

Unidad 7: Normalización FNBC, 4FN y 5FN

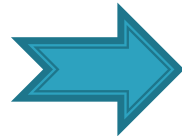
Prof. Ania Cravero Leal

Introducción

- ▶ 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 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**.
- ▶ Las bases de datos relacionales se normalizan para:
 - Evitar la redundancia de los datos.
 - Evitar problemas de actualización de los datos en las tablas.
 - Proteger la integridad de los datos.

Ejemplo Diseño Inadecuado

Proveedores	
PK	IDProveedor
Nombre DireccionFisicaCalle DireccionFisicaNoExt DireccionFisicaNoInt DireccionFisicaColonia DireccionFisicaCiudad DireccionFisicaEstado DireccionFisicaPais CorreoElectronico DatosFiscalesCalle DatosFiscalesNoExt DatosFiscalesNoInt DatosFiscalesColonia DatosFiscalesCiudad DatosFiscalesEstado DatosFiscalesPais RFC Telefono Fax	



Telefonos	
PK	IDTelefono
FK1	Numero
FK2	IDProveedor IDTipoTelefono

TiposDeTelefono	
PK	IDTipoTelefono
U1	Nombre Descripcion

Proveedores	
PK	IDProveedor
U1	Nombre

TiposDeDireccion	
PK	IDTipoDireccion
U1	Nombre Descripcion

Paises	
PK	IDPais
U1	Nombre

Direcciones	
PK	IDDireccion
Calle NumExt NumInt Colonia FK1 IDCiudad FK2 IDEstado FK3 IDPais FK4 IDTipoDireccion FK5 IDProveedor RFC	

Estados	
PK	IDEstado
U1 FK1	Nombre IDPais

Ciudades	
PK	IDCiudad
U1 FK1	Nombre IDEstado

CorreosElectronicos	
PK	IDCorreoElectronico
FK1	Direccion
FK2	IDProveedor IDTipoCorreoElectronico

TiposDeCorreoElectronico	
PK	IDTipoCorreoElectronico
U1	Nombre Descripcion

Repaso 1FN, 2FN y 3FN

Pedidos

<u>Artículo</u>	<u>cliente</u>	cantidad	precio	ciudad	distancia
A1	C1	12	100	Madrid	400
A1	C2	30	100	Valencia	200
A1	C3	15	100	Alicante	80
A2	C1	35	250	Madrid	400
A2	C2	20	250	Valencia	200
A2	C4	10	250	Madrid	400
A3	C3	25	175	Alicante	80



pedidos

<u>art</u>	<u>cli</u>	cant
A1	C1	12
A1	C2	30
A1	C3	15
A2	C1	35
A2	C2	20
A2	C4	10
A3	C3	25

artículos

<u>art</u>	precio
A1	100
A2	250
A3	175

ciudades

<u>ciudad</u>	dist
Madrid	400
Valencia	200
Alicante	80

clientes

<u>cli</u>	ciudad
C1	Madrid
C2	Valencia
C3	Alicante
C4	Madrid

Formas Normales

- ▶ Un esquema de relación está en una determinada forma normal si *satisface un determinado conjunto específico de restricciones* definidas sobre los atributos del esquema (dependencias).
 - 1ª FN (Codd, 1970)
 - Concepto de relación normalizada.
 - 2ª, 3ª FN (Codd, 1970), FNBC (Boyce/Codd, 1974)
 - Basadas en análisis de dependencias funcionales.
 - 4ª FN. Fagin, 1977
 - Basada en análisis de dependencias multivaluadas.
 - 5ª FN. Fagin, 1979
 - Basada en análisis de dependencias de proyección / combinación.

Tipos de Dependencia

- ▶ Dependencia Funcional:
- ▶ Dependencia Funcional Completa:
- ▶ Dependencia Transitiva:
- ▶ Dependencia Multivaluada: Una tabla con una dependencia multivaluada es una donde la existencia de dos o más relaciones independientes muchos a muchos causa redundancia. Se representa como sigue:
 - $A \twoheadrightarrow B$
- ▶ Dependencias de Combinación: Las DF y DMV analizadas, permiten la descomposición sin pérdida de una relación en dos de sus proyecciones.
- ▶ Sin embargo, existen relaciones donde no se puede llevar a cabo una descomposición binaria sin pérdida.
- ▶ Aun no existiendo DF ni DMV pueden existir redundancias y anomalías.

Dependencias multivaluadas

AUTORES (no normalizada)

AUTOR	MATERIA	INSTITUCION
DATE	LENGUAJE SQL DISEÑO DE BD	RELATIONAL INST. CODD&DATE CONS.
ULLMAN	DISEÑO DE BD BASES CONOCIMIENTO	STANFORD UNIV.

AUTORES (NORMALIZADA)

AUTOR	MATERIA	INSTITUCION
DATE	LENGUAJE SQL	RELATIONAL INST
DATE	DISEÑO DE BD	RELATIONAL INST
DATE	LENGUAJE SQL	CODD&DATE CONS
DATE	DISEÑO DE BD	CODD&DATE CONS.
ULLMAN	DISEÑO DE BD	STANFORD UNIV.
ULLMAN	BASES CONOCIMIENTO	STANFORD UNIV.

Dependencias de Combinación

SPJ		
S#	P#	J#
S1	P1	J2
S1	P2	J1
S2	P1	J1
S1	P1	J1

SP	
S#	P#
S1	P1
S1	P2
S2	P1

PJ	
P#	J#
P1	J2
P2	J1
P1	J1

JS	
J#	S#
J2	S1
J1	S1
J1	S2

SP * PJ		
S#	P#	J#
S1	P1	J2
S1	P2	J1
S2	P1	J1
S2	P1	J2
S1	P1	J1

tupla espúrea

$(SP * PJ) * JS \rightarrow SPJ \text{ original}$

Forma Normal de Boyce–Codd FNBC

- ▶ 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. 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.

Ejemplos FNBC

TUTORÍAS		
<u>DNI</u>	<u>Asignatura</u>	<u>Tutor</u>
12121219A	Lenguaje	Eva
12121219A	Matemáticas	Andrés
3457775G	Lenguaje	Eva
5674378J	Matemáticas	Guillermo
5674378J	Lenguaje	Julia
5634823H	Matemáticas	Guillermo

Esa tabla está en tercera forma normal (no hay dependencias transitivas), pero no en FNBC, ya que $(\text{DNI}, \text{Asignatura}) \rightarrow \text{Tutor}$ y $\text{Tutor} \rightarrow \text{Asignatura}$. En este caso la redundancia ocurre por mala selección de clave. La redundancia de la asignatura es completamente evitable. La solución sería:

TUTORÍAS	
<u>DNI</u>	<u>Tutor</u>
12121219A	Eva
12121219A	Andrés
3457775G	Eva
5674378J	Guillermo
5674378J	Julia
5634823H	Guillermo

ASIGNATURASTUTOR	
<u>Asignatura</u>	<u>Tutor</u>
Lenguaje	Eva
Matemáticas	Andrés
Matemáticas	Guillermo
Lenguaje	Julia

Otro Ejemplo

- ▶ Supongamos que disponemos de la Tabla Clientes.
- ▶ En este caso la clave candidata es el Rut
- ▶ Observamos que el **atributo ciudad no facilita información del Rut, sino que de la ubicación**. Por tanto no está en FNBC

Una solución sería almacenar el dato donde Vive en otra tabla o relación. Así evitamos la redundancia de las ciudades.

Datos Cliente			
RUT	NOMBRE	APELLIDO	CIUDAD
12,654,261-8	JUAN	PEREZ	TEMUCO
13,352,123-5	PEDRO	MONTES	TEMUCO
11,358,523-2	ANA	MORAGA	SANTIAGO

Datos Cliente		
RUT	NOMBRE	APELLIDO
12,654,261-8	JUAN	PEREZ
13,352,123-5	PEDRO	MONTES
11,358,523-2	ANA	MORAGA

Vive	
RUT	CIUDAD
12,654,261-8	TEMUCO
13,352,123-5	TEMUCO
11,358,523-2	SANTIAGO

Ciudades	
CIUDAD	Nº HABITANTES
TEMUCO	300000
SANTIAGO	5000000



4ta Forma Normal

4FN

- ▶ Una tabla está en 4NF si y solo si esta en Tercera forma normal o en BCNF (Cualquiera de ambas) y **no posee dependencias multivaluadas no triviales**. La definición de la 4NF confía en la noción de una dependencia multivaluada.
- ▶ Una tabla con una dependencia multivaluada es una donde la existencia de dos o más relaciones independientes muchos a muchos causa redundancia; y es esta redundancia la que es suprimida por la cuarta forma normal.

Ejemplos 4FN

1º Se presenta la siguiente tabla de restaurantes de pizza, el tipo y el lugar de envío:

<u>N- Restaurante</u>	<u>N-Tipo de pizza</u>	<u>N-Lugar de envío</u>
Pizza Hut	Americana	Surco
Pizza Hut	Hawaiana	Miraflores
Papa Johns	Americana	San Borja
Papa Johns	Hawaiana	San Luis

Se puede observar que hay dependencia multivalor entre el nombre del restaurante y el número de pizza, por ello se aplica la cuarta forma normal que daría paso a dos tablas:

<u>N- Restaurante</u>	<u>N-Tipo de pizza 1</u>	<u>N-Tipo de pizza 2</u>
Pizza Hut	Americana	Hawaiana
Papa Johns	Americana	Hawaiana

<u>N- Restaurante</u>	<u>N-Lugar de envío</u>
Pizza Hut	Surco
Pizza Hut	Miraflores
Papa Johns	San Borja
Papa Johns	San Luis

Ejemplos 4FN

2º Se presenta una tabla de nombre de especialidades con su código y el nombre del curso extraacadémico que existen.

<u>C-Clave Área</u>	<u>N-Especialidad</u>	<u>N-Curso</u>
S01	Sistemas	Natación
B01	Bioquímica	Danza
B01	Bioquímica	Guitarra
C03	Civil	Natación

Como es claro, se puede observar una dependencia de datos entre la clave y la especialidad, siendo más específico con la especialidad de Bioquímica. Por ello, se procede a realizar la cuarta forma normal.

<u>C-Clave Área</u>	<u>N-Especialidad</u>
S01	Sistemas
B01	Bioquímica
C03	Civil

<u>C-Clave Área</u>	<u>N-Curso</u>
S01	Natación
B01	Danza
B01	Guitarra
C03	Natación

3º En la tabla se observa al nombre del colegio, nombre de algunos cursos y el nombre del profesor que dicta el curso.

<u>N-Colegio</u>	<u>N-Curso</u>	<u>N-Profesor</u>
San Agustín	Matemática	Aldo Rojas
San Agustín	Lenguaje	José Landa
América	Física	Alberto Pérez
América	Química	Rodrigo Vivar

Sin embargo, se genera una dependencia multivalor entre el nombre del curso y el colegio donde éstos se dictan. Se procede a aplicar la cuarta forma normal para que no haya redundancia de datos, la forma que se aplica es una opción de resolver el problema.

<u>N-Colegio</u>	<u>N-Curso1</u>	<u>N-Curso2</u>
San Agustín	Matemática	Lenguaje
América	Física	Química

<u>N-Colegio</u>	<u>N-Profesor</u>
San Agustín	Aldo Rojas
San Agustín	José Landa
América	Alberto Pérez
América	Rodrigo Vivar

5ta Forma Normal

5FN

- ▶ Una tabla esta en 5FN o Forma Normal de Proyección–Reunión si está en 4FN y las únicas dependencias que existen son las dependencias de reunión de una tabla con sus proyecciones relacionándose entre las distintas proyecciones mediante la clave primaria o cualquier clave alternativa.
- ▶ La 5FN se emplea cuando en una misma tabla tenemos mucha información redundante, con pocos atributos o cuando una tabla posee una gran cantidad de atributos y se hace por ello inmanejable.
- ▶ La dificultad de la 5ª FN está en la identificación de las dependencias de combinación, las cuales no tienen una interpretación intuitiva.

Para conseguir que una tabla 4FN con gran cantidad de atributos esté en 5FN, se parte la tabla original en tantas tablas como se desee, teniendo cada una de ellas en común con las demás los campos que forman la clave primaria en la tabla original.

Descomposición a 5ª FN de una relación R

- ▶ Ejemplo para el caso de una tabla que posee una gran cantidad de atributos:

Tabla

Id	Datos Familiares			Datos Profesionales			Datos Personales			Datos Clínicos		
1	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12

- En este caso tenemos una empresa donde se guardan los datos personales, familiares, profesionales y clínicos de cada empleado en una única tabla llamada Empleados.
- Si esta tabla está ya en 4FN, se puede partir en las tablas empleados–personal, empleados–familia, empleados–profesional, empleados–clínicos; de este modo, la velocidad de acceso y la gestión de datos por cada departamento de la empresa se simplifica, al no tenerse que crear ningún tipo de restricción sobre determinados atributos que no han de ser vistos por el personal que no los necesita.
- El resultado sería:

Tabla en quinta forma normal

Id	Datos Familiares		
1	D1	D2	D3

Tabla en quinta forma normal

Id	Datos Profesionales		
1	D4	D5	D6

Tabla en quinta forma normal

Id	Datos Personales		
1	D7	D8	D9

Tabla en quinta forma normal

Id	Datos Clínicos		
1	D10	D11	D12