

# Normalización

El objetivo principal es eliminar la redundancia y la inconsistencia en los datos, lo que ayuda a evitar problemas como la duplicación de datos y la pérdida de integridad.

Existen diferentes niveles de normalización, conocidos como formas normales, que establecen ciertas reglas y restricciones sobre cómo deben organizarse los datos. Las formas normales más comunes son:

## PRIMERA FORMA NORMAL (1NF)

ID Empleado	Nombre	Habilidades
1	Pedro	Marketing, Desarrollo
2	Juan	Marketing
3	Robert	Analista de datos, Desarrollo

TABLA "EMPLEADOS" ✓

ID Empleado	Nombre
1	Pedro
2	Juan
3	Robert

Muchas "Habilidades" separadas por comas en una sola celda viola la 1NF, ya que la columna "Habilidades" contiene valores multivaluados.

TABLA "HABILIDADES" ✓

ID Empleado	Habilidad
1	Marketing
1	Desarrollo
2	Marketing
3	Analista de datos
3	Desarrollo



APLICANDO 1NF: la tabla "Empleados" solo contiene los datos de identificación de los empleados (ID Empleado y Nombre), y las habilidades se han dividido en registros individuales en la tabla "Habilidades", vinculadas a los empleados a través de la clave primaria (ID Empleado). Ahora cada columna contiene un solo valor atómico.



## SEGUNDA FORMA NORMAL (2NF)

Todos los atributos deben depender completamente de la clave primaria, si hay atributos que depende parcialmente se deben separar en una nueva tabla.

Número Pedido	ID Cliente	Nombre Cliente	Producto	Cantidad
1	101	Juan Pérez	Camiseta	2
2	102	María Gómez	Pantalón	1
3	103	Pedro Rodríguez	Zapatos	3
4	101	Juan Pérez	Gorra	2

En esta tabla, la columna "Nombre Cliente" y "Producto" dependen solo de una parte de la clave primaria, ya que "Nombre Cliente" depende solo de "ID Cliente" y "Producto" depende solo de "Número Pedido". Esto viola la Segunda Forma Normal, ya que hay dependencias parciales en la tabla.

TABLA "PEDIDOS"

Número Pedido	ID Cliente	Cantidad
1	101	2
2	102	1
3	103	3
4	101	2

TABLA "CLIENTES"

ID Cliente	Nombre Cliente
101	Juan Pérez
102	María Gómez
103	Pedro Rodríguez



TABLA "PRODUCTOS"

Número Pedido	Producto
1	Camiseta
2	Pantalón
3	Zapatos
4	Gorra

✓ APLICANDO 2NF: la tabla "Pedidos" contiene solo los datos relacionados con los pedidos, la tabla "Clientes" contiene los datos de los clientes, y la tabla "Productos" contiene los datos de los productos. Las dependencias parciales se han eliminado y cada columna depende completamente de la clave primaria correspondiente, cumpliendo así con la Segunda Forma Normal (2NF)

TERCERA FORMA NORMAL (3NF)

Se eliminan las dependencias transitivas, lo que significa que no debe haber dependencias indirectas en las tablas. Si un atributo depende de otro atributo que a su vez depende de la clave primaria, entonces el primer atributo se debe mover a una nueva tabla.

Número Factura	ID Cliente	Nombre Cliente	Dirección Cliente	Ciudad Cliente	Producto	Cantidad
1	101	Juan Pérez	Calle 123	Ciudad A	Camiseta	2
2	102	María Gómez	Calle 456	Ciudad B	Pantalón	1
3	103	Pedro Rodríguez	Calle 789	Ciudad C	Zapatos	3
4	101	Juan Pérez	Calle 123	Ciudad A	Gorra	2





En esta tabla, la columna "Nombre Cliente", "Dirección Cliente" y "Ciudad Cliente" dependen de "ID Cliente", y la columna "Ciudad Cliente" depende también de "Dirección Cliente". Esto crea una dependencia transitiva, ya que "Ciudad Cliente" depende indirectamente de la clave primaria "Número Factura" a través de "Dirección Cliente". Esto viola la Tercera Forma Normal, ya que hay dependencias transitivas en la tabla.

Para llevar la tabla a la Tercera Forma Normal, debemos dividir la tabla en varias tablas separadas, de la siguiente manera:



Aplicando 3NF: la tabla "Facturas" contiene solo los datos relacionados con las facturas, la tabla "Clientes" contiene los datos de los clientes, y se ha eliminado la dependencia transitiva "Ciudad Cliente" en la tabla "Facturas" al moverla a la tabla "Clientes".

TABLA "FACTURAS"

Número Factura	ID Cliente	Producto	Cantidad
1	101	Camiseta	2
2	102	Pantalón	1
3	103	Zapatos	3
4	101	Gorra	2

TABLA "CLIENTES"

ID Cliente	Nombre Cliente	Dirección Cliente	Ciudad Cliente
101	Juan Pérez	Calle 123	Ciudad A
102	María Gómez	Calle 456	Ciudad B
103	Pedro Rodríguez	Calle 789	Ciudad C



CUARTA FORMA NORMAL (4NF)

Se aplica a bases de datos que tienen múltiples relaciones de muchos a muchos y buscan evitar redundancias y anomalías de actualización.

Supongamos que tienes una tabla llamada "Estudiantes" con la siguiente estructura:

ID Estudiante	Nombre Estudiante	Materias	Calificaciones
101	Juan Pérez	Matemáticas, Física	95, 88
102	María Gómez	Física, Química	92, 89
103	Pedro Rodríguez	Matemáticas, Química	97, 85

En esta tabla, la columna "Materias" contiene una lista de materias separadas por comas, y la columna "Calificaciones" contiene una lista de calificaciones correspondientes a esas materias también separadas por comas. Esto crea una dependencia multivaluada, ya que las columnas "Materias" y "Calificaciones" dependen de la misma clave primaria "ID Estudiante", y además, tienen dependencias entre sí. Esto viola la Cuarta Forma Normal, ya que hay dependencias multivaluadas en la tabla.

TABLA "ESTUDIANTES"

ID Estudiante	Nombre Estudiante
101	Juan Pérez
102	María Gómez
103	Pedro Rodríguez



TABLA "MATERIAS"

ID Estudiante	Materia	Calificación
101	Matemáticas	95
101	Física	88
102	Física	92
102	Química	89
103	Matemáticas	97
103	Química	85

✓ Aplicando 4NF: se ha creado una tabla separada "Materias" que contiene las materias y calificaciones de cada estudiante, en lugar de tenerlas como listas separadas en la tabla "Estudiantes". De esta manera, se eliminan las dependencias multivaluadas y se cumple con la Cuarta Forma Normal (4NF).

QUINTA FORMA NORMAL (5NF)

La 5NF busca evitar redundancias y anomalías de actualización en bases de datos temporales. Esta forma de normalización se aplica a situaciones muy específicas y complejas, y su implementación no es común en bases de datos cotidianas.

Supongamos que tienes dos tablas: "Estudiantes" y "Materias", que se relacionan entre sí en una relación de muchos a muchos, es decir, un estudiante puede tomar varias materias y una materia puede ser tomada por varios estudiantes.

TABLA "ESTUDIANTES"

ID Estudiante	Nombre Estudiante
101	Juan Pérez
102	María Gómez
103	Pedro Rodríguez

TABLA "MATERIAS"

ID Materia	Nombre Materia
1	Matemáticas
2	Física
3	Química

TABLA "INSCRIPCIONES"

ID Estudiante	ID Materia	Calificación
101	1	95
101	2	88
102	2	92
102	3	89
103	1	97
103	3	85

En este diseño, la tabla "Inscripciones" contiene las inscripciones de los estudiantes en las diferentes materias, junto con las calificaciones obtenidas. Sin embargo, esta tabla no cumple con la Quinta Forma Normal (5NF), ya que tiene una dependencia de unión o combinación, es decir, la tabla depende de la combinación de los valores de "ID Estudiante" y "ID Materia" para determinar las calificaciones.





Para llevar la tabla a la Quinta Forma Normal, podemos descomponerla en tablas separadas:

TABLA "ESTUDIANTES"

ID Estudiante	Nombre Estudiante
101	Juan Pérez
102	María Gómez
103	Pedro Rodríguez

TABLA "MATERIAS"

ID Materia	Nombre Materia
1	Matemáticas
2	Física
3	Química

TABLA "INSCRIPCIONES"

ID Estudiante	ID Materia
101	1
101	2
102	2
102	3
103	1
103	3

TABLA "CALIFICACIONES"

ID Estudiante	ID Materia	Calificación
101	1	95
101	2	88
102	2	92
102	3	89
103	1	97
103	3	85







Aplicando 5NF: En esta nueva estructura, la tabla "Inscripciones" contiene solo las claves foráneas "ID Estudiante" e "ID Materia" que vinculan a los estudiantes con las materias en las que están inscritos, eliminando la dependencia de unión o combinación. Por otro lado, la tabla "Calificaciones" contiene las calificaciones asociadas a cada inscripción. De esta manera, se ha eliminado la redundancia y las anomalías de actualización que podrían surgir al modificar las calificaciones en la tabla "Inscripciones".

