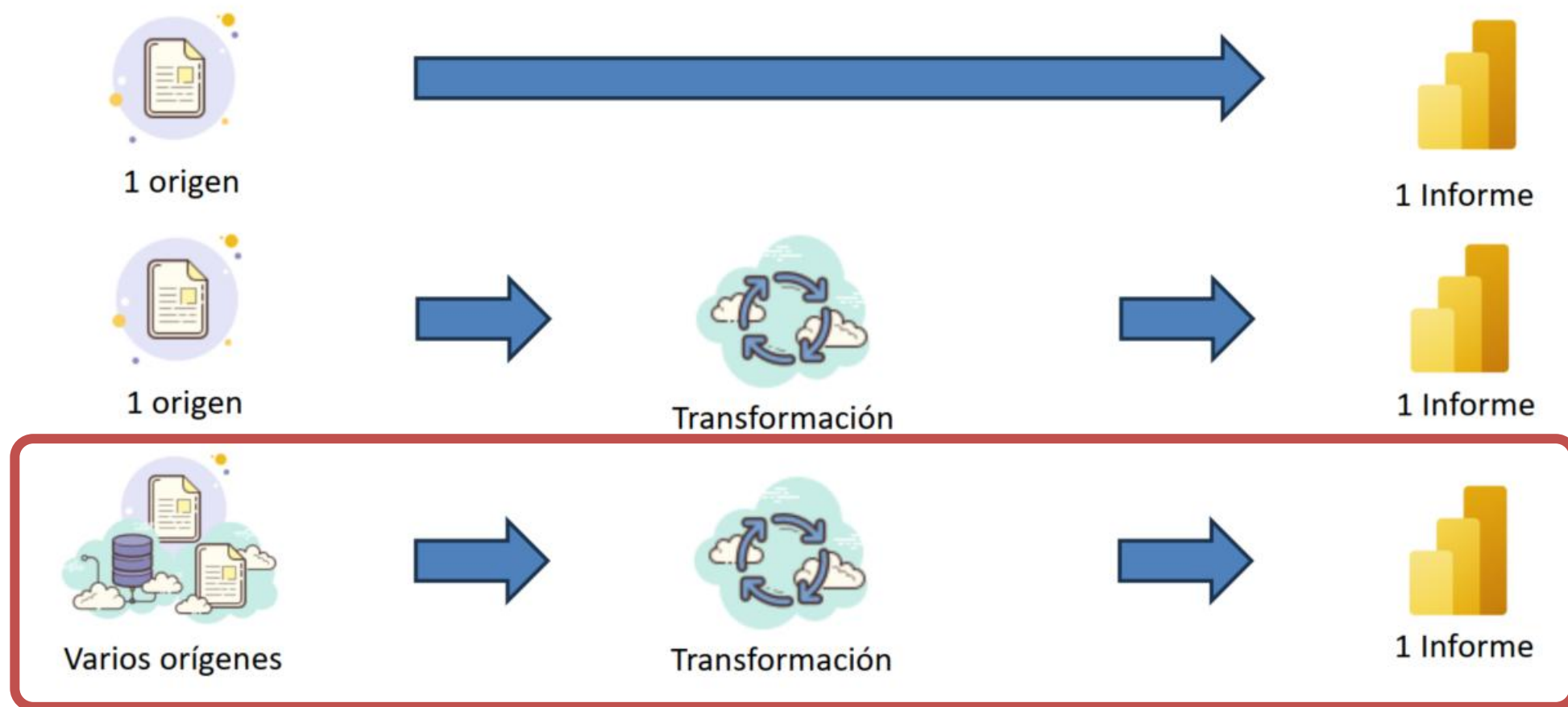


MODELADO DE DATOS EN POWER BI

Transformación digital con Power BI Básico y bases de datos relacionales

Rubén Hermoso Díez

¿Dónde estamos?



RELACIONES ENTRE TABLAS

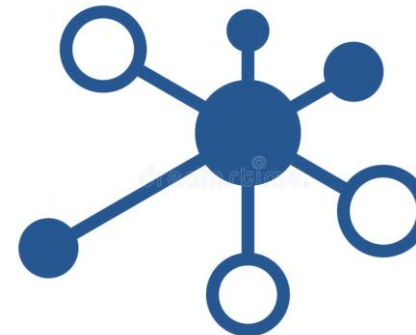
¿Qué hay detrás de un informe en Power BI?

Power BI organiza los datos en **tablas**, similares a las de Excel, pero con la posibilidad de conectarlas entre sí para analizar información de manera más eficiente.

- ◆ **Tablas:** Son conjuntos de datos organizados en columnas y filas. Cada tabla contiene información sobre un tema específico, como clientes, ventas o productos.



- ◆ **Relaciones:** Son los vínculos entre tablas que permiten combinar información sin repetir datos. Se crean conectando columnas en común, como un "ID de Cliente" en una tabla de ventas y en una tabla de clientes.



Ejemplo de relaciones entre tablas

💡 Ejemplo sencillo

Una tabla de "Ventas" solo guarda el **ID del producto** y la **cantidad vendida**, pero no el nombre ni el precio.

Una tabla de "Productos" tiene el **ID del producto**, el **nombre** y el **precio**.

Con una **relación** entre ambas, Power BI puede mostrar cuánto dinero se ha ganado con cada producto.

♦ Tabla Ventas

ID Venta	ID Producto	Cantidad Vendida
001	P001	3
002	P002	2
003	P001	1

♦ Tabla Productos

ID Producto	Nombre	Precio
P001	Camisa	20€
P002	Zapatos	50€

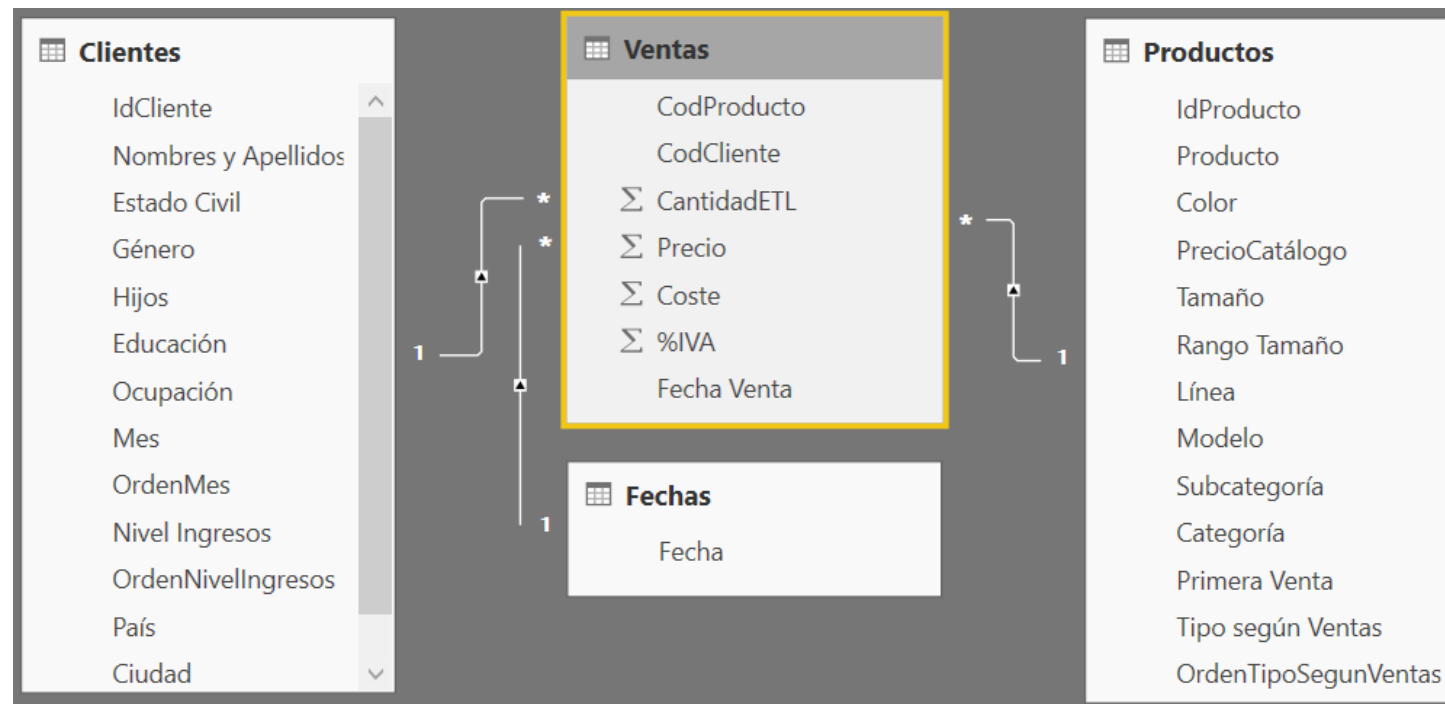
Ejemplo de relaciones entre tablas en Power BI

💡 Ejemplo sencillo

Una tabla de "Ventas" solo guarda el **ID del producto** y la **cantidad vendida**, pero no el nombre ni el precio.

Una tabla de "Productos" tiene el **ID del producto**, el **nombre** y el **precio**.

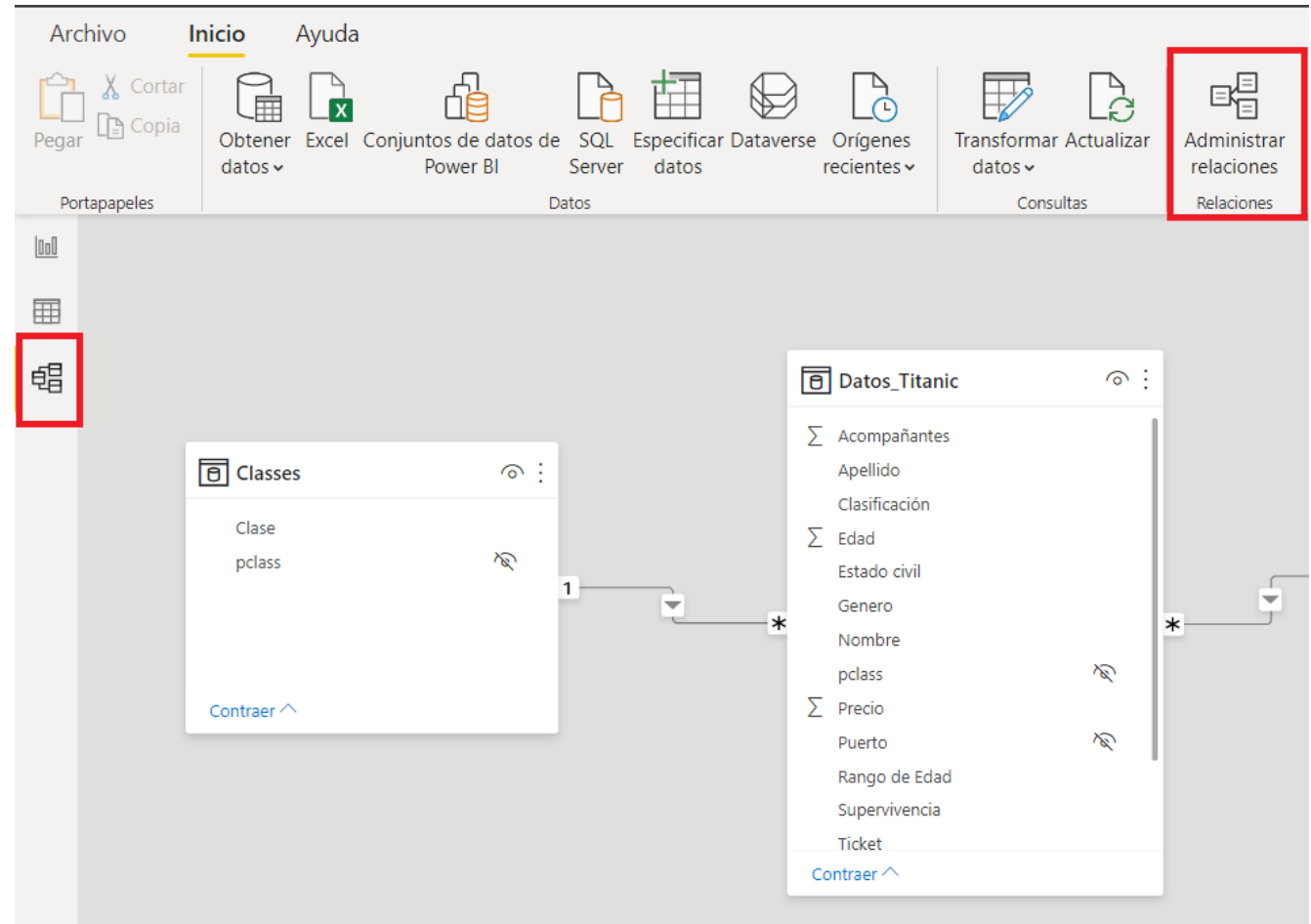
Con una **relación** entre ambas, Power BI puede mostrar cuánto dinero se ha ganado con cada producto.



Relacionamiento de tablas en Power BI

Para relacionar varias tablas, lo haremos desde la vista modelo.

Ahí, bien podremos arrastrar los campos que tengan datos comunes entre tablas diferentes y Power BI creará automáticamente la relación, o podremos crear relaciones a través del botón Administrar relaciones, que nos dará acceso a un Wizard en el que se nos ayudará a relacionar tablas.



Relacionamiento de tablas en Power BI

Para relacionar varias tablas, lo haremos desde la vista modelo.

Ahí, bien podremos arrastrar los campos que tengan datos comunes entre tablas diferentes y Power BI creará automáticamente la relación, o podremos crear relaciones a través del botón Administrar relaciones, que nos dará acceso a un Wizard en el que se nos ayudará a relacionar tablas.

×

Edit relationship

Select tables and columns that are related.

Sales

SalesOrderLineKey	ResellerKey	CustomerKey	ProductKey	OrderDateKey	DueDateKey	ShipDateKey
46638001	203	-1	333	20180718	20180728	20180725
46638002	203	-1	325	20180718	20180728	20180725
46642010	4	-1	321	20180720	20180730	20180727

Product

ProductKey	Product	Standard Cost	Color	List Price	Model	Subcategory	Category
210	HL Road Frame - Black, 58	\$868.63	Black	\$1,431.50	HL Road Frame	Road Frames	Compc
215	Sport-100 Helmet, Black	\$12.03	Black	\$33.64	Sport-100	Helmets	Access
216	Sport-100 Helmet, Black	\$13.88	Black	\$33.64	Sport-100	Helmets	Access

Cardinality

Many to one (*:1)

Cross filter direction

Single

☒ Make this relationship active

☐ Apply security filter in both directions

☐ Assume referential integrity

OK

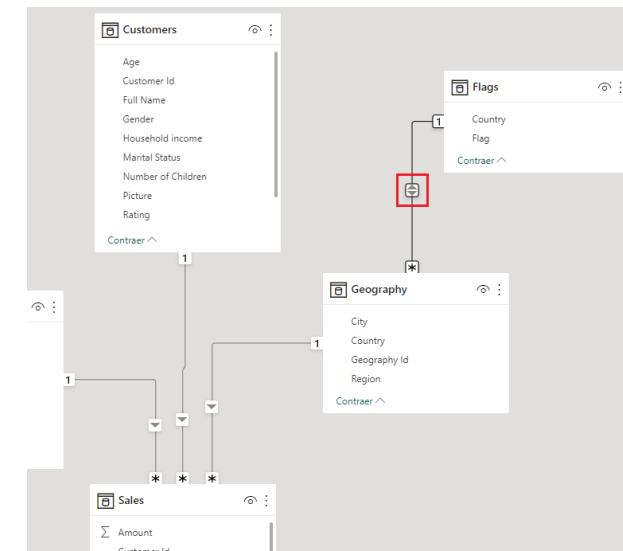
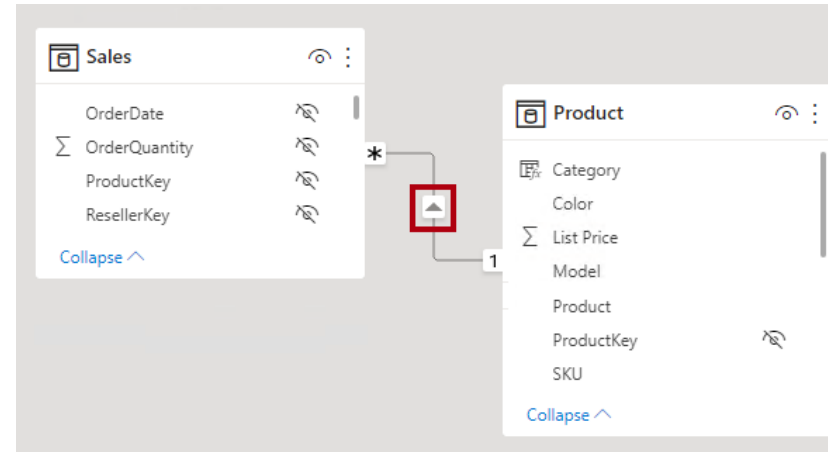
Cancel

Relacionamiento de tablas en Power BI

La **dirección del filtro** se refiere a cómo se aplican los filtros entre las tablas en un informe.

La dirección del filtro cruzado Único significa "**dirección única**" y Ambos se aplica a "**ambas direcciones**".

Una relación que filtra en ambas direcciones se describe normalmente como bidireccional.

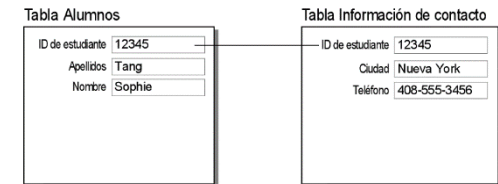


¿Qué cosas afectan a que un informe funcione rápido o lento?

Relación de Uno a Uno (1:1)

Cada fila en una tabla se relaciona con una única fila en otra tabla.

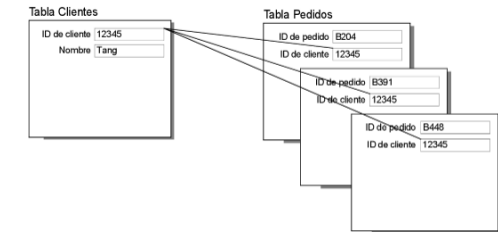
Cliente → Cuenta Bancaria



Relación de Uno a Muchos (1:*)

Una fila en una tabla se relaciona con varias filas en otra tabla.

Cliente → Transacciones

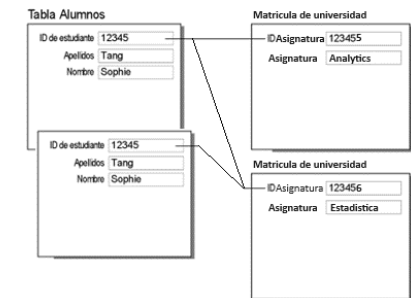


Relación de Muchos a Muchos (*:*)

Varias filas en una tabla se relacionan con varias filas en otra tabla

Cliente → Producto Financiero

*(un cliente, puede tener contratado varios productos
hay muchos clientes que se benefician de un mismo producto)*



INTRODUCCIÓN AL MODELADO DIMENSIONAL



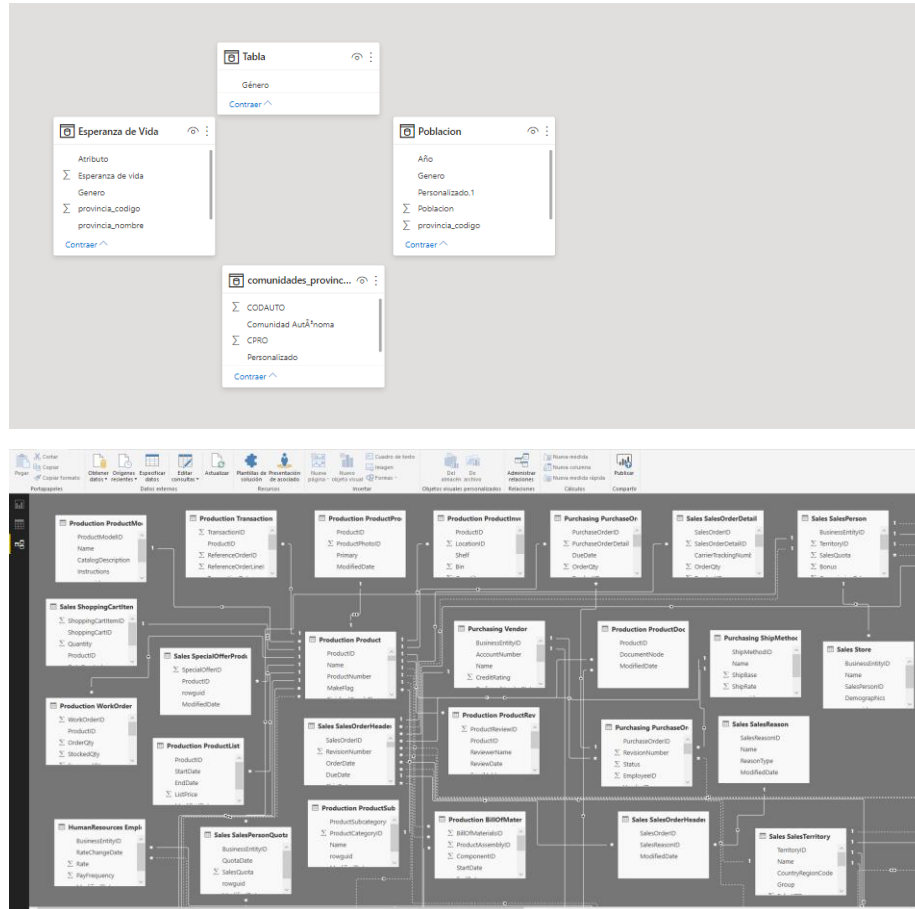
Concepto de Modelado de datos



Un modelo de datos es el conjunto de fuentes de datos, tablas o conjuntos de datos, que **al relacionarlos** nos permiten analizar de forma concreta un proceso, o procesos de negocio para poder sacar conclusiones de negocio.

El modelado de datos, son un conjunto de técnicas que permiten a los diseñadores de informe técnicos, preparar la información correctamente para que cualquier usuario consuma el informe cómodamente

¿Qué modelo tiene mejor pinta?



¿Cuándo es necesario un modelo de datos?

SIEMPRE, pero sobre todo es imprescindible cuando...

Millones de registros

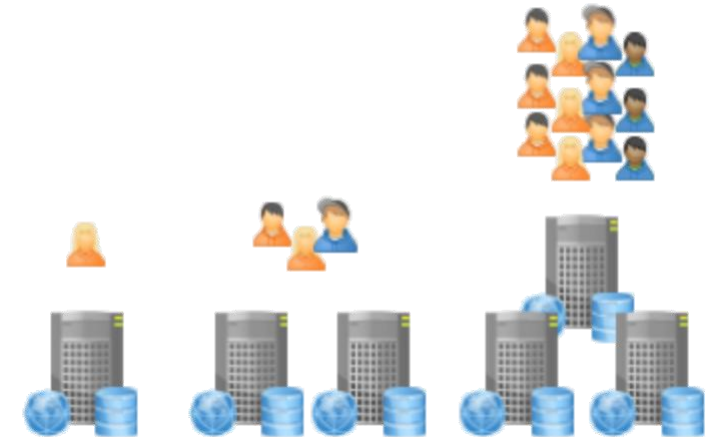
Cuando se manejan cantidades importantes de información es imprescindible que el modelo sea eficiente

Muchas tablas

Cuando se tienen varias tablas diferentes que contribuyen al modelo y se quiere enseñar información combinada

Previsión de crecimiento

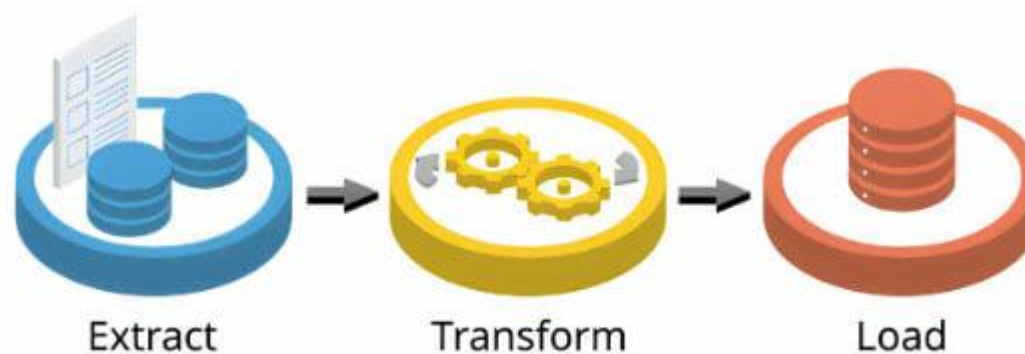
Cuando las fuentes de información empiezan a crecer y se prevé una gran cantidad de información en nuestra empresa, con necesidad de ser analizada.



Introducción al modelado dimensional

Durante el **proceso de modelado** es donde se establece la organización de la información y las relaciones entre tablas. Podríamos diferenciar las siguientes fases:

- Cargar tablas
- Transformar datos
- Relacionar tablas



Modelado de datos básico



Habitualmente una entidad de negocio se corresponde con una tabla

1 entidad negocio = 1 tabla

- **Fábrica:** Operarios, Lineas de producción, Producción de piezas, Máquinas...
- **Hospital:** Visitas, Quejas, Hospitales, Doctores, Medicamentos, Recetas...
- **Productos:** Ventas, Devoluciones, Vendedores, Clientes, Marcas, Productos...

Modelado de datos básico

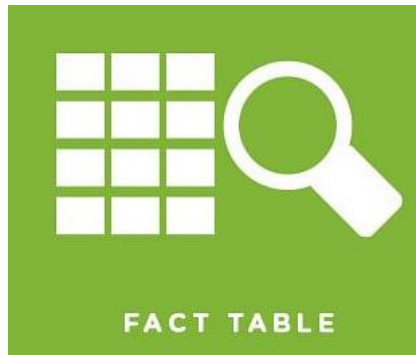
Tablas de Hechos

“Lo que voy a medir”

- **Fábrica:** Producción de piezas
- **Hospital:** Visitas, Quejas,
- **Productos:** Ventas, Devoluciones

Hace referencia a **eventos** que pueden **medirse**. Suelen tener millones de registros y pocas columnas.

Suele contener columnas de **IDENTIFICADOR** que servirán para unirse con las dimensiones (atributos del dato)



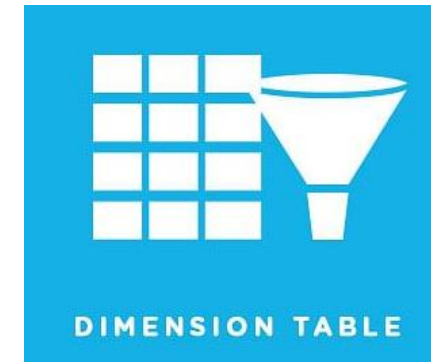
Tablas de dimensiones

“Por lo que voy a filtrar”

- **Fábrica:** Operarios, Lineas de producción, Máquinas...
- **Hospital:** Hospitales, Doctores, Medicamentos, Recetas...
- **Productos:** Vendedores, Clientes, Marcas, Productos...

Se usan para **filtrar o segmentar** las tablas de hechos.

Suelen responder a las preguntas ¿Qué? ¿Quién? ¿Cómo? ¿Cuándo? O ¿Dónde? Sucedió el evento medido. Hace referencia a datos poco cambiantes, maestros



Modelado de datos básico

Calendario

Fecha	Año	Mes	Día	Día de la Semana
2025-03-01	2025	3	1	Lunes
2025-03-02	2025	3	2	Martes
2025-03-03	2025	3	3	Miércoles

Productos

ProductoID	Nombre Producto	Categoría	Precio
1	Laptop X	Electrónica	500
2	Smartphone Y	Electrónica	300
3	Auriculares Z	Accesorios	50

Ventas

VentaID	Fecha	ProductoID	ClienteID	Cantidad	Total
1001	2025-03-01	1	101	2	1000
1002	2025-03-02	2	102	1	300
1003	2025-03-03	3	103	3	150

Clientes

ClienteID	Nombre Cliente	Ciudad	Edad
101	Juan Pérez	Madrid	32
102	Ana Gómez	Barcelona	28
103	Luis Rodríguez	Valencia	45

Modelado de datos básico

Fecha	Año	Mes	Día	Día de la Semana
2025-03-01	2025	3	1	Lunes
2025-03-02	2025	3	2	Martes
2025-03-03	2025	3	3	Miércoles

ProductoID	Nombre Producto	Categoría	Precio
1	Laptop X	Electrónica	500
2	Smartphone Y	Electrónica	300
3	Auriculares Z	Accesorios	50

VentaID	Fecha	ProductoID	ClienteID	Cantidad	Total
1001	2025-03-01	1	101	2	1000
1002	2025-03-02	2	102	1	300
1003	2025-03-03	3	103	3	150

ClienteID	Nombre Cliente	Ciudad	Edad
101	Juan Pérez	Madrid	32
102	Ana Gómez	Barcelona	28
103	Luis Rodríguez	Valencia	45

¿Podría analizar la tabla de ventas por vendedor?

Calendario

Fecha	Año	Mes	Día	Día de la Semana
2025-03-01	2025	3	1	Lunes
2025-03-02	2025	3	2	Martes
2025-03-03	2025	3	3	Miércoles

Productos

ProductoID	Nombre Producto	Categoría	Precio
1	Laptop X	Electrónica	500
2	Smartphone Y	Electrónica	300
3	Auriculares Z	Accesorios	50

Ventas

VentaID	Fecha	ProductoID	CienteID	Cantidad	Total
1001	2025-03-01	1	101	2	1000
1002	2025-03-02	2	102	1	300
1003	2025-03-03	3	103	3	150

Clientes

CienteID	Nombre Cliente	Ciudad	Edad
101	Juan Pérez	Madrid	32
102	Ana Gómez	Barcelona	28
103	Luis Rodríguez	Valencia	45

Vendedores

VendedorID	Nombre Vendedor	Región	Fecha Contratación
1	Carlos García	Norte	2022-06-15
2	Marta López	Sur	2021-11-01
3	Andrés Martínez	Este	2023-02-20



¿Podría analizar la tabla de ventas por vendedor?

Calendario

Fecha	Año	Mes	Día	Día de la Semana
2025-03-01	2025	3	1	Lunes
2025-03-02	2025	3	2	Martes
2025-03-03	2025	3	3	Miércoles

Productos

ProductoID	Nombre Producto	Categoría	Precio
1	Laptop X	Electrónica	500
2	Smartphone Y	Electrónica	300
3	Auriculares Z	Accesorios	50

Ventas

VentaID	Fecha	ProductoID	ClienteID	VendedorID	Cantidad	Total
1001	2025-03-01	1	101	1	2	1000
1002	2025-03-02	2	102	2	1	300
1003	2025-03-03	3	103	3	3	150

Clientes

ClienteID	Nombre Cliente	Ciudad	Edad
101	Juan Pérez	Madrid	32
102	Ana Gómez	Barcelona	28
103	Luis Rodríguez	Valencia	45

Vendedores

VendedorID	Nombre Vendedor	Región	Fecha Contratación
1	Carlos García	Norte	2022-06-15
2	Marta López	Sur	2021-11-01
3	Andrés Martínez	Este	2023-02-20



¿Podría analizar las ventas y las devoluciones?

Calendario

Fecha	Año	Mes	Día	Día de la Semana
2025-03-01	2025	3	1	Lunes
2025-03-02	2025	3	2	Martes
2025-03-03	2025	3	3	Miércoles

Productos

ProductoID	Nombre Producto	Categoría	Precio
1	Laptop X	Electrónica	500
2	Smartphone Y	Electrónica	300
3	Auriculares Z	Accesorios	50

Clientes

ClienteID	Nombre Cliente	Ciudad	Edad
101	Juan Pérez	Madrid	32
102	Ana Gómez	Barcelona	28
103	Luis Rodríguez	Valencia	45

Vendedores

VendedorID	Nombre Vendedor	Región	Fecha Contratación
1	Carlos García	Norte	2022-06-15
2	Marta López	Sur	2021-11-01
3	Andrés Martínez	Este	2023-02-20

Ventas

VentaID	Fecha	ProductoID	ClienteID	VendedorID	Cantidad	Total
1001	2025-03-01	1	101	1	2	1000
1002	2025-03-02	2	102	2	1	300
1003	2025-03-03	3	103	3	3	150

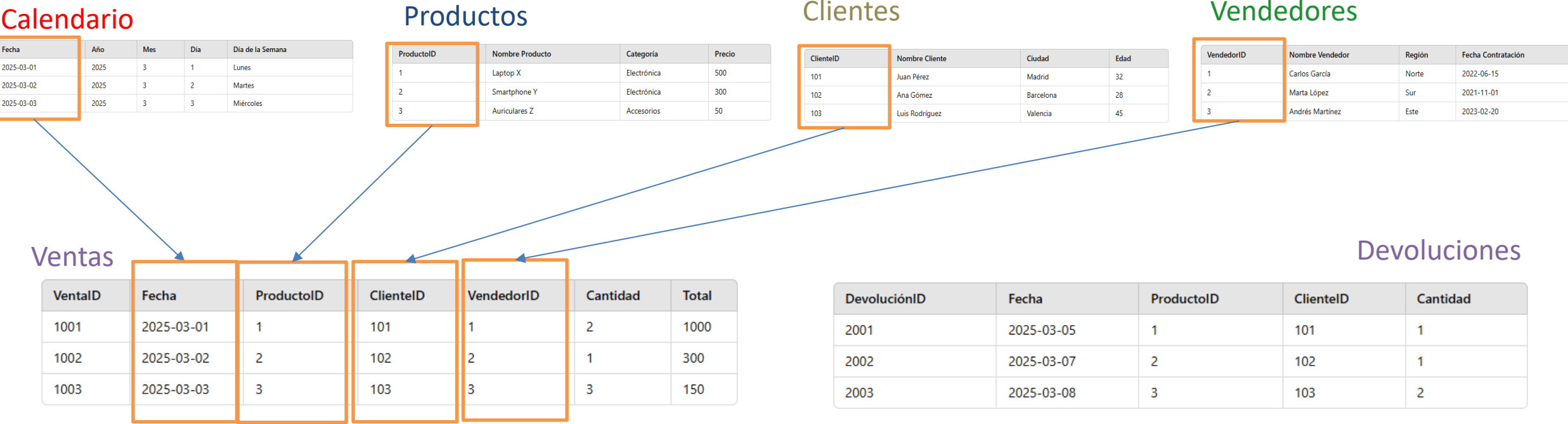
Devoluciones

DevoluciónID	Fecha	ProductoID	ClienteID	Cantidad
2001	2025-03-05	1	101	1
2002	2025-03-07	2	102	1
2003	2025-03-08	3	103	2

¿Por qué dimensiones podría analizar cada tabla?



¿Podría analizar las ventas y las devoluciones?



¿Por qué dimensiones podría analizar cada tabla?

¿Podría analizar las ventas y las devoluciones?

Calendario

Fecha	Año	Mes	Día	Día de la Semana
2025-03-01	2025	3	1	Lunes
2025-03-02	2025	3	2	Martes
2025-03-03	2025	3	3	Miércoles

Productos

ProductoID	Nombre Producto	Categoría	Precio
1	Laptop X	Electrónica	500
2	Smartphone Y	Electrónica	300
3	Auriculares Z	Accesorios	50

Clientes

ClienteID	Nombre Cliente	Ciudad	Edad
101	Juan Pérez	Madrid	32
102	Ana Gómez	Barcelona	28
103	Luis Rodríguez	Valencia	45

Vendedores

VendedorID	Nombre Vendedor	Región	Fecha Contratación
1	Carlos García	Norte	2022-06-15
2	Marta López	Sur	2021-11-01
3	Andrés Martínez	Este	2023-02-20

Ventas

VentaID	Fecha	ProductoID	ClienteID	VendedorID	Cantidad	Total
1001	2025-03-01	1	101	1	2	1000
1002	2025-03-02	2	102	2	1	300
1003	2025-03-03	3	103	3	3	150

Devoluciones

DevoluciónID	Fecha	ProductoID	ClienteID	Cantidad
2001	2025-03-05	1	101	1
2002	2025-03-07	2	102	1
2003	2025-03-08	3	103	2

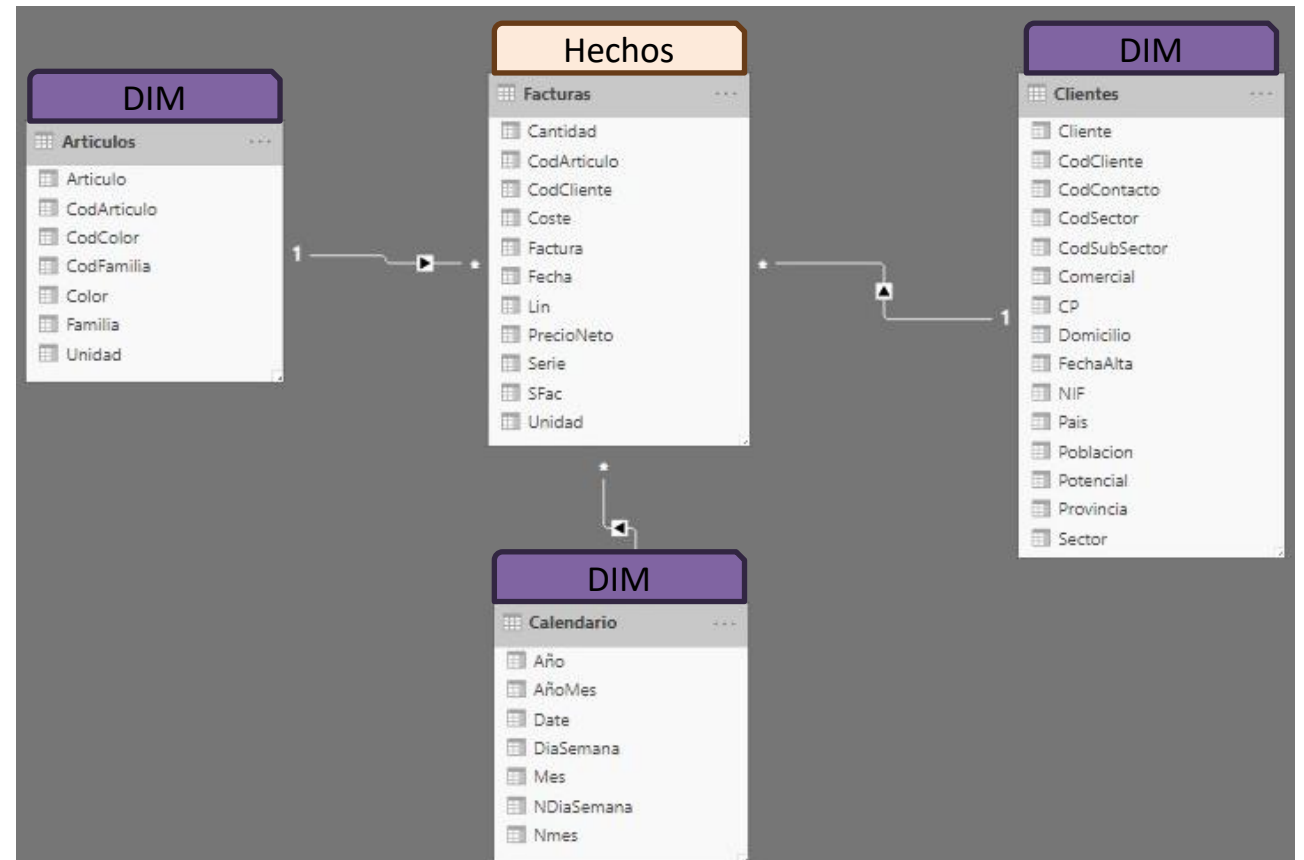
¿Por qué dimensiones podría analizar cada tabla?

Modelo en estrella

Con este esquema de tablas se pueden procesar y analizar grandes cantidades de información rápidamente.

Un BI emplea cierto tiempo en traducir las relaciones entre tablas. **Cuantas menos relaciones** y más rápido sea el camino a la tabla de hechos, **más rápidas son las consultas.**

No es recomendable tener una tabla central enorme tampoco

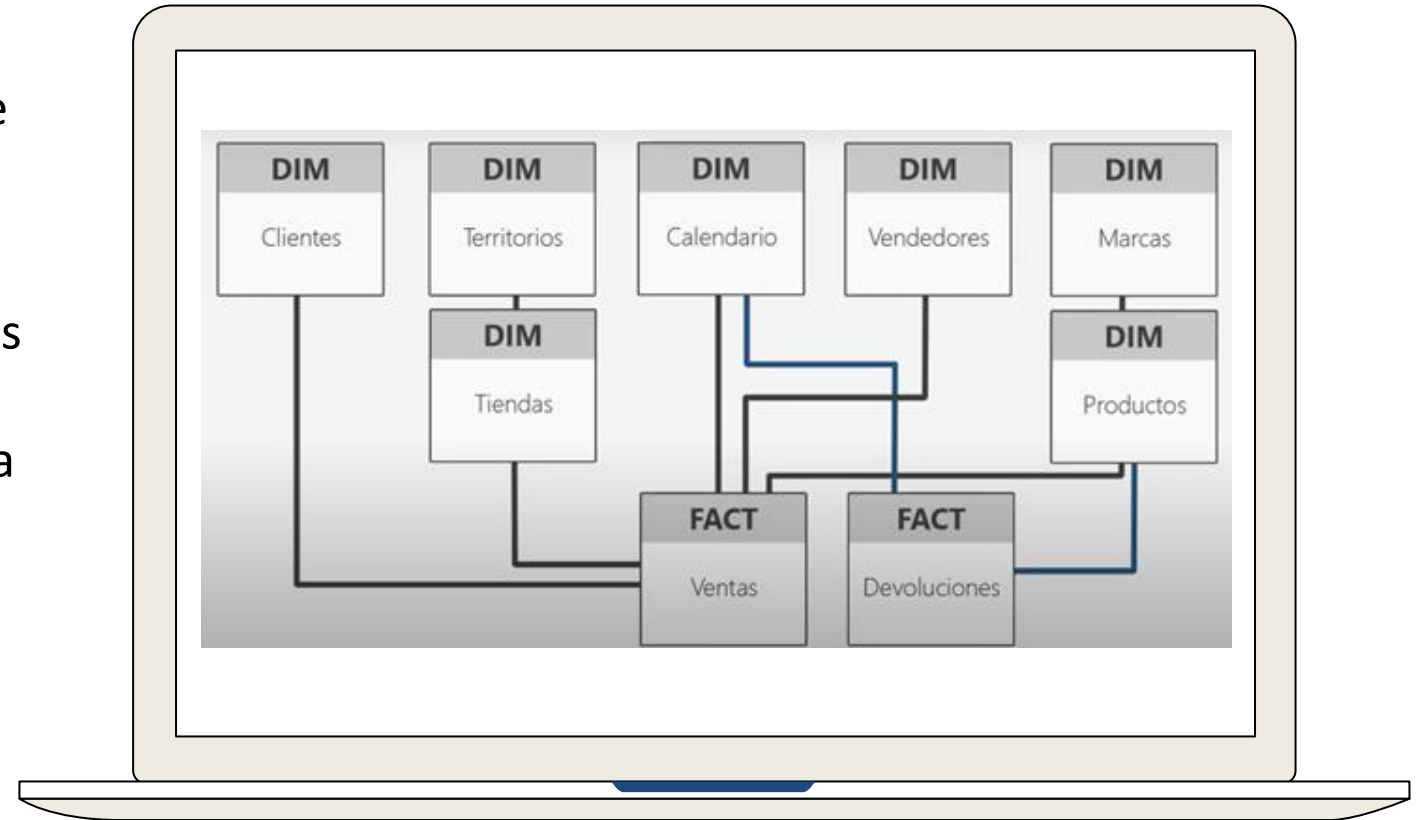


Modelo en estrella

Con este esquema de tablas se pueden procesar y analizar grandes cantidades de información rápidamente.

Un BI emplea cierto tiempo en traducir las relaciones entre tablas. **Cuantas menos relaciones** y más rápido sea el camino a la tabla de hechos, **más rápidas son las consultas.**

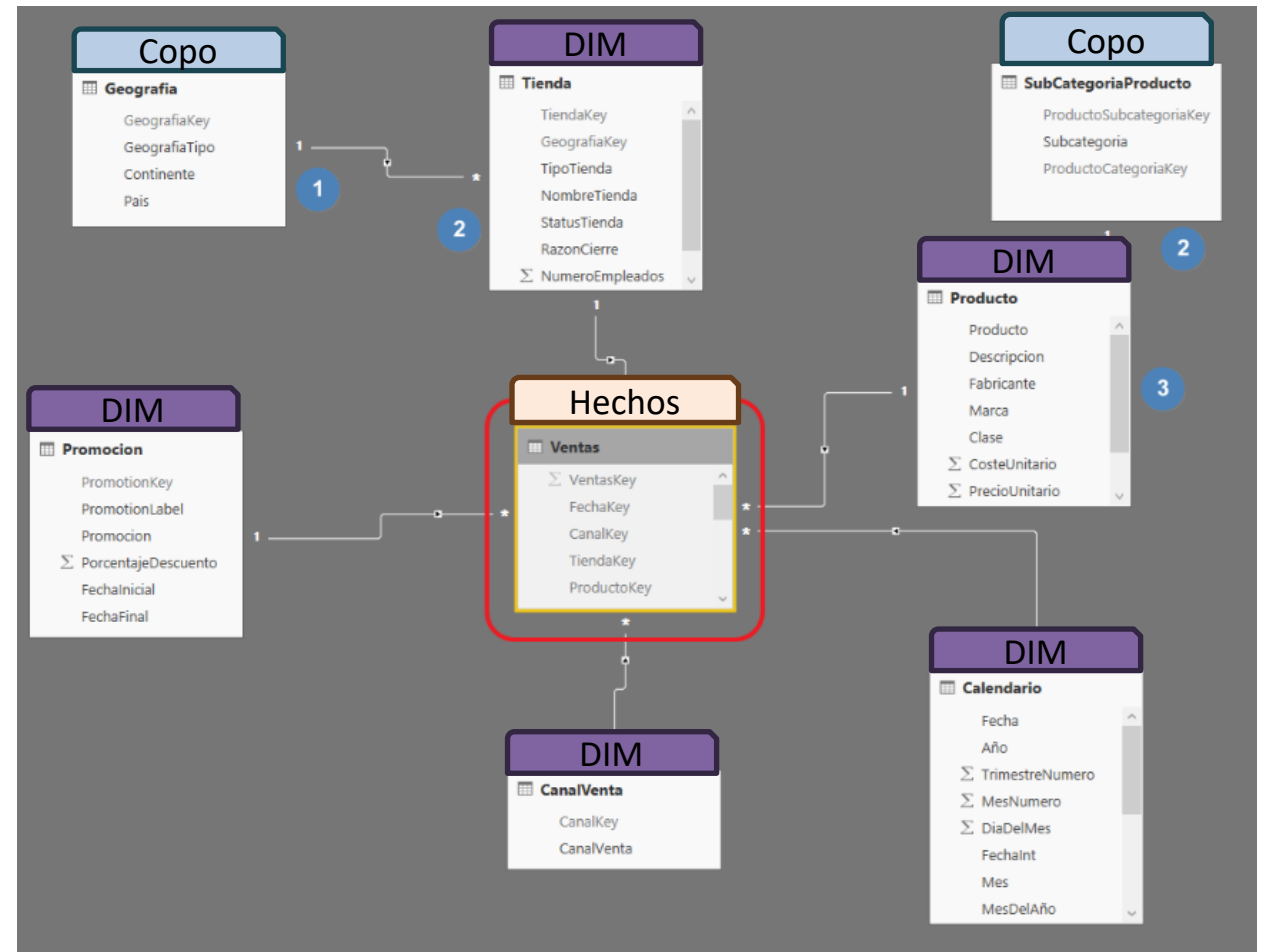
No es recomendable tener una tabla central enorme tampoco



Modelo en copo de nieve

Así como en el modelo estrella, al tener pocas tablas, es posible que tengamos bastante información repetida, en el modelo copo de nieve se intenta que cada dato exista una única vez.

De cara al uso de BI, se recomienda menos ya que la herramienta tiene que navegar a lo largo de todas las relaciones y eso implica invertir más tiempo, por tanto menor eficiencia.



¿Cómo funciona entonces Power BI?

Para entender bien el modelo ideal para el análisis en Power BI, haremos un símil con los diccionarios de toda la vida:



Modelo con una sola tabla de hechos

Con una sola tabla grande, cuando Power BI quiera buscar, siempre tendrá que leerla entera, registro por registro para poder mostrar los datos.

Imagina que además hay textos y son muy largos, igual que una persona, Power BI tendrá que invertir tiempo en leerlo.

Si la tabla es muy grande, puede ser que el informe tarde mucho en responder a ciertas consultas

¿Cómo funciona entonces Power BI?

Para entender bien el modelo ideal para el análisis en Power BI, haremos un símil con los diccionarios de toda la vida:



Modelo estrella

Si tenemos una tabla central y un diccionario. Conseguimos tener los eventos registrados en la tabla de hechos, y la información adicional y aquellas descripciones relevantes, aunque sean largas, las tendremos referenciadas con un ID, donde Power BI es capaz de buscar mediante **relaciones entre tablas**

¿Cómo funciona entonces Power BI?

Para entender bien el modelo ideal para el análisis en Power BI, haremos un símil con los diccionarios de toda la vida:



Modelo copo de nieve (con muchas relaciones entre si)

Aunque es bueno tener relaciones entre tablas y que Power BI haga búsquedas, lo importante es llegar a un equilibrio, si tenemos demasiadas tablas unidas entre si, se puede llegar a un punto en el que también le cueste a Power BI resolver las búsquedas de referencias

¿Cómo funciona entonces Power BI?

Para entender bien el modelo ideal para el análisis en Power BI, haremos un símil con los diccionarios de toda la vida:



¿Y la realidad en el día a día?

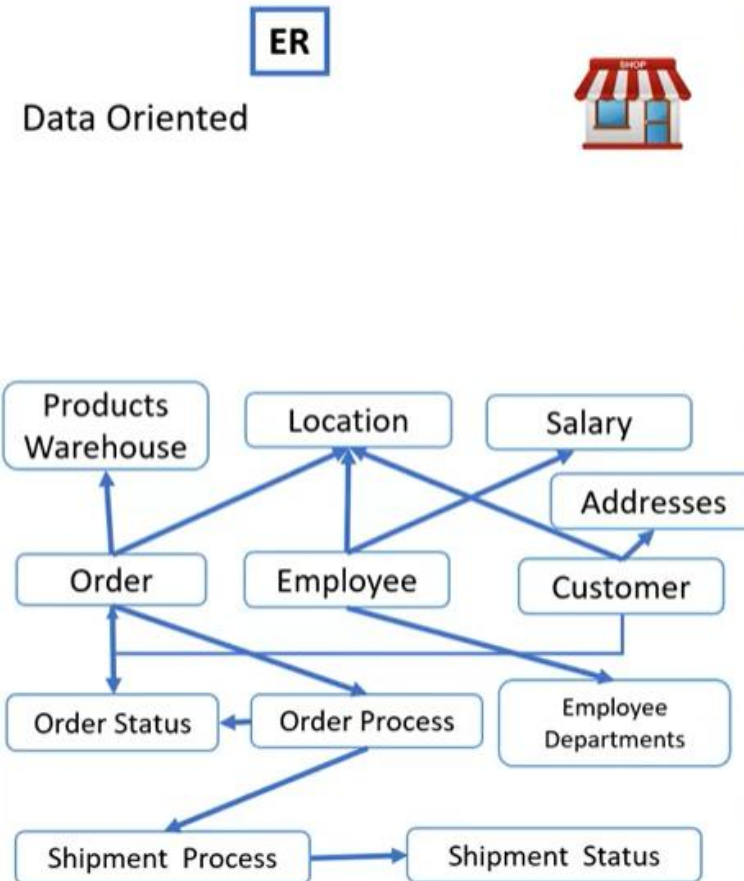
Haremos lo que podamos con lo que nos proporcionen. Los usuarios de negocio, muchas veces no estamos en esa posición de modelar o definir un modelo de datos, quizá incluso nos lo den hecho. O incluso para informes sencillos tengamos que usar una tabla grande con todos los datos.

Lo importante es no preocuparnos y entender nuestro caso. Y si se trata de un informe con mucha información, ahí si necesitaremos tener un buen modelo

CASOS DE USO DE MODELADO

¿Cuál es el escenario habitual en una base de datos?

ER vs Dimensional



Muchas tablas

En muchas ocasiones, los conceptos de negocio no son complicados de entender pero si de encontrar

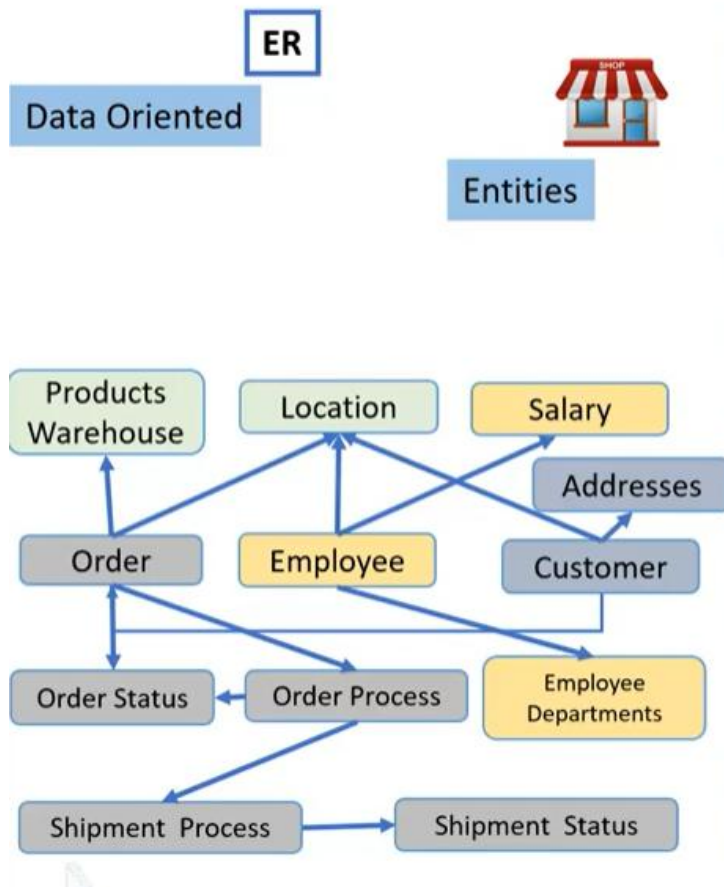
Nombres técnicos

En muchas ocasiones las bases de datos no permiten poner nombres intuitivos a las columnas

Optimizado para la introducción de datos

¿Cuándo es necesario un modelo de datos?

ER vs Dimensional



¿Cuál es el escenario habitual en una base de datos?



Sencillo de entender

Simplificado y agrupado en entidades de negocio con sentido para los usuarios

Nombres adaptados

Se intenta usar nombres con los que el usuario está familiarizado siempre que sea posible

Optimizado para el análisis de datos en herramientas de visualización

Ejercicio: modelado dimensional

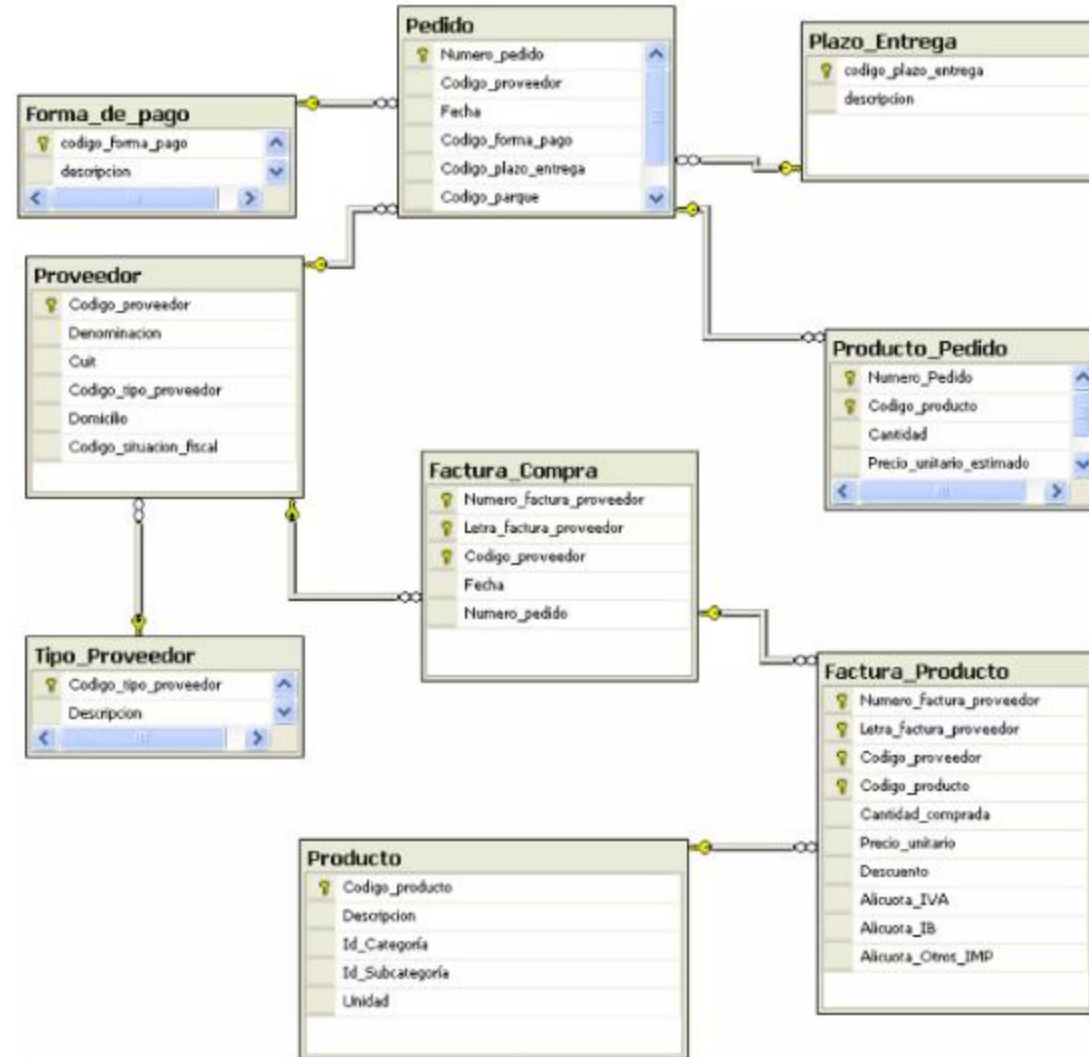
Vamos a modelar un departamento de compras. En concreto vamos a analizar los pedidos realizados a los proveedores y las facturas recibidas. Donde un pedido puede ser generado sin que exista una factura asociada de inmediato

Los usuarios nos han pedido analizar la información de los Pedidos y Facturas:

- **Pedidos:** permite evaluar la demanda de productos y el comportamiento de los proveedores antes de que se realice la facturación. Se pueden responder preguntas como:
 - ¿Cuántos pedidos se han generado en un período determinado?
 - ¿Qué productos han sido solicitados con mayor frecuencia?
- **Facturas:** El análisis de facturas permite evaluar el gasto real y las condiciones comerciales aplicadas por los proveedores. Se pueden responder preguntas como:
 - ¿Cuánto se ha facturado en comparación con lo pedido?
 - ¿Se han aplicado correctamente descuentos e impuestos en la facturación?



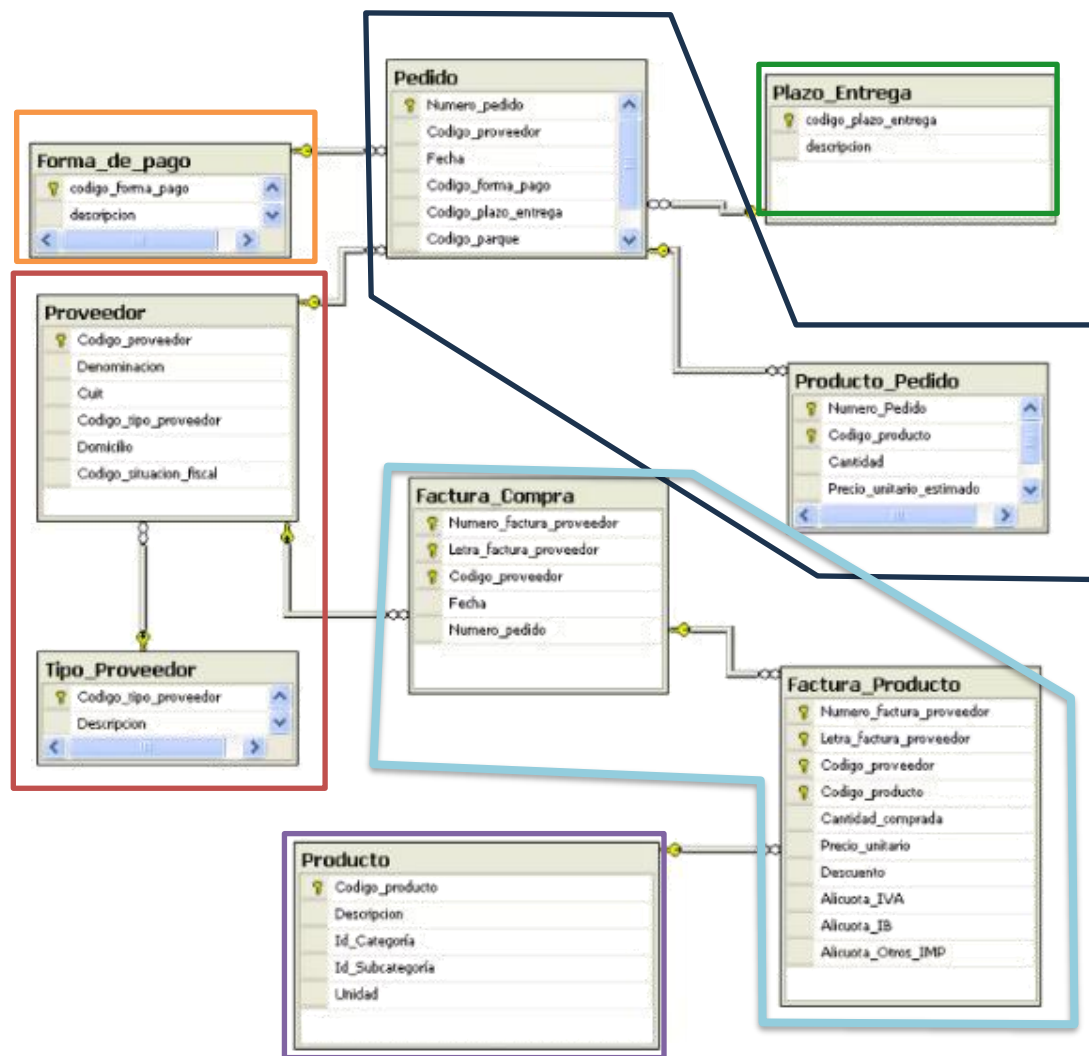
Ejercicio: modelado dimensional



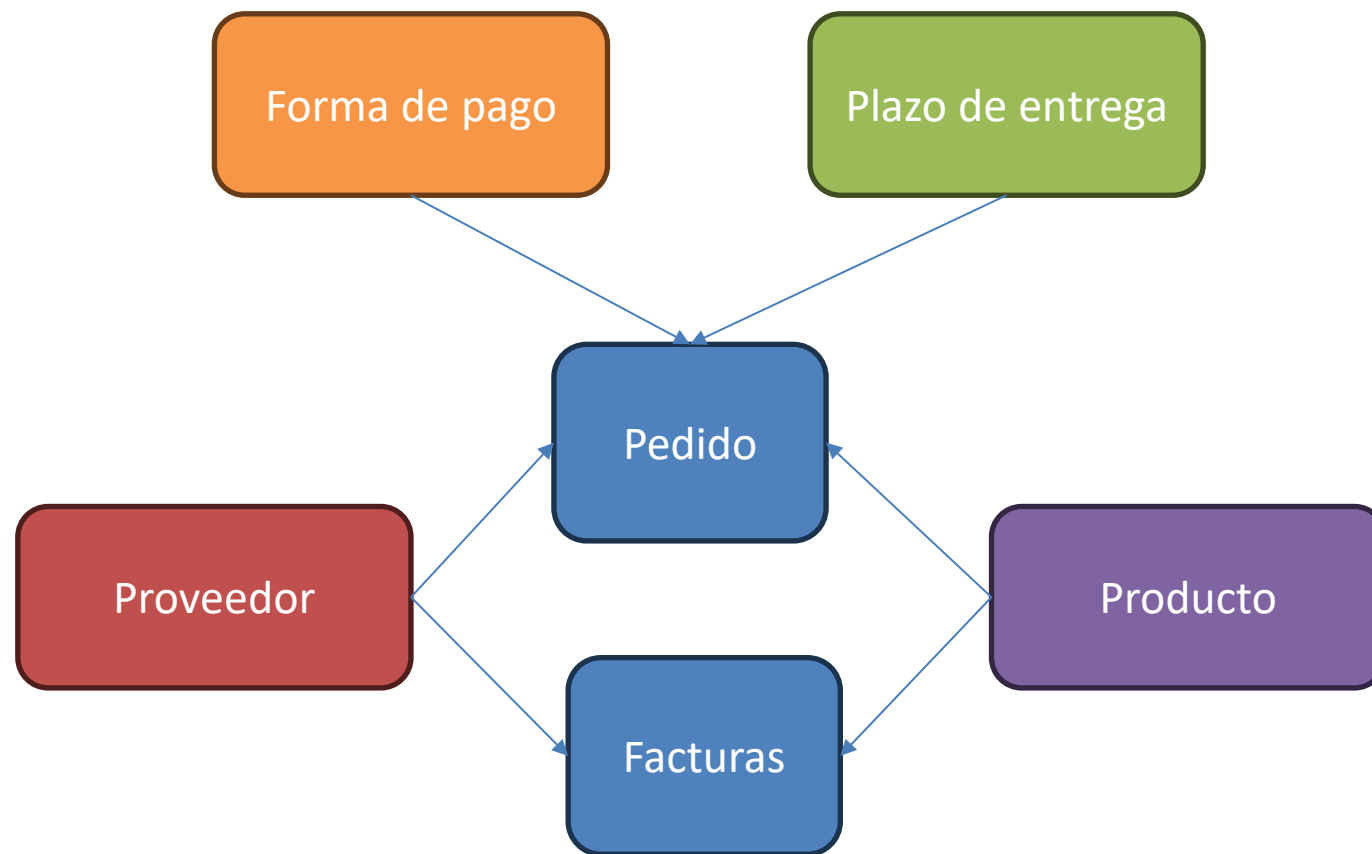
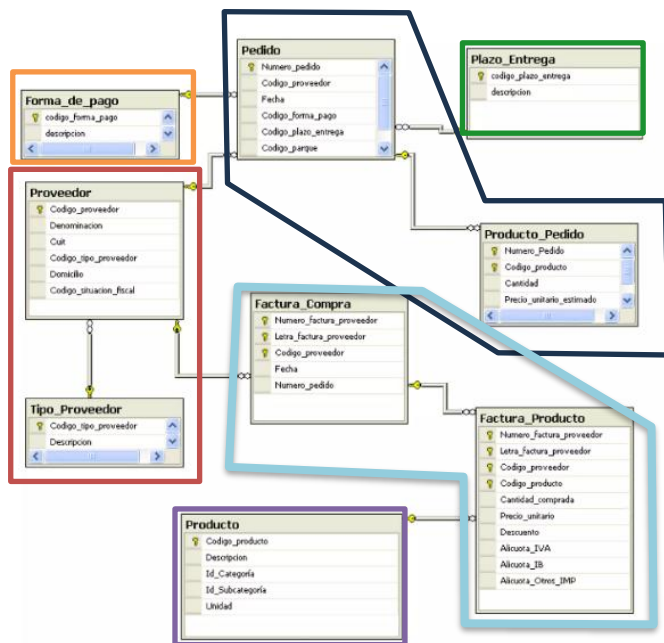
Agrupar por
entidades de
negocio que
tengan cosas en
común

¿Qué tablas tienen cosas en común?

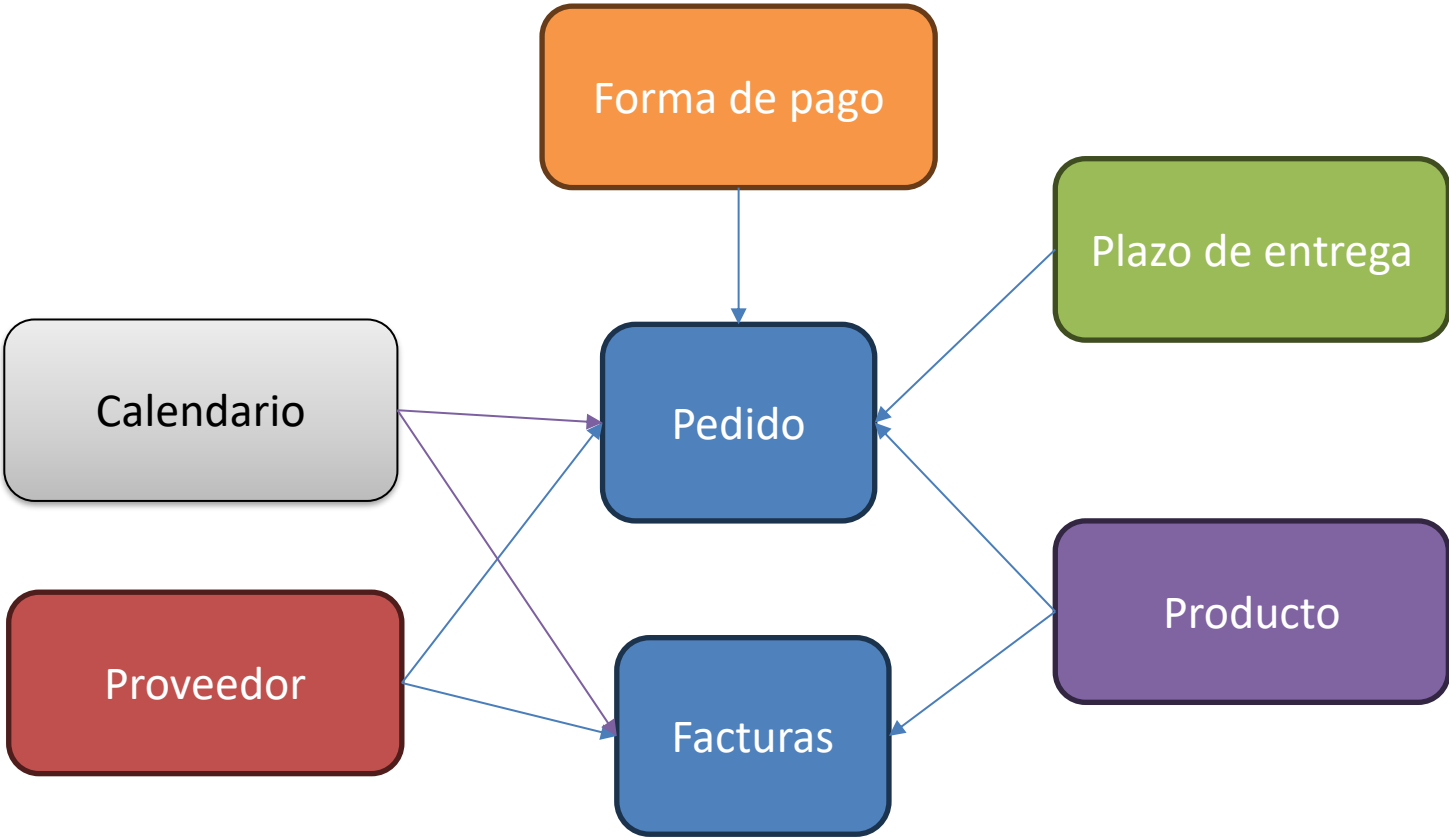
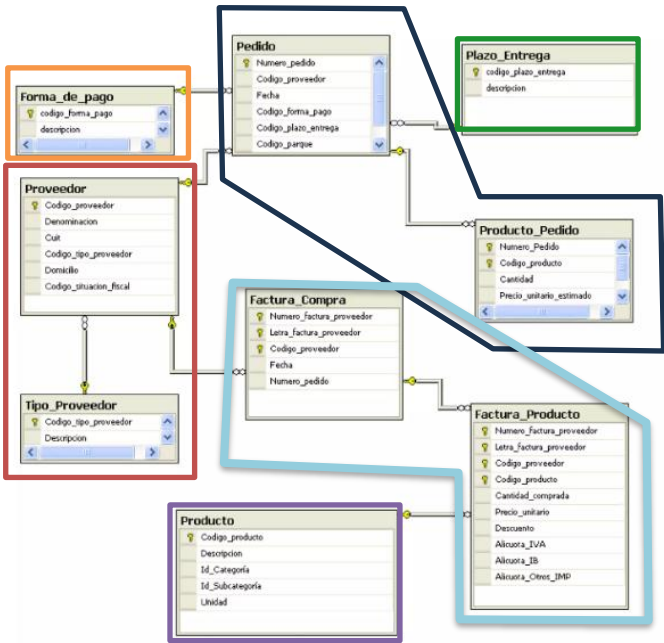
Ejercicio: modelado dimensional



Resolución de ejercicio de modelado dimensional



EXTRA: Para nota



RENDIMIENTO EN LOS INFORMES

¿Qué cosas afectan a que un informe funcione rápido o lento?

Granularidad: Nivel de detalle de los datos

Indica qué tan específicos son los registros en una tabla. Cuanto más detallada sea la información, más filas tendrá la tabla.

Ejemplo

- Alta granularidad: Ventas por cada ticket individual.
- Baja granularidad: Ventas totales por día.

Cardinalidad: Cantidad de valores únicos en una columna

Se refiere a la cantidad de valores distintos en una columna. Afecta cómo Power BI almacena y procesa los datos.

Ejemplo

- Alta cardinalidad: Un millón de números de factura distintos.
- Baja cardinalidad: Solo cinco categorías de producto.

¿Qué cosas afectan a que un informe funcione rápido o lento?

Granularidad: Nivel de detalle de los datos

Indica qué tan específicos son los registros en una tabla. Cuanto más detallada sea la información, más filas tendrá la tabla.

Rendimiento

- ✓ Permite análisis más específicos (ventas por hora, cliente, producto).
- ✗ Aumenta el tamaño del modelo y el tiempo de procesamiento.
- ✗ Puede ralentizar cálculos complejos como medidas DAX.

Cardinalidad: Cantidad de valores únicos en una columna

Se refiere a la cantidad de valores distintos en una columna. Afecta cómo Power BI almacena y procesa los datos.

Rendimiento

- ✓ Necesaria cuando se requiere mucha cantidad de información y diferentes análisis
- ✗ Una cardinalidad alta aumenta el uso de memoria RAM.
- ✗ Filtrados y cálculos DAX pueden ser más lentos.

**Fundación
iberCaja** 