



# 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

## Modelo de Datos

El modelado de datos como resultado final del proceso de ETL

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

Creando nuestro modelo de datos en estrella.

¿Qué quiero analizar?

El modelo de estrella

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

Que queremos analizar. Tablas de hechos y tablas de dimensiones

Modelo Estrella vs Modelo Copo de Nieve

Tabla de Hechos (Fact Tables)

Tabla de Dimensiones

Pases para seguir para crear nuestro modelo de datos.

Las jerarquías en las tablas de dimensión

Las relaciones entre tablas y la propagación de filtros

Ejemplos de Modelos de Datos por Procesos de Negocio

Ejemplo de ventas sencillo

Ejemplo de ventas detallado

Modelo de Datos – Financiero

Modelo de recursos humanos sobre las horas trabajadas

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

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.

## Creando nuestro modelo de datos en estrella

### ***¿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, etc

## 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...*

### **El modelo estrella**

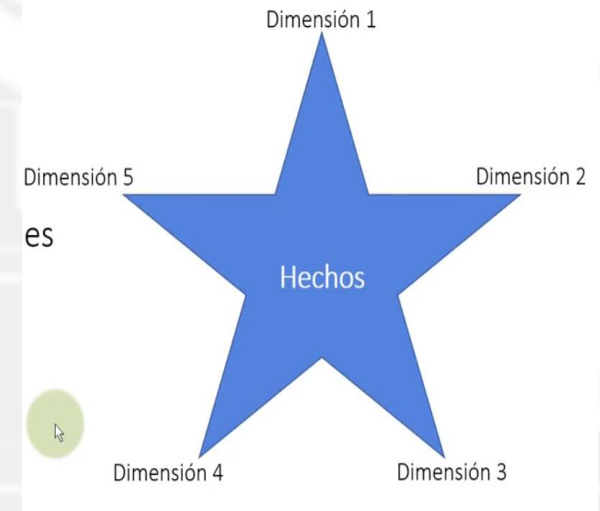
Al final tendremos que disponer de diferentes tablas que debemos relacionar y que contienen los campos necesarios para dar respuesta a las preguntas planteadas tal y como se muestra en la siguiente ilustración.

- 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



### Ejemplo del proceso de negocio de las ventas

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.

## Ejemplo del proceso de negocio de las compras

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.



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

### Tabla de Hechos (Fact Tables)

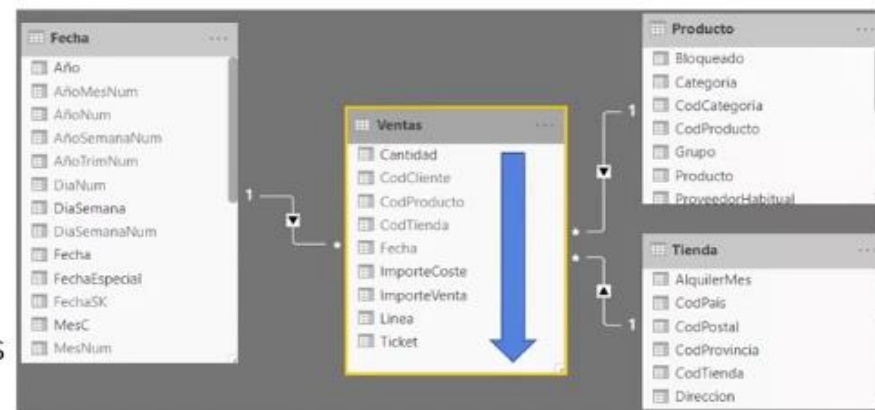
Las historias de hechos ocurridos en la empresa es diversa ventas, compras, envíos,, incidencia, , llamadas (call center), asistencia al trabajo, apuntes contables, etc. y todas ellas se caracterizan por estar registrada en una tabla tipo de hechos

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..

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



## Tabla de Dimensión

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 la 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..

## Sobre las jerarquías

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.

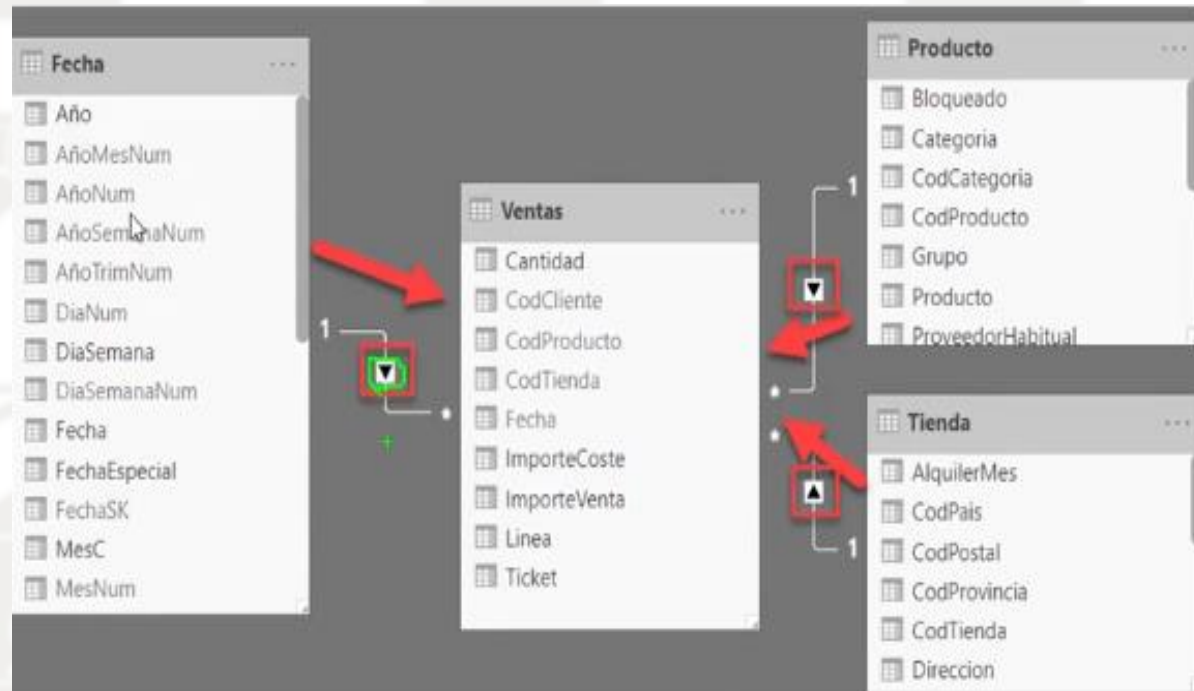
## Las relaciones entre tablas y la propagación de filtros

En los modelos de datos basados en Power Pivot, las relaciones son algo diferentes a las del mundo de las bases de datos relacionales, ya que en este entorno 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

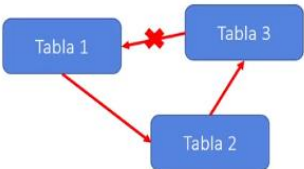
Aunque las más frecuentes en un modelo estrellas son las dos primeras y que tienen un carácter unidireccional, es decir, tal y como muestra el triangulito que define la dirección de la relación apuntando del 1 al asterisco que indica muchos.



***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.



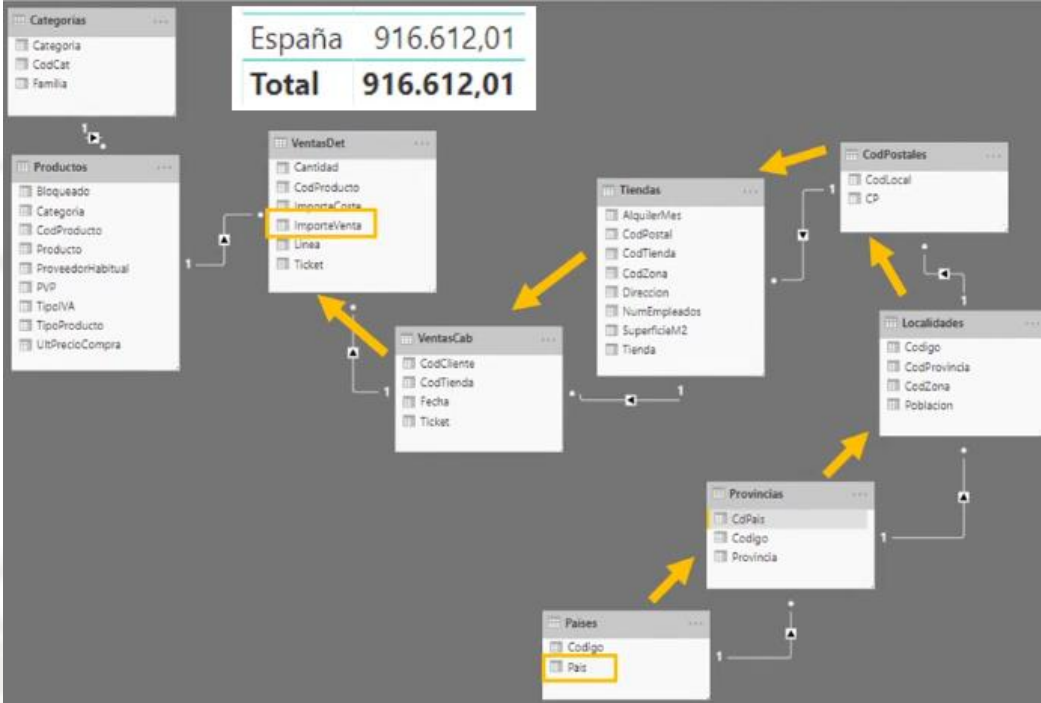
Por tanto, no podremos tener dos caminos activos a la vez en la relación de dos tablas se han de evitar las llamadas referencias circulares.

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.

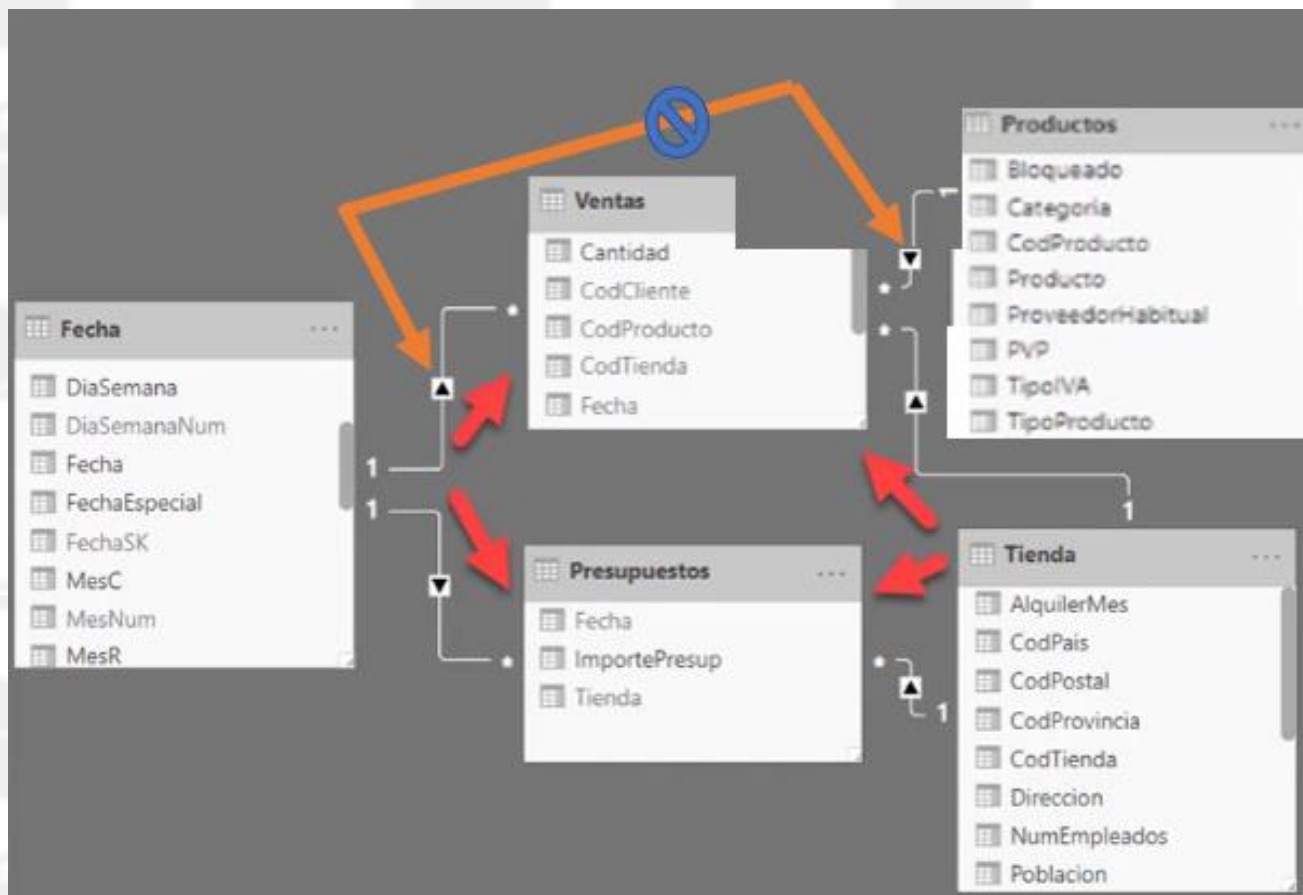
Veamos que tipos de limitaciones tienen estos conceptos de relaciones en estos modelos basados en Power Pivot.

- 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

# 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

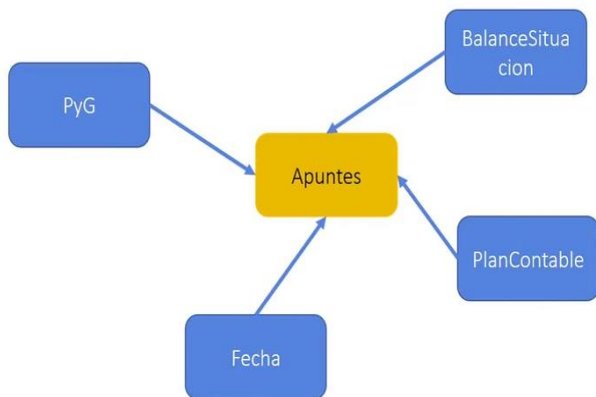
## Ejemplo de ventas sencillo



## Ejemplo de ventas detallado



## Ejemplo de modelo de datos financiero



## Ejemplo de modelo de recursos humanos

