



## Inteligencia de Negocios

INY6001



# Data Warehousing



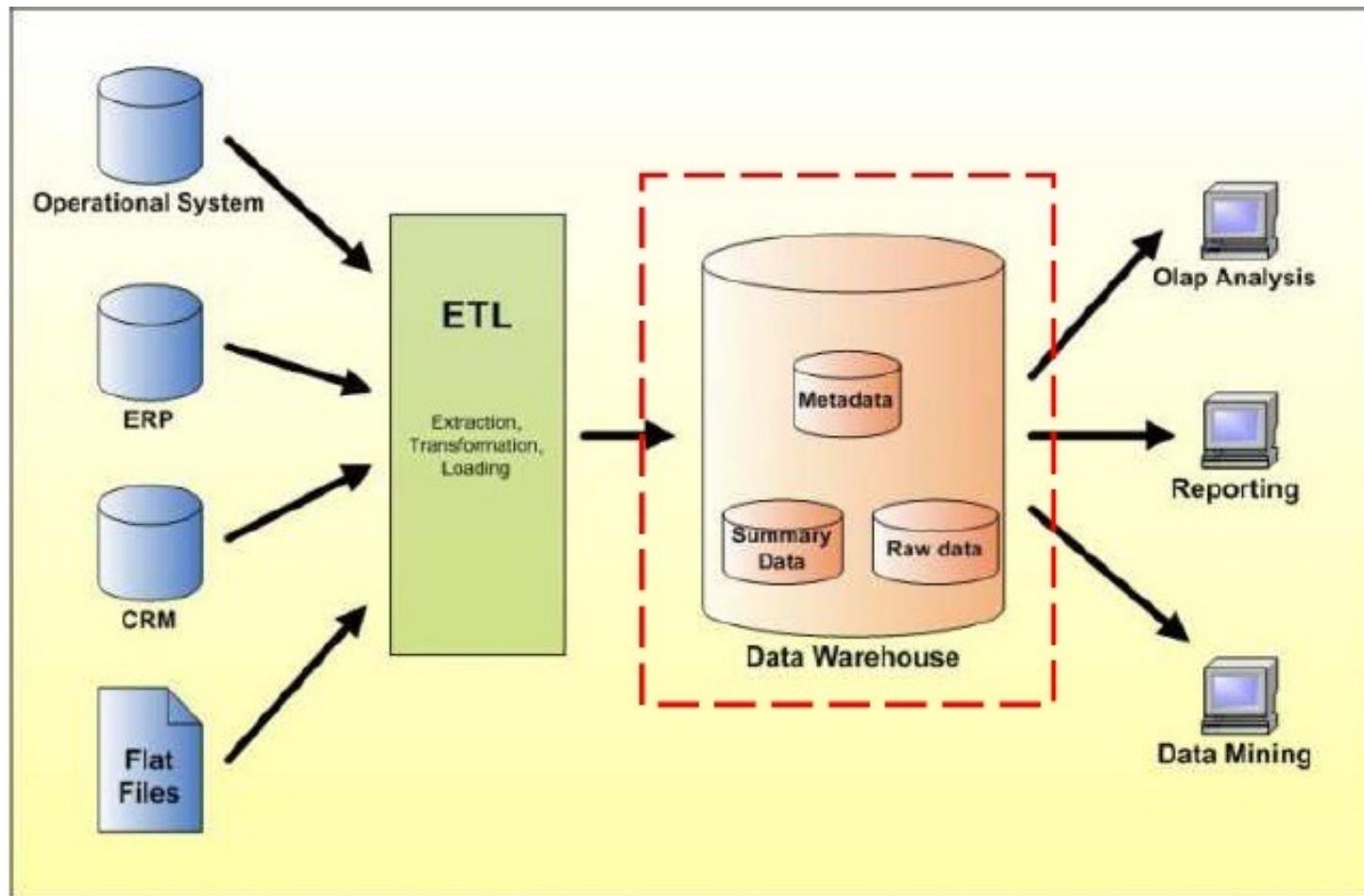
## ¿Qué es Data Warehousing?

Es el proceso para construir un Data Warehouse

Un **Data Warehouse (DWH)** es una colección de datos “orientada a un tema específico”, integrada, no volátil y variante en el tiempo.

De manera práctica es una base de datos que reúne datos de diversas fuentes para apoyar el proceso de toma de decisiones.

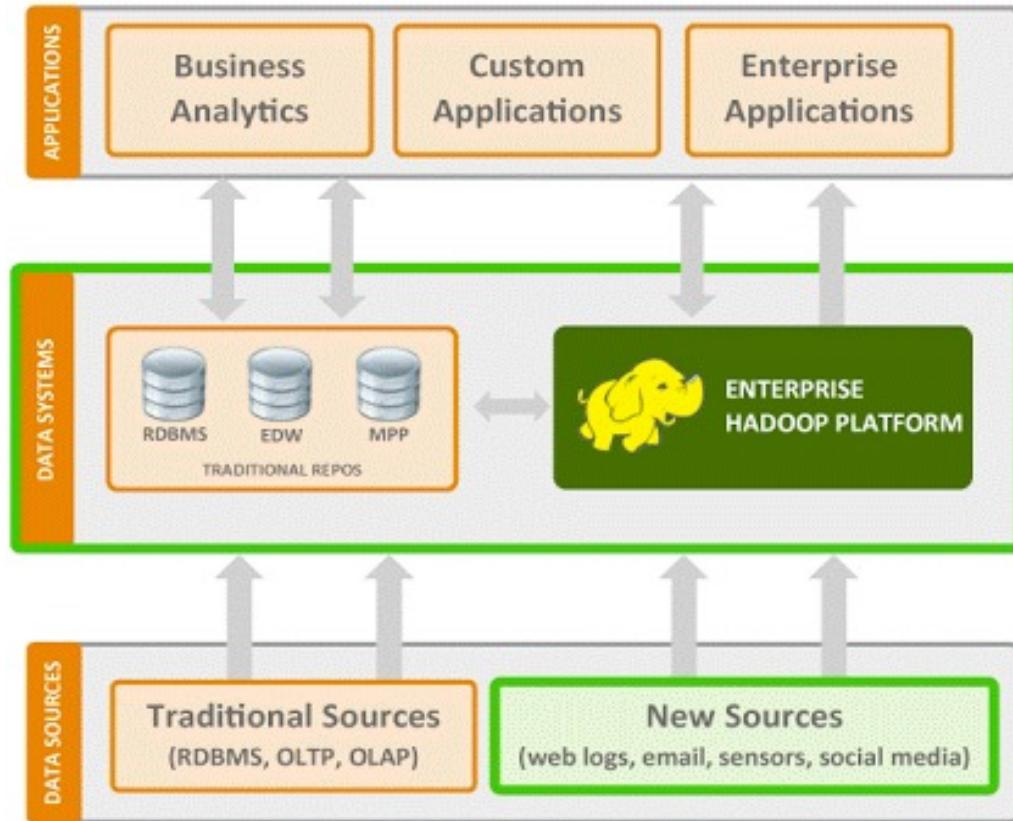




En esta arquitectura se puede apreciar de izquierda a derecha:

- ✓ Diversas fuentes de datos, que generalmente contienen los datos transaccionales u operacionales de la organización
- ✓ Procesos de Extracción, Transformación y Carga de Datos (ETL)
- ✓ Data Warehouse o almacén de datos
- ✓ Capa de explotación, la cual puede ser representada como reportes, minería de datos o herramientas OLAP

## Arquitectura de un DWH en la era del Big Data

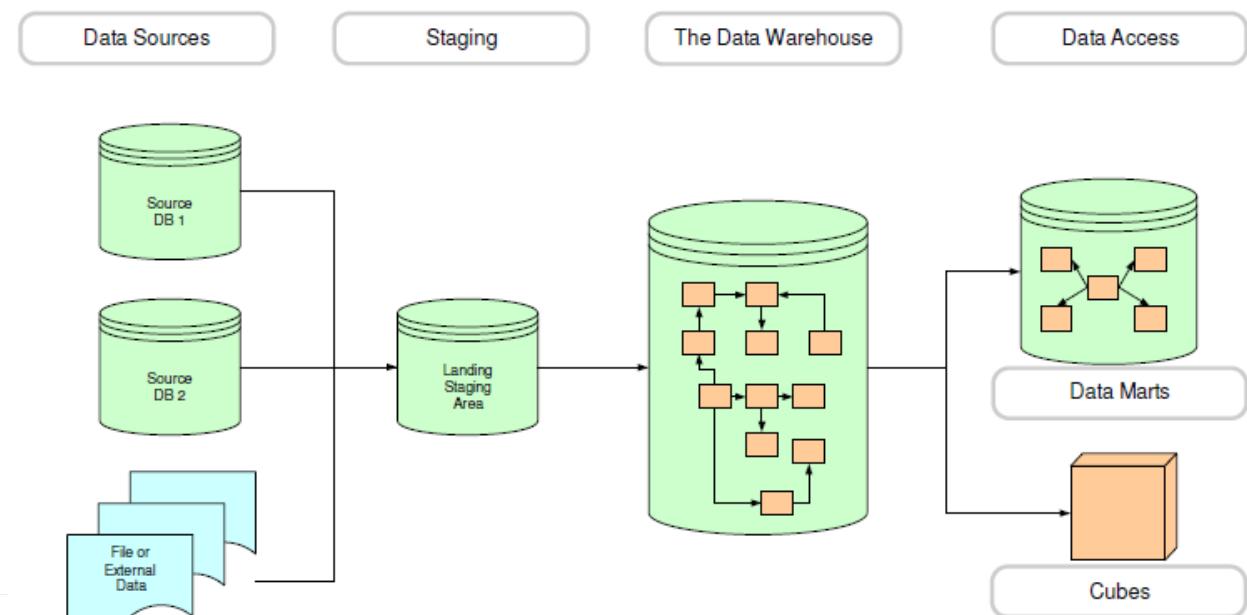


En esta arquitectura se incorpora al ecosistema un repositorio de Big Data, para capturar datos semiestructurados o no estructurados.

- Existen dos grandes autores respecto metodologías para proyectos de Data Warehouse, **Bill Inmon** y **Ralph Kimball**.

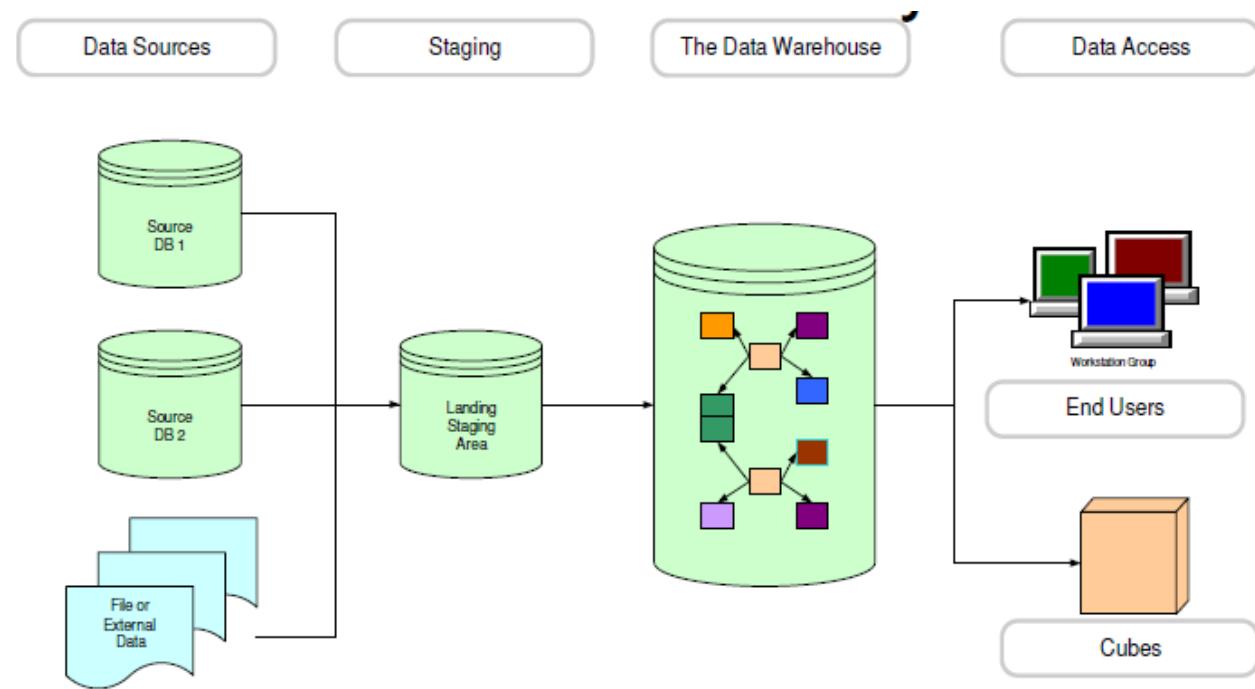
### Enfoque de Data Warehouse Corporativo

- Top-down
- Metodología de Bill Inmon (Padre del DWH)
- Pensado para proyecto que desean tener todo el universo de información



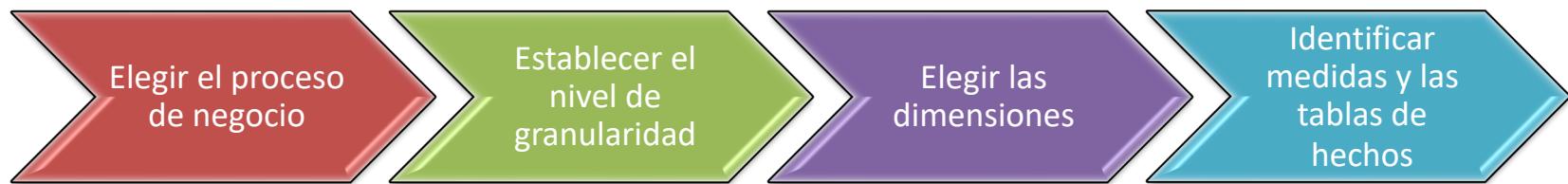
## Enfoque de Data Mart

- Bottom-up
- Metodología de Ralph Kimball (Padre del BI)
- Diseñada para realizar consultas rápidas y orientadas al negocio



## Modelamiento Dimensional

Este enfoque se basa en el modelamiento Dimensional. Para ellos se deben realizar las siguientes actividades:



### I. Seleccionar un Proceso de Negocio

- Por ejemplo:
  - Proceso de Venta
  - Proceso de Compras
  - Proceso de Selección de Personal

### II. Determinar la Granularidad

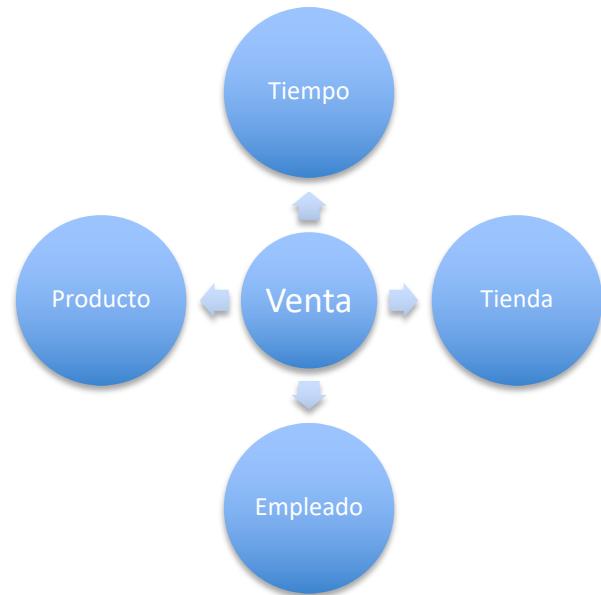
- Definir el nivel de detalle que se desea almacenar los datos del proceso.
- Depende de los requerimientos del negocio y los datos que existe en los sistemas que soportan el proceso
- La sugerencia es comenzar a diseñar el DW al mayor nivel de detalle posible
- Se puede luego realizar Sumarizaciones (Agregaciones) al nivel deseado

- Por ejemplo en la dimensión Sucursal:

Dimensión Sucursal	Dimensión Sucursal	Dimensión Sucursal	Dimensión Sucursal
* Sucursal	* Sucursal	* Sucursal	* Sucursal
** Tipo Sucursal	** Tipo Sucursal	** Tipo Sucursal	** Tipo Sucursal
*** País	*** País	*** País	*** País
**** Provincia	**** Provincia	**** Provincia	**** Provincia
			***** Ciudad

### III. Identificar las Dimensiones

- Surgen naturalmente de las sesiones de análisis y facilitadas por la elección del nivel de granularidad y de la matriz de procesos/dimensiones
- Las dimensiones corresponden a tablas que sirven para representar cada uno de los factores por los que se pueden analizar las diferentes áreas del negocio.
- Por ejemplo:
  - Tiempo: ¿cuándo se produce la actividad?
  - Producto: ¿cuál es el objeto de la actividad?
  - Almacén: ¿dónde se produce la actividad?
  - Cliente: ¿quién es el destinatario de la actividad?



### III. Identificar las Métricas y Tablas de Hecho

- Decidir la información a almacenar sobre el proceso.
  - **Hechos:** información (sobre la actividad) que se desea almacenar en cada registro de la tabla de hechos y que será objeto del análisis.
  - La tabla hechos, representará la transacción realizada en el negocio y que responderá a las preguntas generadas por los ejecutivos de la empresa. En esta tabla siempre están los id de las dimensiones y las **métricas** o cálculos que se utilizan,
  - Ejemplos: cantidades, sumas, máximos,...

#### Métricas Directas:

- Unidades Vendidas
- Monto Vendido
- Monto Pago
- Puntos
- Vuelto
- Métricas del Contexto
  - Costo Venta
  - Margen Venta

- Esquema Estrella

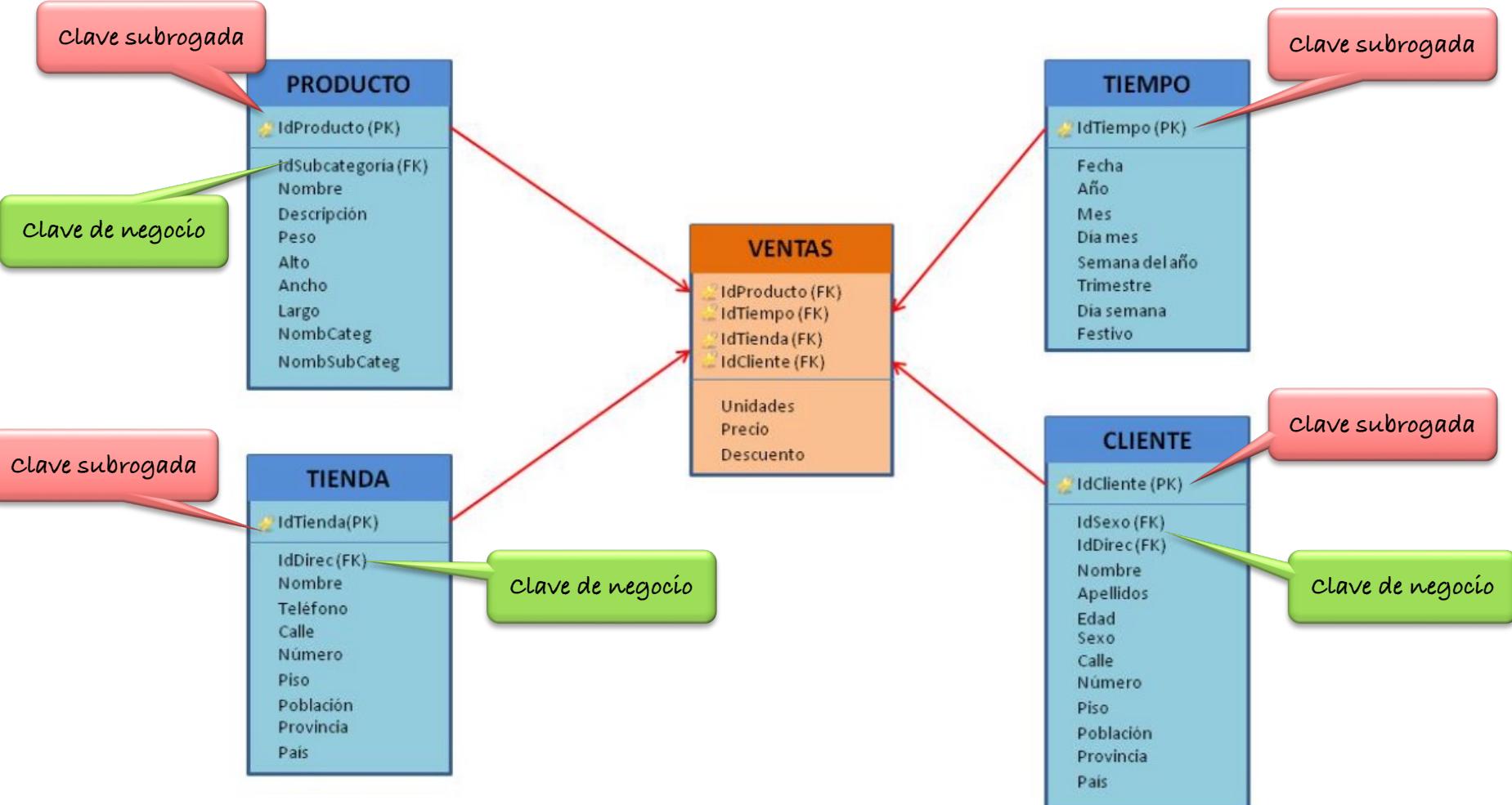
- El centro del modelo es la tabla de hechos (Fact Table)
- Alrededor de la tabla de hechos se organizan las tablas de Dimensiones (Dimension Table)



- Esquema Estrella

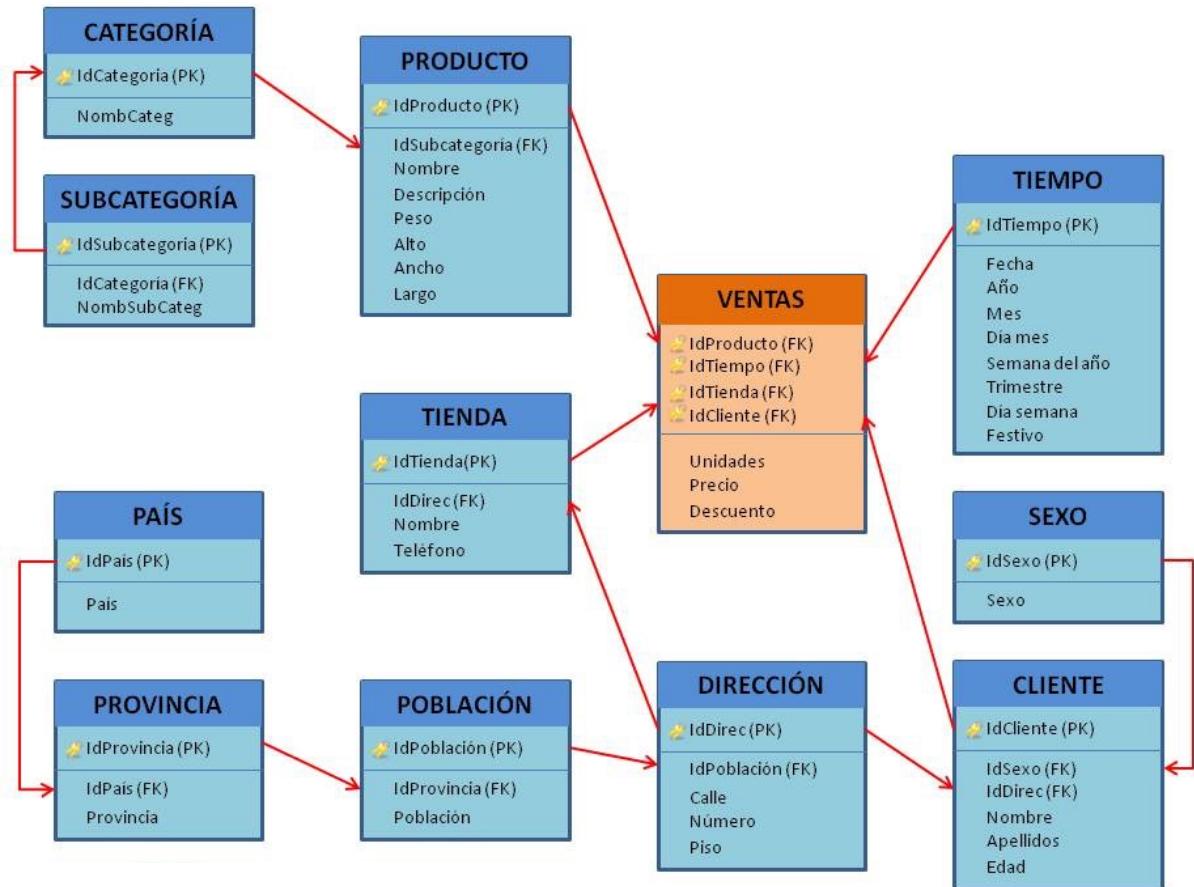
- La tabla de hechos de arma en base a las Primary Key de cada tabla de dimensión y con atributos numéricos (medidas), que corresponden a los valores medibles sobre el proceso que se está modelando
- Se sugiere que cada dimensión tenga como **Primary Key un clave subrogada**, para independizarla de cambios del negocio o para generar distintas versiones de la información **SCD (Slowly Changing Dimension)**
- Las Dimensiones se utilizan para filtrar o agrupar las medidas

## Diseño Físico



# Esquema Copo de Nieve

- Es una extensión del modelo Estrella
- Cada dimensión se normaliza, por lo que cada nivel (atributo) se lleva a una tabla (Lookup Table)



- Comparación de Esquemas

