



Sistemas de Inteligencia de Negocios – DW Modelado dimensional

Prof. Lorena Zúñiga S.

Febrero, 2020

Ejemplo#1 de diseño dimensional *Retail Sales*



El proceso de modelado dimensional

1. Seleccionar el proceso de negocio que se va a modelar
2. Definir la granularidad del proceso de negocio
3. Elegir las dimensiones que aplican a cada fila en la tabla de hechos
4. Identificar los hechos numéricos (medidas) que compondrán cada fila de la tabla de hechos

Ventas al detalle en una cadena de supermercados

- 100 locales en 5 estados
- 60mil productos en cada tienda
- Cada tienda tiene los mismos departamentos:
 - abarrotes, alimentos congelados, lácteos, carnes, vegetales y frutas, panadería, floristería, artículos de salud y belleza
- Los productos individuales son llamados stock keeping unit (SKUs)

Ventas al detalle en una cadena de supermercados

- Alrededor de 55mil SKUs provienen de fabricantes externos y tienen códigos de barras en sus empaques. A los códigos de barras se les llama universal product codes (UPCs)
- Los UPCs y los SKUs tienen la misma granularidad
- Cada variación de un producto tiene su propio SKU y UPC

Ventas al detalle en una cadena de supermercados

- Los 5mil SKUs restantes vienen de departamentos como carnes, frutas y vegetales, panadería y floristería.
- Esos SKUs no tienen UPCs reconocidos a nivel nacional.
- Estos productos también llevan etiquetas con códigos de barras, pero el número es un SKU

Ventas al detalle en una cadena de supermercados

- Los datos más útiles se recolectan en las cajas registradoras.
- Los códigos de barras son escaneados e ingresados directamente al sistema de punto de venta (POS).

¿Qué necesitan los gerentes?

- Están preocupados por:
 - la logística de las órdenes
 - el inventario y la venta de productos
 - la maximización de las ganancias.
- Las decisiones más importantes de los gerentes tienen que ver con la asignación de precios y las promociones

1. Seleccionar el proceso de negocio

- Valorar los requerimientos del negocio y la disponibilidad de los datos.
- El primer modelo debe responder a las preguntas que más presión generan
- En este caso la gerencia quiere comprender las compras de sus clientes (que se registran en el POS)

1. Seleccionar el proceso de negocio

el proceso a modelar es Ventas al detalle en el POS



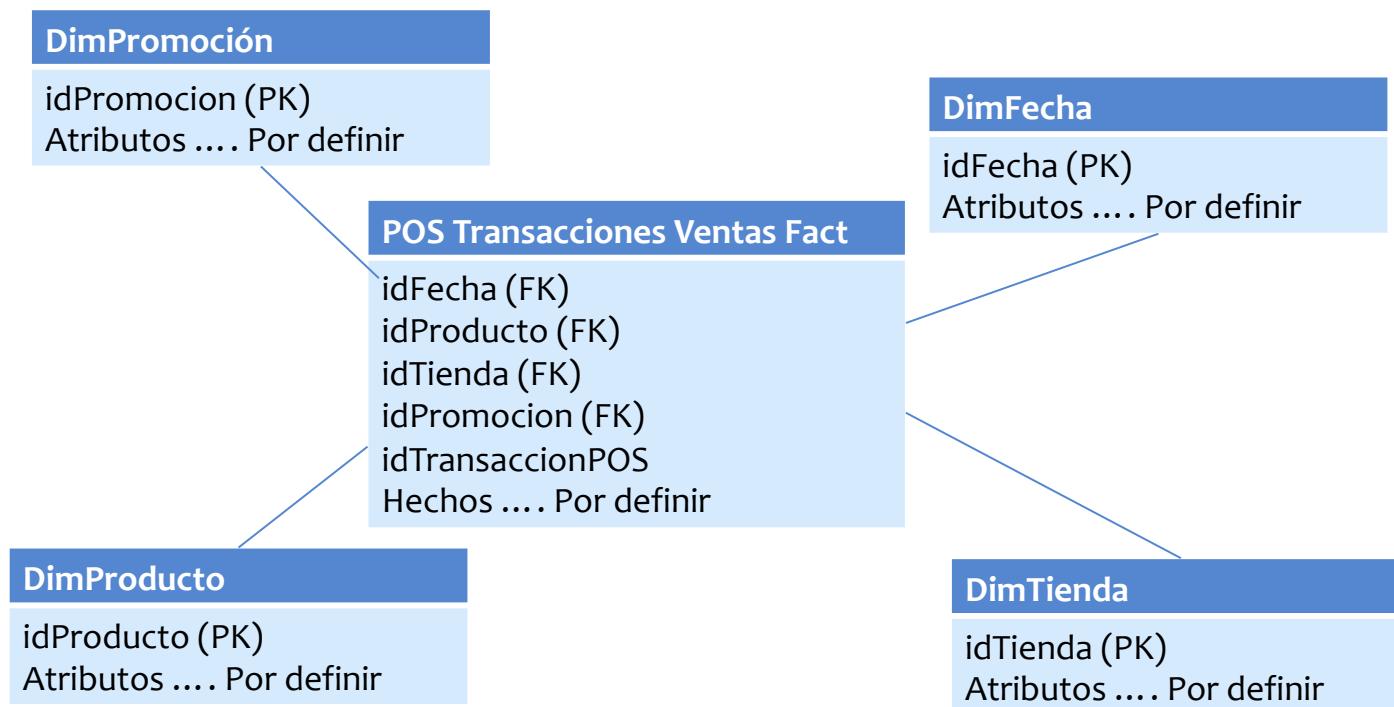
2. Definir la granularidad del proceso de negocio

- ¿Qué nivel de detalle debe tenerse en el modelo dimensional?
- Los datos más atómicos capturados durante el proceso de negocio

2. Definir la granularidad del proceso de negocio

- Se puede definir niveles menores de granularidad (menos detalle)
- En este caso de las ventas:
 - una línea de detalle en una transacción del POS

3. Elegir las dimensiones



4. Identificar los hechos

- Deben tener ***la misma granularidad*** que se definió en el paso 2.
- Los hechos que se recolectan mediante el POS incluyen:
 - cantidad vendida
 - precio unitario del producto
 - total en dólares de la venta
 - costo del producto

4. Identificar los hechos

- Monto de ventas, costo en dólares y cantidad de ventas son **medidas aditivas** a lo largo de todas las dimensiones.
- Margen bruto = ganancia bruta / ingresos.
- Margen bruto es una **medida no aditiva**

4. Identificar los hechos

POS Transacciones Ventas Fact
idFecha (FK)
idProducto (FK)
idTienda (FK)
idPromocion (FK)
idTransaccionPOS
Cantidad Ventas
Monto de Ventas
Costo
Ganancia Bruta

Dimensión Fecha

- Algunos atributos

DimFecha
idFecha (PK)
Fecha
Descripción de la fecha
Día de la semana
Número de día en el mes
Número de día en el año
Número de día en año fiscal
Número de semana en el año
Nombre del mes
Número del mes calendario
Mes-año (AAAA,MM)
Trimestre calendario
Trimestre – año
Semestre
Año calendario
Temporada de ventas
Indicador de fin de semana
Indicador de feriado
.....

Dimensión Fecha

idFecha	Fecha	Descripción	Día de la semana	Mes calendario	Año Calendario	Mes año fiscal	Indicador Feriado	Indicador Fin de semana
1	01/01/2013	1 de enero , 2013	Martes	Enero	2013	2013-1	Feriado	FinSema na
2	01/02/2013	2 de febrero , 2013	Jueves	Febrero	2013	2013-1	NoFeriado	FinSema na
3	21/09/2013	21 de setiembre, 2013	Lunes	Setiembre	2013	2013-2	NoFeriado	FinSema na

Dimensión Producto

- Algunos atributos

DimProducto
idProducto (PK)
Descripción
Número SKU
Marca
Categoría
Departamento
Tipo de empaque
Tamaño de empaque
Contenido grasa
Peso
Unidades de medida de peso
Tipo almacenamiento
Ancho estante
Alto estante
Profundidad estante
....

Dimensión Tienda

- Algunos atributos

DimTienda
idTienda (PK)
Nombre Tienda
Número de la tienda
Calle
Ciudad
Estado/Provincia
Región
Código postal
Gerente
Metros cuadrados ventas
Metros cuadrados totales
Fecha primera apertura
Fecha última remodelación
.....

Dimensión Promoción

Describe las condiciones de la promoción bajo la cual se vendió un producto

DimPromoción

idPromoción (PK)
Nombre de la promoción
Tipo de reducción de precio
Tipo de anuncio
Tipo de cupón
Tipo de medio
Tipo de display
Proveedor del display
Costo de la promoción
Fecha inicio promoción
Fecha final promoción
.....

Ejemplo #2

Etapas para el diseño del datamart

1. Identificar el proceso a modelar
2. Identificar el nivel de detalle
3. Diseño de las Dimensiones
4. Diseño de la Tabla de Hechos
5. Definición de las Medidas

Diseño de las dimensiones

- Se determina la necesidad de analizar los datos de la empresa según el siguiente cuadro:

Medidas	Dimensiones				
	Tiempo	Sucursal	Vendedor	Cliente	Producto
Importe_Ventas	X	X	X	X	X
Costo_Ventas	X	X	X	X	X
Unidades_Ventas	X	X	X	X	X
ImporteTotal_Ventas	X	X	X	X	X
Ganancia	X	X	X	X	X
Promedio_Ventas	X	X	X	X	X

Diseño de las dimensiones

Tiempo

- La empresa necesita analizar sus datos diariamente
- Se pueden definir entonces los niveles: Año, Semestre, Trimestre, Mes, Día o fecha específica

DimTiempo
IdTiempo
Año
Semestre
Trimestre
NombreMes
Fecha

Diseño de las dimensiones

Sucursal

- Tendrá un esquema estrella y su estructura jerárquica será:

DimSucursal
IdSucursal
CodSucursal
Tipo Sucursal
País
Provincia
Ciudad

Diseño de las dimensiones

Vendedor

- Tendrá un esquema estrella y quedará definida por los niveles siguientes:

DimVendedor
IdVendedor
CodVendedor
Sucursal
Departamento
NombreVend

Diseño de las dimensiones

Cliente

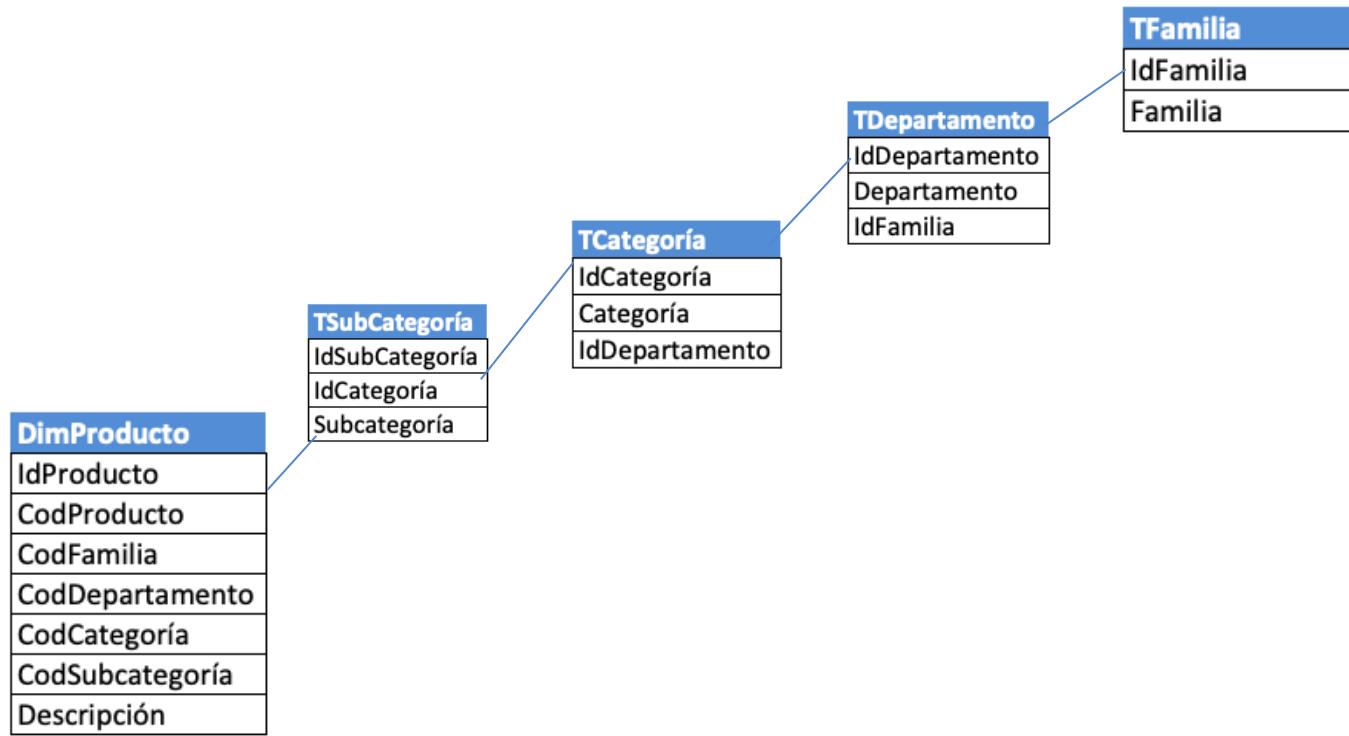
- Contiene también una jerarquía

DimCliente
IdCliente
CodCliente
País
Provincia
Ciudad
Razón Social

Diseño de las dimensiones

Producto

- Tendrá un esquema copo de nieve.
- El cliente puede clasificar sus productos según la categoría, subcategoría, el departamento y la familia de producto a la que pertenece



Diseño de la tabla de hechos

- Primero se colocan las columnas llaves de la tabla
- La empresa necesita medir:
 - El costo de cada unidad vendida
 - el valor de venta de cada producto
 - la ganancia obtenida en la venta de cada producto.

Diseño de la tabla de hechos

- La medida “*ganancia obtenida en la venta de cada producto*” puede ser calculada a partir de las medidas ventas importe y ventas costo.

Diseño de la tabla de hechos

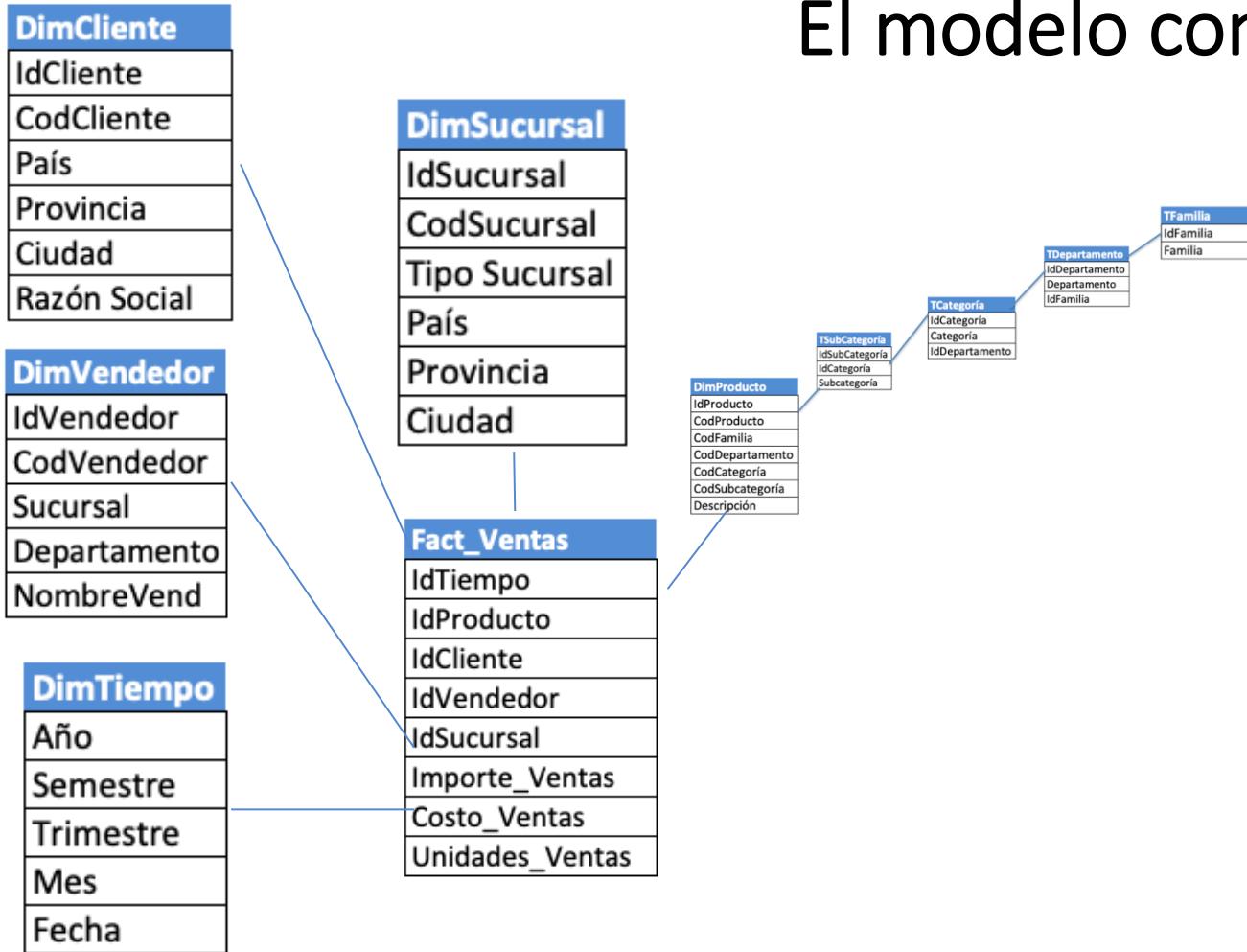
- Se agregan a la tabla de hechos ventas estas medidas:

Fact_Ventas
IdTiempo
IdProducto
IdCliente
IdVendedor
IdSucursal
Importe_Ventas
Costo_Ventas
Unidades_Ventas

Diseño de la tabla de hechos

- Se tienen también las siguientes medidas calculadas:
- **Ventas_Ganacia** que se calcula como Ventas_Importe - Ventas_Costo
- **Ventas_Promedio** que será el resultado de la suma de Ventas_Unidades / cantidad de días, comprobando la condición del numerador diferente de cero.

El modelo completo



Integracion, limpieza y
transformacion del almacen de
datos

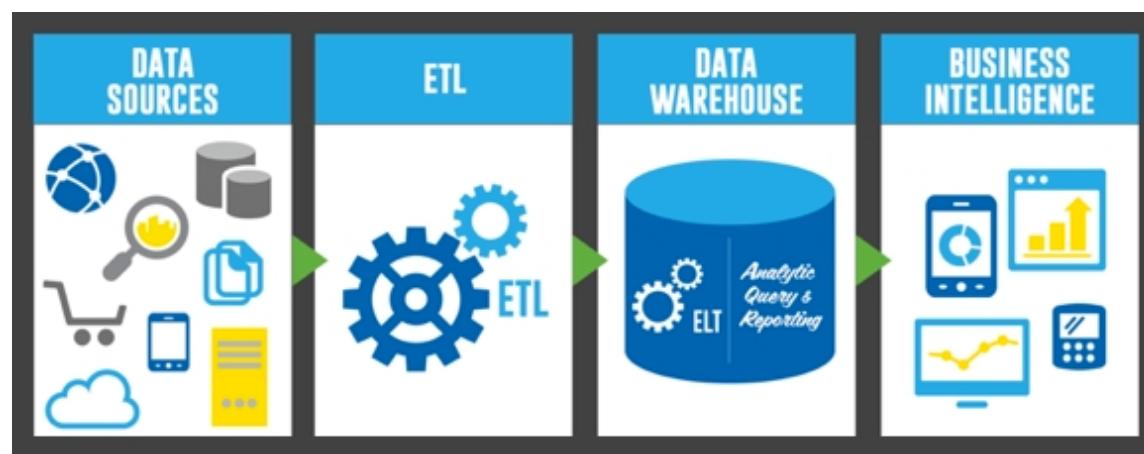
ETL

- Establecer las reglas que serán utilizadas para realizar la transformación.
- Detectar las inconsistencias que pueden originarse al tomar los datos desde distintas fuentes.
- Planificar la transformación de los datos para obtener como resultado final conjuntos de datos consistentes.

ETL

- Codificación, unidades de medida, formatos, unión de columnas (ej: Nombre, Apellido1,Apellido2 en Nombre), una columna a varias.
- Granularidad de los datos

ETL



Extract – Transform - Load

- La extracción supone que hay previamente una selección (requerimientos de los usuarios)
- Las tres funciones son igualmente importantes
- Es un proceso que consume mucho tiempo y esfuerzo

ETL - Dificultades

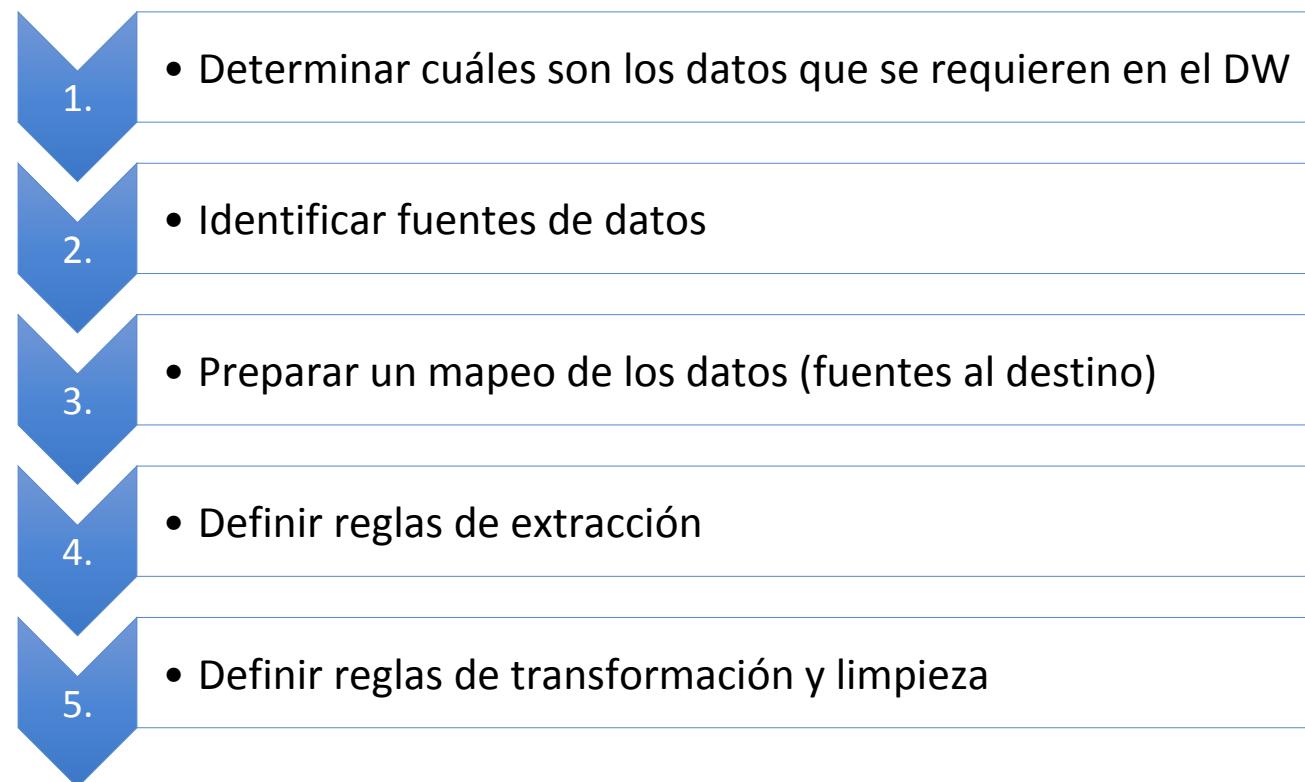
- Algunas razones:
- Diversidad de sistemas transaccionales
- Múltiples plataformas
- Sistemas legados
- No se actualizan las versiones existentes de los datos
- Problemas de calidad de datos



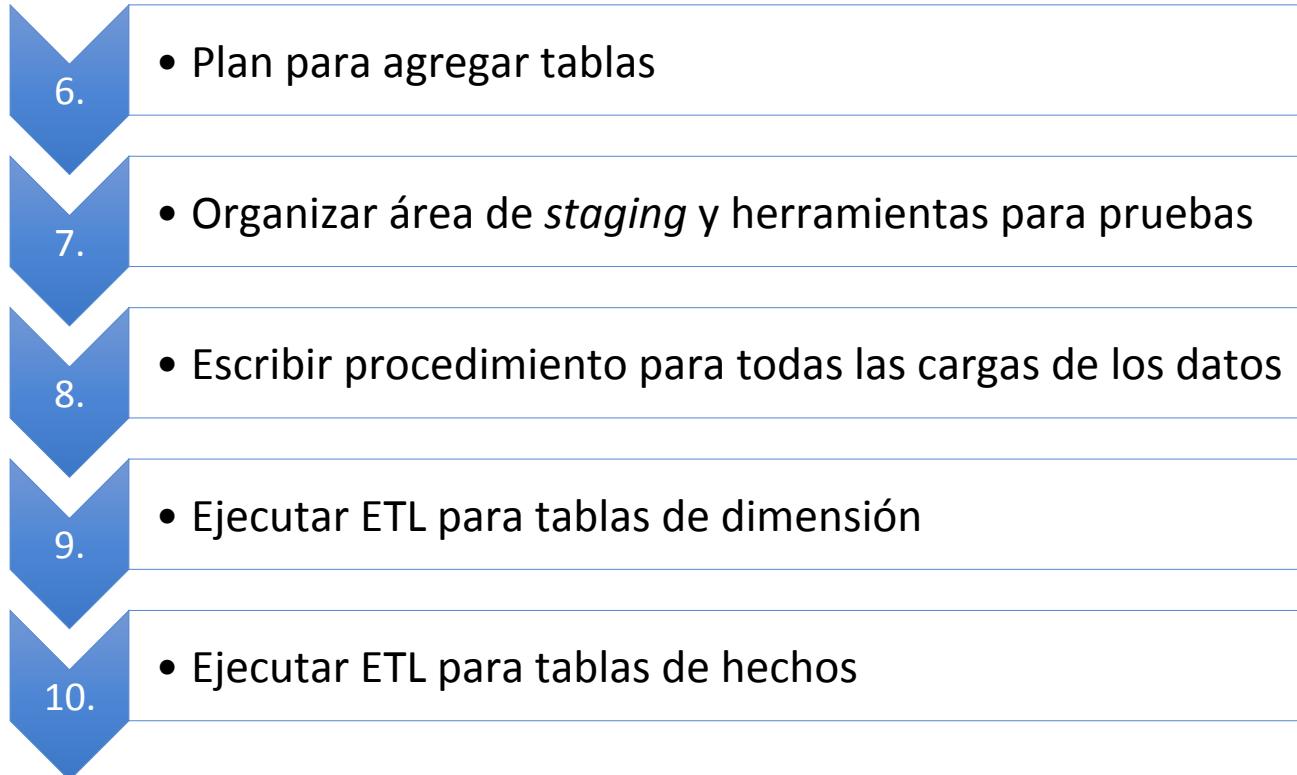
Proceso de ETL

Extract- Transform - Load

Pasos del proceso de ETL



Pasos del proceso de ETL

- 
6. • Plan para agregar tablas
 7. • Organizar área de *staging* y herramientas para pruebas
 8. • Escribir procedimiento para todas las cargas de los datos
 9. • Ejecutar ETL para tablas de dimensión
 10. • Ejecutar ETL para tablas de hechos

Fase de Extracción de los datos

Fase de Extracción de los datos

- **Identificación de las fuentes de datos**
- **Método de extracción:**
 - Para cada fuente indicar si será automatizada o manual
- **Frecuencia de extracción**
- **Ventana de tiempo**
 - Para cada fuente de datos
- **Manejo de excepciones**
 - Registros que no se pudieron extraer

Fase de Transformación de los datos

Principales tareas

Las principales tareas de esta fase son:

- Selección
- Dividir/unir
- Conversión
- Resumir
- Derivar nuevos datos



Tipos de transformaciones

Los principales son:

- Revisiones de formatos
- Decodificación de campos
 - ¿ qué significado tienen los códigos ?
- Valores calculados y campos derivados
- División o unión de campos
- Conversiones de unidades de medida,
formatos

Tipos de transformaciones

Los principales son (continuación) :

- Resumir datos
- Reestructuración de llaves
- Deduplicación de datos



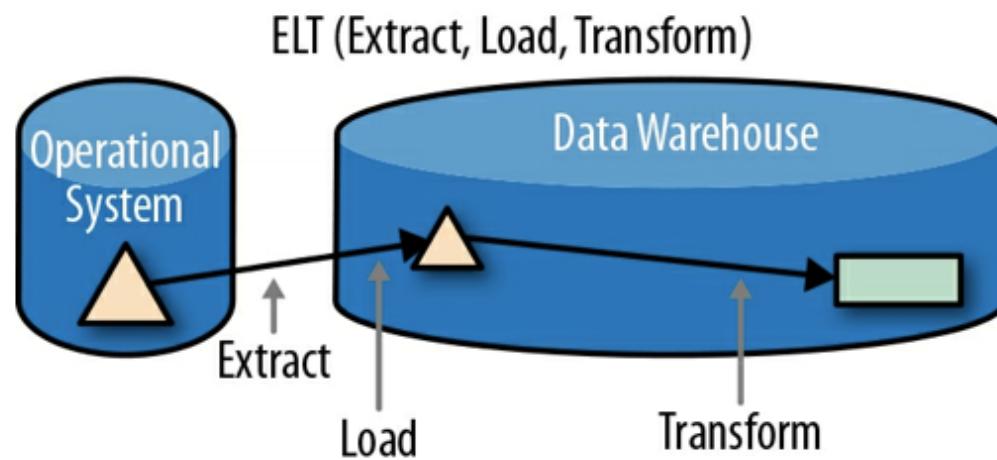
Integración y consolidación de los datos

- Combinar todos los **datos relevantes** que vienen de las fuentes operacionales
- Listos para ser cargados en el DW

Fase de Carga de datos

- Carga de datos, aplicación de datos:
- Principales métodos
 - Carga completa
 - Carga incremental
 - Streaming
 - Por lotes (Batch)

Del ETL al ELT



¿PREGUNTAS?