

# MODELADO DIMENSIONAL

Jorge Iván Triviño Arbeláez  
Bodegas de Datos

# Modelo Entidad Relación

- El modelo entidad relación (ER) es una técnica poderosa para diseñar lógicamente sistemas para el procesamiento de transacciones OLTP (procesamiento transaccional en línea). Siempre va encaminado a la eliminación de la redundancia, lo que permite que la manipulación sobre la base de datos tenga que hacerse en un solo lugar y sea mucho más rápido ya que en el momento en que la transacción requiera insertar, adicionar, borrar o modificar un dato es necesario que lo haga en un solo lugar y no en múltiples.

# Modelo Entidad-Relación(2)



- Esto contribuye también al almacenamiento de grandes cantidades de datos dentro de las bases de datos relacionales, a través del proceso de normalización. Por eso es perfecto para la inserción y actualización de la información.
- Este es un modelo excelente para registrar transacciones y administración de tareas operativas. Sin embargo, para el modelamiento de una bodega de datos presenta varios problemas. Los usuarios finales no entienden ni recuerdan un diagrama entidad relación.

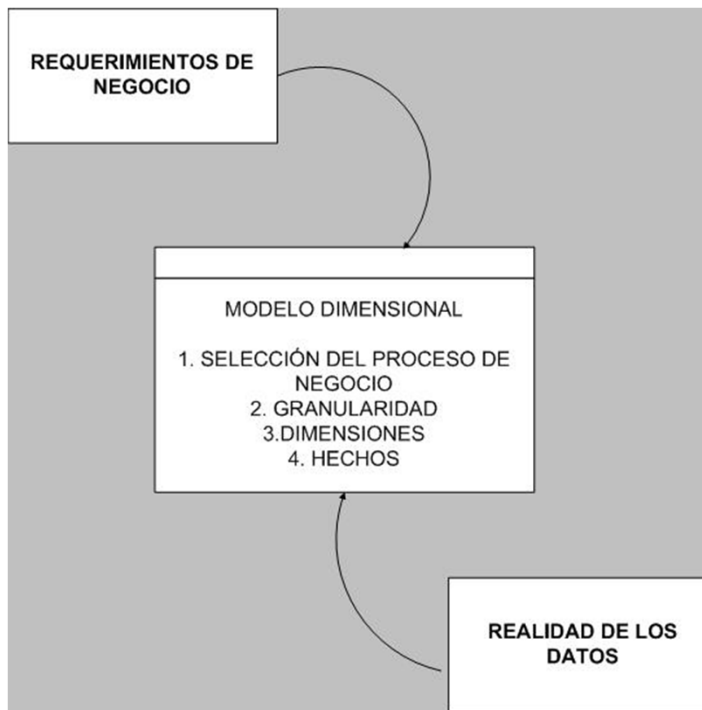
# Modelo Entidad-Relación (3)



- Nos es posible que los usuarios finales naveguen sobre el modelo. El uso del modelo entidad relación va en contra del objetivo principal de una bodega de datos, de proporcionar datos de forma intuitiva y con un buen desempeño y tiempos de respuesta.

# Modelo Dimensional

- El modelo dimensional es una técnica de diseño lógico que busca presentar los datos de una forma intuitiva y que proporcione acceso de alto desempeño.



Según Kimball

# Modelado Dimensional



- Cada modelo dimensional se compone de una tabla con múltiples llaves foráneas, llamada tabla de hechos (fact table), y un conjunto de tablas más pequeñas, llamadas tablas de dimensión.
- Los atributos de las tablas de dimensión son las fuentes de las restricciones de búsqueda necesarias para consultar una bodega de datos. Son utilizadas como título de atributo de las filas resultantes de queries de SQL

# Tabla de Dimensión

- las tablas de dimensiones son elementos que contienen atributos (o campos) que se utilizan para restringir y agrupar los datos almacenados en una tabla de hechos cuando se realizan consultas sobre dicho datos en un entorno de almacén de datos o data mart.
- Estos datos sobre dimensiones son parámetros de los que dependen otros datos que serán objeto de estudio y análisis y que están contenidos en la tabla de hechos. Las tablas de dimensiones ayudan a realizar ese estudio/análisis aportando información sobre los datos de la tabla de hechos

# Condiciones de las dimensiones



- 1. Una tabla de dimensión puede ser usada con cualquier tabla de hechos de la misma base de datos.
- 2. Las interfaces de usuario y contenido de datos son consistentes para cualquier uso de la dimensión.
- 3. Hay una interpretación consistente de atributos, por lo tanto se obtiene la misma interpretación de la tabla en cualquier datamart.



# Tabla de Hecho



- una tabla de hechos (o tabla fact) es la tabla central de un esquema dimensional (en estrella o en copo de nieve) y contiene los valores de las medidas de negocio. Cada medida se toma mediante la intersección de las dimensiones que la definen, dichas dimensiones estarán reflejadas en sus correspondientes tablas de dimensiones que rodearán la tabla de hechos y estarán relacionadas con ella.

# Tabla de Hecho(2)



- ▣ La tabla de hechos dentro de un esquema de estrella, es la que contiene los movimientos, eventos o hechos asociados a una o más actividades de un proceso de negocio
- ▣ Las tablas de hechos contienen un conjunto de claves dimensionales que a menudo conforman una clave única compuesta para identificar un registro. Estas claves dimensionales, son claves foráneas que se relacionan con las tablas dimensionales

# MODELADO MULTIDIMENSIONAL

## □ Modelo dimensional

- Propuesto por Ralph Kimball, es un mecanismo metodológico para construir una bodega de datos con todas las características mencionadas

### ▣ Características

- Usa esquemas de estrella
  - Tablas de hechos
  - Tablas dimensionales
- Alto rendimiento en las consultas
- Registra información como eventos que ocurren a través del tiempo
  - Orientado a recibir siempre inserciones de nuevos registros
- Originalmente diseñados para almacenar información resumida

# MODELADO MULTIDIMENSIONAL

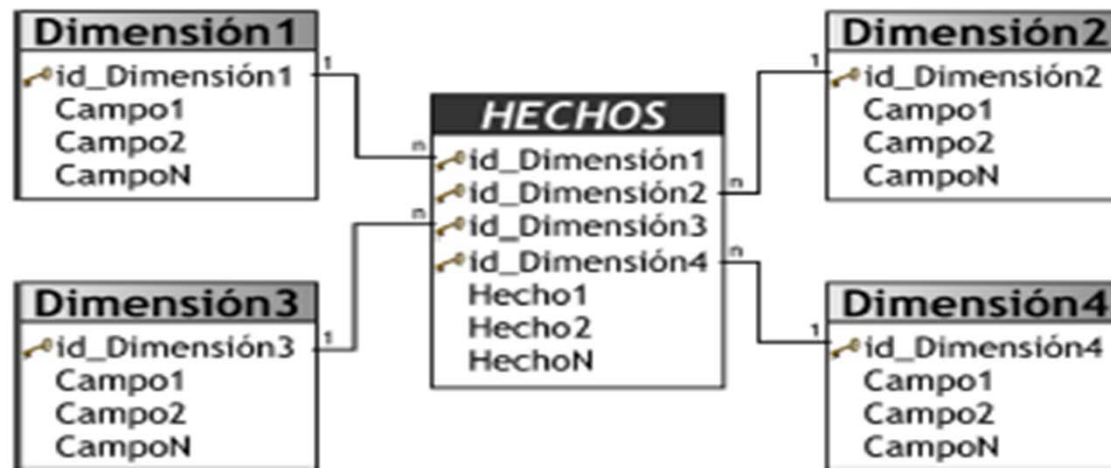
## □ Datamarts

- Subconjunto de una bodega de datos
  - Puede haber varios criterios para dividir una bodega en datamarts: Geografía, Tiempo, Producto, Personas, etc.
- Orientado a temas
- Tienen los mismos principios de diseño de las bodegas de datos
- Podría decirse que:
  - Varios Datamart pueden llegar a formar una bodega de datos (táctica para construir una bodega)
  - De una bodega se pueden obtener varios Datamarts (Distribución de información)

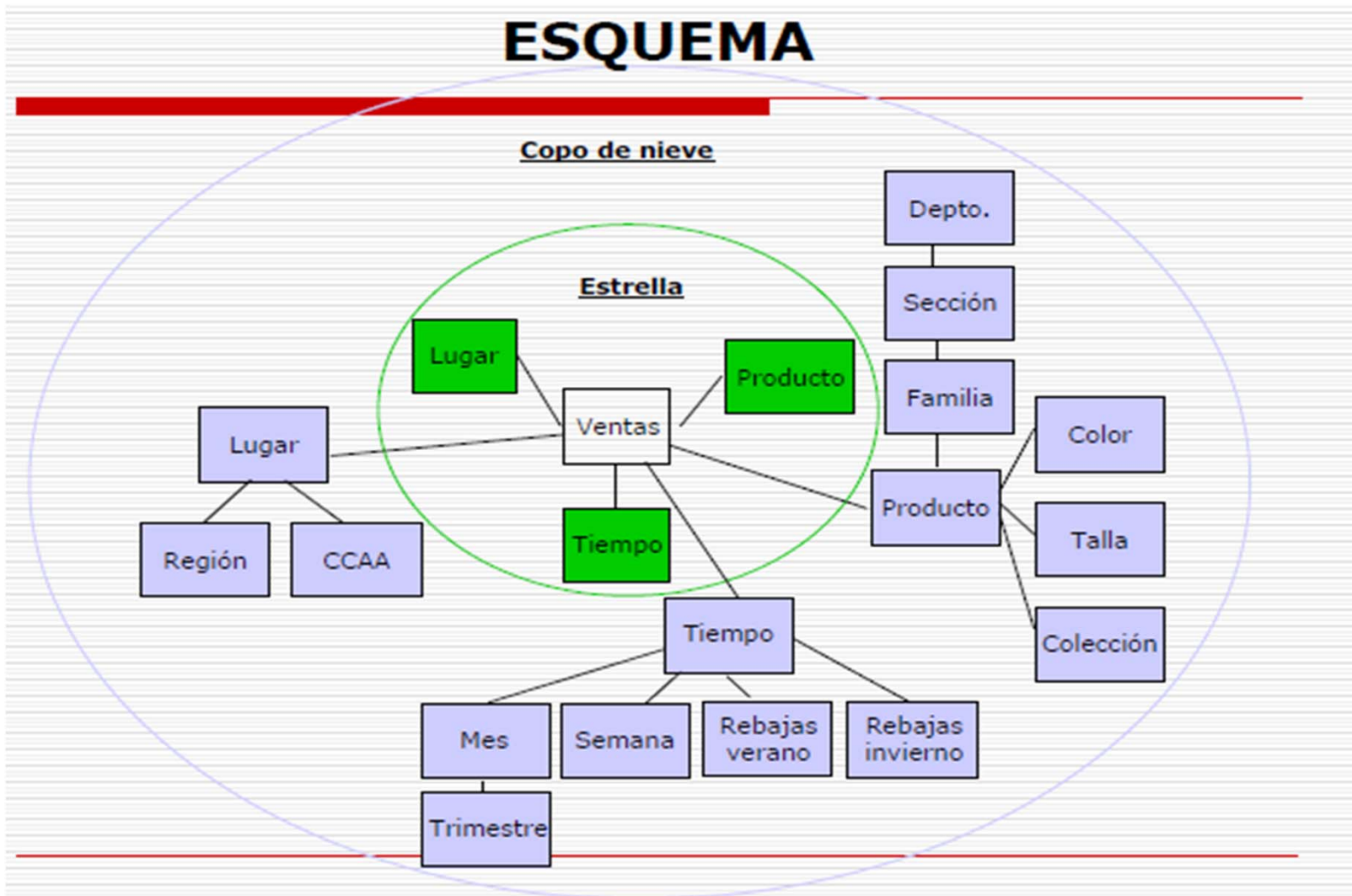
# MODELADO MULTIDIMENSIONAL

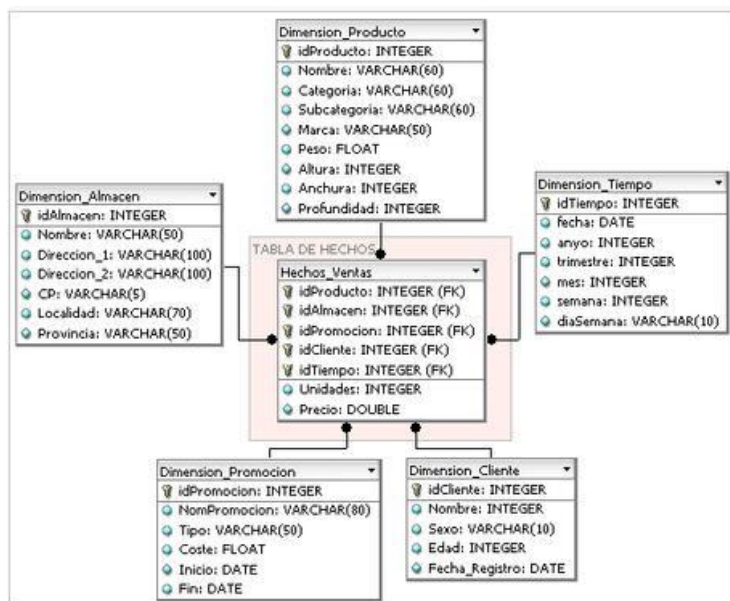
## □ Esquemas de estrella

- Contiene una tabla central llamada Tabla de Hechos (factResellerSales) que registra los eventos del proceso
- La rodean un conjunto de tablas dimensionales que relacionan con la tabla de hechos a través de una clave foránea
- Su nombre se debe a su parecido con una estrella donde el centro es la tabla de hechos y las puntas son las tablas dimensionales

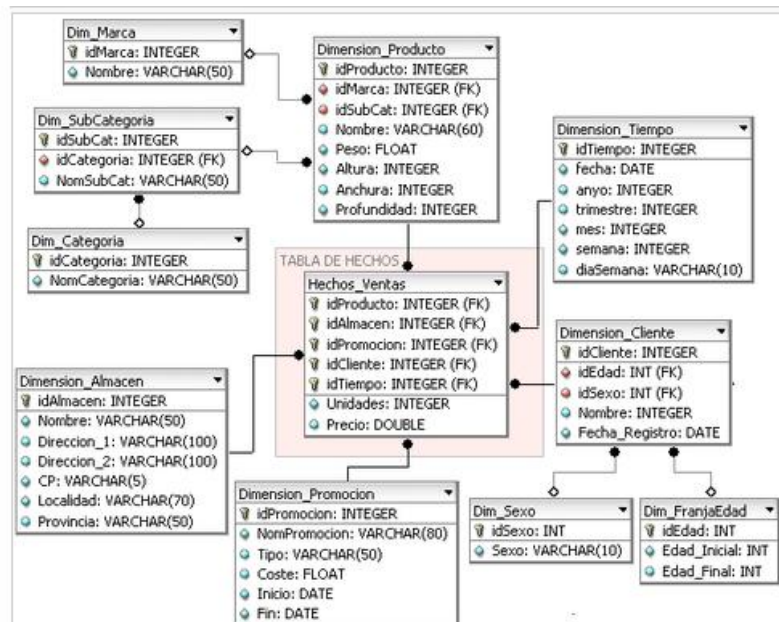


# Modelado copo de nieve





Ejemplo de modelo de datos en estrella.



Ejemplo de modelo de datos en copo de nieve.

# MODELADO MULTIDIMENSIONAL

## □ Modelamiento de tablas dimensionales

- Sus columnas son altamente correlacionadas y descriptivas
- Son des normalizadas
- Las tablas dimensionales pueden ser “anchas”
- Deben tener una clave subrogada (Clave Primaria)
- Deben tener al menos una clave alterna (es la clave primaria en los sistemas de origen)
- Deben evitarse columnas con abreviaciones
- Deben evitarse las columnas con valores nulos
- Cree columnas útiles para establecer niveles de agregación
- Minimice el número de columnas que cambien con el tiempo
- Clave sustituta de la dimensión de tiempo
  - Aaaammdd - Por ejemplo: 20090325





















# Errores Comunes



















- ❑ Colocar en la tabla de hechos atributos para filtrar / agrupar
- ❑ Reducir atributos para ahorrar espacio
- ❑ Dividir jerarquías y niveles en varias dimensiones
- ❑ Ignorar la necesidad de los cambios en las dimensiones
- ❑ Ante un problema de rendimiento añadir más hardware en lugar de
- ❑ revisar diseño de cubos, índices, o crear nuevos
- ❑ Utilizar claves operacionales para el *join* con la tabla de hechos
- ❑ Negarse a comprender la granularidad de la tabla de hechos
- ❑ Crear el modelo dimensional para un informe concreto
- ❑ Esperar que los usuarios consulten los datos en la *staging area*
- ❑ Fallar al conformar las dimensiones



















# MODELADO MULTIDIMENSIONAL

## □ Modelamiento de tablas dimensionales

Ej

DimCustomer	
	CustomerKey
	GeographyKey
	CustomerAlternateKey
	Title
	FirstName
	MiddleName
	LastName
	NameStyle
	BirthDate
	MaritalStatus
	Suffix
	Gender
	EmailAddress
	YearlyIncome
	TotalChildren
	NumberChildrenAtHome
	EnglishEducation
	SpanishEducation

DimProduct	
	ProductKey
	ProductAlternateKey
	ProductSubcategoryKey
	WeightUnitMeasureCode
	SizeUnitMeasureCode
	EnglishProductName
	SpanishProductName
	FrenchProductName
	StandardCost
	FinishedGoodsFlag
	Color
	SafetyStockLevel
	ReorderPoint
	ListPrice
	Size
	SizeRange
	Weight
	DaysToManufacture

DimEmployee	
	EmployeeKey
	ParentEmployeeKey
	EmployeeNationalIDAlter...
	ParentEmployeeNationall...
	SalesTerritoryKey
	FirstName
	LastName
	MiddleName
	NameStyle
	Title
	HireDate
	BirthDate
	LoginID
	EmailAddress
	Phone
	MaritalStatus
	EmergencyContactName
	EmergencyContactPhone

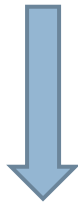
# MODELADO MULTIDIMENSIONAL

- Modelamiento de tablas dimensionales
  - ▣ **Dimensiones padre-hijo:** Tienen relaciones reflexivas
  - ▣ **Dimensión de tiempo:** Permiten acomodar diferentes períodos de tiempo incluyendo períodos no naturales y calendarios especiales
  - ▣ **Dimensiones degeneradas:** Son dimensiones que se crean a partir de la tabla de hechos
  - ▣ **Dimensiones basura:** Son dimensiones que agrupan atributos que no tienen correlación entre ellos. Usadas para disminuir el espacio en las tablas de hechos mediante el uso de una clave subrogada.
  - ▣ **Dimensiones lentamente cambiantes:** Son dimensiones que tienen atributos que cambian lentamente con el tiempo y para las cuales se debe guardar trazabilidad de los cambios en dichos atributos.

# MODELADO MULTIDIMENSIONAL

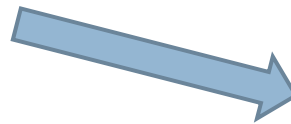
- Dimensiones lentamente cambiantes
  - ▣ Supóngase que después de un tiempo el cliente 123 se cambia a la ciudad de Bogotá. ¿Cómo debería entonces quedar registrado el cliente?

IdCliente	CodCliente	Cliente	Ciudad
1	123	Compañía X	Medellín



Solución de tipo II

IdCliente	CodCliente	Cliente	Ciudad	FechaInicioVigencia	FechaFinVigencia
1	123	Compañía X	Medellín	2000-01-01	2002-10-15
50	123	Compañía X	Bogotá	2002-10-15	



Solución de tipo I

IdCliente	CodCliente	Cliente	Ciudad
1	123	Compañía X	Bogotá

# MODELADO MULTIDIMENSIONAL

## □ Modelamiento de tablas de hechos

- ▣ La tabla de hechos dentro de un esquema de estrella, es la que contiene los movimientos, eventos o hechos asociados a una o más actividades de un proceso de negocio
- ▣ Las tablas de hechos contienen un conjunto de claves dimensionales que a menudo conforman una clave única compuesta para identificar un registro. Estas claves dimensionales, son claves foráneas que se relacionan con las claves subrogadas de las tablas dimensionales

FactResellerSales	
	ProductKey
	OrderDateKey
	DueDateKey
	ShipDateKey
	ResellerKey
	EmployeeKey
	PromotionKey
	CurrencyKey
	SalesTerritoryKey
	SalesOrderNumber
	SalesOrderLineNumber
	RevisionNumber
	OrderQuantity
	UnitPrice
	ExtendedAmount
	UnitPriceDiscountPct
	DiscountAmount
	ProductStandardCost
	TotalProductCost
	SalesAmount
	TaxAmt
	Freight
	CarrierTrackingNumber
	CustomerPONumber

# MODELADO MULTIDIMENSIONAL

## □ Modelamiento de tablas de hechos

- Es importante anotar que la combinación de las claves dimensionales determina la **granularidad** de la tabla de hechos. Es decir, a qué nivel de detalle máximo se puede llegar en los datos que quedan registrados en la tabla de hechos.
- Si se adicionara una nueva clave dimensional a una tabla de hechos, esto implicaría que la tabla se vuelve más granular o más detallada

FactResellerSales	
	ProductKey
	OrderDateKey
	DueDateKey
	ShipDateKey
	ResellerKey
	EmployeeKey
	PromotionKey
	CurrencyKey
	SalesTerritoryKey
	SalesOrderNumber
	SalesOrderLineNumber
	RevisionNumber
	OrderQuantity
	UnitPrice
	ExtendedAmount
	UnitPriceDiscountPct
	DiscountAmount
	ProductStandardCost
	TotalProductCost
	SalesAmount
	TaxAmt
	Freight
	CarrierTrackingNumber
	CustomerPONumber

# MODELADO MULTIDIMENSIONAL

## □ Modelamiento de tablas de hechos: Medidas

- Una tabla de hechos puede contener uno o más atributos diferentes de las claves dimensionales. Estos atributos se conocen como medidas, son los valores de datos que se analizan (son numéricos). Las medidas pueden ser de varios tipos:

- **Aditivas:** son las más comunes. Estas son atributos numéricos que tiene sentido sumarlos con el fin de obtener totales. Normalmente se usan con la función de agregación SUMA Y CONTADOR.
- **Semiaditivas:** Solo se pueden sumar a través de algunas dimensiones. Típicamente no se pueden sumar a través de la dimensión de tiempo. Esto sucede por ejemplo en casos como los saldos de una cuenta bancaria, los cuales no se pueden sumar a través del tiempo. De forma similar ocurre con los saldos de inventario o con el nivel de un tanque.
- **No aditivas:** sencillamente son medidas que no tiene sentido sumar y a menudo son simples atributos complementarios del hecho que se está registrando. Pueden ser atributos como números de factura, fechas de control, u algún otro tipo de atributo que pueda ser usado para trazabilidad para poder encontrar información complementaria en los sistemas transaccionales.

- Deben ser “delgadas”. Pocos campos

- Se recomienda tener índices por cada clave foránea.

## Dimensiones

FactResellerSales	
ProductKey	
OrderDateKey	
DueDateKey	
ShipDateKey	
ResellerKey	
EmployeeKey	
PromotionKey	
CurrencyKey	
SalesTerritoryKey	
SalesOrderNumber	
SalesOrderLineNumber	
RevisionNumber	
OrderQuantity	
UnitPrice	
ExtendedAmount	
UnitPriceDiscountPct	
DiscountAmount	
ProductStandardCost	
TotalProductCost	
SalesAmount	
TaxAmt	
Freight	
CarrierTrackingNumber	
CustomerPONumber	

## Medidas o hechos

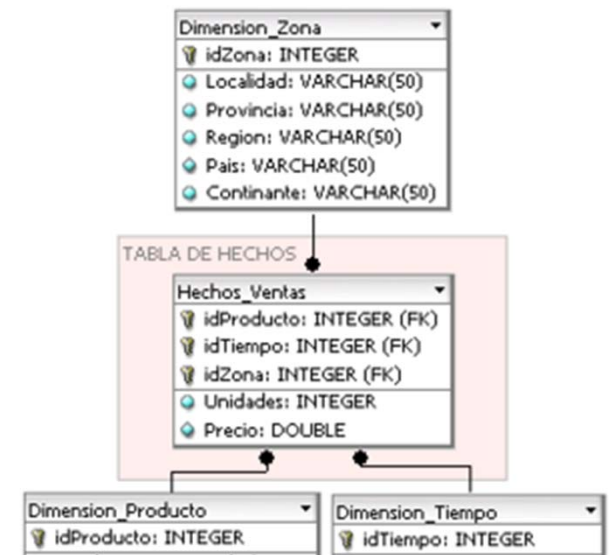
# MODELADO MULTIDIMENSIONAL

## Granularidad:

Es el nivel de detalle en que se almacena la información.

### □ Por ejemplo:

- Datos de ventas o compras de una empresa, pueden registrarse día a día
- Datos pertinentes a pagos de sueldos o cuotas de socios, podrán almacenarse a nivel de mes.



- A mayor nivel de detalle, mayor posibilidad analítica, ya que los mismos podrán ser resumidos o sumariados.
- Los datos con granularidad fina (nivel de detalle) podrán ser resumidos hasta obtener una granularidad media o gruesa. No sucede lo mismo en sentido contrario.



# MODELADO MULTIDIMENSIONAL

## □ Estrategia de modelado y arquitectura

### ▣ Otras consideraciones

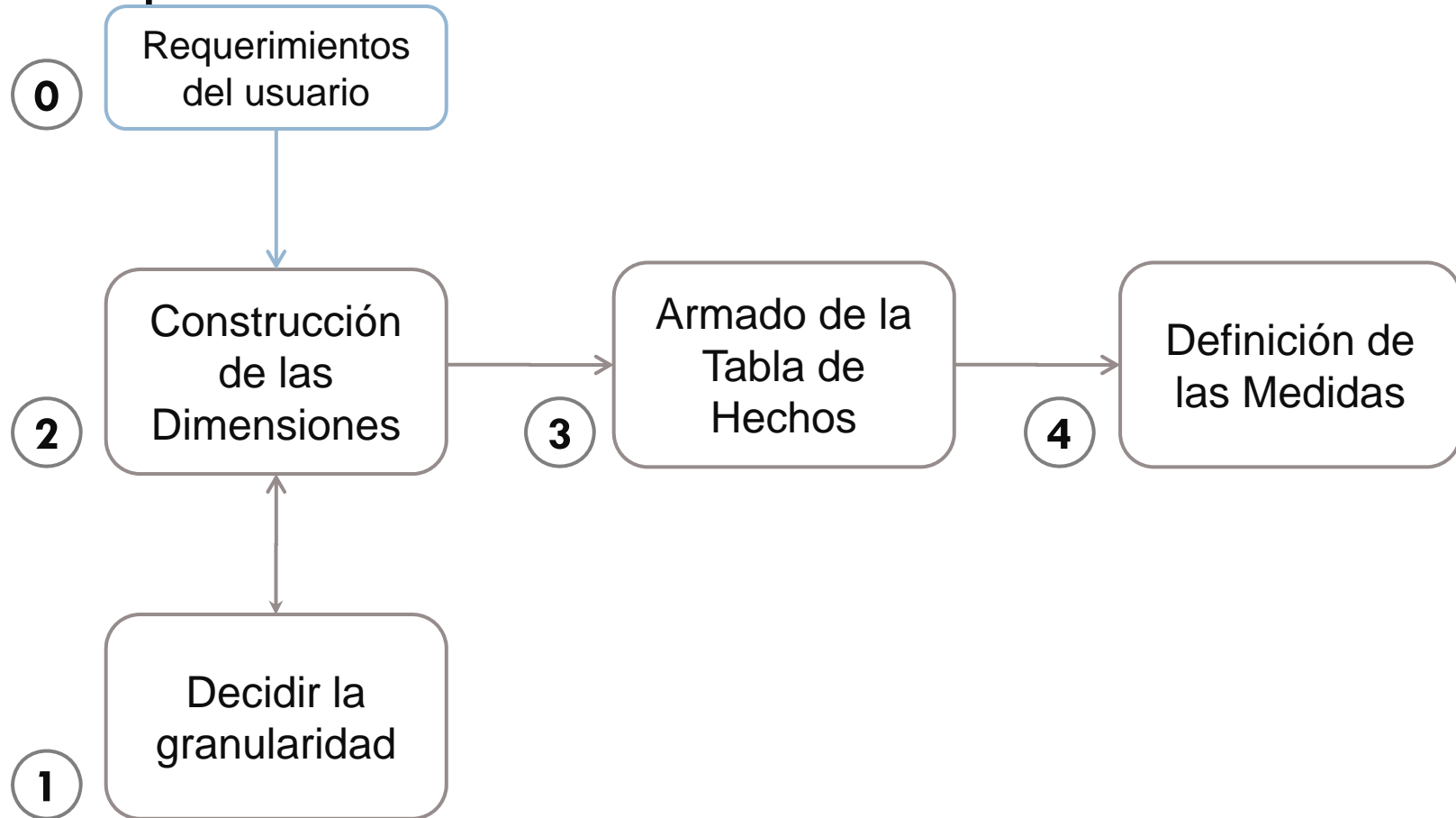
- Una de las mejores estrategias para la construcción de una bodega de datos consiste en la construcción sucesiva de datamarts integrándolos mediante dimensiones conformes
  - Cada datamart no debe ser independiente
  - En cada iteración se debe pensar en como realizar la integración a nivel corporativo
  - Piense en la bodega de datos y construya un datamart
- Las dimensiones conformes unifican conceptos a través de toda la organización y dichos conceptos se estandarizan como parte del proceso de ETL

# EJERCICIO



# Ejercicio

## □ Etapas en la construcción de un modelo dimensional:



# Requerimientos del usuario

0

	Dimensiones				
Medidas	Tiempo	Sucursal	Vendedor	Cliente	Producto
Ventas_Importe	X	X	X	X	X
Ventas_Costo	X	X	X	X	X
Ventas_Unidades	X	X	X	X	X
Ventas_ImporteTotal	X	X	X	X	X
Ventas_Ganancia	X	X	X	X	X
Ventas_Promedio	X	X	X	X	X

# Decidir la granularidad

1

- La granularidad:
  - ▣ Es el nivel de detalle al que se desea almacenar información sobre la actividad a modelar.
  - ▣ Define el nivel atómico de datos en el almacén de datos.
  - ▣ Determina el significado de las tuplas de la tabla de hechos.
  - ▣ Determina las dimensiones básicas del esquema.
- Por ejemplo en la dimensión Sucursal:

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

# Decidir la granularidad

- Ejemplo de la dimensión fecha. Se desea los datos por:

- Información anual
- Información semestral
- Información trimestral
- Información mensual. ....
- Información semanal
- Información diaria

+ granularidad  
+ detalle

Dimensión Tiempo	Dimensión Tiempo	Dimensión Tiempo	Dimensión Tiempo	Dimensión Tiempo
* Año	* Año ** Semestre	* Año ** Semestre *** Trimestre	* Año ** Semestre *** Trimestre **** Mes	* Año ** Semestre *** Trimestre **** Mes ***** Día

# Construcción de las dimensiones

- Identificar las dimensiones que caracterizan el proceso al nivel de detalle (gránulo) que se ha elegido.
- De cada dimensión se debe decidir los atributos (propiedades) relevantes para el análisis de la actividad.
- Entre los atributos de una dimensión existen jerarquías naturales que deben ser identificadas (día-mes-año)
  - ▣ Tiempo. Cuándo se produce la actividad
  - ▣ Sucursal. Donde está ubicado el almacén
  - ▣ Vendedor. Quién ha vendido
  - ▣ Cliente. Quién es el destinatario de la actividad
  - ▣ Producto. Cuál es el objeto de la actividad

