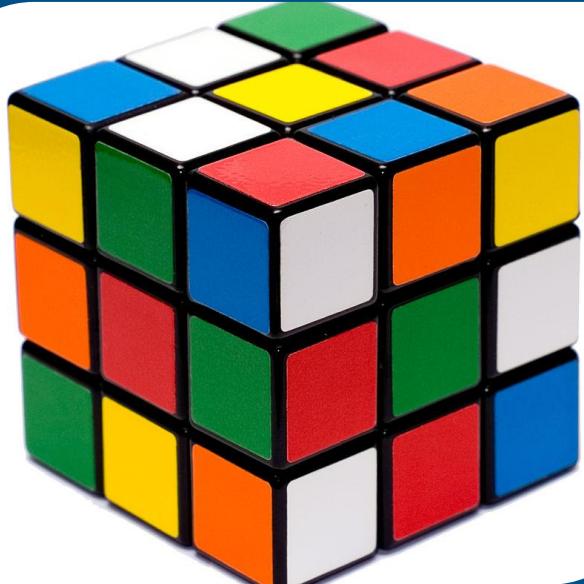


Visualización de Datos

Modelado de Datos



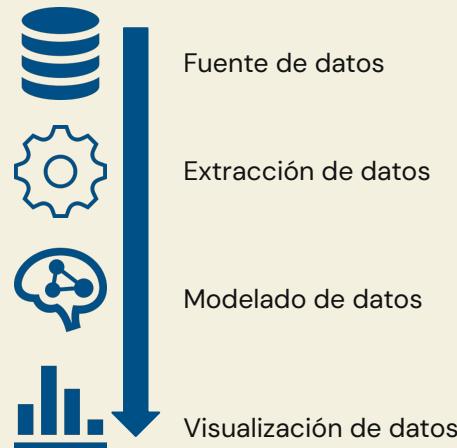
¿Que vimos ayer?

Visualización de Datos.

La visualización de datos sirve para **explorar y entender** los datos.

La visualización de datos sirve para **comunicar** los descubrimientos que hacemos en los datos.

Proceso de DataViz.

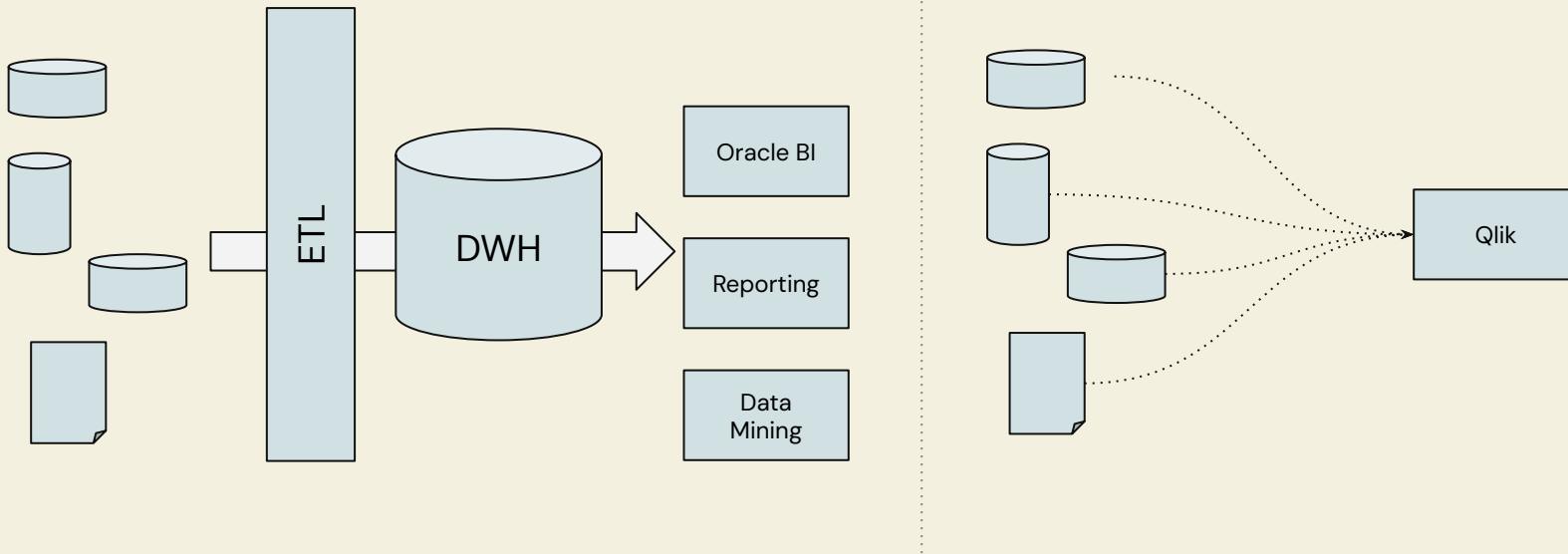


Qlik Sense.

Qlik Sense es una plataforma de BI que permite a los usuarios de negocio **explorar datos de forma interactiva** gracias a su motor asociativo.

Además, es una herramienta para científicos de datos que busquen crear **aplicaciones analíticas y cuadros de mando** empresariales.

Modelo Intermedio



¿Qué es el modelado de datos?

El **modelado de datos** es el proceso de **estructurar y organizar las tablas** de datos de una manera lógica y eficiente para facilitar el análisis y la generación de visualizaciones en herramientas como Qlik o Power BI. Es un paso crítico porque un buen modelo de datos:

- Simplifica la creación de informes y dashboards.
- Optimiza el rendimiento de las consultas.
- Ayuda a evitar errores en los cálculos y relaciones.

¿Qué es el modelado de datos?

El **modelado de datos** es el proceso de **estructurar y organizar las tablas** de datos de una manera lógica y eficiente para facilitar el análisis y la generación de visualizaciones en herramientas como Qlik o Power BI. Es un paso crítico porque un buen **modelo de datos**:



Simplifica informes y dashboards.

Facilita la creación de informes y visualizaciones claras.



Optimiza rendimiento.

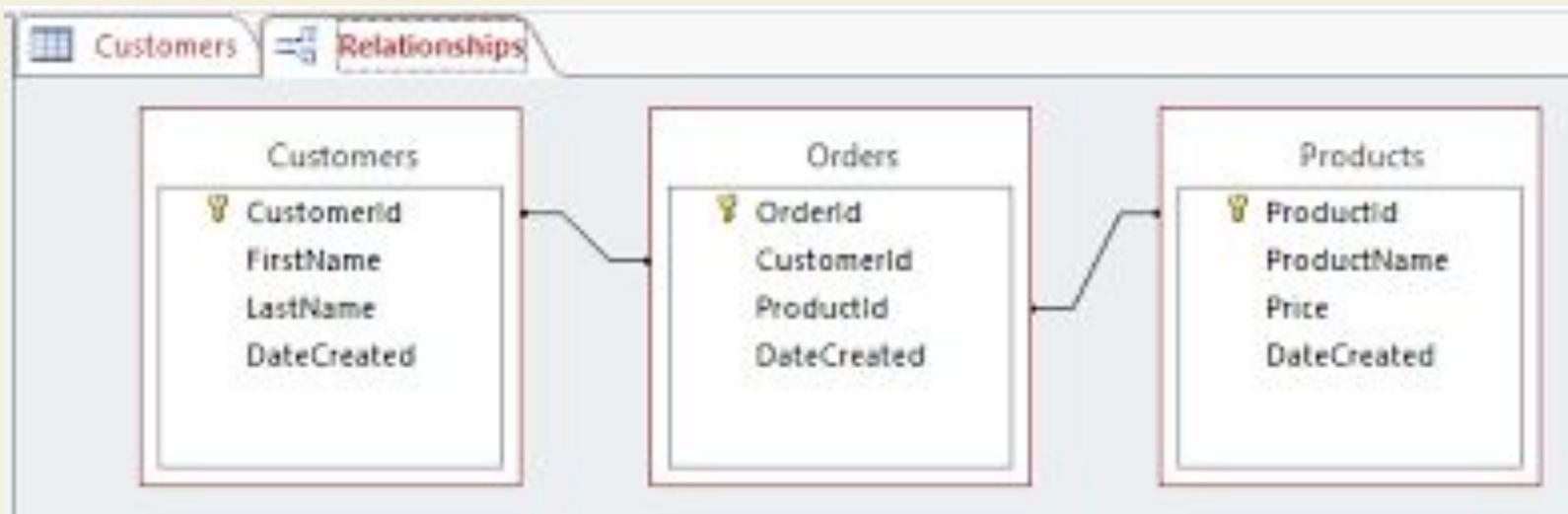
Mejora la velocidad y eficiencia de las consultas de datos.



Evita errores.

Asegura la precisión en los cálculos y relaciones entre tablas.

Tablas y Relaciones



Cardinalidad de las Relaciones

1:1



1:N



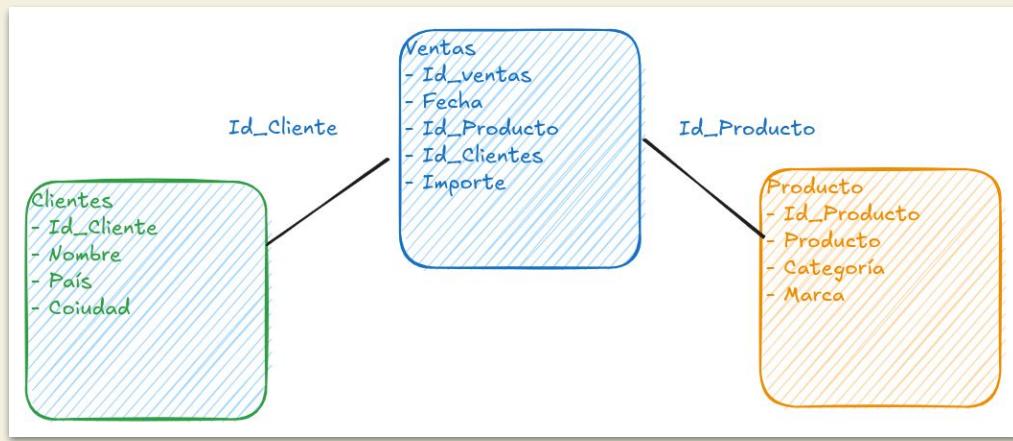
N:1



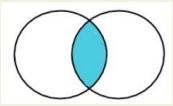
N:M



Normalizado Vs Desnormalizado



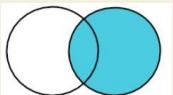
Joins



Inner Join

Solo devuelve los registros que coinciden en ambas tablas.

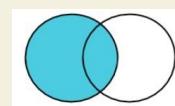
Ejemplo: Queremos ver solo los clientes que han hecho pedidos.



Right Join

Es como el LEFT JOIN, pero priorizando la segunda tabla.

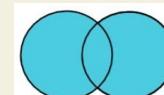
Ejemplo: Queremos ver todos los pedidos, incluso si no sabemos a qué cliente pertenecen.



Left Join

Devuelve todos los registros de la primera tabla y los coincidentes de la segunda.

Ejemplo: Queremos ver todos los clientes, incluyendo los que no han hecho pedidos.

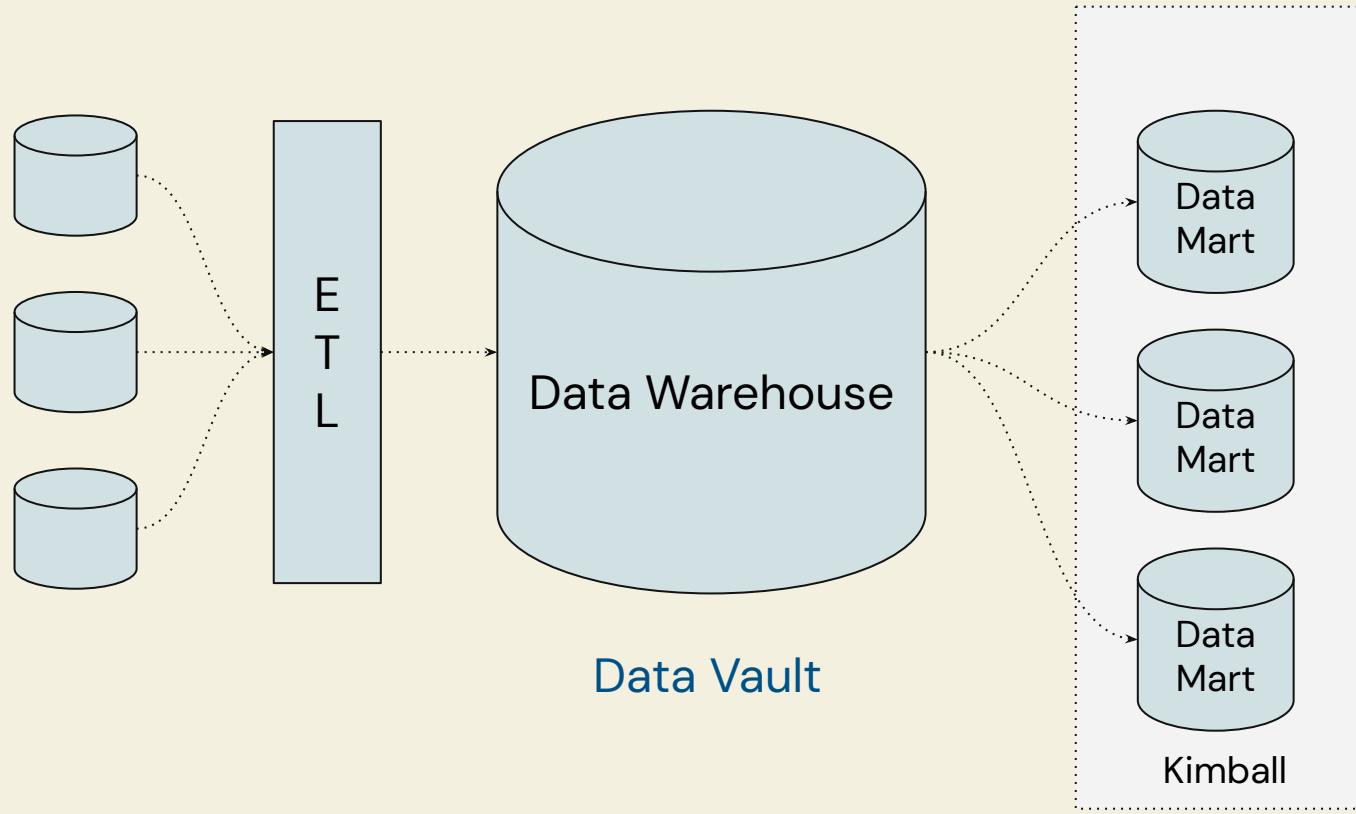


Full Outer Join

Devuelve todos los registros de ambas tablas, poniendo NULL donde no haya coincidencias.

Ejemplo: Queremos ver todos los clientes y todos los pedidos, aunque no tengan relación.

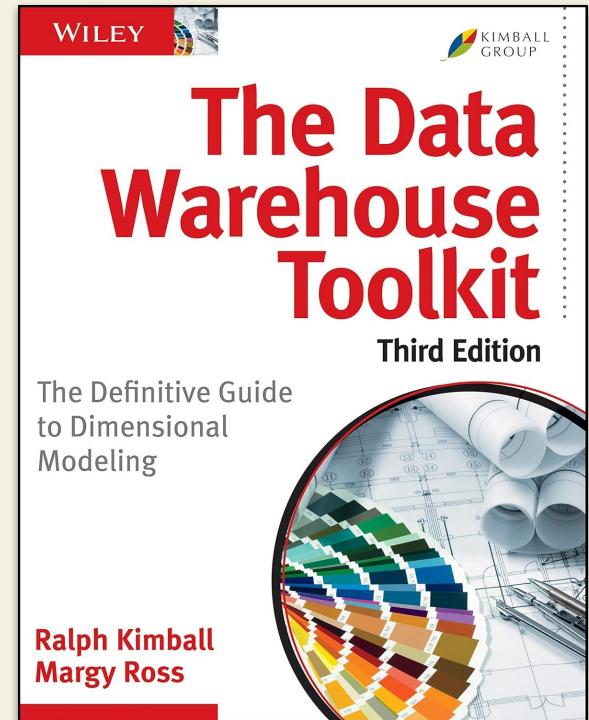
Data Warehouse



El Modelado Dimensional

The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd Edition

Ralph Kimball y Margy Ross



Modelado Dimensional

-  **Facilita el análisis de datos** mediante consultas rápidas y eficientes.
-  **Mejora la comprensión** de la información al estructurarse en un formato intuitivo.
-  **Optimiza el rendimiento** en sistemas de reporting y dashboards.
-  **Permite el almacenamiento histórico** de datos para análisis de tendencias.

¿Cómo visualizamos las ventas de forma que satisfaga a todos?

OrderID	Sales	Costs	Profit	Quantity	Discount	Freight
10251	21.36	16.84	4.52	6	1.02	43.41
10251	332.64	288.14	44.5	15	15.84	43.41
10251	185.2	156.44	28.76	20	0	43.41
10277	889.6	768.61	120.99	20	0	77.65
10277	360.96	311.87	49.09	12	0	77.65
10289	616.2	532.4	83.8	30	0	25.25
10289	320.4	276.83	43.57	9	0	25.25
10290	131.4	113.53	17.87	20	0	58.77
10290	1890.45	1633.35	257.1	15	0	58.77
10290	294.9	254.79	40.11	15	0	58.77
10290	134.9	116.55	18.35	10	0	58.77

Un analista quiere ver las ventas **por trimestre**, un gerente quiere ver las ventas **por región**, y un ejecutivo quiere ver el rendimiento de **los 10 mejores productos**.

Modelado Dimensional

¿Qué hemos vendido?



Tabla de Dimensión

¿Qué ha pasado?

RECEIPTIFY		
LAST 6 MONTHS		
QTY	ITEM	AMT
00	SOFIA - CLAIRO	3:08
01	HEAT WAVE - SNAIL MAIL	5:08
02	BAGS - CLAIRO	4:21
03	I KNOW THE END - PHOEBE BRIDGERS	5:45
04	CHINESE SATELLITE - PHOEBE BRIDGERS	3:37
05	PRISTINE - SNAIL MAIL	4:55
06	BETTY - TAYLOR SWIFT	4:55
07	KYOTO - PHOEBE BRIDGERS	3:05
08	KYOTO - PHOEBE BRIDGERS	3:05
09	SAVIOR COMPLEX - PHOEBE BRIDGERS	4:02
ITEM COUNT:		10
TOTAL :		41:60
CARD #: **** * 2020		
AUTH CODE: 123420		
CARDHOLDER: MICHELLE LIU		
THANK YOU FOR VISITING!		
 receiptify.herokuapp.com		

Tabla de Hechos

¿Cuándo hemos vendido?



Tabla de Dimensión

¿Quién ha comprado?



Tabla de Dimensión

¿Quién ha vendido?



Tabla de Dimensión

Tabla de Hechos

Id Venta	Id Cliente	Id Prod	Fecha	Importe
1	2	2	3/13	100
2	3	6	2/12	60
3	2	3	10/12	120
4	8	2	6/12	90

- Las tablas de hechos contienen los **datos numéricos** que se van a analizar, como ventas, ingresos, cantidad vendida, costos, etc.
- Son las tablas donde se almacenan las métricas.
- Estas tablas suelen tener claves que hacen **referencia a las dimensiones** (como IDProducto, IDCliente, IDFecha, etc.), lo que permite hacer un análisis más detallado.

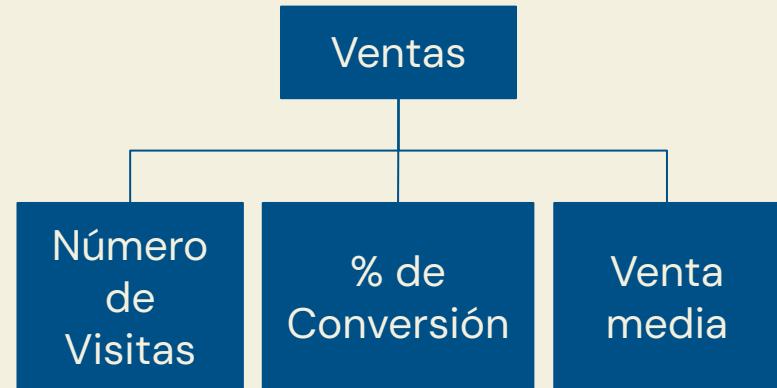
Tabla de Dimensiones

- Las tablas de dimensiones contienen los **atributos descriptivos** o características que ayudan a interpretar los datos en las tablas de hechos.
- Las dimensiones **permiten segmentar y clasificar los datos** para realizar análisis más detallados, como análisis por producto, cliente, tiempo, etc.

Id Cliente	Nombre	Ciudad
1	Paco	BCN
2	Marta	MAD
3	Ana	BCN
4	Luís	GRN

No todas las métricas son iguales: La Jerarquía del Dato

1. **La Métrica Estrella (The North Star):** Es nuestro Objetivo. Nos dice si ganamos o perdemos el partido. (Ej. Facturación Total).
2. **Las Métricas Palanca (Drivers):** Son la Explicación. Son las variables que componen matemáticamente el objetivo. Si movemos estas, movemos la estrella."



Jerarquías, dimensiones

Una jerarquía es una **estructura lógica que organiza los datos en niveles**. Las jerarquías permiten navegar de forma intuitiva entre diferentes niveles de detalle, como Año > Mes > Día o Región > País > Ciudad.

Ventajas:

- Navegación rápida: Facilita el análisis a diferentes niveles sin necesidad de aplicar filtros manualmente.
- Claridad visual: Mejora la organización de datos en gráficos e informes.
- Uso en visualizaciones dinámicas: Permite crear gráficos con drill-down (profundización) para explorar los datos en detalle.

Granularidad de los datos

La granularidad se refiere al nivel de detalle en el que los datos son capturados y almacenados.

- **Alta granularidad:** Los datos se almacenan con mucho detalle, como en el caso de tickets de compra individuales.
 - Ejemplo: Cada ticket incluye el ID del producto, precio, cantidad y fecha exacta de compra.
- **Baja granularidad:** Los datos están agregados, lo que reduce el nivel de detalle pero también disminuye el volumen de información.
 - Ejemplo: Ventas totales por día o por categoría de producto.

Granularidad de los Datos

Alta granularidad (tickets individuales):



A detailed illustration of a stack of sales receipts. One receipt is prominently displayed in the foreground, titled "SALES RECEIPT". It contains columns for "Id Pedido", "Fecha", "Id Cliente", "Producto", "Cantidad", and "Precio". The receipt shows four separate purchases: a TV for \$500, two phones for \$800 total, a washing machine for \$400, and a vacuum cleaner for \$200. A barcode is at the bottom. A magnifying glass is shown focusing on the first purchase.

Id Pedido	Fecha	Id Cliente	Producto	Cantidad	Precio
1	1/12/24	1	Televisor	1	500
2	1/12/24	2	Phone	2	800
3	2/12/24	1	Lavadora	1	400
4	2/12/24	3	Aspiradora	1	200

Baja granularidad (datos agregados por día):



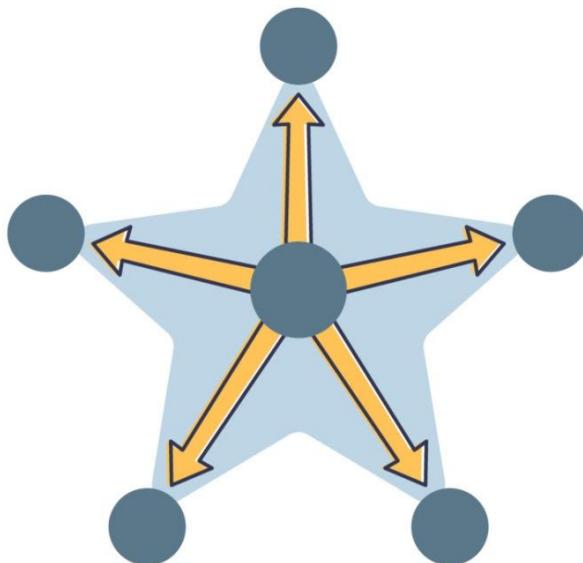
Impacto de la Granularidad

- Ventajas de reducir granularidad:
 - Mejora en el rendimiento: Menor volumen de datos facilita consultas y cálculos más rápidos.
 - Simplificación: Los datos son más fáciles de interpretar y analizar en niveles altos.
- Desventajas de reducir granularidad:
 - Pérdida de detalle: No se pueden realizar análisis específicos, como qué productos se vendieron más.
 - Menor flexibilidad: No se pueden realizar consultas detalladas sin recurrir a los datos originales.
 - Las causas que explican las desviaciones están en los detalles.

Ejemplo, granularidad de los datos

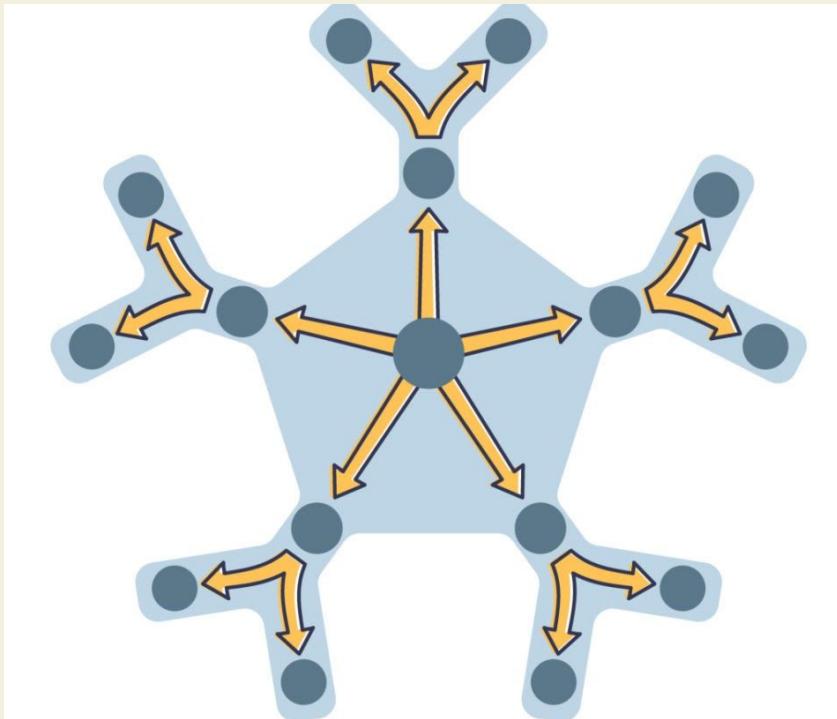
A	B	C	D	E	F
Fecha	Hora	Descripción actividad	Importe	Titular	Nombre comercio
1/3/2003	13.07	"TOYS""R""US"	361.8999939	José María de la Riva Amez	TOYS R US IBERIA,
4/4/2003	19.34	"TOYS""R""US"	199.9900055	Ricardo Romero de Tejada y Picatoste	TOYS R US IBERIA,
3/31/2004	13.53	"TOYS""R""US"	37.97999954	Miguel Ángel Araujo Serrano	TOYS R US IBERIA,
12/27/2004	21.21	"TOYS""R""US"	34.99000168	José María de la Riva Amez	TOYS R US IBERIA,
10/18/2007	20.18	"TOYS""R""US"	70.01000214	José María de la Riva Amez	TOYS R US IBERIA,
1/2/2008	19.50	"TOYS""R""US"	154.9499969	José María de la Riva Amez	TOYS R US IBERIA,
7/10/2008	20.19	"TOYS""R""US"	34.99000168	José María de la Riva Amez	TOYS R US IBERIA,
10/5/2008	14.35	"TOYS""R""US"	44.99000168	José María de la Riva Amez	TOYS R US IBERIA,
12/27/2008	16.49	"TOYS""R""US"	189.9400024	Carlos María Martínez Martínez	TOYS R US IBERIA,
12/31/2009	02.25	"TOYS""R""US"	59.99000168	José María de la Riva Amez	TOYS R US IBERIA,
12/31/2009	03.29	"TOYS""R""US"	29.98999977	José María de la Riva Amez	TOYS R US IBERIA,
1/2/2010	18.03	"TOYS""R""US"	583.75	Carlos María Martínez Martínez	TOYS R US IBERIA,
1/25/2003	07.22	3 C COMUNICATIC	257.8200073	Francisco José Moure Bourio	-
1/26/2003	09.24	ACCESORIOS Y REI	587.5499878	Ricardo Romero de Tejada y Picatoste	MAJAURO S.A.
3/22/2003	11.58	ACCESORIOS Y REI	181.0200043	Juan Manuel Astorqui Portera	NEUMATICOS AR
4/19/2003	11.33	ACCESORIOS Y REI	60.43999863	Miguel Ángel Araujo Serrano	CYA GIL Y BLAYA,S
4/19/2003	14.19	ACCESORIOS Y REI	40.63000107	Joaquín García Pontes	TOP-FIT S.A.
8/12/2003	11.48	ACCESORIOS Y REI	11.81999969	Juan Manuel Astorqui Portera	DON MARINO DU

Modelo en Estrella



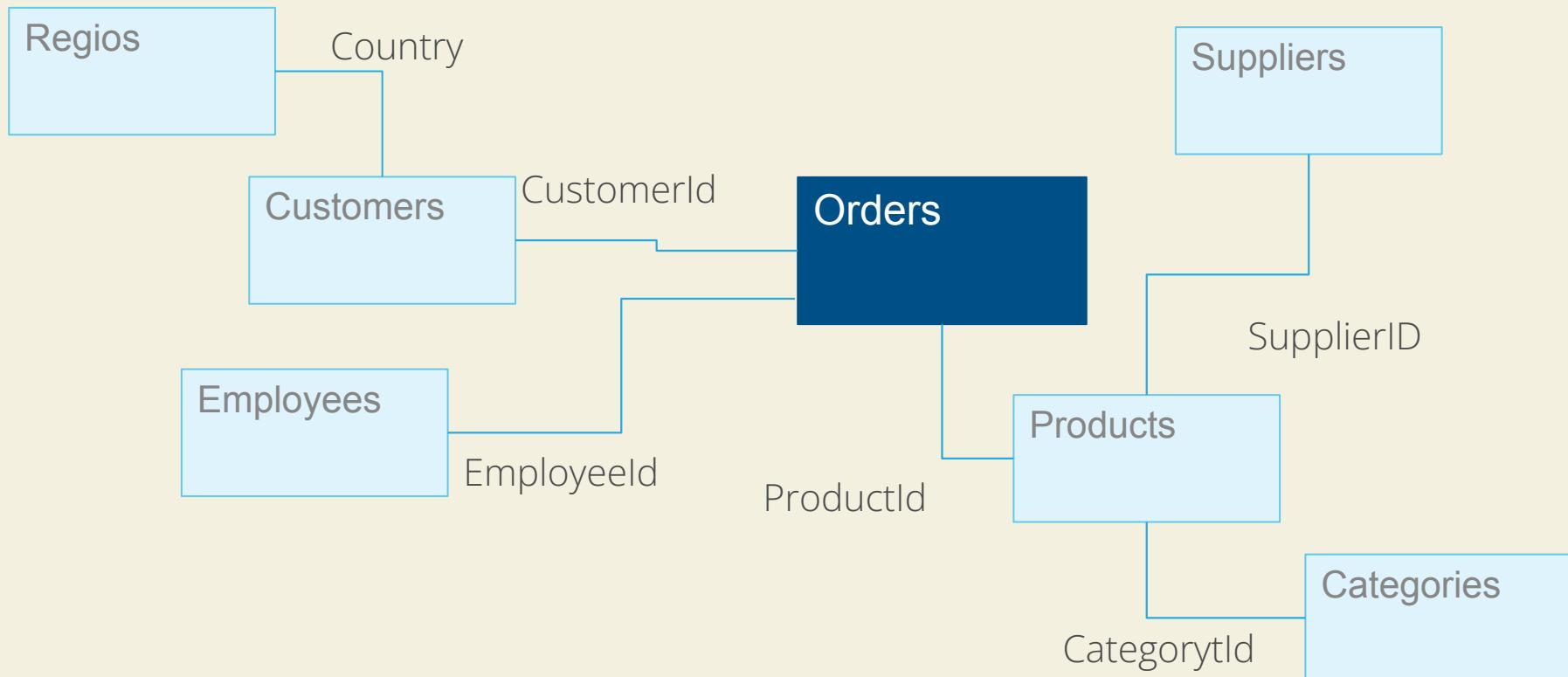
- Consta de una **tabla central de hechos** (contiene métricas o transacciones) conectada directamente a varias tablas de dimensiones (describen las métricas).
- Las dimensiones no tienen relaciones entre sí.
- Tablas de dimensiones son **desnormalizadas**.

Modelo en Copo de Nieve

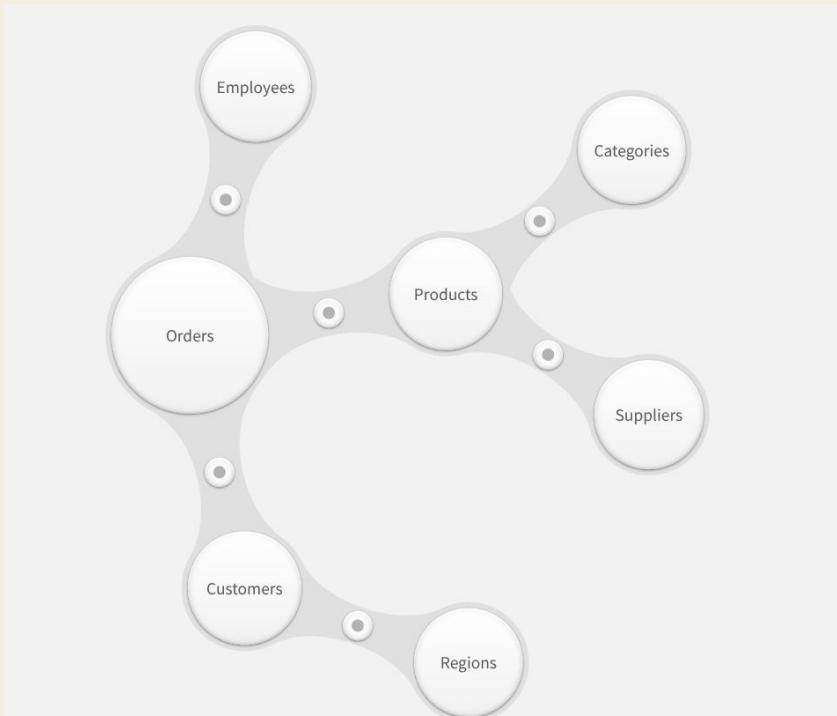


- La tabla central de hechos está conectada a **dimensiones normalizadas**.
- Las dimensiones pueden estar **divididas en varias tablas** relacionadas jerárquicamente.

Fashion Retail



Fashion Retail, Modelado de Datos



Fashion Retail

1. Cargar Datos.
 - a. ¿Es necesario cargar todos los datos?
2. Limpiar Datos.
 - a. Formatos.
3. Modelado de datos.
 - a. Desnormalización.
 - b. Modelo en estrella.
4. Elementos Maestros.
 - a. Medidas.
 - b. Dimensiones (simples, jerarquicas, cíclicas).

Visualización de Datos

Modelado de Datos

