

# **UNIVERSIDAD NACIONAL DE LA MATANZA**

## **INTELIGENCIA DE NEGOCIOS**

---

### **Modelo Dimensional – Parte 1**

**Docentes: ING. LORENA R. MATTEO**

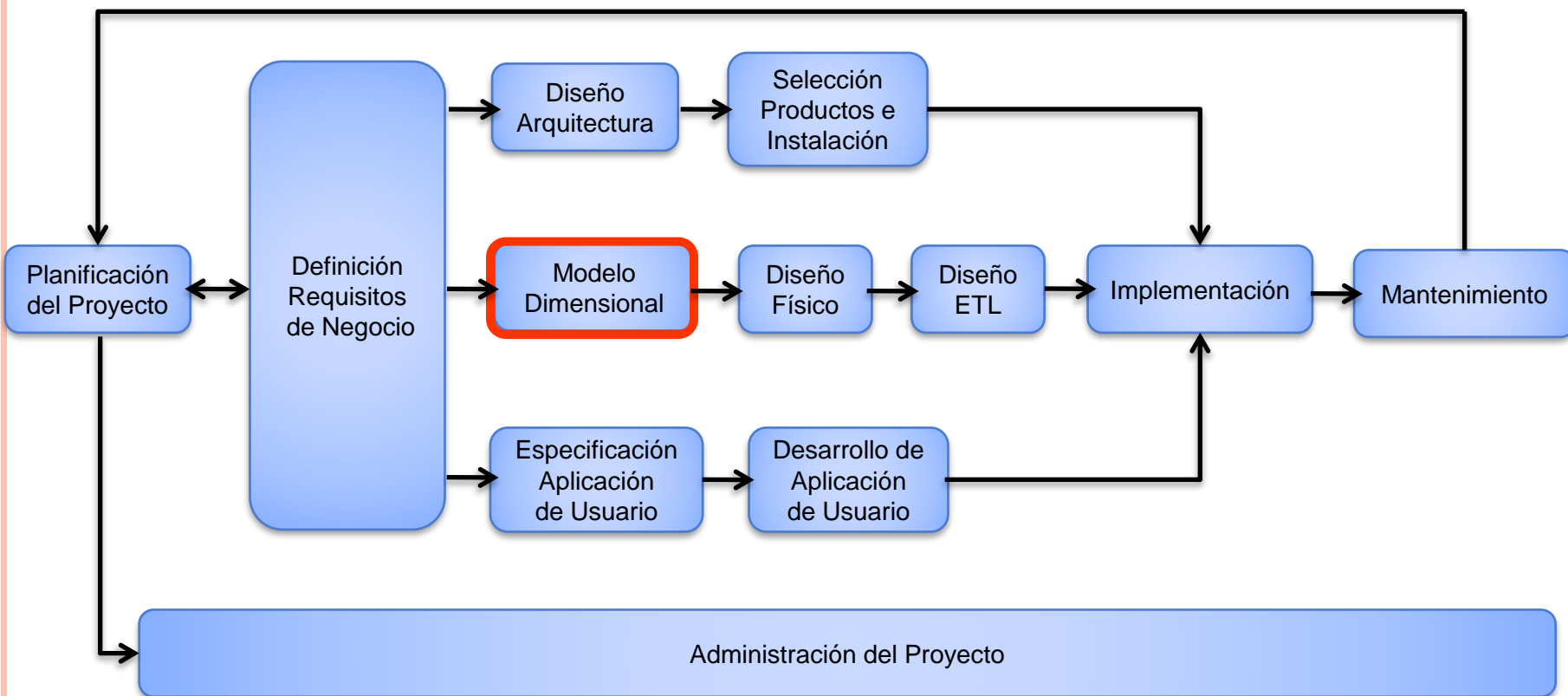
**Autores ppt orig.: Lic. HUGO M. CASTRO / MG. DIEGO BASSO**

**Fecha última actualización.: 15/04/2025**



# CICLO DE VIDA DE UN PROYECTO DE BI

15/4/2025 IN2025





# MODELO DE DATOS

## ○ Modelo de Entidad-Relación (OLTP)

- Orientado a la implementación de los procesos transaccionales
- Tareas operacionales
- Entidades, Atributos y Relaciones

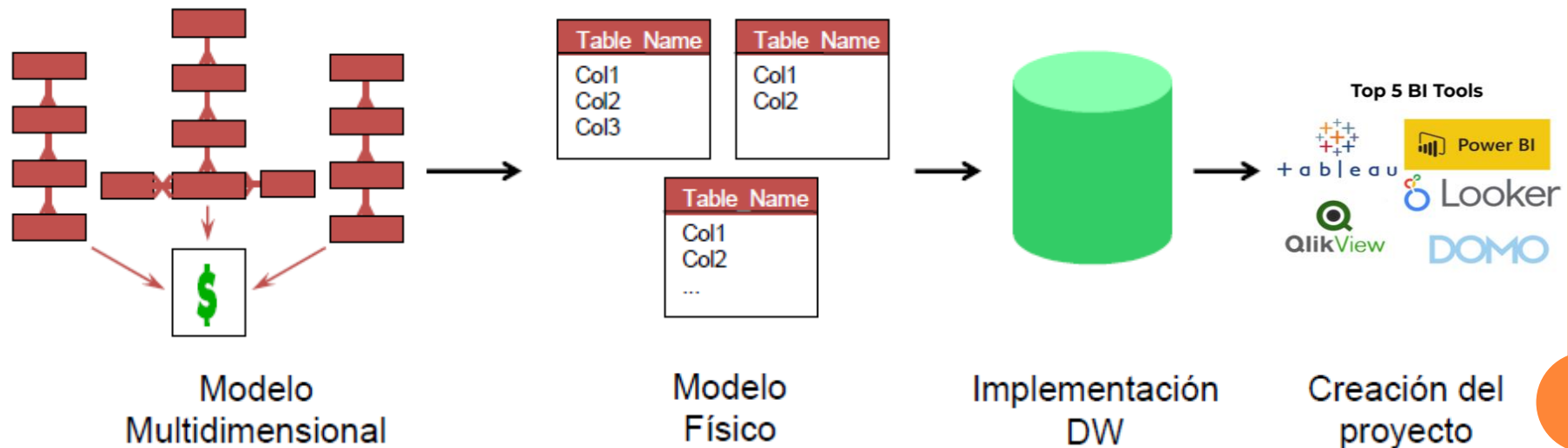
## ○ Modelo Dimensional (OLAP)

- Orientado a las características del negocio (variables del negocio)
- Ofrece una visión de los datos orientada hacia el análisis y una rápida y flexible navegación por estos.



# MODELO DIMENSIONAL

- “El modelo dimensional es una técnica de diseño que busca presentar los datos en un framework estándar, intuitivo y escalable, que permite un acceso a los datos altamente performante, basándose en el modelado relacional, pero con algunas restricciones de diseño importantes”.





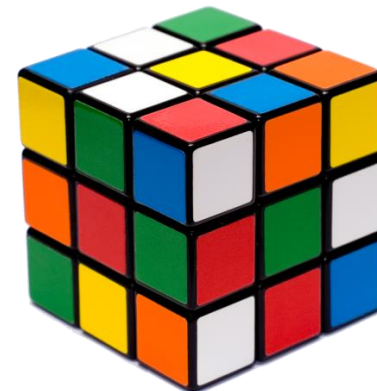
# MODELO DIMENSIONAL

- Busca acercar los datos a la manera en que estos serán convertidos en información útil para los usuarios del negocio.
- Componentes
  - **Hechos** (Fact) / Medidas / Indicadores de negocio
  - **Dimensiones**
  - **Atributos**
    - Elementos
    - Jerarquías
  - **Relaciones**



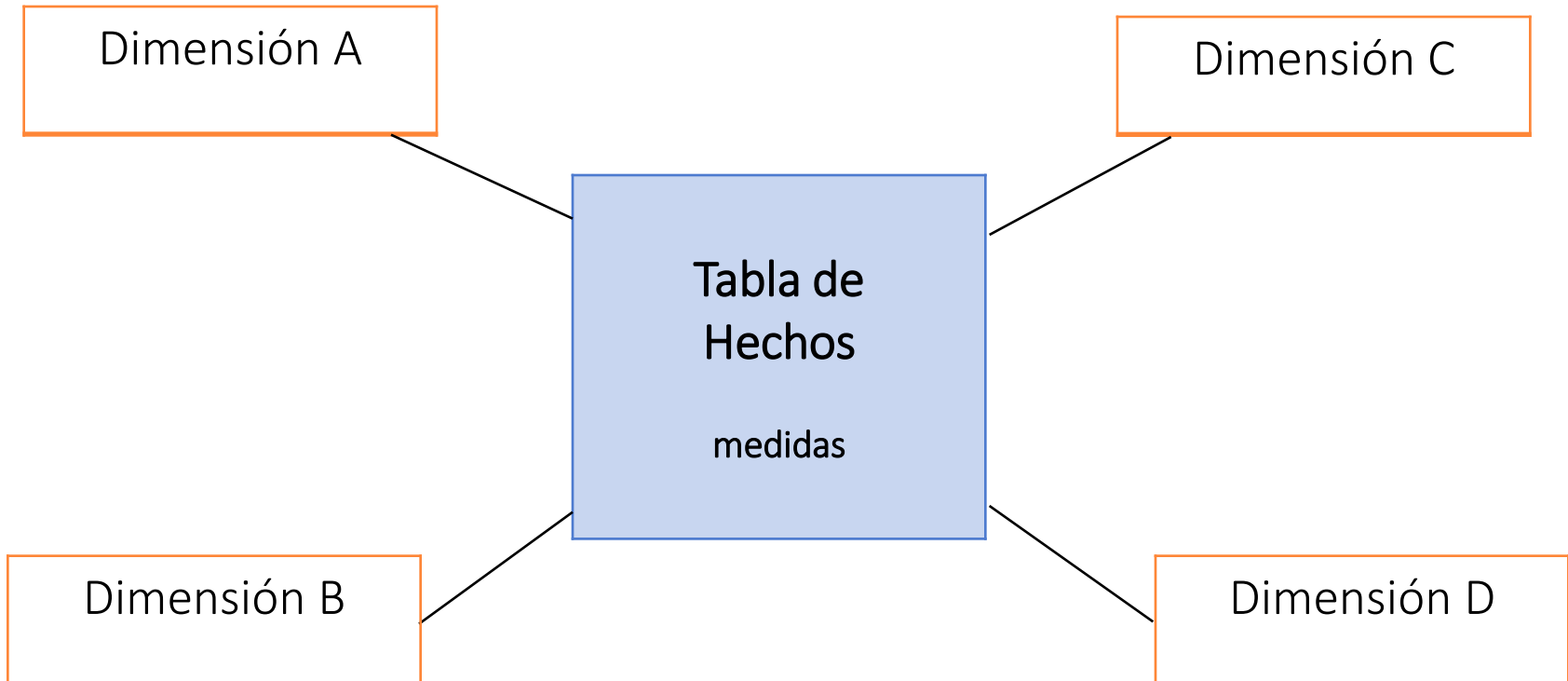
# MODELO DIMENSIONAL

- Representación de la estructura que va a tener el DW.
- Define un nivel mínimo de detalle (**Granularidad**).
- Se puede revisar con el profesional de negocios.
- El modelo de datos dimensional es lo que hace que un DW sea una base de datos **orientada al negocio**.
- La forma de representar la organización de los datos en un modelo dimensional es a través de un **cubo**.





# ESQUEMA DE UN MODELO DIMENSIONAL





# HECHOS O MEDIDAS (FACTS)

## ○ Hecho

- Es algo que ocurre en el tiempo (operación o actividad).

## ○ Medidas / Indicadores

- Son valores **numéricos** que describen el **hecho** analizado.
- Las medidas más usadas son aditivas.
- Ayudan a medir el desempeño del negocio.
- Tipos
  - **Básicas**: existen físicamente en el DW junto a los atributos que las caracterizan. Proviene de las fuentes de datos OLTP.
  - **Calculadas o derivadas**: construidas a partir de medidas básicas y pueden o no estar almacenadas en el DW.





# HECHOS O MEDIDAS (FACTS)

## ○ Medidas / Indicadores

- Columnas agregables del DW
- Admiten sumas, promedios, consolidaciones, etc.
- Dan lugar a las métricas.

## ○ Ejemplos

- ✓ \$ Venta (en un supermercado)
- ✓ Unidades vendidas (en un supermercado)
- ✓ Minutos hablados (en una telefónica)
- ✓ Duración de la llamada (en una telefónica)
- ✓ Saldo \$ (en un banco)

- Todos los hechos, medidas o indicadores de negocio se guardan en la **tabla de hechos** y deben tener el **mismo nivel de granularidad**.



# DIMENSIONES

- Categorías para describir el contexto (variables del negocio) en el que se pueden analizar las medidas.
- Contienen uno o más **atributos** de tipo texto agrupados lógicamente. Las dimensiones no comparten atributos.
- Sirven para **mostrar**, **agrupar** y **filtrar** información.
- Ejemplos
  - ✓ TIPO DE LLAMADA
  - ✓ EMPLEADO
  - ✓ SUCURSAL
  - ✓ GEOGRAFÍA
  - ✓ CLIENTE
  - ✓ VENDEDOR
  - ✓ PRODUCTO
  - ✓ AÑO
- Si hay valores numéricos deben ser categorizados.
  - *Ejemplos:* Rangos de edades, niveles de precios



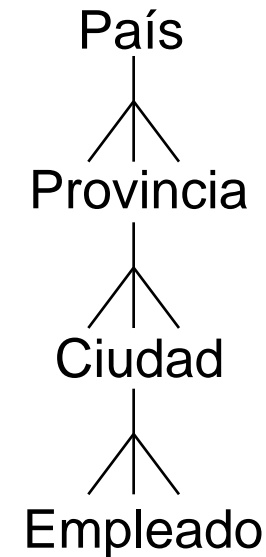
# DIMENSIONES CON ATRIBUTOS

- Las dimensiones definen los niveles de análisis (**jerarquías**). Esto permite obtener sumariaización o agregación.

## DIM Tiempo



## DIM Geografía





# ATRIBUTOS

- Representan categorías o clases de elementos que tienen el mismo nivel lógico dentro de una dimensión.
- Valores cualitativos de una transacción.
- Visualizar la información de las dimensiones a distintos **niveles de detalle** y **agrupar** datos para analizar las métricas.
- Son una agrupación de **elementos** (valores que toma).
- Son parecidos a las entidades del modelo E-R.
- Ejemplos:

- ✓ Provincia
- ✓ Ciudad
- ✓ Mes
- ✓ Ítem
- ✓ Empleado
- ✓ Edad
- ✓ Cliente
- ✓ Año

Provincia	
Buenos Aires	} <i>Elementos</i>
Córdoba	
La Pampa	



# ELEMENTOS

- Son los valores o instancias que toma un **atributo**.
- Ejemplos:

- ✓ Provincia
- ✓ Ciudad
- ✓ Mes
- ✓ Ítem
- ✓ Empleado
- ✓ Edad
- ✓ Cliente
- ✓ Año

Provincia
Buenos Aires
Córdoba
La Pampa

} *Elementos*

Mes
Enero 2024
...
Diciembre 2024

} *Elementos*

Edad
0-15 años
16-30 años
31-45 años

} *Elementos*



# JERARQUÍAS DE ATRIBUTOS

- Ordenamiento lógico de atributos dentro de la dimensión.
- Contiene relaciones entre los atributos de una misma dimensión.
  - 1:1
  - 1:N
  - N:N
- Las **jerarquías** definidas permiten bajar el nivel de complejidad de los datos y ayudar al usuario a ver los datos en el nivel de detalle que necesita.



# JERARQUÍAS DE ATRIBUTOS

- **Importante:** si en el modelo de la dimensión aparece un sólo nivel, prestar atención a la cardinalidad.
  - *Cardinalidad alta:* Hace necesaria la definición de jerarquías.
    - Ejemplo: Si hay 20.000 productos, hace falta definir niveles jerárquicos para poder analizar.
  - *Cardinalidad baja:* Aceptable no tener jerarquías.
    - Ejemplo: Dimensión Mercado tiene dos posibles valores (Local o Extranjero).



# GRANULARIDAD

- Es el nivel más detallado (mínimo nivel de detalle o nivel atómico) que tiene cada tabla de hechos en el modelo del proceso de negocio. Se expresa como:

**<AtribDim1> x ... x <AtribDimN>**

siendo AtribDimx el nivel más detallado de cada Dimensión

- Ejemplos:

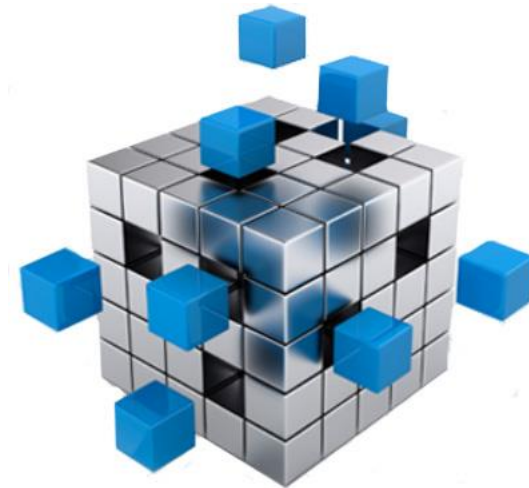
- la granularidad de un modelo dimensional que consta de las dim. de fecha, almacén y producto es producto vendido en el almacén por día, es decir:  
**Producto x Almacén x Fecha**
- si en el modelo existe la dimensión Fecha (Tiempo) (con las jerarquías de año y trimestre) la granularidad de esa dimensión en la tabla de hechos estará hasta el nivel trimestral, es decir no se tendrá información para los días o meses individuales: **<AtribDim1>x ... x Trimestre**





# CUBO OLAP

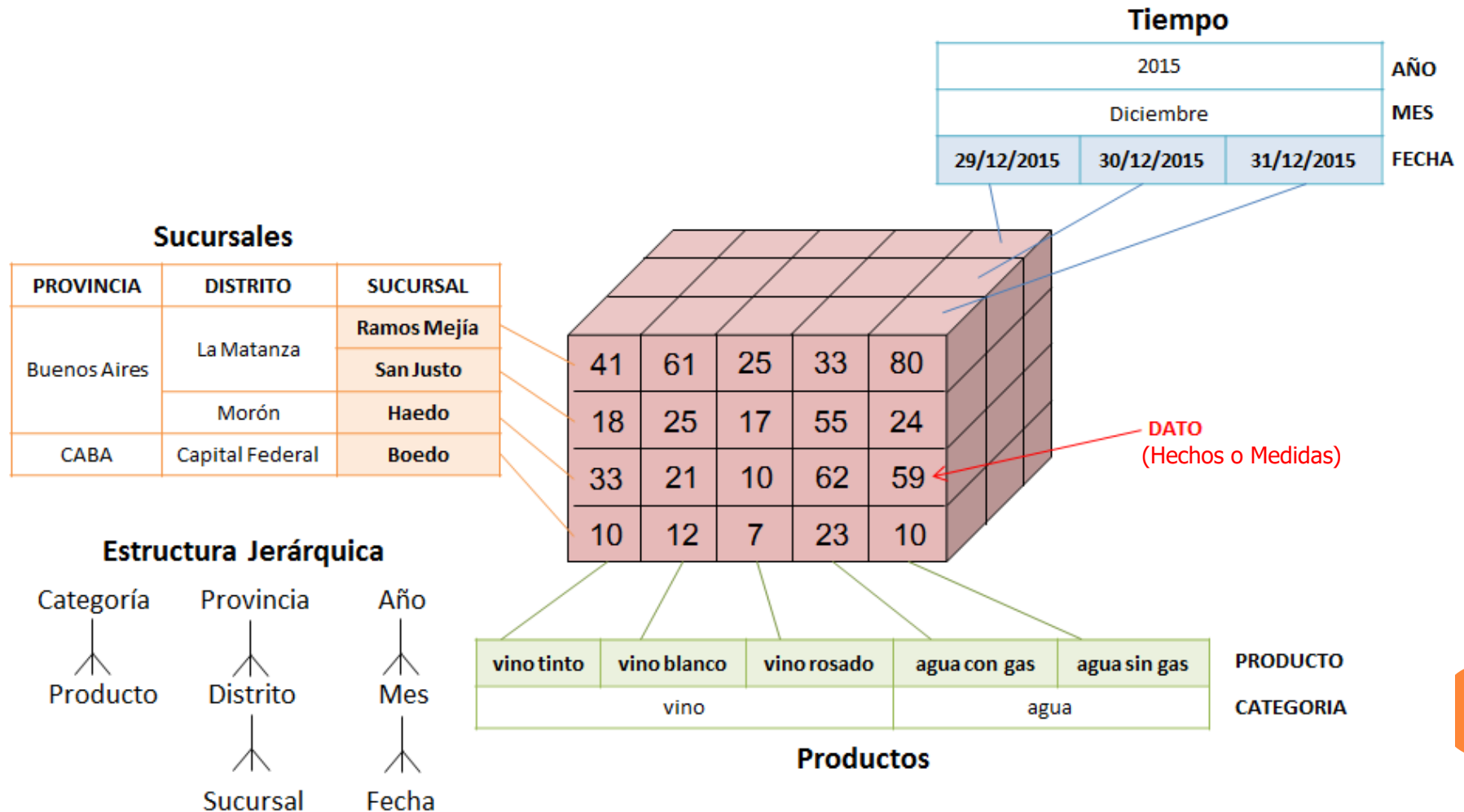
- Es una base de datos que posee diversas dimensiones.
- Cada dimensión es una “arista” del cubo.
- En las celdas del cubo están las medidas (hechos).
- Es independiente de la forma en que realmente se almacenan los datos.





# EJEMPLO DE CUBO OLAP

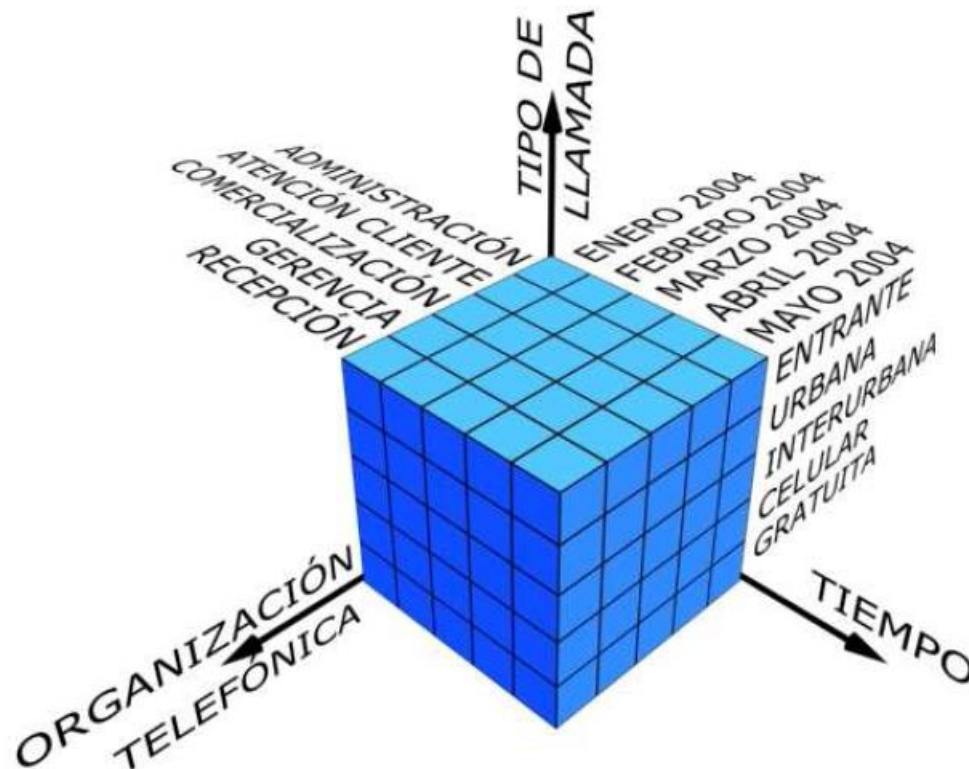
- Cubo que posee como medida la “Cantidad de pedidos” por Fecha, Producto y Sucursal.





## EJEMPLO DE CUBO OLAP

- Cubo que posee como medida la “*Duración de las Llamadas*” y como dimensiones Tiempo, Tipo de Llamada y Organización Telefónica.





## ¿POR DÓNDE EMPEZAMOS?

- La variedad y disparidad de datos dentro de una empresa requiere un enfoque incremental.
- No se puede cargar todo y ver después quién lo va a usar.
- Este enfoque puede llevar al fracaso.
- Se pierde tiempo y esfuerzo sin obtener resultados.





# ¿POR DÓNDE EMPEZAMOS?

- Por las necesidades de uso de información de los usuarios de negocios.
  - Es decir, por los *requerimientos*.
- Priorizar
  - Comenzar por lo más importante para la empresa.
  - ¿Qué valores numéricos necesitan analizar?
  - ¿Cuáles son el/los procesos de negocios que generan esos valores?





# PROCESO DE DISEÑO DIMENSIONAL

## Pasos

1. Elegir el **proceso de negocios** a modelar (ventas diarias, manejo de stock, etc.).
2. Elegir las **dimensiones** que van a intervenir (tener en claro los **atributos**, **jerarquías** y elementos de atributos).
3. Elegir la **granularidad** (nivel de detalle) del proceso de negocios con que se van a guardar los datos en el DW.
4. Elegir los **hechos** y **medidas** que se van a utilizar en la tabla de hechos (básicas y calculadas).

*Metodología de Ralph Kimball para  
el armado incremental del DW*



# CASO DE ESTUDIO

## ○ Relevamiento

- Un supermercado posee distintas sucursales en distintos lugares del país en las que comercializa distintos productos.
- Cada producto pertenece a una familia y cada familia pertenece a un tipo de producto. Ejemplo: el producto Yogur A x 170gr pertenece a la familia yogur y al tipo lácteo.
- Registra los tickets de las ventas de sus sucursales en una BBDD.
- La información que se registra de cada producto vendido es :

• Fecha	• Costo de cada producto
• Hora	• Cantidad de unidades
• Sucursal	• Precio unitario
• Nro. de ticket	• Importe
• Producto	



# CASO DE ESTUDIO

## ○ Relevamiento

- Quiere aprovechar esta información del año actual y los dos anteriores para ver, entre otras cosas.
  1. ¿Cuánto se vendió (en unidades e importes) por sucursal y tipo de producto en los distintos trimestres, del mismo año o año actual y alguno de los anteriores.
  2. ¿Cómo fueron las ventas en las distintas provincias en las que están las sucursales?
  3. ¿Cuál fue la ganancia obtenida en la venta de cada producto?
  4. ¿Cómo fueron las ventas mensuales con tarjetas (crédito y débito) respecto al pago en efectivo (por sucursal, por producto, por mes, o trimestre o año)?
  5. ¿Cómo evolucionaron las ventas en los últimos 12 meses?



# CASO DE ESTUDIO

## ANÁLISIS DE LOS REQUERIMIENTOS



1. Ventas por sucursal y tipo de producto en los distintos trimestres, del mismo año o año actual y alguno de los anteriores.
  - ❑ Medidas básicas: **cantidad de unidades, importe de venta**
  - ❑ Ver el detalle de las medidas por **sucursal**. El supermercado posee distintas sucursales.
  - ❑ Las medidas están referidas a un **producto** y a su vez se desea verlos agrupados por **tipo de producto**.
  - ❑ A su vez cada producto pertenece a una **familia de producto** y cada familia pertenece a un tipo de producto.
  - ❑ Ver la evolución de las ventas en el transcurso del **tiempo**, detallados por **trimestre** o **año**.
- ✓ Para satisfacer estas necesidades debemos crear las dimensiones **SUCURSAL, PRODUCTO** y **TIEMPO**.

# CASO DE ESTUDIO

## ANÁLISIS DE LOS REQUERIMIENTOS



2. Ventas en las distintas provincias en las que están las sucursales.
  - ❑ La información de las ventas por **sucursal** requiere que también pueda visualizarse por **provincia** asociada.
3. Ganancia obtenida en la venta de cada producto.
  - ❑ Medidas básicas: **costo del producto**
  - ❑ Medidas derivadas o calculadas: **ganancia de venta**
4. Ventas mensuales con tarjetas (crédito y débito) y pago en efectivo (por sucursal, por producto, por mes, o trimestre o año)
  - ❑ Analizar las ventas por diferentes **formas de pago**.
5. Ver evolución de las ventas en el transcurso del **tiempo**, detallados por **trimestre**, **año** o **mes**. La información del ticket permite conocer hasta la **fecha** y **hora**.

# CASO DE ESTUDIO

## ANÁLISIS DE LOS REQUERIMIENTOS



- Resumen de los requerimientos, agrupar por apertura de las dimensiones y comenzar a definir los cubos a crear.

Hecho a medir: Venta de Productos			
	Dimensiones		
Medidas	Tiempo	Sucursal	Producto
Cantidad	X	X	X
Importe	X	X	X
Costo	X	X	X
Ganancia	X	X	X

- Proceso de negocio a modelar: **Ventas**

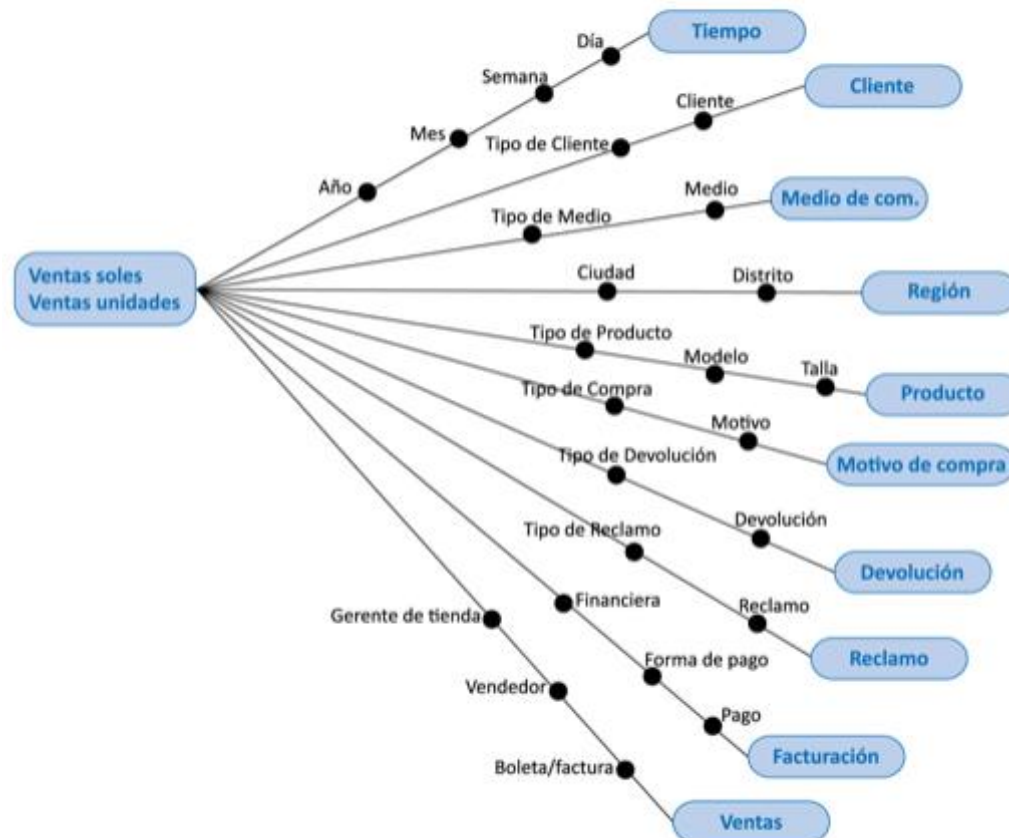
# CASO DE ESTUDIO

## ANÁLISIS DE LOS REQUERIMIENTOS STARNET

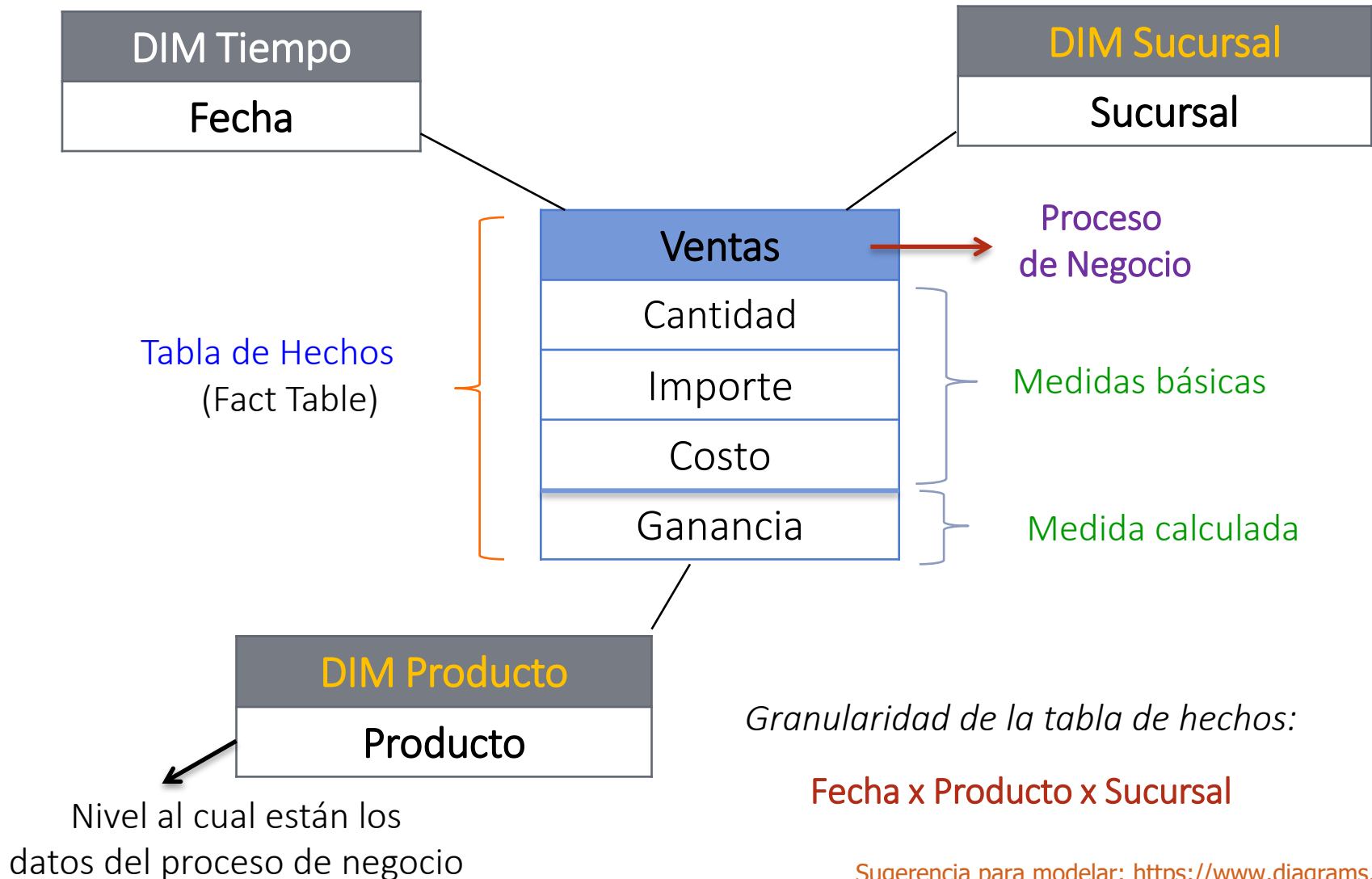


- Otra forma de mostrar los requerimientos gráficamente si se cuenta con ese nivel de detalle:

Diagrama Starnet de requerimientos para un sistema de Business Intelligence



# MODELO DIMENSIONAL “CONCEPTUAL” (GRUPO DE HECHOS Y GRANULARIDAD)





# RELACIONES ENTRE ATRIBUTOS

## 1:1 (uno-a-uno)

- A cada código de artículo le corresponde una descripción.
- A cada descripción le corresponde un código de artículo.
- Ambos atributos forman parte de la misma dimensión.

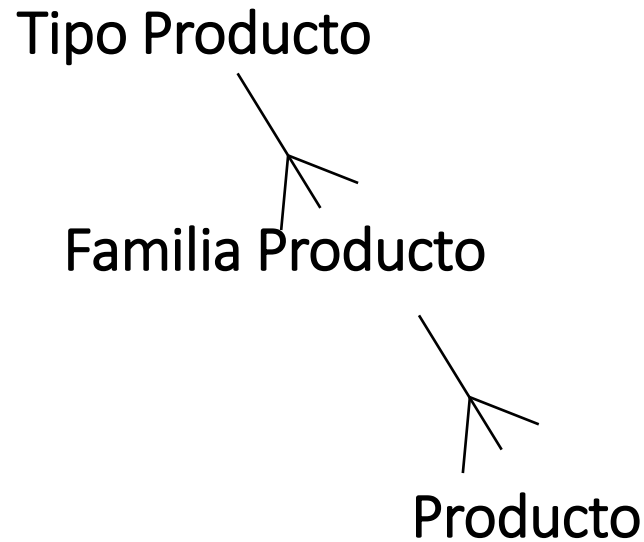
## 1:N (uno-a-muchos)

- Analicemos la dimensión *Producto*
- Cada **producto** pertenece a una **familia de producto** y cada familia pertenece a un **tipo de producto**.
  - Una familia de productos comprende varios productos
  - Un tipo de producto comprende varias familias



# DIMENSIÓN PRODUCTO

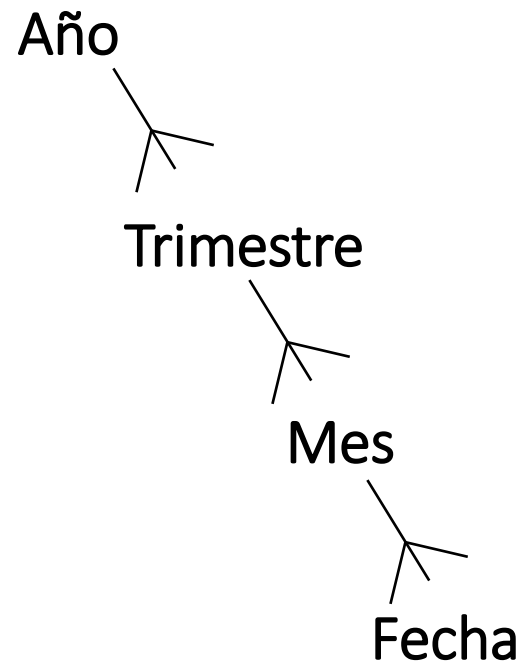
- La relación jerárquica es:





# DIMENSIÓN TIEMPO

- ¿De qué manera se quiere analizar la información?
  - Monto de ventas por sucursal y tipo de producto en distintos **trimestres**, del mismo **año** o año actual.
  - Evolución de ventas en los últimos 12 **meses**.
- La relación **jerárquica** es:



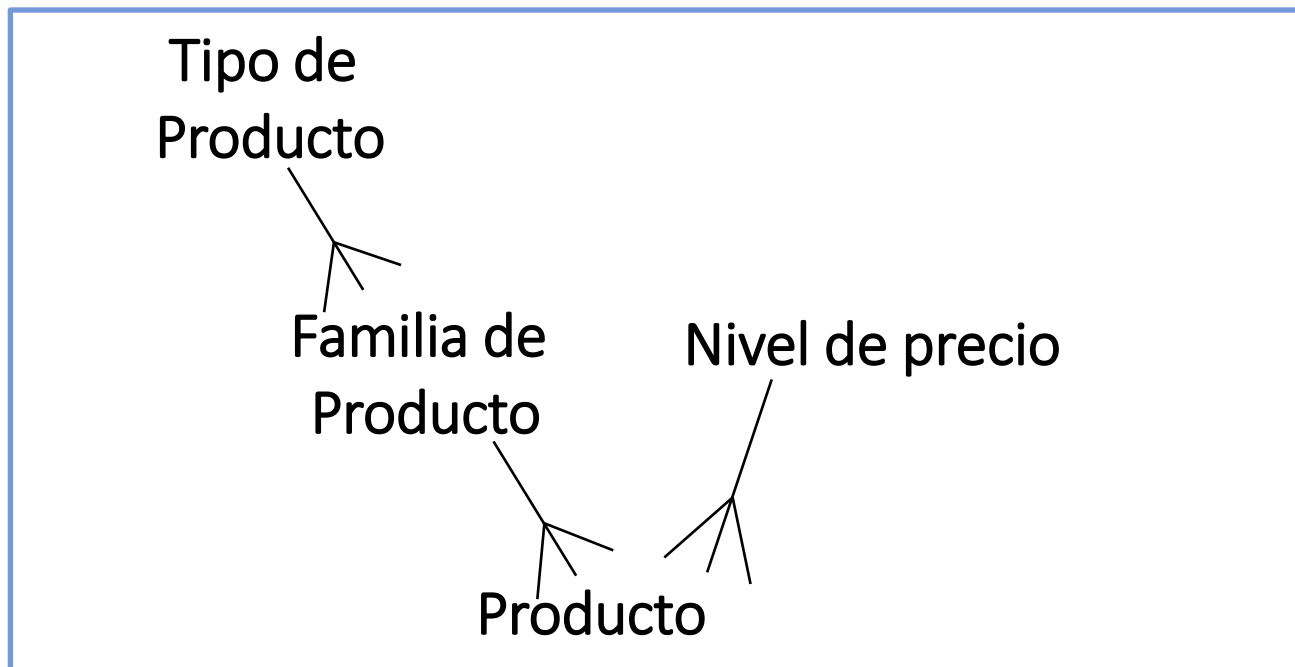




# DIMENSIONES CON JERARQUÍAS MÚLTIPLES

- Puede ocurrir que además de agrupar los productos por **tipo de producto** y **familia de producto** sea útil agruparlos por **nivel de precio** (caro, mediano, barato)
- Esto se representa con una doble jerarquía:

## DIM PRODUCTO



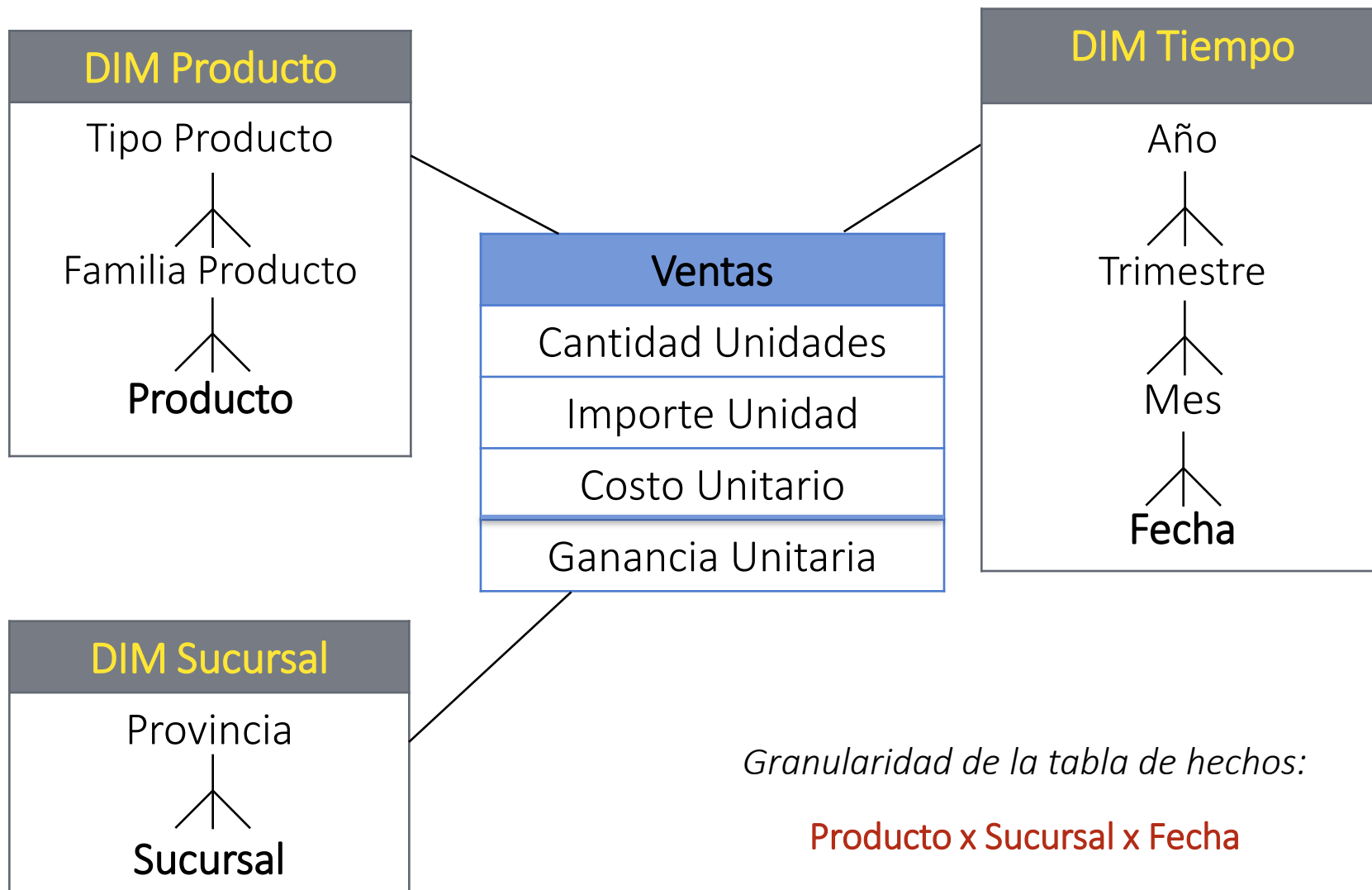


# RELACIONES ENTRE ATRIBUTOS

## N:N (muchos-a-muchos)

- En el caso de estudio: “El supermercado posee distintas **sucursales** en distintos lugares del país en las que comercializa distintos **productos**.”
  - Cada producto se vende en varias sucursales
  - En cada sucursal se venden varios productos
- Los atributos *Sucursal* y *Producto* corresponden a dimensiones diferentes.
- Se relacionan a través la **tabla de hechos** y representan la relación muchos a muchos que existe entre las dimensiones.

# MODELO DIMENSIONAL “LÓGICO” (MUESTRA JERARQUÍAS)



*Granularidad de la tabla de hechos:*

**Producto x Sucursal x Fecha**



# CASO DE ESTUDIO - CONCLUSIONES

- Al consultar los datos y usar los metadatos con las estructuras jerárquicas de las dimensiones se pueden obtener totales por varias cosas:
  - ✓ Importes totales por tipo de producto, provincia, mes.
  - ✓ Importes totales mensuales por Familia de productos de la provincia de Córdoba o Buenos Aires.
  - ✓ Comparar el total de unidades vendidas de un mismo producto en diferentes sucursales o provincias.
- Recordar que las **dimensiones** sirven para:
  - Mostrar
  - Filtrar
  - Agrupar

# CASO DE ESTUDIO - CONCLUSIONES



- Si un usuario nos pregunta:
  - ¿Voy a poder obtener los totales trimestrales de los importes de todas las provincias? SI
  - ¿Voy a poder obtener los totales trimestrales de los impuestos de todas las provincias? NO
    - No hay una medida para el valor del impuesto
  - ¿Voy a poder obtener los totales mensuales de los descuentos de todas las provincias? NO
    - No hay una medida para el descuento
  - ¿Voy a poder obtener los totales anuales de los importes de los productos “Yogur A” vendidos en cada sucursal de CABA? SI

# CASO DE ESTUDIO - CONCLUSIONES



- Si un usuario nos pregunta:
  - ¿Voy a poder obtener los totales mensuales de los importes de todas las provincias por forma de pago? **NO**
    - El ticket no contiene información sobre la Forma de Pago.
    - No hay una dimensión Forma de Pago.
  - ¿Voy a poder obtener los totales mensuales de los importes de todas las sucursales grandes de todas las provincias? **NO**
    - En este modelo no está la información sobre las sucursales para saber si son grandes o chicas.
    - Se puede resolver fácilmente agregando otra jerarquía para la dimensión Sucursal y teniendo una fuente de datos que traiga esa información.



# VENTAJAS DEL MODELO DIMENSIONAL

- Presenta la información de una manera estándar, sencilla y sobre todo intuitiva para los usuarios.
- Resiste cambios inesperados en la conducta del usuario.
- Todas las dimensiones son equivalentes y pueden ser pensadas como puntos de acceso a la tabla de hechos.
- Facilidad para adaptarse a cambios.
  - ❑ Agregar hechos o medidas a la tabla de hechos, siempre que sean consistentes con el mayor nivel de detalle de las dimensiones.
  - ❑ Agregar atributos a las dimensiones.
  - ❑ Agregar nuevas dimensiones.



# VENTAJAS DEL MODELO DIMENSIONAL

## Facilidad para adaptarse a cambios

1. Agregar un nuevo hecho no planeado, siempre y cuando sea compatible con la granularidad definida.

Fecha	SKU	Volumen
05/04/2011	Agua 1.5 L	10
05/04/2011	Agua 2.25 L	35
05/04/2011	Cola 1 L	15
06/04/2011	Cola 1.5 L	10





# VENTAJAS DEL MODELO DIMENSIONAL

## Facilidad para adaptarse a cambios

- Supongamos que quiero agregar el *importe*:
  - Agrego una columna a la tabla de hechos.
  - Completo los valores para los registros existentes.
    - Con algún valor por defecto.
    - Voy al transaccional a buscar el valor correcto (muy costoso).

Fecha	SKU	Volumen	Importe
05/04/2011	Agua 1.5 L	10	0
05/04/2011	Agua 2.25 L	35	0
05/04/2011	Cola 1 L	15	0
06/04/2011	Cola 1.5 L	10	0
10/04/2011	Agua 1.5 L	5	20

- ✓ El nuevo hecho respeta la granularidad ya definida.



# VENTAJAS DEL MODELO DIMENSIONAL

## Facilidad para adaptarse a cambios

- Supongamos que trabajo todas bebidas línea Coca Cola y me doy cuenta, que tener la *línea de productos* es importante:
  - ❑ ¿La línea de productos respeta la granularidad de la tabla?

Fecha	SKU	Volumen
05/04/2011	Agua 1.5 L	10
05/04/2011	Agua 2.25 L	35
05/04/2011	Cola 1 L	15
06/04/2011	Cola 1.5 L	10



# VENTAJAS DEL MODELO DIMENSIONAL

## Facilidad para adaptarse a cambios

- Si agrego la línea de productos a la tabla, los hechos no varían, entonces respeta la granularidad.

Fecha	Línea	SKU	Volumen
05/04/2011	Coca Cola	Agua 1.5 L	10
05/04/2011	Coca Cola	Agua 2.25 L	35
05/04/2011	Coca Cola	Cola 1 L	15
06/04/2011	Coca Cola	Cola 1.5 L	10

- ✓ Simplemente es agregar una columna y completarla con los valores correspondientes.



# VENTAJAS DEL MODELO DIMENSIONAL

## Facilidad para adaptarse a cambios

2. Agregar en una dimensión un atributo que no había sido planteado.
  - ❑ Supongamos que tengo la dimensión País y quiero agregar la densidad poblacional:

### **DIM PAÍS**

Id_pais	Desc_pais
1	Argentina
2	Chile
3	Perú



# VENTAJAS DEL MODELO DIMENSIONAL

## Facilidad para adaptarse a cambios

- Es un atributo directamente relacionado con la dimensión.
- No tiene ningún impacto en la tabla de hechos.
- Se agrega un nuevo campo a la dimensión con el nuevo valor.

### **DIM PAÍS**

Id_pais	Desc_pais	Desc_densi
1	Argentina	1000
2	Chile	700
3	Perú	500



# VENTAJAS DEL MODELO DIMENSIONAL

## Facilidad para adaptarse a cambios

3. Romper una dimensión existente llevándola a un nivel de granularidad menor, a partir de un momento dado.
  - ❑ Quiero quedarme con menos nivel de detalle.
  - ❑ Se puede ver que en la tabla de hechos la dimensión Tiempo está a nivel de fecha.
  - ❑ Dejar sólo el nivel superior de la dimensión (ej. mes)

Id_mes	Desc_mes
04	Abril



Fecha	Id_mes
05/04/2011	04
05/04/2011	04

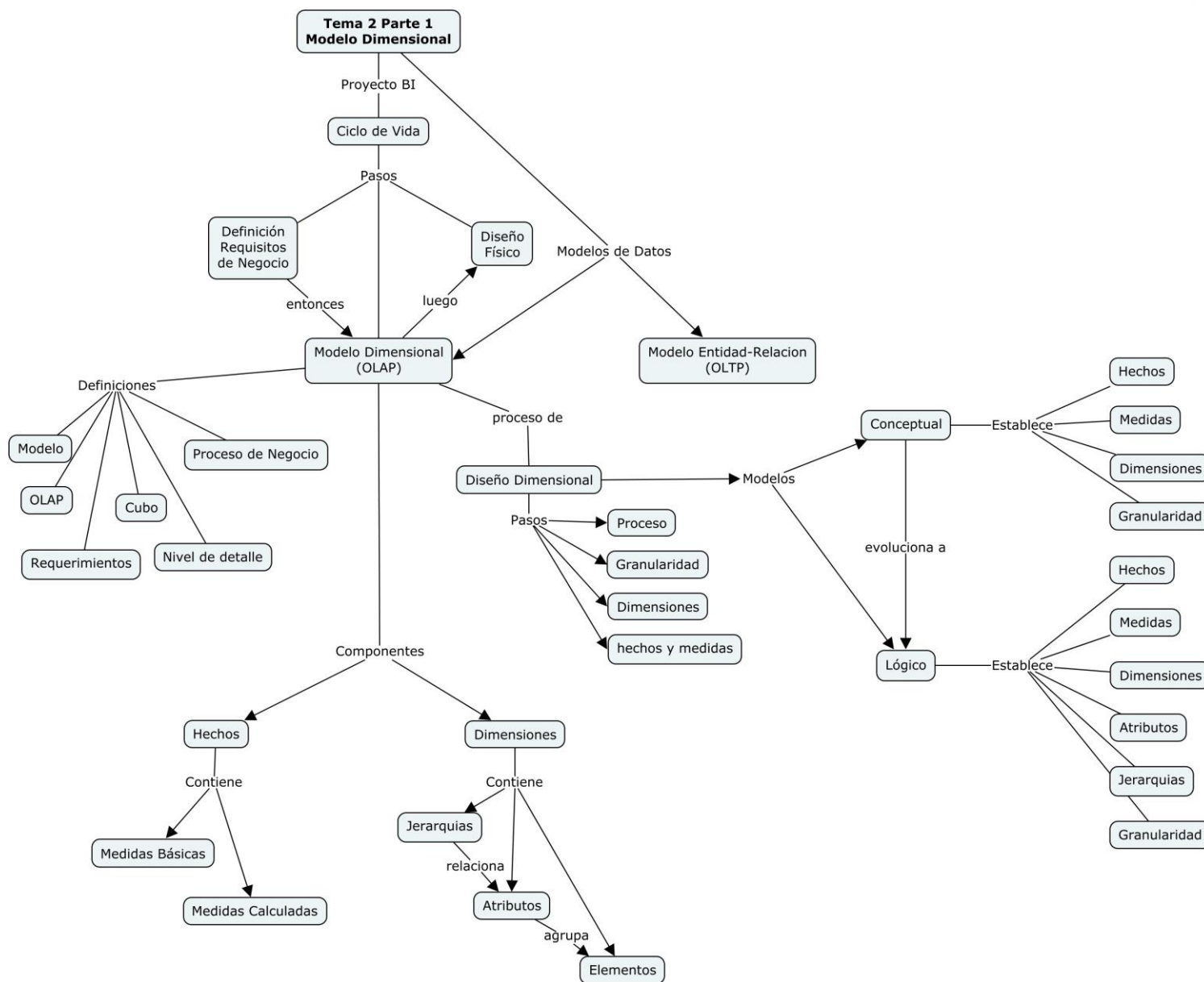
Dimensión Tiempo

Group by MES

SUM

Fecha	SKU	Volumen
05/04/2011	Agua 1.5 L	10
05/04/2011	Agua 2.25 L	35
05/04/2011	Cola 1 L	15
06/05/2011	Cola 1.5 L	10

# RESUMEN TEMA 2 PARTE 1



# TP CASO ESTUDIO: DORSA

## ENTREGA OPCIONAL (\*)



- Listar las posibles dimensiones a partir de las necesidades de información enunciadas.
- Identificar los hechos con las medidas básicas y calculadas.
- Identificar el proceso de negocio a modelar.
- Construir el modelo dimensional conceptual.
- Identificar atributos y jerarquías para cada dimensión.
- Construir el modelo dimensional lógico.

(\*)

- *Los TPs Opcionales sirven para afianzar conceptos necesarios para realizar los TPs de Aplicación, serán corregidos en clase y/o mediante Autoevaluación.*
- *Los TPs de Aplicación tiene una Fecha Límite de Entrega que deberá ser cumplida sin excepción, serán corregidos en detalle por los docentes.*
- *Ver documentos: "Condiciones de Cursada en MIeL - sección: Plazos y condiciones de Entrega Trabajos Prácticos y Casos de Estudio" + "Circuito Entrega TPs Teams/MIeL"*



