

# CI2355 – Almacenes de datos y OLAP

---



UNIVERSIDAD DE  
**COSTA RICA**

# Examen

---

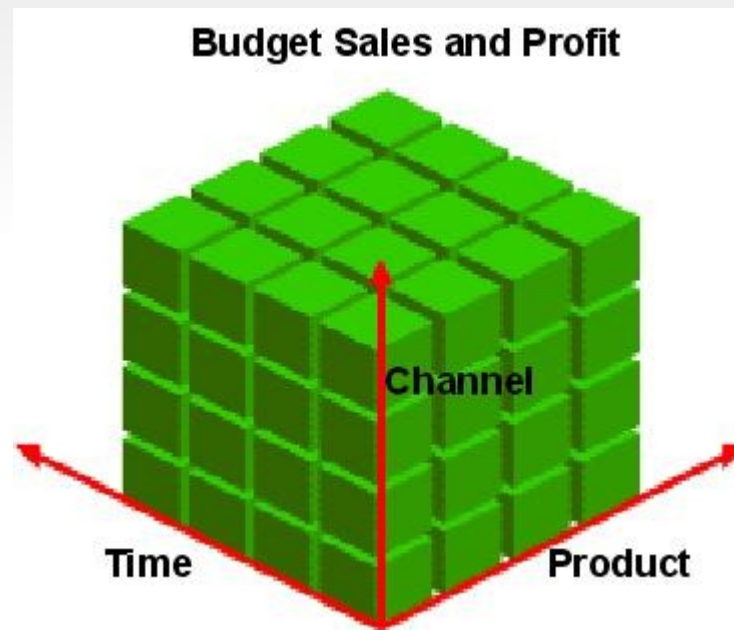
- Lunes, 7 de mayo de 2012

# Trabajo de Investigación

---

- Presentación
  - 40-45 minutos
  - Fechas:
    - Entre el 14 de mayo y el 11 de junio de 2012, inclusive.
- Artículo
  - 4 de junio de 2012

# Modelo multidimensional

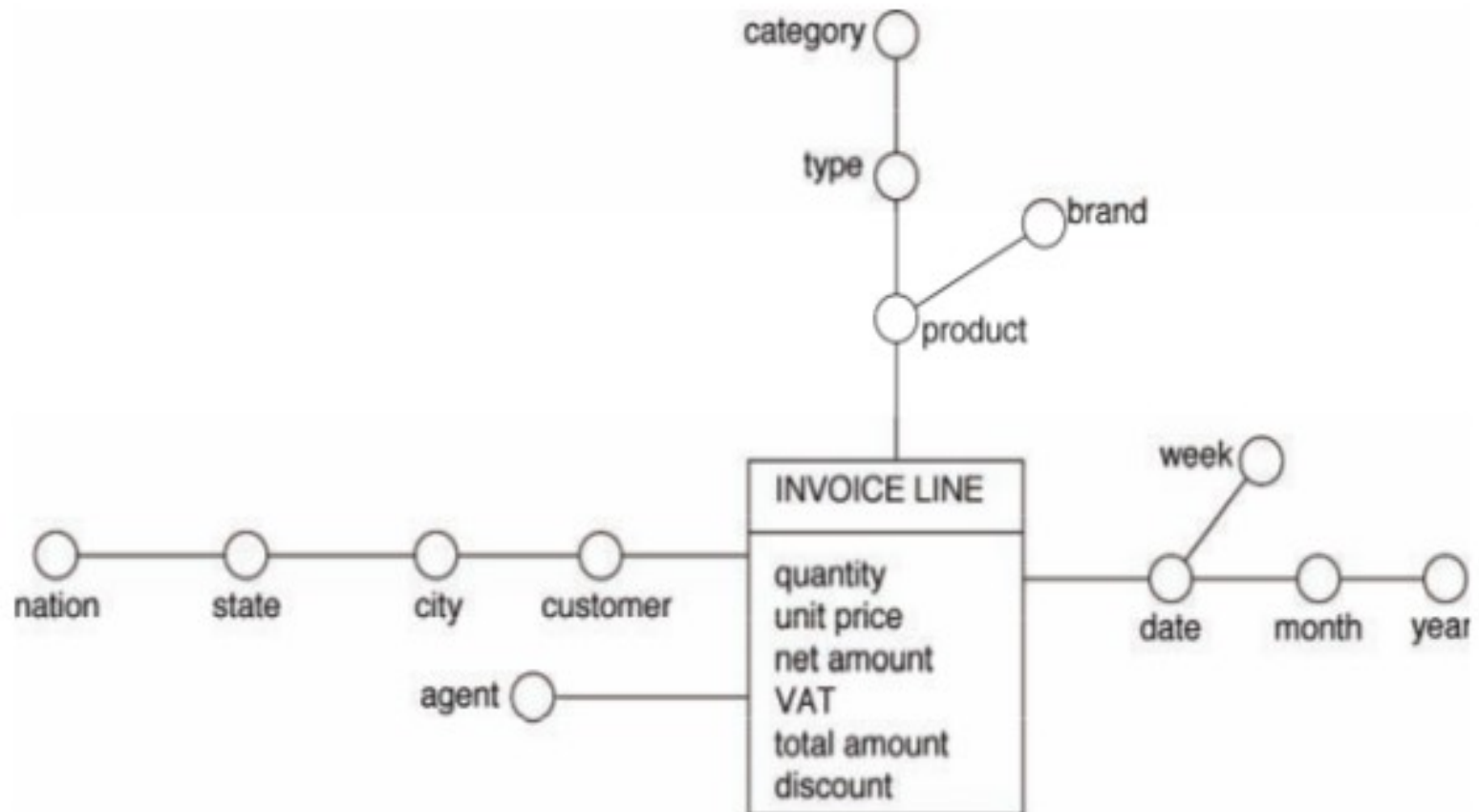


# Modelado básico

---

- Conceptos básicos que modelaremos:
  - Hechos
  - Medidas
  - Dimensiones
  - Atributos
  - Jerarquías

# Línea de factura



# Conceptos básicos

---

- Un hecho es un foco de interés para el proceso de toma de decisiones; típicamente, modela un conjunto de eventos que ocurren en el mundo de los negocios.
  - **Ejemplos:**
    - Ventas, envíos, compras, reclamos (comercio)
    - Transacciones de bolsa, contratos de pólizas de seguros, préstamos, estados de cuenta, compras con tarjeta de crédito (finanzas)

# Conceptos básicos

---

- Los conceptos que en las fuentes de datos son actualizados frecuentemente son buenos candidatos para ser modelados como hechos; los que representan registros casi estáticos, no lo son.
- Aunque, en realidad, pocas cosas son completamente estáticas, considere los siguientes casos:
  - Cambio de gerente en el departamento de ventas
  - Relación entre productos y promociones



# Conceptos básicos

---

- Una medida es una propiedad numérica de un hecho que describe uno de sus aspectos cuantitativos de interés para análisis.
  - Un hecho podría no tener medidas, si la única cosa interesante a ser registrada es la ocurrencia de eventos.
- Una dimensión es una propiedad de un hecho con un dominio finito y describe una de sus coordenadas de análisis.
  - El conjunto de dimensiones de un hecho determina su granularidad más fina de representación.

# Conceptos básicos

---

- Al menos una de las dimensiones de un hecho debe representar tiempo, en algún grado de granularidad.
- La relación entre medidas y dimensiones es expresada, a nivel de instanciación, por el concepto de evento.
- Un evento primario es una ocurrencia de un hecho y es identificado por una tupla de valores, uno para cada dimensión.
  - Cada evento primario es descrito por un valor para cada medida.
  - Representan las celdas del cubo.

# Conceptos básicos

---

- Si la granularidad de eventos primarios determinados por un conjunto de dimensiones es más gruesa que la granularidad de las tuplas en la fuente de datos, las medidas deberán ser definidas como agregados de atributos numéricos o como sumas de tuplas.
- Agregación es la operación básica de OLAP.
- Un atributo dimensional es una propiedad, con un dominio finito, de una dimensión.
  - **Ejemplo: un producto es representado por su tipo, categoría y marca.**
- Las relaciones entre atributos se expresan como jerarquías.

# Conceptos básicos

---

- Una jerarquía es un árbol dirigido, con raíz en una dimensión, cuyos nodos son todos los atributos de dimensión que describen la dimensión, y cuyas aristas modelan asociaciones *muchos-a-uno* entre pares de atributos de dimensiones.
  - Las jerarquías deben reproducir los patrones de dependencia funcional entre atributos expresados en la fuente de datos.
  - Las jerarquías determinan cómo eventos primarios pueden ser agregados como eventos secundarios.

# Conceptos básicos

---

- La dimensión en la que una jerarquía tiene su raíz define su granularidad más fina de agregación.
- Dado un conjunto de atributos de dimensiones, cada tupla de sus valores identifica un evento secundario que agrega todos los eventos primarios correspondientes.
  - Cada evento secundario es descrito por un valor para cada medida que resume los valores tomados por la misma medida en los eventos primarios correspondientes.

# Conceptos avanzados

---

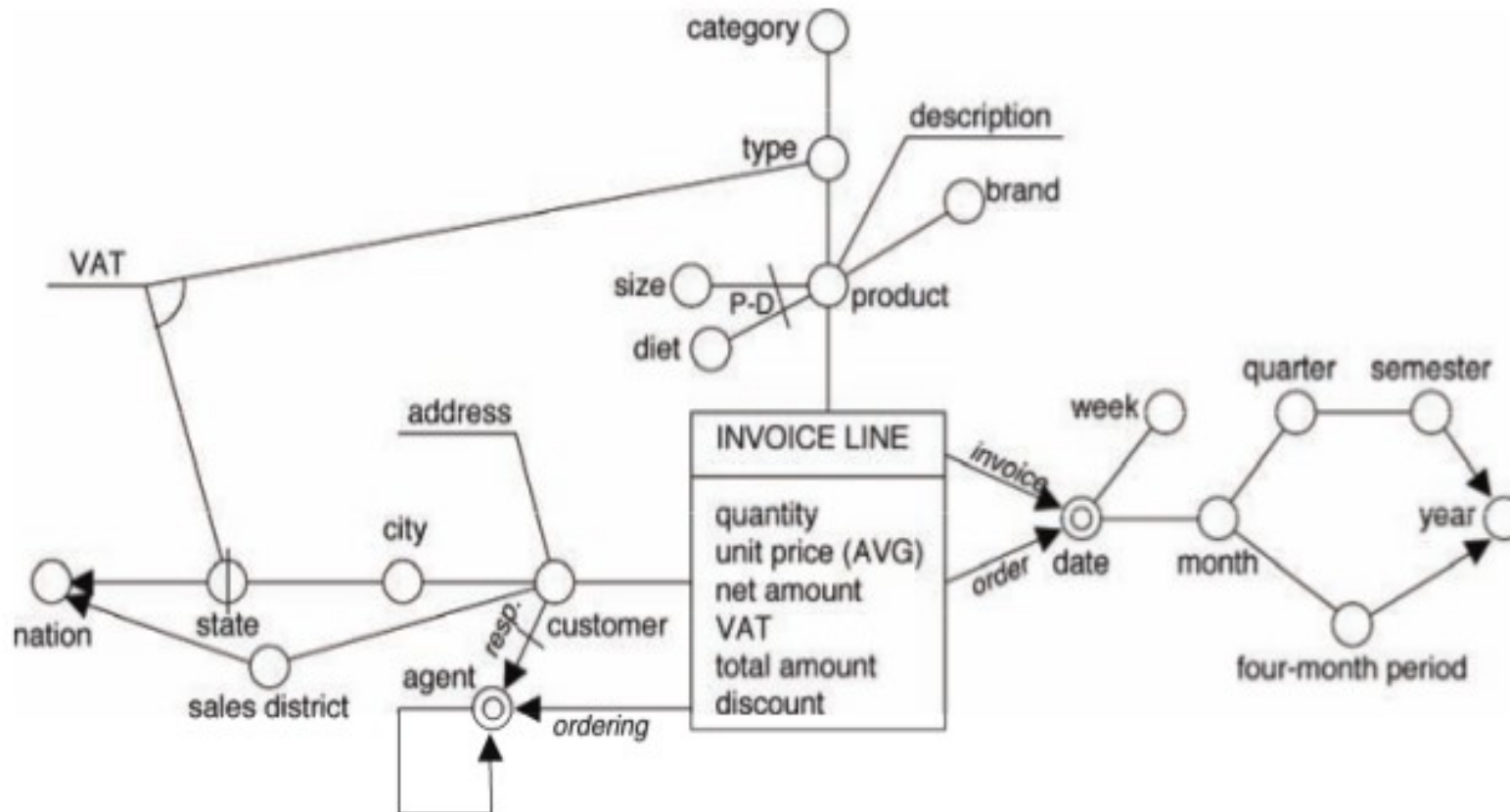
- Atributos
  - Descriptivos
  - Interdimensionales
- Convergencias
- Jerarquías
  - Compartidas
  - Incompletas
  - Recursivas
  - Dinámicas

# Conceptos avanzados

---

- Aristas
  - Múltiples
  - Opcionales
- Aditividad

# Conceptos avanzados





# Atributos descriptivos

---

- Especifican una propiedad de un atributo de dimensión.
- No son utilizados para agregados porque:
  - Tienen un dominio de valores continuos, e.g. peso de un producto
  - Tienen una asociación uno-a-uno, e.g. dirección de cliente

# Atributos interdimensionales

---

- Es un atributo cuyo valor es determinado por la combinación de dos o más atributos de dimensiones, que posiblemente pertenecen a diferentes jerarquías

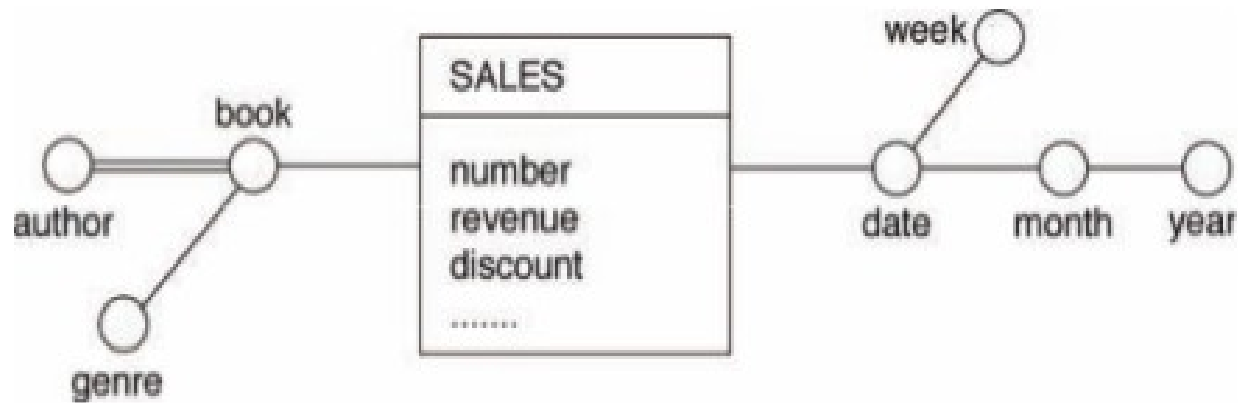
# Convergencia

---

- Tiene lugar cuando dos atributos de dimensión dentro de una jerarquía están conectados por dos o más rutas alternativas de asociación muchos-a-uno

# Arcos

- Opcionales
  - Asociación no definida para un subconjunto de eventos
- Múltiples

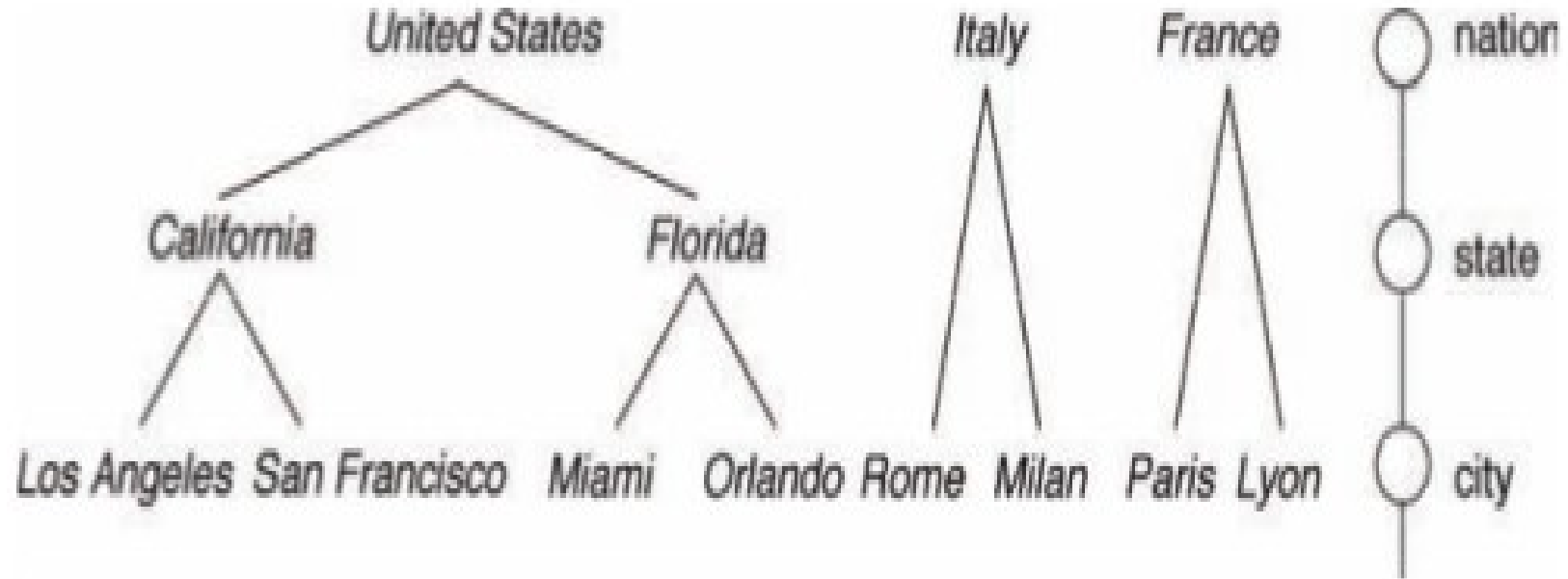


# Jerarquías

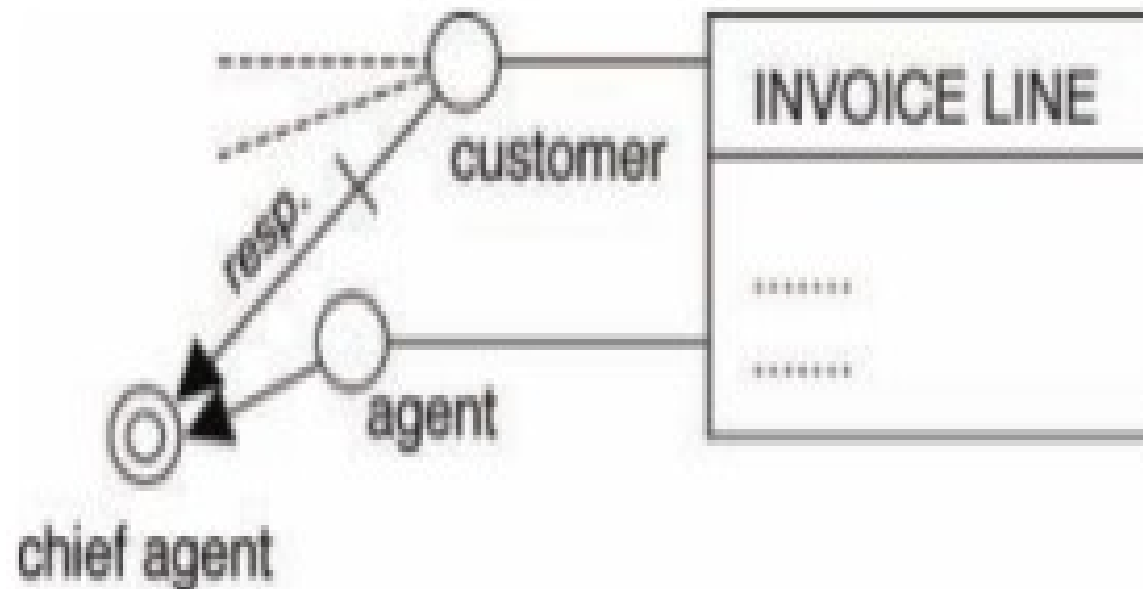
---

- Compartidas
  - Ej. tiempo y agente
- Incompletas
  - Ej. estado
  - Consejo: fragmentar hechos
- No balanceadas o recursivas
- Dinámicas
  - Hoy por ayer, ayer por hoy, verdad histórica

# Ejemplo de jerarquía incompleta



# Solución jerarquías recursivas

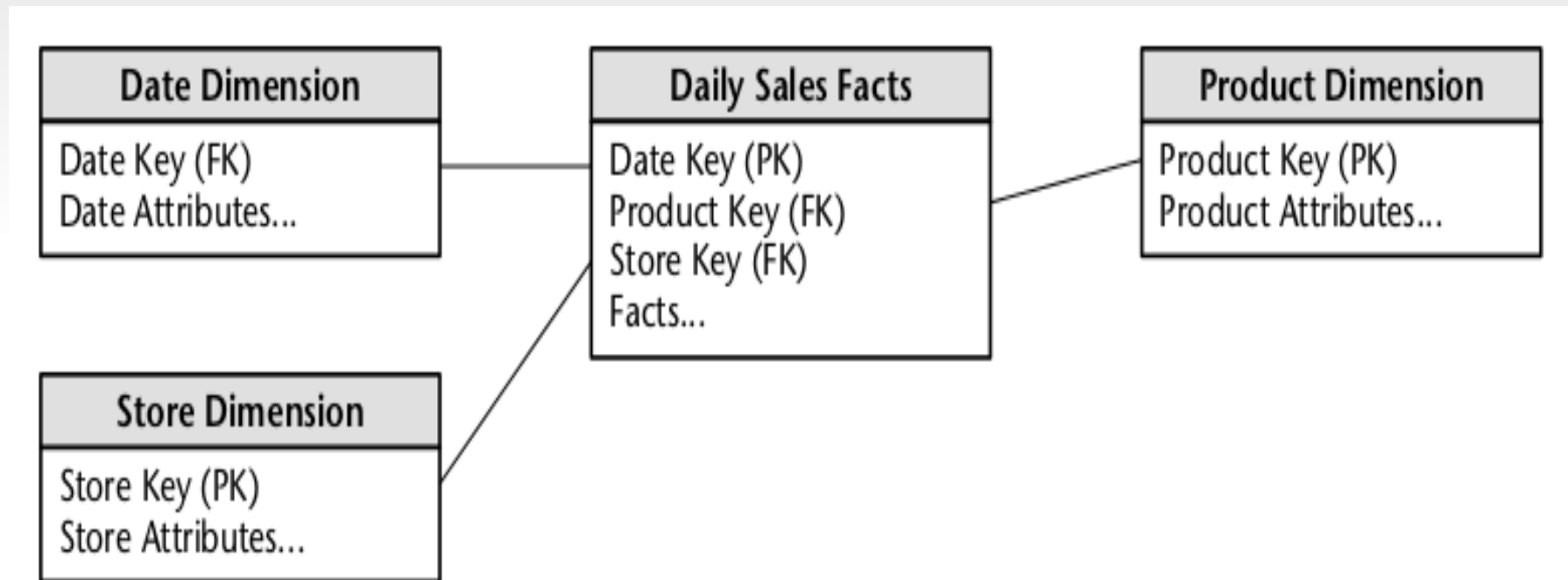


# Aditividad

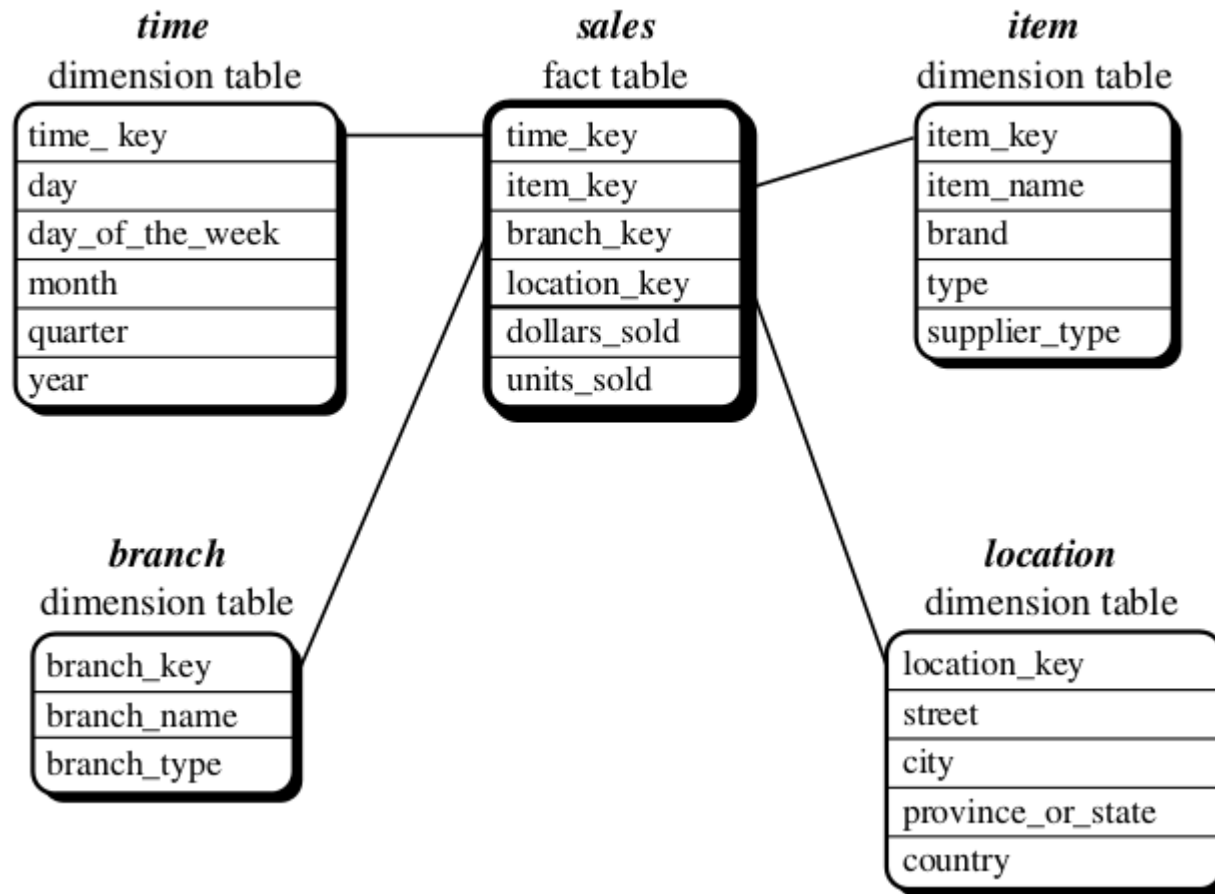
- Se requiere definir un operador apropiado para convertir los valores de medida que caracterizan eventos primarios en valores de medida que caracterizan eventos secundarios.
- Podemos distinguir tres tipos de medidas:
  - **Medidas de flujo**
    - Acumulativas en un período de tiempo
  - **Medidas de inventario**
    - Evaluadas en un momento dado
  - **Medidas de unidades**
    - Expresadas en términos relativos. Ej. TC



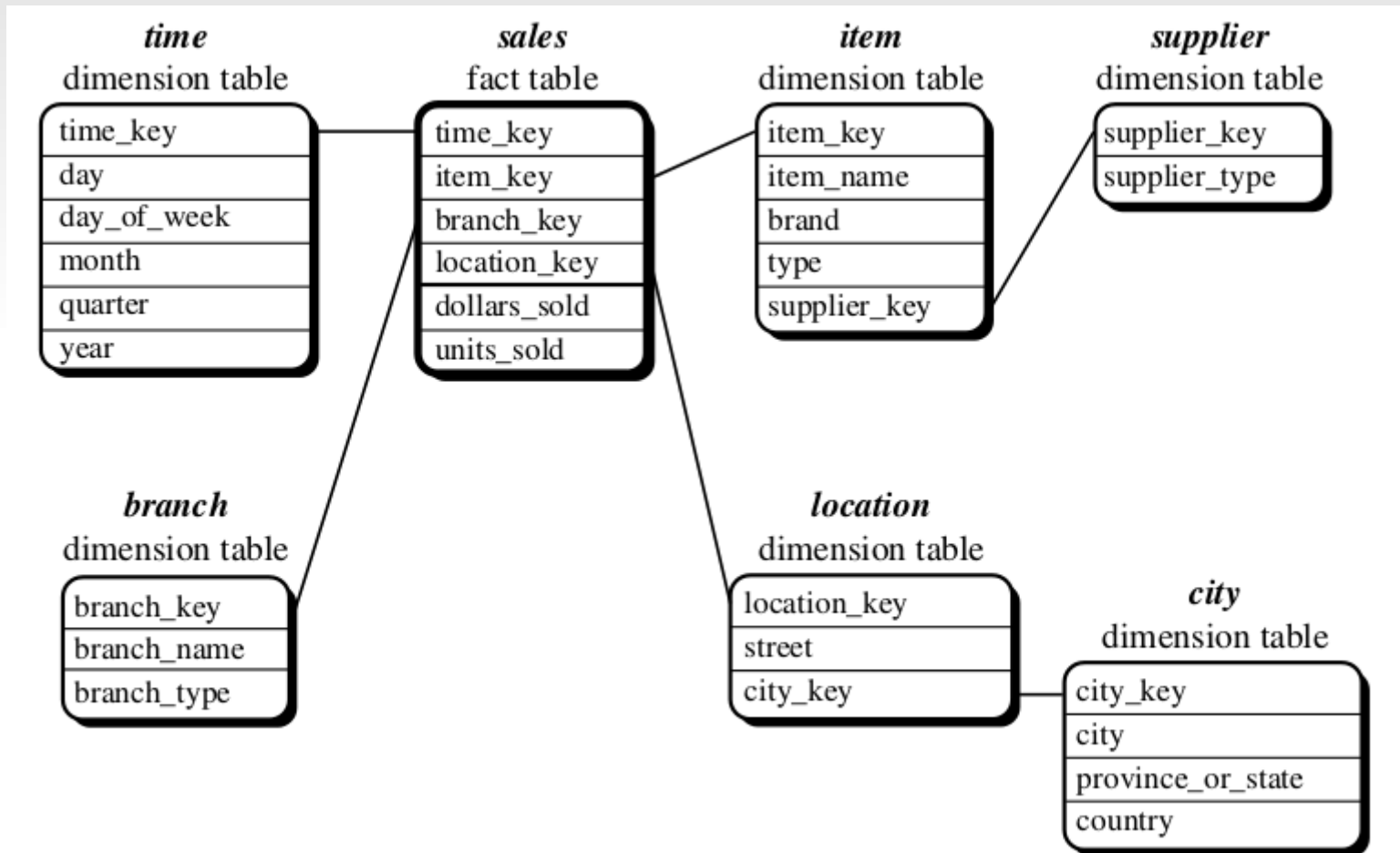
# Tablas en un modelo multidimensional



# Esquema de estrella



# Esquema de copo de nieve



# Referencias

---

- Golfarelli, M., Rizzi, S. *Data Warehouse Design: Modern Principles and Methodologies*. McGraw-Hill, 2009
- Jiawei, H., Kamber, M. *Data Mining: Concepts and Techniques (Second Edition)*. Morgan-Kaufmann, 2006
- Kimball, R., Ross, M. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition)*. John Wiley & Sons, 2002