



# El Modelado de Datos

*Fundamentos básicos relacionados con el modelado de datos*

Jose Ignacio González Gómez

Departamento de Economía Contabilidad y Finanzas - Universidad de La Laguna

[www.jggomez.eu](http://www.jggomez.eu)

Ver. 2.1

## INDICE

|       |   |    |
|-------|---|----|
| 1     | Modelo de Datos .....   | 1  |
| 1.1   | <i>El modelado de datos como resultado final del proceso de ETL .....</i>                 | 1  |
| 1.2   | <i>Las bases de nuestros datos, los sistemas transaccionales ERP, CRM, TPV, etc .....</i> | 1  |
| 1.3   | <i>Creando nuestro modelo de datos en estrella. ....</i>                                  | 2  |
| 1.3.1 | <i>¿Qué quiero analizar? .....</i>  | 2  |
| 1.3.2 | <i>El modelo de estrella.....</i>   | 3  |
| 1.3.3 | <i>Ejemplo del proceso de negocio de las ventas y del de compras.....</i>                 | 3  |
| 2     | Que queremos analizar. Tablas de hechos y tablas de dimensiones.....                      | 5  |
| 2.1   | <i>Introducción.....</i>  | 5  |
| 2.2   | <i>Modelo Estrella vs Modelo Copo de Nieve.....</i>                                       | 5  |
| 2.3   | <i>Tabla de Hechos (Fact Tables) .....</i>  | 6  |
| 2.3.1 | <i>Características generales.....</i>   | 6  |
| 2.4   | <i>Tabla de Dimensiones.....</i>  | 6  |
| 2.5   | <i>Pases para seguir para crear nuestro modelo de datos. ....</i>                         | 7  |
| 2.6   | <i>Las relaciones entre tablas y la propagación de filtros.....</i>                       | 8  |
| 2.7   | <i>Ejemplos de Modelos de Datos por Procesos de Negocio .....</i>                         | 10 |
| 2.7.1 | <i>Ejemplo de ventas sencillo.....</i>  | 10 |
| 2.7.2 | <i>Ejemplo de ventas detallado.....</i>   | 10 |
| 2.7.3 | <i>Modelo de Datos – Financiero.....</i>  | 10 |
| 2.7.4 | <i>Modelo de recursos humanos sobre las horas trabajadas .....</i>                        | 10 |

# 1 Modelo de Datos

## 1.1 El modelado de datos como resultado final del proceso de ETL

En este tema nuestro principal objetivo es conocer la importancia y fundamento del modelado de datos, que es la base sobre la que se sustentan las analíticas e informes.

En muchas ocasiones nos podemos encontrar analíticas e informes basadas en una sola tabla lo que da entender que no se ha diseñado un modelo de datos correcto, perdiendo la oportunidad de disponer de mayor información y capacidad de análisis.

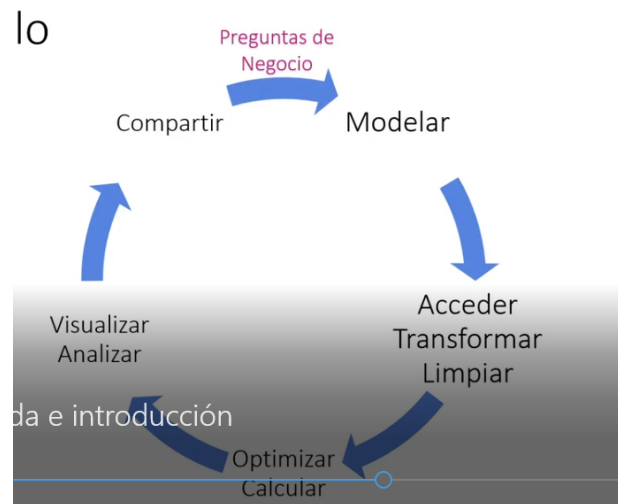
Por tanto nuestra área de atención está vinculada con el proceso ETL (Extract, Transform & Load), es decir obtener, transformar y cargar datos en la base de datos, y para ello estudiaremos los fundamentos de modelo dimensional “Esquema de Estrella).

De forma simplificada debemos de partir en primer lugar con unas preguntas de negocio clara y crear un modelo de datos (tablas y relaciones) que cuente con los datos necesarios que nos permita analizar y aproximarnos a dar respuesta a esas preguntas.

En los modelos de datos vamos a trabajar con un conjunto de nuevos conceptos que son fundamentales como son: Hechos – Dimensiones – Métricas - Claves Principales- Perspectivas- Segmentaciones- Copo de Nieve – Estrellas - etc..

Destacar que es necesario conocer estos términos y aplicarlos con rigor, esta será la base de nuestras analíticas e informes.

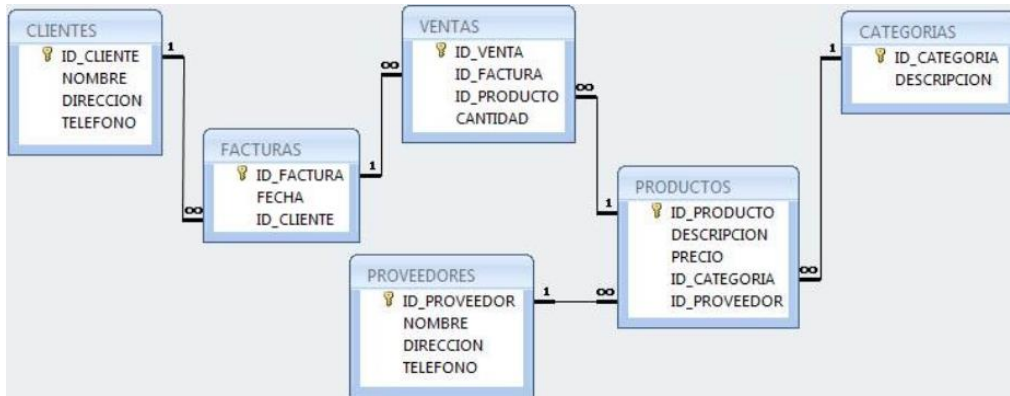
Presentamos finalmente el ciclo de desarrollo de una solución para el análisis de datos



En esta lo primero que necesitamos son las preguntas de negocio que nos va a permitir el hacer una estructura de tablas y relaciones acordes a las cuestiones planteadas y que nos facilite el diseño final de los informes.

## 1.2 Las bases de nuestros datos, los sistemas transaccionales ERP, CRM, TPV, etc.

Recordemos que las fuentes de información de la empresa se encuentran en diferentes aplicaciones como son los ERP, CRM, TPV, aplicaciones a medida, etc y estas contienen diferente información agrupada en tablas que deberían relacionarse para crear el modelo de datos.



### 1.3 Creando nuestro modelo de datos en estrella.

#### 1.3.1 ¿Qué quiero analizar?

El primer paso que debemos plantearnos en el momento de diseño de un modelo de datos es que quiero analizar con el mismo y esto implica una serie de pasos como son:

1. Identificar el proceso de negocio que se ve afectada para el análisis que queremos hacer, por ejemplo, ventas, compras, finanzas, gestión de stock, recursos humanos, etc..
2. Contar con una batería de preguntas a las que intentamos que el modelo nos de respuesta.
3. Identificar donde está la historia de lo que ha ocurrido, en que tablas, y que nivel de detalle necesito, campos que contiene la tabla
4. Y debemos también definir desde que perspectivas o puntos de vista queremos analizar la información, es decir por características del cliente, características del producto,

Estos son los cuatro pasos que debemos definir crear un modelo de datos de un proceso de negocio.

#### Veamos el siguiente ejemplo

- ✓ *Proceso de negocio – Ventas.* Contamos con el proceso de negocio en este caso ventas.
  - *Proceso y personas que necesita información* Analizamos que departamentos y personas necesitan información de el
  - *Batería de preguntas.* Hacemos las preguntas a cada persona y a cada departamento. Esta es la parte mas compleja y que mas tiempo requiere pero es la parte fundamental.
- ✓ *Determinamos la variable que queremos analizar y punto de vista.* Por ejemplo, la variable que queremos analizar son las ventas y el punto de vista que queremos que queremos, por ejemplo

| Cliente              | Fecha    | Tienda  | Producto | Empleado              |
|----------------------|----------|---------|----------|-----------------------|
| ¿A quién he vendido? | ¿Cuándo? | ¿Dónde? | ¿Qué?    | ¿Quién lo ha vendido? |

- ✓ *Que vamos a medir por cada venta, por cada hecho que ocurre.* Nos podría interesar el pvp, importe, descuento, coste, beneficio...
- ✓ *Desde que punto de vista vamos a analizar cada venta o hecho.*
  - Cliente: nombre, dirección, edad, salario, población, tipo, estado...
  - Fecha: día, hora, día de la semana, trimestre, es festivo, quincena, semestre, ...
  - Producto: categoría, ubicación, familia, tarifa...

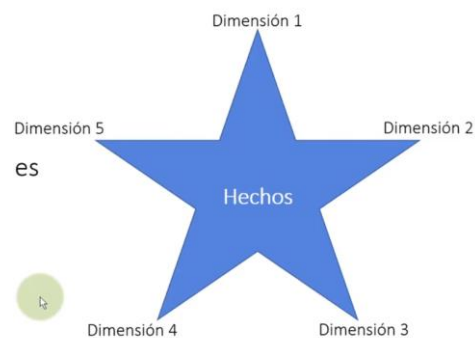
### 1.3.2 El modelo de estrella

tendremos que disponer de diferentes tablas que debemos crear y que contienen los datos necesarios para dar respuesta a las preguntas que nos surtan tal y como se muestra en la siguiente ilustración.

- Hechos ocurridos → Ventas → 1 Tabla
- Puntos de vista (Perspectivas)
  - Cliente → 1 Tabla
  - Fecha → 1 Tabla
  - Tienda → 1 Tabla
  - Producto → 1 Tabla
  - Empleado → 1 Tabla

Vemos que contamos con una tabla de hechos que refleja lo que ha ocurrido y 5 tablas de perspectiva, dimensión o punto de vista desde la que voy a analizar la información.

Con este conjunto de tablas podemos diseñar un modelo de estrella que es aquel que por su forma relacionamos las tablas anteriores siempre permaneciendo en el centro o eje principal la tabla de hechos y el resto de tablas que son las perspectivas desde que la quiero analizar se relaciona con la principal directamente o indirectamente formando una especie de estrella, de ahí su nombre, tantos picos de la estrella como dimensiones o perspectiva queramos a analizar.



Así cada vez que hagamos un diseño de una base de datos para nuestros análisis en Excel o Power BI, seguiremos este esquema de estrella definiendo cuales son los hechos y cual es la tabla o tablas que los almacena y por cada una de las perspectivas que queramos analizar tendremos una tabla de dimensión.

### 1.3.3 Ejemplo del proceso de negocio de las ventas y del de compras

Así para el ejemplo de análisis de las ventas vamos a plantarnos el modelo estrella a construir y para ello damos respuesta a las siguientes cuestiones:

**Que quiero analizar: Las ventas**

**Que ha ocurrido ~ Hechos ~ Diario de ventas o histórico de ventas (Tabla de hecho)**

**Puntos de vista (perspectivas) ~ Dimensiones~ Cliente, Fecha, Tienda... (Tablas de dimensiones)**

Así según el modelo estrella, nos planteamos que partiendo de la tabla de hecho diario de ventas o histórico de ventas que contienen diversos campos como son importe de la venta, descuento, fecha, tienda empleada, cliente y producto vamos a analizarla bajo 5 perspectivas tal y como muestra la ilustración.



Ahora supongamos que tenemos otro proceso de negocio como son las compras y en este caso el historial de hechos ocurridos está en mis tablas de compras y estas las vamos a analizar desde diferentes perspectivas como son el proveedor,

fecha, tienda, y producto es decir solo 4 dimensiones y por tanto tendríamos estas perspectivas.

Así y para este caso nos va a salir una estrella para cada proceso de negocio y las ventas contamos con 5 dimensiones y en las compras con solo 4.

Como conclusión hemos de tener presente siempre cuando vamos a trabajar con análisis de datos tanto en Excel como en Power BI, el modelar siempre en estrellas.



Concepto relacionado con modelado de datos son, normalización de bases de datos, modelo dimensional, Data Warehouse, Data Marts, etc.

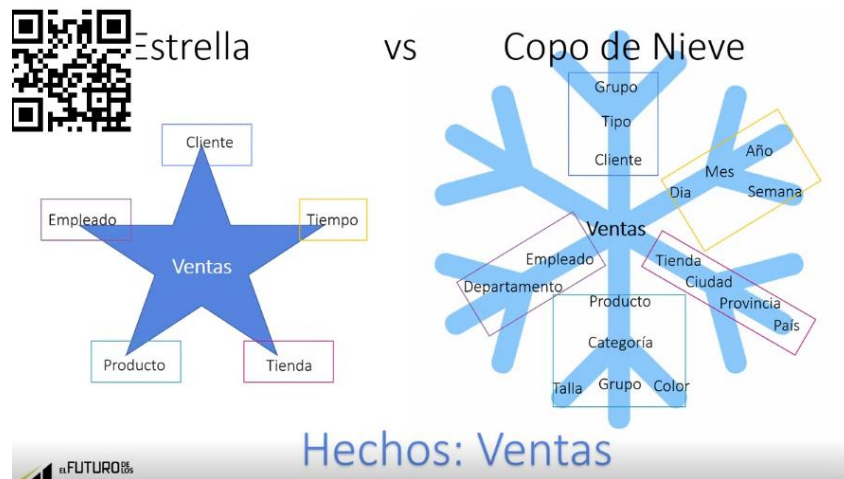
## 2 Que queremos analizar. Tablas de hechos y tablas de dimensiones

### 2.1 Introducción

En este apartado vamos a profundizar un poco más en lo visto anteriormente, veremos como a través de las Preguntas de Negocio, que es el pilar fundamental del análisis, determinamos tablas de hechos y dimensiones son necesarias y con que campos para dar respuestas a las preguntas planteadas.

### 2.2 Modelo Estrella vs Modelo Copo de Nieve

Vamos a poner un esquema comparativo entre estos dos modelos



Como podemos observar en ambos modelos tenemos una sola tabla de hechos, ventas, pero las tablas de dimensiones se configuran de forma distinta, así por ejemplo en la dimensión cliente, vemos que esa tabla en el modelo estrella está configurada por una sola dimensión en el modelo estrella y que contiene todos los campos necesarios, grupo, tipo, cliente. Mientras en el modelo Copo de Nieve, la dimensión cliente se configura por tres tablas relacionadas, una para el grupo, otra para el tipo y otra para cliente, es decir, el modelo Copo de Nieve incluye la subcategorías de las propiedades de los clientes.

El modelo copo de nieve es una estrella en que cada pico de la estrella se compone de varias tablas al contrario del modelo en estrella donde cada pico de la estrella es una tabla o hace referencia a una tabla.

Es decir en el modelo Copo de Nieves, colocó igualmente en el centro la tabla de hechos, ventas, y luego en cada una de las ramas del copo de nieve en cada dimensión relaciono todas las características vinculadas con esa dimensión, por ejemplo para la dimensión empleado, vinculo las dos características o propiedades vinculadas, empleado y departamento, mientras que en el modelo de estrella, solo utilizo una tabla de dimensión empleado que contiene todas las características o las dos propiedades anteriores.

Al final tendríamos los mismos datos, pero con una estructura diferente. Ambos modelos son muy buenas soluciones para el análisis de datos, pero en la medida de lo posible es mejor trabajar con el modelo de estrella, por diversas razones entre las que destacamos las siguientes:

1. Es mucho más simplificado, tiene una tabla por cada punta, punto de vista o perspectiva
2. Esto permite mayor rendimiento a la hora de trabajar
3. Mas claro para el usuario, mayor usabilidad.

## 2.3 Tabla de Hechos (Fact Tables)

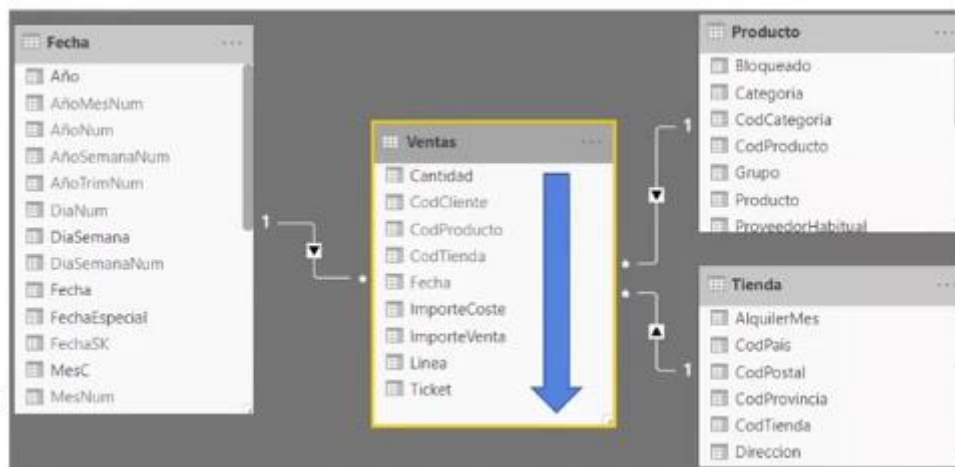
### 2.3.1 Características generales

Las historias de hechos ocurridos en la empresa es diversa ventas, compras, envíos,, incidencia, , llamadas (call center), asistencia al trabajo, apuntes contables, etc. y todas ellas se caracterizan por estar registrada en una tabla tipo de hechos

Pero debemos tener en cuenta que estos hechos o procesos (ventas, compras, etc) se pueden dividir en subprocesos, por ejemplo para el caso del proceso de Ventas este contiene los subprocesos de pedidos, albaranes, facturas, presupuestos, etc..

Por tanto la tabla de hechos de un proceso o subproceso contiene los datos que queremos medir, por ejemplo importe vendido, porcentaje de beneficio, etc.. y estas existen tanto en los modelos de estrellas como en el de copo de nieve,

En general la tabla de hechos se suele caracterizar por:



- Pocas columnas
- Casi todas de tipo numérico
- Contiene muchas filas o registros y con crecimiento periódico

## 2.4 Tabla de Dimensiones

Recordemos que las tablas de dimensiones son las que nos van a permitir segmentar y filtrar la información por muy diversos criterios. Es decir son los puntos de vista (perspectiva) desde la que queremos analizar al información.

Así las tablas de dimensión dan contexto a la tabla de hechos, así cuando preguntamos cuanto hemos vendido, esa variable se determina en la tabla de hechos pero, cuando, donde, a quien, etc.. en ese contexto me la da la tabla de dimensiones.

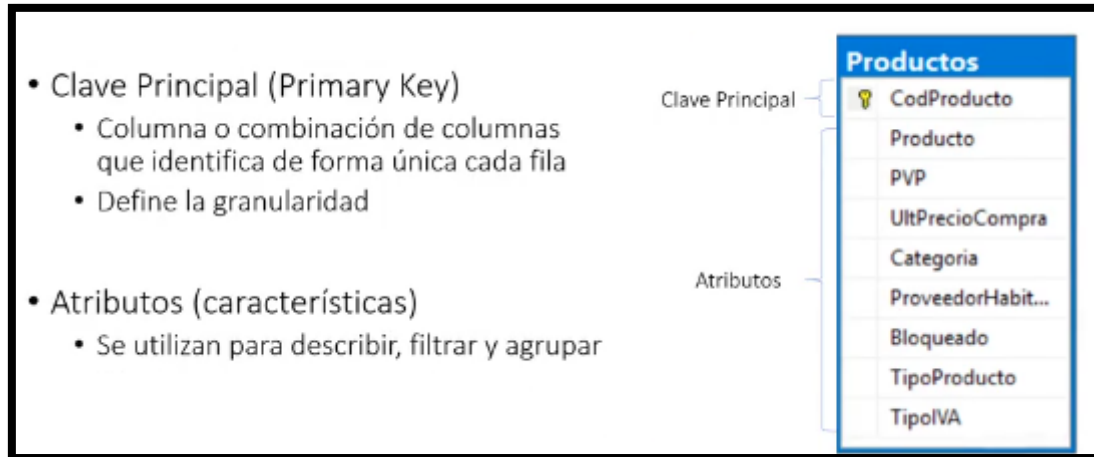
Por tanto las tablas de dimensión permite poner en contexto los hechos.



Una de las características de las tablas de dimensión es que normalmente contiene pocas filas en relación a los hechos, por ejemplo en el caso de la tabla de hechos de registro TPV contiene normalmente muchos mas registros que la tabla de dimensión de clientes, o que la tabla de dimensión forma de pago, o de proveedores, productos, almacenes, etc..

Respecto a los tipos de columnas que contiene podemos distinguir entre:

- Clave principal (Primary Key): columnas o combinación de columnas que identifican de forma única cada fila.
- Atributos o características: Se utilizan, para describir, filtrar y agrupar.



Además, las dimensiones pueden tener características denominada Jerarquías, que son grupos de atributos organizados en niveles, por ejemplo los productos se pueden agrupar en familias y subfamilias, las fechas en años, meses, trimestre, ciudad en provincia, región, país, etc. Por tanto, dentro de las dimensiones podemos definir jerarquías que nos va a permitir navegar de nivel mas detallado a más agregado o viceversa.

Debemos tener en cuenta que para crear una jerarquía es necesario modelar en estrellas.

## 2.5 Pasos para seguir para crear nuestro modelo de datos.

Como resumen, presentamos a continuación las fases o pasos a seguir para crear nuestro modelo de datos:

### 1. Creamos las tablas de hechos

- Definimos la clave principal de cada tabla
- Y el grado de desarrollo o nivel de detalle que vamos a tener
- Definimos las claves externas, tantas como dimensiones vamos a vincular después
- Definimos las medidas



### 2. Creamos las tablas de dimensión

- Una vez definida las tablas de hechos creamos las tablas de dimensión y definimos la clave principal.
- Definimos las columnas con las características a analizar.
- Defino la jerarquía necesaria.

### 3. Validar las relaciones entre hechos y dimensiones.



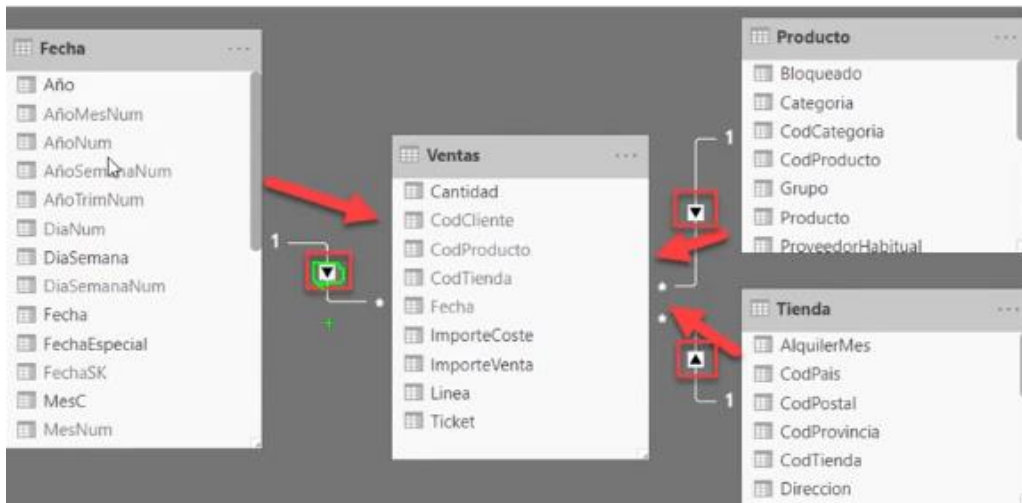
## 2.6 Las relaciones entre tablas y la propagación de filtros

En los modelos de datos basados en Power Pivot, las relaciones es mas que las relaciones en el mundo de las bases de datos relacionales, ya que en este entorno de Power Pivot las relaciones también afectan a los cálculos DAX por que afecta a la propagación de filtros.

Teniendo en cuenta estas consideraciones podemos señalar que las relaciones posibles en los modelos de datos son:

- Uno a muchos
- Uno a Uno
- Muchos a muchos

Aunque las más frecuentes en un modelo estrellas son las dos primeras y que tienen un carácter unidireccional, es decir, tal y como muestra el triangulito que define la dirección de la relación apuntando del 1 al asterisco que indica muchos.

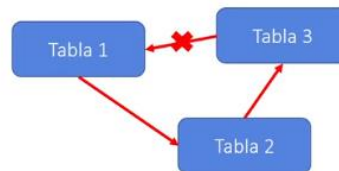


***Siempre en nuestro modelo de estrella esta relación va de la tabla de dimensión (1) hacia la tabla de hechos (\*) para garantizar correctamente la propagación de filtros a través de las relaciones.***

Es decir, cuando filtremos a través de un campo la tabla de dimensión se propague a la tabla de hechos correctamente.

Por tanto, y como hemos comentado al comienzo, el concepto de relación que nos ocupa no es el mismo que en las Bases de Datos Relacionales (SQL Server, Access...).

Señalar que en el caso que necesitemos más de una relación entre dos tablas, por ejemplo entre la fecha de venta y fecha de factura, destacar que “solo puede haber una relación activa entre dos tablas” aunque podre crear relaciones inactivas para usarlas en cálculos DAX y las podre activar en caso necesario.



Por tanto, no podremos tener dos caminos activos a la vez en la relación de dos tablas se han de evitar las llamadas referencias circulares.

En el siguiente ejemplo va a funcionar bien siempre que haya un camino correctamente establecido para llevar desde país filtro hasta total ventas,

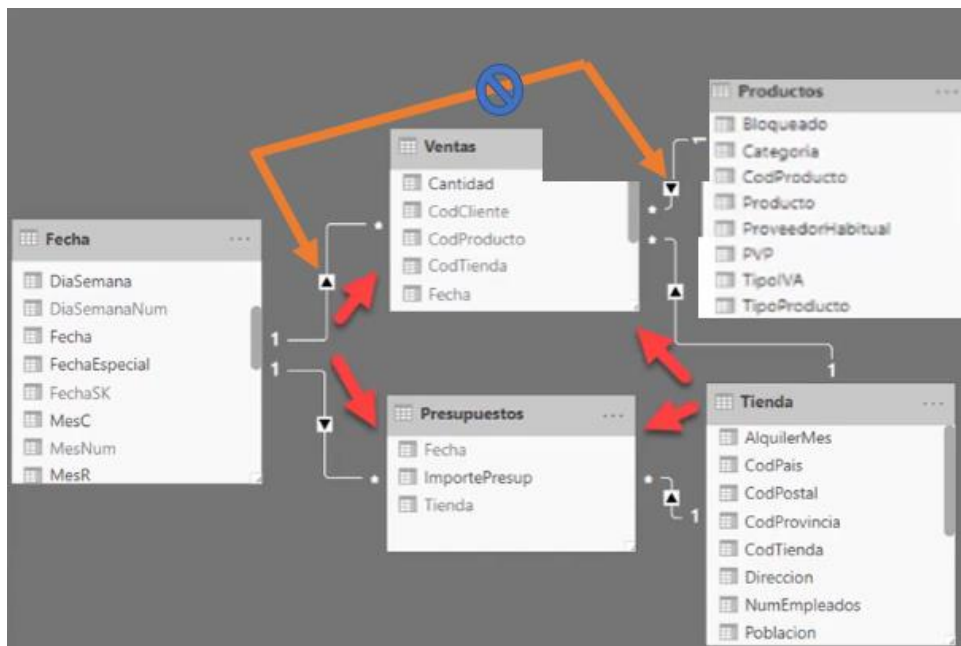
Si hay un camino único establecido de las relaciones 1 a \* y el triangulito esta siempre en ese sentido y nos lleva a esa conexión y extensión del filtro, en cuanto se rompa una de estas relaciones o no tenga el triangulito apuntado correctamente el filtro no se propagaría correctamente y causaría un error.

Veamos que tipos de limitaciones tienen estos conceptos de relaciones en estos modelos basados en Power Pivot.

- No es posible relaciones entre columnas de la misma tabla, aunque esta si puede hacerse en las clásicas bases de datos relacionales.
- No permite relaciones circulares, es decir más de un camino o relación para llegar de una tabla a otra



Hay que destacar que en el siguiente caso aparentemente tiene referencias circulares pero no las tiene por que no se propaga el filtro se rompe cuando la dirección o triangulo cambia de 1 a \* . a la de \* a 1.



Vemos que tenemos dos tablas de hechos (ventas y presupuestos) y dos tablas de dimensiones y una tabla especial de fechas o calendario. Si todas las direcciones fueran bidireccionales, por ejemplo de fechas pasando por presupuestos a tienda y de fechas pasando por ventas a tienda, esto nos daría una referencia circular, pero no se cumple por que como hemos señalado se rompe la propagación del filtro cuando la dirección o triangulo cambia de 1 a \* . a la de \* a 1.

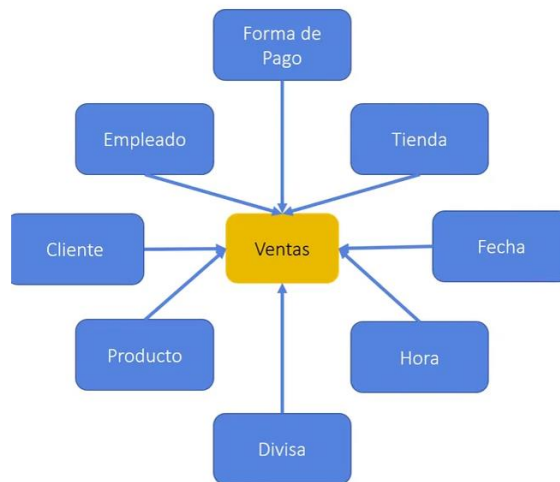
## 2.7 Ejemplos de Modelos de Datos por Procesos de Negocio

Para finalizar, mostraremos esquemas de modelos de datos aplicados a diversos Procesos de Negocio: Ventas – Finanzas-RRHH Horas Trabajadas

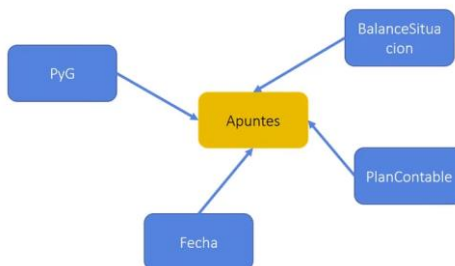
### 2.7.1 Ejemplo de ventas sencillo



### 2.7.2 Ejemplo de ventas detallado



### 2.7.3 Modelo de Datos – Financiero



### 2.7.4 Modelo de recursos humanos sobre las horas trabajadas

