

Universidad Autónoma del Estado de Hidalgo

Escuela Superior Huejutla





Área Académica: Sistemas Computacionales

Tema: Arquitectura de un sistema de almacén de datos

Profesor: Mtro Felipe de Jesús Núñez Cárdenas

Periodo: Agosto Noviembre 2011

Keywords

Almacen de Datos, Datawarehouse, Arquitectura





Tema: Arquitectura de un sistema de almacén de datos

Abstract

The construction of a datawarehouse involve determining the dimensions that must have, as well as the attributes that each will have, all within multidimensional architecture

Keywords:

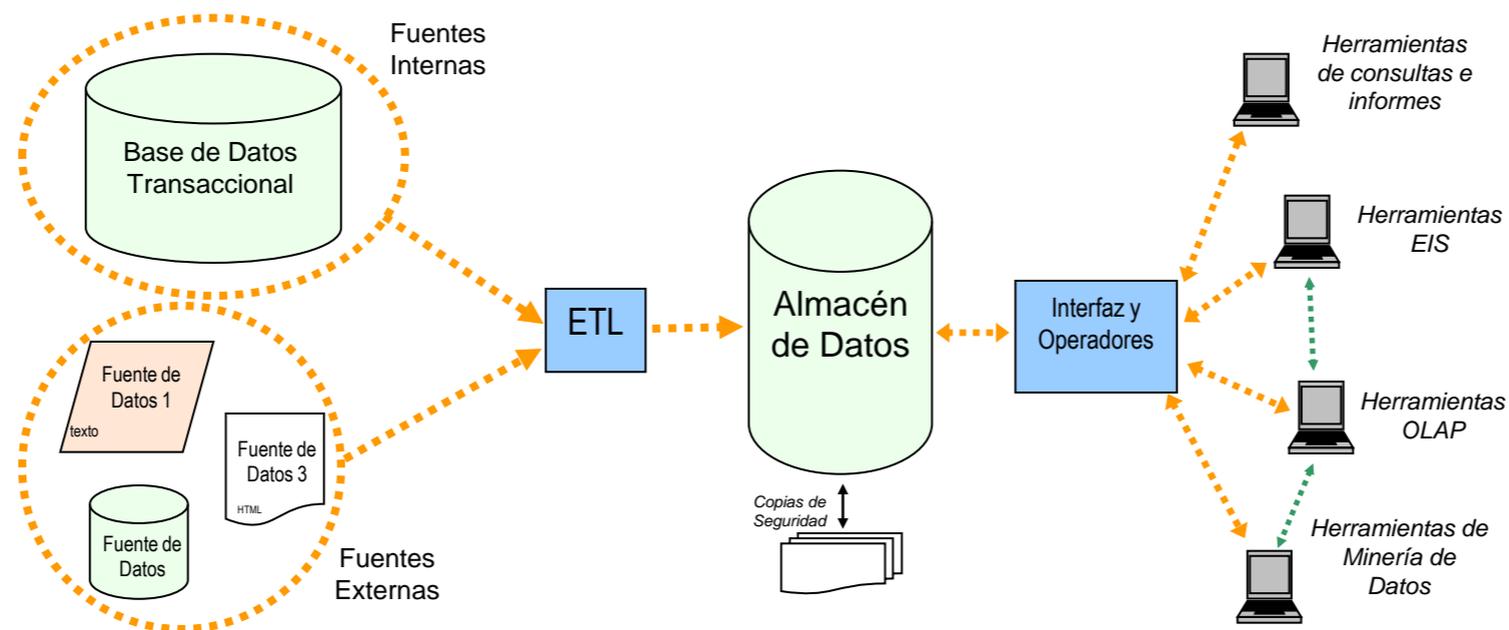
Almacen de Datos, Datawarehouse, Arquitectura





Arquitectura de un Almacén de Datos

- La Arquitectura de un AD viene determinada por su situación central como fuente de información para las herramientas de análisis.





Arquitectura de un Almacén de Datos

- Componentes:
 - Sistema ETL (*Extraction, Transformation, Load*): realiza las funciones de *extracción* de las fuentes de datos (transaccionales o externas), *transformación* (limpieza, consolidación, ...) y la *carga* del AD, realizando:
 - extracción de los datos.
 - filtrado de los datos: limpieza, consolidación, etc.
 - carga inicial del almacén: ordenación, agregaciones, etc.
 - refresco del almacén: operación periódica que propaga los cambios de las fuentes externas al almacén de datos
 - Repositorio Propio de Datos: información relevante, metadatos.
 - Interfaces y Gestores de Consulta: permiten acceder a los datos y sobre ellos se conectan herramientas más sofisticadas (OLAP, EIS, minería de datos).
 - Sistemas de Integridad y Seguridad: se encargan de un mantenimiento global, copias de seguridad, ...





Arquitectura de un Almacén de Datos

- Organización (Externa) de Los Datos...

Las herramientas de explotación de los almacenes de datos han adoptado un **modelo multidimensional de datos**.



Se ofrece al usuario una visión multidimensional de los datos que son objeto de análisis.





Arquitectura de un Almacén de Datos

EJEMPLO

Organización: Cadena de supermercados.

Actividad objeto de análisis: ventas de productos.

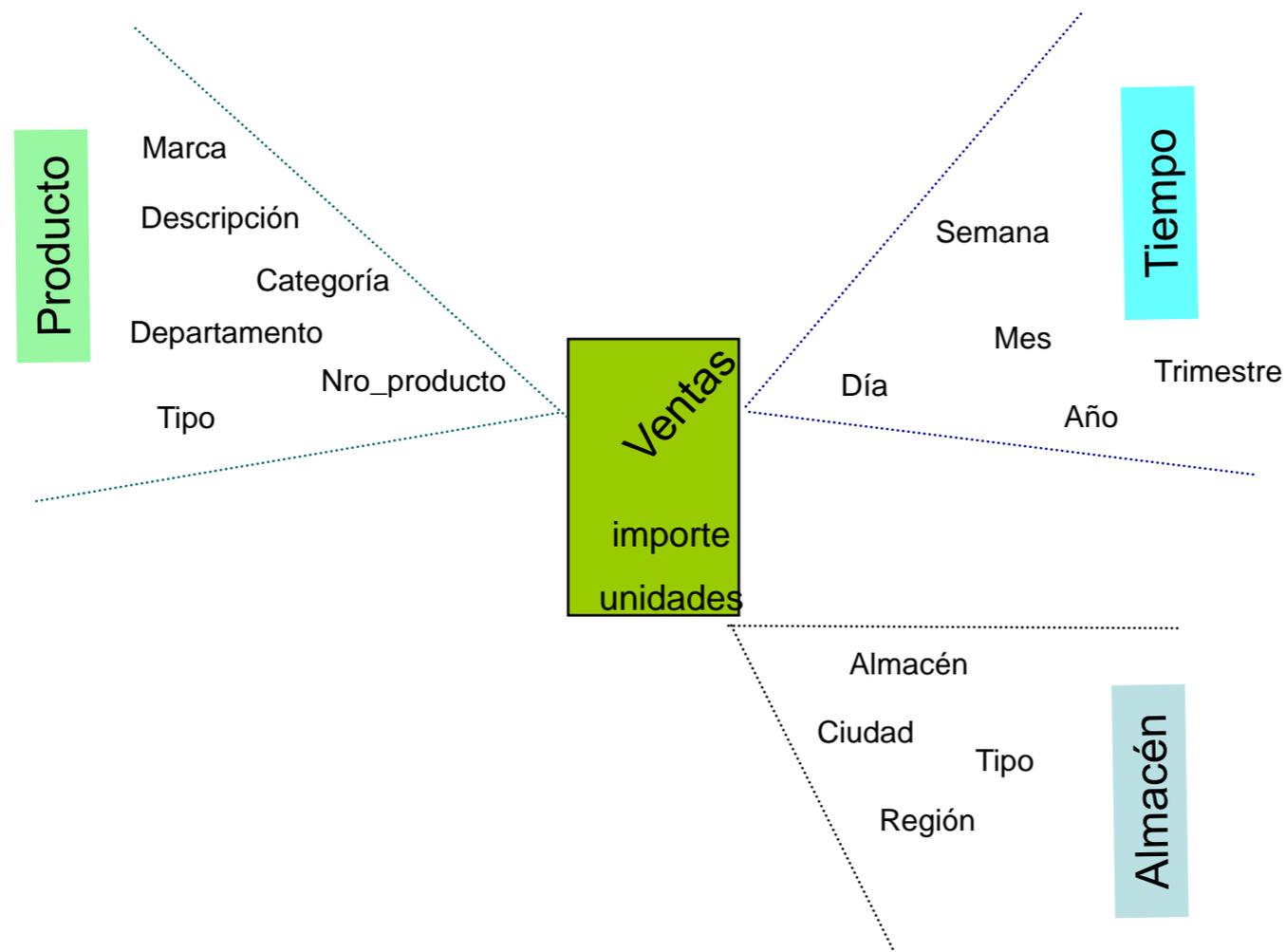
Información registrada sobre una venta: “del **producto** “Tauritón 33cl” se han vendido en el **almacén** “Almacén nro.1” el **día** 17/7/2003, 5 **unidades** por un **importe** de 103,19 euros.”

Para hacer el análisis no interesa la venta individual (ticket) realizada a un cliente sino las ventas diarias de productos en los distintos almacenes de la cadena.



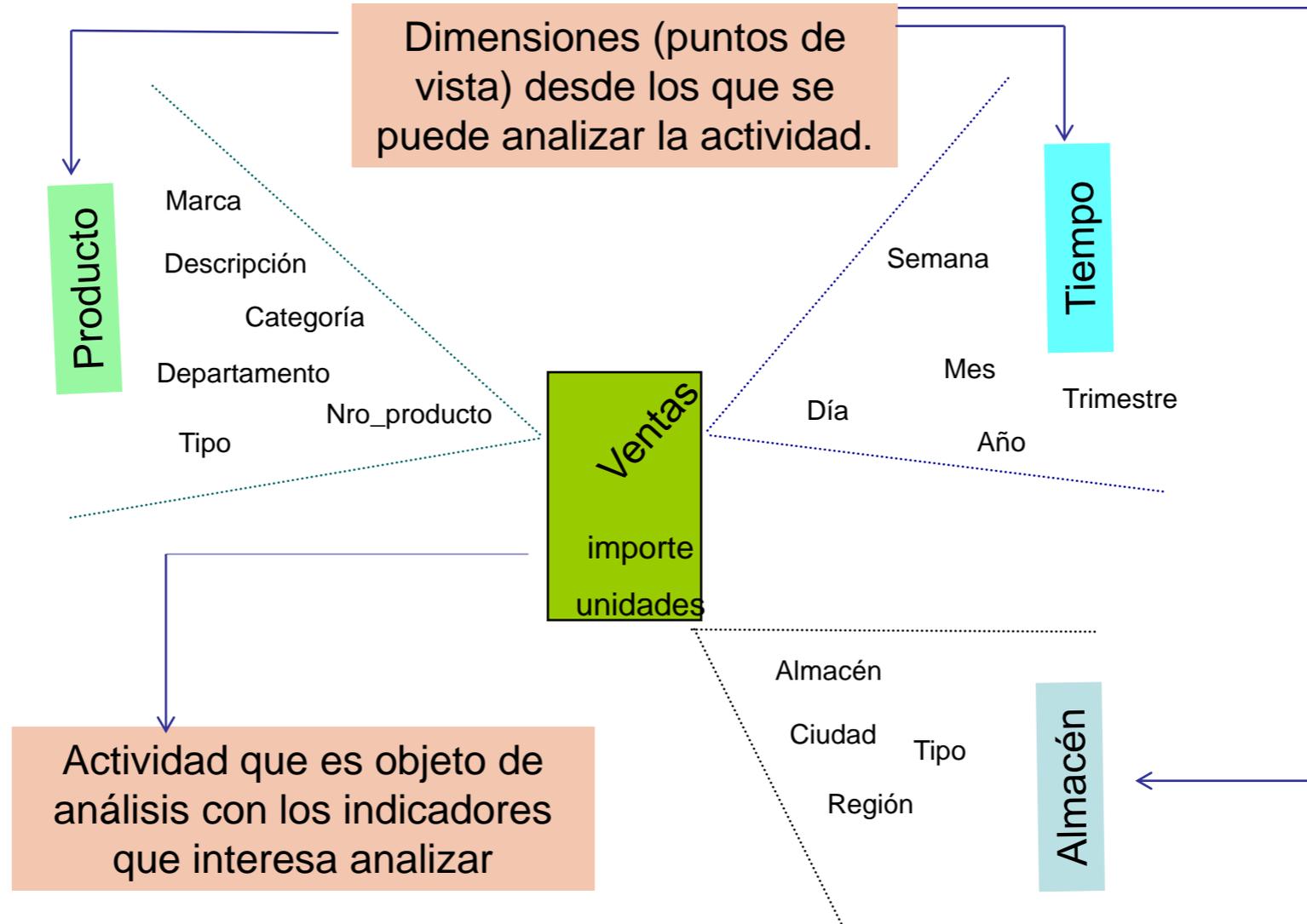


Arquitectura de un Almacén de Datos





Arquitectura de un Almacén de Datos





Arquitectura de un Almacén de Datos

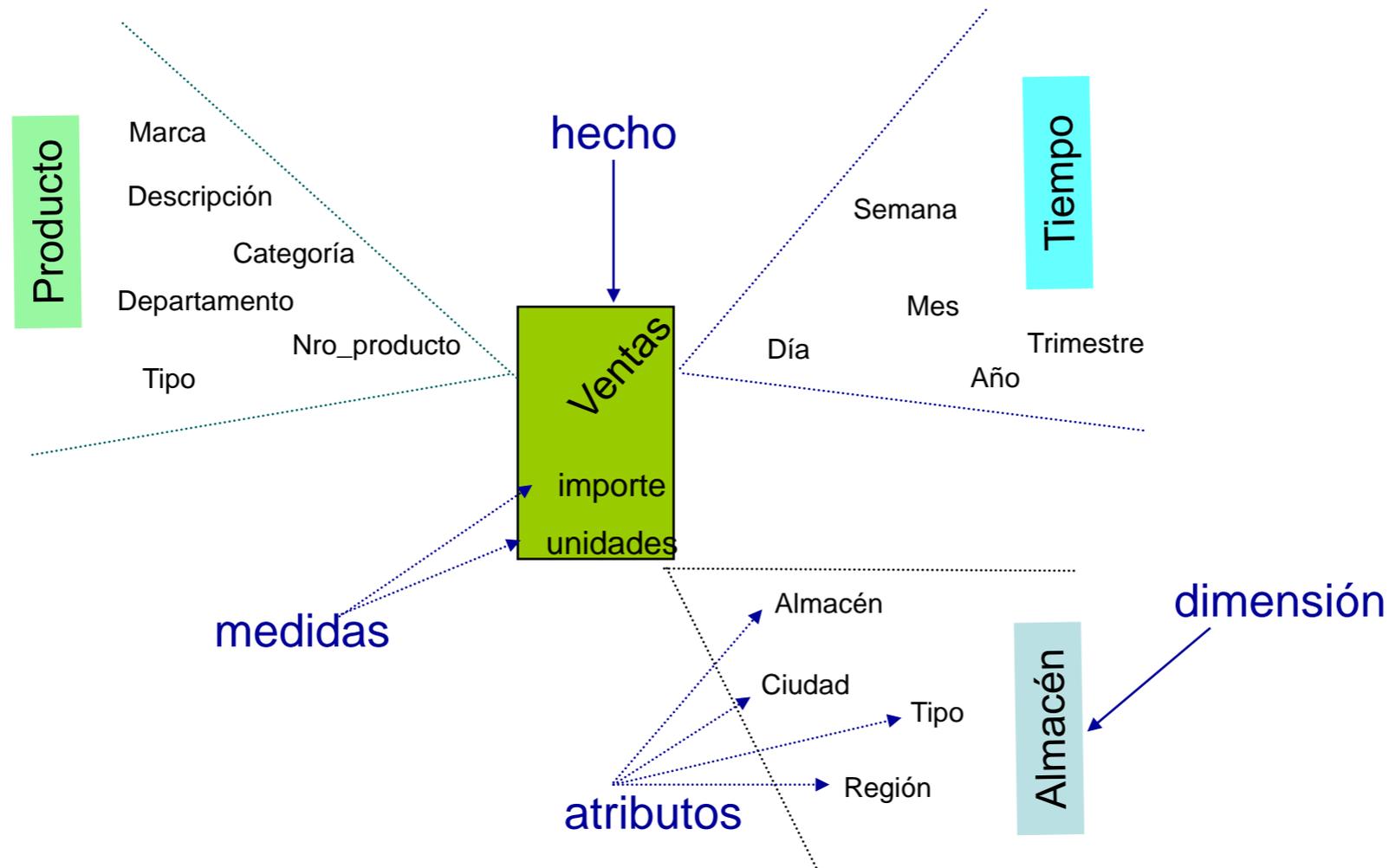
Modelo multidimensional:

- ✓ en un esquema multidimensional se representa una actividad que es objeto de análisis (**hecho**) y las dimensiones que caracterizan la actividad (**dimensiones**).
- ✓ la información relevante sobre el **hecho** (actividad) se representa por un conjunto de indicadores (**medidas o atributos de hecho**).
- ✓ la información descriptiva de cada **dimensión** se representa por un conjunto de atributos (**atributos de dimensión**).





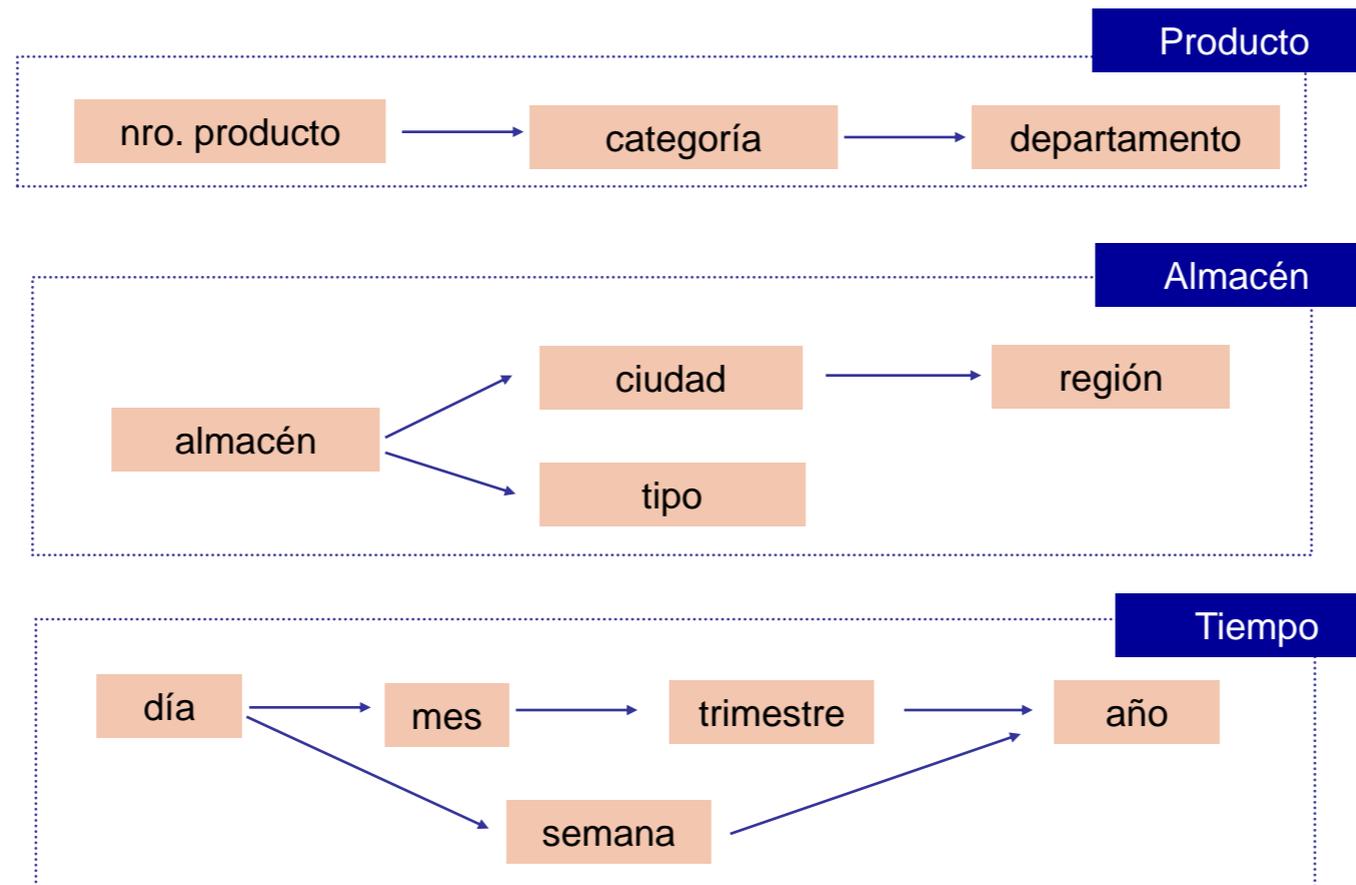
Arquitectura de un Almacén de Datos





Arquitectura de un Almacén de Datos

Entre los atributos de una dimensión se definen **jerarquías**

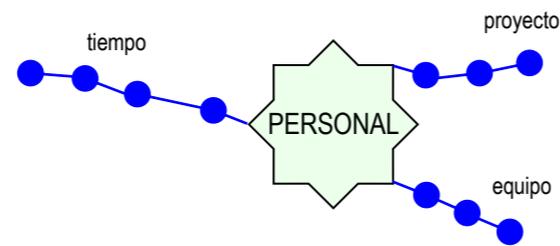




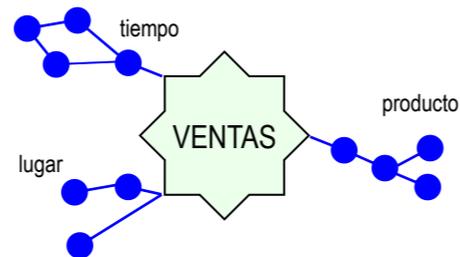
Arquitectura de un Almacén de Datos

Este esquema multidimensional recibe varios nombres:

- estrella: si la jerarquía de dimensiones es lineal



- estrella jerárquica o copo de nieve: si la jerarquía no es lineal.

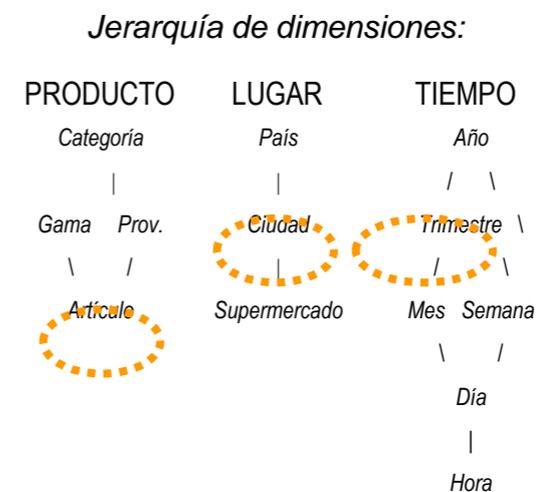
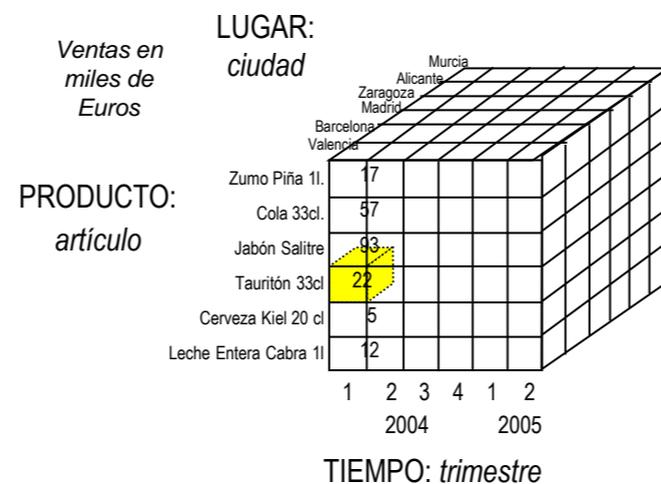




Arquitectura de un Almacén de Datos

- Se pueden obtener hechos a diferentes niveles de agregación:
- obtención de **medidas** sobre los **hechos** parametrizadas por atributos de las **dimensiones** y restringidas por condiciones impuestas sobre las dimensiones

HECHO: “El primer trimestre de 2004 la empresa vendió en Valencia por un **importe** de 22.000 euros del producto tauritón 33 cl.”



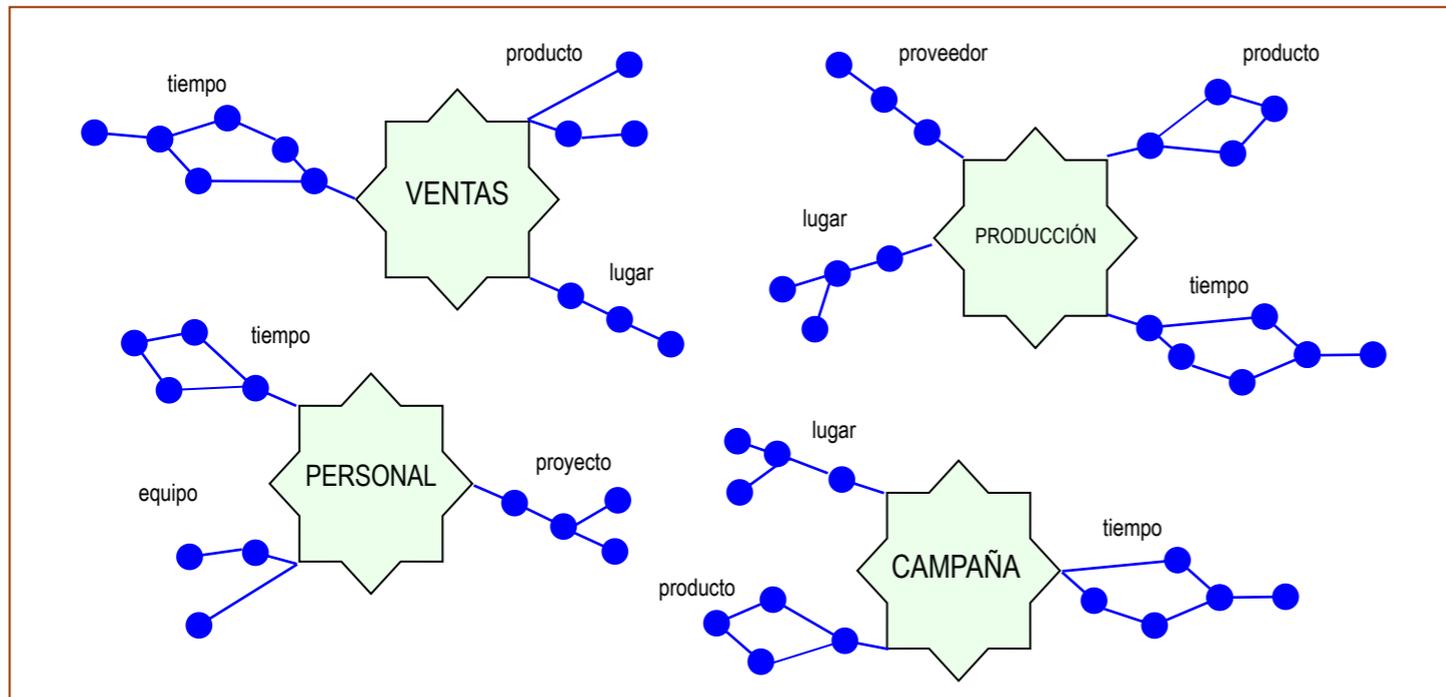
- Un nivel de agregación para un conjunto de dimensiones se denomina cubo.





Arquitectura de un Almacén de Datos

- ¿Se puede recopilar toda la información necesaria en un único esquema estrella o copo de nieve?
- NO : necesidad de varios esquemas.
- Cada uno de estos esquemas se denomina datamart.



*Almacén
formado por
4 datamarts.*





Arquitectura de un Almacén de Datos

- El almacén de datos puede estar formado por varios datamarts y, opcionalmente, por tablas adicionales.

Data mart → subconjunto de un almacén de datos, generalmente en forma de estrella o copo de nieve.

- ✓ se definen para satisfacer las necesidades de un departamento o sección de la organización.
- ✓ contiene menos información de detalle y más información agregada.





Bibliografía

- Hand, D.J.; Mannila, H. and Smyth, P. “Principles of Data Mining”, The MIT Press, 2000.
- Hernández, J.; Ramírez, M.J.; Ferri, C. “Introducción a la Minería de Datos” Pearson Prentice Hall, 2004.
- Kosala, R.; Blockeel, H. “Web Mining Research: A Survey” ACM SIGKDD Explorations, Newsletter of the ACM SIG on Knowledge Discovery and Data Mining, June 2000, Vol. 2, nº1, pp. 1-15.
- Mena, Jesus “Data Mining Your Website”, Digital Press, July 1999.
- Mitchell, T.M. “Machine Learning” McGraw-Hill 1997.
- Pyle, D. “Data Preparation for Data Mining” Morgan Kaufmann, Harcourt Intl., 1999.
- Thuraisingham, B. “Data Mining. Technologies, Techniques, Tools, and Trends”, CRC Press, 1999.
- Witten, I.H.; Frank, E. “Tools for Data Mining”, Morgan Kaufmann, 1999.
- Wong, P. C. “Visual Data Mining”, Special Issue of *IEEE Computer Graphics and Applications*, Sep/ Oct 1999, pp. 20- 46.
- Material extraído del Análisis y Extracción de Conocimiento en Sistemas de Información: Datawarehouse y Datamining de **José Hernández Orallo**, **Universidad Politécnica de Valencia**

