

NORMALIZACIÓN DE BASES DE DATOS.

Definiciones y Ejemplos del Proceso de Normalización.

En que consiste la normalización.

El proceso de normalización de bases de datos consiste en designar y aplicar una serie de reglas a las relaciones obtenidas tras el paso del modelo entidad-relación al modelo relacional.

Las bases de datos relacionales se normalizan para:

- ▣ Evitar la redundancia de los datos.
- ▣ Disminuir problemas de actualización de los datos en las tablas.
- ▣ Proteger la integridad de los datos.

Importante: Se han definido muchas formas normales, las tres primeras fueron definidas por Codd 1FN, 2FN y 3FN.

La definición original de Codd de la 3FN fue complementada por una versión más sólida debida a Boyce y Codd a la nueva 3FN se la conoce como **Forma Normal de Boyce/Codd (FNBC)**

Primera Forma Normal

Id Empleado	Nombres	Apellidos	Telefono	Documento
1	Eduardo Andrés	Gamarra Solis	0981234234 / 21-204523	1321234
2	Pedro Pablo	Leguizamon Duarte	0961900900 / 0981555666	4232123
3	Valeria Elizabeth	Ramirez Gonzalez	21-234090	2341512

Figura 1

- Una Relación esta en 1FN si y solo si, en cada valor válido de esa relación, toda tupla contiene exactamente un valor para cada atributo.
- En el ejemplo de la figura 1 se ve la relación de *Empleados*.
- Los atributos de *Nombres*, *Apellidos* y *Teléfono* contienen valores multivaluados, es decir pueden descomponerse en valores atómicos.

Primera Forma Normal

- Para lograr la primera forma normal se deben crear los atributos que permitan identificar en forma atómica a cada valor contenido en el mismo.
- Notar que para el atributo *Telefono* la estructura esta definida para manejar 2 (dos) números.

Id Empleado	Primer Nombre	Segundo Nombre	Primer Apellido	Segundo Apellido	Telefono 1	Telefono 2	Documento
1	Eduardo	Andrés	Gamarra	Solis	0981234234	21-204523	1321234
2	Pedro	Pablo	Leguizamon	Duarte	0961900900	0981555666	4232123
3	Valeria	Elizabeth	Ramirez	Gonzalez	21-234090		2341512

Figura 2

Primera Forma Normal

- En este caso es necesario crear una nueva relación que permita resolver el grupo de atributos repetitivos creado por el atributo *Telefono*

Id Empleado	Primer Nombre	Segundo Nombre	Primer Apellido	Segundo Apellido
1	Eduardo	Andrés	Gamarra	Solis
2	Pedro	Pablo	Leguizamon	Duarte
3	Valeria	Elizabeth	Ramirez	Gonzalez

Id Empleado	Tipo Telefono	Telefono
1	Particular	0981234234
1	Laboral	21-204523
2	Laboral	0961900900
2	Laboral	0981555666
3	Particular	21-234090

Figura 3

Segunda Forma Normal

- Una Relación esta en 2FN si y solo si esta en 1FN y todo atributo que no sea clave es dependiente irreduciblemente de la clave primaria.
- En el ejemplo de la figura 4 se ve la relación de *Empleados* con los datos referentes al grado de especialización del funcionario.

Id Empleado	Primer Nombre	Segundo Nombre	Primer Apellido	Segundo Apellido	Documento	Especialización
1	Eduardo	Andrés	Gamarra	Solis	1321234	Licenciatura
2	Pedro	Pablo	Leguizamon	Duarte	4232123	Administración
3	Valeria	Elizabeth	Ramirez	Gonzalez	2341512	Ingeniería

Figura 4

Segunda Forma Normal

- La regla de la Segunda Forma Normal (2FN) establece que todas las dependencias parciales se deben eliminar y separar dentro de sus propias tablas. Una dependencia parcial es un término que describe a aquellos datos que no dependen de la clave de la tabla para identificarlos.

Id Empleado	Primer Nombre	Segundo Nombre	Primer Apellido	Segundo Apellido	Documento	Id Grado
1	Eduardo	Andrés	Gamarra	Solis	1321234	1
2	Pedro	Pablo	Leguizamon	Duarte	4232123	2
3	Valeria	Elizabeth	Ramirez	Gonzalez	2341512	3

Id Grado	Nombre Área
1	Licenciatura
2	Administración
3	Ingeniería

Figura 5

Tercera Forma Normal

- Una relación esta en 3FN sí y solo sí esta en 2FN y todos los atributos que no son clave son dependientes en forma no transitiva de la clave primaria.
- En el ejemplo de la figura 6 se ve la relación de *Empleados* con los datos referentes al área donde se desempeña el funcionario.

Id Empleado	Primer Nombre	Segundo Nombre	Primer Apellido	Segundo Apellido	Documento	Area	Nombre Área
1	Eduardo	Andrés	Gamarra	Solis	1321234	101	Contabilidad
2	Pedro	Pablo	Leguizamon	Duarte	4232123	102	Finanzas
3	Valeria	Elizabeth	Ramirez	Gonzalez	2341512	103	Tecnología

Figura 6

Tercera Forma Normal

- La Tercera Forma Normal 3FN, consiste en que ningún atributo. que depende de la clave, dependa de otro atributo. Es decir, no debe tener dependencia transitiva.

Id Empleado	Primer Nombre	Segundo Nombre	Primer Apellido	Segundo Apellido	Documento	Id Area
1	Eduardo	Andrés	Gamarra	Solis	1321234	101
2	Pedro	Pablo	Leguizamon	Duarte	4232123	102
3	Valeria	Elizabeth	Ramirez	Gonzalez	2341512	103

Id Area	Nombre Área
101	Contabilidad
102	Finanzas
103	Tecnología

Figura 7

Forma Normal Boyce Codd (FNBC)

- Una relación esta en FNBC sí y solo sí toda Dependencia Funcional no trivial, irreducible a la izquierda tiene una clave candidata como su determinante.
- La Forma Normal de Boyce-Codd (o FNBC) es una versión ligeramente más fuerte de la Tercera forma normal (3FN).

IdEmpleado	Primer Nombre	Primer Apellido	IdArea	IdResponsable
1	Eduardo	Gamarra	101	21
2	Pedro	Leguizamon	102	25
3	Valeria	Ramirez	103	32
21	Ramiro	Centurion	101	
25	Maria	Cardozo	102	
32	Nilse	Rodriguez	103	

Figura 8

Forma Normal Boyce Codd (FNBC)

- Si añadimos la limitación de que el responsable sólo puede serlo de un área, este detalle produce una dependencia funcional ya que: Responsable \rightarrow Area
- Por lo tanto hemos encontrado un determinante (IdResponsable) que sin embargo no es clave primaria. Por ello, esta tabla no está en FNBC, para ello se debe crear una nueva entidad con IdResponsable, IdArea ambos atributos componen una clave primaria concatenada.

Id Empleado	Primer Nombre	Primer Apellido	Id Area
1	Eduardo	Gamarra	101
2	Pedro	Leguizamon	102
3	Valeria	Ramirez	103
21	Ramiro	Centurion	101
25	Maria	Cardozo	102
32	Nilse	Rodriguez	103

Id Area	Id Responsable
101	21
102	25
103	32

Figura 9

En que casos se puede Desnormalizar...

- Otra situación se presenta cuando se crea redundancia para ganar en rendimiento (performance), en este caso se busca registrar la información en una sola tupla en vez de recorrer un gran volumen de información.
- En la figura 11 se muestra el Saldo del Salario del Empleado, asumiendo que todas las extracciones del mismo son registradas en nuestro sistema.

Id Empleado	Primer Nombre	Primer Apellido	Salario Neto	Saldo Empleado	Id Extracción	Id Empleado	Fecha Liquidacion	Monto Extraccion
1	Eduardo	Gamarra	3.000.000	2.700.000	210134	1	07/01/2015	300.000
2	Pedro	Leguizamon	5.000.000	4.950.000	210135	2	08/01/2015	50.000
3	Valeria	Ramirez	2.000.000	0	210136	3	05/01/2015	2.000.000
21	Ramiro	Centurion	4.000.000	3.900.000	210137	21	20/01/2015	100.000
25	Maria	Cardozo	9.000.000	8.750.000	210138	25	25/01/2015	250.000
32	Nilse	Rodriguez	2.500.000	2.100.000	210139	32	15/01/2015	400.000

Figura 11

En que casos se puede Desnormalizar...

- Uno de los casos en los que se debe considerar manejar redundancia es cuando se evalúa que puede haber pérdida de información.
- Como vemos en la figura 10 existe redundancia de información referente al atributo *salario_net* sin embargo si se modificara el salario de un empleado estaríamos perdiendo la información histórica del mismo.

Id Empleado	Primer Nombre	Primer Apellido	Salario Neto	Id Liquidacion	Id Empleado	Fecha Liquidacion	Salario Neto
1	Eduardo	Gamarra	3.000.000	210134	1	05/01/2015	3.000.000
2	Pedro	Leguizamon	5.000.000	210135	2	05/01/2015	5.000.000
3	Valeria	Ramirez	2.000.000	210136	3	05/01/2015	2.000.000
21	Ramiro	Centurion	4.000.000	210137	21	05/01/2015	4.000.000
25	Maria	Cardozo	9.000.000	210138	25	05/01/2015	9.000.000
32	Nilse	Rodriguez	2.500.000	210139	32	05/01/2015	2.500.000

Figura 10