



Agencia de
Aprendizaje
a lo largo
de la vida

BIG DATA

Data Analytics

Actividad: Normalización de tablas

¿Qué es la normalización?

La normalización de bases de datos es un proceso de optimización que aplica una serie de reglas para minimizar la redundancia de datos y eliminar ambigüedades o problemas de inconsistencia.

Ejemplo

Partimos de la siguiente tabla:

id_cliente	nombre	direccion	fecha_nac	producto	precio	unidades	total	fecha_compra
1	Allix Ravenshaw	658 Clarendon Crossing - 7th Floor , Wynwood	14/11/1971	Wine - Stoneliegh Sauvignon	\$1.100,00	2	\$2.200,00	12/11/2023
1	Allix Ravenshaw	658 Clarendon Crossing - 7th Floor , Wynwood	14/11/1971	Milk Powder	\$1.500,00	3	\$4.500,00	5/2/2024
3	Curcio Samet	9793 Express Circle - Apt 535 , Española Way	20/9/1997	Salmon - Smoked, Sliced	\$2.300,00	2	\$4.600,00	21/3/2023
3	Curcio Samet	9793 Express Circle - Apt 535 , Española Way	20/9/1997	Bagel - Plain	\$800,00	3	\$2.400,00	11/11/2023
5	Drew Damrell	918 Sugar Trail - Room 463 , Little Havana	14/2/1996	Milk Powder	\$1.500,00	4	\$6.000,00	16/4/2023
6	Kimberley Mandry	142 Derek Park - 1st Floor , Ybor City	16/9/1970	Milk Powder	\$1.500,00	5	\$7.500,00	10/5/2023
7	Rainer Thirlwell	89475 Moose Place - Room 1996 , Old City	6/1/1978	Wine - Stoneliegh Sauvignon	\$1.100,00	4	\$4.400,00	3/7/2023
7	Rainer Thirlwell	89475 Moose Place - Room 1996 , Old City	6/1/1978	Salmon - Smoked, Sliced	\$2.300,00	3	\$6.900,00	8/8/2023
7	Rainer Thirlwell	89475 Moose Place - Room 1996 , Old City	6/1/1978	Milk Powder	\$1.500,00	2	\$3.000,00	29/3/2023
10	Catherin Tootell	17 Fairview Crossing - 15th Floor , Hyde Park	22/12/1995	Wine - Stoneliegh Sauvignon	\$1.100,00	3	\$3.300,00	26/12/2023

Primera forma normal (1FN)

Es la forma normal obligatoria y propia para lograr un esquema relacional.

- Los atributos deben ser atómicos (no se pueden dividir)
- No deben existir valores repetidos en las columnas

En el ejemplo:

- *Paso 1: Atomizar los datos.*
 - *Separar la columna nombre en nombre, apellido*
 - *Separar la columna direccion en barrio, nro, calle, depto*
- *Paso 2: Buscar patrones de repetición en las columnas y crear tablas separadas*

Primera forma normal (1FN) – Paso 1

PASO 1: Atomizar datos

id_cliente	nombre	apellido	barrio	nro	calle	depto	fecha_nac	producto	precio	unidades	total	fecha_compra
1	Allix	Ravenshaw	Wynwood	658	Clarendon Crossing	7th Floor	14/11/1971	Wine - Stoneliagh Sauvignon	\$1.100,00	2	\$2.200,00	12/11/2023
1	Allix	Ravenshaw	Wynwood	658	Clarendon Crossing	7th Floor	14/11/1971	Milk Powder	\$1.500,00	3	\$4.500,00	5/2/2024
3	Curcio	Samet	Española Way	9793	Express Circle	Apt 535	20/9/1997	Salmon - Smoked, Sliced	\$2.300,00	2	\$4.600,00	21/3/2023
3	Curcio	Samet	Española Way	9793	Express Circle	Apt 536	20/9/1997	Bagel - Plain	\$800,00	3	\$2.400,00	11/11/2023
5	Drew	Damrell	Little Havana	918	Sugar Trail	Room 463	14/2/1996	Milk Powder	\$1.500,00	4	\$6.000,00	16/4/2023
6	Kimberley	Mandry	Ybor City	142	Derek Park	1st Floor	16/9/1970	Milk Powder	\$1.500,00	5	\$7.500,00	10/5/2023
7	Rainer	Thirlwell	Old City	89475	Moose Place	Room 1996	6/1/1978	Wine - Stoneliagh Sauvignon	\$1.100,00	4	\$4.400,00	3/7/2023
7	Rainer	Thirlwell	Old City	89475	Moose Place	Room 1997	6/1/1978	Salmon - Smoked, Sliced	\$2.300,00	3	\$6.900,00	8/8/2023
7	Rainer	Thirlwell	Old City	89475	Moose Place	Room 1998	6/1/1978	Milk Powder	\$1.500,00	2	\$3.000,00	29/3/2023
10	Catherin	Tootell	Hyde Park	17	Fairview Crossing	15th Floor	22/12/1995	Wine - Stoneliagh Sauvignon	\$1.100,00	3	\$3.300,00	26/12/2023

Primera forma normal (1FN) – Paso 2

id_cliente	nombre	apellido	Barrio	nro	calle	depto	fecha_nac	producto	precio	unidades	total	fecha_compra
1	Allix	Ravenshaw	Wynwood	658	Clarendon Crossing	7th Floor	14/11/1971	Wine - Stoneliegh Sauvignon	\$1.100,00	2	\$2.200,00	12/11/2023
1	Allix	Ravenshaw	Wynwood	658	Clarendon Crossing	7th Floor	14/11/1971	Milk Powder	\$1.500,00	3	\$4.500,00	5/2/2024
3	Curcio	Samet	Old City	9793	Express Circle	Apt 535	20/9/1997	Salmon - Smoked, Sliced	\$2.300,00	2	\$4.600,00	21/3/2023
3	Curcio	Samet	Old City	9793	Express Circle	Apt 536	20/9/1997	Bagel - Plain	\$800,00	3	\$2.400,00	11/11/2023
5	Drew	Damrell	Wynwood	918	Sugar Trail	Room 463	14/2/1996	Milk Powder	\$1.500,00	4	\$6.000,00	16/4/2023
6	Kimberley	Mandry	Old City	142	Derek Park	1st Floor	16/9/1970	Milk Powder	\$1.500,00	5	\$7.500,00	10/5/2023
7	Rainer	Thirlwell	Old City	89475	Moose Place	Room 1996	6/1/1978	Wine - Stoneliegh Sauvignon	\$1.100,00	4	\$4.400,00	3/7/2023
7	Rainer	Thirlwell	Old City	89475	Moose Place	Room 1997	6/1/1978	Salmon - Smoked, Sliced	\$2.300,00	3	\$6.900,00	8/8/2023
7	Rainer	Thirlwell	Old City	89475	Moose Place	Room 1998	6/1/1978	Milk Powder	\$1.500,00	2	\$3.000,00	29/3/2023
10	Catherin	Tootell	Hyde Park	17	Fairview Crossing	15th Floor	22/12/1995	Wine - Stoneliegh Sauvignon	\$1.100,00	3	\$3.300,00	26/12/2023

Los patrones de repetición de datos en las diferentes columnas, nos dan una pista de cómo llegar al siguiente paso

Primera forma normal (1FN) – Paso 2

PASO 2: Crear tablas en 1FN

CLIENTES							
id_cliente	nombre	apellido	id_barrio	nro	calle	depto	fecha_nac
1	Allix	Ravenshaw	1	658	Clarendon Crossing	7th Floor	14/11/1971
3	Curcio	Samet	2	9793	Express Circle	Apt 535	20/9/1997
5	Drew	Damrell	1	918	Sugar Trail	Room 463	14/2/1996
6	Kimberley	Mandry	2	142	Derek Park	1st Floor	16/9/1970
7	Rainer	Thirlwell	2	89475	Moose Place	Room 1996	6/1/1978
10	Catherin	Tootell	3	17	Fairview Crossing	15th Floor	22/12/1995

Se agrega una ID como PK para enlazarla con la tabla de compras

PRODUCTOS		
id_prod	producto	precio
1	Wine - Stoneliegh Sauvignon	\$1.100,00
2	Milk Powder	\$1.500,00
3	Salmon - Smoked, Sliced	\$2.300,00
4	Bagel - Plain	\$800,00

BARRIOS	
id_barrio	barrio
1	Wynwood
2	Old City
3	Hyde Park

Se agrega una ID como PK para enlazarla con la tabla de clientes

COMPRAS					
id_cliente	id_prod	unidades	total	fecha_compra	
1	1	2	\$2.200,00	12/11/2023	
1	2	3	\$4.500,00	5/2/2024	
3	3	2	\$4.600,00	21/3/2023	
3	4	3	\$2.400,00	11/11/2023	
5	2	4	\$6.000,00	16/4/2023	
6	2	5	\$7.500,00	10/5/2023	
7	1	4	\$4.400,00	3/7/2023	
7	3	3	\$6.900,00	8/8/2023	
7	2	2	\$3.000,00	29/3/2023	
10	1	3	\$3.300,00	26/12/2023	

Segunda forma normal (2FN)

- El diseño debe estar en la 1FN.
- Los valores de las columnas dependen únicamente de la clave primaria de la tabla.
- Las tablas deben tener una única clave primaria que identifique a la tabla y de la que van a depender sus atributos (dependencia funcional).

En el ejemplo:

Las dos claves primarias de la tabla COMPRAS no son buenas candidatas para identificar inequívocamente las operaciones. Creamos una nueva columna con clave primaria.

Segunda forma normal (2FN) – Paso 3

CLIENTES							
id_cliente	nombre	apellido	id_barrio	nro	calle	depto	fecha_nac
1	Allix	Ravenshaw	1	658	Clarendon Crossing	7th Floor	14/11/1971
3	Curcio	Samet	2	9793	Express Circle	Apt 535	20/9/1997
5	Drew	Damrell	1	918	Sugar Trail	Room 463	14/2/1996
6	Kimberley	Mandry	2	142	Derek Park	1st Floor	16/9/1970
7	Rainer	Thirlwell	2	89475	Moose Place	Room 1996	6/1/1978
10	Catherin	Tootell	3	17	Fairview Crossing	15th Floor	22/12/1995

PRODUCTOS		
id_prod	producto	precio
1	Wine - Stoneliegth Sauvignon	\$1.100,00
2	Milk Powder	\$1.500,00
3	Salmon - Smoked, Sliced	\$2.300,00
4	Bagel - Plain	\$800,00

BARRIOS	
id_barrio	barrio
1	Wynwood
2	Old City
3	Hyde Park

COMPRAS					
id_cliente	id_prod	unidades	total	fecha_compra	
1	1	2	\$2.200,00	12/11/2023	
1	2	3	\$4.500,00	5/2/2024	
3	3	2	\$4.600,00	21/3/2023	
3	4	3	\$2.400,00	11/11/2023	
5	2	4	\$6.000,00	16/4/2023	
6	2	5	\$7.500,00	10/5/2023	
7	1	4	\$4.400,00	3/7/2023	
7	3	3	\$6.900,00	8/8/2023	
7	2	2	\$3.000,00	29/3/2023	
10	1	3	\$3.300,00	26/12/2023	

id_cliente identifica a los clientes, pero las otras columnas no se relacionan con esta clave
Lo mismo sucede con la columna id_prod.

Creamos una nueva columna para otra clave primaria

Segunda forma normal (2FN) – Paso 3

PASO 3: Crear clave primaria en tabla compras

CLIENTES							
id_cliente	nombre	apellido	id_barrio	nro	calle	depto	fecha_nac
1	Allix	Ravenshaw	1	658	Clarendon Crossing	7th Floor	14/11/1971
3	Curcio	Samet	2	9793	Express Circle	Apt 535	20/9/1997
5	Drew	Damrell	1	918	Sugar Trail	Room 463	14/2/1996
6	Kimberley	Mandry	2	142	Derek Park	1st Floor	16/9/1970
7	Rainer	Thirlwell	2	89475	Moose Place	Room 1996	6/1/1978
10	Catherin	Tootell	3	17	Fairview Crossing	15th Floor	22/12/1995

PRODUCTOS		
id_prod	producto	precio
1	Wine - Stoneliegh Sauvignon	\$1.100,00
2	Milk Powder	\$1.500,00
3	Salmon - Smoked, Sliced	\$2.300,00
4	Bagel - Plain	\$800,00

BARRIOS	
id_barrio	barrio
1	Wynwood
2	Old City
3	Hyde Park

COMPRAS						
id_compra	id_cliente	id_prod	unidades	total	fecha_compra	
1	1	1	2	\$2.200,00	12/11/2023	
2	1	2	3	\$4.500,00	5/2/2024	
3	3	3	2	\$4.600,00	21/3/2023	
4	3	4	3	\$2.400,00	11/11/2023	
5	5	2	4	\$6.000,00	16/4/2023	
6	6	2	5	\$7.500,00	10/5/2023	
7	7	1	4	\$4.400,00	3/7/2023	
8	7	3	3	\$6.900,00	8/8/2023	
9	7	2	2	\$3.000,00	29/3/2023	
10	10	1	3	\$3.300,00	26/12/2023	

Tercera forma normal (3FN)

- El diseño debe estar en la 2FN.
- Los valores de las columnas de la tabla no deben depender de otras columnas que no sean la clave primaria (DEPENDENCIA TRANSITIVA).

En el ejemplo:

Se elimina la columna de total, ya que se puede calcular a partir de las columnas unidades y precio.

Tercera forma normal (3FN) – Paso 4

CLIENTES						
id_cliente	nombre	apellido	id_barrio	nro	calle	depto
1	Allix	Ravenshaw	1	658	Clarendon Crossing	7th Floor
3	Curcio	Samet	2	9793	Express Circle	Apt 535
5	Drew	Damrell	1	918	Sugar Trail	Room 463
6	Kimberley	Mandry	2	142	Derek Park	1st Floor
7	Rainer	Thirlwell	2	89475	Moose Place	Room 1996
10	Catherin	Tootell	3	17	Fairview Crossing	15th Floor

PRODUCTOS		
id_prod	producto	precio
1	Wine - Stoneliegh Sauvignon	\$1.100,00
2	Milk Powder	\$1.500,00
3	Salmon - Smoked, Sliced	\$2.300,00
4	Bagel - Plain	\$800,00

BARRIOS	
id_barrio	barrio
1	Wynwood
2	Old City
3	Hyde Park

COMPRAS						
id_compra	id_cliente	id_prod	unidades	total	fecha_compra	
1	1	1	2	\$2.200,00	12/11/2023	
2	1	2	3	\$4.500,00	5/2/2024	
3	3	3	2	\$4.600,00	21/3/2023	
4	3	4	3	\$2.400,00	11/11/2023	
5	5	2	4	\$6.000,00	16/4/2023	
6	6	2	5	\$7.500,00	10/5/2023	
7	7	1	4	\$4.400,00	3/7/2023	
8	7	3	3	\$6.900,00	8/8/2023	
9	7	2	2	\$3.000,00	29/3/2023	
10	10	1	3	\$3.300,00	26/12/2023	

El precio total puede calcularse a partir del precio y la unidades pedidas. No depende de la clave primaria

Esa columna puede eliminarse

Tercera forma normal (3FN) – Paso 4

PASO 4: Eliminar dependencia transitiva

CLIENTES							
id_cliente	nombre	apellido	id_barrio	nro	calle	depto	fecha_nac
1	Allix	Ravenshaw	1	658	Clarendon Crossing	7th Floor	14/11/1971
3	Curcio	Samet	2	9793	Express Circle	Apt 535	20/9/1997
5	Drew	Damrell	1	918	Sugar Trail	Room 463	14/2/1996
6	Kimberley	Mandry	2	142	Derek Park	1st Floor	16/9/1970
7	Rainer	Thirlwell	2	89475	Moose Place	Room 1996	6/1/1978
10	Catherin	Tootell	3	17	Fairview Crossing	15th Floor	22/12/1995

PRODUCTOS		
id_prod	producto	precio
1	Wine - Stoneliegh Sauvignon	\$1.100,00
2	Milk Powder	\$1.500,00
3	Salmon - Smoked, Sliced	\$2.300,00
4	Bagel - Plain	\$800,00

BARRIOS	
id_barrio	barrio
1	Wynwood
2	Old City
3	Hyde Park

COMPRAS				
id_compra	id_cliente	id_prod	unidades	fecha_compra
1	1	1	2	12/11/2023
2	1	2	3	5/2/2024
3	3	3	2	21/3/2023
4	3	4	3	11/11/2023
5	5	2	4	16/4/2023
6	6	2	5	10/5/2023
7	7	1	4	3/7/2023
8	7	3	3	8/8/2023
9	7	2	2	29/3/2023
10	10	1	3	26/12/2023

Diagrama entidad-relación

Un diagrama entidad-relación es un diagrama que muestra cómo las "entidades", como personas, objetos o conceptos, se relacionan entre sí dentro de un sistema.

DER: Diagrama Entidad/Relación

