

# Data Warehouse Diseño

PhD María Hallo

# DISEÑO LOGICO

Sistemas OLTP

Modelo E - R

Sistemas OLAP

Esquema en Estrella

Esquema Copo de Nieve

# Diseño de un Data Mart

- Objetivos del Negocio
- Procesos del Negocio
- Requerimientos del Data Mart
- Medidas
- Dimensiones
- Jerarquías

# DISEÑO DIMENSIONAL

## PASOS A SEGUIR PARA UN DISEÑO DIMENSIONAL

- Definir un Data Mart OLAP
  - Se establece durante la identificación de requerimientos
  - Debe ser para un proceso del negocio.
  - Establecer los requerimientos históricos .
- Seleccionar Medidas
- Establecer Dimensiones
- Diseñar Agrupaciones

# Selección de Medidas - Diseño Dimensional

- Se debe definir el nivel de detalle a ser modelado.
- La granularidad de la tabla de hechos determina la configuración de sus atributos.
- El nivel de detalle más bajo corresponde al nivel de una transacción.
- Un repositorio con todas las transacciones sirve para realizar análisis a detalle pero se pueden tener modelos con indicadores agregados.

# Selección de Medidas - Diseño Dimensional

Para seleccionar medidas considerar que:

- Todas las medidas en una tabla deben corresponder a la granularidad seleccionada (no contener atributos con cantidades totales y a la vez detalles).
- Todas las medidas en una tabla deben corresponder a la misma especificación de tiempo (meses o días, etc.).

# Selección de Medidas - Diseño Dimensional

Para seleccionar medidas :

- Seleccionar el nivel más bajo posible de granularidad que soporte el sistema.

Con una granularidad baja es más fácil la introducción de nuevos elementos.

Con niveles bajos de granularidad se puede responder mejor a consultas de usuarios

# Selección de Medidas - Diseño Dimensional

## Estilos Comunes de Medidas:

- Medidas por Transacción
- Medidas por Snapshot
- Medidas por Item
- Medidas por Evento



# Selección de Medidas - Diseño Dimensional

## Estilos Comunes de Medidas:

### Medidas por Transacción:

- Se basan solo en una instancia de un evento del negocio
  - Representan el nivel más bajo de granularidad posible
  - Normalmente contienen solo una medida: la cantidad de la transacción
  - Se basan en tipos de transacción clave
- Ejemplo: Una llamada por teléfono

# Selección de Medidas - Diseño Dimensional

## Estilos Comunes de Medidas:

### Medidas por Snapshot:

- Capturan el estado del negocio en un tiempo en particular, al final de una venta, del día, del mes, etc.
- Contienen sumatorios o valores instantáneos  
Ejemplo: Cuentas de ventas por producto mensuales

# Selección de Medidas - Diseño Dimensional

## Estilos Comunes de Medidas:

### Medidas por Item

- Útiles para detallar información sobre los elementos de un documentos del negocio.
- Representan ítems individuales relacionados con los eventos del negocio, contienen todas las medidas sobre el ítem (precio de venta, cantidad, etc)

Ejemplo: Ordenes de compra

# Selección de Medidas - Diseño Dimensional

## Estilos Comunes de Medidas:

### Medidas por Evento

- Representan un acontecimiento del evento pero no los detalles
- Útiles cuando el análisis incluye unas actividades y otras no para un conjunto específico de dimensiones,

Ejemplo: Productos que están en promoción y no han sido vendidos hoy

# Características de las Tablas de Hechos - Diseño Dimensional

- Toda tabla de hechos debe tener una clave que identifique a cada registro de medida. La clave está compuesta de una o más claves foráneas que vienen de las tablas de dimensiones.
- Las medidas están representadas por columnas numéricas.
- No es necesario que exista datos para todas las claves foráneas.
- El número de filas de la tabla de hechos depende de la granularidad establecida.

# Dimensiones

Las Dimensiones determinan la forma como los datos deberían estar organizados y agrupados

# Características de las Dimensiones

## Atributos con Información de Alta Calidad

- Nombres Descriptivos
- No usar abreviaciones en los nombres.
- Datos útiles, no nulos o inválidos

## Claves de Identificación

- Independencia de claves del sistema origen
- Usar claves simples, no compuestas
- Resulta eficiente usar secuencias numéricas

# Características de las Dimensiones

- Cardinalidad de uno a muchos con la Tabla de Hechos
- Contiene al menos un atributo descriptivo
- Contiene atributos útiles para la Agrupación
- Contiene un número limitado de registros que crecen lentamente en el tiempo



# Agregación - Diseño Dimensional

Determinan el rendimiento del DW

La Agregación es una estructura - tabla - que contiene datos pre-calculados.

Las Agregaciones pueden ser creadas dinámicamente o durante la carga de datos.

Pueden ser almacenadas para usos repetitivos o creadas dinámicamente para cada consulta.

Los requerimientos del almacenamiento dependen del número de dimensiones, medidas y del número de niveles de jerarquías en las dimensiones.

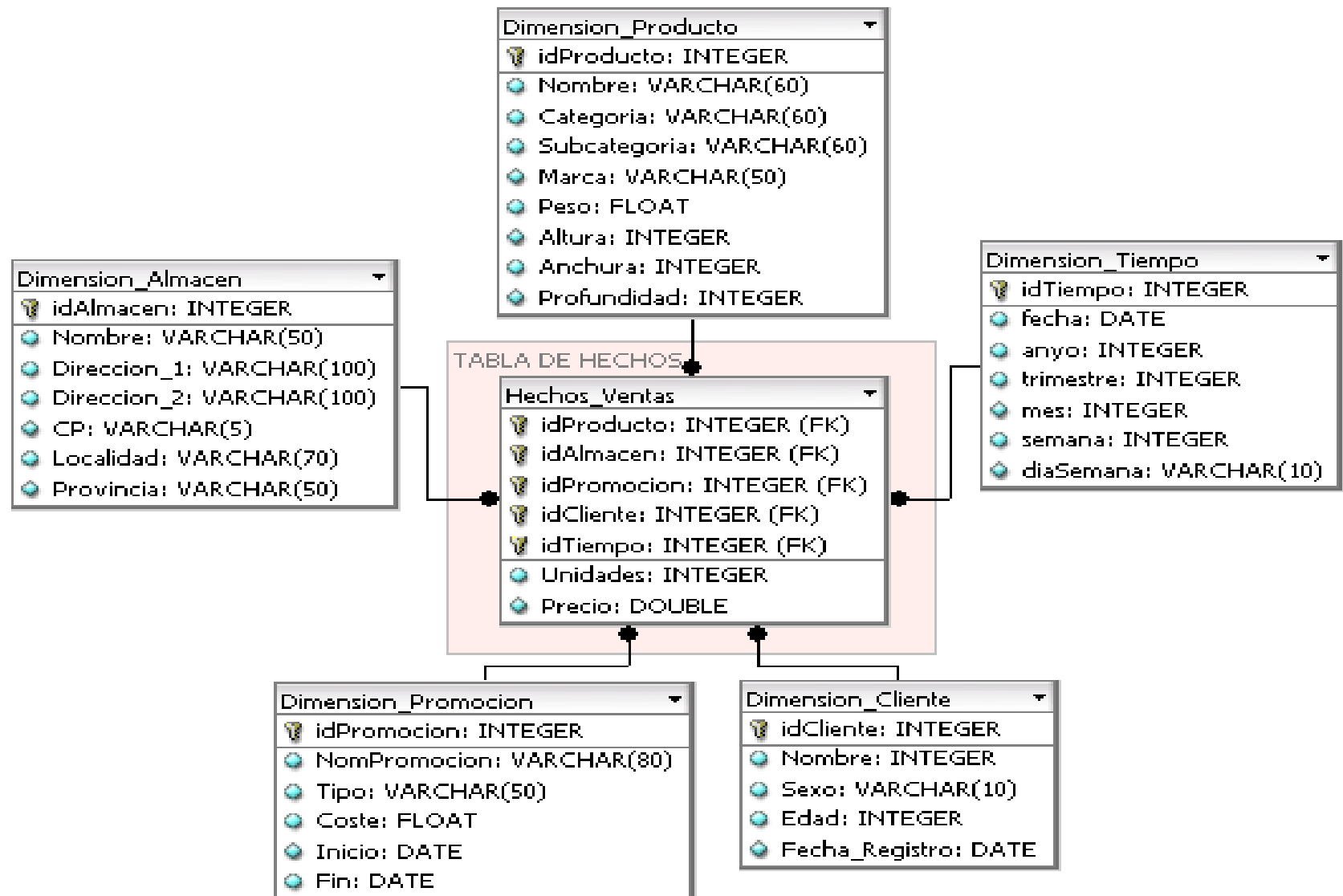
# Agregación - Diseño Dimensional

Cada nivel de Agrupación es almacenado en una Tabla de Hechos Distinta  
Estas tablas son inicialmente generadas desde la tabla de hechos base durante un proceso de separación y son actualizadas desde ésta.

# Estrella vs. Copo de Nieve

Criterio	Estrella	Copo de Nieve
Espacio requerido	Alto	Bajo
Comprensión del Modelo	Fácil	Mayor complejidad
Número de Tablas	Menor	Mayor
Complejidad de Consultas	Simple	Mayor complejidad
Búsqueda dimensional	Rápida	Lenta
Indexación bitmap	Soporta	No reconoce

# Ejemplo de un modelo dimensional



# Esquema de copo de nieve

