalho	McGraw Hill	Pérez	Gómez	Juan	18/04/2005

Tabla 3. Tabla que cumple la 1FN

2. SEGUNDA FORMA NORMAL (2FN)

Se dice que una tabla se encuentra en segunda forma normal si se encuentra en 1FN y si no existan dependencias parciales, o dicho de otra manera, todos los atributos no clave deben depender por completo de la clave primaria.

En nuestro ejemplo actualmente tenemos varias dependencias parciales si consideramos como atributo clave el código del libro.

idLibro	Título	Autor	Editorial	Paterno	Materno	Nombre	FechaDevolucion
1001	Variable compleja	Murray Spiegel	McGraw Hill	Pérez	Gómez	Juan	15/04/2005
1004	Visual Basic 5	E. Petroustsos	Anaya	Ríos	Terán	Ana	17/04/2005
1005	Estadística	Murray Spiegel	McGraw Hill	Roca		René	16/04/2005
1006	Oracle University	Nancy Greenberg	Oracle Corp.	García	Roque	Luis	20/04/2005
1006	Oracle University	Priya Nathan	Oracle Corp.	García	Roque	Luis	20/04/2005
1007	Clipper 5.01	Ramalho	McGraw Hill	Pérez	Gómez	Juan	18/04/2005

Tabla 4 Tabla que no cumple con la 2FN

Por ejemplo, el título es completamente identificado por el código del libro, pero el nombre del lector en realidad no tiene dependencia de este código, por tanto estos datos deben ser trasladados a otra tabla quedando de la siguiente manera.

idLibro	Título	Autor	Editorial
1001	Variable compleja	Murray Spiegel	McGraw Hill
1004	Visual Basic 5	E. Petroustsos	Anaya
1005	Estadística	Murray Spiegel	McGraw Hill
1006	Oracle University	Nancy Greenberg	Oracle Corp.
1006	Oracle University	Priya Nathan	Oracle Corp.
1007	Clipper 5.01	Ramalho	McGraw Hill

Tabla 5 Tabla Libro que cumple la 2FN

idLector	Paterno	Materno	Nombre
501	Pérez	Gómez	Juan
502	Ríos	Terán	Ana
503	Roca		René
504	García	Roque	Luis

Tabla 6 Tabla Lector que cumple la 2FN

Hemos creado una tabla para contener los datos del lector y también tuvimos que crear la columna idLector para identificar a cada uno. Sin embargo, esta nueva disposición de la base de datos necesita que exista otra tabla para mantener la información de qué libros están prestados a qué lectores. Esta tabla se muestra a continuación:

idLibro	idLector	FechaDevolucion
1001	501	15/04/2005
1004	502	17/04/2005
1005	503	16/04/2005
1006	504	20/04/2005
1007	501	18/04/2005

Tabla 7 Tabla Prestamo que cumple la 2FN

3. TERCERA FORMA NORMAL (3FN)

Para la Tercera Forma Normal la relación debe estar en 2NF y además los atributos no clave deben ser mutuamente independientes y dependientes por completo de la clave primaria. 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.

En nuestro ejemplo en 2NF, la primera tabla conserva información acerca del libro, los autores y editoriales, por lo que debemos crear nuevas tablas para satisfacer los requisitos de 3NF.

idLibro	Título
1001	Variable compleja
1004	Visual Basic 5
1005	Estadística
1006	Oracle University
1007	Clipper 5.01

Tabla 8 Tabla Libro

idAutor	Autor
801	Murray Spiegel
802	E. Petroustos
803	Murray Spiegel
804	Nancy Greenberg
805	Priya Nathan
806	Ramalho

Tabla 9 Tabla Autor

idEditorial	Editorial
901	McGraw Hill
902	Anaya
903	Oracle Corp.

Tabla 10 Tabla Editorial

Aunque hemos creado nuevas tablas para que cada una tenga sólo información acerca de una entidad, también hemos perdido la información acerca de qué autor ha escrito qué libro y las editoriales correspondientes, por lo que debemos crear otras tablas que relacionen cada libro con sus autores y editoriales.

idLibro	idAutor
1001	801
1004	802
1005	801
1006	803
1006	804
1007	806

Tabla 11 Tabla LibroAutor

idLibro	idEditorial
1001	901
1004	902
1005	901
1006	903
1007	901

Tabla 12 Tabla LibroEditorial