

<b>Unidades del Examen 1 – Inteligencia de Negocios v3</b>	<b>4</b>
<b>Introducción:</b>	<b>4</b>
<b>Inteligencia de Negocios (Business Intelligence - BI):</b>	<b>4</b>
Problema para BI	4
Lo que busca BI	4
<b>Sistema de Información:</b>	<b>4</b>
<b>Pirámide de Inteligencia de Negocios:</b>	<b>5</b>
<b>Nueva estructura de Base de datos</b>	<b>5</b>
Arquitectura de BI:	5
Fuentes de Datos:	6
<b>Sistemas OLTP vs OLAP:</b>	<b>6</b>
Data Warehouse (DW):	6
<b>Alternativas (Evolución de la Arquitectura BI):</b>	<b>7</b>
<b>Los 5 Estilos de BI:</b>	<b>7</b>
<b>Procesos de BI:</b>	<b>9</b>
Alcance dentro de la organización	10
<b>Aplicaciones de BI:</b>	<b>10</b>
<b>Ciclo de vida de un proyecto BI</b>	<b>11</b>
<b>Planificación del Proyecto BI:</b>	<b>11</b>
<b>Factores de Éxito del Proyecto:</b>	<b>11</b>
<b>Definición de Requisitos del Negocio:</b>	<b>11</b>
<b>Modelo Dimensional</b>	<b>11</b>
Componentes del Modelo Dimensional	12
Esquemas del Modelo Dimensional	13
<b>Jerarquías de Atributos</b>	<b>14</b>
<b>Granularidad</b>	<b>15</b>
<b>Cubo OLAP</b>	<b>15</b>
<b>Relaciones entre Atributos</b>	<b>16</b>
<b>Ventajas del Modelo Dimensional</b>	<b>16</b>
<b>Data Mart</b>	<b>17</b>
<b>Paradigmas de Construcción</b>	<b>17</b>
Kimball:	17
Inmon:	18
<b>Dimensión Conformada</b>	<b>18</b>
<b>Método de la Matriz (Bus Matrix)</b>	<b>18</b>

<b>Sistemas OLAP</b>	<b>19</b>
<b>Operaciones OLAP</b>	<b>19</b>
<b>ROLAP (Relational OLAP)</b>	<b>22</b>
<b>MOLAP (Multidimensional OLAP)</b>	<b>23</b>
<b>HOLAP (Hybrid OLAP)</b>	<b>23</b>
<b>Modelo Físico</b>	<b>24</b>
<b>Estructura de Tablas</b>	<b>24</b>
Tablas Look Up	24
Tablas Relación	25
Tablas de Hechos (Fact Tables)	25
Tabla de Hechos Base	25
Tabla de Hechos Agregada	26
<b>Esquemas de Modelado</b>	<b>27</b>
Esquema Estrella (Star Schema)	27
Esquema Copo de Nieve (Snowflake Schema)	28
<b>Tipos de Medidas (Hechos)</b>	<b>30</b>
<b>Formas de Consolidación de Medidas</b>	<b>30</b>
<b>Conexión entre Hechos y Dimensiones</b>	<b>30</b>
<b>Clave Subrogada (Surrogate Key - SK)</b>	<b>30</b>
<b>Buenas Prácticas para Nombres en el DW</b>	<b>31</b>
<b>Dimensión Múltiple</b>	<b>32</b>
<b>Dimensión Degenerada</b>	<b>32</b>
<b>Dimensiones de Cambio Lento (SCD – Slowly Changing Dimensions)</b>	<b>33</b>
SCD Tipo 0 – Mantener valor original	34
SCD Tipo 1 – Reemplazo	34
SCD Tipo 2 – Nuevo registro	34
SCD Tipo 3 – Nueva columna	35
Atributos con diferente SCD	36
SCD Tipo 4 – Mini Dimensiones	36
<b>Dimensión Junk</b>	<b>37</b>
<b>Hechos o Medidas – Casos Especiales</b>	<b>38</b>
<b>Tablas de Hechos sin Medidas (Factless Fact Table)</b>	<b>38</b>
<b>Tipos de Tablas de Hechos Complementarias</b>	<b>39</b>
a. Tabla de Transacciones	39
b. Captura Periódica (Snapshot)	40
c. Captura Acumulativa	40
<b>Dimensionamiento de un Modelo</b>	<b>41</b>
<b>¿Qué es Power BI?</b>	<b>41</b>

<b>Arquitectura Básica de Power BI</b>	<b>42</b>
<b>Componentes Principales</b>	<b>43</b>
<b>Diferencias entre Power BI Desktop y Power BI Service</b>	<b>43</b>
<b>Fases de Trabajo en Power BI</b>	<b>43</b>
<b>Power Query Editor</b>	<b>44</b>
<b>DAX (Data Analysis Expressions)</b>	<b>44</b>
<b>Visualizaciones</b>	<b>44</b>
<b>Dashboards (Tableros de control)</b>	<b>45</b>
<b>Reportes en Línea (Power BI Service)</b>	<b>45</b>
<b>Buenas Prácticas para Proyectos Power BI</b>	<b>45</b>
<b>Recursos y Enlaces Útiles</b>	<b>45</b>
<b>Integración de Datos</b>	<b>45</b>
<b>ETL vs ELT</b>	<b>46</b>
<b>Proceso ETL</b>	<b>46</b>
Funciones del ETL	47
Subprocesos Clave del ETL	47
Staging Area	47
<b>Transformaciones Comunes</b>	<b>48</b>
<b>Estrategias de Carga</b>	<b>48</b>
<b>9. Errores Típicos en los Datos</b>	<b>48</b>
<b>Herramientas ETL</b>	<b>49</b>
<b>Cuadrante Mágico de Gartner (2024)</b>	<b>49</b>
<b>Arquitecturas y Tendencias BI</b>	<b>50</b>
Data Lake	50
Data Lakehouse	50
Data Vault	51
Data Mesh	51
Data Fabric	52
<b>Modelos de Implementación</b>	<b>52</b>
<b>Modelos de Servicio en la Nube</b>	<b>52</b>
<b>Comparación entre Soluciones BI</b>	<b>53</b>
<b>Ejemplos de Examen</b>	<b>54</b>
Ej. 1	54
Ej. 2	57
<b>Algunas auto evaluaciones de tarea</b>	<b>60</b>

# Unidades del Examen 1 – Inteligencia de Negocios v3

Recomiendo igual más la practica

## Introducción:

- La **información es el recurso más valioso** para las organizaciones, pero a menudo está dispersa en distintas áreas y formatos.
- El **objetivo de BI** (Business Intelligence) es **integrar estos datos y transformarlos en información útil para la toma de decisiones**.

## Inteligencia de Negocios (Business Intelligence - BI):

- Significa Business Intelligence (Inteligencia de Negocios /Empresarial)
- BI es el conjunto de procesos, tecnologías y herramientas que permiten transformar datos en información útil.
- Su objetivo es ayudar a las organizaciones a tomar decisiones basadas en datos

## Problema para BI

En general son datos que por cada área del sistema no están conectados entre sí, en distintas codificaciones, los nombres significan conceptos diferentes.

## Lo que busca BI

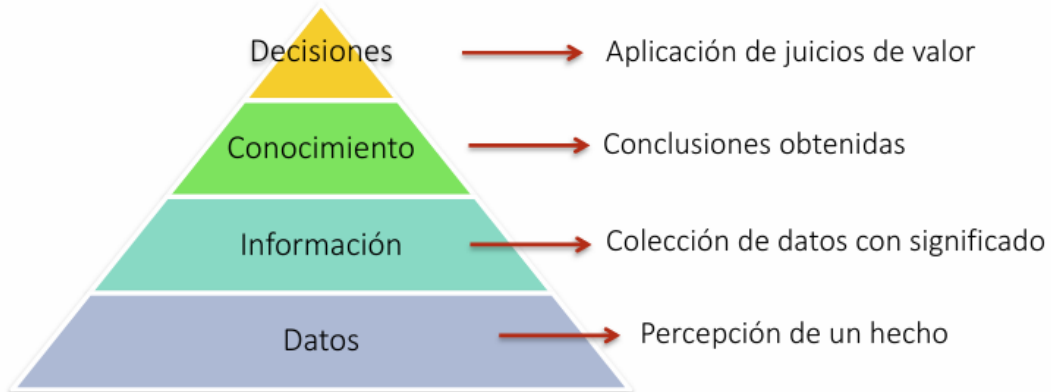
- Tener una única verdad de datos confiable
- Lograr visión integra del negocio
- Poder acceder a la información en el momento adecuado, sin intervenciones del área de sistemas

## Sistema de Información:

- Es un **conjunto de subsistemas que recopilan, almacenan, procesan y distribuyen información**.
- Permite que las organizaciones utilicen los datos para la operación y toma de decisiones.
- La información debe estar bien estructurada para ser útil.

## Pirámide de Inteligencia de Negocios:

- **Datos:** Información sin procesar (nombres, fechas, números).
- **Información:** Datos organizados que tienen significado.
- **Conocimiento:** Comprensión basada en información analizada.
- **Decisiones:** Aplicación del conocimiento para tomar acciones estratégicas.

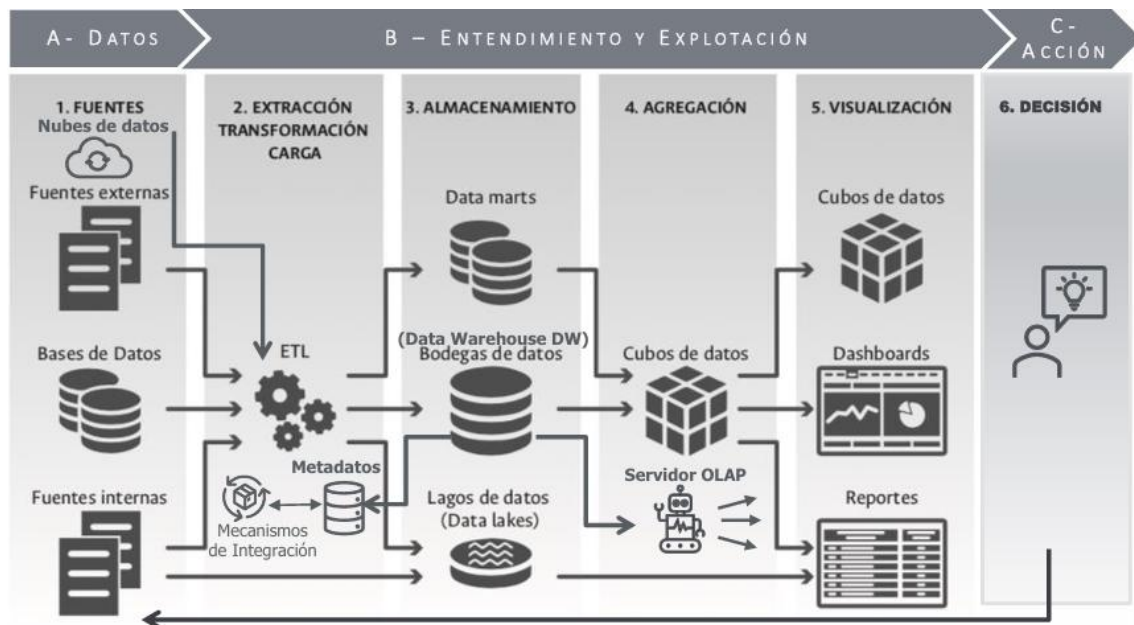


## Nueva estructura de Base de datos

Se necesita crear una base de datos nueva con la forma adecuada

Para esa estructura que se usa es DATA WAREHOUSE

## Arquitectura de BI:



- **Datos:** Centralizar información de múltiples fuentes en un DW

- **Entendimiento y Explotación:** Se integran y transforman en información útil (ETL) con herramientas BU y DM analizando y mejorando el entendimiento del negocio
- **Acción:** Actuar sobre los hallazgos realizados en el análisis

### Fuentes de Datos:

- **Internas:** Sistemas transaccionales (ERP, CRM), bases de datos locales.
- **Externas:** APIs, redes sociales, archivos externos (CSV, JSON, XML), datos abiertos.
- **Web 2.0:** Datos no estructurados como publicaciones, comentarios y multimedia.
- **IoT:** Datos de sensores y dispositivos conectados.
- **Sistemas OLTP:**
  - OLTP (Online Transaction Processing) se centra en el procesamiento de transacciones en tiempo real.
  - Características: Alta velocidad, múltiples usuarios, datos detallados y actualizados.
  - Ejemplos: Sistemas de ventas, reservas, inventarios.

### Sistemas OLTP vs OLAP:

**OLTP** (Online Transaction Processing): Sistemas **transaccionales** para operaciones diarias (ventas, reservas).

- Enfocado en **velocidad y eficiencia** para **transacciones** rápidas.
- Información actualizada y detallada.

**OLAP** (Online Analytical Processing): Sistemas para **análisis de datos históricos** y estratégicos.

- Enfocado en el **análisis y toma de decisiones**.
- Información histórica, agregada y optimizada para consultas complejas

### Data Warehouse (DW):

- Un DW es un **repositorio centralizado** de **datos** organizados para facilitar su **análisis**. Contiene información histórica e información consolidada
- Características: **Orientado al negocio** (fecha, cliente, producto, sucursal, ventas), **integrado** (Datos heterogéneas unificando nombres,

codificaciones y formatos), **no volátil** (No se modifican, se renueva el DW) y **variable en el tiempo** (Maneja datos históricos con análisis en el tiempo)

**Metadatos:** Datos que describen objetos en DW

## Alternativas (Evolución de la Arquitectura BI):

- **Data Mart:** Almacén de datos específico para un área (ventas, finanzas).
- **Data Lake:** Almacén para datos estructurados y no estructurados.
- **Data Lakehouse:** Combinación de Data Warehouse y Data Lake.
- **Data Vault:** Modelo para trazabilidad y auditoría de datos.
- **Data Mesh:** Modelo distribuido donde cada dominio gestiona sus propios datos.
- **Data Fabric:** Enfoque automatizado para gestionar y mover datos entre entornos

## Los 5 Estilos de BI:

- **Reportes:** Informes estándar y personalizados para usuarios de negocio.



**Reportes Operacionales Print-perfect**

- Via Web e Impresión
- Facil navegación a través de cientos de páginas de reportes
- Prompts que permiten a los usuarios definir el contenido



**Reportes de Negocios Pixel-perfect**

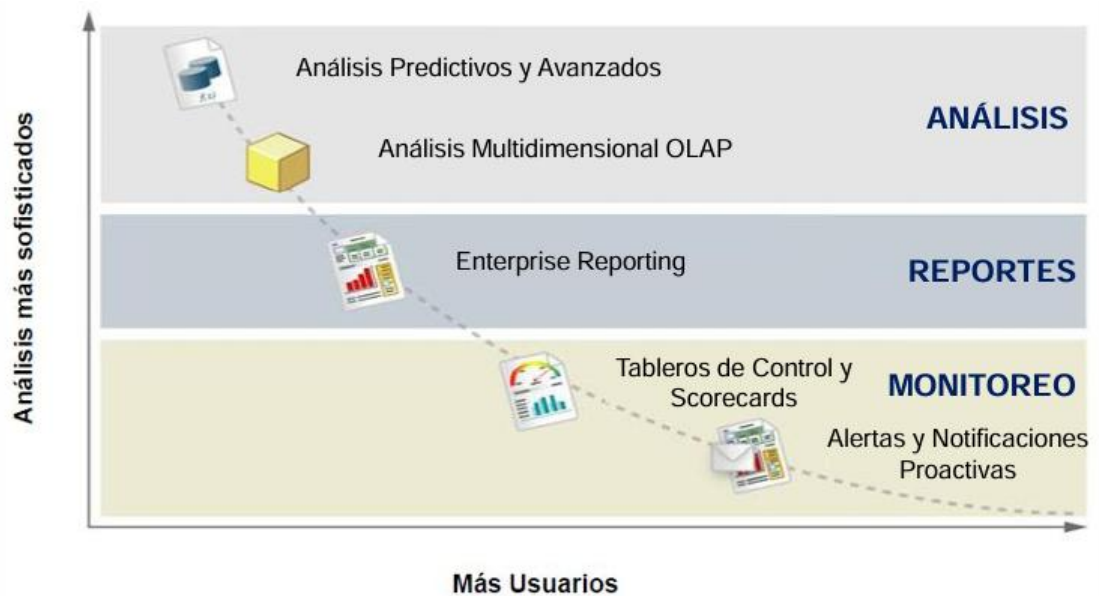
- Creado por usuarios de negocio, no por el departamento de sistemas
- Tablas y gráficos integrados



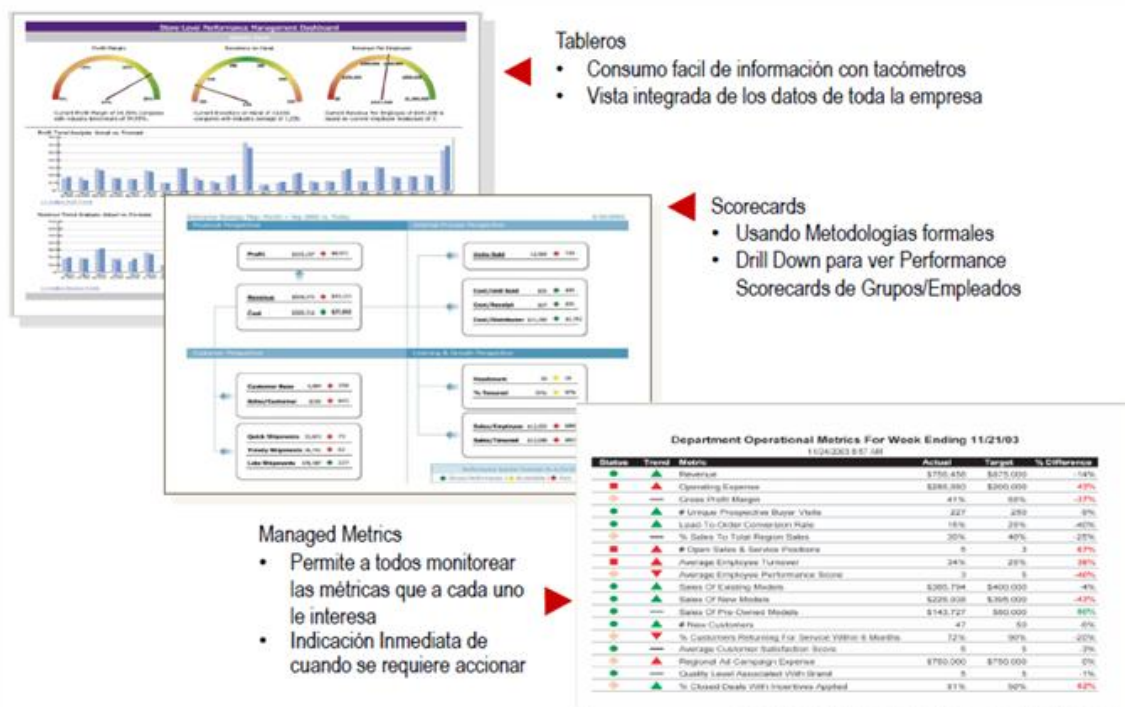
**Page-perfect Invoices and Statements**

- On-line Billing Applications
- Statements
- Other Page Forms

- **Análisis:** Exploración y descubrimiento de datos, análisis predictivo.



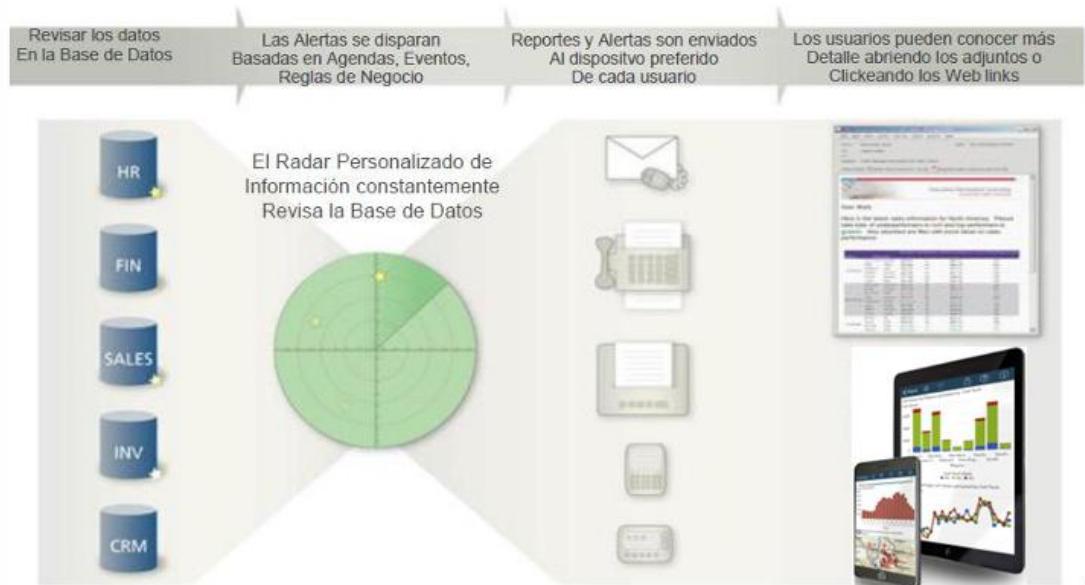
- **Monitoreo: Tableros de control (dashboards) para seguir métricas clave.**



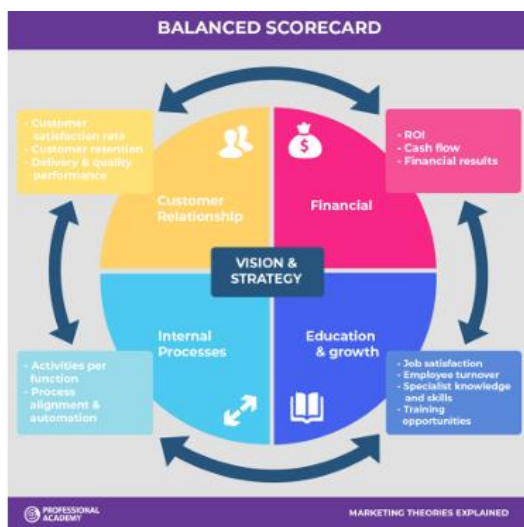
- **Alertas y Notificaciones: Avisos automáticos basados en condiciones.**



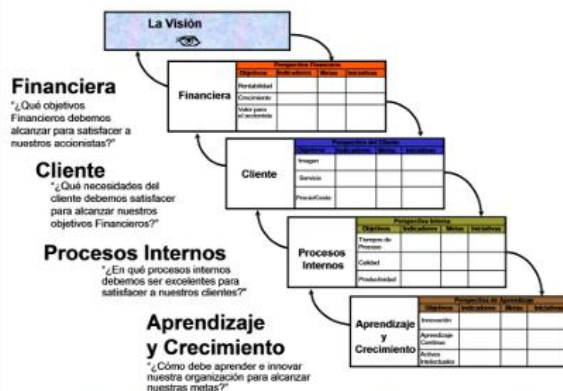
## Radar Personalizado de Información



- **Scorecards (Cuadros de Mando):** Medición del rendimiento con indicadores clave.



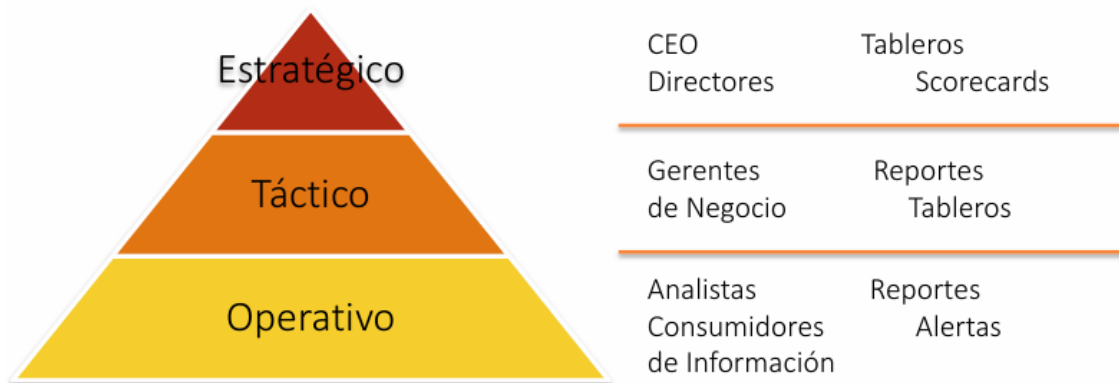
## Cuatro Perspectivas de Negocio



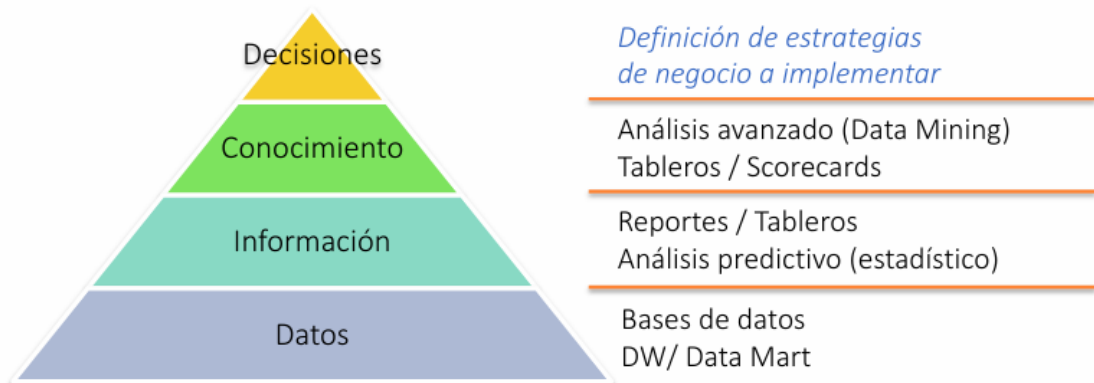
Fuentes: <https://balancedscorecard.org/> / <https://blog.bismart.com/tipos-de-cuadros-de-mando/> / <https://gestion.pensem.com/cuadro-de-mando-integral-ejemplo-definitivo-6-plantillas/> / <https://bscdesigner.com/es/bsc-plantillas-y-ejemplos.htm> / <https://margaritaberdugo.wordpress.com/2015/10/29/el-balanced-scorecard/>

## Procesos de BI:

- **Recolección de datos:** Extraer información de múltiples fuentes.
- **Transformación:** Organizar y limpiar los datos.
- **Análisis:** Generar reportes, dashboards y análisis predictivos.



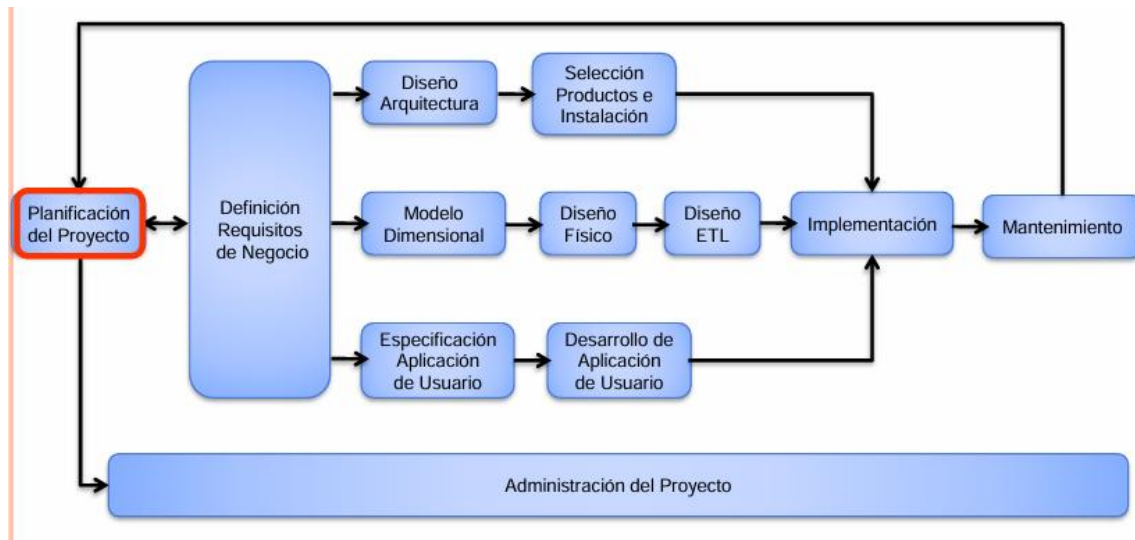
### Alcance dentro de la organización



### Aplicaciones de BI:

- **Reportes:** Información clara y accesible para los usuarios de negocios.
- **Análisis Predictivo:** Identificación de patrones y tendencias.
- **Monitoreo:** Control de indicadores clave en tiempo real.

## Ciclo de vida de un proyecto BI



### Planificación del Proyecto BI:

- Definir objetivos claros y realistas.
- Identificar las necesidades del negocio y los usuarios clave.
- Seleccionar herramientas y tecnologías adecuadas.
- Establecer un cronograma de implementación.

### Factores de Éxito del Proyecto:

- Patrocinio de la alta dirección.
- Participación activa de usuarios y técnicos.
- Calidad y disponibilidad de los datos.
- Capacitación y soporte continuo.

### Definición de Requisitos del Negocio:

- Identificar los procesos clave a analizar (ventas, marketing, finanzas).
- Establecer indicadores clave de rendimiento (KPIs).
- Priorizar las necesidades de los usuarios.

### Modelo Dimensional

El **Modelo Dimensional** es una técnica de diseño para bases de datos orientadas al análisis de información (OLAP). Se caracteriza por su simplicidad y eficiencia para responder preguntas de negocio

## Componentes del Modelo Dimensional

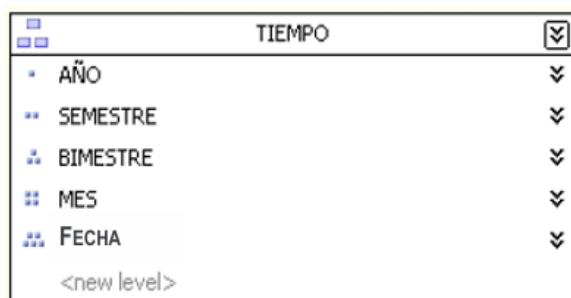
### 1. Hechos (Facts)

- Son eventos o transacciones que ocurren en el tiempo, como ventas, llamadas, pedidos, etc.
- Contienen **medidas o indicadores**: valores numéricos que describen el hecho (como cantidad vendida, importe, duración).
- **Tipos de medidas**:
  - **Básicas**: Existen directamente en el Data Warehouse (DW).
  - **Calculadas**: Se obtienen a partir de medidas básicas (ejemplo: ganancia = ingresos - costos).

### 2. Dimensiones

- Son categorías que describen el **contexto de los hechos**.
- Ejemplos: **Producto, Sucursal, Tiempo, Cliente**.
- Cada dimensión tiene **atributos** que permiten describirla:
  - Ejemplo: En la dimensión **Producto**, los atributos podrían ser **Nombre, Categoría, Precio**.
- Sirven para **mostrar, agrupar y filtrar** la información.

#### DIM Tiempo



#### DIM Geografía



### 3. Atributos

- Son las propiedades o características de las dimensiones.

- Pueden ser de tipo texto, fechas o incluso valores numéricos categorizados.
- Ejemplos:
  - En la dimensión **Producto**: Tipo de producto, Marca.
  - En la dimensión **Tiempo**: Año, Mes, Día.

✓ Provincia

✓ Ciudad

✓ Mes

✓ Ítem

✓ Empleado

✓ Edad

✓ Cliente

✓ Año

Provincia	
Buenos Aires	} Elementos
Córdoba	
La Pampa	

#### 4. Elementos

- Son los valores específicos que puede tener un atributo.
- Ejemplo: en el atributo **Provincia**, los elementos serían **Buenos Aires, Córdoba, La Pampa**.

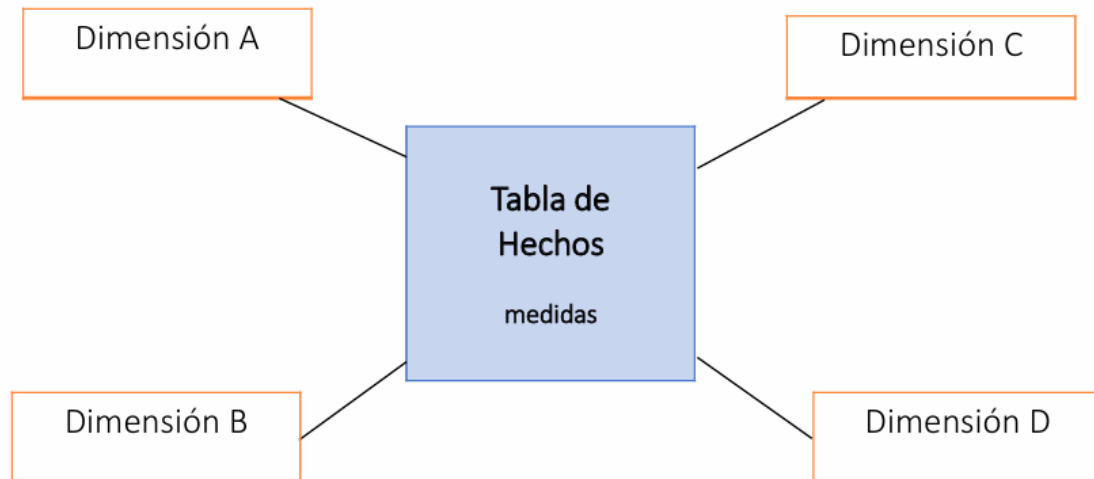
Mes	
Enero 2024	} Elementos
...	
Diciembre 2024	

Edad	
0-15 años	} Elementos
16-30 años	
31-45 años	

## Esquemas del Modelo Dimensional

### 1. Esquema Estrella (Star Schema)

- Una **tabla de hechos** en el centro conectada directamente a las **tablas de dimensiones**.
- Simplicidad y facilidad de interpretación.
- Ejemplo: Tabla de hechos de **Ventas** con dimensiones **Producto, Sucursal, Tiempo**.



## 2. Esquema Copo de Nieve (Snowflake Schema)

- Similar al esquema estrella, pero las dimensiones están normalizadas.
- Las dimensiones están divididas en tablas más pequeñas.
- Ofrece mayor flexibilidad, pero es más complejo.

## 3. Esquema Constelación (Galaxy Schema)

- Varias tablas de hechos que comparten las mismas dimensiones.
- Útil para manejar múltiples procesos de negocio relacionados.

## Jerarquías de Atributos

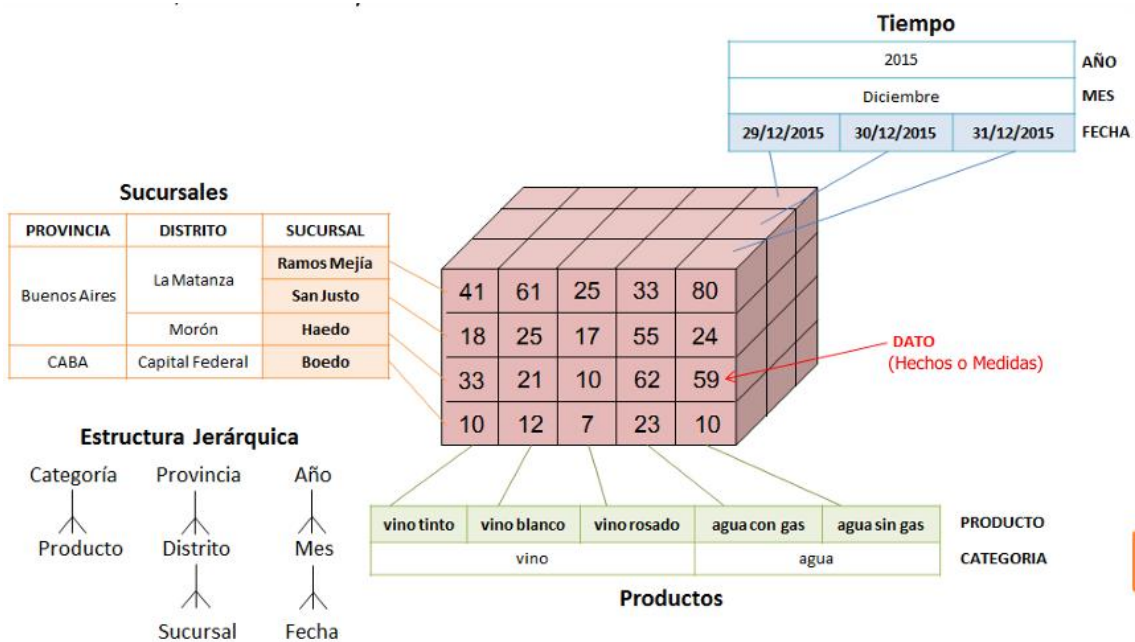
- Son **niveles de organización dentro de las dimensiones**.
- Permiten organizar los datos desde el nivel **más alto al más detallado**.
- Son relaciones entre atributos de una misma dimensión
  - 1:1
  - 1:N
  - N:N
- Ejemplo de la dimensión **Tiempo**:
  - Año > Trimestre > Mes > Día.
- Ejemplo de la dimensión **Producto**:
  - Tipo de Producto > Familia > Producto.

## Granularidad

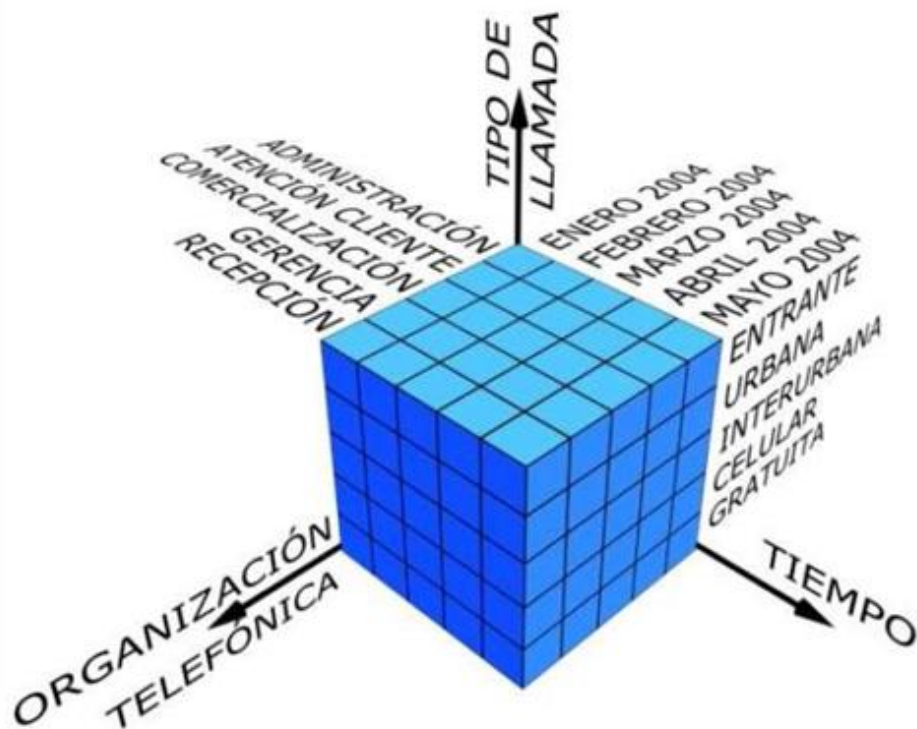
- Es el **nivel de detalle que se almacena en la tabla de hechos**
- Debe definirse correctamente para cumplir los requisitos del negocio.
- Ejemplo: Si las ventas se analizan por día, la granularidad es **Producto x Sucursal x Fecha**.

## Cubo OLAP

- Es una representación multidimensional del modelo dimensional.
- Cada **arista del cubo** representa una dimensión.
- En las **celdas del cubo** se encuentran las medidas o hechos (valores numéricos).
- Permite analizar los datos en múltiples perspectivas al mismo tiempo.
- Ejemplos:
  - Cubo de **Ventas** por **Producto, Tiempo y Sucursal**.



- Cubo de **Duración de Llamadas** por **Tipo de Llamada, Tiempo y Organización Telefónica**.



## Relaciones entre Atributos

- **1:1 (Uno a Uno):** Cada elemento de un atributo se asocia a un único valor de otro atributo.
  - Ejemplo: Cada código de producto tiene una descripción única.
- **1: N (Uno a Muchos):** Un elemento de un atributo se asocia a múltiples valores de otro.
  - Ejemplo: Cada familia de producto puede tener varios productos.
- **N: N (Muchos a Muchos):** Un elemento de un atributo puede estar asociado a muchos elementos de otro, y viceversa.
  - Ejemplo: Un producto se puede vender en varias sucursales y una sucursal puede vender muchos productos.

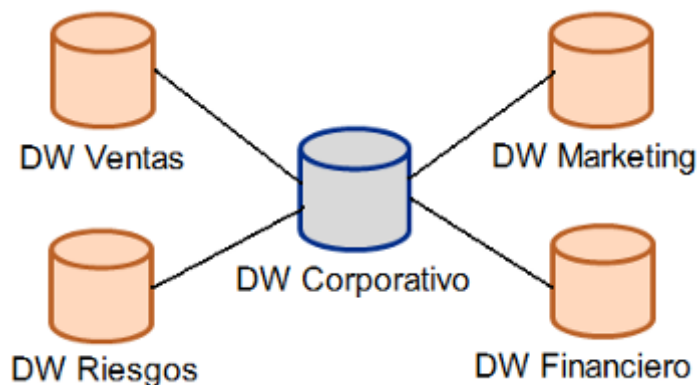
## Ventajas del Modelo Dimensional

- Presenta la información de manera clara y sencilla para los usuarios.
- Facilita el análisis y la generación de reportes.
- Resiste cambios en las necesidades del usuario (es flexible).
- Permite agregar nuevas dimensiones, hechos o atributos sin afectar el diseño general.



## Data Mart

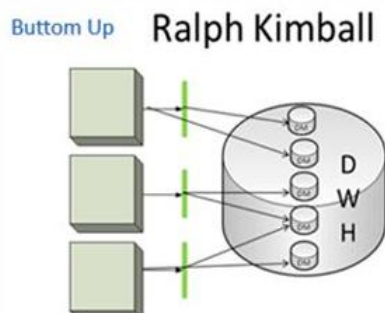
- Es un subconjunto del Data Warehouse (DW) que sirve a un área o función específica de una organización.
- **Tipos:**
  - **Dependiente:** Derivado del DW central.
  - **Independiente:** Se construye directamente desde fuentes a través de ETL.
- **Ventajas:** más enfocado, rápido de construir, ideal para necesidades puntuales.



## Paradigmas de Construcción

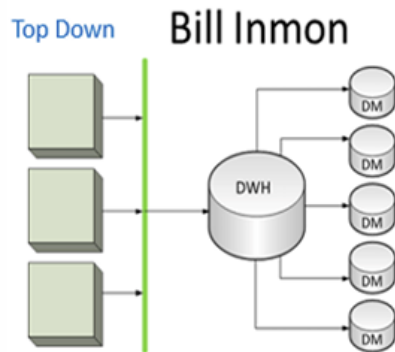
### Kimball:

- Construye Data Marts por prioridad de negocio y luego los integra en un DW.
- **Ventajas:** soluciones rápidas, más fáciles de implementar.
- **Desventajas:** mantenimiento más complejo, riesgo de datos desincronizados.



## Inmon:

- Construye primero un DW centralizado y luego genera Data Marts derivados.
- **Ventaja:** datos normalizados, estructura limpia.
- **Desventajas:** requiere más tiempo y mayor inversión.



## Dimensión Conformada

- Es una dimensión estándar compartida por múltiples Data Marts.
  - Ejemplo: dimensión **Cliente** usada en ventas, cobranzas, etc.
- **Ventajas:**
  - No se duplican dimensiones.
  - Reducción de costos de mantenimiento.
- **Desventajas:**
  - Cambios afectan todos los cubos que la utilizan.
  - Necesita control centralizado y coherente.

## Método de la Matriz (Bus Matrix)

- Herramienta para planificar un DW integrando múltiples procesos.
- **Cómo funciona:**
  - Filas = procesos o hechos (ej. ventas, stock, facturación).
  - Columnas = dimensiones (ej. tiempo, producto, región).
  - Las “X” indican relación entre hecho y dimensión.
- Ayuda a identificar dimensiones conformadas y granularidad mínima común

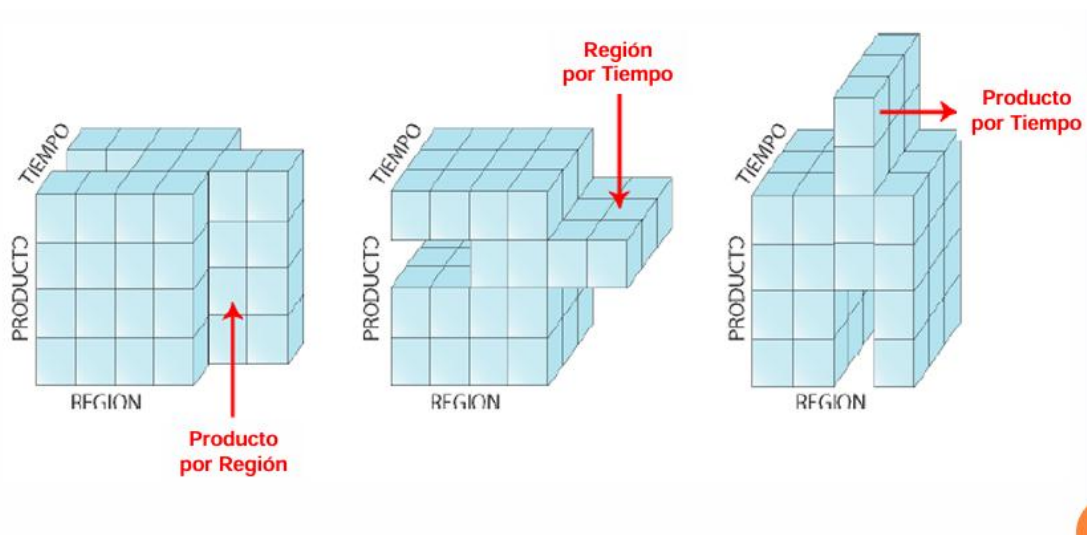
	Tiempo	Llamada	Documento	Geografía	Centro de Distribución
Reclamos	X	X		X	
Facturación	X		X	X	X
....					
Stock				X	X

## Sistemas OLAP

- OLAP (Online Analytical Processing) permite analizar grandes volúmenes de datos desde múltiples perspectivas.
- Se basa en dimensiones de negocio (tiempo, producto, región).
- Los datos se almacenan de forma estructurada para facilitar su exploración y visualización.

### Herramientas de Explotación DW

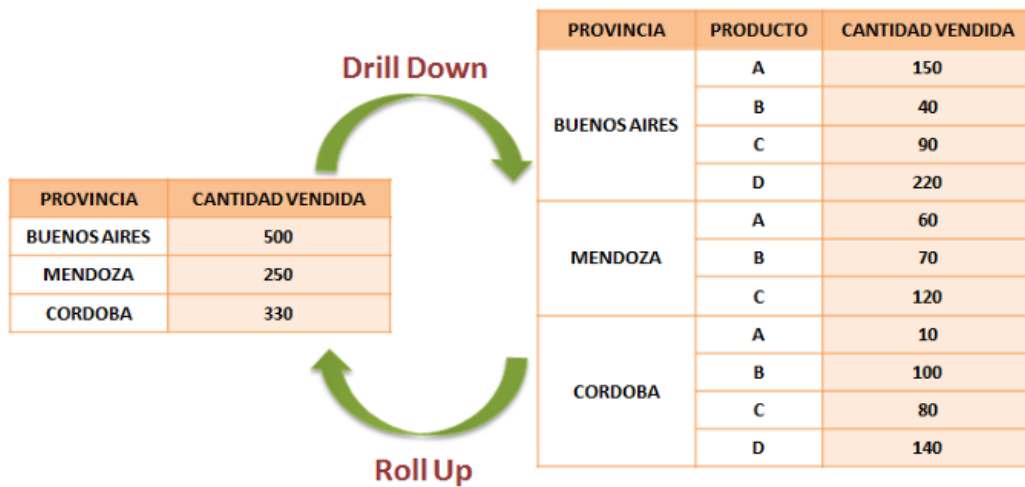
- Facilitan la consulta del DW sin necesidad de conocer su estructura interna.
- El usuario selecciona atributos y medidas desde una interfaz gráfica.
- La herramienta traduce esto a consultas SQL complejas automáticamente.
- Permite usar filtros, agrupar datos, mostrar por niveles jerárquicos, etc.



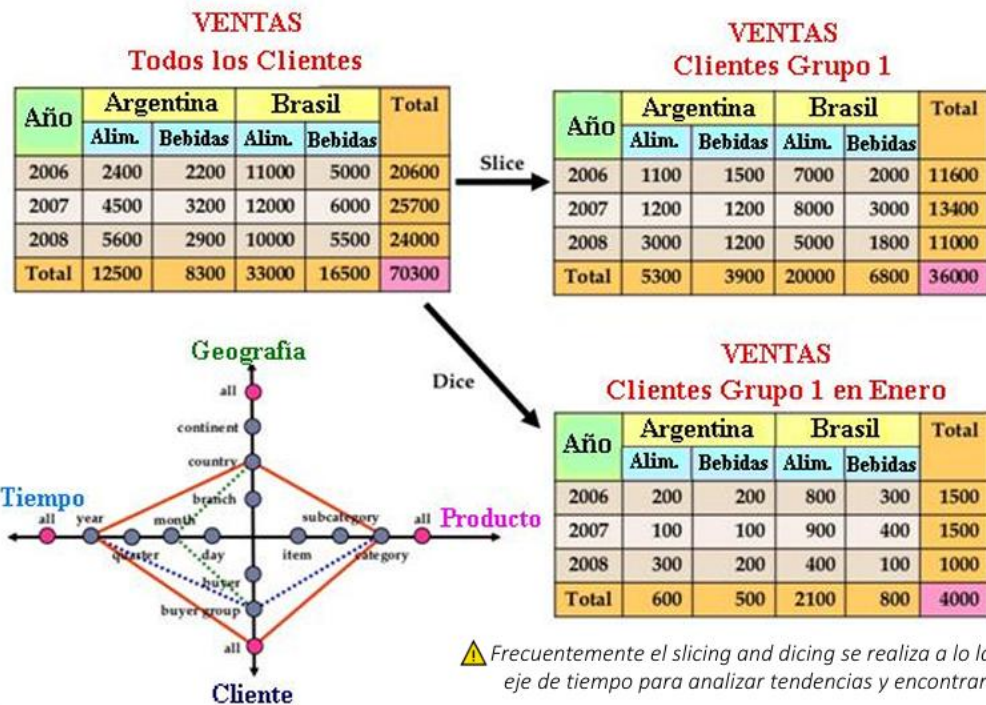
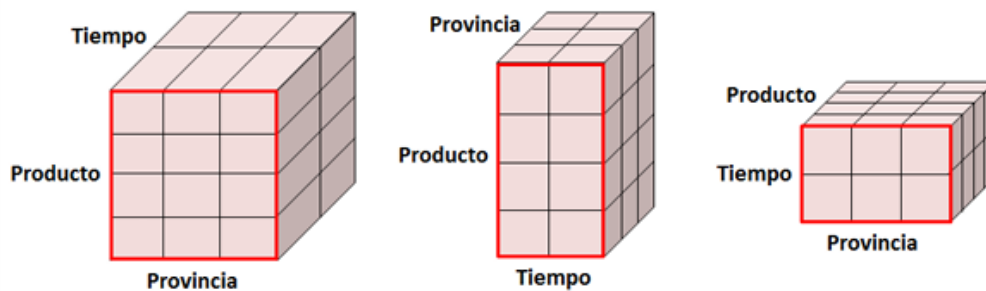
## Operaciones OLAP

- Permiten navegar y explorar el cubo de datos de forma dinámica:
  - **Drill Down / Roll Down:** ir a más detalle (de año a mes).

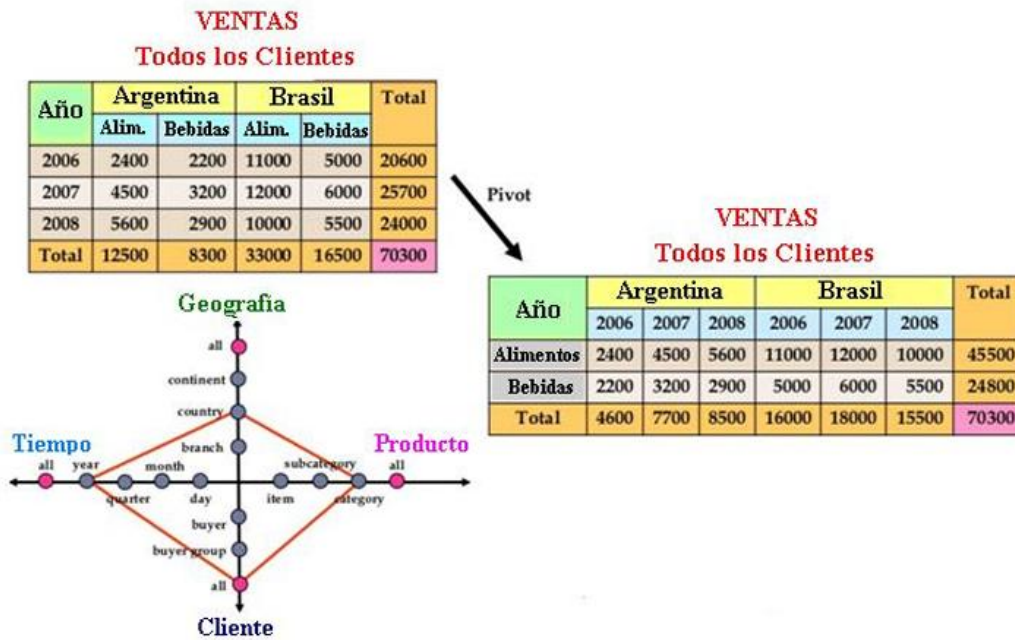
- **Roll Up / Drill Up:** consolidar niveles (de mes a trimestre).



- **Slice and Dice:** cortar y descomponer los datos desde distintas vistas (ej. solo “Mendoza”, o los 2 productos más vendidos).



- **Pivot:** cambiar la orientación del análisis (filas por columnas).



- **Drill Across:** combinar información entre cubos que comparten dimensiones.

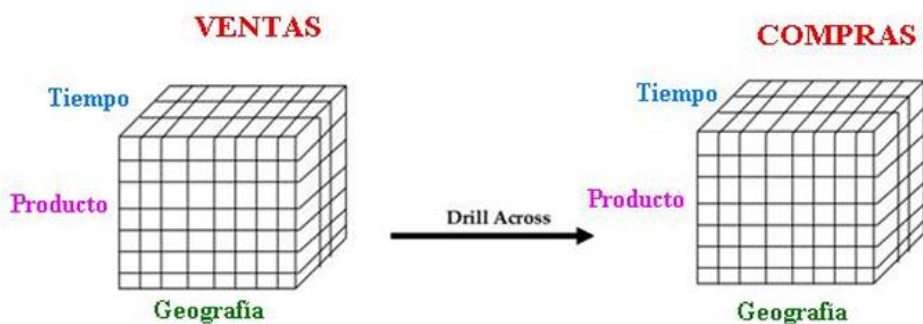
**VENTAS**  
Todos los Clientes

Año	Argentina		Brasil		Total
	Alim.	Bebidas	Alim.	Bebidas	
2006	2400	2200	11000	5000	20600
2007	4500	3200	12000	6000	25700
2008	5600	2900	10000	5500	24000
Total	12500	8300	33000	16500	70300

Drill Across

**COMPRAS**  
Todos los Clientes

Time	Argentina		Brasil		Total
	Alim.	Bebidas	Alim.	Bebidas	
2006	1500	1500	6000	3000	12000
2007	3500	2500	8000	4000	18000
2008	4000	1500	5000	3000	13500
Total	9000	5500	19000	10000	43500

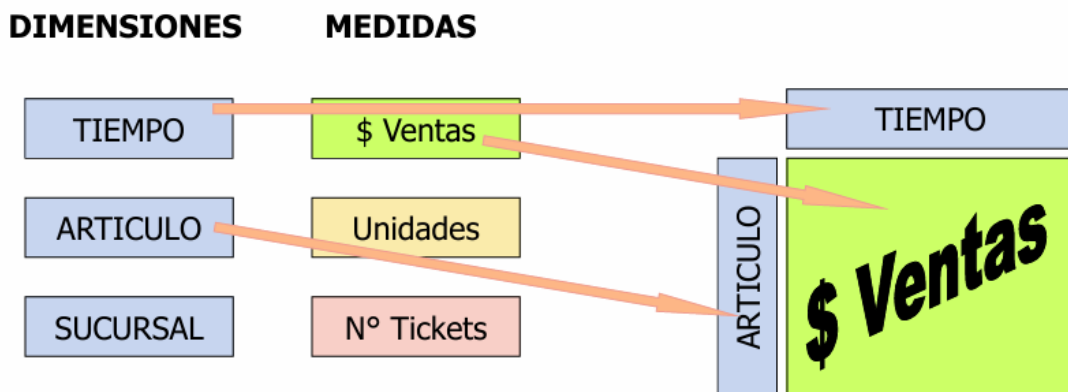


- **Drill Through:** ir desde el cubo hasta los datos fuente a nivel más bajo (SQL en tablas relacionales).

## Análisis OLAP

- Implica seleccionar:
  - Una **dimensión para filas** (ej. Producto),

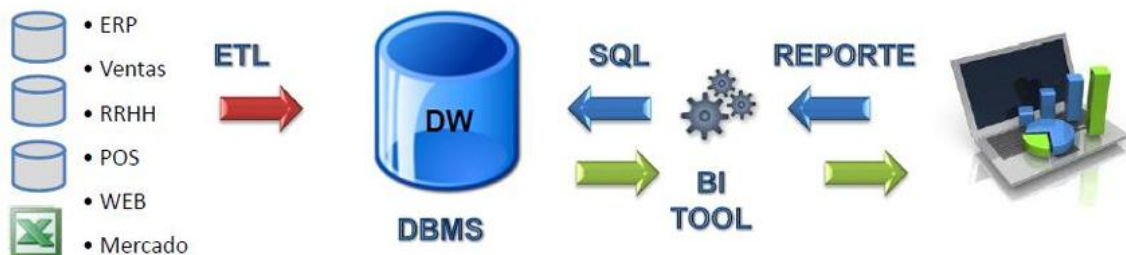
- Una **dimensión para columnas** (ej. Tiempo),
- Y una o más **medidas** (ej. Ventas, Unidades).



- Permite realizar comparaciones, aplicar filtros avanzados (ej. solo productos de perfumería entre 2011 y 2014) y generar múltiples vistas interactivas.
- Se pueden combinar operaciones OLAP en los análisis (Slice, Pivot, Drill Down).

## ROLAP (Relational OLAP)

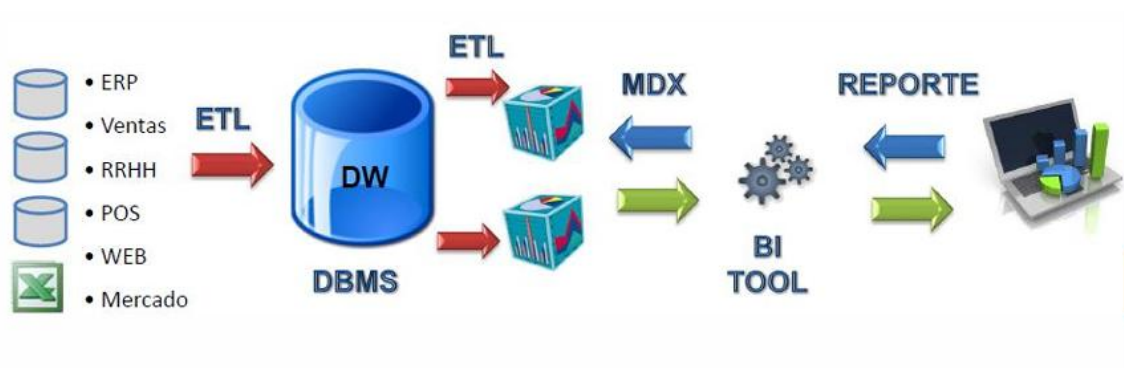
- Utiliza bases de datos relacionales como almacenamiento de datos OLAP.
- **Ventajas:**
  - Escalable.
  - Usa SQL estándar.
  - No necesita duplicar los datos.
- **Desventajas:**
  - Respuesta más lenta.
  - Necesita conexión constante a la base de datos.
  - Limitado a funciones propias del gestor relacional.





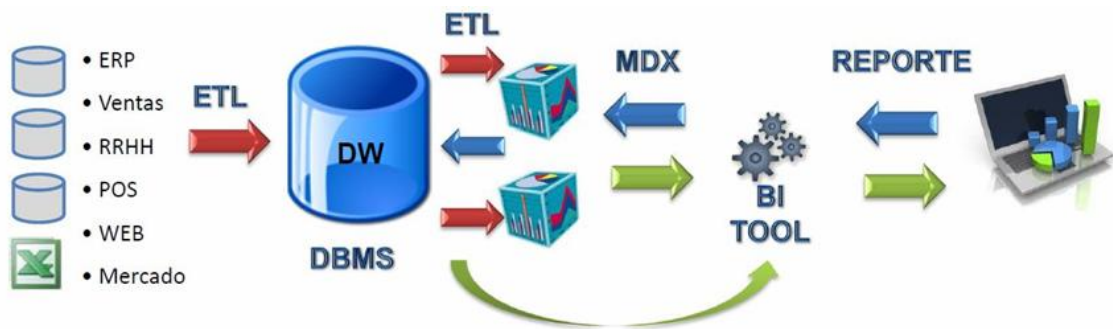
## MOLAP (Multidimensional OLAP)

- Almacena los datos en bases multidimensionales propietarias.
- **Ventajas:**
  - Consultas muy rápidas.
  - Datos precalculados y comprimidos.
  - Optimizado para el análisis multidimensional.
- **Desventajas:**
  - Mayor tiempo y recursos para cargar los cubos.
  - Dificultad con muchos niveles o dimensiones.
  - No accede a datos que no estén previamente cargados



## HOLAP (Hybrid OLAP)

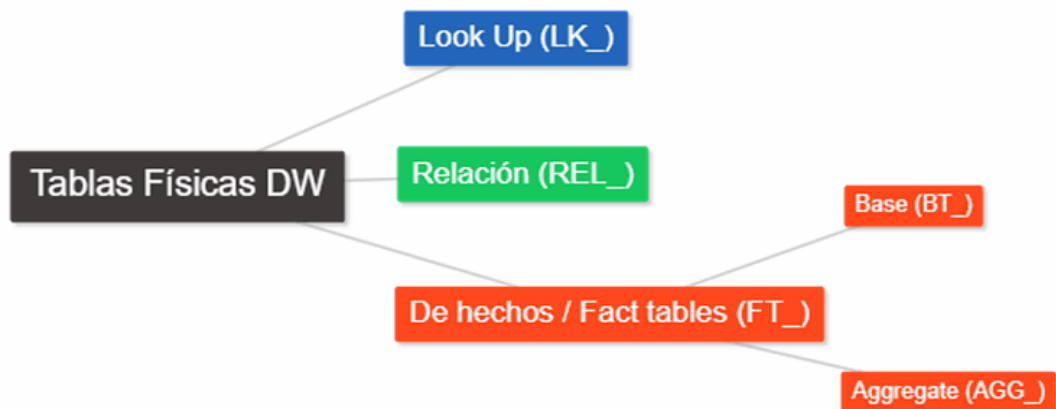
- Combina ventajas de ROLAP y MOLAP.
- **Cómo funciona:**
  - Datos agregados se almacenan en MOLAP.
  - Datos detallados en ROLAP.
- **Ventajas:**
  - Flexibilidad y velocidad.
  - Reduce tiempos de carga y espacio de almacenamiento.



## Modelo Físico

- Representa la estructura real de almacenamiento de datos en el Data Warehouse.
- Deriva del modelo dimensional lógico.
- Está compuesto por tablas, registros y columnas que soportan dimensiones y hechos.

## Estructura de Tablas



## Tablas Look Up

- Almacenan elementos de atributos descriptivos (texto, discretos).
- Se usan para **mostrar, agrupar y filtrar**.
- No participan en cálculos.
- Pueden contener:
  - Una columna por cada “attribute form” (forma del atributo).



- Una columna por cada jerarquía padre si hay jerarquía entre atributos.

#### ○ Provincia

- ID\_Provincia
- Desc\_Provincia

} *Attribute form*

1, Buenos Aires  
2, Córdoba  
3, La Pampa

} *Elementos*

**LK\_PROVINCIA**

ID_Provincia	Desc_Provincia
1	Buenos Aires
2	Córdoba
3	La Pampa
...	...

Una fila por **elemento** del atributo

Una columna por **attribute form**

#### Ejemplo:

- Tabla de Provincia: ID y descripción de cada provincia.
- Tabla de Ciudad: ID, nombre de la ciudad y referencia a su provincia.

#### Tablas Relación

- Se utilizan cuando hay relaciones de **muchos a muchos** entre atributos.

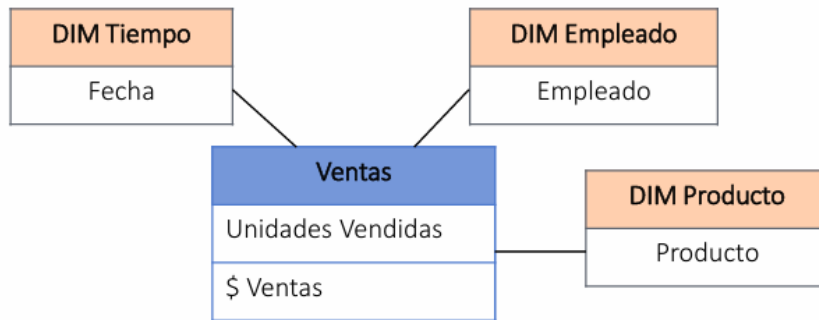
#### Tablas de Hechos (Fact Tables)

- Guardan los **valores numéricos** de los hechos (ventas, cantidad, etc.).
- Tienen una fila por cada evento o transacción.
- Son las tablas más grandes del DW.
- Contienen:
  - Columnas para cada **medida**.
  - Claves foráneas (FK) que se enlazan a las dimensiones.

#### Tabla de Hechos Base

- Almacena los datos al **nivel más detallado (granularidad)**.
- Se llena desde los sistemas OLTP.

- **Es esencial** en el DW.



**BT\_VENTAS**

ID_Fecha	ID_Empleado	ID_Producto	Unid_Vend	\$ Ventas
01/04/2020	111	17	20	\$ 400
01/04/2020	112	17	45	\$ 900
01/04/2020	113	18	30	\$ 450
...	...	...	...	...

Dimensiones

Medidas

### Tabla de Hechos Agregada

- Contiene **resúmenes** (agregaciones) para acelerar las consultas.
- Se calculan a partir de la tabla base.
- Tienen menos filas y son prescindibles a nivel informativo, pero muy útiles en rendimiento.

### BT\_VENTAS

ID_Fecha	ID_Empleado	ID_Producto	Unid_Vend	\$ Ventas
01/04/2020	111	17	20	\$ 400
08/04/2020	112	17	45	\$ 900
15/04/2020	112	17	30	\$ 450
19/04/2020	111	18	15	\$ 300

### AGG\_VENTAS\_EMP\_MES

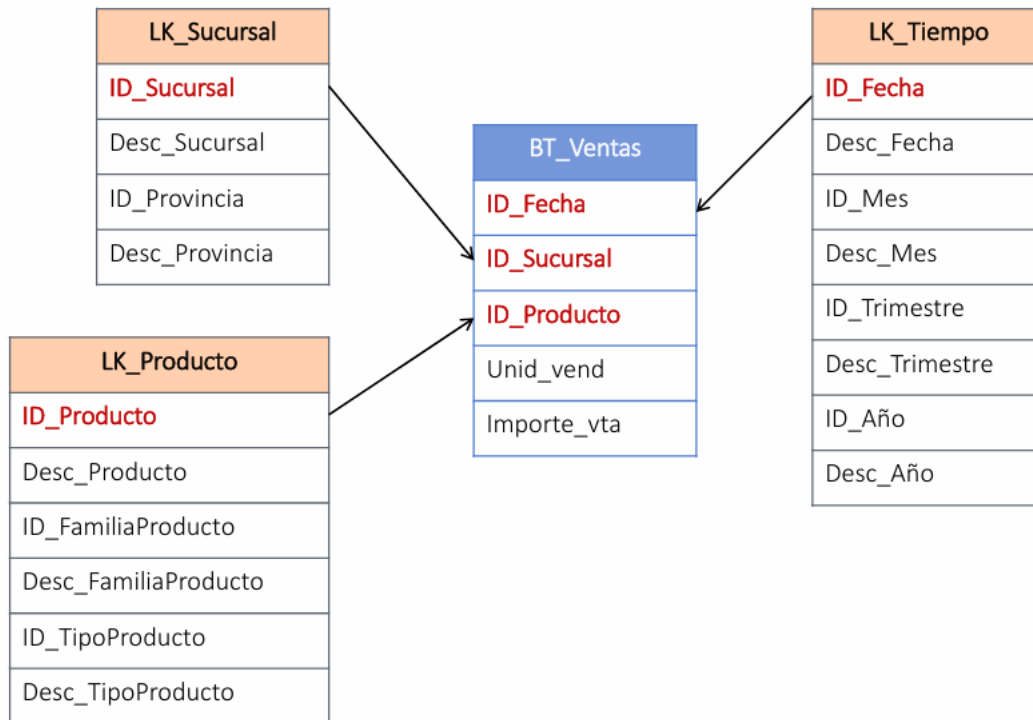
ID_Mes	ID_Empleado	Unid_Vend	\$ Ventas
202004	111	35	\$ 700
202004	112	75	\$ 1350
...	...	...	...

Contiene un atributo de alguna dimensión, por ejemplo TIEMPO

## Esquemas de Modelado

### Esquema Estrella (Star Schema)

- Una tabla por dimensión.
- Menos tablas y joins simples con la tabla de hechos.
- Mayor rendimiento, pero más espacio ocupado.
- Las dimensiones están desnormalizadas.

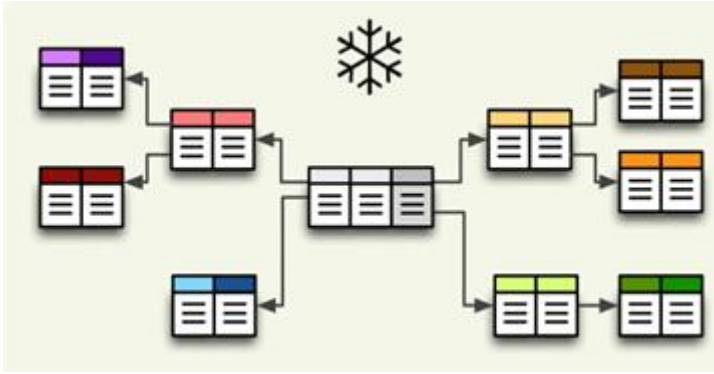


```

SELECT p.Desc_Producto, sum(v.Importe_vta) AS Importe
FROM
  BT_Ventas v,
  LK_Sucursal s,
  LK_Producto p,
  LK_Tiempo t
WHERE
  v.ID_Fecha = t.ID_Fecha and
  v.ID_Sucursal = s.ID_Sucursal and
  v.ID_Producto = p.ID_Producto and
  s.Desc_Provincia = 'CABA' and
  t.ID_Mes = '022020' and
  p.Desc_TipoProducto = 'Bazar'
GROUP BY p.Desc_Producto
  
```

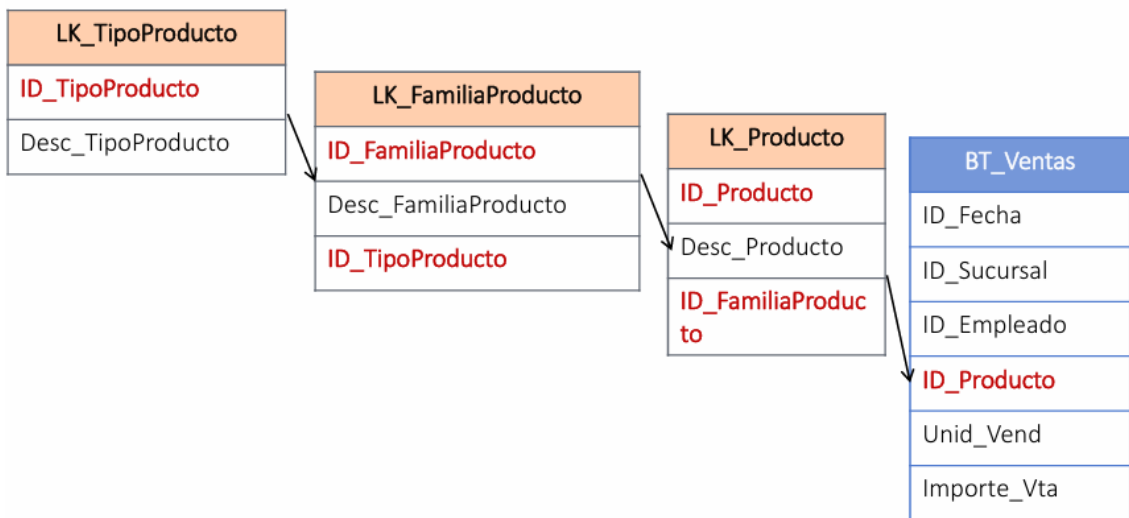
## Esquema Copo de Nieve (Snowflake Schema)

- Una tabla por cada atributo.
- Todo en 3FN: menor redundancia, pero más joins.
- Mejor para modelos ROLAP y cuando hay muchas filas en las dimensiones.

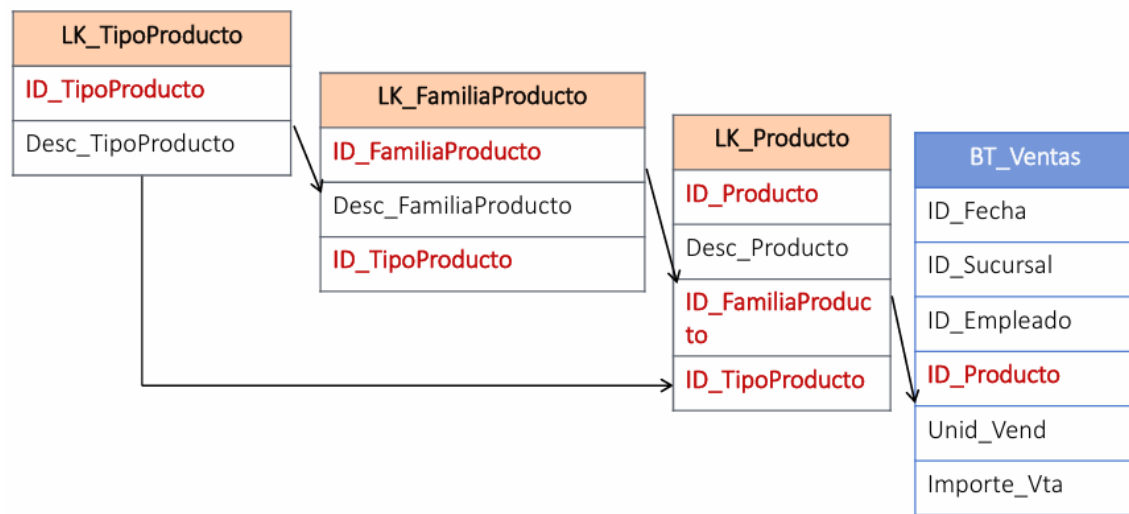


## Normalización en el Modelo Físico

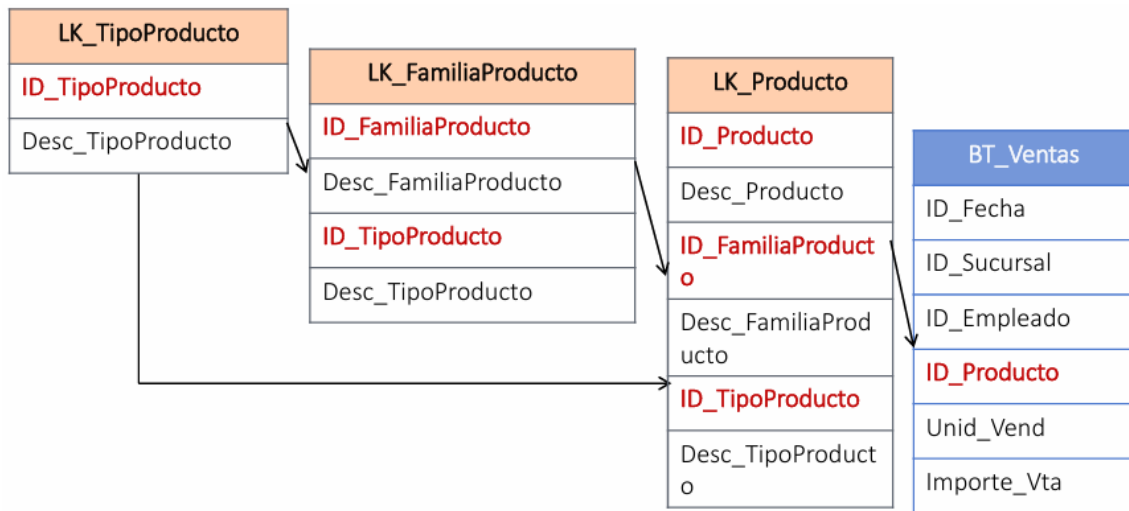
- **Completamente normalizado:** mínimas redundancias, muchos joins.



- **Moderadamente normalizado:** algún nivel de redundancia para reducir joins



- **Desnormalizado:** mucha redundancia, pero consultas muy rápidas.



## Tipos de Medidas (Hechos)

- **Aditivas:** se pueden sumar por cualquier dimensión (ej. importe, cantidad).
- **Semiaditivas:** solo se pueden sumar en algunas dimensiones (ej. stock, saldo).
- **No aditivas:** no se deben sumar, se analizan por otras funciones (ej. precio, edad, porcentaje).

## Formas de Consolidación de Medidas

- Suma, Promedio, Máximo, Mínimo, Cantidad de casos, Casos distintos.

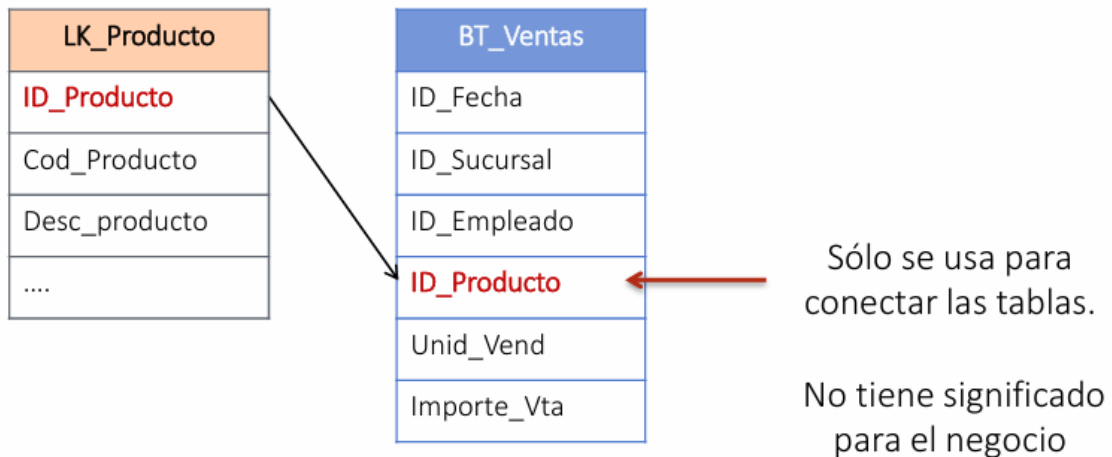
## Conexión entre Hechos y Dimensiones

- Las claves primarias (PK) de las tablas de dimensión se referencian como claves foráneas (FK) en la tabla de hechos.
- Todas las FK juntas forman la PK de la tabla de hechos.

## Clave Subrogada (Surrogate Key - SK)

- Es una clave numérica sin significado de negocio.
- Se usa para mantener integridad y control en el DW.
- Ventajas:
  - Independencia del sistema fuente.
  - Permite manejar cambios lentos.
  - Mejora el rendimiento y uso de espacio.
- Desventaja:

- Hay que gestionarlas desde el ETL.

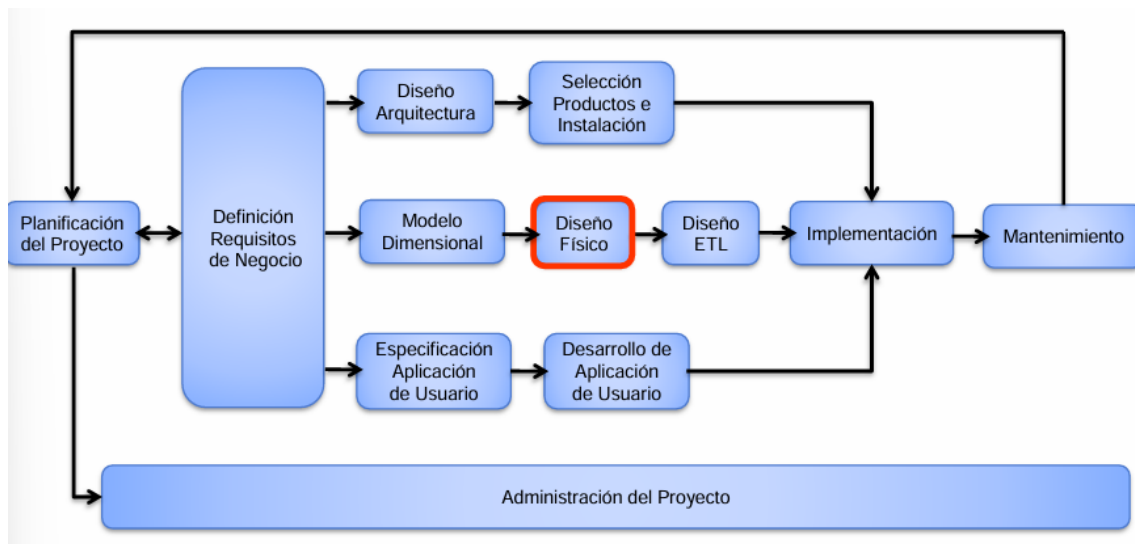


## Las 10 Reglas Esenciales del Modelado (Kimball)

1. Cargar datos atómicos en estructuras dimensionales.
2. Modelar según procesos de negocio.
3. Toda tabla de hechos debe tener dimensión tiempo.
4. Todos los hechos deben tener la misma granularidad.
5. Resolver relaciones muchos a muchos en hechos.
6. Resolver relaciones muchos a muchos en dimensiones.
7. Guardar descripciones en dimensiones.
8. Usar claves subrogadas en dimensiones.
9. Crear dimensiones conformadas para integración.
10. Escuchar siempre al usuario y adaptar el diseño.

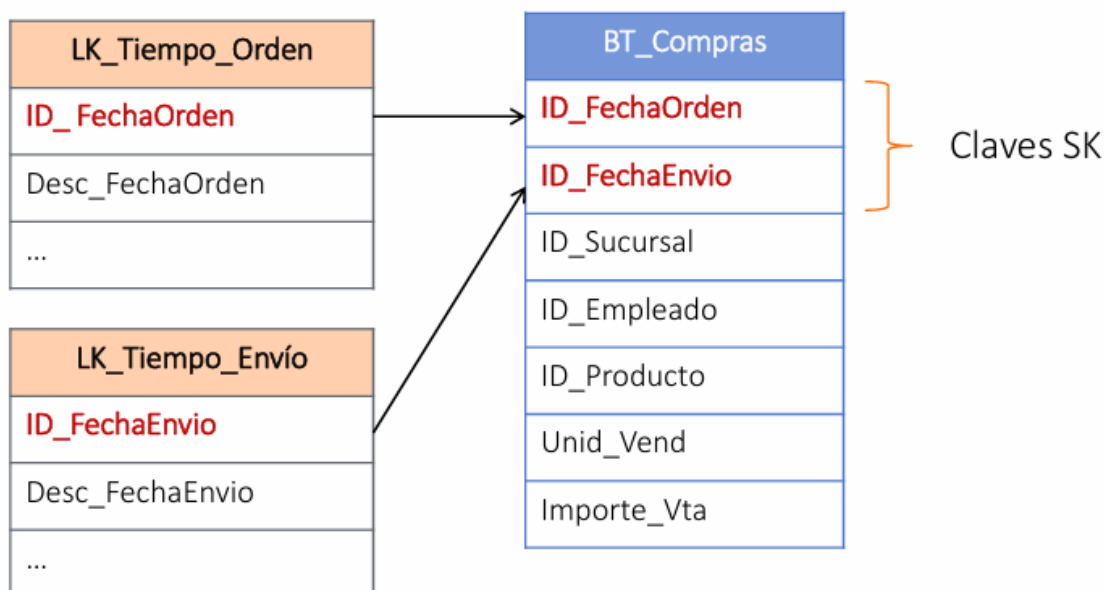
## Buenas Prácticas para Nombres en el DW

- Nombres claros y orientados al negocio.
- Sin abreviaciones ni tecnicismos.
- Coherentes y únicos.
- Pensados para ser entendidos por usuarios no técnicos.



## Dimensión Múltiple

- Una dimensión puede tener **más de un rol** en una tabla de hechos.
- Ejemplo: una compra tiene **fecha de orden** y **fecha de envío** (dos usos distintos de la dimensión Tiempo).
- Solución: crear **alias de la dimensión** por cada rol (ej. LK\_Tiempo\_Orden, LK\_Tiempo\_Envío)

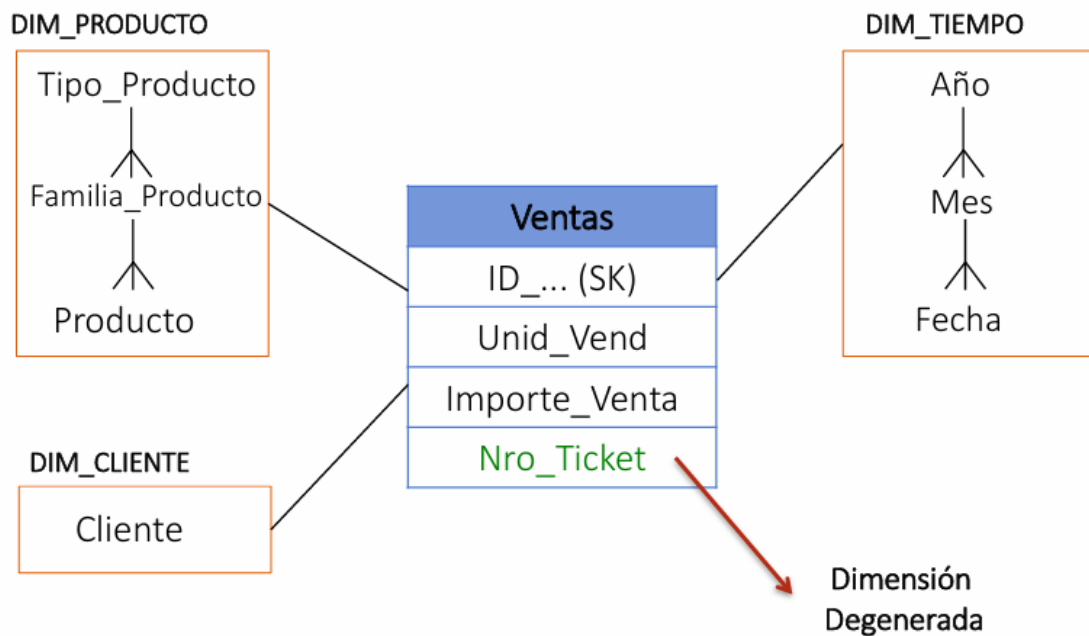


## Dimensión Degenerada

- Es un **dato transaccional** sin atributos propios, como el número de factura o número de ticket.
- No tiene tabla de dimensión asociada.
- Se almacena directamente en la tabla de hechos.



- Sirve para:
  - Vincular ítems de una misma transacción.
  - Facilitar búsquedas o análisis específicos.



**BT\_VENTAS**

ID_Fecha	ID_Cliente	ID_Producto	Unid_Vend	\$ Venta	Nro_Ticket
15/04/2017	1	10	3	\$ 90	1234
15/04/2017	1	11	2	\$ 100	1234
15/04/2017	1	12	5	\$ 150	1234
15/04/2017	3	10	2	\$ 60	2411
15/04/2017	3	13	10	\$ 200	2411

Dimensiones
Medidas
Dimensión Degenerada

## Dimensiones de Cambio Lento (SCD – Slowly Changing Dimensions)

- Algunas dimensiones cambian con el tiempo (nombre, ubicación, estado civil).
- El DW debe reflejar la historia sin perder datos.

- Estrategias SCD:

### SCD Tipo 0 – Mantener valor original

- Nunca se actualiza.
- Ejemplo: CUIT de cliente, fechas.

### SCD Tipo 1 – Reemplazo

- Se sobrescribe el valor anterior.
- No guarda historial.
- Simple de implementar, pero se pierde la historia

ID_Cliente	Cod_Cliente	NombreCliente	EstadoCivil
1	JC001	Jorge Cornejo	Soltero
2	CG015	Carlos González	Casado

ID_Cliente	Cod_Cliente	NombreCliente	EstadoCivil
1	JC001	Jorge Cornejo	Casado
2	CG015	Carlos González	Casado

ID_Producto	Cod_Producto	Desc_Producto	ID_TipoProducto	ID_RubroProducto
11	A101	Jeans	1	1

ID_Producto	Cod_Producto	Desc_Producto	ID_TipoProducto	ID_RubroProducto
11	A101	Jeans	1	2

### SCD Tipo 2 – Nuevo registro

- Se agrega una nueva fila por cada cambio.
- Se conserva el historial completo.
- Requiere columna de fechas o versiones.
- **Es el tipo más común.**

ID_Cliente	Cod_Cliente	NombreCliente	EstadoCivil
1	JC001	Jorge Cornejo	Soltero
2	CG015	Carlos González	Casado

ID_Cliente	Cod_Cliente	NombreCliente	EstadoCivil
1	JC001	Jorge Cornejo	Soltero
2	CG015	Carlos González	Casado
3	<u>JC001</u>	<u>Jorge Cornejo</u>	Casado

ID_Cliente	Cod_Cliente	NombreCliente	EstadoCivil	FDesde	FHasta
1	JC001	Jorge Cornejo	Soltero	02/06/2005	10/09/2009
3	JC001	Jorge Cornejo	Casado	11/09/2009	05/02/2017
10	JC001	Jorge Cornejo	Divorciado	06/02/2017	

○ Otra forma:

ID_Cliente	Cod_Cliente	NombreCliente	EstadoCivil	Versión
1	JC001	Jorge Cornejo	Soltero	1
3	JC001	Jorge Cornejo	Casado	2
10	JC001	Jorge Cornejo	Divorciado	3

### SCD Tipo 3 – Nueva columna

- Guarda solo **un valor histórico** (anterior) en una columna extra.
- No escala si hay muchos cambios.

ID_Producto	Cod_Producto	Desc_Producto	ID_TipoProducto	ID_RubroProducto	ID_RubroProducto Ant
11	A101	Jeans	1	1	-

ID_Producto	Cod_Producto	Desc_Producto	ID_TipoProducto	ID_RubroProducto	ID_RubroProducto Ant
11	A101	Jeans	1	2	1

ID_Cliente	Cod_Cliente	NombreCliente	EstadoCivil	FDesde	EstadoCivil_Ant
1	JC001	Jorge Cornejo	Soltero	02/06/2005	-

- Jorge se casa, entonces:

ID_Cliente	Cod_Cliente	NombreCliente	EstadoCivil	FDesde	EstadoCivil_Ant
1	JC001	Jorge Cornejo	Casado	11/09/2009	Soltero

- Jorge se divorcia, entonces:

ID_Cliente	Cod_Cliente	NombreCliente	EstadoCivil	FDesde	EstadoCivil_Ant
1	JC001	Jorge Cornejo	Divorciado	06/02/2017	Casado

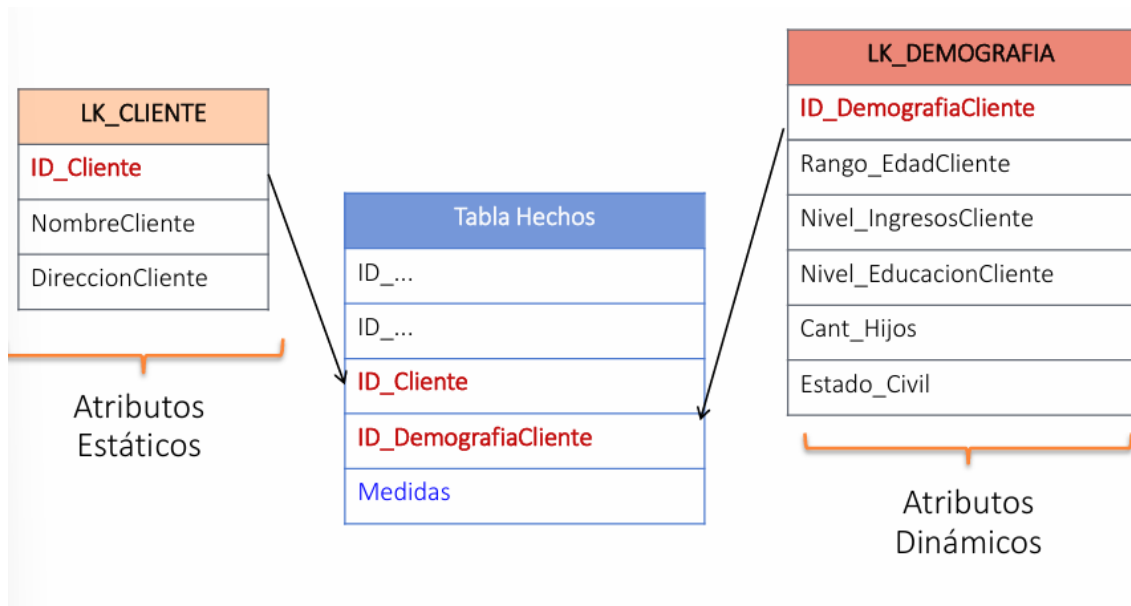
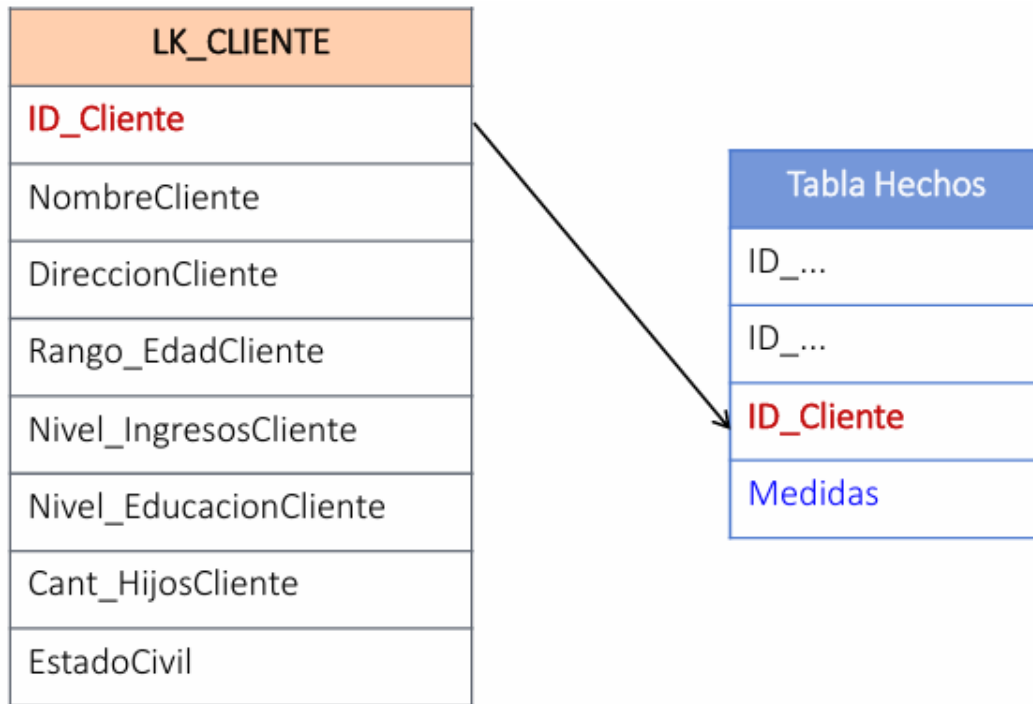
- **Nota:** Este tipo de SCD no es el más utilizado.

### Atributos con diferente SCD

- Se pueden usar distintos tipos de SCD para diferentes atributos en la misma dimensión.
- Importante para definir correctamente el ETL.

### SCD Tipo 4 – Mini Dimensiones

- Separación de atributos **frecuentemente cambiantes** en una tabla nueva (mini dimensión).
- Ejemplo: edad, ingresos, educación.
- La tabla de hechos almacena dos claves: la de la dimensión primaria y la mini.



## Dimensión Junk

- Agrupa atributos poco relevantes o de baja cardinalidad (como indicadores S/N).
- En lugar de crear muchas dimensiones pequeñas, se unifican en una sola.
- Se asigna una clave por combinación de valores posibles.

LK_Junk
ID_Junk
Promo1_IND
Descuento_IND

- Combinación de Valores:

ID_Junk	Promo1_IND	Descuento_IND
1	S	S
2	S	N
3	N	S
4	N	N

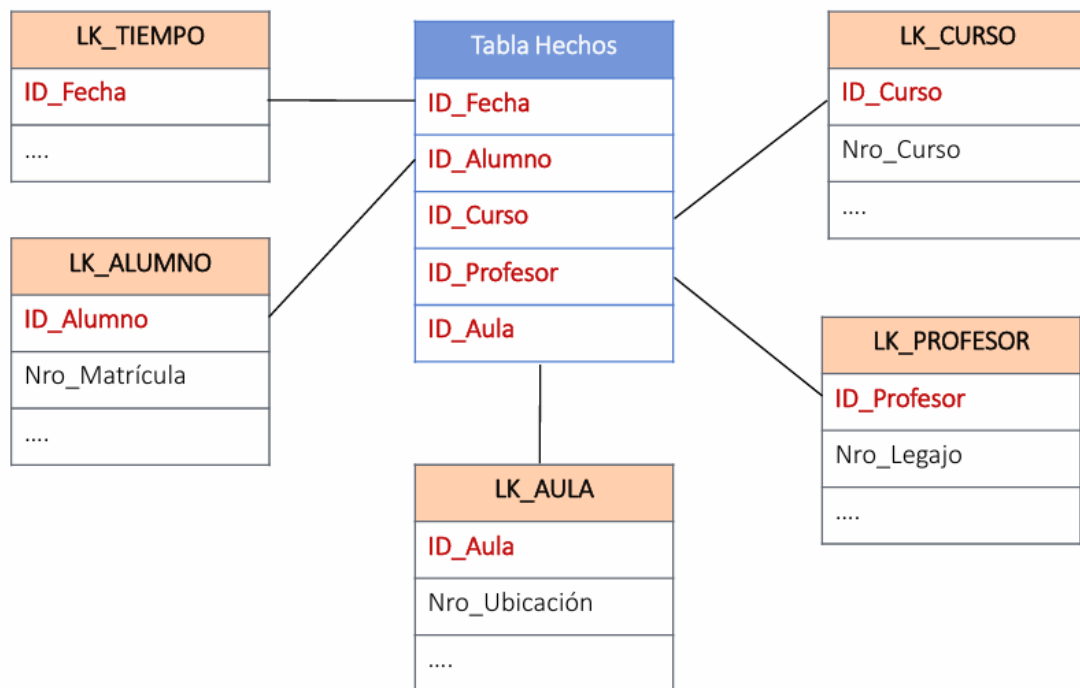
BT_Ventas
ID_Fecha
ID_Cliente
ID_Sucursal
ID_Producto
ID_Junk
Cant_UnidVendidas
Importe_Venta

## Hechos o Medidas – Casos Especiales

- A veces se necesita contar eventos sin medidas naturales.
- Solución: agregar una columna “veces” con valor **1** por fila (ej. accesos a un sitio).

## Tablas de Hechos sin Medidas (Factless Fact Table)

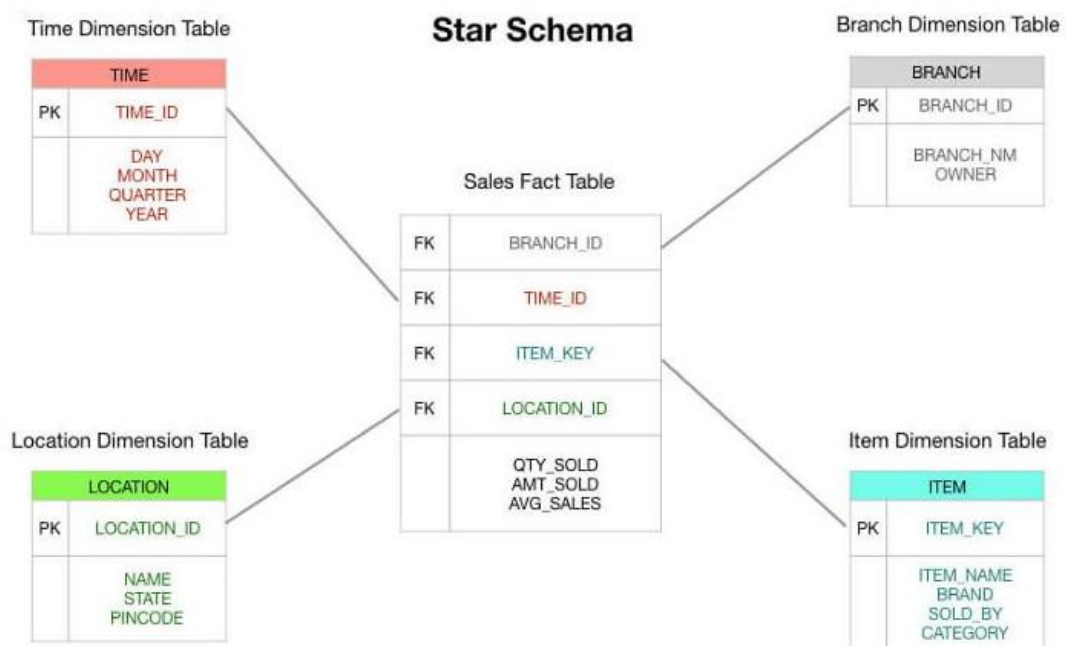
- No tienen medidas numéricas.
- Registran la ocurrencia de eventos.
- Ejemplos: asistencia, inscripciones, atención médica.
- Se usa un contador como medida derivada.



## Tipos de Tablas de Hechos Complementarias

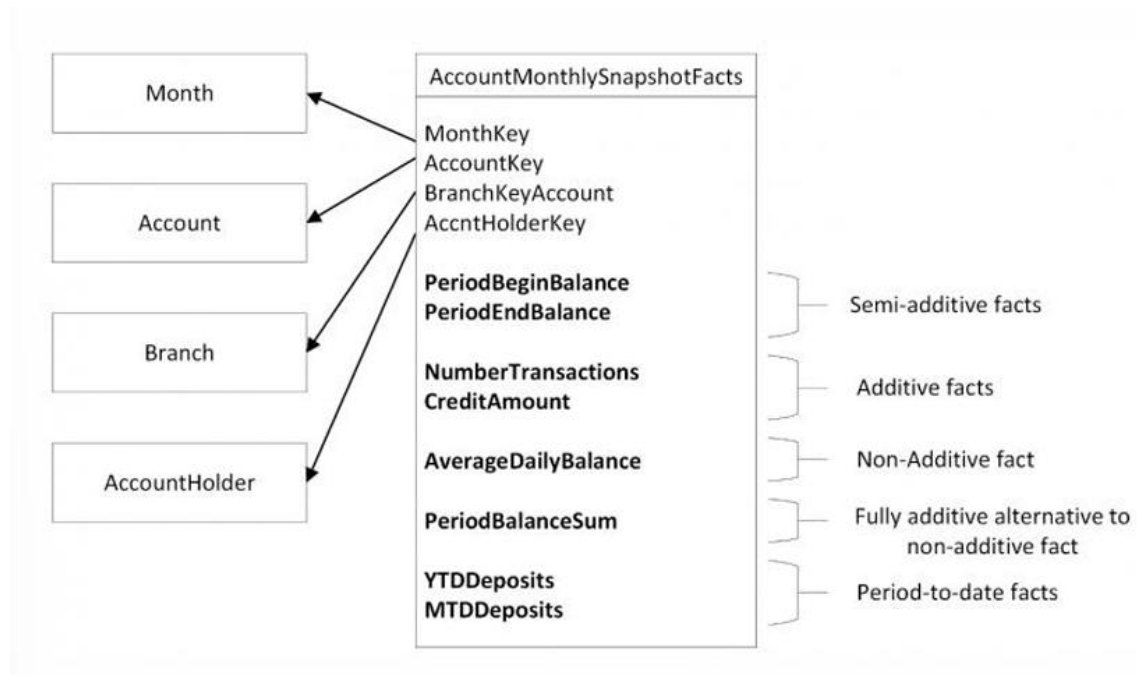
### a. Tabla de Transacciones

- Registra eventos que ocurren en un momento puntual.
- Ejemplos: venta, llamada, inscripción.



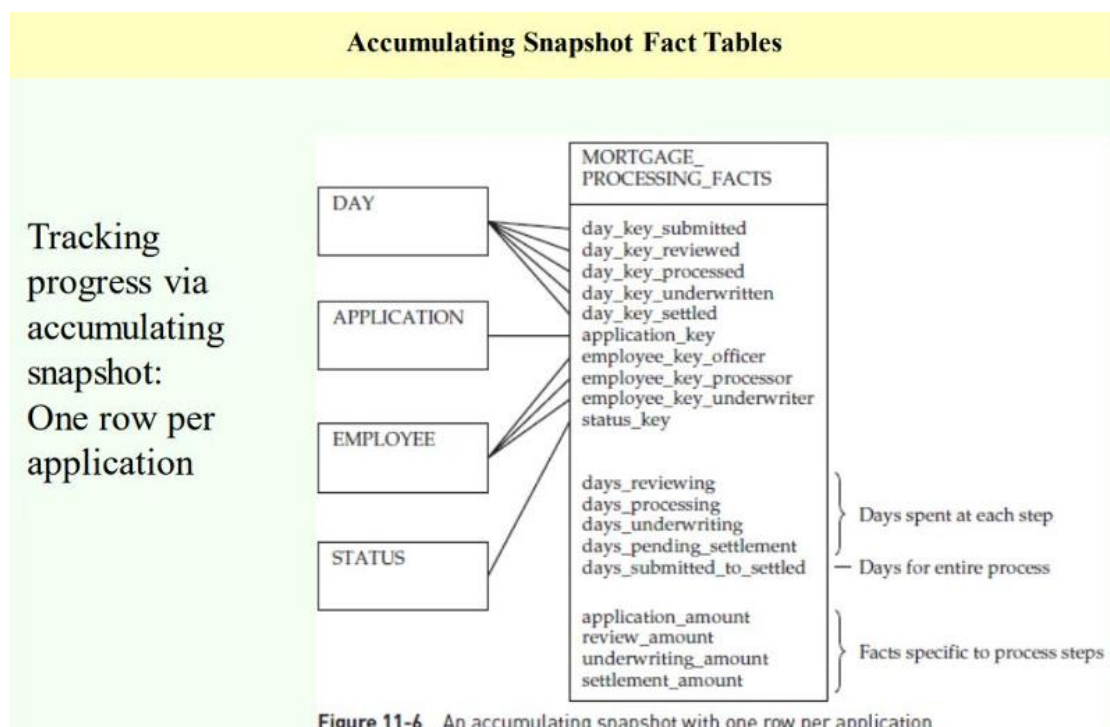
### b. Captura Periódica (Snapshot)

- Representa una “foto” en el tiempo.
- Ejemplos: stock mensual, saldo diario, notas al cierre del trimestre.



### c. Captura Acumulativa

- Registra eventos a lo largo del tiempo dentro de un mismo proceso.
- Ejemplos: seguimiento de un pedido, estado de una solicitud.
- Permite medir duración entre pasos.



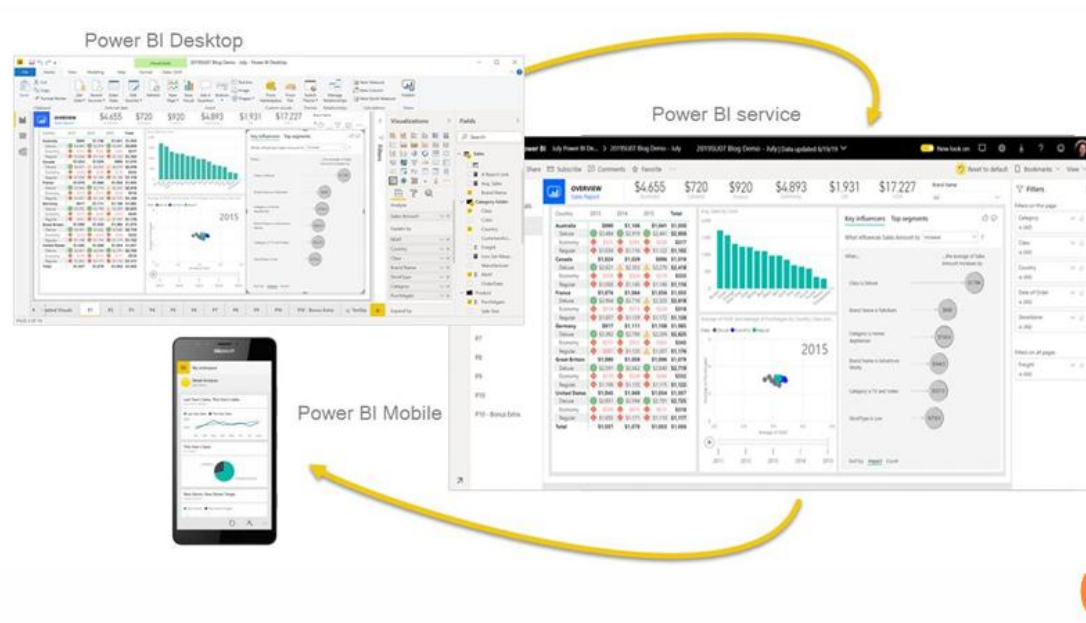


## Dimensionamiento de un Modelo

- Ejemplo realista con millones de registros.
- Cálculo del espacio requerido en disco según cantidad de registros y columnas.
- Considerar granularidad, volumen, campos índice, etc.

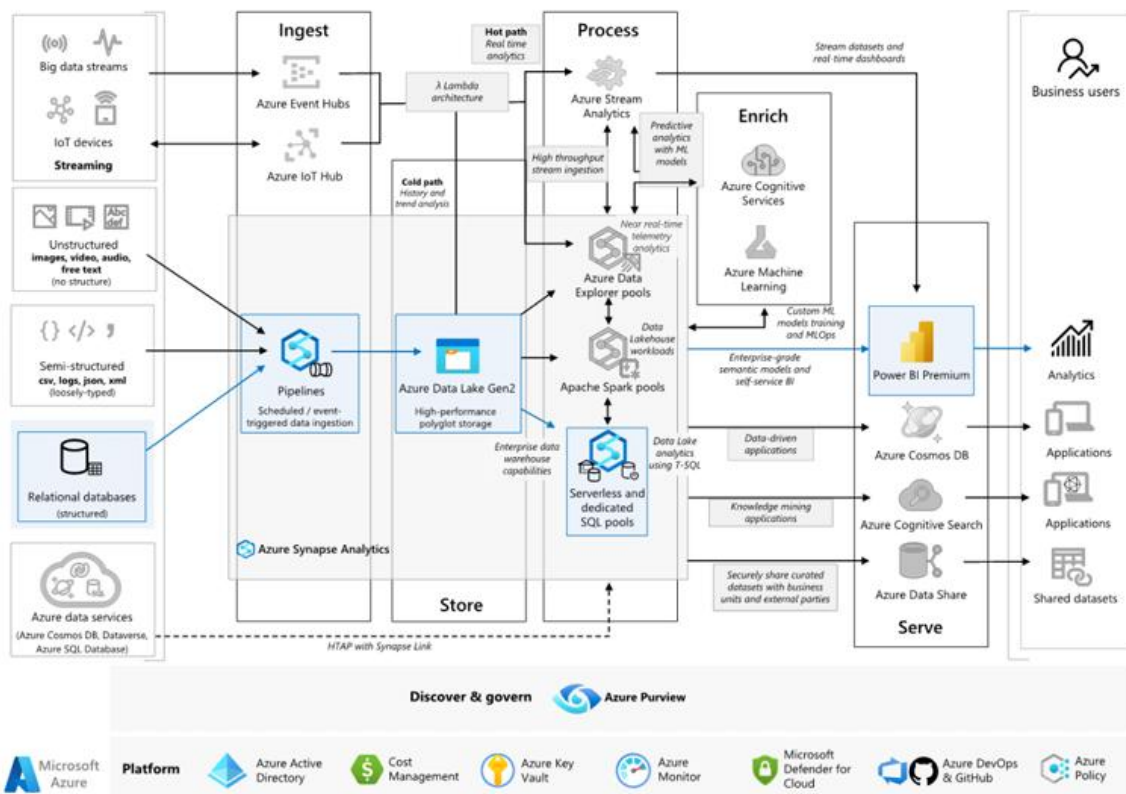
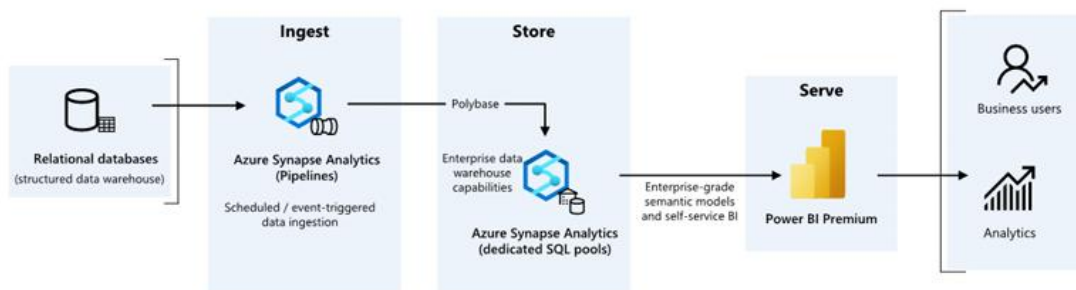
## ¿Qué es Power BI?

- Conjunto de servicios, apps y conectores de Microsoft que permiten transformar datos sin relación entre sí en información visual e interactiva.
- Sus componentes principales son:
  - **Power BI Desktop:** aplicación de escritorio para modelar y diseñar reportes.
  - **Power BI Service (SaaS):** plataforma online para compartir dashboards e informes.
  - **Power BI Mobile:** apps para visualizar informes desde dispositivos móviles.
  - **Power BI Report Builder:** para crear informes paginados.
  - **Power BI Report Server:** permite publicar informes en servidores locales.



## Arquitectura Básica de Power BI

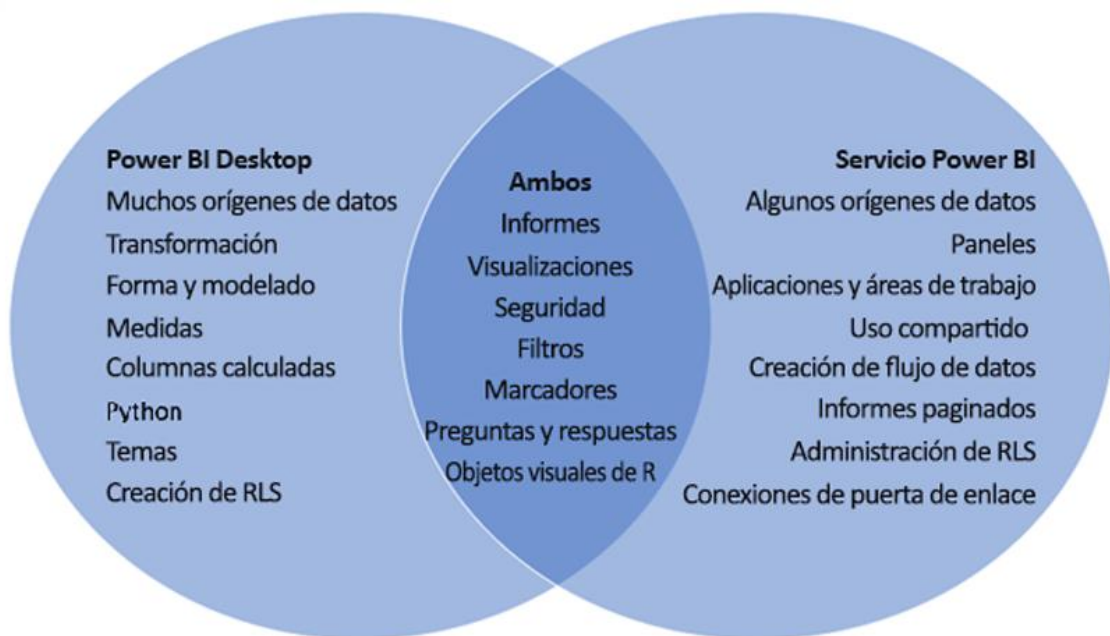
- Se compone de:
  - **Data Sources:** conectividad con una gran variedad de fuentes de datos, locales y en la nube.
  - **Gateway:** canal seguro para acceder a datos locales.
  - **Power BI Service:** se ejecuta sobre Microsoft Azure.
    - **Web Front-End:** gestiona el acceso inicial y autenticación.
    - **Backend Cluster:** administra todas las operaciones internas del usuario.



## Componentes Principales

- **Power BI Desktop:** para conectar, transformar y modelar datos.
- **Power BI Gateway:** conexión segura a datos on-premise.
- **Power BI Mobile:** visualización de dashboards en dispositivos móviles.
- **Embedded Power BI:** permite incrustar informes en aplicaciones.
- **Power BI Service:** servicio online para compartir y colaborar.

## Diferencias entre Power BI Desktop y Power BI Service



Característica	Power BI Desktop	Power BI Service
Edición del modelo	Completa	Limitada
ETL (Power Query)	Sí	Parcial
Creación de relaciones	Sí	Limitada
Publicación de informes	No	Sí
Crear áreas de trabajo	No (versión Free)	Sí (requiere Pro)

## Fases de Trabajo en Power BI

1. **GD (Get Data):** Conectar y actualizar datos.

2. **DP (Data Preparation):** Limpiar, transformar y cargar (con Power Query).
3. **DM (Data Modeling):** Relacionar tablas, crear medidas (DAX).
4. **DV (Data Visualization):** Crear informes visuales.
5. **Publicación:** Subir a Power BI Service.



## Power Query Editor

- Herramienta para transformar datos antes del modelado.
- Usa el lenguaje M.
- Se recomienda:
  - Filtrar filas y eliminar columnas innecesarias.
  - Evitar cálculos complejos aquí si pueden hacerse en origen.

## DAX (Data Analysis Expressions)

- Lenguaje para crear medidas, columnas calculadas y KPIs.
- Usar medidas en lugar de columnas cuando sea posible.
- Utilizar funciones como DIVIDE, COUNTROWS, KEEPFILTERS, etc.

## Visualizaciones

- Tipos de visuales:
  - Comparación, tendencias, partes del total, distribución, mapas, rankings, correlaciones, filtros, narrativas.
- Buenas prácticas:

- Usar pocos colores.
- No sobrecargar los informes.
- Priorizar el recorrido visual natural.

## Dashboards (Tableros de control)

- Resumen visual de KPIs y métricas clave.
- Deben ser claros, organizados y diseñados para el usuario final.
- Similares al tablero de un auto: datos clave, en un solo vistazo.

## Reportes en Línea (Power BI Service)

- Crear y compartir reportes vía web.
- Se recomienda hacer el modelo en Desktop y publicarlo al Service.
- Las versiones gratuitas tienen funcionalidades limitadas.

## Buenas Prácticas para Proyectos Power BI

- Usar modelos en estrella.
- Evitar relaciones 1:1 y N:N.
- Separar fechas y horas.
- Aprovechar Direct Query y modelos duales si es necesario.
- Usar variables DAX, documentar medidas, ocultar campos irrelevantes.
- Usar dataflows para homogeneizar y reutilizar fuentes de datos.

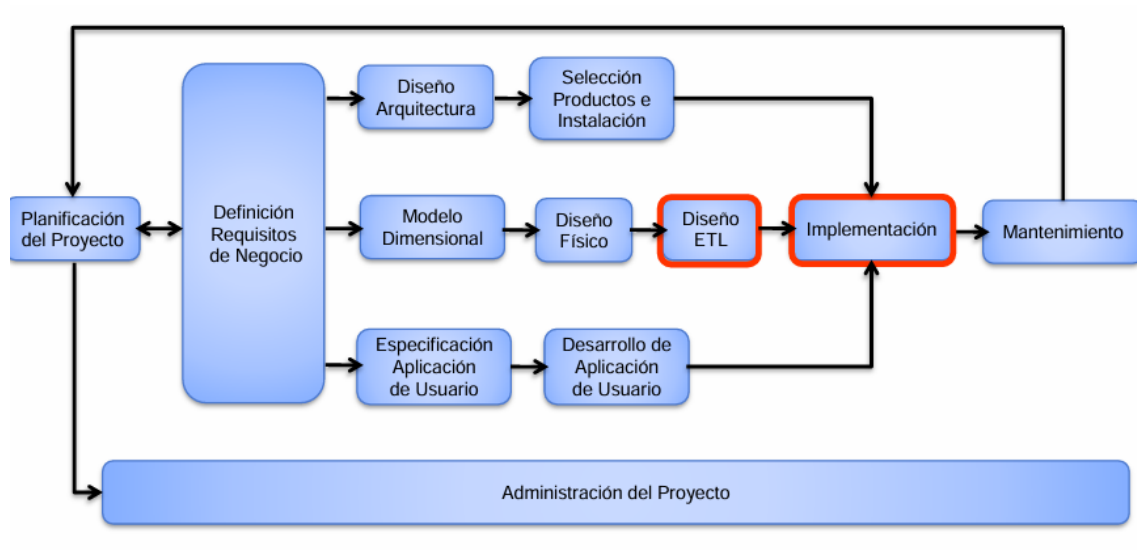
## Recursos y Enlaces Útiles

- [Documentación oficial](#)
- [Power BI Desktop](#)
- [Blog de Power BI](#)
- Canales de YouTube: Power BI, Guy in a Cube, SQLBI
- Ejemplos en GitHub: [PBIX samples](#)

## Integración de Datos

- Unifica datos desde múltiples fuentes (ERP, CRM, redes sociales, sensores).

- Clave para convertir datos crudos en información útil para la toma de decisiones.



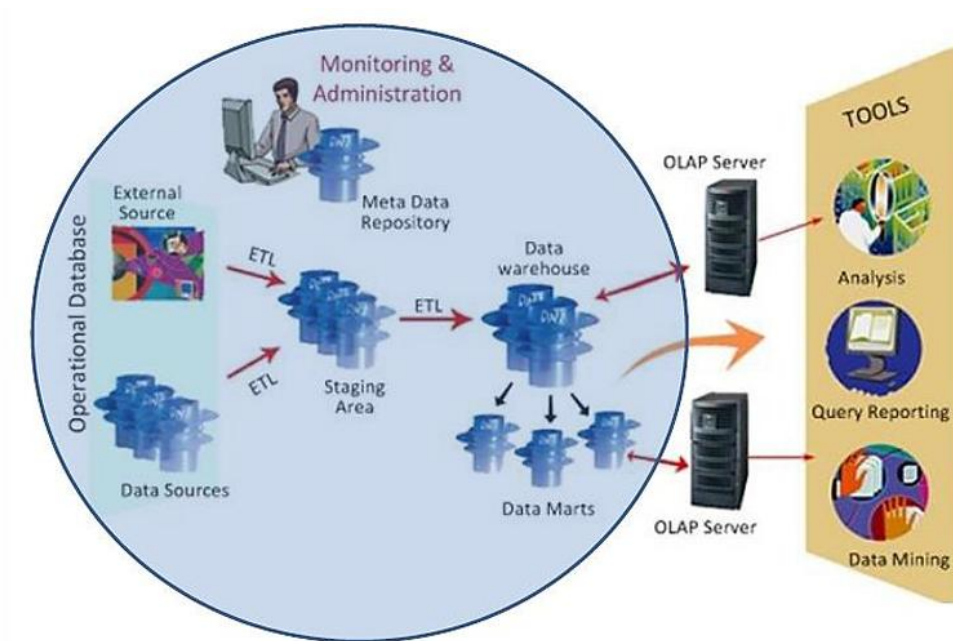
## ETL vs ELT

- **ETL (Extract, Transform, Load):** transforma los datos antes de cargarlos al DW. Ideal para estructuras tradicionales.
- **ELT (Extract, Load, Transform):** carga primero los datos y transforma después (común en la nube).

## Proceso ETL

- **Extract:** toma datos de distintas fuentes.
- **Transform:** limpia, normaliza y aplica reglas de negocio.
- **Load:** inserta en el Data Warehouse (DW).

Consume entre el 70 y 80% del tiempo de un proyecto BI.



## Funciones del ETL

- Unifica fuentes heterogéneas.
- Corrige errores y completa datos.
- Mantiene dimensiones actualizadas.
- Agrega filas en tablas de hechos.
- Genera métricas para control y auditoría.

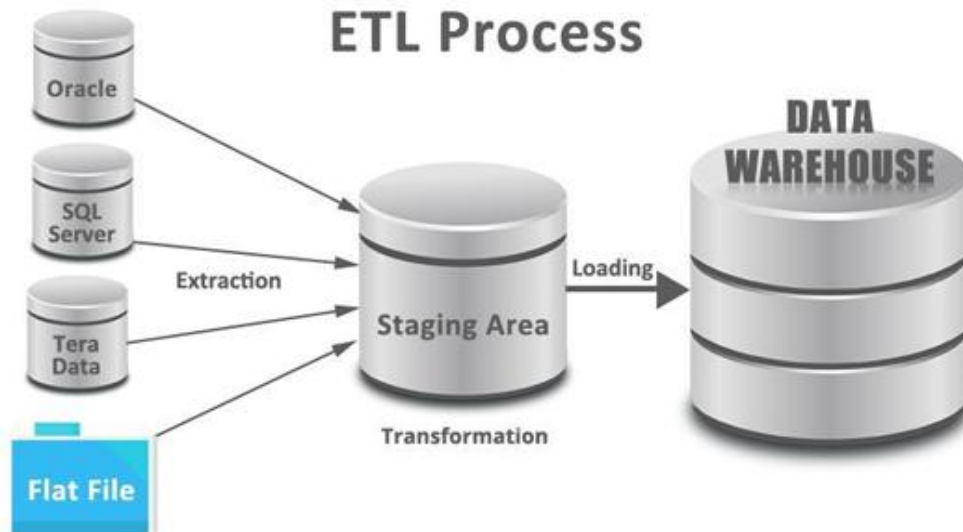
## Subprocesos Clave del ETL

- **Data Quality:** garantiza calidad alineada a objetivos del negocio.
- **Data Profiling:** analiza estructura y calidad de los datos fuente.
- **Data Cleansing:** corrige errores, inconsistencias y vacíos.

## Staging Area

- Zona temporal de carga para minimizar impacto en el sistema fuente.
- Permite recuperación de procesos ETL fallidos.





## Transformaciones Comunes

- Selección de columnas, traducción de códigos, codificación, cálculos derivados y agrupaciones.
- ELT (Extract, Load, Transform): Carga de datos sin transformar en almacenamiento en la nube para su procesamiento posterior.
- Data Mesh: Arquitectura distribuida que descentraliza la gestión de datos y permite su acceso mediante dominios organizacionales.
- Data Fabric: Enfoque unificado que automatiza el movimiento y transformación de datos en entornos híbridos y multi-cloud.

## Estrategias de Carga

- **Dimensiones:** inserts/updates.
- **Hechos:** preferentemente inserts (carga incremental).
- Preparado para cargas periódicas y recuperación ante errores.

---

## 9. Errores Típicos en los Datos

- **Incompletos:** campos faltantes.
  - **Incorrectos:** códigos erróneos, duplicados.
  - **Incomprensibles:** estructuras anómalas o formatos desconocidos.
  - **Inconsistentes:** codificaciones y reglas contradictorias.
-



# Herramientas ETL

## Comparación:

Herramientas ETL	Codificación Manual
Costosa pero completa	Menor costo inicial
Auto documentación	Manual
Conectores múltiples	Deben desarrollarse
Mantenimiento sencillo	Mantenimiento complejo
Alta trazabilidad	Difícil de implementar

Herramientas ETL	Codificación Manual
Específicas	Flexibilidad
Metodologías ETL	Técnicas de programación estándar
Personal capacitado en herramienta	Recursos propios
Mayor costo inicial	Menor costo inicial
Conectores múltiples	Se deben desarrollar conectores
Simplifica el mantenimiento	El mantenimiento se complejiza con el tiempo
Auto documentación (meta data)	Se debe desarrollar
Trazabilidad de los datos	Muy difícil de conseguir trazabilidad

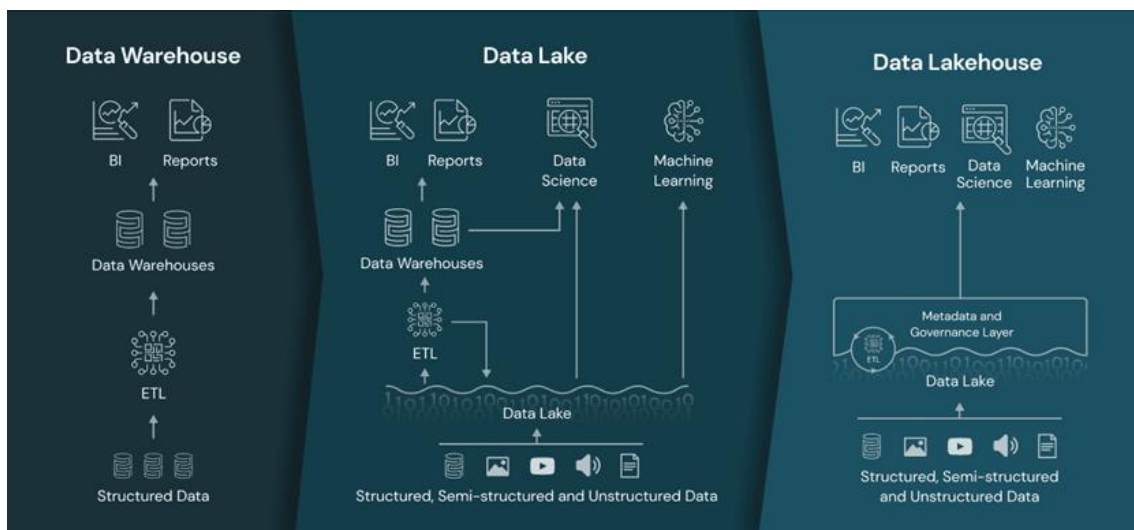
## Cuadrante Mágico de Gartner (2024)

Principales proveedores:

- **Informatica** (PowerCenter), **Microsoft** (Azure Synapse), **IBM** (DataStage), **Oracle**, **Talend**, **Denodo**, **SAP**, **SAS**.



## Arquitecturas y Tendencias BI



### Data Lake

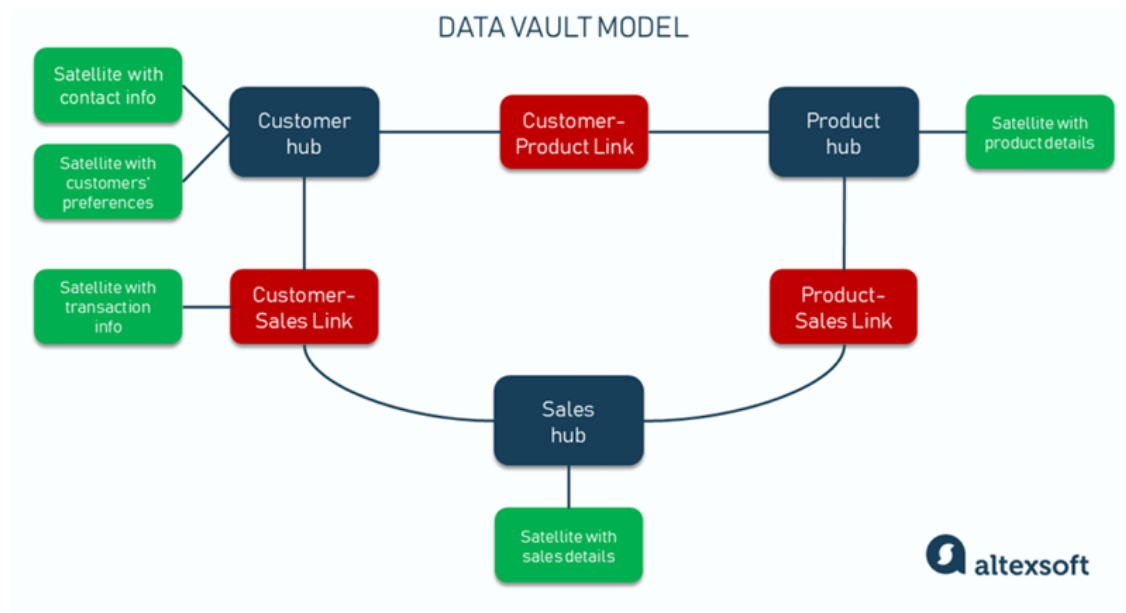
- Repositorio de datos en su forma bruta (estructurados y no estructurados).

### Data Lakehouse

- Combina la estructura de un DW con la flexibilidad del Data Lake.

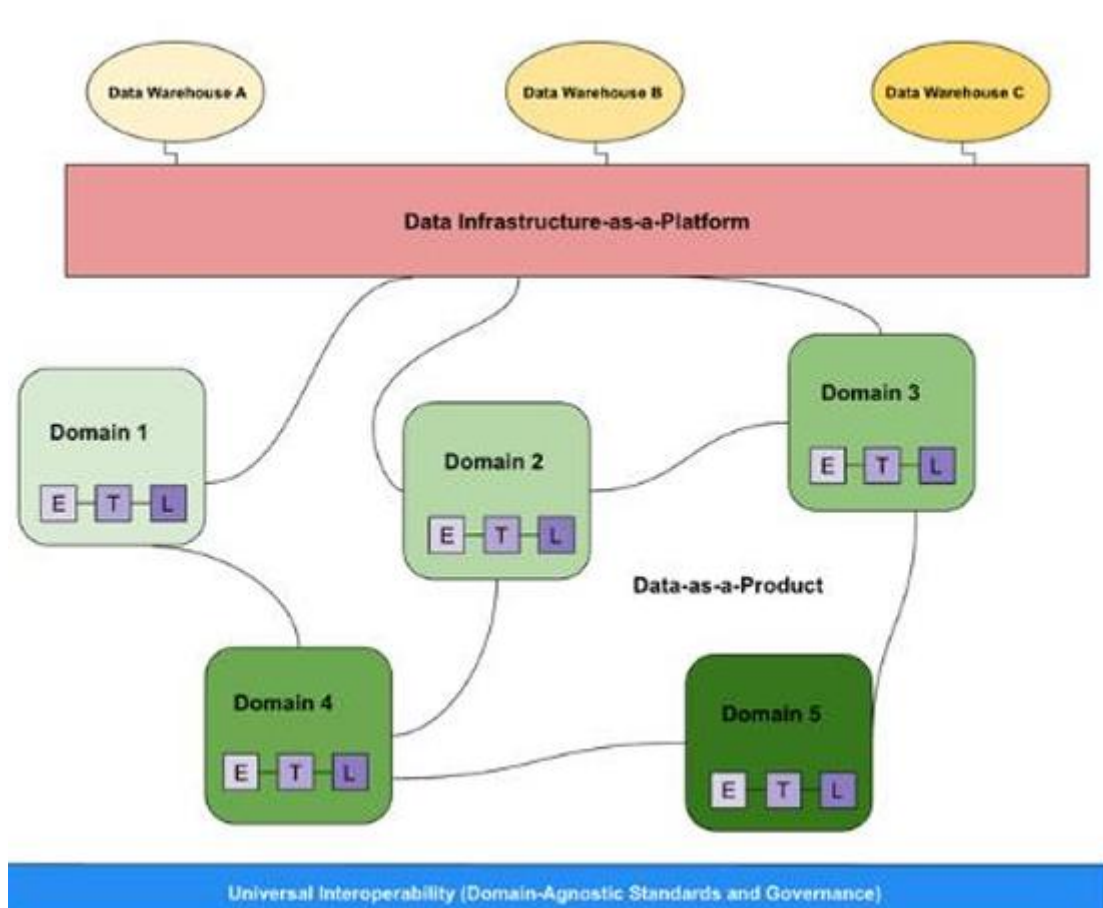
## Data Vault

- Modelo con **Hubs** (claves), **Satélites** (atributos con historial) y **Links** (relaciones).
- Ideal para auditoría y trazabilidad.



## Data Mesh

- Descentraliza el control de los datos, dando responsabilidad a los equipos de dominio



## Data Fabric

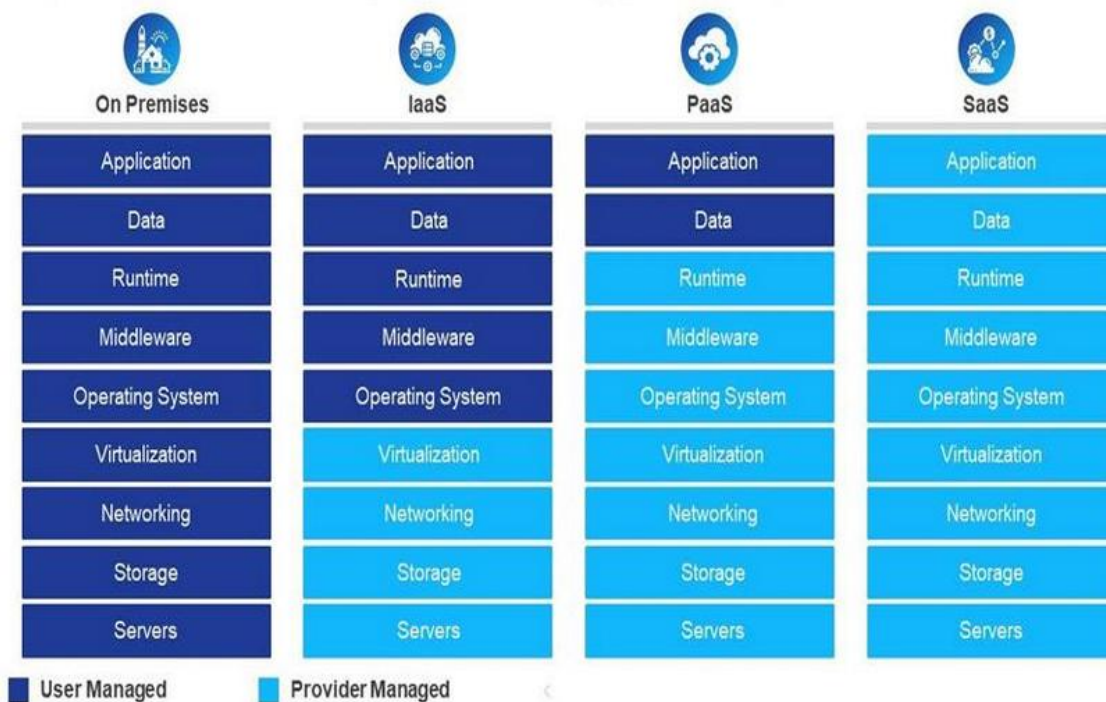
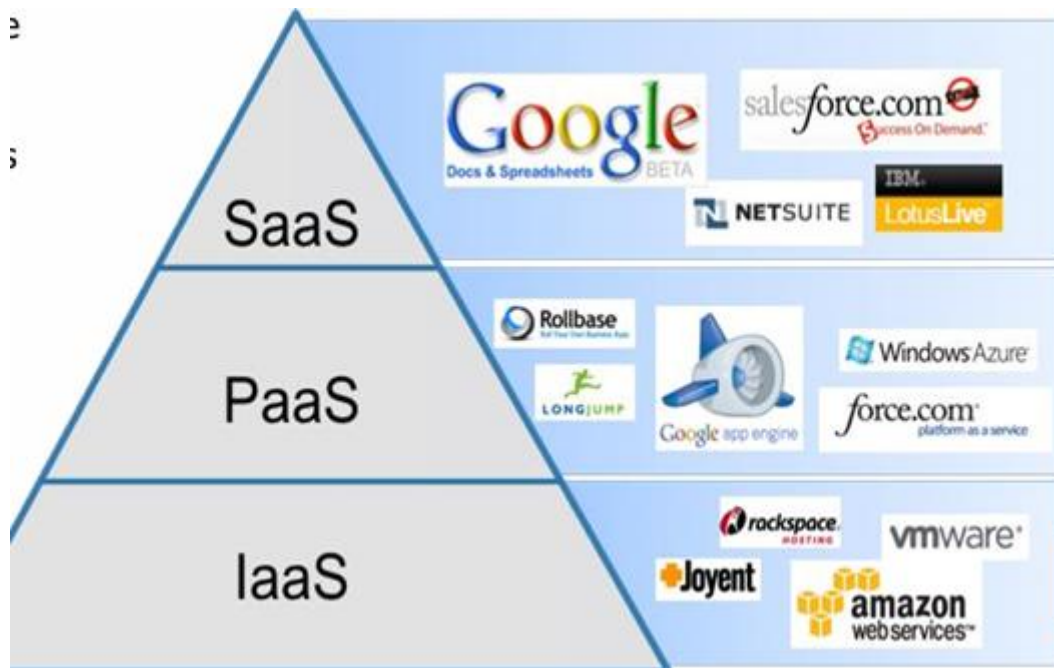
- Plataforma unificada y automatizada para integrar datos en tiempo real desde múltiples fuentes.

## Modelos de Implementación

- **On-Premise:** infraestructura local, más control y costos.
- **Cloud:** acceso remoto, escalabilidad, menor costo inicial.
- **Híbrido:** combinación flexible de ambas.

## Modelos de Servicio en la Nube

- **IaaS:** Infraestructura como servicio.
- **PaaS:** Plataforma como servicio para desarrollo.
- **SaaS:** Aplicaciones listas para usar desde el navegador.



## Comparación entre Soluciones BI

Característica	Data Warehouse	Data Lake	Data Vault	Data Mesh	Data Fabric
Tipo de datos	Estructurados	Todos	Históricos	Dominios	Multifuentes
Escalabilidad	Media	Alta	Alta	Muy Alta	Muy Alta
Gobernanza	Alta	Baja	Media	Baja	Alta

## Característica Data Warehouse Data Lake Data Vault Data Mesh Data Fabric

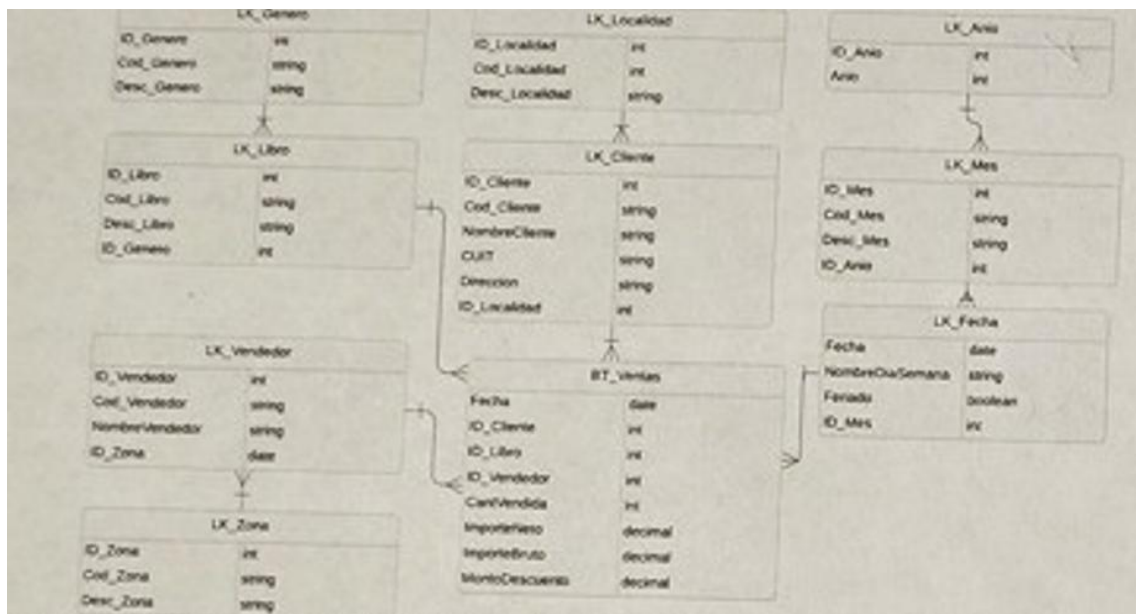
Procesamiento ETL	Sin ETL	ELT	Distribuido	Tiempo real
-------------------	---------	-----	-------------	-------------

## Ejemplos de Examen

### Ej. 1

1) Como consultor responsable del proyecto BI de una conocida cadena de venta de libros online le plantean el objetivo de brindar una solución global de las ventas incluyendo el análisis de evolución de ventas y descuentos llegando al nivel más detallado posible en base a la información disponible.

Para ello cuentan con un modelo físico ya avanzado, por lo que el Lider de Proyecto de la compañía le solicita que indique cuál sería la granularidad del mismo, usando la expresión vista en clase.



La granularidad:

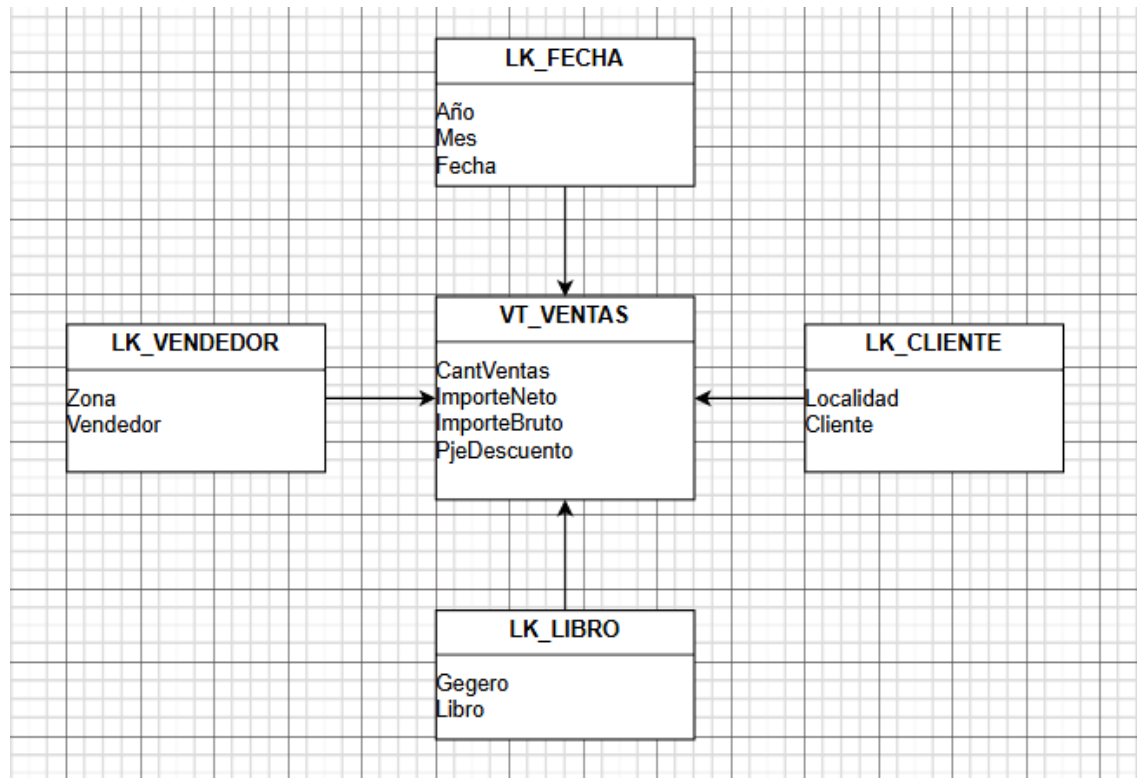
Libro x Cliente x Fecha x Vendedor

2) Continuando con sus tareas de consultoría se le pide que describa brevemente la utilidad que tiene la granularidad para el negocio. ¿Qué incluirá en dicho escrito?

La granularidad es lo que rodea a BT en nivel más bajo de detalle que se almacena en el DWH. Determina el alcance de análisis, mejor trazabilidad, flexibilidad analítica, mejores decisiones y rendimiento. Hace que este **alineado con el objetivo del negocio**

3) Dado que el proyecto comenzó con un equipo de baja experiencia en modelado dimensional aún no cuentan con la documentación correspondiente para los usuarios, por lo que el Gerente de Organización y Métodos le solicita que arme el modelo dimensional lógico.

Los atributos están ordenados en forma jerárquica



4) En una reunión de avance el DBA del proyecto le pide que indique en que tabla y que campos debería agregar para mantener el historial de cambios de la Dirección del Cliente usando el tipo 3.

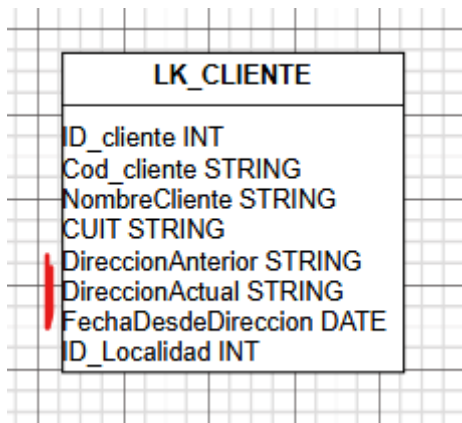
SCD1 (Pisa) Se sobrescribe el valor anterior. No guarda historial.

SCD 2 (Nuevo registro) Se agrega una nueva fila por cada cambio. Se conserva el historial completo.

SCD Tipo 3 (Nueva columna) Guarda solo un valor histórico (anterior) en una columna extra. No escala si hay muchos cambios.

Si quiere a Cliente de tipo SCD 3, se debería crear una nueva columna sobre el mismo registro, en este caso estará DireccionActual y DireccionAnterior





5) Uno de los pasantes del proyecto le pregunta: "¿Qué característica define mejor a un esquema copo de nieve comparado con el esquema estrella?"

Marque que le respondería:

- a) Menos tablas involucradas (Copo de nieve tiene más tablas)
- b) Mayor redundancia de datos (Copo de nieve reduce redundancia)
- c) Menor espacio de almacenamiento debido a la normalización
- d) Mayor facilidad en la construcción de consultas (Muchos joins)
- e) Todas las anteriores son correctas.
- f) Ninguna de las anteriores es correcta.

6) En el cuestionario de una certificación en Bi pretenden validar cuál de las siguientes afirmaciones describe mejor el concepto de Staging Area, debiendo elegir la correcta entre las siguientes opciones:

Staging Area: Zona temporal de carga para minimizar impacto en el sistema fuente. Permite recuperación de procesos ETL fallidos

- a) Sirve para insertar los registros de la tabla de hechos (No directamente)
- b) Es un espacio de almacenamiento temporal de los datos del Data Warehouse.
- c) Modifica información en las bases de datos transaccionales para el Data Warehouse. (Solo lee)
- d) Todas las anteriores son correctas.
- e) Ninguna de las anteriores es correcta.

7) Un colega le consulta que problema identifica cuando datos con el mismo nombre significan conceptos distintos. ¿Con que opción le respondería ?:

- a) Hay redundancia de datos. (No son datos repetidos, sino que el mismo nombre se usa para otra cosa)
- b) Hay inconsistencia de datos.
- c) Ha faltante de datos (No faltan datos, solo significan otra cosa)



- d) Todas las anteriores son correctas.
- e) Este problema no ocurre en el ámbito de BI. (Es común en BI)

**8) Un compañero que estudió con Ud. para este examen le pregunta cuál de los siguientes es un objetivo principal de un proyecto de BI. ¿Qué le respondería?**

- a) Aumentar la volatilidad de los datos
- b) Satisfacer necesidades tecnológicas específicas.
- c) Satisfacer las necesidades de negocio a través del análisis de datos
- d) Todas las anteriores son correctas
- e) Ninguna de las anteriores es correcta

**9) En un blog de DWH encuentra cuáles son las características correctas de una dimensión conformada. Indique las más acertadas:**

Dimensión Conformada: Es una dimensión estándar compartida por múltiples Data Marts. Ej.: dimensión **Cliente** usada en ventas, cobranzas, etc.

- a) Tiene el mismo significado en todos los Data Marts.
- b) Requiere cambios individuales en cada cubo cuando se actualiza. (Solo una vez, no en cada cubo)
- c) Puede usarse de manera diferente en cada cubo que la emplea. (No debe ser distinta)
- d) Reduce los costos de mantenimiento
- e) Ninguna de las anteriores es correcta

**10) Su jefe pretende saber cuál de las siguientes es una ventaja del enfoque de Kimball para la construcción de Data Marts. ¿Qué le respondería?**

Kimball: Construye Data Marts por prioridad de negocio y luego los integra en un DW. **Ventajas:** soluciones rápidas, más fáciles de implementar.

- a) Tener una única verdad con datos confiables. (Eso es Inmon)
- b) Poder acceder a la información sin la intervención de sistemas.
- c) Incrementar la capacidad de almacenamiento. (No necesariamente)
- d) Todas las anteriores son correctas.
- e) Ninguna de las anteriores es correcta.

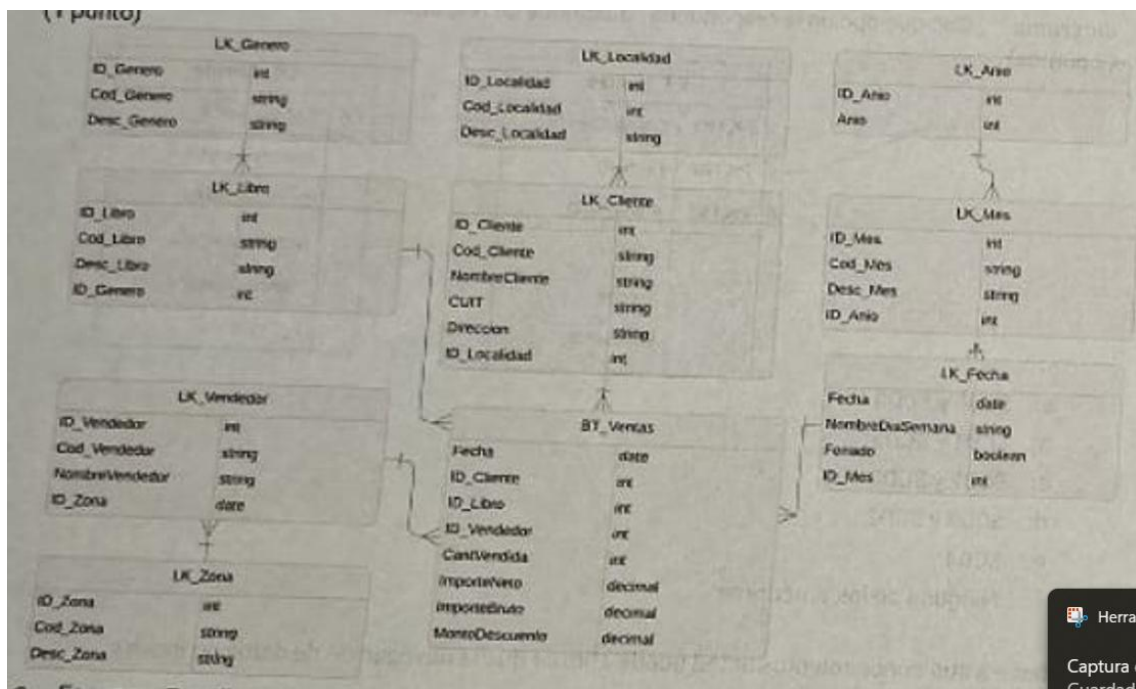
## Ej. 2

**1. Su jefe encontró un artículo referido a soluciones de Inteligencia de Negocios y le consulta: "Creo que un de las siguientes afirmaciones sobre claves subrogadas es incorrecta pero no me doy cuenta cuál puede ser." ¿Qué le respondería?**

Las claves subrogadas (Surrogate Keys) son identificadores artificiales, normalmente números enteros sin significado de negocio. Se usan para:

- Mejorar la performance del DW.
  - Simplificar las relaciones entre tablas.
  - Permitir el control de cambios históricos (SCD).
  - Asegurar independencia del sistema fuente.
- a) Son más performantes que las claves de los sistemas OLTP.
  - b) Permiten la aplicación de técnicas SCD.
  - c) **Contienen información útil para el negocio.**
  - d) Requieren menor espacio de almacenamiento.
  - e) Todas las anteriores son incorrectas.

**2. Como responsable del modelo físico de una empresa textil que está implementando soluciones de Inteligencia de Negocios les solicitan que indique a que esquema corresponde el siguiente modelo físico. ¿Con que opción respondería?**



- a) Esquema Estrella
- b) **Esquema Copo de Nieve**
- c) Esquema Dimensional
- d) Esquema Lógico
- e) Ninguno de los anteriores

**3. En base al modelo expuesto anteriormente indique cuál sería la granularidad.**

Libro x Cliente x Fecha x Vendedor

**4. Un compañero que estudió con Ud. para el recuperatorio afirma que una dimensión degenerada es aquella que no tiene atributos ¿Qué opina al respecto? Justifique su respuesta.**

Dimensión Degenerada: Es un **dato transaccional** sin atributos propios, como el número de factura o número de ticket. No tiene tabla de dimensión asociada.

Directamente en la tabla de hechos.

- a) Está en lo cierto
- b) Es incorrecto

justificación: Si es un atributo, pero no tiene tabla propia. Los datos transaccionales son como: Nro. de factura/pedido/ticket. Directamente en la tabla de hechos porque no tienen más atributos asociados

**5. Una compañera de curso que está repasando para el recuperatorio afirma que "La tabla de hechos almacena los valores de las medidas de los hechos al nivel más detallado." ¿Está de acuerdo con dicha afirmación? ¿A qué concepto se refiere? De un ejemplo de los campos básicos de una tabla de hechos (física) que permita analizar las compras de un negocio de provisión de viandas nutricionales**

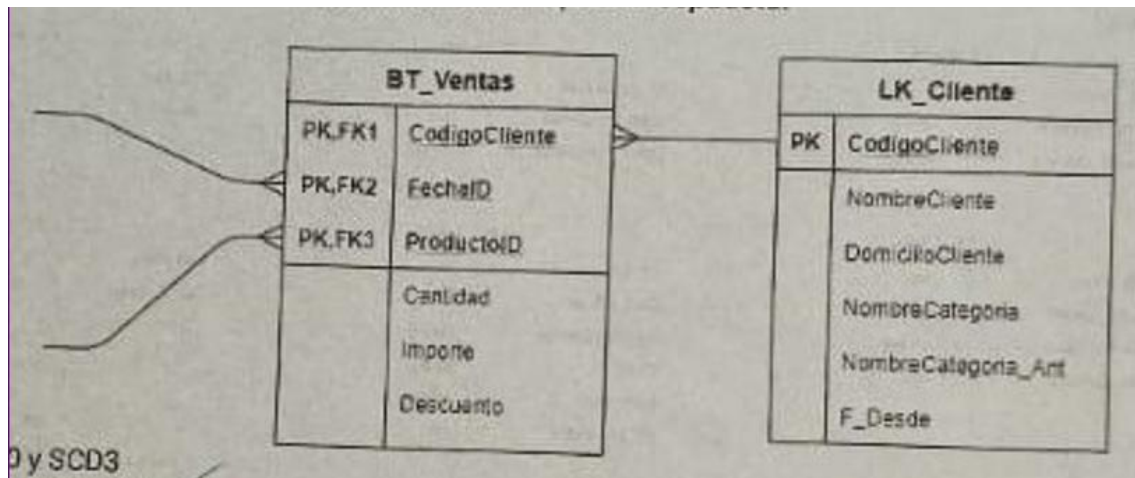
Es correcta, en la de hecho se almacenan los valores de las medidas al nivel mas detallado por la **granularidad**. Se refiere a eso, la granularidad

**6. Como responsable del proyecto Ud. sabe que al comienzo del diseño se debe identificar el proceso de negocios del Data Warehouse que va a recibir los datos porque:**

- a) Define el tamaño y la estructura de la base de datos del Data Warehouse (Se define más adelante, después del proceso)
- b) Define el grado de detalle que va a tener la información en el Data Warehouse (Esto es granularidad, después del proceso)
- c) Define las pautas de validación que se van a usar para depurar los datos
- d) Define la identidad y la finalidad del Data Warehouse (Es un paso recomendado por Kimball)
- e) Ninguna es correcta

**7. Un compañero que estudió con Ud. para el primer parcial le pregunta: "¿Qué tipos de técnicas SCD se estarán aplicando a los atributos Nombre Cliente y**

Nombre Categoría (respectivamente) en este segmento del diagrama?" ¿Con que opción le respondería? Justifique su respuesta.



- a) SCDO y SCD3
- b) SCD1 y SCD3
- c) SCD2 y SCDO
- d) SCD3 y SCD2
- e) SCD4
- f) Ninguna de las anteriores

Justificación: El nombre de cliente es SCD1 ya que solo reemplaza el nombre, pero con el Nombre de Categoría vemos que usa SCD3 ya que agrego un campo nuevo para ver la categoría Actual y la Anterior, además que la Fecha Desde

**8. En base a sus conocimientos usted puede afirmar que la navegación de datos no incluye:**

En OLAP implica moverse dentro del modelo de datos ya cargado

- a) Cambio en la granularidad de los datos (Si lo hace, pasar de ver ventas por año a ventas por mes. Drill down/roll up)
- b) Cambio de dimensiones y medidas (Si lo hace, elegir analizar por producto en vez de cliente)
- c) Conexiones a otras fuentes externas (Eso ya es extracción de datos o integración)
- d) Todas son correctas
- e) Ninguna es correcta

### Algunas auto evaluaciones de tarea

**Según la presentación, ¿qué herramienta o concepto de BI permite analizar los datos desde múltiples perspectivas y dimensiones?**

- a) Data Mining

- b) Data Warehouse
- c) OLAP (Procesamiento Analítico en Línea)

**Marca algunos de los retos para convertirse en una organización basada en datos según la presentación.**

- a) v/ Unificación de datos
- b) v/ Capacidad de almacenamiento escalable
- c) Consistencia de los sistemas transaccionales
- d) v/ Analítica ágil
- e) v/ Seguridad de datos

"El modelo dimensional está diseñado exclusivamente para la implementación de procesos operacionales."

- a) Verdadero
- b) v/ Falso

**¿Qué elemento del modelo dimensional es una categoría para describir el contexto en el que se pueden analizar las medidas?**

- a) Atributo
- b) v/ Dimensión

**¿Cuáles son los componentes principales de un modelo dimensional según la presentación?**

- a) Hechos/Medidas
- b) Dimensiones
- c) Cubos OLAP
- d) Atributos

**¿cuál de las siguientes describe mejor el concepto de "Atributos" en un modelo dimensional?**

- a) v/ Valores cualitativos de una transacción
- b) Operaciones o actividades que ocurren en el tiempo
- c) Ninguna de las definiciones es correcta

**"La granularidad en un modelo dimensional se refiere a la capacidad de agregar más dimensiones al modelo,"**

- a) Verdadero
- b) v/ Falso

**¿Qué tipo de relación entre atributos indica que cada elemento de un atributo se asocia con múltiples elementos de otro atributo dentro de la misma dimensión?**

- a) 1:1
- b) 1:N
- c) NIN