Materia: Bases de datos

Docente: Erick Giovanni Varela Guzmán

Instructor: Diego Eduardo Castro Quintanilla

Ciclo 01-2025

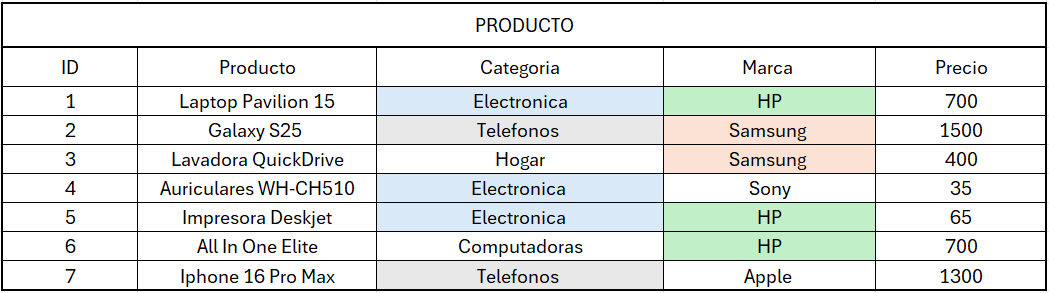
Normalización en bases de datos

Consiste en utilizar una serie de normas que se aplican en el modelado de bases de datos con el objetivo de organizar la información para: reducir la redundancia, mejorar la integridad, facilitar el mantenimiento y minimizar errores.

En palabras simples podemos decir que:

“**Normalizar** una base de datos significa **dividir los datos en varias tablas** para que cada una tenga un propósito claro, **evitando datos repetidos** y **mejorando la integridad**.”

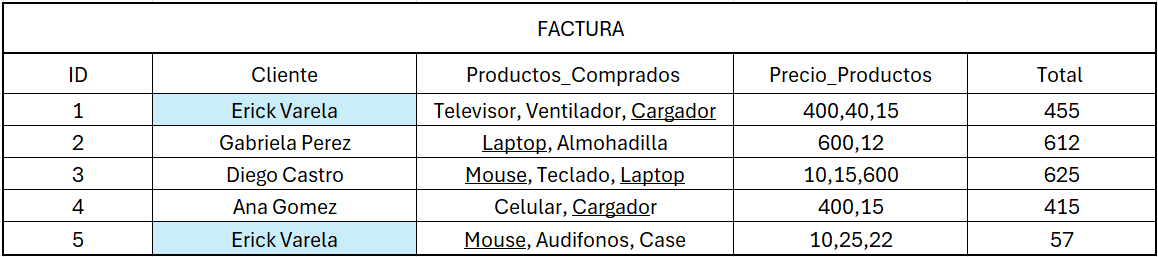
**Redundancia de datos:** La misma información se almacena en varios registros lo que aumenta significativamente el espacio que utiliza la base al manejar grandes volúmenes de datos.



“Este es un ejemplo donde claramente hay redundancia de datos en las columnas Categoría y Marca ya que se repiten valores los cuales están marcados”

**IMPORTANTE:** Los valores numéricos como precio a pesar que se pueden llegar a repetir por lo general no se normalizan tampoco campos como edad, precios, fecha de nacimiento y fechas en general debemos poner más atención en los campos de tipo String como veremos en los próximos ejemplos.

Aquí tenemos otro caso de una tabla que **no está normalizada**



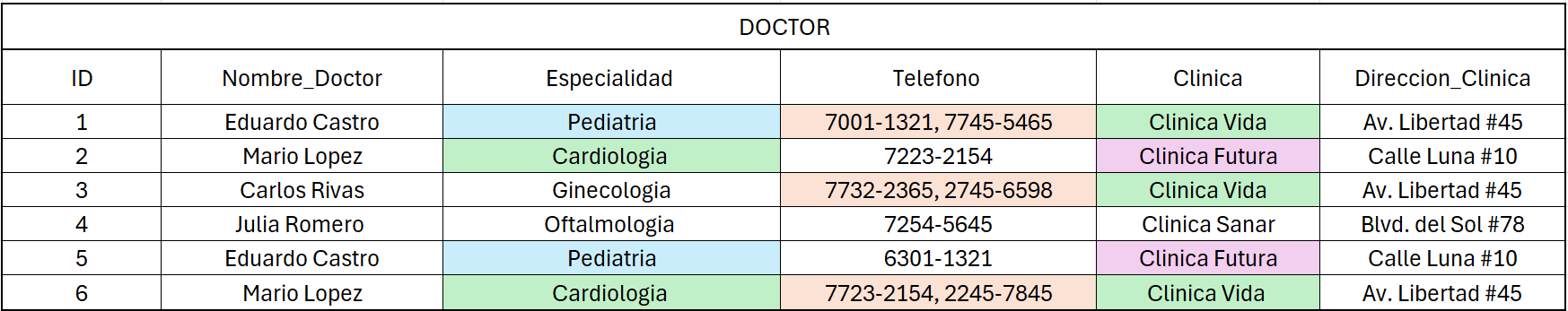
¿Es una forma valida de guardar información?

Si es una forma valida y hasta funcional, pero al no respetar la normalización provoca que las consultas se vuelvan mucho más complejas y poco practicas ya que para saber el precio de un producto tenemos que buscar la posición a través de las comas lo que dificultaría enormemente poder realizar cálculos o estadísticas en general, en caso de querer obtener una lista de productos que adquirió un cliente tendríamos desarrollar un algoritmo que separara los productos y si quisiéramos modificar los productos de una factura seria un caso peor ya que tendríamos que desarmar la cadena por comas por esas razones se considera una mala práctica manejar atributos multivaluados de esa forma.

¿En qué debemos fijarnos para normalizar?

Principalmente podemos visualizar que existen campos (Productos\_Comprados y Precio\_Productos) que contienen varios valores separados por comas lo cual es una forma de representar los atributos multivaluados además en la columna Cliente tiene valores repetidos.

A continuación, tenemos un último caso más complejo donde tenemos muchos factores que se pueden normalizar



La lista de factores que tenemos que tener en cuenta es la siguiente:

* Una especialidad puede tenerla uno o muchos doctores
* Los doctores pueden tener uno o varios teléfonos
* Un doctor puede operar en una o varias clínicas
* ¿Sera que todos los atributos están relacionados directamente con el doctor?
* Existen múltiples direcciones repetidas

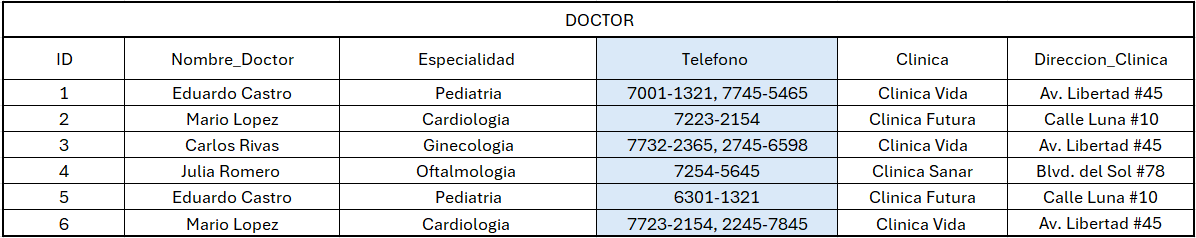
Para estos casos es que utilizamos la normalización para evitar esa redundancia de datos, existen 5 formas normales (FN) pero nosotros solo vamos a trabajar hasta la forma normal 3 (FN3).

Algunas consideraciones previas antes de aplicar normalización:

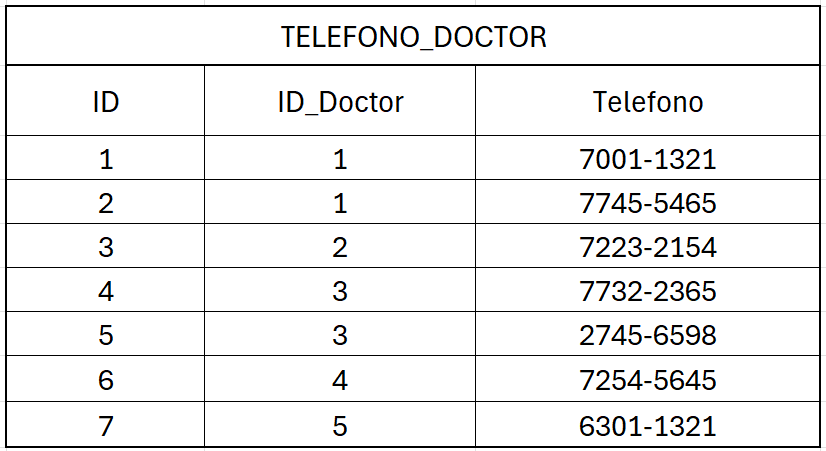
* Se debe aplicar en orden desde la FN1 luego la FN2 y por último la FN3.
* En caso que no se pueda aplicar alguna de las FN se omite y verifica la siguiente.
* Se aplica en todo el modelo relacional a excepción de las tablas cruz o tablas de muchos a muchos (N: N).
* Debemos evitar que existan campos en la base de datos donde no se les asigna ningún valor a lo largo del tiempo es decir que queden NULL para siempre.
* Al crear el diagrama la pata de gallo dentro del modelo relacional siempre apunta hacia la tabla que contiene la llave foránea.
* Si hay atributos que no pertenecen principalmente a la entidad utilizamos forma normal 3.

**FORMA NORMAL 1: Atributos Multivaluados**

Basada en la regla de que “cada celda debe tener un solo valor “, por lo tanto, lo que tenemos que hacer es detectar si la tabla tiene algún campo es multivaluado (que pueda contener más de un valor en una misma celda).

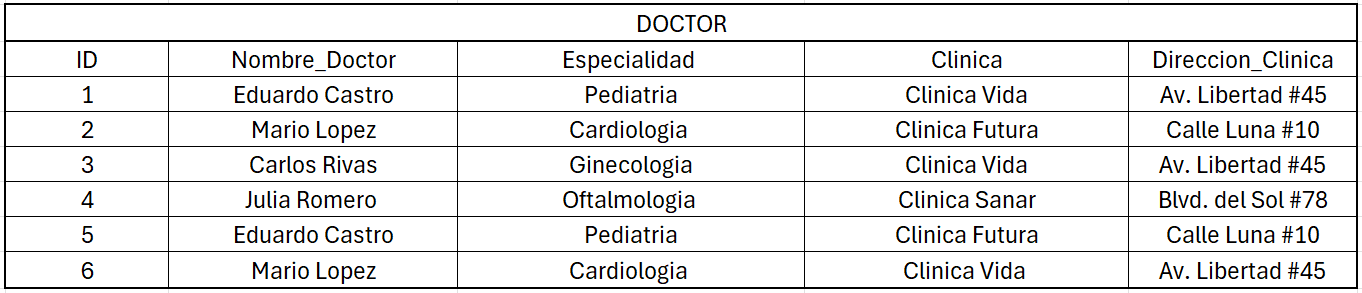


En el caso de la tabla Doctor podemos visualizar que el campo Teléfono es multivaluado por lo tanto lo que hacemos es lo siguiente



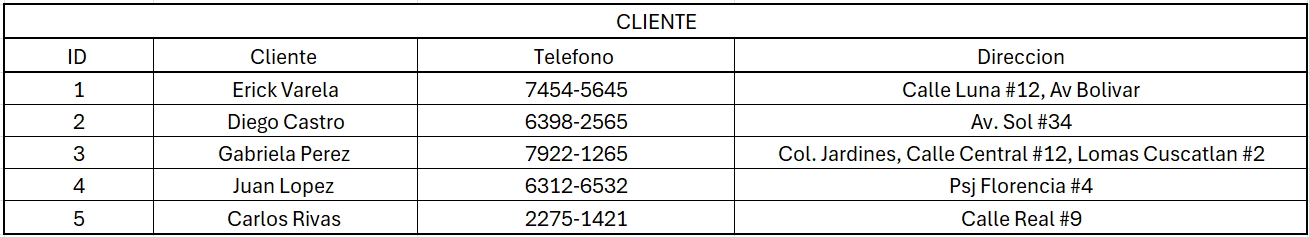
Creamos una nueva tabla con el nombre del atributo multivaluado y la tabla original (en este caso Doctor), la nueva tabla contendrá tres campos ID (identificador de cada registro), id\_doctor (del doctor al que pertenece el teléfono) y el campo teléfono que contiene un valor atómico.

IMPORTANTE: Debemos eliminar el atributo multivaluado de la entidad original por lo tanto la tabla final nos quedaría así.

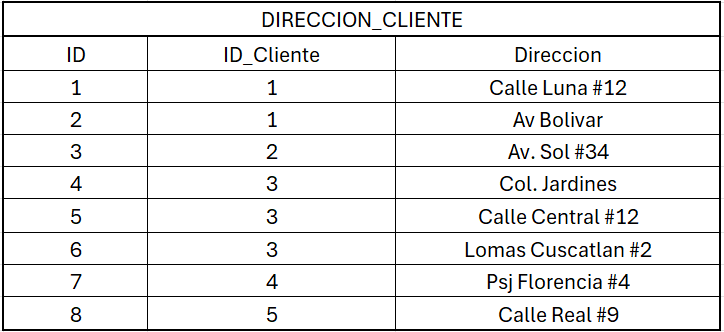


Generalmente se aplica a campos como teléfonos, correos, direcciones y otros atributos que puedan ser multivariados que también pueden colocarse como por ejemplo dirección1, dirección2, direccion3 en esos casos tenemos que normalizarlo.

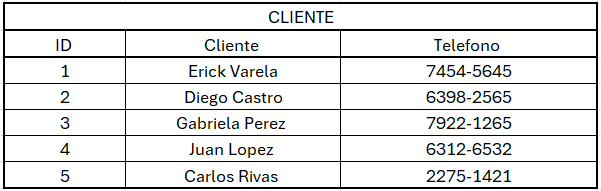
Otro ejemplo practico seria:



En este caso podemos ver que el Teléfono es un campo único para cada cliente, pero la dirección es multivaluada esto es muy común en tiendas en línea donde un cliente puede tener más de una dirección de entrega por lo tanto tenemos que normalizarlo y obtenemos el siguiente resultado



La tabla Direccion\_Cliente permite que el cliente pueda tener una o varias direcciones asociadas y la tabla original se reescribe sin el atributo multivaluado.

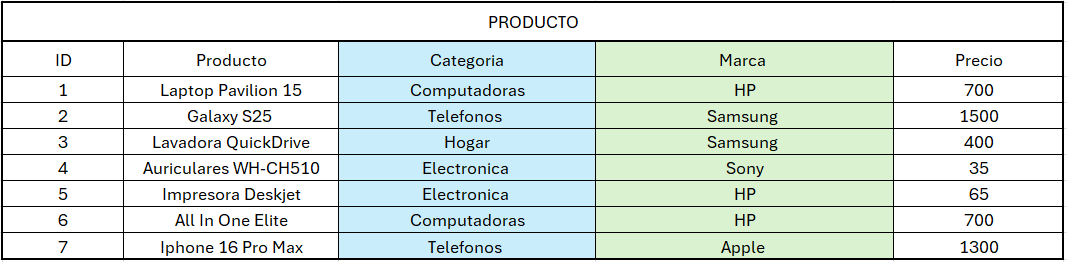


**FORMA NORMAL 2: Tablas Catálogos**

Cuando una tabla usa dos columnas como clave primaria, todos los demás datos deben depender de las dos juntas, no solo de una de ellas.

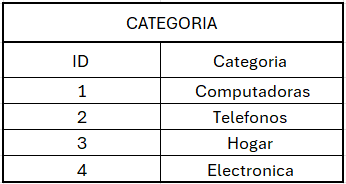
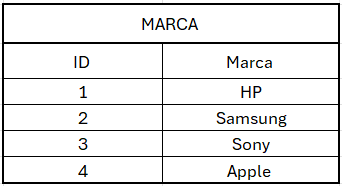
Si algún dato depende solo de una parte, entonces la tabla no cumple la 2FN y ese dato debería estar en otra tabla.

En palabras simples “es cuando separamos los datos repetidos en nuevas tablas llamadas tablas catálogo para evitar duplicar información”. Por lo general son atributos como: categoría, marca, tipo, proveedor, curso, materia, profesión, cargo, entre otros.

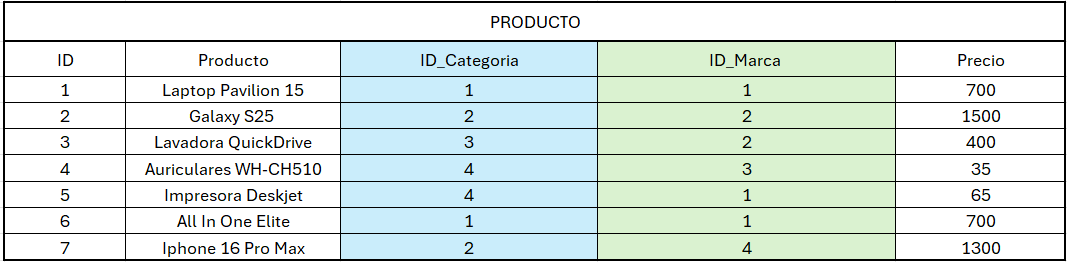


En este caso tenemos dos tablas catalogo “Categoria y Marca” las podemos identificar porque tienen campos con valor atómico pero que pueden ser repetidos en varios registros.

Entonces lo que vamos a hacer es normalizar aplicando la FN2, lo que hacemos es crear una nueva tabla con cada atributo catálogo.

Y dentro de la tabla principal en lugar de colocar el nombre de las categorías colocamos únicamente el ID de la categoría de la que queramos hacer referencia y renombramos el nombre del atributo como ID\_Catalogo nos queda de la siguiente forma.



Con esto únicamente repetimos el ID en lugar de colocar el nombre completo de la categoría con lo que estamos ahorrando muchísima memoria cuando manejamos muchísimos registros.