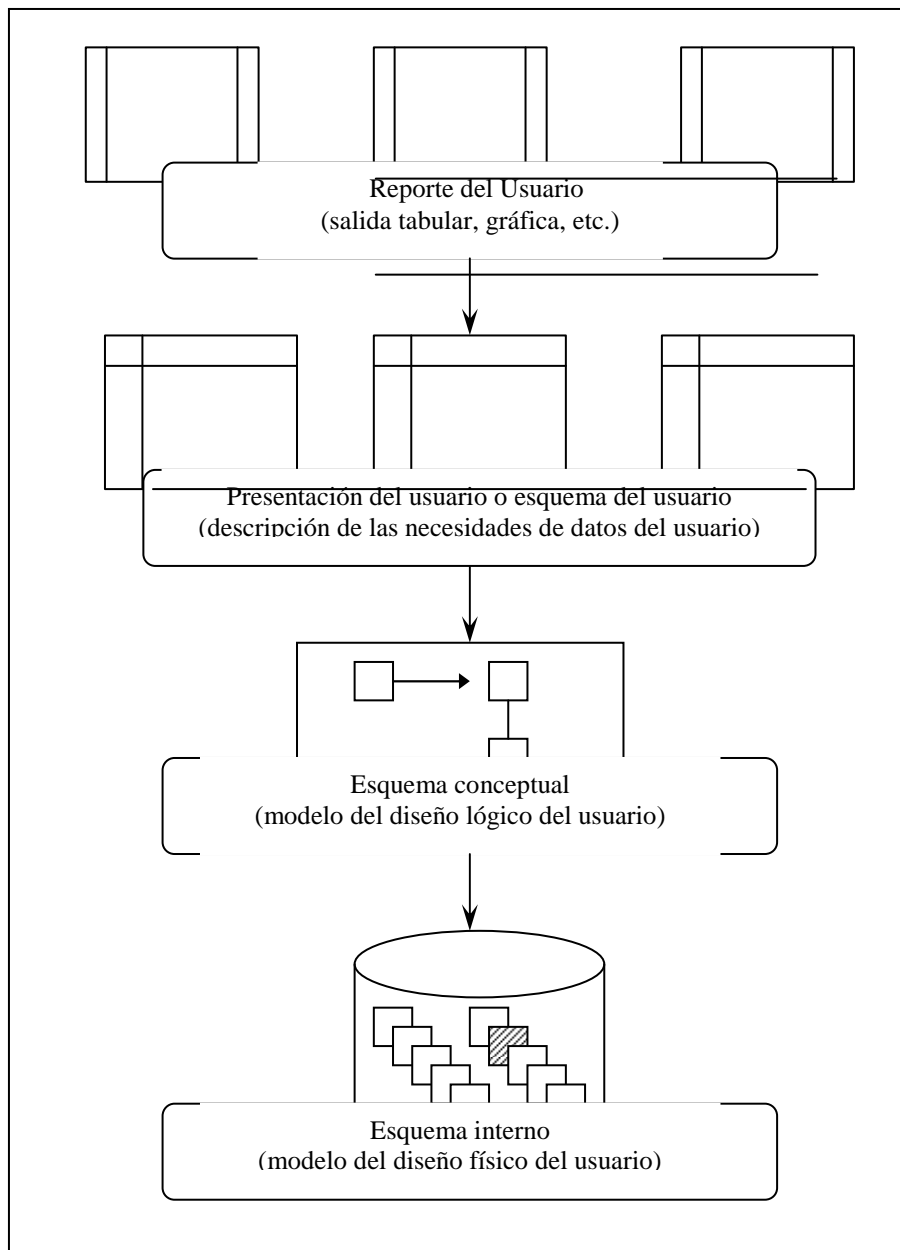


Normalización

El advenimiento de las bases de datos creó la necesidad de utilizar técnicas como la normalización, ya que las bases de datos no son meramente una colección de archivos. Más bien, una base de datos es una fuente central de datos significativos, los cuales son compartidos por numerosos usuarios para diversas aplicaciones. La esencia de una base de datos es el Sistema Administrados de Base de Datos (DBMS: Data Base Management System), el cual permite la creación, modificación y actualización de la base de datos; la recuperación de datos y la emisión de reportes.



El diseño de la base de datos incluye el resumen de los reportes del usuario, las presentaciones para el usuario y los diseños lógicos y físicos.

Términos Básicos

Comencemos por enunciar y explicar distintos elementos y conceptos básicos.

Entidad, Una entidad es una cosa, persona o un hecho sobre el cual el sistema debe guardar información, debido a que con posterioridad interesará su recuperación. Entidades del sistema de una empresa podrían ser clientes, productos, ventas, etc.

Para cada entidad es necesario definir los siguientes conceptos:

- ♦ Atributos
- ♦ Ocurrencias
- ♦ Clave primaria

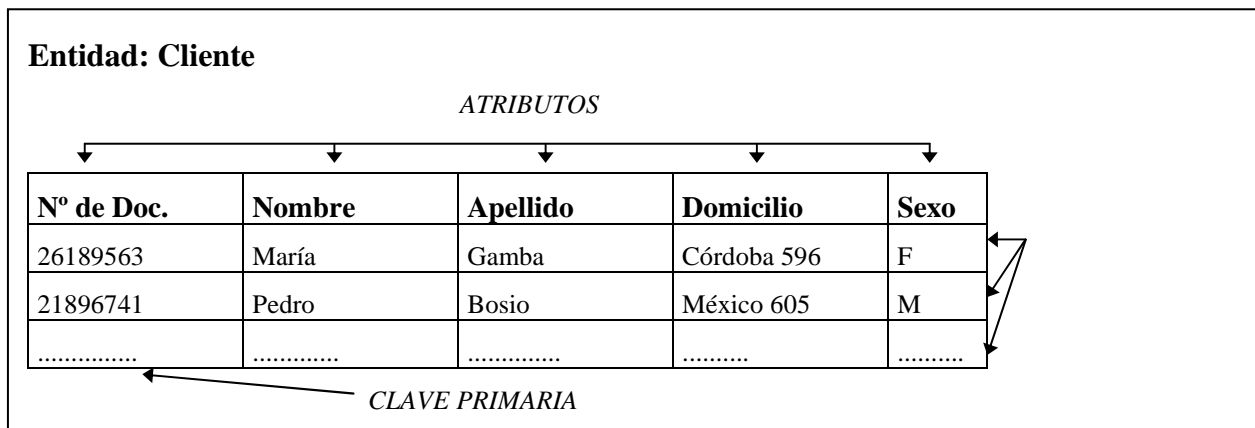
Un **atributo** es una característica o propiedad que interesa representar sobre una entidad. Por ejemplo, algunos atributos para la *entidad cliente* podrían ser: Nombre, Domicilio, Sexo, N° documento, etc. Es importante aclarar que los atributos de una entidad son acordes con las características de la realidad que se desea representar. Por ejemplo, para la *entidad cliente* de una determinada empresa que vende ropa puede ser un atributo la altura de la persona, mientras que para otra que fuese una ferretería, puede no representar una característica que interese modelar. Es por ello que los atributos de las entidades varían para cada diseño en particular, ya que los atributos no son todas las características de esa entidad, sino solamente las que son relevantes; son aquellas propiedades que representan a la entidad para esa organización.

Los atributos a su vez podrán asumir distintos **valores**. Al conjunto de valores que pueden asumir los atributos se lo llama **dominio** de un atributo. Por ejemplo: el dominio del atributo Sexo puede asumir dos valores: Masculino / Femenino.

Una **ocurrencia** es una instancia de la entidad, es decir, es cuando en una entidad los distintos atributos asumen un valor para representar una determinada instancia de la misma en la realidad. Para el ejemplo anterior de la *entidad cliente*, una ocurrencia es un cliente en particular, donde sus atributos Nombre, Apellido, domicilio, etc. asumen los valores que le son propios a dicha persona. Es decir, que en este caso la entidad cliente debiera tener tantas ocurrencias como clientes tuviese dicha organización.

La **clave primaria** es un atributo o un conjunto de atributos que se definen en una entidad, y que permiten identificar en forma unívoca a las ocurrencias de la misma. En la entidad cliente la clave primaria podría ser el número de documento, ya que este podría identificar unívocamente una ocurrencia del resto. El valor que asume la clave primaria en una ocurrencia no podrá repetirse en ninguna otra. Para elegir la clave primaria se evalúan los distintos atributos de la entidad y se analiza cuál tiene la particularidad de no repetirse en más de una ocurrencia. Si no existe ningún atributo que cumpla dicha característica, debe definirse una clave primaria conjunta, es decir más de un atributo tomados en forma conjunta.

El siguiente gráfico ejemplifica en un formato matricial los conceptos explicados a través de un ejemplo con la entidad cliente.



Las **relaciones** son las conexiones que existen entre los datos que se almacenan. Se llamarán relaciones a las asociaciones que existen entre las ocurrencias de las entidades. En la mayoría de los casos una relación es la que permite conectar ocurrencias entre dos entidades. Una relación por ejemplo, es la que existe entre las ocurrencias de las entidades *clientes* y *venta*, ya que las ocurrencias que existen en la entidad cliente están relacionadas con las ocurrencias de la entidad venta donde figuran las ventas que se realizaron a dichos clientes, y es así que un cliente podrá tener muchas ventas relacionadas.

Cada relación une dos tablas en una de tres formas:

- ♦ **Uno en uno:** cada fila de la primer tabla se corresponde con una (y sólo una) fila de la segunda. Por ejemplo, el caso de una tabla con información sobre caballos de carrera y otra con datos sobre jockeys; si se desea establecer una relación que muestre qué jockey jinetea cada caballo, a cada fila de la tabla de caballos le corresponderá una de la de jockey, y viceversa.
- ♦ **Uno a muchos:** cada fila de la primer tabla se corresponde con una o más filas de la segunda. En este caso, se dice que la primera es la tabla "madre" y la otra, la tabla "hija". Por ejemplo, una tabla de compañías discográficas y otra de discos: a cada compañía le corresponderán varios discos, pero a cada disco puede estar relacionado con una única compañía.
- ♦ **Muchos a muchos:** Varias filas de la primera tabla se corresponden con varias filas de la segunda. En este caso se hace necesario crear una tabla intermedia para poder determinar qué fila de la primer tabla se corresponde con cuál de la segunda. Sería el caso de una relación entre una tabla en la cual se almacenan recetas y otra en que se registran los ingredientes de las mismas; una receta normalmente tiene varios ingredientes, y un ingrediente puede formar parte de más de una receta.

Concepto

La normalización es un proceso de transformación, de las complejas presentaciones de usuarios y de almacenamiento de datos en conjuntos estables de estructuras de datos de menor tamaño y absolutamente independiente. Además de ser más sencilla, las estructuras de datos normalizados permiten que el almacenamiento futuro del sistema resulte más fácil y seguro. Es por ello que la normalización de una base de datos es un proceso con tres objetivos básicos: asegurar la integridad de los datos, eliminar información redundante y hacer los sistemas más eficientes.

Los tres pasos de la normalización

Al comenzar, ya sea con la presentación del usuario (formulario o listado) o con el almacenamiento de datos diseñado para un diccionario de datos, el analista normaliza una estructura de datos en tres pasos, tal como veremos en el gráfico. Cada paso involucra un importante procedimiento de simplificación de la estructura de datos.

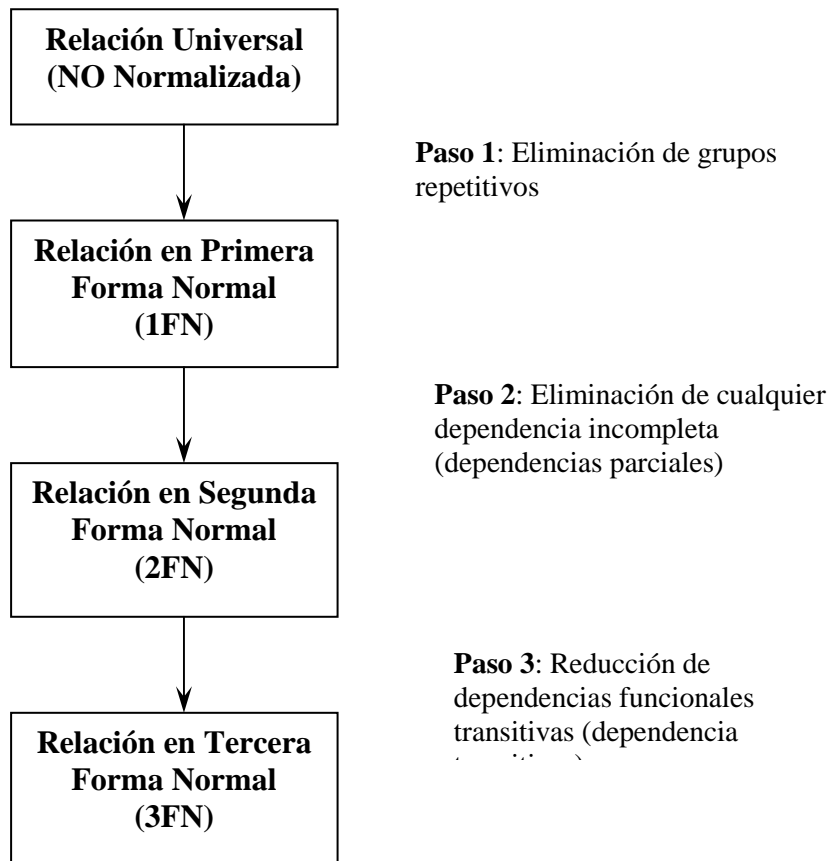
La relación derivada de la presentación del usuario o del almacenamiento de datos, generalmente se encontrará no normalizada.

La primera etapa del proceso incluye la eliminación de grupos repetitivos y la identificación de la clave que define al criterio primario. Con el fin de hacer esto, la relación necesita desglosarse en dos o más relaciones.

En este punto, las relaciones pueden encontrarse en forma normal tercera, pero quizás sea necesario más pasos para transformar las relaciones a la tercera forma normal.

El segundo paso asegura que todos los atributos no-clave, o sin llave, sean completamente dependientes de la clave del criterio primario. Todas las dependencias normales se eliminan y se colocarán en otra relación.

El tercer paso elimina cualquier dependencia transitoria. Una dependencia transitoria es aquella en la cual sus atributos no-clave son dependientes de otros atributos no-clave.



Ejemplo de Normalización

En la figura que se presenta a continuación es una presentación para el usuario de la Compañía Manufacturera de Equipos de AI S. Well Hidraulic.

*Compañía de Equipos
Hidráulicos AI. S. Well
Spring Valley, Minnesota*

Vendedor N° : 3462
Nombre: Walter
Area de Venta: Norte

Cliente Nro	Nombre Cliente	Número Almacén	Localidad Almacén	Ventas
18765	Delta Service	4	Fargo	13.540
18830	M. Levy ans Sons	3	Bismarck	10.600
.....

.....

El principal objetivo del proceso de normalización es simplificar toda la complejidad existente de los datos en las aplicaciones de los usuarios. Por ejemplo, si se considerara utilizar la presentación expuesta del usuario e intentara desarrollar una tabla de relación a partir de ella, la tabla se asemejaría a la que se encuentra más abajo, comúnmente llamado **Archivo Plano o Presentación Plana**, ya que ésta es una relación que se basa en nuestra presentación inicial, nos referiremos a ella como REPORTE-VENTAS.

El REPORTE-VENTA es una relación no normalizada, ya que cuenta con grupos repetitivos.

Nro Vend.	Nombre Vendedor	Area de Venta	Nro de Cliente	Nombre Cliente	Nro de Almacén	Ubicación Almacén	Valor de Venta
3462	Walter	Norte	18765	Delta....	4	Fargo	13540
			18830	Levy	3	Bismark	10600
			19242	Ranner....	3	Bismark	9700
3593	Dryne	Este	18841	Flood Inc.	2	Superior	11560
			18899	Seward...	2	Superior	2590
			19565	Stod Inc.	1	Plymouth	8800
.....

Si los Datos se listaran en una lista no normalizada o archivo plano, habría grupos repetidos. Un mismo dato almacenado en distintas filas, implica ineficiencia en el uso de los medios de almacenamiento (¿cuál es el valor agregado de almacenar múltiples veces el mismo dato?) y acarrea serios inconvenientes en el procesamiento de altas, bajas y modificaciones.

Primera Forma Normal

Para que una entidad se encuentre en **Primer Forma Normal**, debe cumplir dos condiciones:

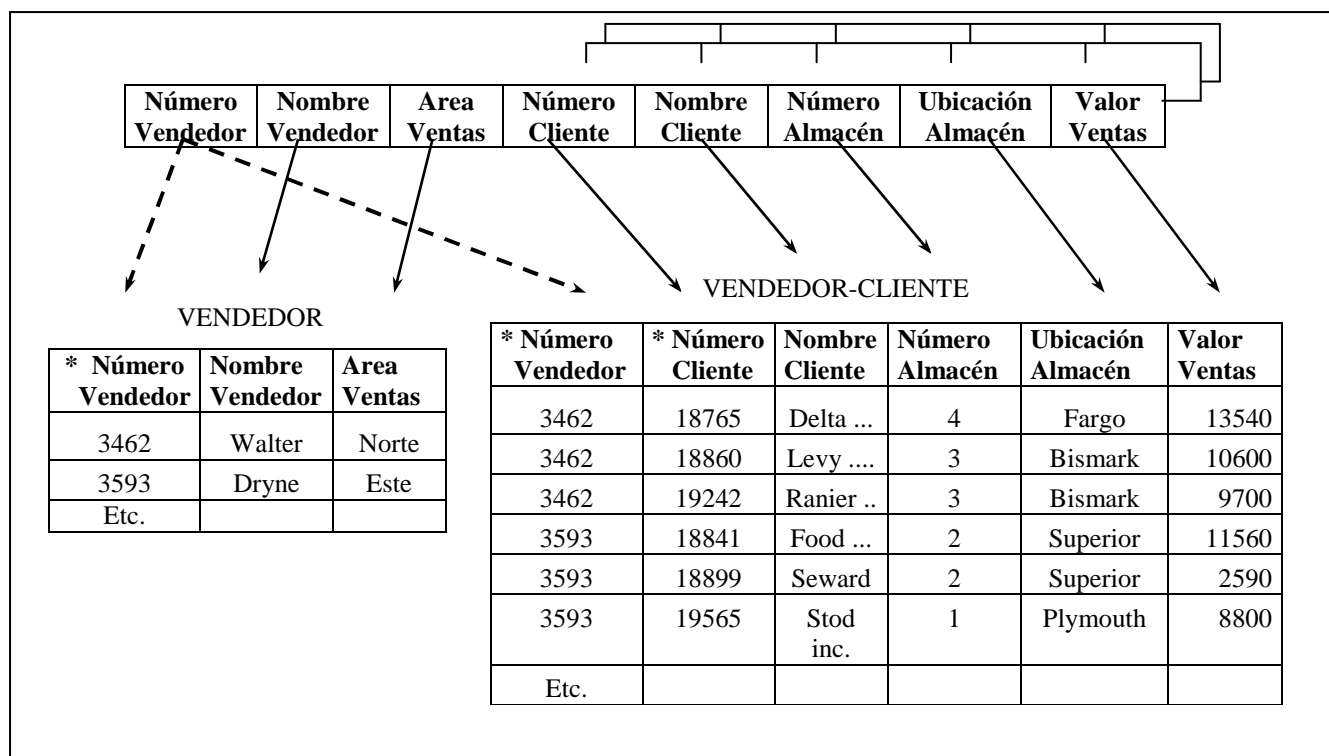
- ♦ No contener *grupos repetitivos*.
- ♦ Tener definida su *clave primaria*.

Grupo Repetitivo

Existe un grupo repetitivo cuando un atributo puede asumir más de un valor en una misma ocurrencia. Por ejemplo, en REPORTE-VENTA el atributo "Nombre de Cliente" puede asumir varios valores para un único "Vendedor" (o ocurrencia).

Clave Primaria

Llamamos clave a un conjunto no-vacío de atributos, que tiene la propiedad de identificar unívocamente y sin redundancia, a cada fila de una relación. **Clave simple** es aquella en la que el conjunto de atributos es unitario, en tanto que **clave compuesta** es aquel conjunto que reúne más de un atributo, concatenados.



La relación VENDEDOR-CLIENTE es una primera relación de normalización, pero no se encuentra en una forma ideal. En otras palabras, ciertos atributos no claves son dependientes sólo de "Número Cliente" y no del criterio concatenador. El diagrama de datos modelo muestra que "Valor Ventas" es dependiente, tanto de "Número Vendedor" como de "Número Cliente" , pero los otros tres atributos son solamente dependientes de "Número Cliente".

Segunda Forma Normal

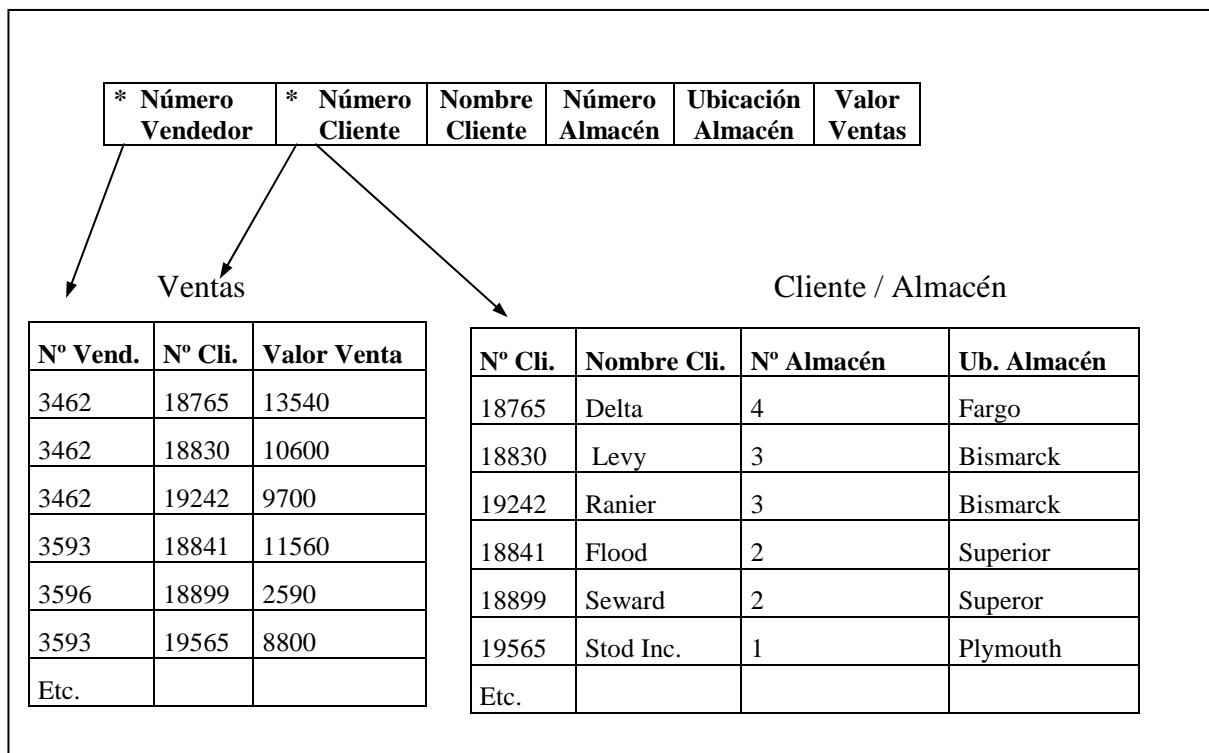
Para que una entidad se encuentre en **Segunda Forma Normal** debe cumplir dos condiciones:

- ♦ Encontrarse en *Primera Forma Normal*.
- ♦ No existir *dependencias funcionales* parciales respecto de la clave principal. Es decir, todo atributo que no forme parte de la clave primaria debe depender funcionalmente del total de la clave primaria.

Dependencia funcional

Se dice que un atributo B depende funcionalmente de un atributo A, si el valor que asume el atributo A determina el valor que asumirá el atributo B.

En nuestro ejemplo, deberemos entonces eliminar todas las dependencias parciales y colocarlas en otra relación. La relación VENDEDOR-CLIENTE se separa en dos nuevas relaciones, VENTAS y CLIENTE-ALMACEN.

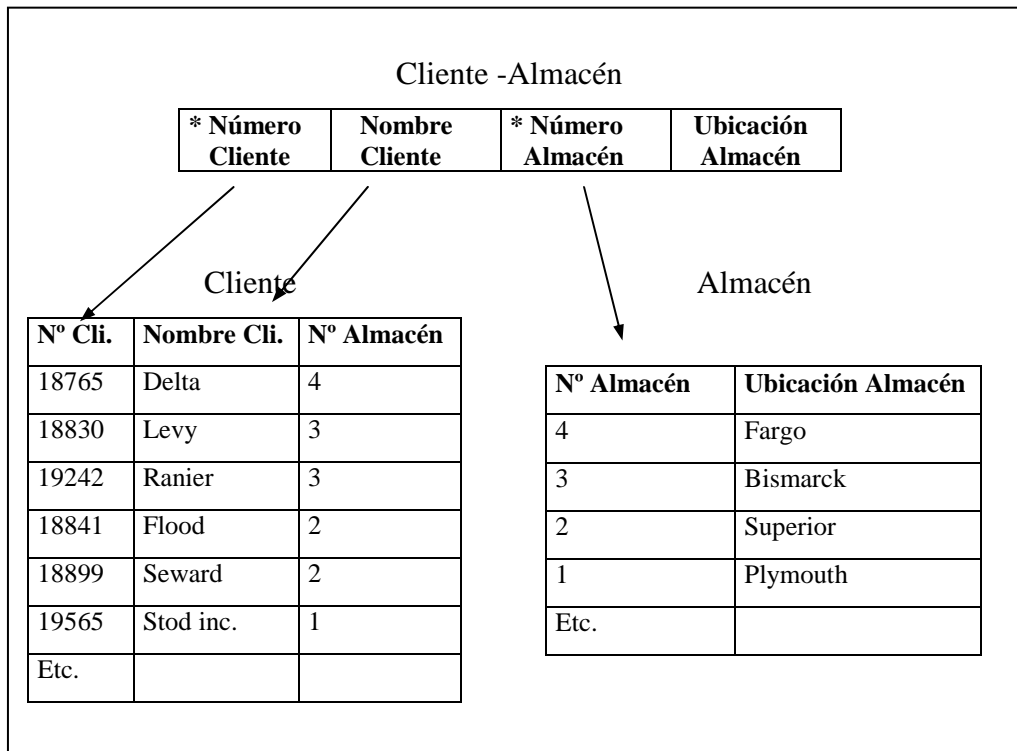


Tercera Forma Normal

Para que una entidad se encuentre en **Tercera Forma Normal** debe cumplir dos condiciones:

- ♦ Encontrarse en *Segunda Forma Normal*.
- ♦ No existir *dependencias funcionales transitivas*. Es decir, que no existan dependencias funcionales entre atributos que no formen parte de la clave primaria.

De manera similar a los pasos anteriores es posible descomponer la relación **CLIENTE – ALMACEN** en dos relaciones, como se muestra en el gráfico siguiente.



La clave para la relación CLIENTE es NUMERO DE CLIENTE y la clave para la relación ALMACEN es NUMERO DE ALMACEN.

Finalmente, la relación no normalizada REPORTE DE VENTAS se transforma en cuatro relaciones normales terciarias (FN3). Al revisar las relaciones que se muestran en la figura siguiente, uno puede observar que la relación sencilla REPORTE – VENTA se transforma en las siguientes cuatro relaciones:

