



UNIVERSIDAD
DE LOS ANDES
MÉRIDA VENEZUELA

Datawarehousing

Jose Aguilar
CEMISID, Escuela de Sistemas
Facultad de Ingeniería
Universidad de Los Andes
Mérida, Venezuela

Introducción a Data Warehouse

Analizando la información de una empresa

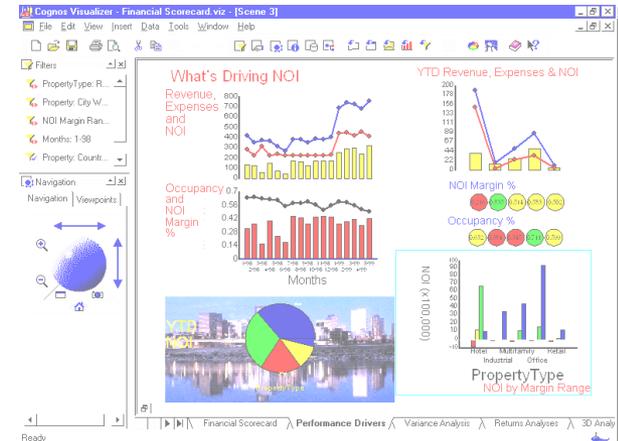
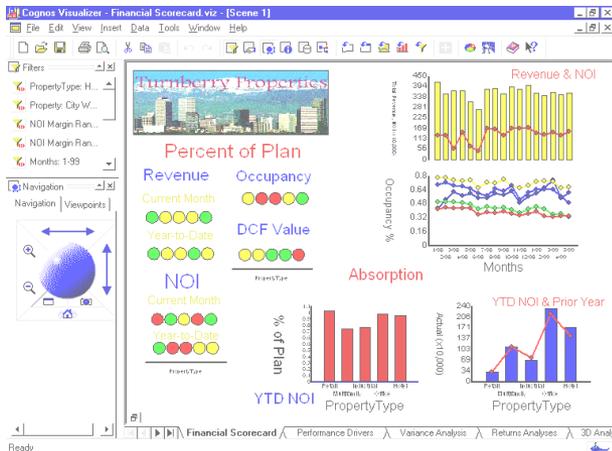
- ✓ Información periódica de las ventas
- ✓ Información del esfuerzo comercial
- ✓ Información sobre los pedidos a los proveedores

Por qué no integrarla y cruzarla para obtener:

- ✓ En qué zonas se está vendiendo más de cada línea de productos?
- ✓ Quienes son los clientes más rentables?
- ✓Cuál es la relación entre el esfuerzo comercial y las operaciones cerradas?
- ✓ De qué proveedores se está comprando la mayor parte de los productos vendidos ?

Introducción a Data Warehouse

- ✓ Se necesita entender no solo **QUE** está pasando, sino **CUANDO, DONDE, QUIEN, COMO Y POR QUE**.
- ✓ Requerimientos de información con **OPORTUNIDAD**.
- ✓ **ESCALAR, ENRIQUECER Y COMPARTIR** a todos los usuarios en la organización



Introducción a Data Warehouse

Ventas

- Número de pedidos
- Productos pedidos
- Clientes que realizaron los pedidos

- Clientes más rentables
- Pedidos más frecuentes
- Productos más rentables
- % de nuevos clientes

Servicio al Cliente

- Datos de llamadas de nuestros clientes
- Información sobre el log de nuestra página web

- Qué clientes visitan nuestra página web ?
- % pedidos realizados por nuestros canales de ventas
- Qué consulta es más frecuente?

Marketing

- Número de campañas realizadas y características de cada una

- % éxito de las campañas
- Qué tipo de clientes han respondido mejor a cada una de las campañas realizadas ?

Distribución

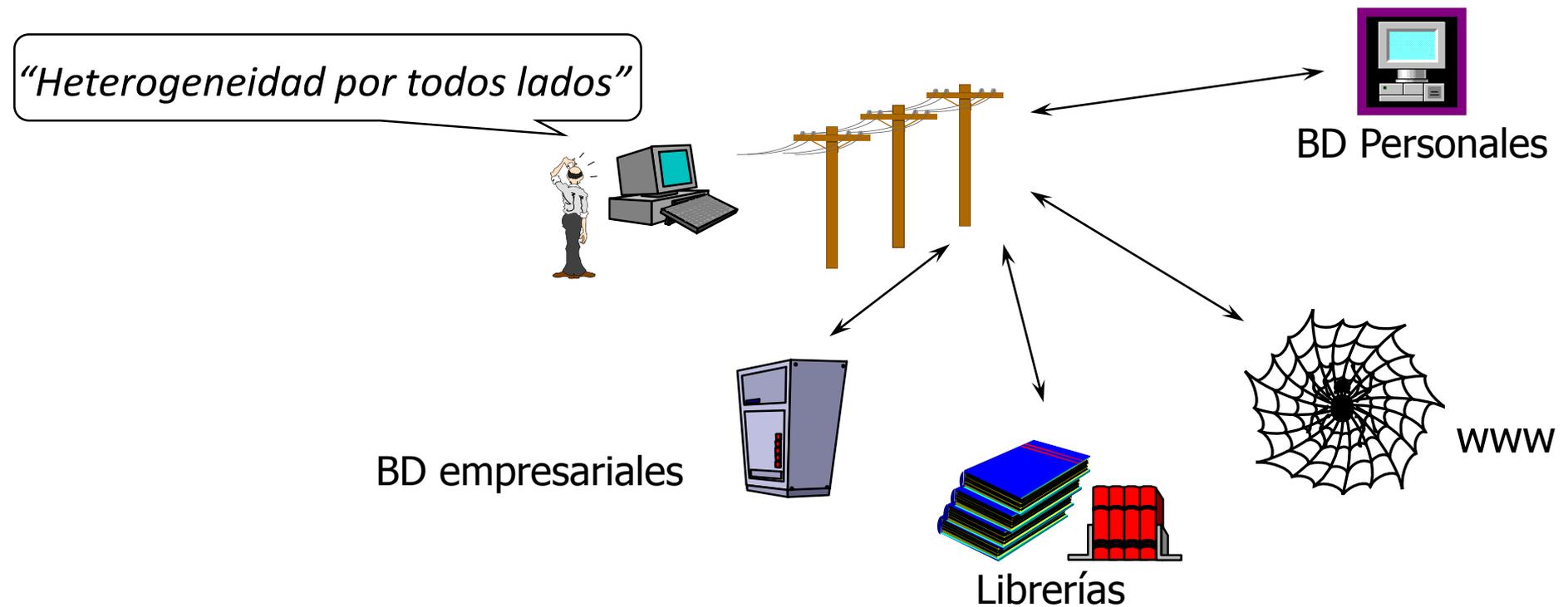
- Productos que salen diariamente del almacén
- Tiempo teórico de entrega

- Número de pedidos retrasados
- Distribuidor que tiene el mayor número de retrasos
- Tiempo medio de entrega

DATO

INFORMACION

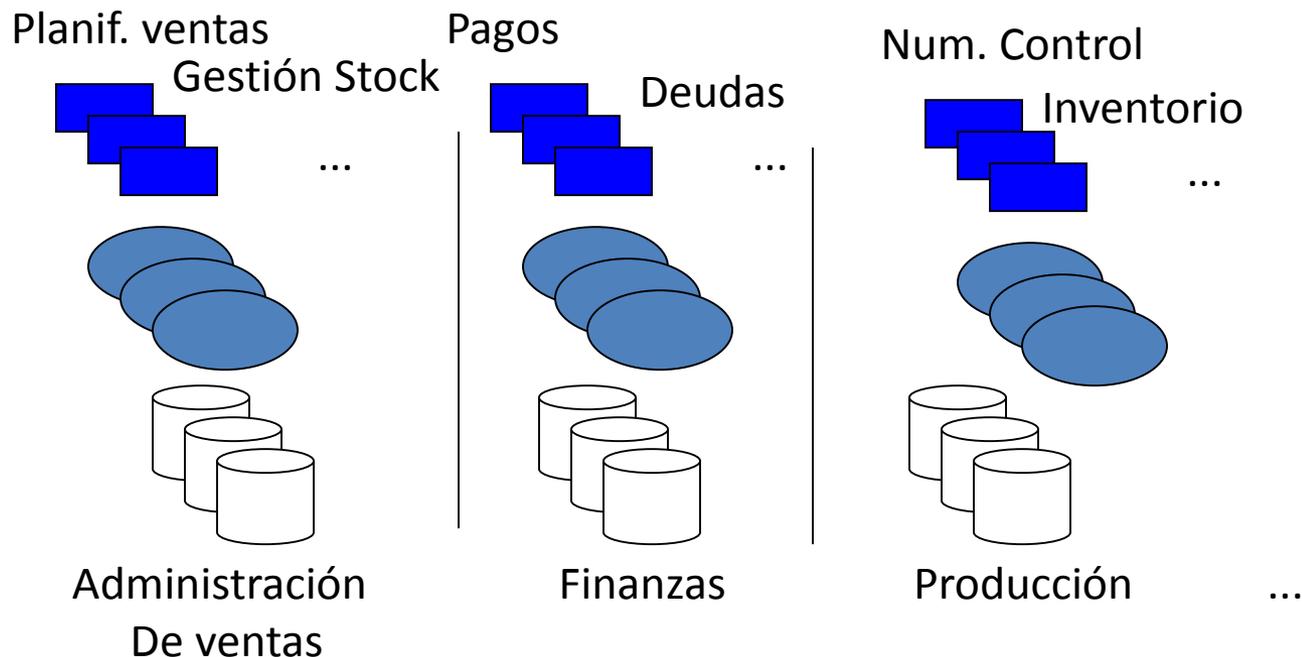
Problemas: Heterogeneidad de las fuentes de Información



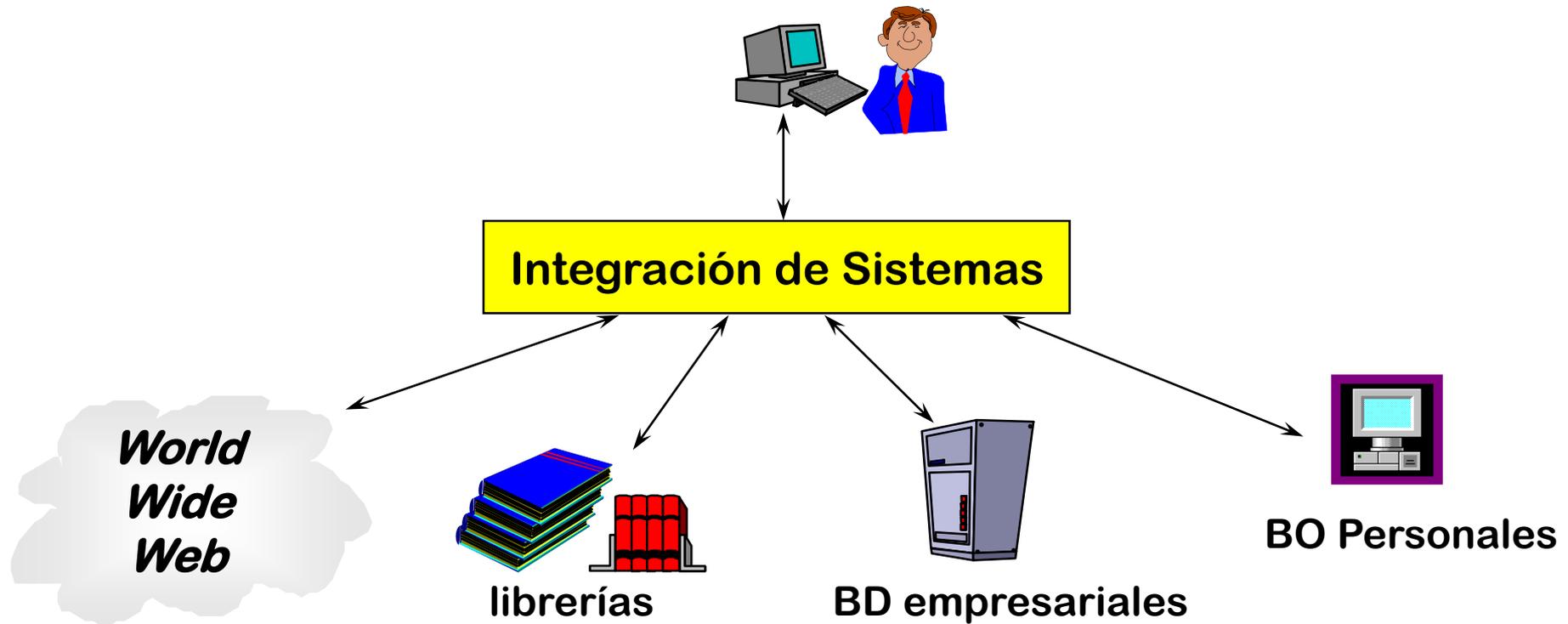
- Diferentes interfaces
- Diferentes representaciones de datos
- Información duplicada e inconsistente

Problema: Gestión de datos en grandes empresas

- Fragmentación vertical de los sistemas de información
- Desarrollo de las aplicaciones guiadas por los sistemas operativos



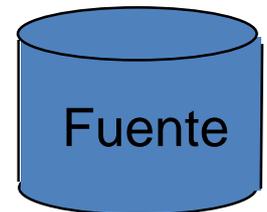
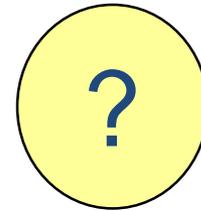
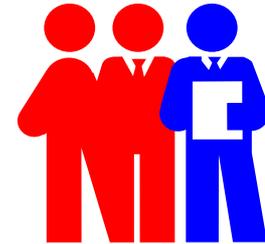
Objetivo: Unificar Acceso a los Datos



- Recopilar y combinar la información
- Proporcionar visión integrada, en una interfaz de usuario uniforme
- Soportar el intercambio

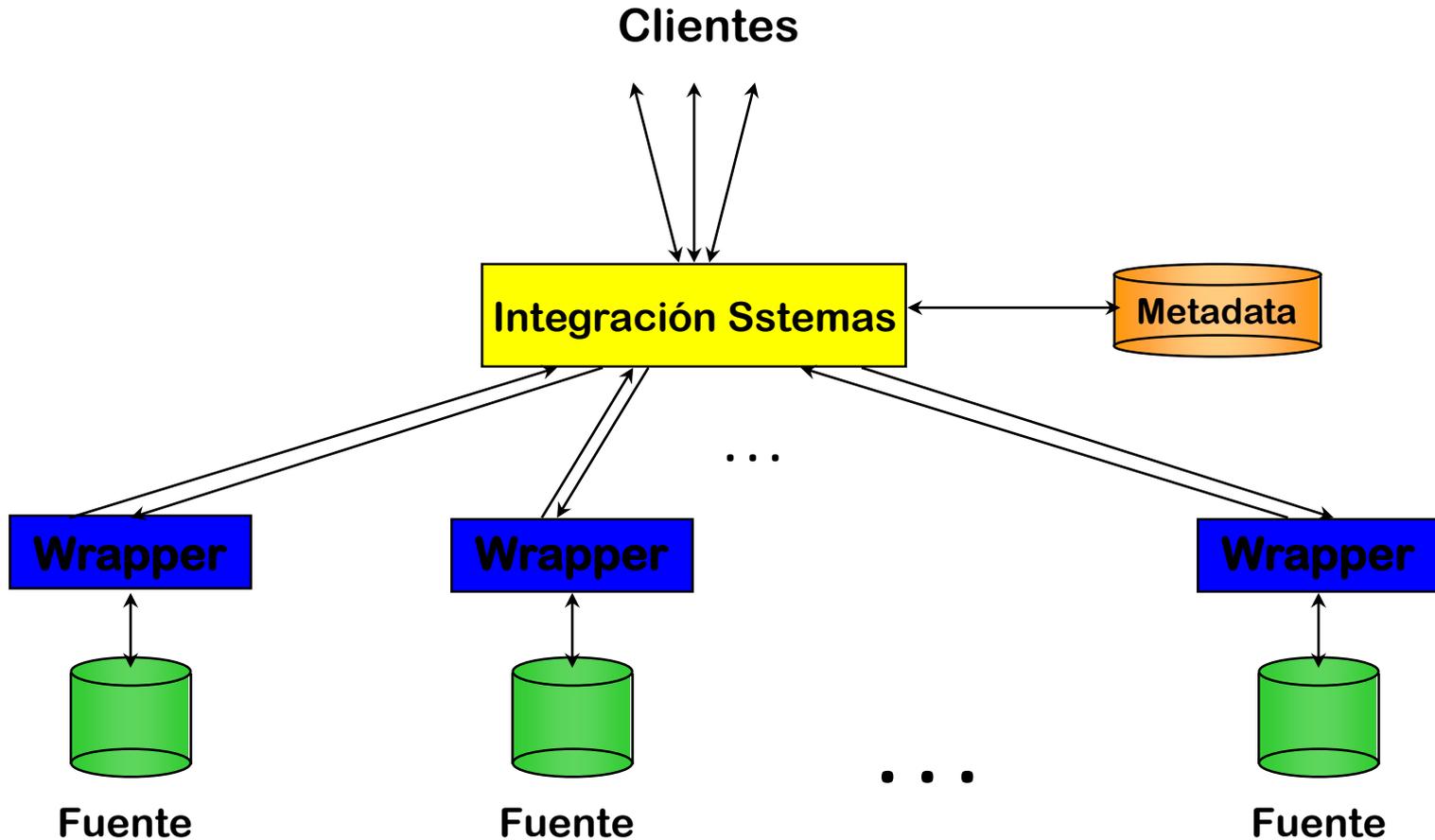
Objetivo: Unificar Acceso a los Datos

- Dos enfoques:
 - Guiado por la consulta (perezoso)
 - Warehouse (ansioso)



Enfoque tradicional

- Guiado por la consulta (perezoso, on-demand)

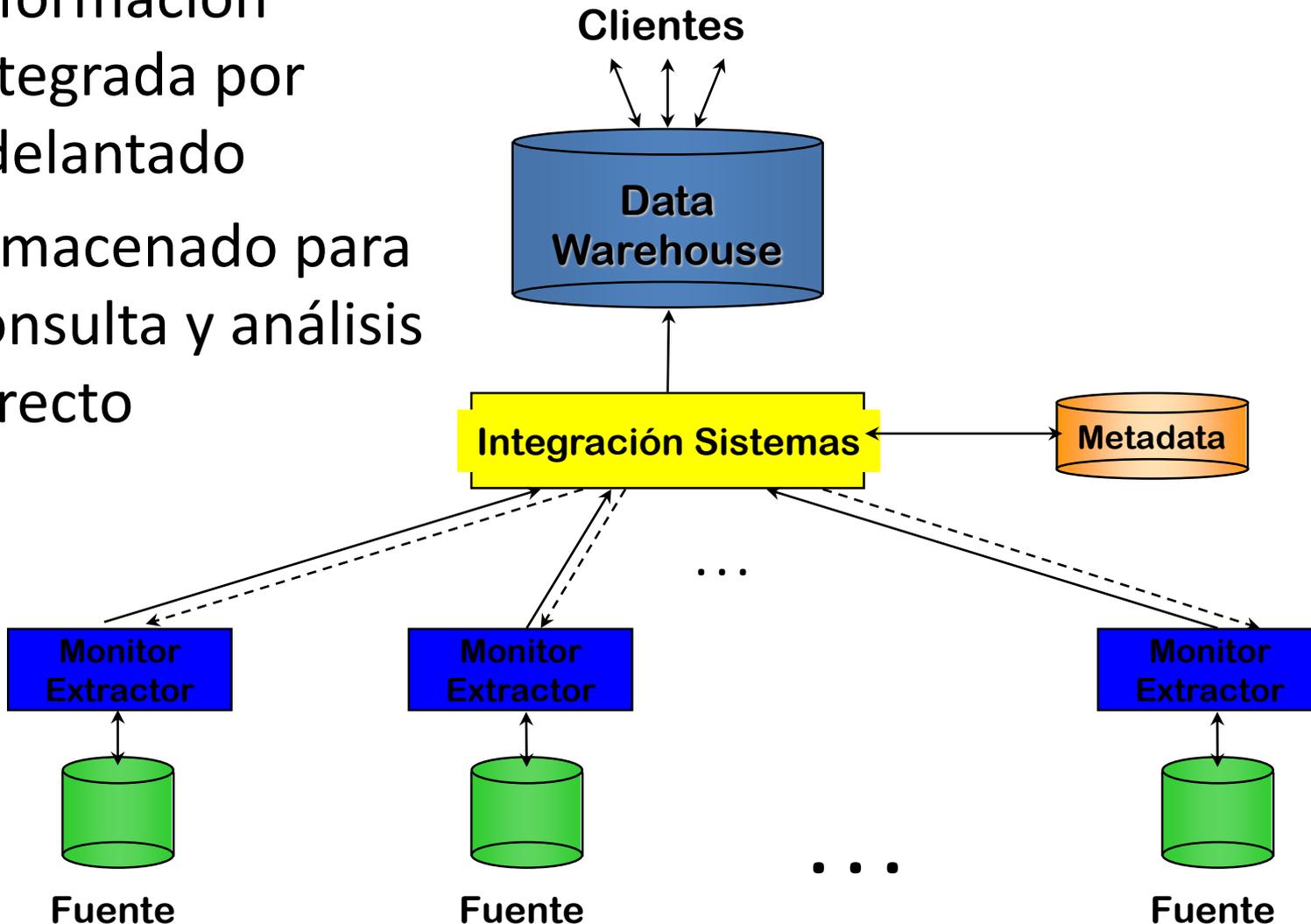


Problema del Enfoque tradicional

- ◆ El retraso en el procesamiento de consultas
- ◆ Fuentes de información lentas o no disponibles
 - ◆ Filtrado e integración son complejas
 - ◆ Ineficiente y potencialmente costoso para las consultas frecuentes
- ◆ Compite con el procesamiento local en el sitio fuente

Enfoque Warehousing

- Información integrada por adelantado
- Almacenado para consulta y análisis directo



Ventaja Enfoque Warehousing

- **Alto rendimiento de la consulta**
 - Pero no necesariamente la información más actualizada
- **No interfiere con el procesamiento local en el sitio origen**
 - Las consultas complejas en warehouse
 - OLTP en las fuentes de información
- **Información copiada en el almacén**
 - Se puede modificar, anotar, resumir, reestructurar, etc.
 - Puede almacenar información histórica
 - Seguridad, sin auditoría
- **Usada en la industria**

Concepto Data Warehousing

*Es un **gran almacén de datos** para consultar

*Es un **repositorio de datos** de muy fácil acceso, alimentado de numerosas fuentes, transformadas en grupos de información sobre temas específicos de negocios, para permitir nuevas consultas, análisis, toma de decisiones.

Se trata, de **un expediente completo de una organización**, más allá de la información transaccional y operacional, almacenado en una base de datos diseñada para favorecer el análisis y la divulgación eficiente de datos.

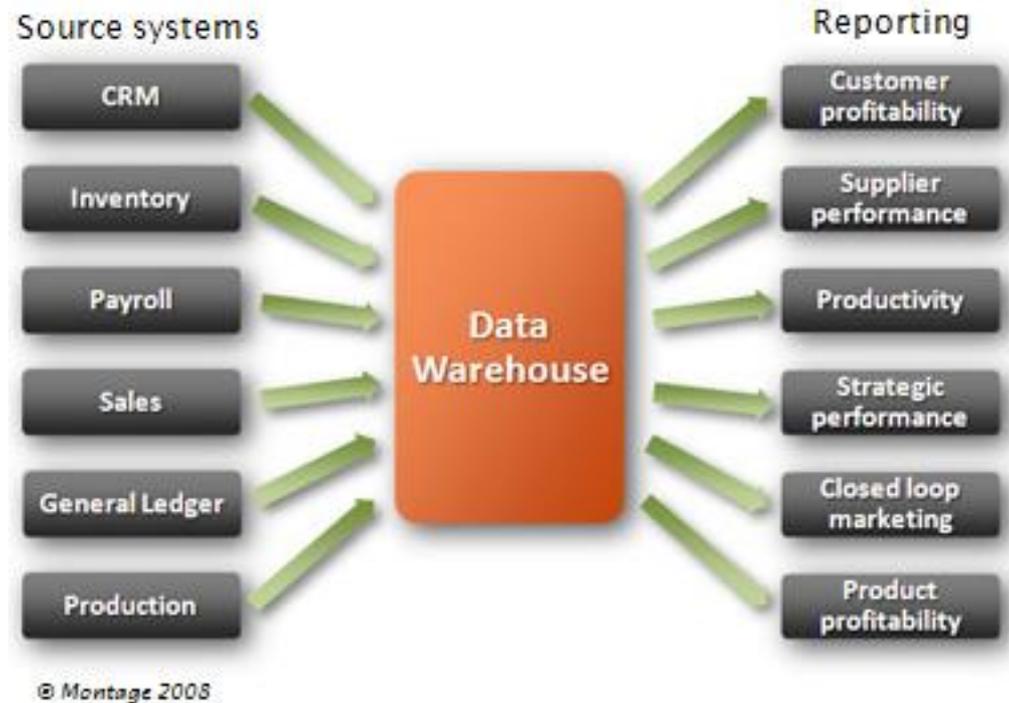
*Tiene **gran capacidad de almacenamiento**, pues los datos pueden ser de grandes periodos de tiempo.



Concepto Data Warehousing

Basado en el **concepto de Metadatos** (datos que describen otros datos). Por ejemplo, en una biblioteca se usan fichas que especifican autores, títulos, casas editoriales y lugares para buscar libros. Así, los metadatos ayudan a ubicar datos.

Cumple el principio de arquitectura fundamental que es **separar los sistemas transaccionales de los informacionales** en dos entornos de manera que el **análisis de los datos existentes no interfiera** con el procesamiento y registro de nuevos datos.



Concepto Data Warehousing (Bodegas de Datos)

- Datawarehouse (Almacén de datos) se define como una colección de datos que verifican las siguientes propiedades:
 - **Está orientado a un tema**
 - **Datos integrados**
 - **No volátiles**
 - **Variante en el tiempo**

que surgieron como una herramienta de soporte para la toma de decisiones a nivel gerencial

Orientado hacia temas



Los datos se almacenan y agrupan por temas de interés

Se agrupa por temas orientados a la organización, tales como :

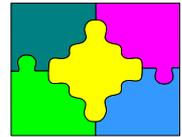
- Clientes
- Productos
- venta

en lugar de las transacciones individuales.

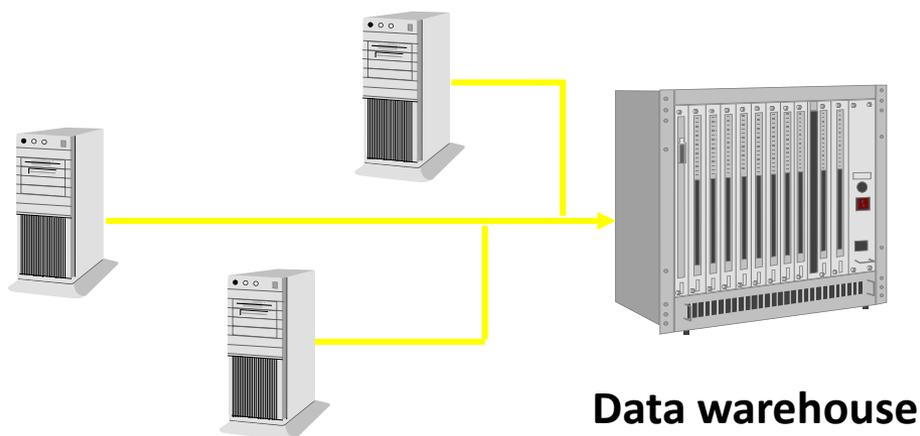
La normalización no es relevante

- **Centrándose en el modelado y análisis de datos** para la toma de decisiones, no en las operaciones diarias o el procesamiento de transacciones.
- Proporcionar una **visión sencilla y concisa en torno a cuestiones temáticas particulares** mediante la exclusión de los datos que no son útiles en el proceso de apoyo a la decisión.

Integración de Datos



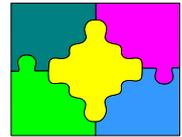
- El almacén de datos **integra datos** que provienen de varias fuentes. Parte de una base de datos operacional y mediante un proceso de **carga de datos** hace el Datawarehouse.
- El proceso de carga es lo más complicado por problemas de codificación, preparación de los datos, ...



Sistemas Operacionales

Los datos en el almacén deben ser:

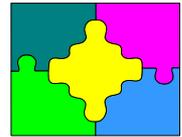
- Limpios
- Validados
- Adecuadamente integrados



Integración de Datos

- Integración de **múltiples fuentes de datos heterogéneas**,
 - bases de datos relacionales, archivos planos, registros de transacciones en línea
- Se aplican **técnicas de limpieza y de integración de datos**.
- Garantiza la **coherencia** en los convenios de denominación, estructuras de codificación, medidas, etc., entre las diferentes fuentes de datos
 - Por ejemplo, el precio del hotel: moneda, impuestos, desayuno, etc.
- Cuando los datos se mueven a la bodega, **se convierten** al estándar usado.

Ejemplo Integración de Datos



Sistema de Cuenta de cheques

Jane Doe (name)
Female (gender)
Bounced check #145 on 1/5/95
Opened account 1994

← *Datos operacionales*

Sistema de cuentas de ahorro

Jane Doe
F (gender)
Opened account 1992

cliente

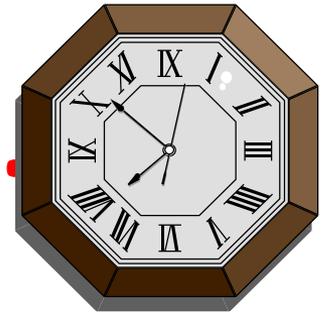
Jane Doe
Female
Bounced check #145
Married
Owns 25 Shares Exxon
Customer since 1992

Sistema de inversiones de clientes

Jane Doe
Owns 25 Shares Exxon
Opened account 1995

↑ *data warehouse*

... variante en el tiempo ...



- Todos los datos en el almacén de datos tienen una **marca de tiempo** en el momento de entrada en el almacén o cuando se usan en el almacén.
- Esta grabación cronológica de datos ofrece posibilidades **históricas y análisis de tendencias**.
- Normalmente, en las BD operativas **los datos se sobrescribe**, ya que los valores anteriores no son de interés.

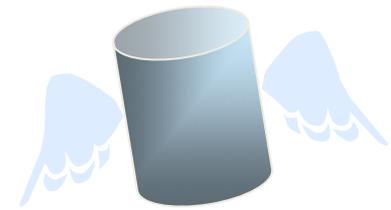
TIEMPO

id_tiempo

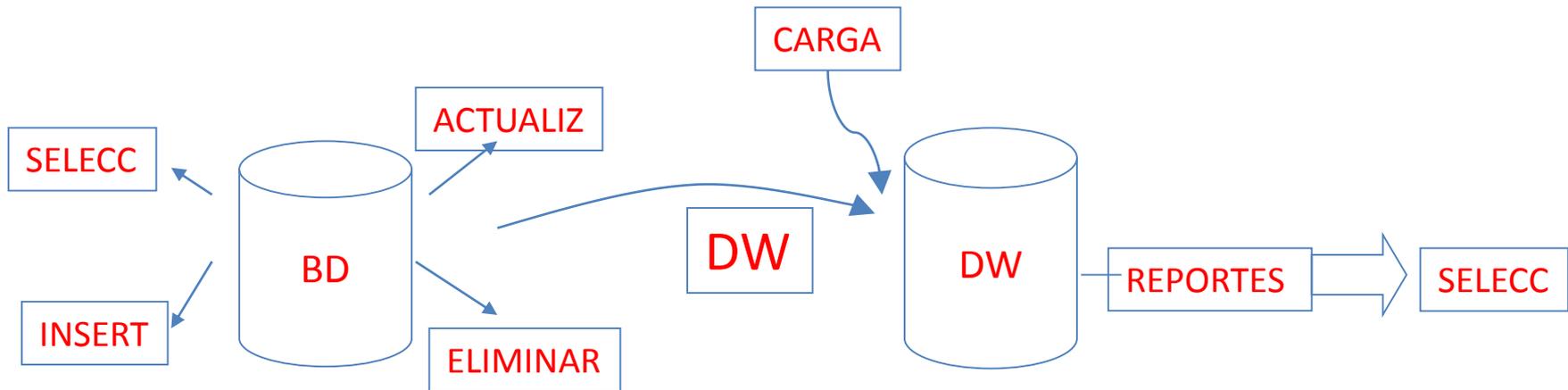
* periodo

porque se hacen cargas de datos continuamente, cuando los datos van cambiando, se actualizan los históricos y se guardan en ficheros temporales.

No volátiles



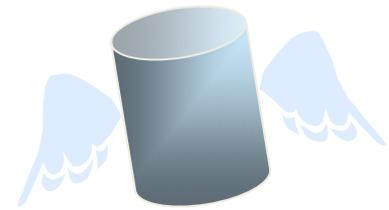
Son estables, una vez almacenados los datos **no se modifican.**



Datos actúan como un recurso estable para informes coherentes y análisis comparativo.

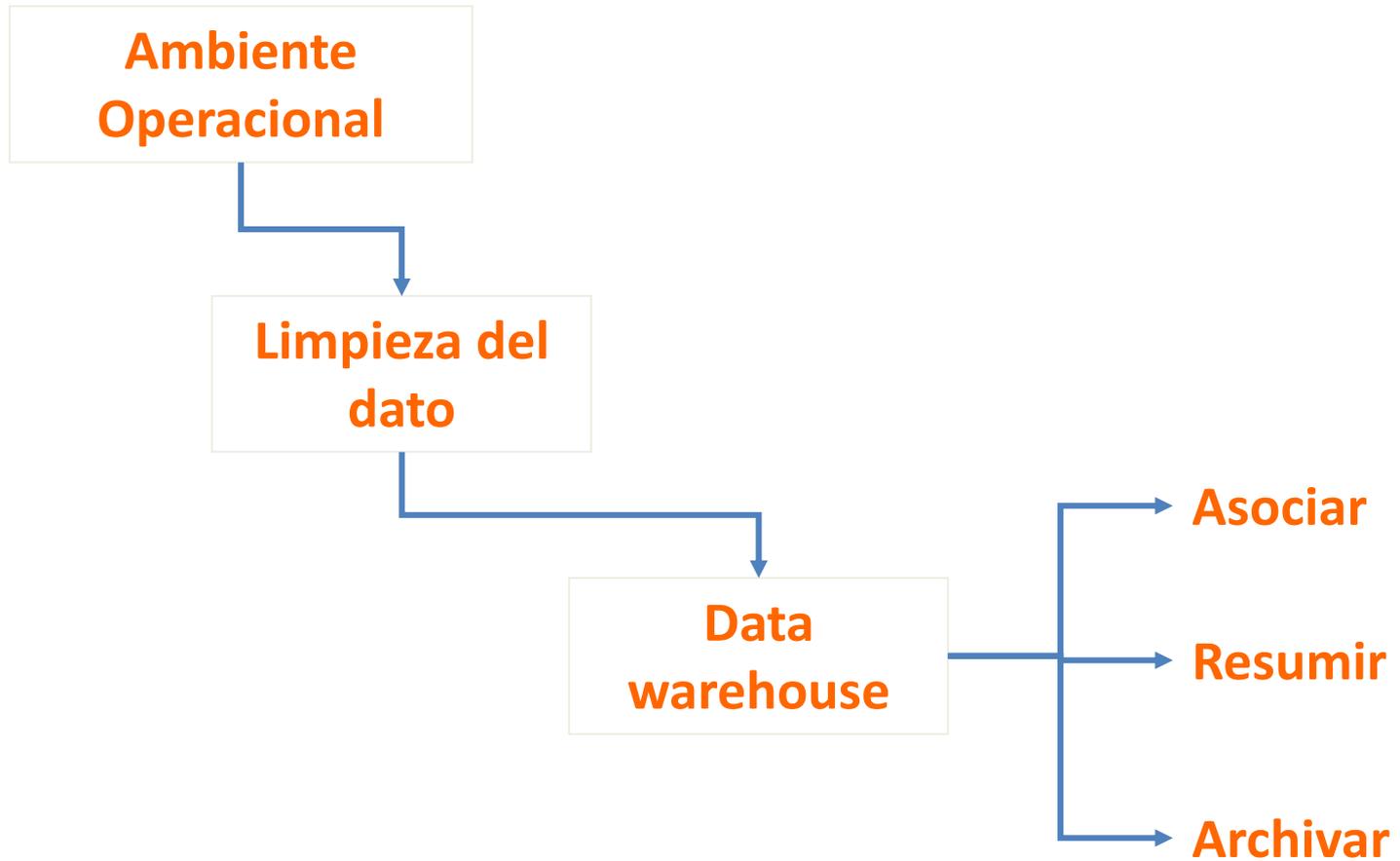
Por el contrario, los datos operativos se actualizan (insertar, eliminar, modificar, etc.)

No volátiles



- **Físicamente separado** de los datos del entorno operacional.
- **Actualización de los datos no se produce** en el entorno de almacén de datos.
- **No requiere el procesamiento de transacciones**, ni mecanismos de recuperación y control de concurrencia
 - Requiere sólo dos operaciones en el Acceso a los datos:
carga inicial de datos y el acceso a los datos.

Flujo del Dato



Funcionamiento de Data warehouse

Tres funciones esenciales:

1. Recopilación de los datos desde Bases de Datos
2. Gestión de los datos en el almacenamiento
3. Análisis de datos para toma de decisiones

Sistemas de
Soporte a
Decisiones
(DSS)



herramientas para
hacer consultas o
informes



Sistemas de
información
ejecutiva (EIS)



Elementos que integran un almacén de datos

Metadatos

"datos acerca de los datos".

Su función es recoger la descripción de la estructura del almacén de datos:

Tablas

Columnas en tablas

Jerarquías y Dimensiones de datos

Relaciones entre tablas

Entidades y Relaciones

Elementos que integran un almacén de datos

Metadato también cuenta con

- **Meta-datos de los datos Operacionales**
 - Historia de los datos (migraciones, transformaciones),
 - Estado de los datos (activo, archivados, o limpiado),
 - Información de seguimiento (estadísticas de uso de almacén, informes de errores, trazas de auditoría)
- **Los algoritmos utilizados para resumir**
- **El mapeo del entorno operativo al almacén de datos**
- **Los datos relacionados con el rendimiento del sistema**
- ...

Elementos que integran un almacén de datos

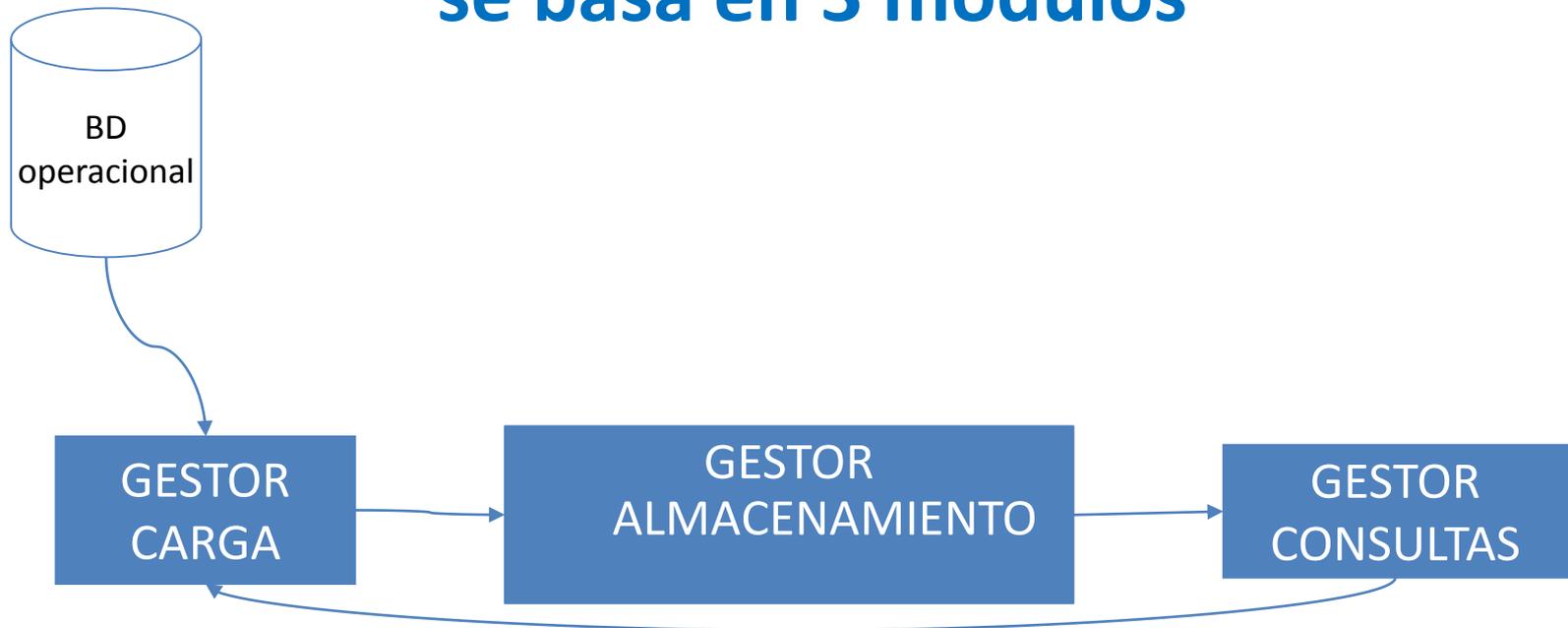
Middleware

- Ofrece servicios a aplicaciones distribuidas que requieren usar el DW.
- Su función es asegurar la conectividad entre todos los componentes de la arquitectura de un almacén de datos (datawarehouse) para brindar servicios.

Funciones ETL, OLAP

ARQUITECTURA PLANA DW

se basa en 3 módulos



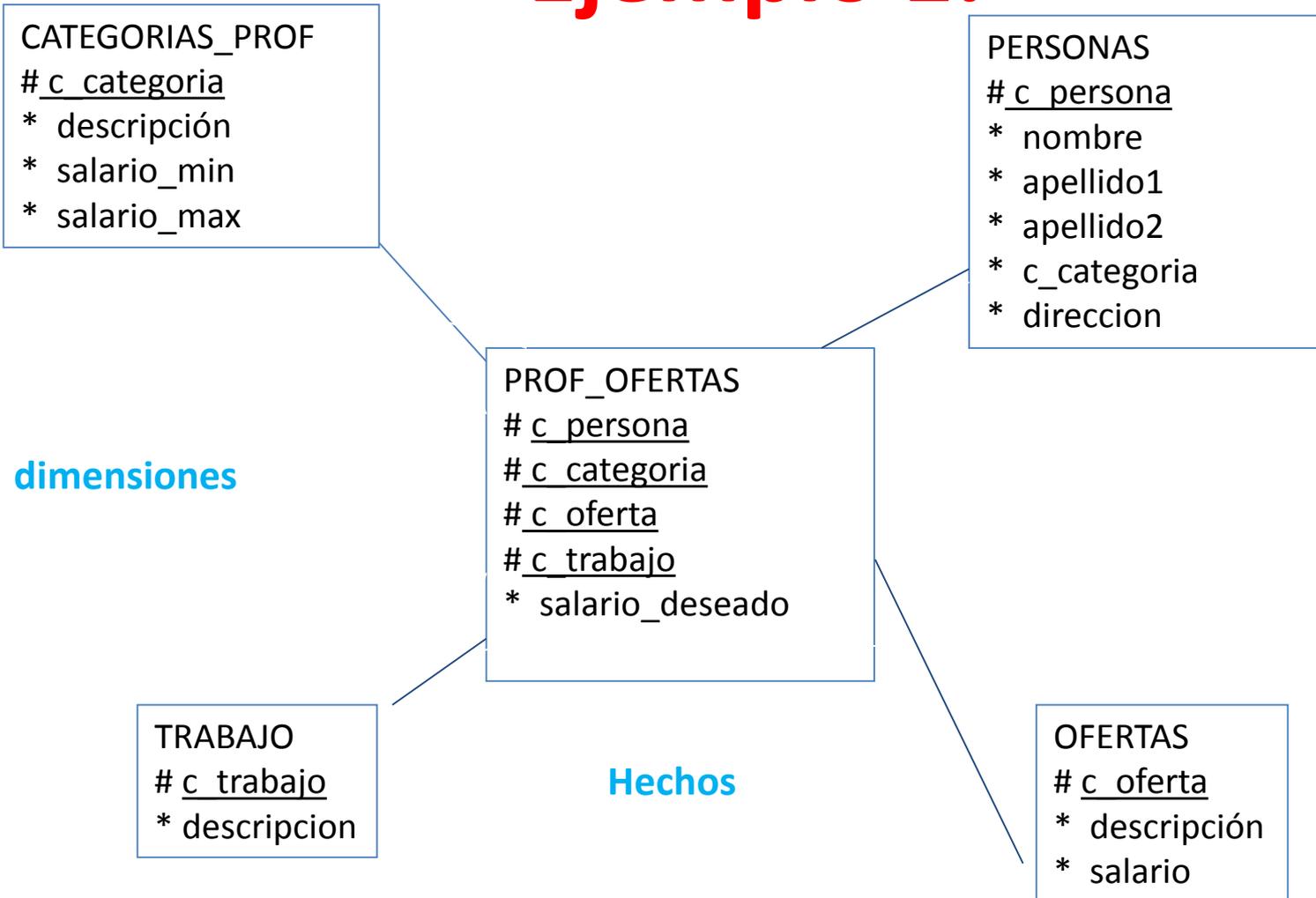
GESTOR DE CARGA

- Permite hacer la carga.
- Dificultades:
 - La integración de los datos
 - La elección del momento de la carga
 - Minimizar el tiempo de carga
 - Buen diccionario de datos o METADATA (para evitar cometer errores en la carga)
 - Diseño de procedimientos SQL

GESTOR DE ALMACENAMIENTO

- Se encarga del almacenamiento, de la estructura,....
- Existe una tabla llamada FACT (Hechos) y unas tablas llamadas dimensiones o tablas dimensionales.
 - Entre la tabla FACT y las tablas dimensionales suele haber relaciones 1:N
 - Este modelo tiene forma de estrella, por eso se denomina MODELO ESTRELLA

Ejemplo 1:



GESTOR DE CONSULTAS

- Las consultas se hacen sobre la tabla FACT.
- También define perfiles de consultas, pues ellas son diferentes dependiendo del usuario y sus necesidades.

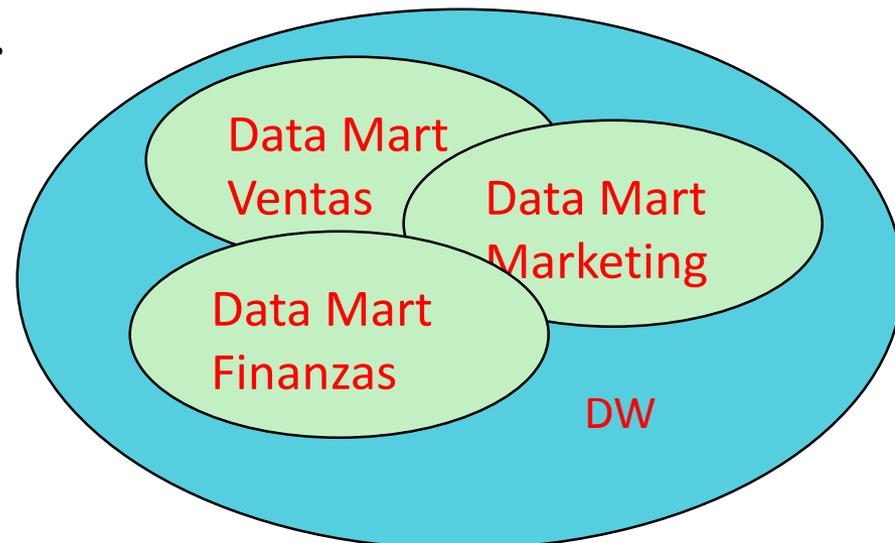
Data Mart

Las grandes cantidades de datos en el Data Warehouse a veces se subdividen en unidades lógicas más pequeñas (**data marts**)

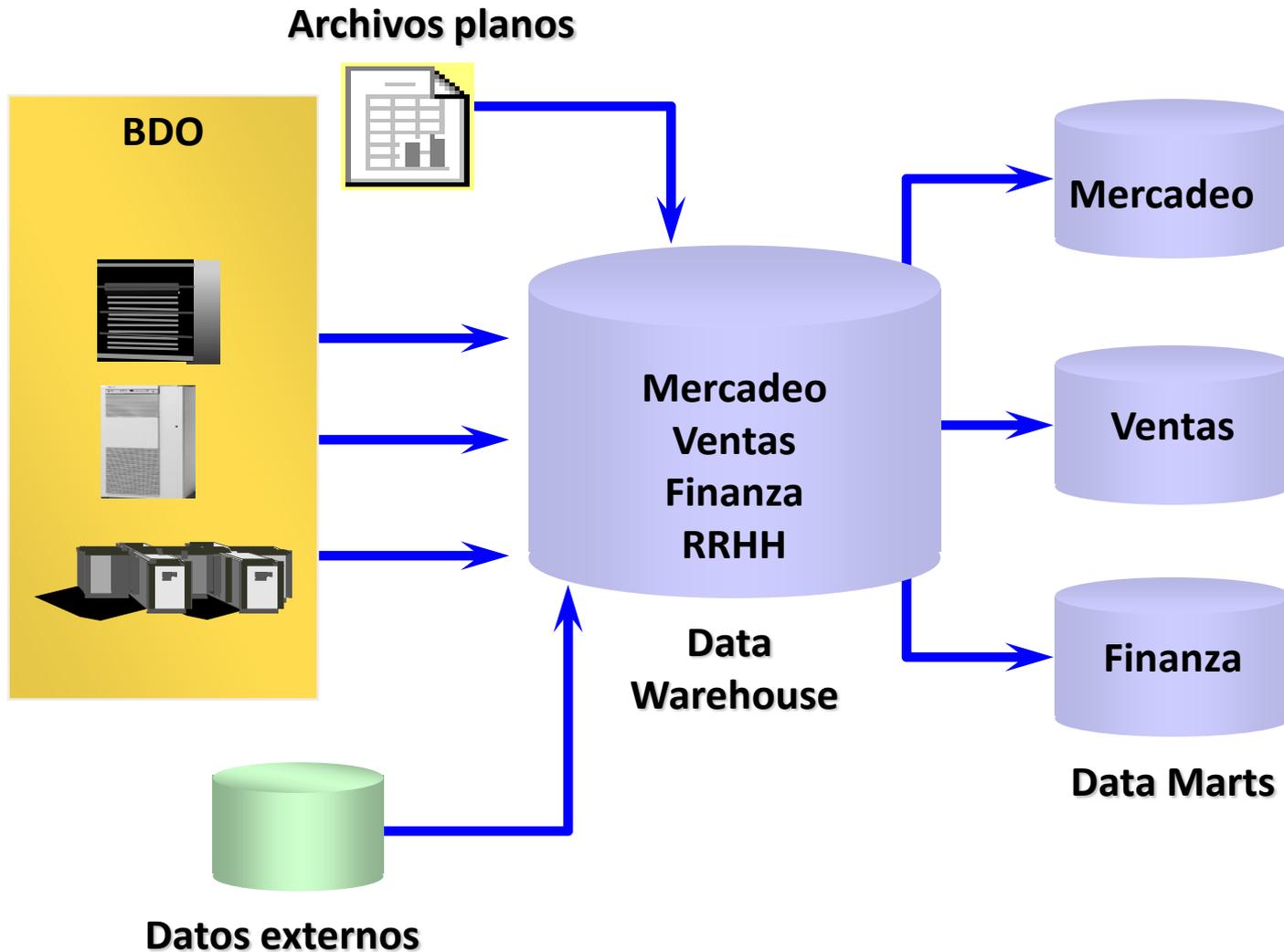


Data Mart

- Una versión reducida de menor costo que el data warehouse, diseñado para una **unidad de negocios estratégica o departamento**.
- Entre las **características** de un data mart destacan:
 - Usuarios limitados.
 - Área específica.
 - Tiene un propósito específico.
 - Tiene una función de apoyo.

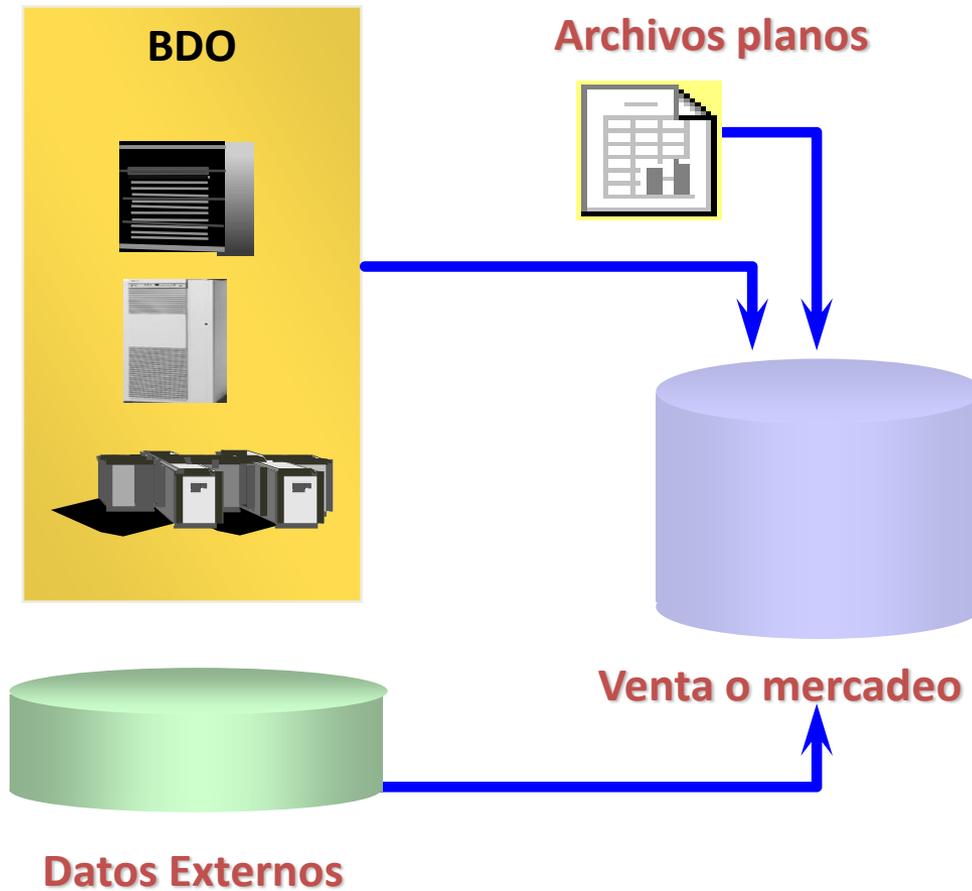


Tipo Data Mart



Dependiente de DW

Tipo Data Mart



Independiente de DW

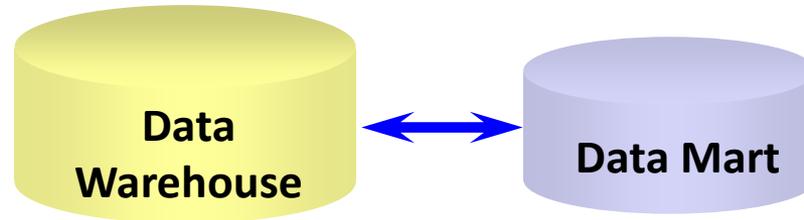
Razones para la creación de un Data Mart

- Para dar a los usuarios un **acceso más flexible** a los datos que necesitan para analizar con más frecuencia.
- Para proporcionar datos en una forma que coincide con el **punto de vista** colectivo de un grupo de usuarios
- Para mejorar el **tiempo de respuesta** del usuario final.
- Los usuarios potenciales de un data mart están **claramente definidos**

Razones para la creación de un Data Mart

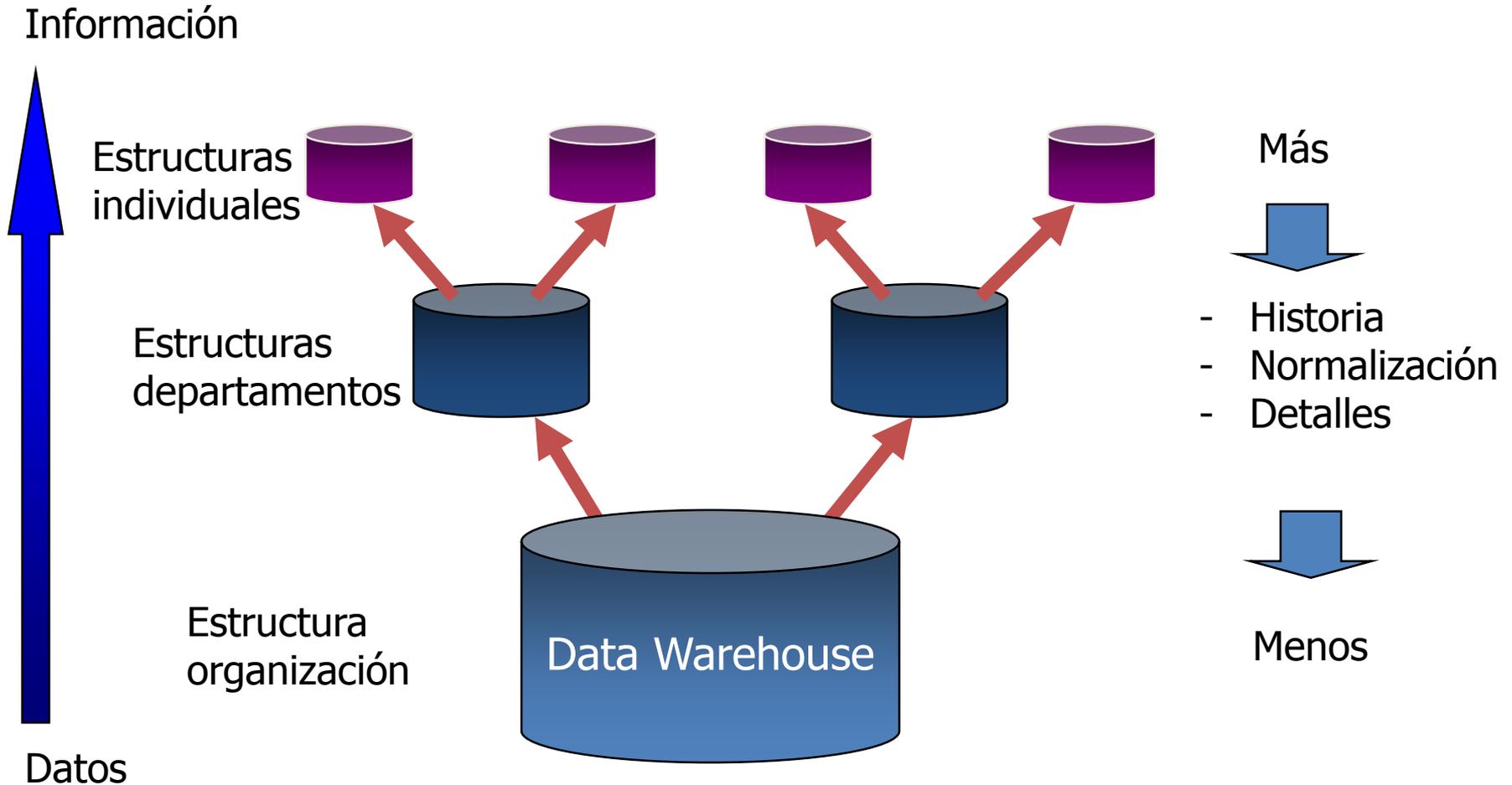
- Proporcionar datos estructurados según las **exigencias de las herramientas** de acceso de los usuarios finales.
- La construcción de un data mart **es más sencilla** en comparación con el establecimiento de un data warehouse corporativo.
- El **costo de la implementación** de data mart es **mucho menor** que el requerido para un data warehouse

Data Warehouses Vs Data Marts

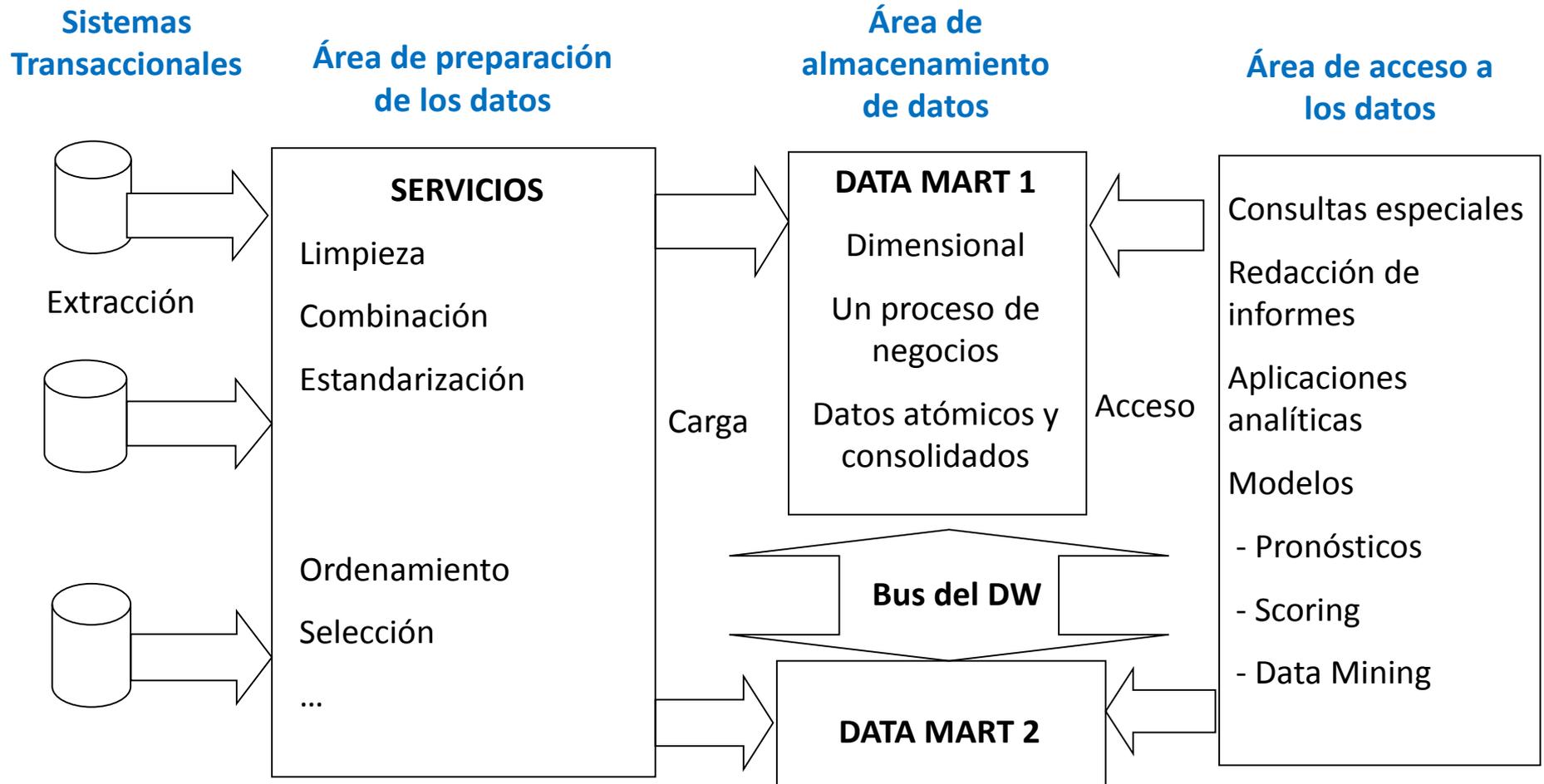


	Data Warehouse	Data Mart
Alcance	Empresa	Departamento
Sujetos	Múltiple	Grupo
Fuente de datos	Muchas	Pocas
Tamaño (típico)	100 GB a > 1 TB	<100 GB
Tiempo de implementación	Meses o años	Meses

Desde Data Warehouse a Data Marts



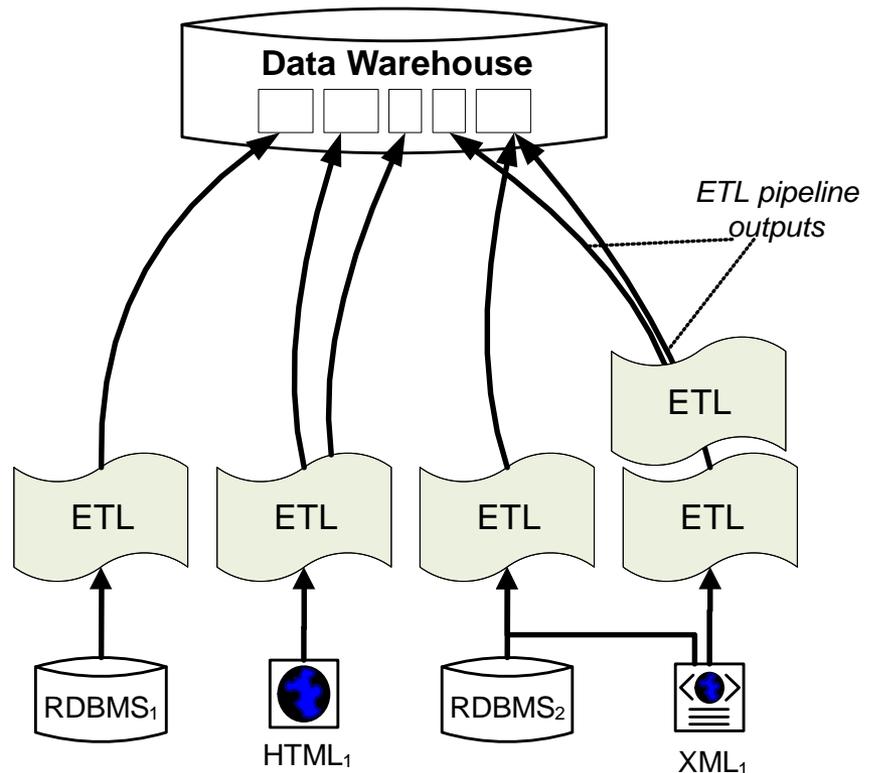
Arquitectura genérica Data Warehouse con Data Mart



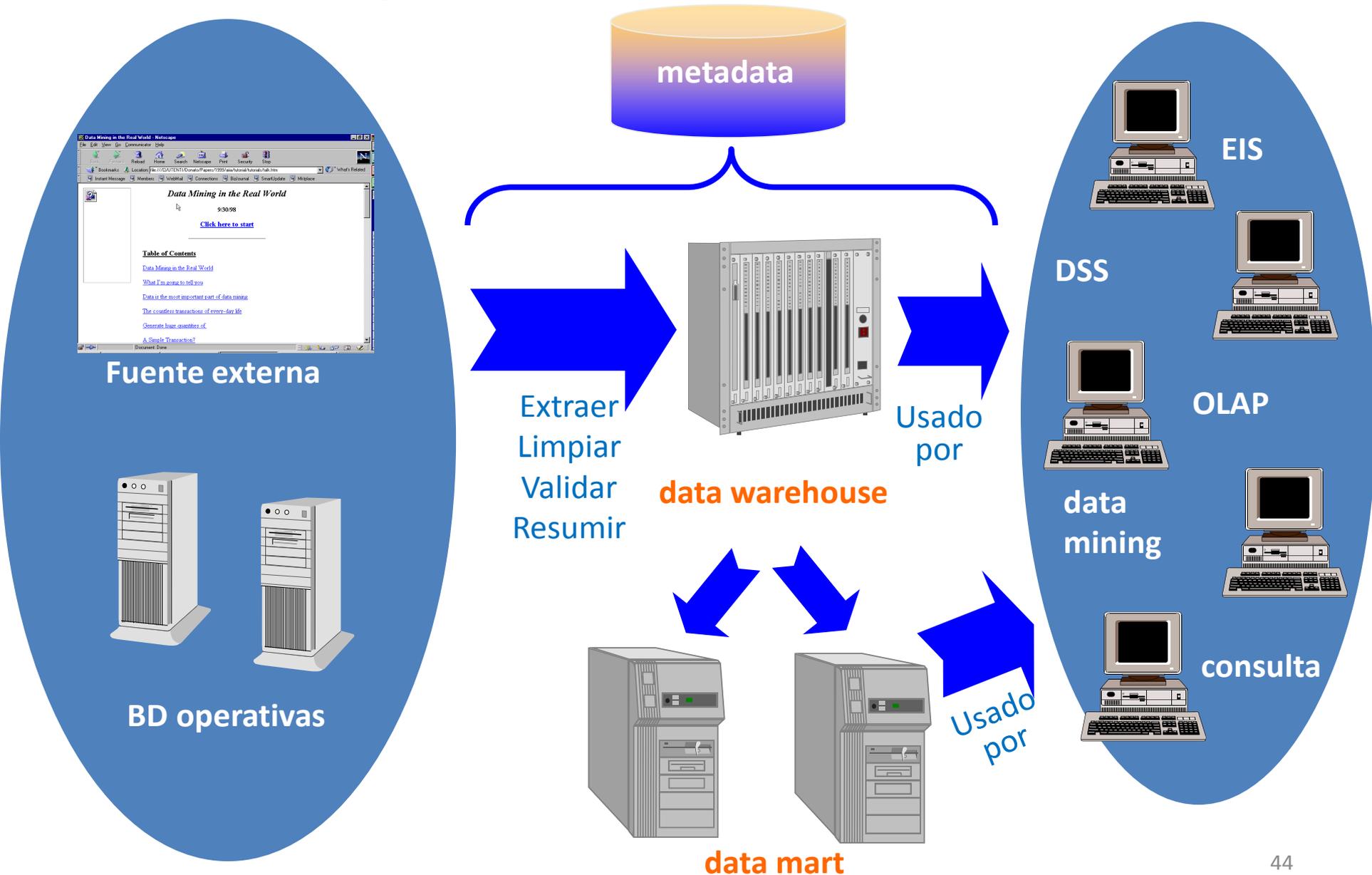
Arquitectura Data Warehouse

primer modelo

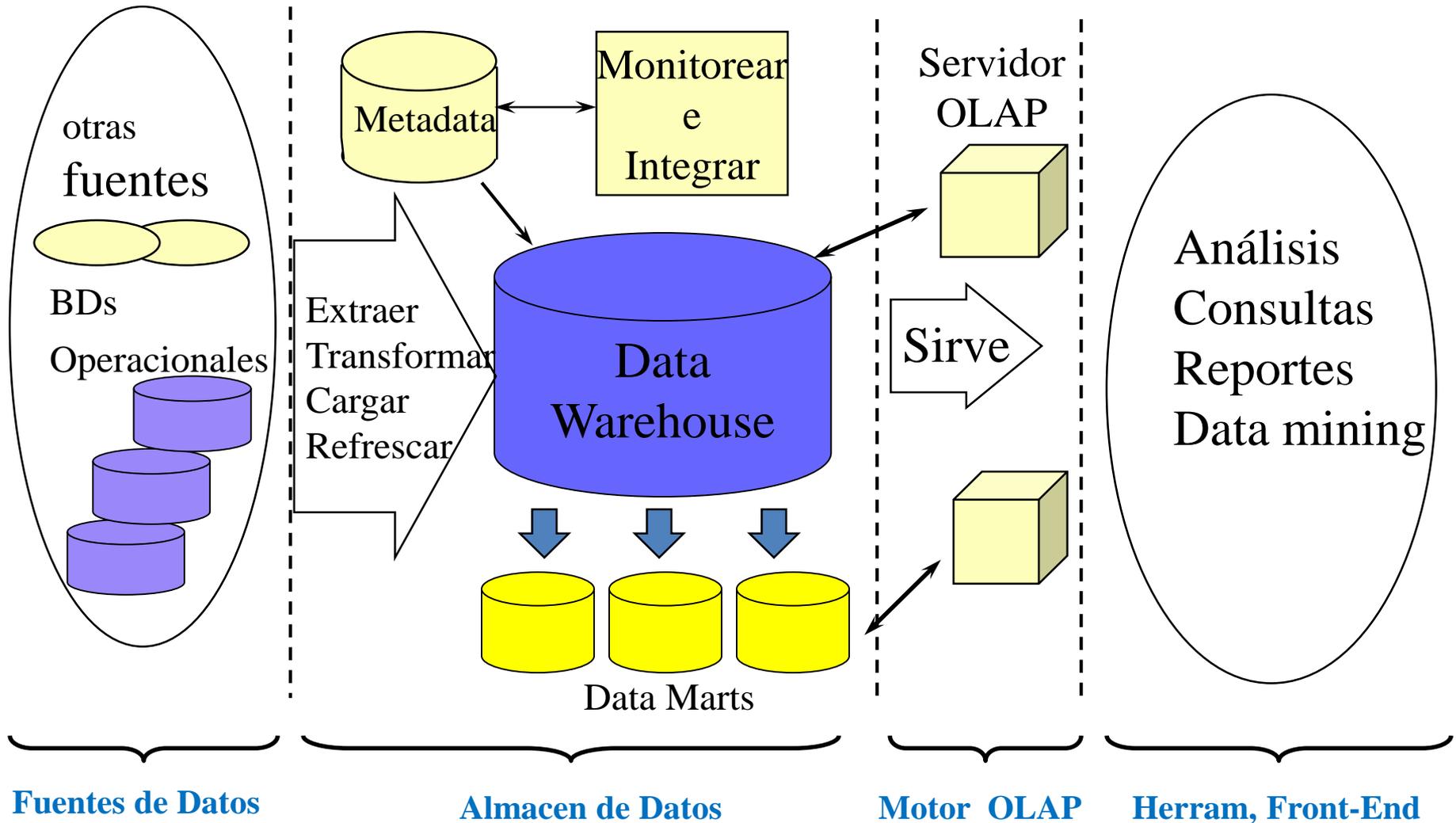
- **En la parte superior una base de datos centralizada**
 - Generalmente configurado para consultas y agregar – no transacciones
 - Muchos índices, vistas , etc.
- **Los datos se cargan y actualizan periódicamente a través de ETL (Extract / Transform / Load)**



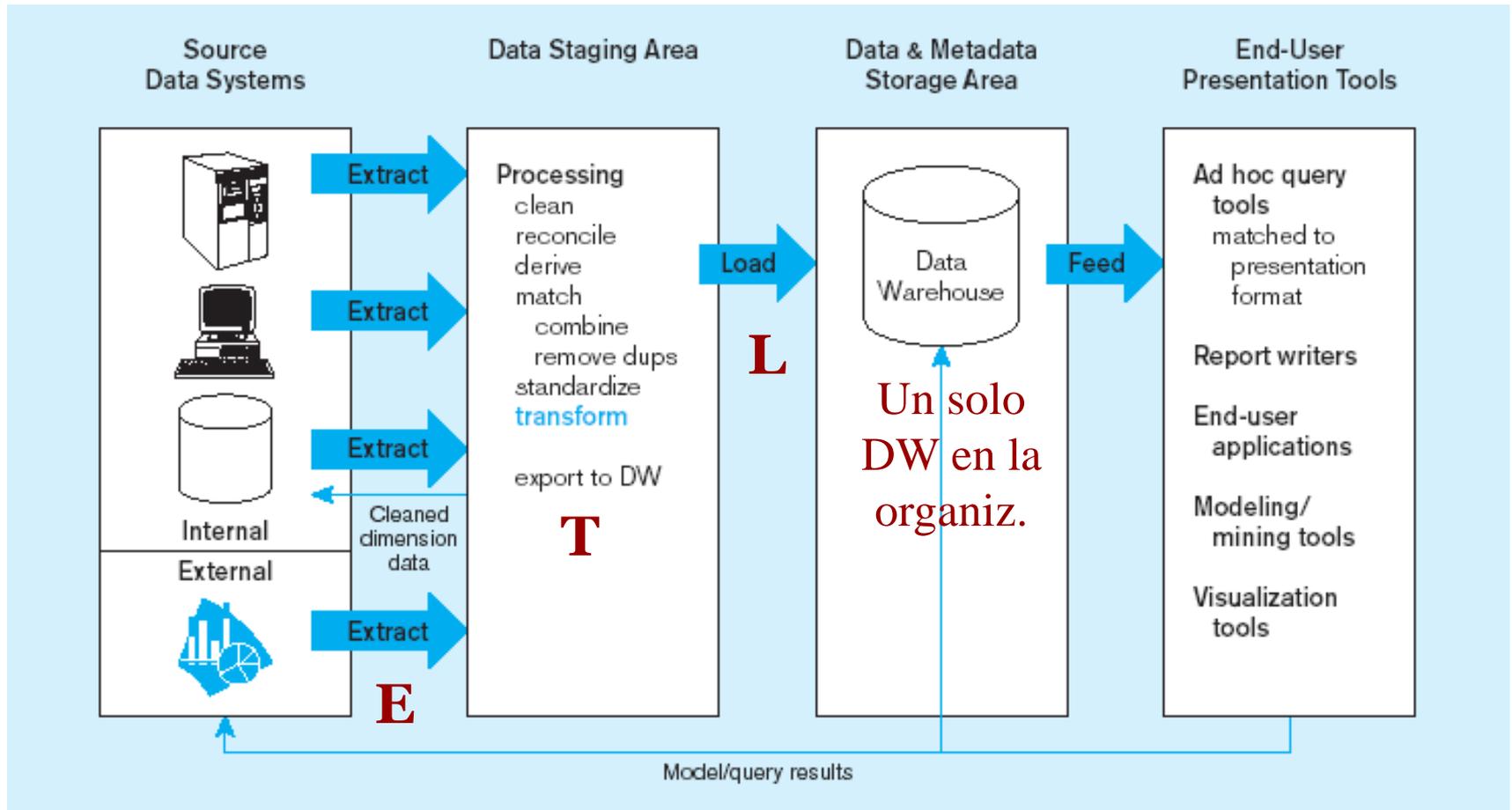
Arquitectura varios niveles



Arquitectura varios niveles



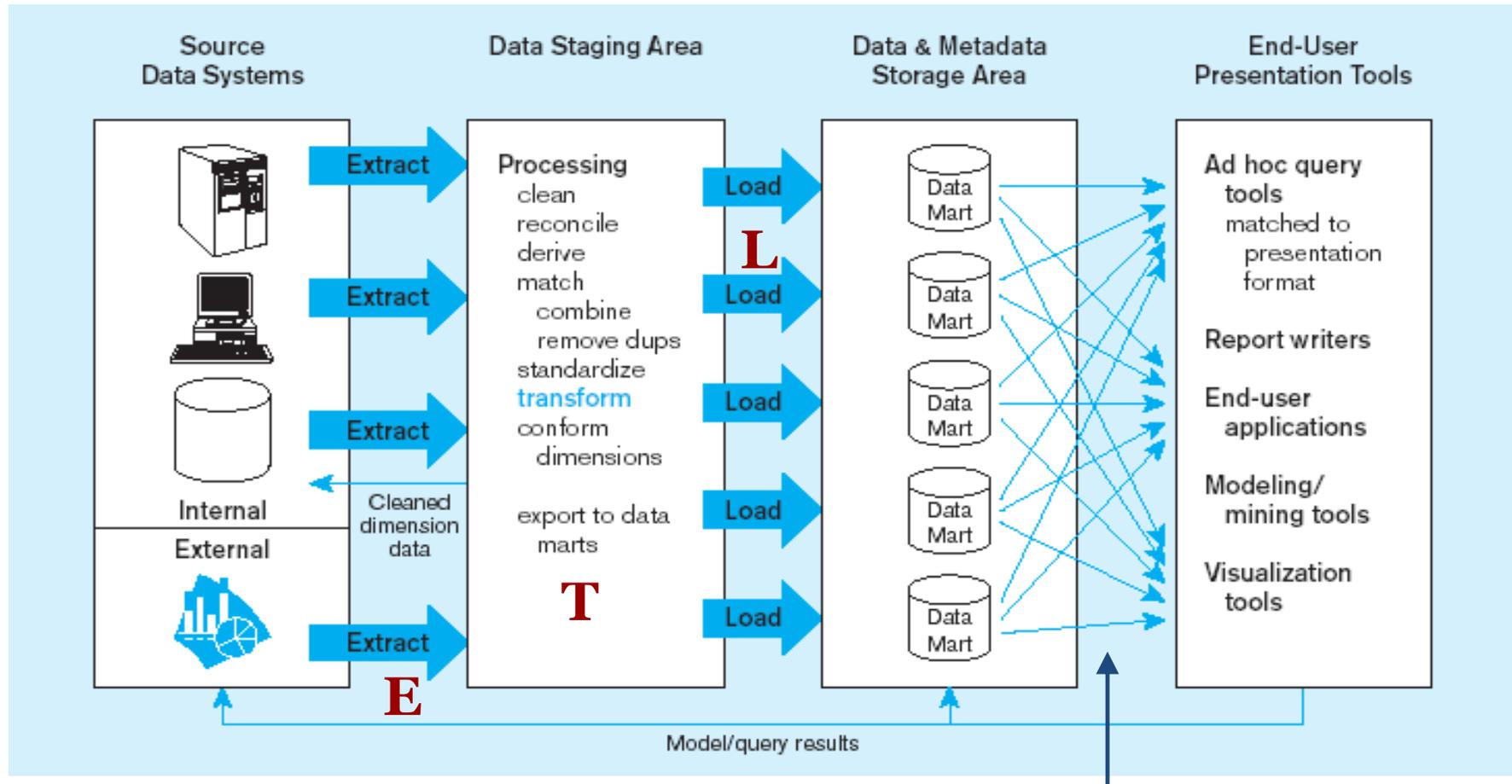
Genérica arquitectura de niveles de DW



Centralizado

Extracción de datos periódica
Datos no están completamente al día

Arquitectura independientes Data mart

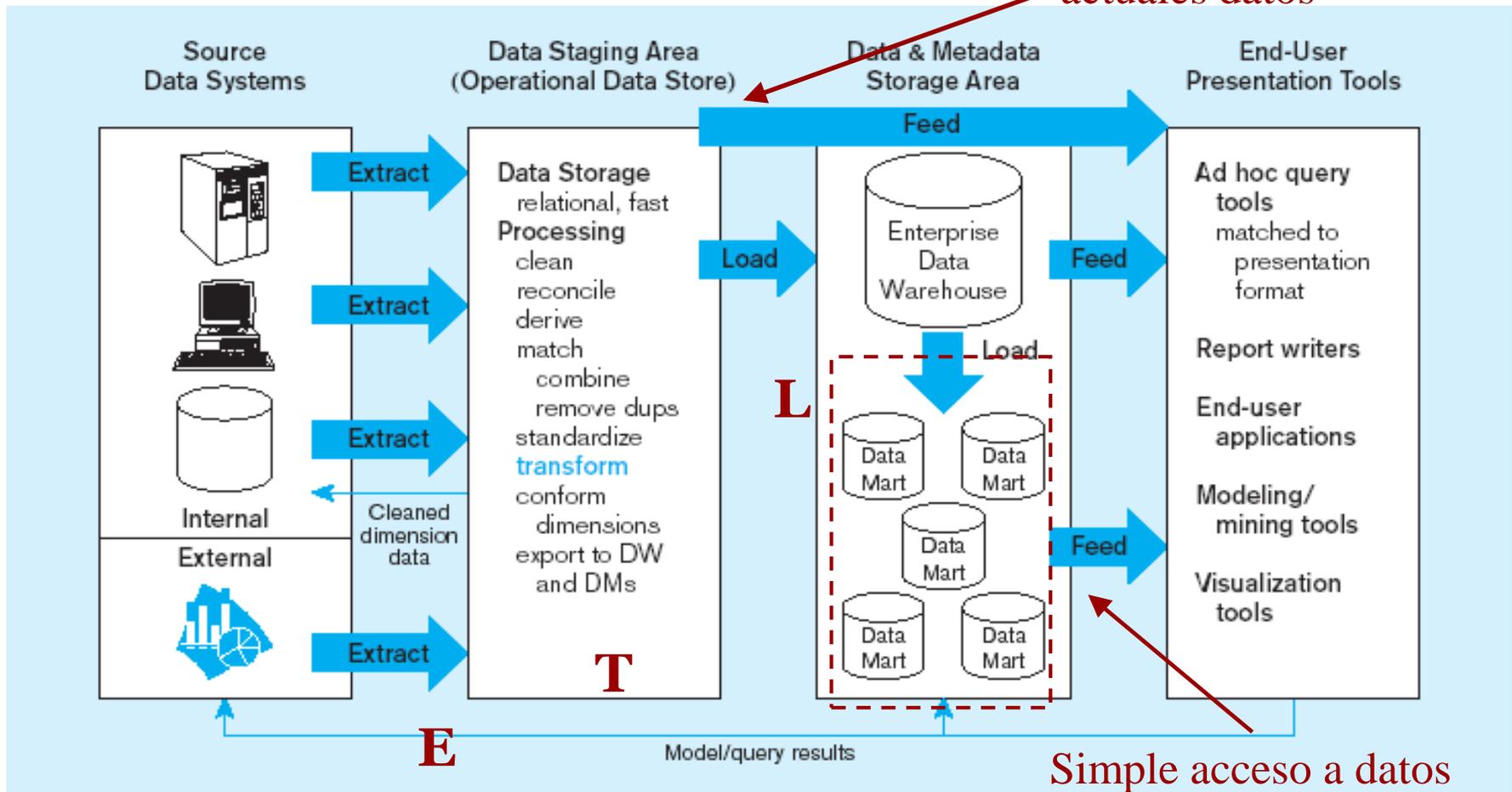


Separado ETL para c/data mart

Acceso datos por *multiples* data marts

Arquitectura Híbrida

Se pueden obtener
actuales datos

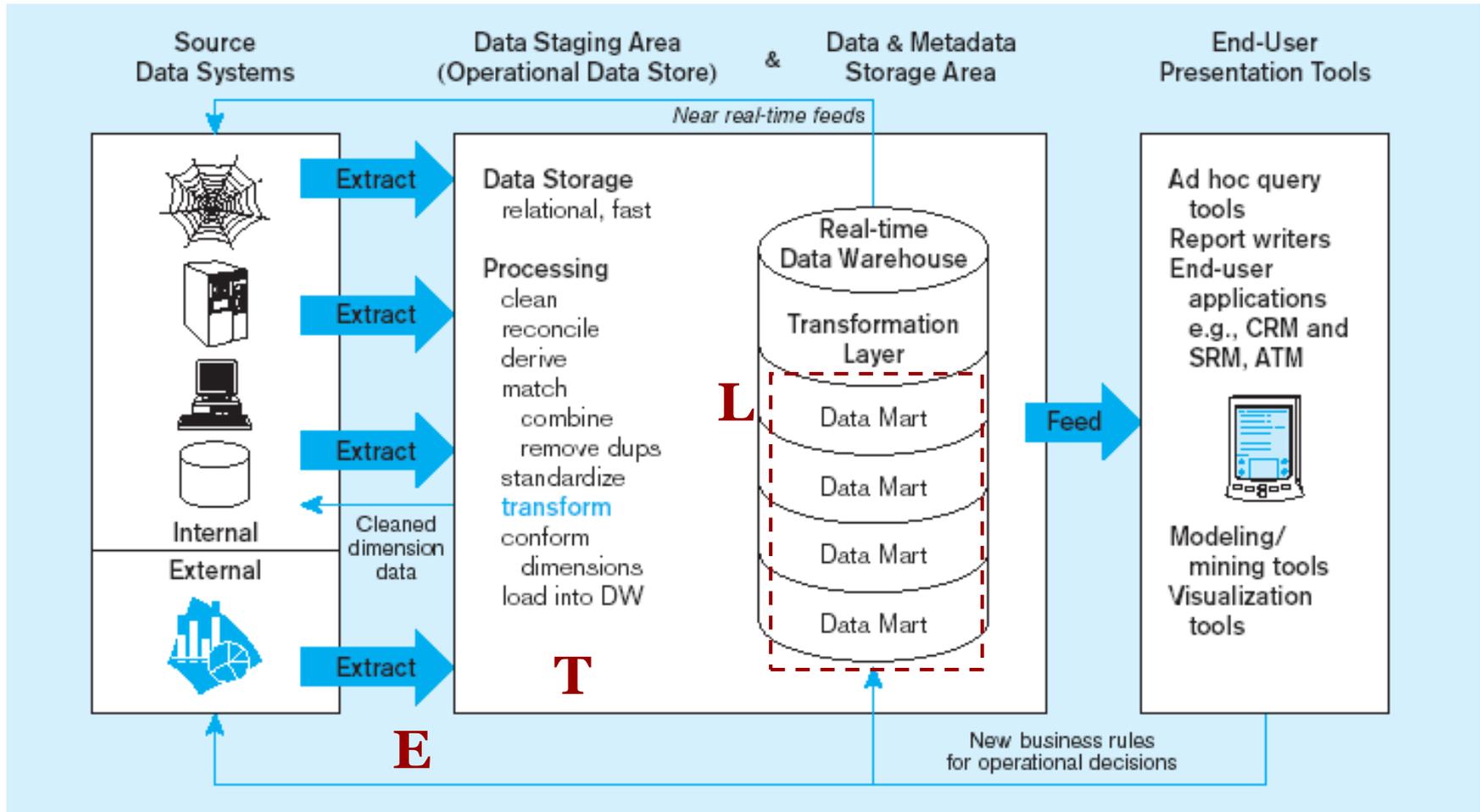


Un solo ETL para
enterprise data warehouse
(*EDW*)

Data marts *dependientes*
del EDW

Federado

Data mart Lógico y DW-RT



ETL RT para
Data Warehouse

Data marts no separado de BDs, vistas lógicas de los datos de data warehouse
→ Fácil crear nuevos data marts

Diferencia Data Warehousing y BD

BD

- Una **colección estructurada** de registros o datos

Data Warehouse

- Una **colección lógica** de información, recogida de muchos diferentes bases de datos operacionales, que apoya las actividades de análisis organizacional y las tareas de toma de decisiones

Formato

Modelos BD

relacional
jerárquica
red
orientada a objetos
Etc.

Data Warehouse

- Tradicionalmente heterogénea => integración
- Se construyen envolturas/ mediadores en la parte superior
- Cuando una consulta se plantea a un sitio, un **meta-diccionario** se usa para traducir la consulta en consultas apropiadas para el sitio
- Los resultados se integran en un conjunto de respuestas globales
Filtradas

La información de fuentes heterogéneas se integra con antelación y se almacena en bodegas para consultas directas y análisis

BD

Diseñado para
tiempos
operaciones en
tiempo real

En su mayoría
permanente
actualizaciones

Muchas pequeñas
transacciones

Mb - Gb de datos

Data Warehouse

Diseñado para el análisis de
la organización por
categorías y atributos

Sobre todo se leen

Las consultas son largas y
complejas

Gb - Tb de datos

BD

- Instantánea
- Los datos en bruto
- Cientos de usuarios (por ejemplo, los usuarios de una organización)

Data Warehouse

- Histórica
- Resumidos y conciliados los datos
- Decenas de usuarios (por ejemplo, los gerentes, los analistas)

Data Warehouse vs. SMBD

- **OLTP (on-line transaction processing)**
 - Major tarea de SMBD relacionales tradicionales
 - Las operaciones del día a día: compras, inventario, financiera, nómina, contabilidad, etc.
- **OLAP (on-line analytical processing)**
 - Majores tareas de data warehouse
 - Análisis de datos y toma de decisiones

Características diferentes (OLTP vs. OLAP):

- **Orientación de la BD:** usuario vs organización
- **Contenido de los datos:** actual, detallada vs histórica, consolidada
- **Diseño de base de datos:** aplicación + ER vs estrella + por tema
- **Vista:** actual, local vs integrada, , evolutiva
- **Patrones de acceso:** actualizar vs consultas complejas

Las diferencias de diseño

BDs

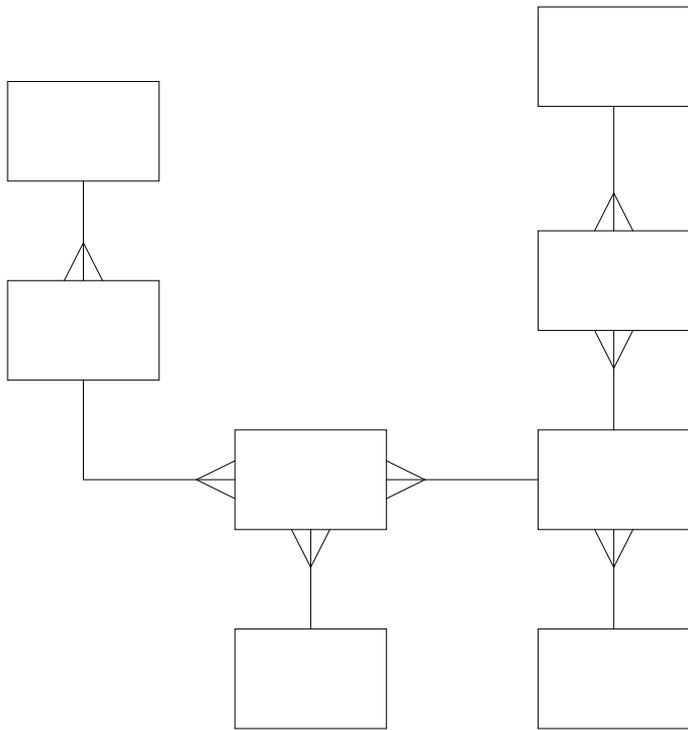
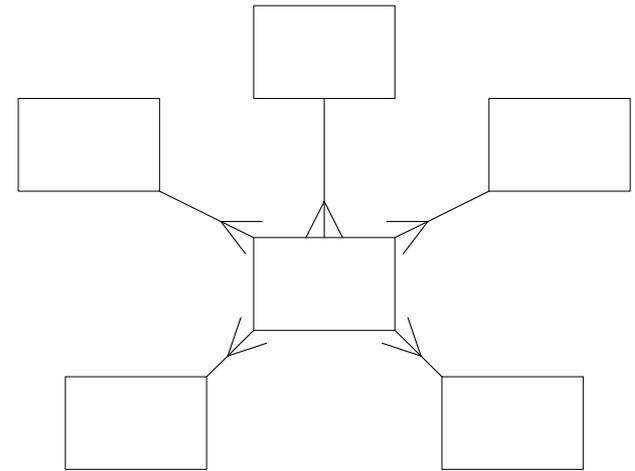


Diagrama ER

Data Warehouse



Estrella

RESUMEN DIFERENCIAS

BD OPERACIONAL

- Datos operacionales
- Orientado a aplicaciones
- Datos Actuales
- Datos Detallados
- Datos en continuo cambio

DATAWAREHOUSE

- Datos de negocio
- Orientado al sujeto
- Actuales + Histórico
- Datos Resumidos
- Datos Estables

Por qué Data Warehouse?

Diferentes funciones y datos:

- **Datos que faltan:** apoyo a las decisiones requiere datos históricos que BDs operacionales no suelen tener
- **Consolidación de datos:** Se requiere de la consolidación (agregación, resumen) de los datos de fuentes heterogéneas
- **Calidad de los datos:** las diferentes fuentes suelen utilizar representaciones de datos inconsistentes, códigos y formatos que deben ser conciliados, etc.



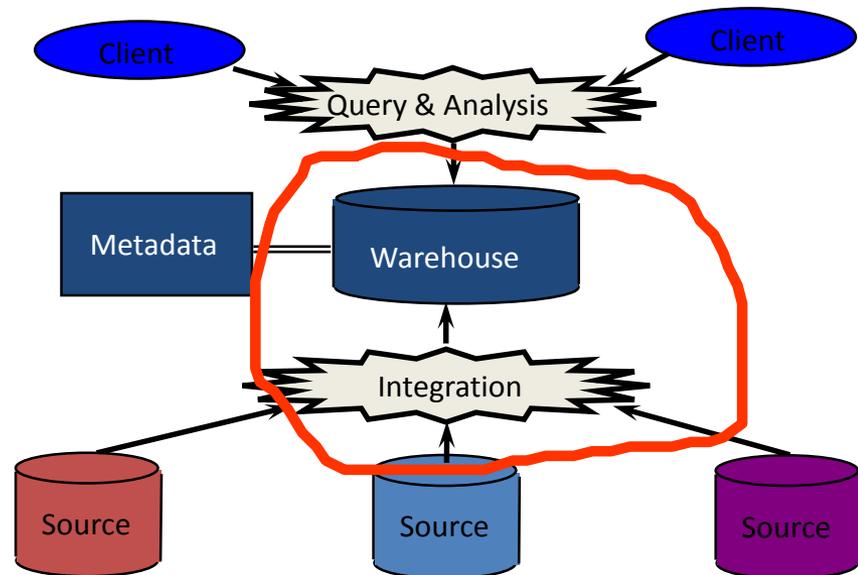
UNIVERSIDAD
DE LOS ANDES
MÉRIDA VENEZUELA

Datawarehousing: ETL

Jose Aguilar
CEMISID, Escuela de Sistemas
Facultad de Ingeniería
Universidad de Los Andes
Mérida, Venezuela

Integración

- Selección de los datos
- Transformación de datos
- Carga de datos



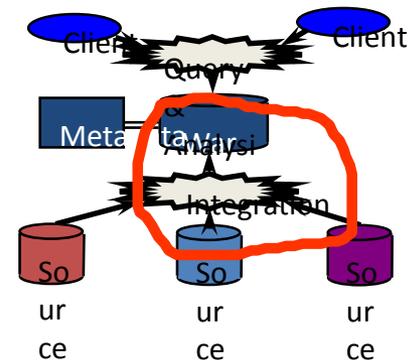
Proceso ETL

ETL (Extracción, Transformación y Carga)

Extracción: Obtención de información de las distintas fuentes, tanto internas como externas.

Transformación: Filtrado, limpieza, depuración, homogeneización y agrupación de la información.

Carga: Organización y actualización de los datos y los metadatos en el DW.

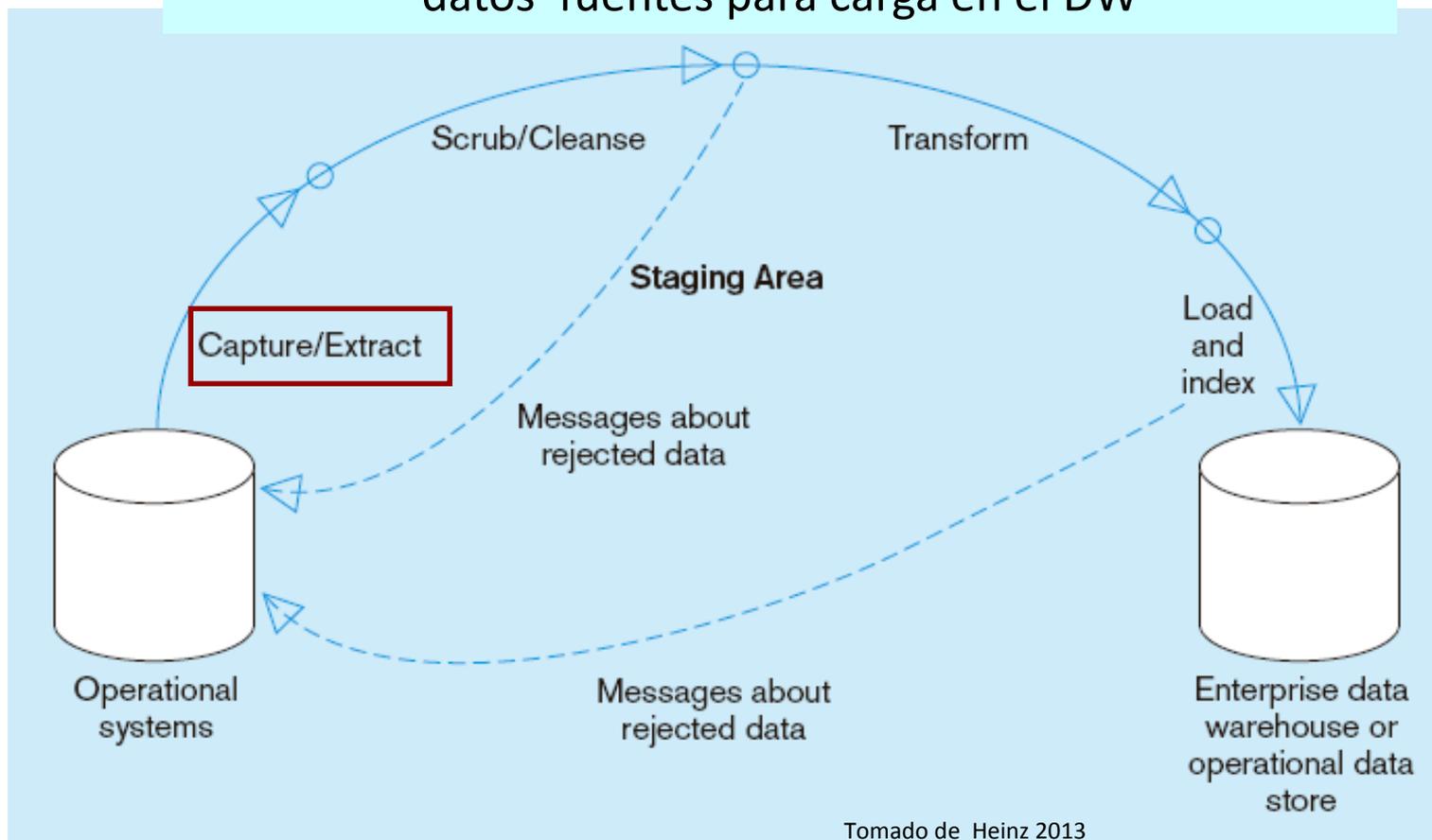


Extracción

Obtener datos de múltiples, heterogéneas fuentes externas

- Periodicidad
- Claves:
 - Manipular los datos sin interrumpir ni paralizar los OLTP, ni tampoco el DW.
 - No depender de la disponibilidad de los OLTP.
 - Almacenar y gestionar los metadatos que se generarán en los procesos ETL.
 - Facilitar la integración de las diversas fuentes, internas y externas.

Captura / Extrae...... obtiene un subconjunto de los datos fuentes para carga en el DW



Extracción estática: captura los datos en un momento puntual

Extracción Incremental: captura cambios que se van produciendo

ETL: Técnicas de Monitoreo

- Instantáneas cada cierto tiempo
- Disparadores de base de datos (reglas (triggers))
- Logs de registros
- Envío de datos
- Envío de transacciones
- Consultas a las BDs

Métodos de Extracción

- **Extracciones a granel**

- Todo el DW se actualiza periódicamente
- Pesado para las conexiones de red entre el origen y destino
- Más fácil de configurar y mantener

- **Extracciones basadas en intercambio**

- Sólo los datos que han sido recién insertadas o actualizadas en los sistemas de origen se cargan en el DW
- Menos carga en la red pero requiere una programación más compleja para determinar cuando un nuevo registro de DW debe insertarse o cuando un registro DW tiene que actualizarse

Limpieza de datos

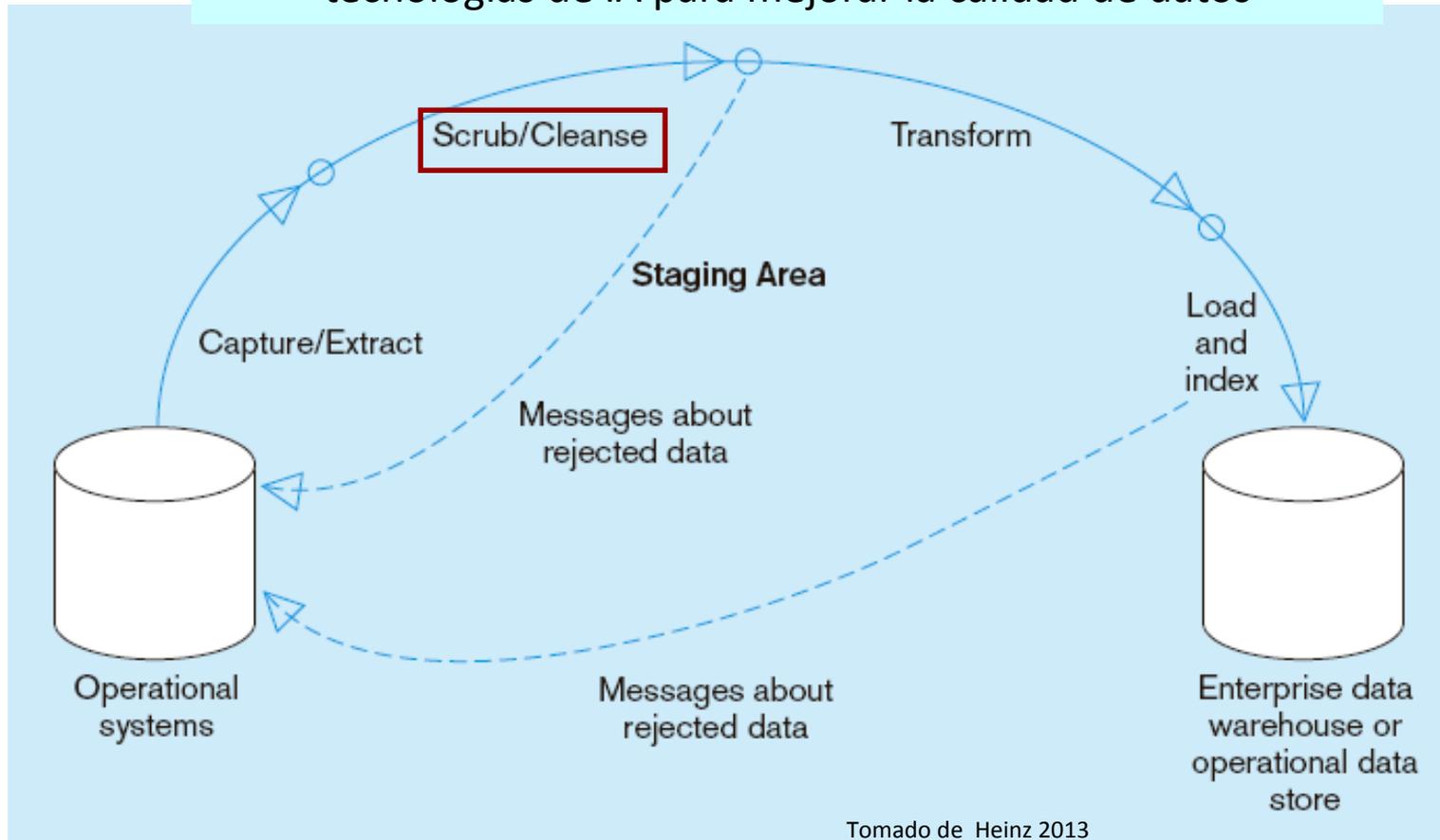
Sistemas de origen tiene "datos sucios" que deben ser limpiados

Razones de datos "sucios"

- valores ficticios
- La falta de datos
- datos encriptados
- datos contradiciéndose
- El uso inapropiado de datos
- Violación de Reglas de Negocio
- Claves principales reutilizados
- Identificadores no únicos

**detectar errores
en los datos y
rectificarlos**

Limpieza... utiliza reconocimiento de patrones y tecnologías de IA para mejorar la calidad de datos



Solución de errores: faltas de ortografía, fechas erróneas, uso de campo incorrecto direcciones no coinciden, datos faltantes, datos duplicados, inconsistencias

También decodifica, reformatea, convierte, genera claves, fusiona, detecta errores registro, localiza datos faltantes

Pasos en la Limpieza de datos

1. Análisis (parsing)
2. Corrección
3. Estandarización
4. Mapeo (matching)
5. Consolidación

Análisis (Parsing)

Localiza e identifica **elementos individuales** en los archivos de origen y luego los aísla.

- **Ejemplos**

- Análisis del primer nombre, segundo nombre y apellido;
- Analizar número y nombre de la calle;
- Analizar la ciudad y el estado.

Corregir

Corrige los componentes de **datos individuales** utilizando algoritmos sofisticados y fuentes de datos secundarias.

Acciones típicas con Datos Anómalos (Outliers):

- Ignorarlos.
- Eliminar la columna.
- Filtrar la columna.
- Filtrar la fila errónea, ya que a veces su origen, se debe a casos especiales.
- Reemplazar el valor.

Acciones contra Datos Faltantes (Missing Values):

- Ignorarlos.
- Eliminar la columna.
- Filtrar la columna.
- Filtrar la fila errónea, ya que a veces su origen, se debe a casos especiales.
- Reemplazar el valor.
- Esperar hasta que los datos faltantes estén disponibles.

Estandarización

Aplica **rutinas de conversión** para transformar los datos en formatos preferidos (y coherentes), utilizando reglas de estandarización.

- **Ejemplos:** la sustitución de un apodo, uso de un nombre de calle preferido, etc.

Mapeo (matching)

búsqueda de **coincidentes en registros** en los datos analizados, corregidos y estandarizados, basados en reglas predefinidas para eliminar la duplicación.

- **Ejemplos:** identificación de nombres y direcciones similares.

Transformación de datos

Transforma los datos de acuerdo con reglas y normas establecidas

- Clásicos problemas:
 - Codificación.
 - Medida de atributos.
 - Convenciones de nombramiento.
 - Fuentes múltiples,

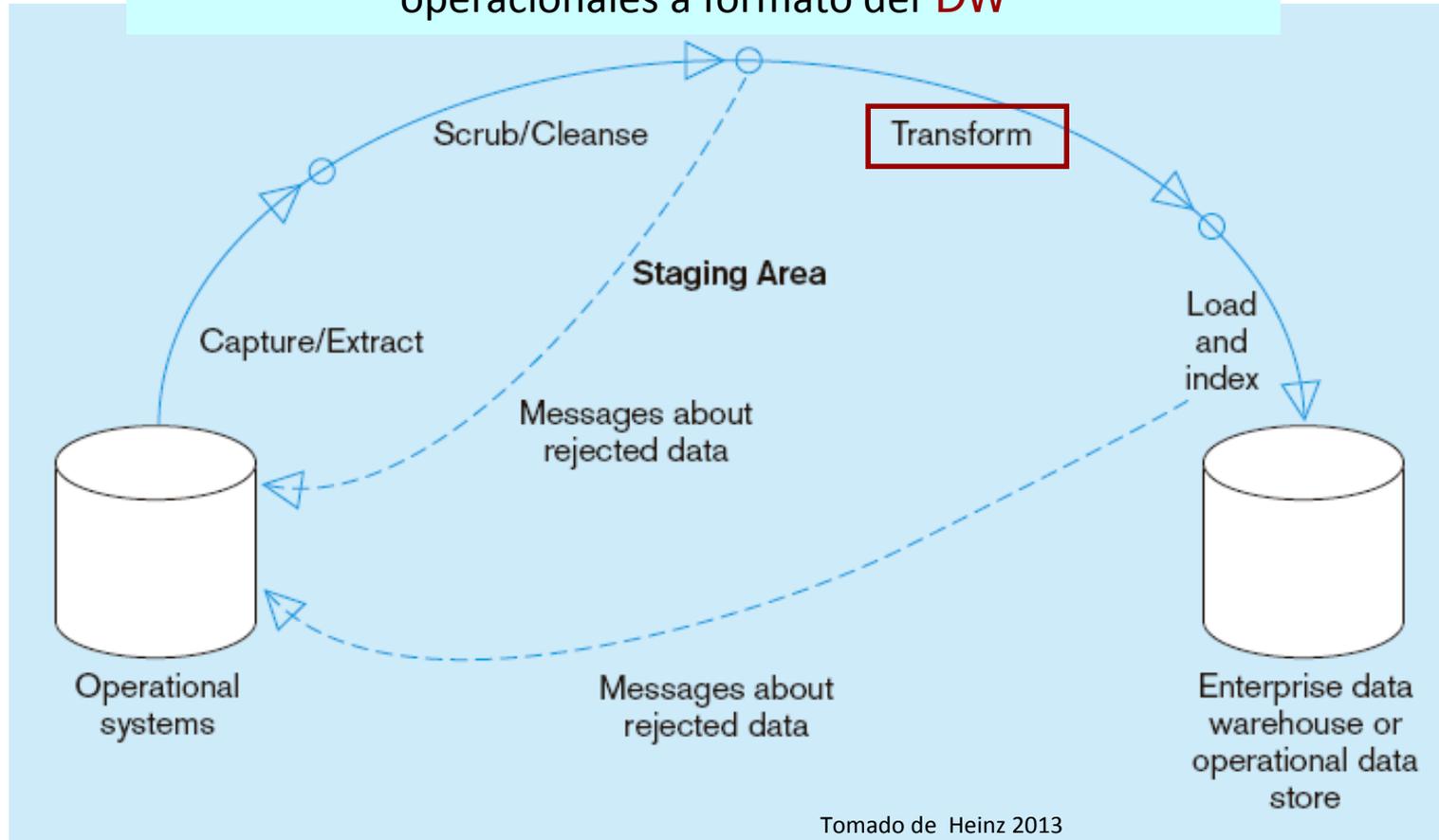
ordena

calcula

resume

consolida

Transforma... convierten los datos desde las BD operacionales a formato del **DW**



Tomado de Heinz 2013

A nivel de registro:

- Partición de Datos (selección)
- Juntar datos (combinación)
- Resumir datos (agregación)

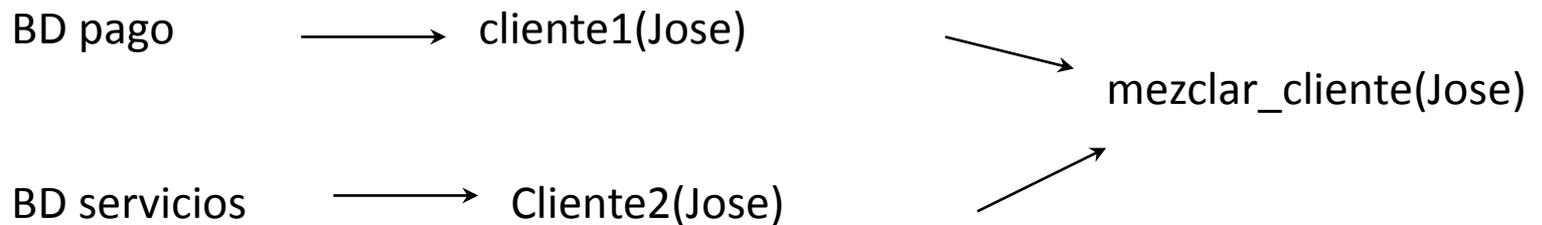
A nivel de campo:

- de un solo campo: de un campo a un campo
- multi-campo: de muchos campos a uno, o uno a muchos campos

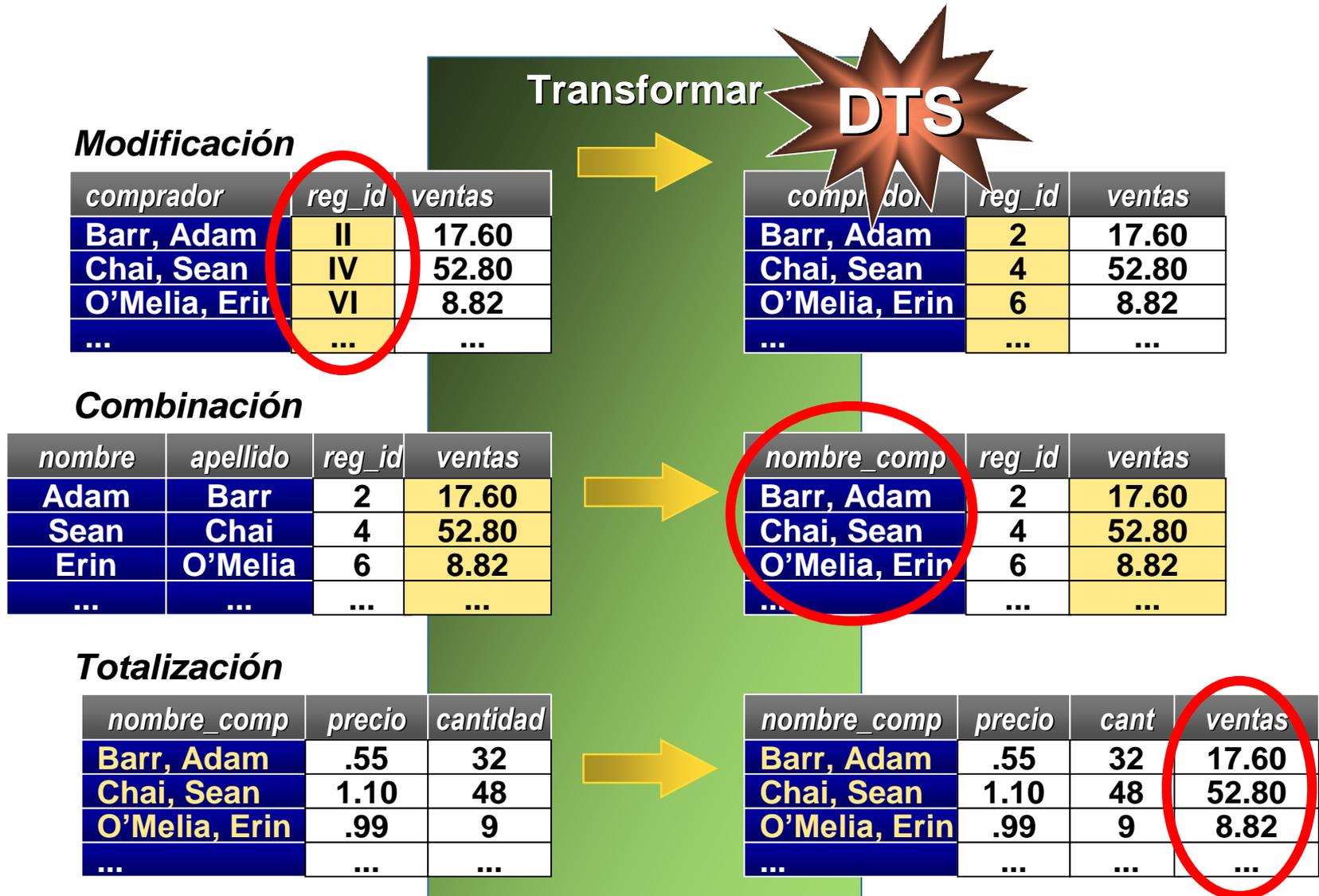
ETL: Transformación de datos

Posibles tareas

- **Migrar** (por ejemplo, yen a dólares)
- **Refinar**: utilizar el conocimiento específico de dominio (por ejemplo, números de seguro social)
- **Fusionar** (por ejemplo, lista de correo con la de clientes)

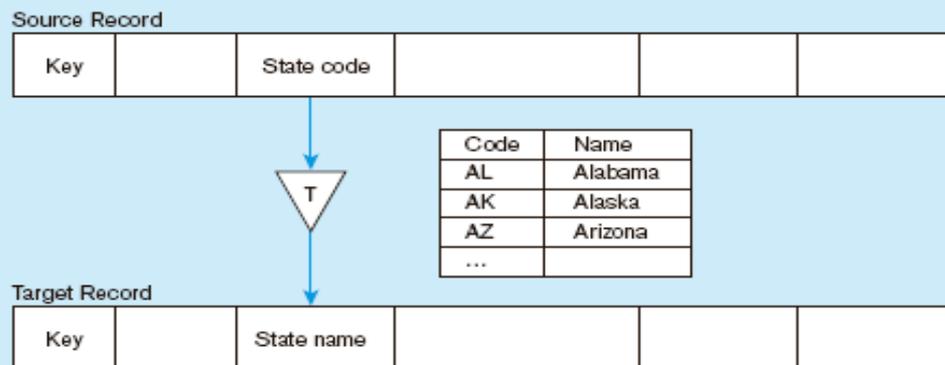
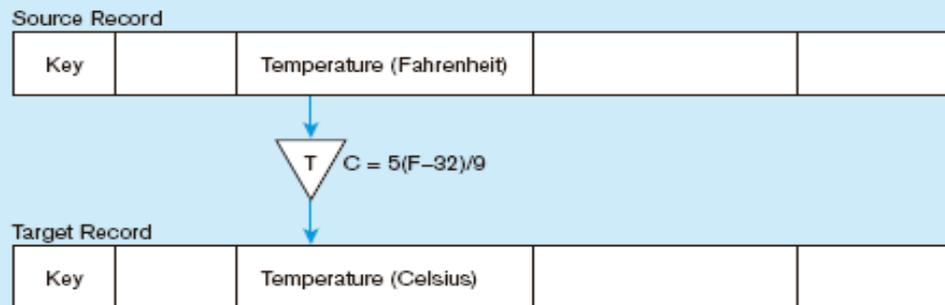
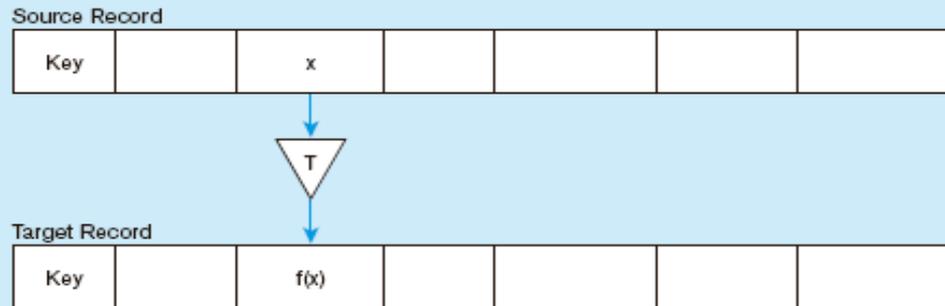


Proceso ETL: Transformar



Transformación de un solo campo

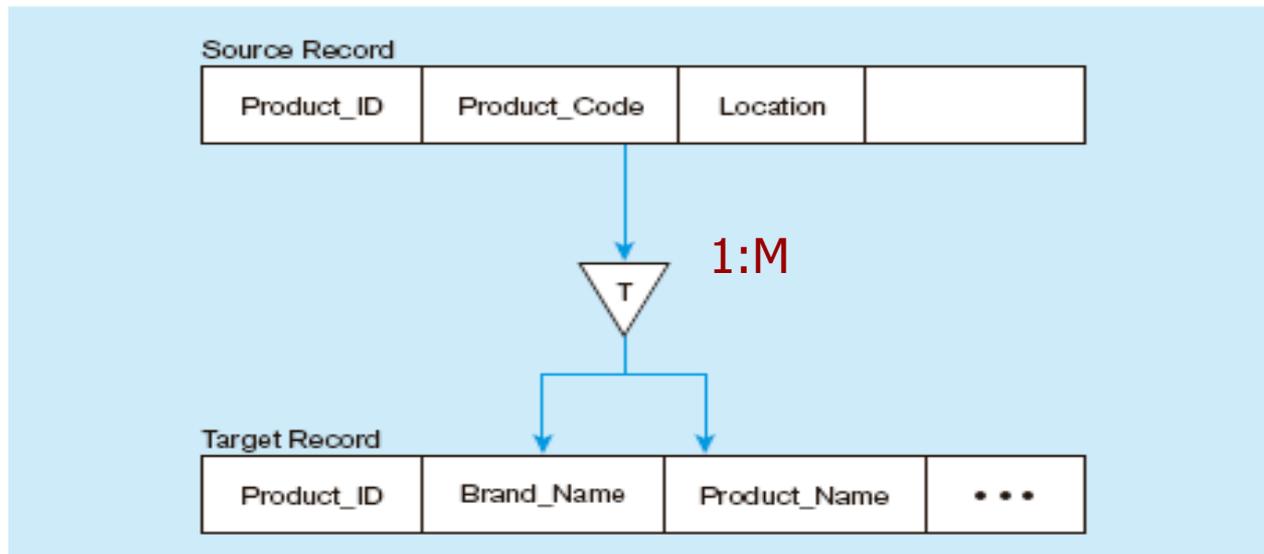
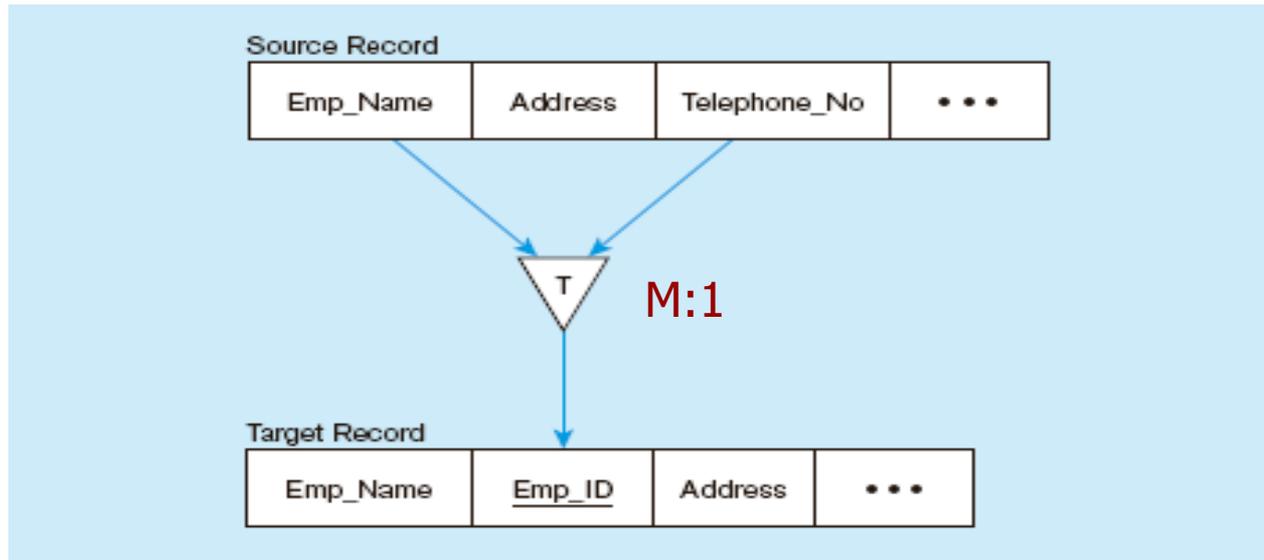
Una función de transformación traduce los datos de una forma antigua a una nueva



Transformación algorítmica: utiliza una fórmula o expresión lógica

Table lookup: utiliza una tabla separada basada en códigos

Transformación multicampos

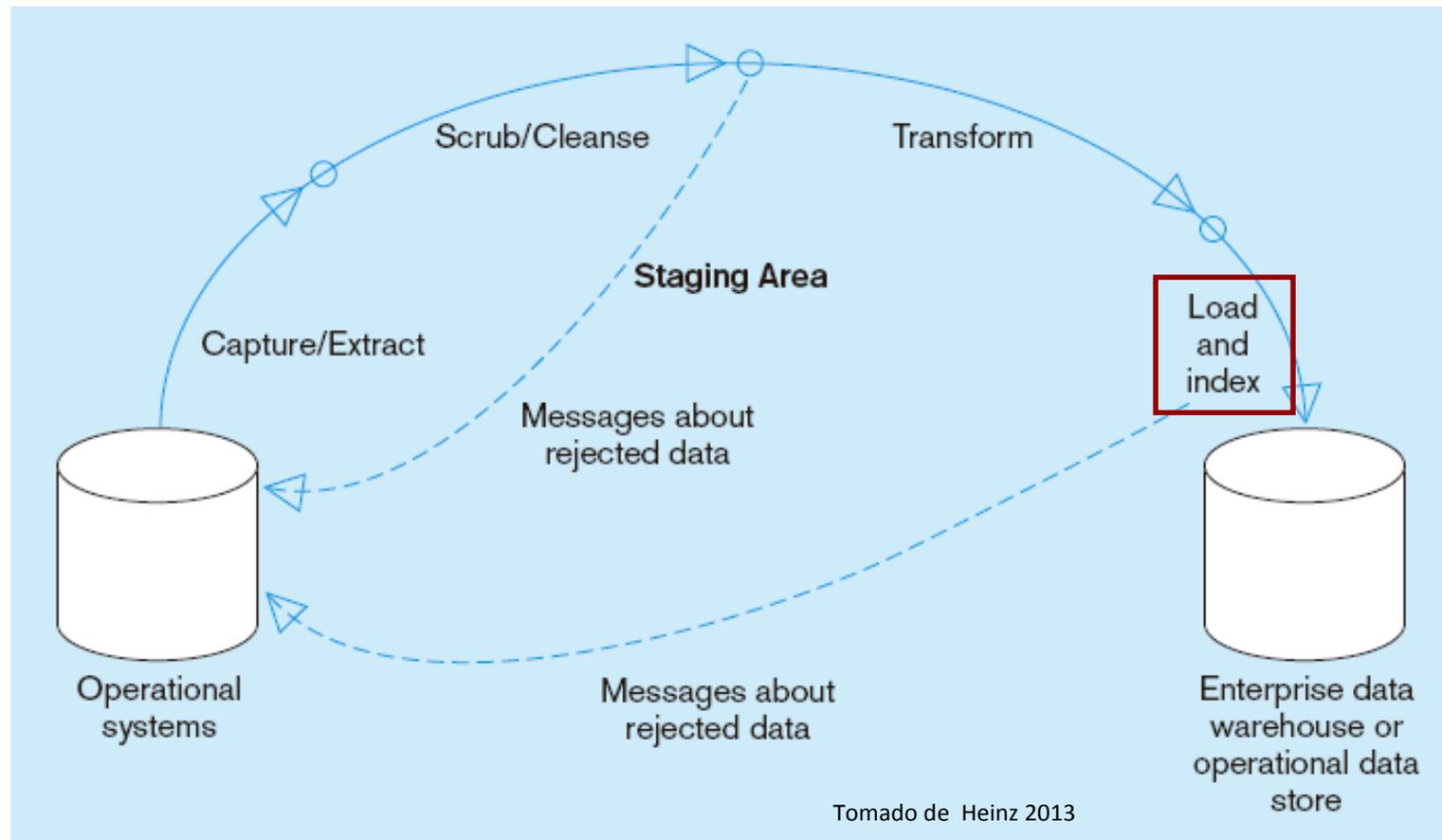


Carga de Datos

Los datos físicamente se almacena en el almacén de datos

- La carga ocurre en una "ventana de carga"
- La tendencia cada vez mayor es actualizaciones en tiempo real

Cargar/Indizar... Transforma datos y crea índices



Modo de Actualización 1:

reescritura masiva de datos en destino a intervalos periódicos

Modo de actualización 2:

solamente cambios en los datos de origen se escriben en el warehouse



UNIVERSIDAD
DE LOS ANDES
MÉRIDA VENEZUELA

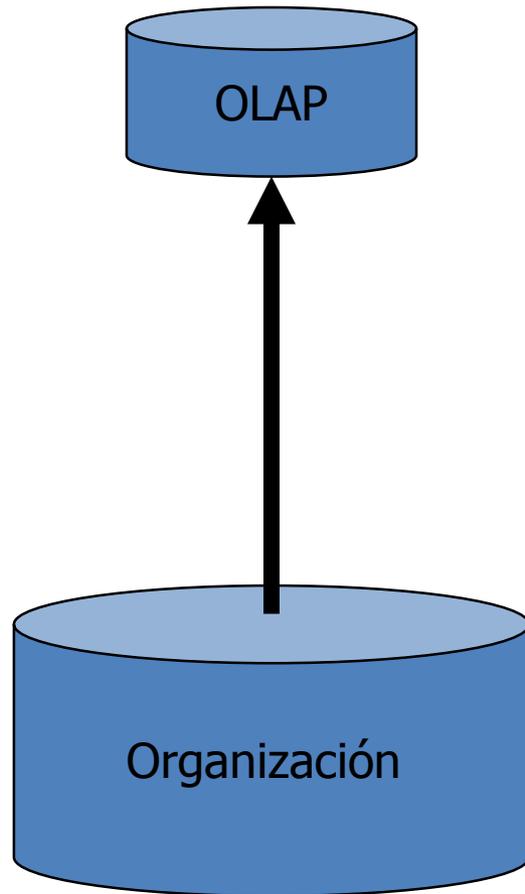


Modelado de Datos

Jose Aguilar
CEMISID, Escuela de Sistemas
Facultad de Ingeniería
Universidad de Los Andes
Mérida, Venezuela

Modelado de Datos y de Información

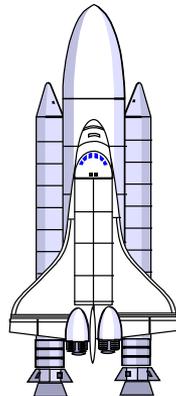
Usuarios tienen diferentes vistas de los datos



Turistas: Navegar por la información recolectada



Agricultores: información de los caminos a los datos, etc.



Exploradores: Busca cosas desconocidos que se esconden en los datos detallados

Consideraciones para el Diseño Data warehouse

Para abordar un proyecto de data warehouse es necesario hacer un estudio de algunos temas generales de la organización:

- **Situación actual:** Cualquier solución propuesta de data warehouse debe estar muy orientada por las necesidades del negocio, debe ser compatible con la arquitectura técnica existente y planeada de la compañía.
- **Tipo y características del negocio:** Tener el conocimiento exacto sobre el tipo de negocios de la organización y el soporte que representa la información dentro de todo su proceso de toma de decisiones.

Consideraciones para el Diseño Data warehouse

Para abordar un proyecto de data warehouse es necesario hacer un estudio de algunos temas generales de la organización:

- **Entorno técnico:** hardware (servidores, redes,...) así como aplicaciones y herramientas. Se dará énfasis a los Sistemas de Soporte a Decisiones (DSS).
- **Expectativas de los usuarios:** Es una forma de vida de las organizaciones y como tal, tiene que contar con el apoyo de todos los usuarios y su convencimiento sobre su bondad.

Modelos dimensionales

Es una técnica de **diseño lógico** comúnmente utilizada para Data Warehouses, que busca presentar los datos en una arquitectura estándar y permita una **alta performance de acceso** a los usuarios finales.

El modelo se basa en **esquemas estrella**, conformados por **Tablas de Hechos y Tablas Dimensionales** (p.ej. cubos).

Modelos dimensionales

- Un **modelo relacional desnormalizado**
 - Compuesto por tablas con atributos
 - Las relaciones son definidas por claves nuevas y claves externas
- Organizado para **la comprensibilidad y facilidad de presentación** de informes en lugar de facilitar la actualización
- Consultado y mantenido por **herramientas especiales de gestión analítica**

Diseño de Esquemas

- **Los datos se organizan por temas importantes:**

Los clientes, los productos, las ventas, ...

- **Tema = datos + dimensiones**

- Recopilación de datos útiles sobre un tema

Ejemplo: ventas

- Sintetizar una visión única de los temas a analizar

Ejemplo: Ventas (producto, período, tienda, número)

- Detallar la vista según dimensiones

Ejemplo:

Productos (IDprod, descripción, color, tamaño ...)

Tiendas (IDmag, nombre, ciudad, departamento, país)

Periodo (IDper, año, trimestre, mes, día)



Diseño de Esquemas

Los tipos de Esquema

- En estrella
- Constelación
- Copo de nieve

1. Aislar Datos a tener en cuenta

- Esquemas de las Tablas de hechos

2. Definir las dimensiones

- Ejes de análisis

3. Estandarizar dimensiones

- Dividir en varias tablas unidas por referencias

4. Integrar todo

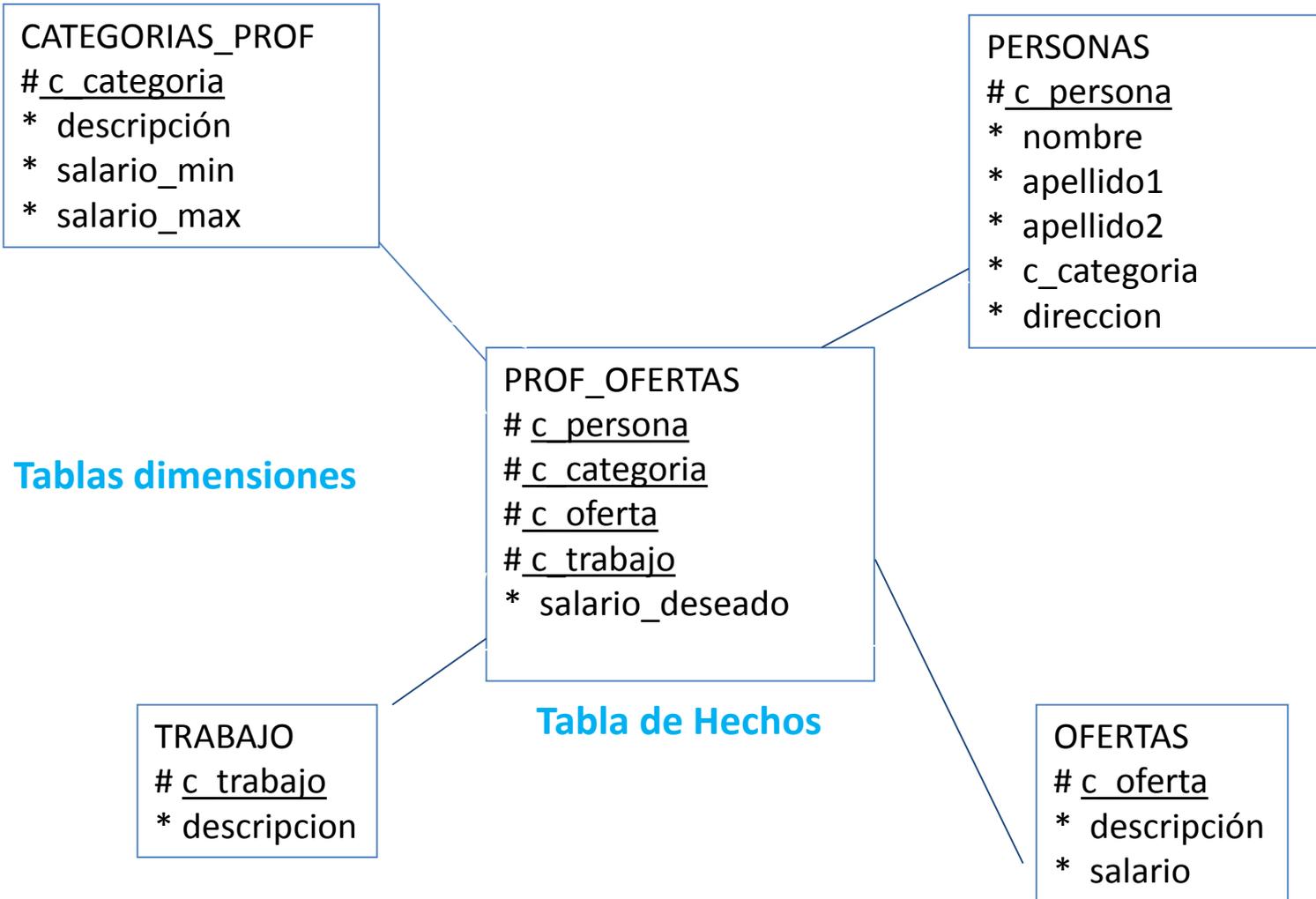
- Varias tablas de hechos comparten algunas tablas de dimensiones (constelación de la estrella)



Esquema en estrella: Componentes

- Datos (hechos)
- Dimensiones
- Atributos
- Jerarquías de atributos

Esquema en estrella



Esquema en estrella

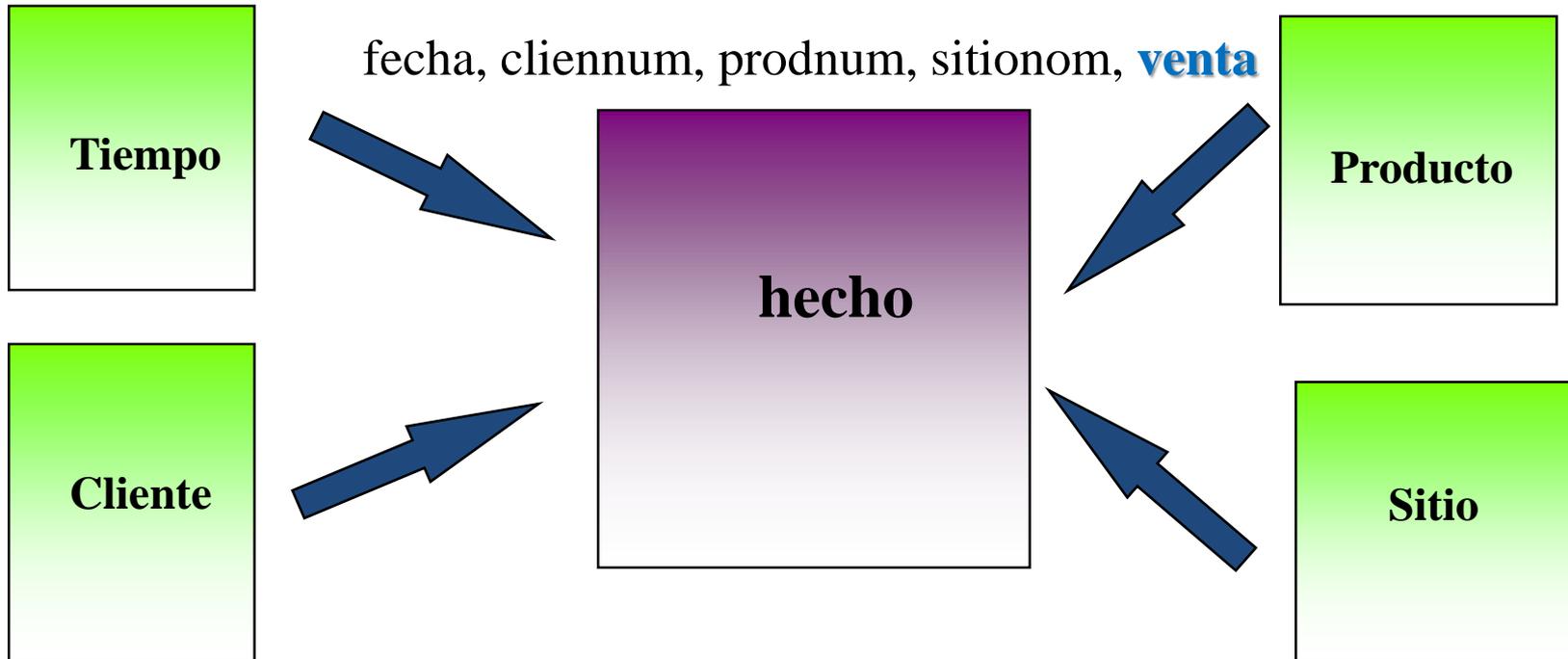
- Modelado relacional actual no satisface las necesidades actuales
- Representaciones de datos multidimensionales
- Optimizar las operaciones de consulta de datos en lugar de las operaciones de actualización de datos
- Los datos no son usados para realizar transacciones del negocio.
- Los datos pueden obtenerse mediante cálculos o agregaciones.

Esquema en estrella

- El modelo estrella es una representación de una vista de la organización.
 - Ventas
 - Mercadeo
- El modelo estrella consolida hechos en relación a dimensiones o filtros.
- Esquema en estrella
 - Hecho rodeado de varias dimensiones (4-15)
 - Las dimensiones se de-normalizan
- Una tabla de hechos en el medio conectado a un conjunto de tablas de dimensiones

Esquema en estrella

- Una sola tabla de hechos y para cada dimensión una tabla de dimensiones
- No captura jerarquías directamente



Tablas de hechos

Contienen los hechos que serán utilizados por los analistas para apoyar el proceso de toma de decisiones.

- Toma los datos desde los sistemas transaccionales
- Recibe enlace dimensiones
- Realiza las transformaciones requeridas en los datos
- Enlaza dimensiones a través de sus claves



Esquema en estrella: Hechos

- Mediciones numéricas (valores) que representan un aspecto del negocio o actividad específica
- Almacenado en una tabla de hechos en el centro del esquema de estrella
- Contiene hechos caracterizados a través de sus dimensiones
- Se pueden calcular o derivar en tiempo de ejecución
- Actualizado periódicamente con los datos de las bases de datos operacionales

Esquema en estrella: Tabla de Hechos

Tabla central

- Representa un proceso o reporta el entorno que es de valor para la organización
- Especifica exactamente lo que representa.
- Por lo general, corresponden a una entidad asociativa en el modelo ER
- Guarda Medidas de interés del negocio
- Varían bastante sus datos

Esquema en estrella: Tabla de Hechos

Tabla central

- Gran número de filas (millones a un mil millones)
- Algunas columnas como máximo
- Acceso por dimensiones: Enlaces directos a las dimensiones
 - Contiene dos o más claves foráneas
- Clave principal de varias partes

Tablas de dimensión

Definen como están los datos organizados lógicamente y proveen el medio para analizar el contexto organizacional.

- Toma los datos desde los sistemas transaccionales
- Depura los valores de los atributos para incorporarlos al modelo dimensional
- Mantiene las claves
- Mantiene la tabla de referencias cruzadas



Esquema en estrella: Dimensiones

- Características cualitativas que proveen perspectivas adicionales a un hecho
- Las dimensiones se **almacenan en tablas de dimensiones**
- Dimensiones comunes:
períodos de tiempo, áreas geográficas (mercados, ciudades), productos, clientes, vendedores, etc.
- **Típicamente contienen atributos para consultas**

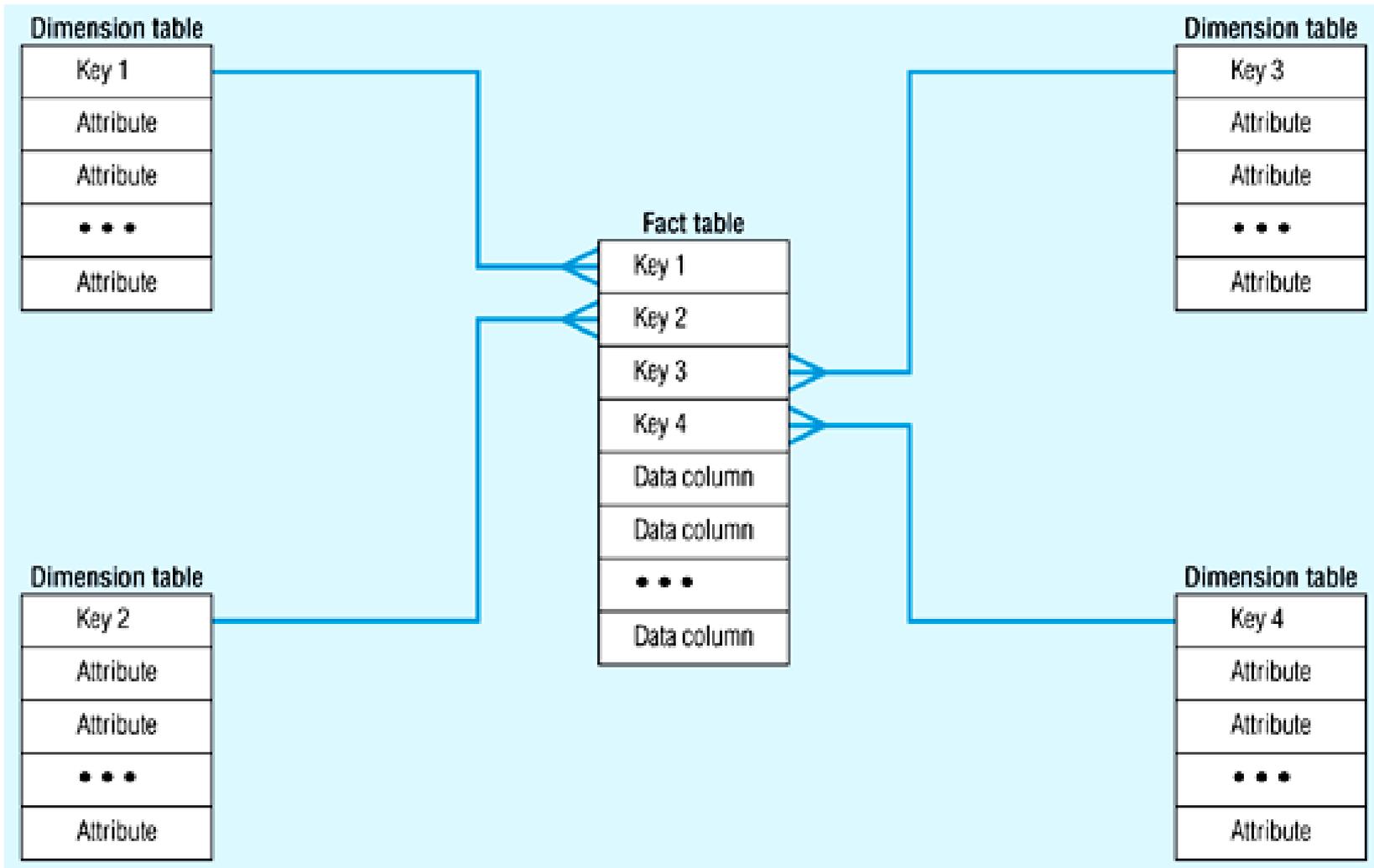
Esquema en estrella: Tabla de Dimensiones

- Se enlaza a la tabla de hechos (clave primaria única)
- Guarda los Atributos del negocio
- Más o menos constante los datos
- Contiene información textual descriptiva
- Filas anchas (muchos campos, incluso descriptivos)
- Tablas pequeñas (alrededor de un millón de filas)
- Ingresó a la tabla de hechos mediante una clave externa
- Fuertemente indexados

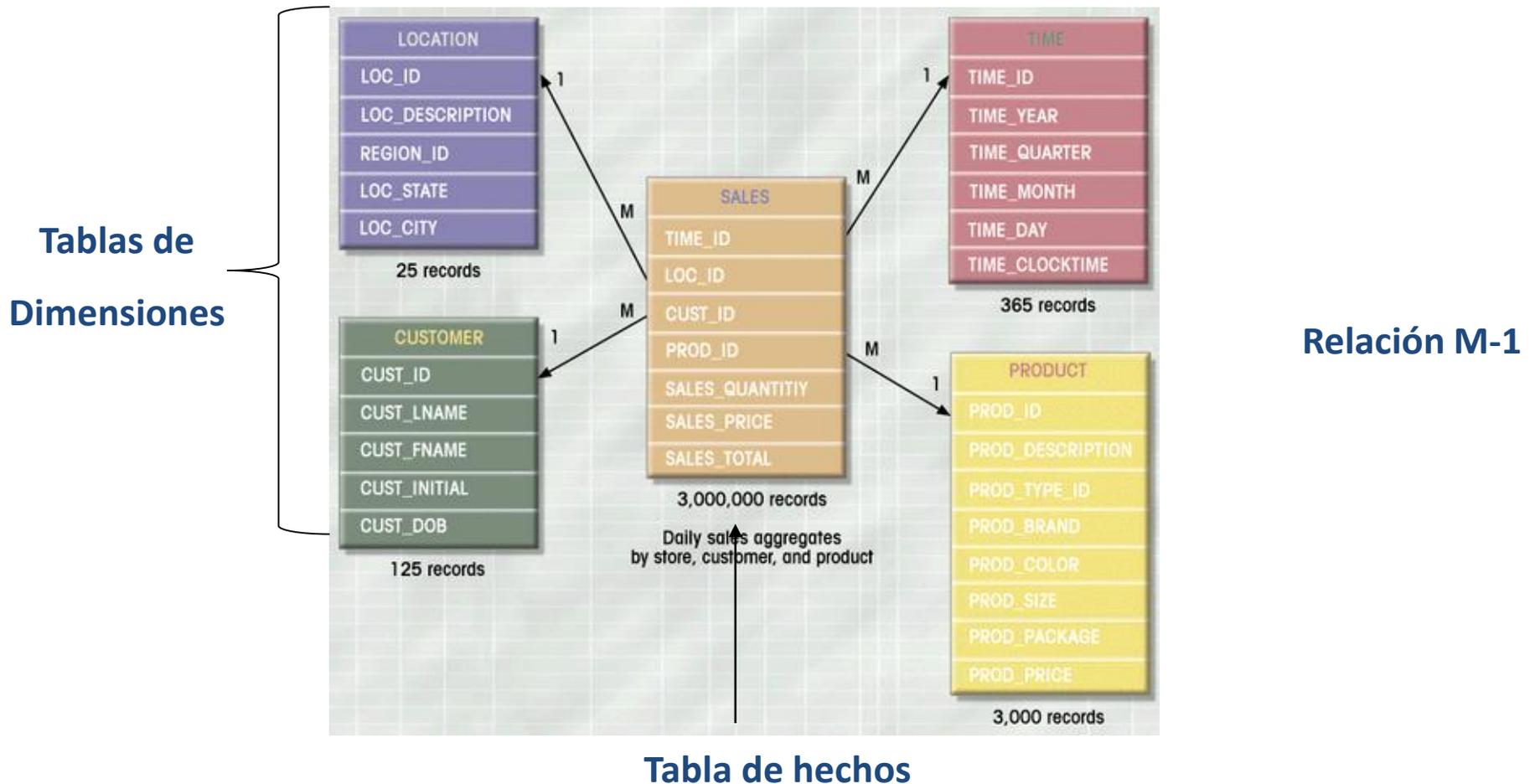
Esquema en estrella: Atributos

- Tablas de dimensiones contienen atributos
- Los atributos se utilizan para buscar, filtrar o clasificar los hechos
- Dimensiones proporcionan características descriptivas acerca de los hechos a través de sus atributos
- Debe definir los atributos comunes que se utilizará para reducir la búsqueda, agrupar información, o describir las dimensiones (por ejemplo, tiempo/lugar/producto)
- Sin límite matemático para el número de dimensiones (3D hace que sea fácil modelar)

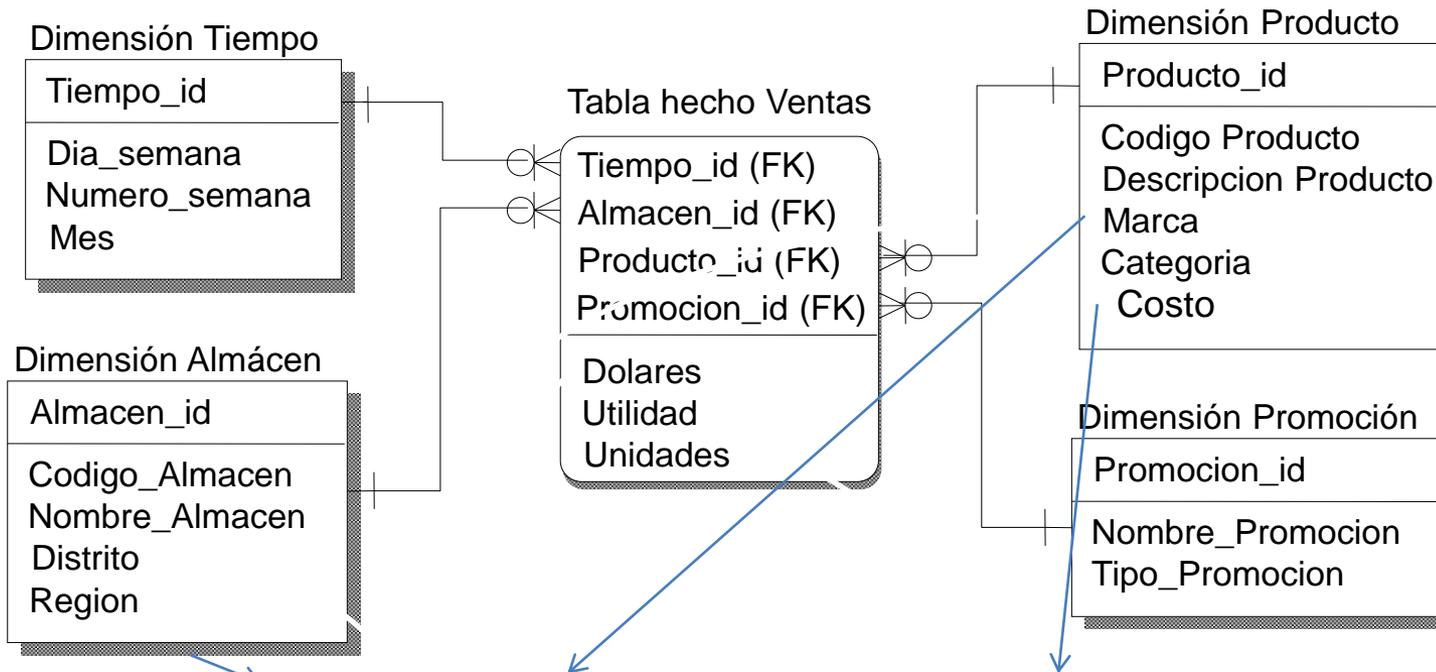
Esquema en estrella: Atributos



Ejemplo de esquema en estrella para ventas



Ejemplo de esquema en estrella para Ventas



Distrito	Marca	Total Dolares	Total Costo	Utilidad	Unidades
Atherton	Clean Fast	\$ 1,233	\$ 1,058	\$ 175	10
Atherton	More Power	\$ 2,239	\$ 2,200	\$ 39	2
Atherton	Zippy	\$ 848	\$ 650	\$ 198	4
Belmont	Clean Fast	\$ 2,097	\$ 1,848	\$ 249	6
Belmont	More Power	\$ 2,428	\$ 2,350	\$ 78	3
Belmont	Zippy	\$ 633	\$ 580	\$ 53	5

Conclusiones Esquema en Estrella

- Las tablas de hechos están relacionados a cada tabla de dimensión en una **relación Muchos a Uno**
- Tabla de hechos está **relacionado con muchas tablas** de dimensiones
- La clave principal de la tabla de hechos es compuesta de las **claves principales de las tablas de dimensiones**
- Cada tabla de hecho está diseñada para **responder a una(s) pregunta(s) específica(s) de IN**

Esquema Copo de nieve

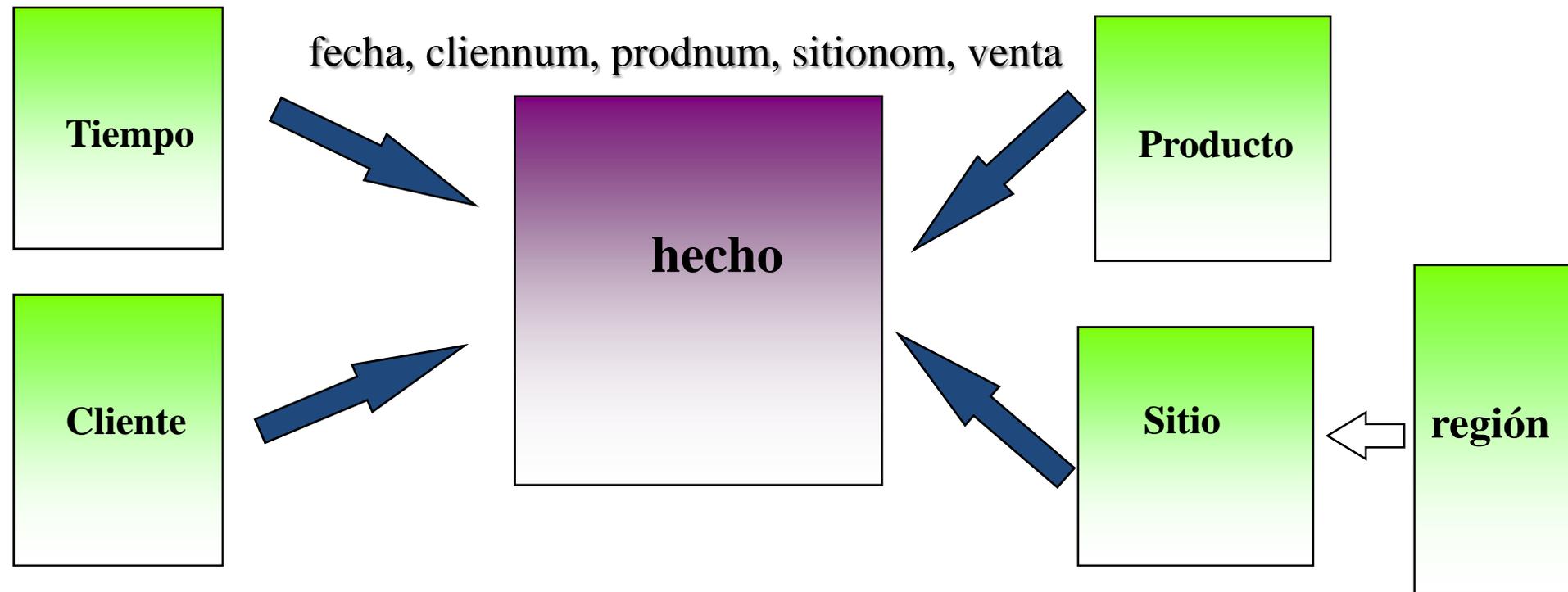
Un refinamiento del esquema en estrella donde alguna jerarquía dimensional se despliega en un conjunto de tablas de dimensiones más pequeñas

- Forma:
 - Esquema en estrella con dimensiones secundarias
 - Fácil de mantener y ahorra almacenamiento

Copo de nieve cuando las dimensiones tienen muchos atributos

Esquema Copo de nieve

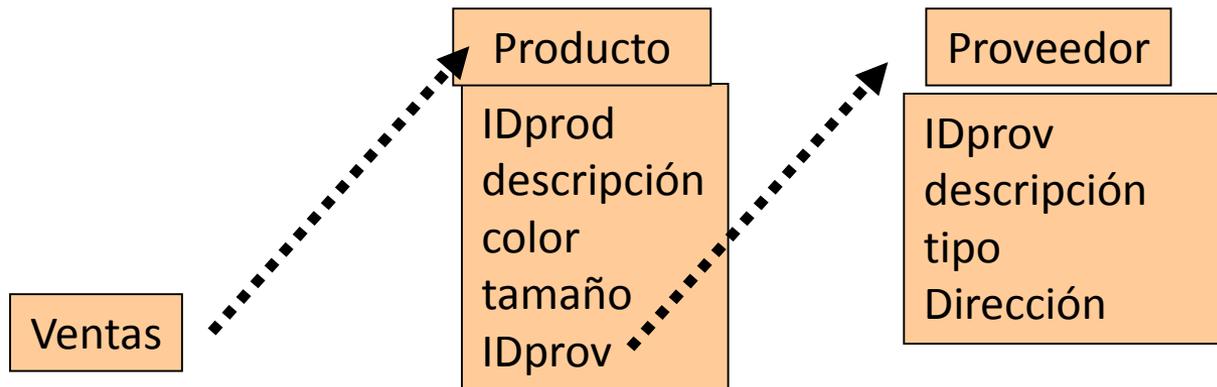
Representa jerarquía dimensional



Esquema Copo de nieve

Beneficios

- Evita la duplicación
- Conduce a las constelaciones (varias tablas de hechos con dimensiones compartidas)



Esquema Copo de nieve

Dimensión del almacén

Clave Almacen

Descripción
Ciudad
Estado
ID Distrito

ID Distrito

Desc_distrito
ID_Región

ID Región

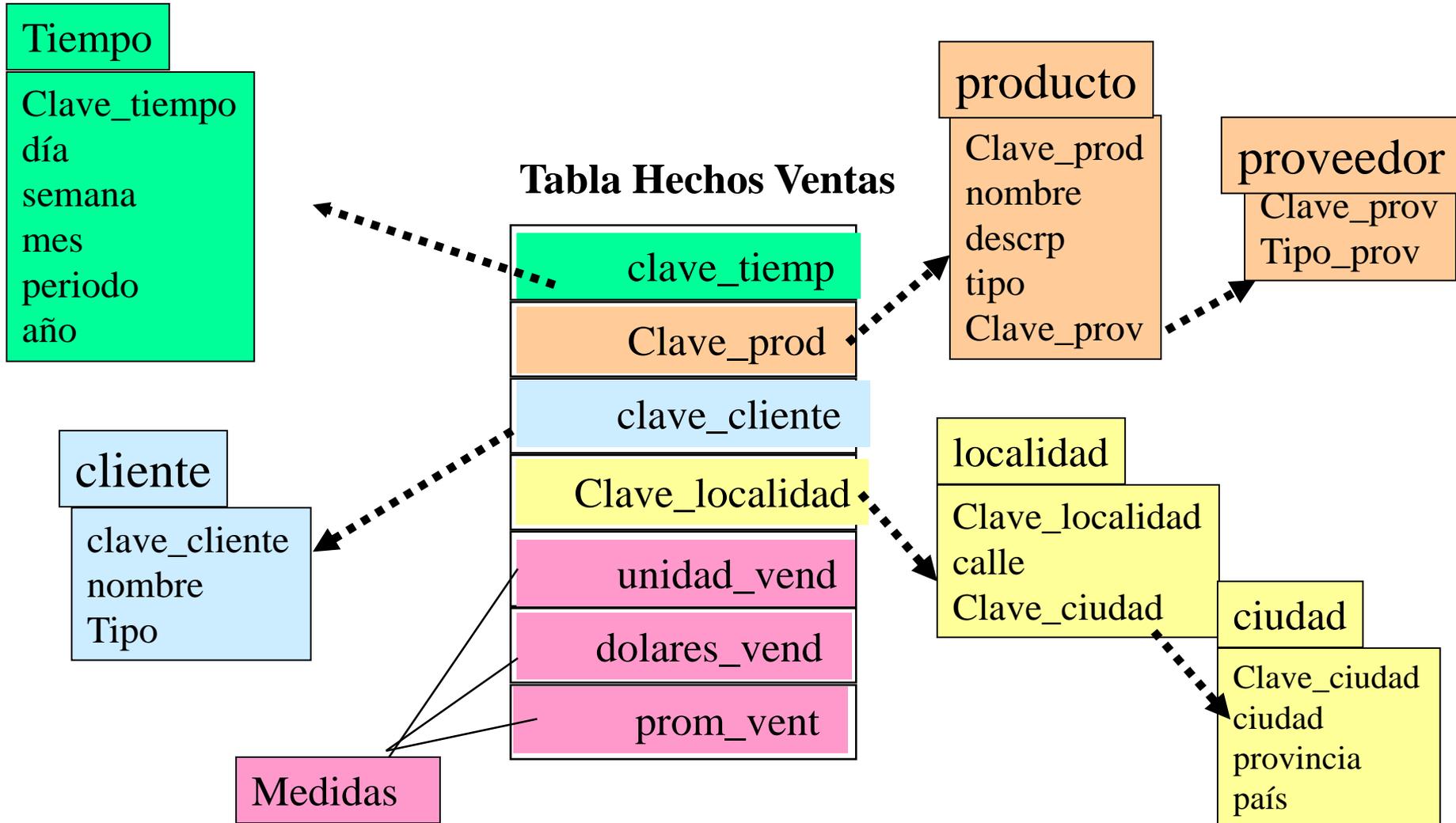
Descripción Región
Gerente Región..

Tabla hecho almacén

Clave Almacen
Clave producto
Clave Periodo

Monto
Unidades
Utilidad

Esquema Copo de nieve



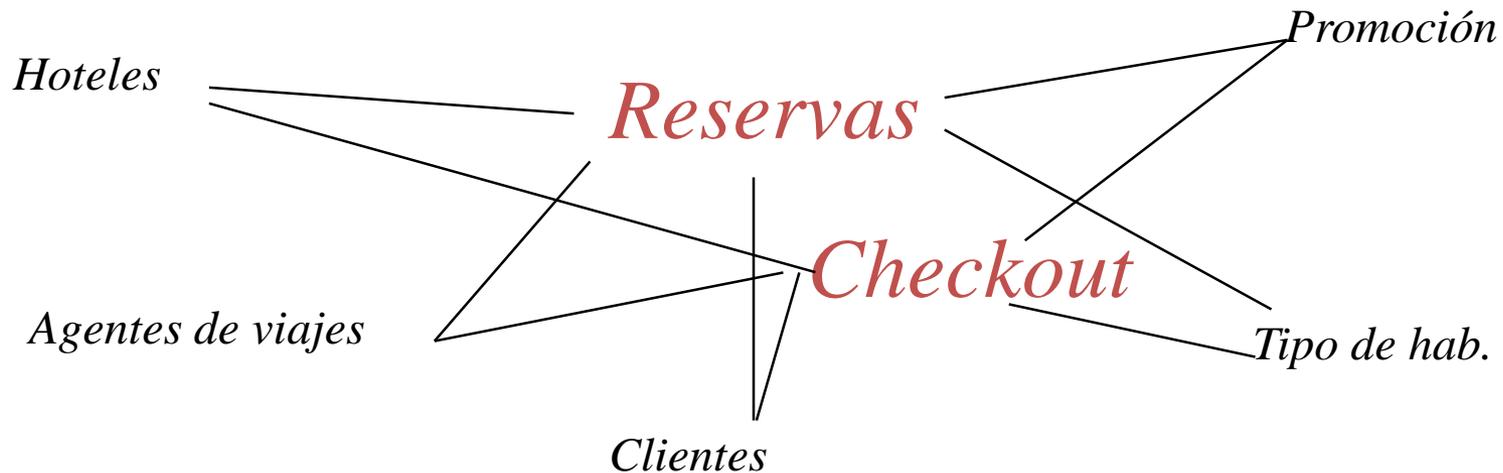
Esquema de Constelación de hechos

Varias tablas de hechos **comparten** tablas de dimensiones

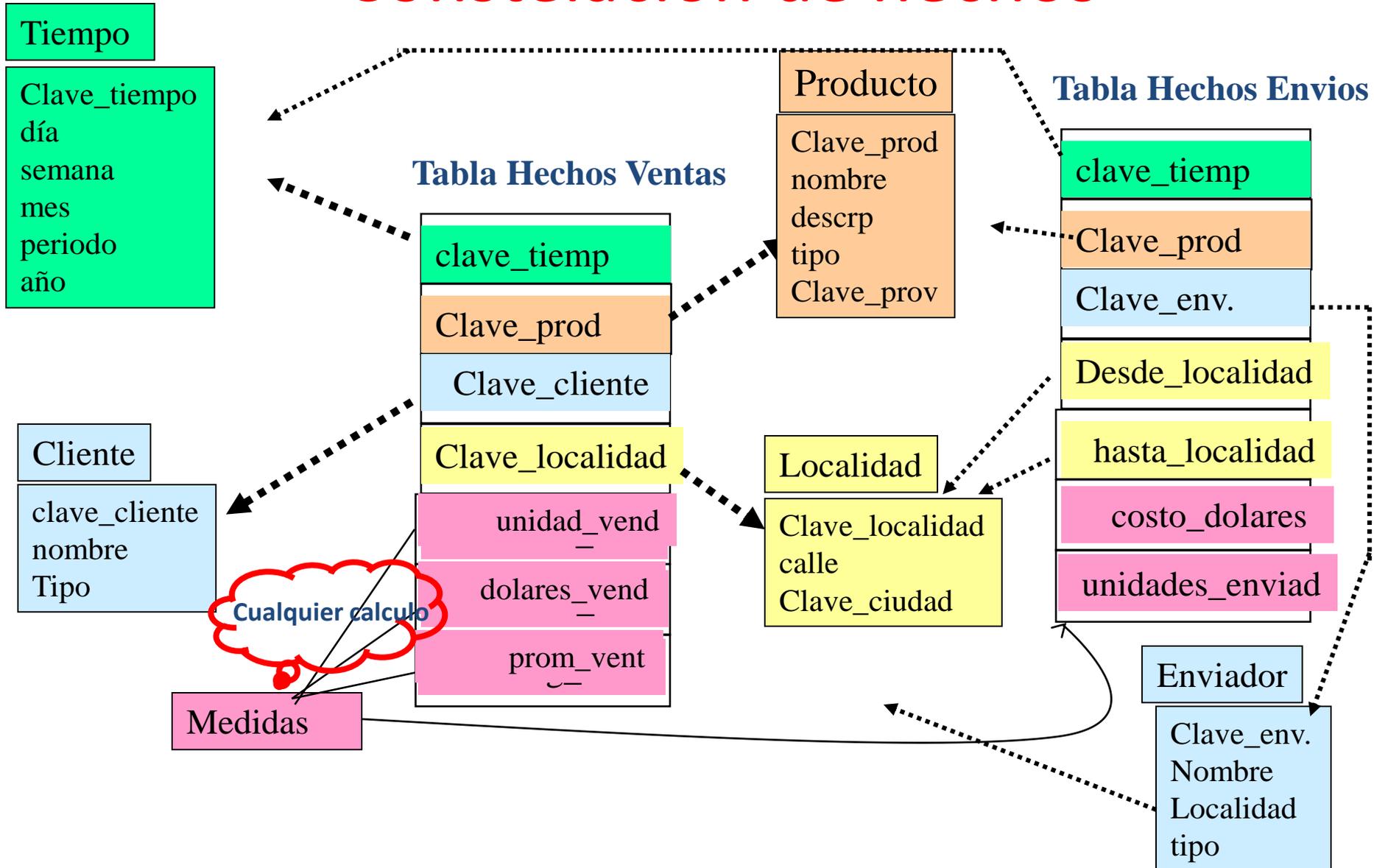
- vistos como una **colección de estrellas**, por lo tanto, llamados **esquema de galaxia o constelación de hecho**

Esquema de Constelación de hechos

Reservas y Checkout pueden compartir tablas de dimensiones en la industria hotelera



Esquema de Constelación de hechos



De las Tablas a los cubos de datos

- Un data warehouse se basa en un **modelo de datos multidimensional**
- Todo los datos se pueden ver en la forma de un **cubo de datos**
- Un cubo de datos permite ver múltiples dimensiones

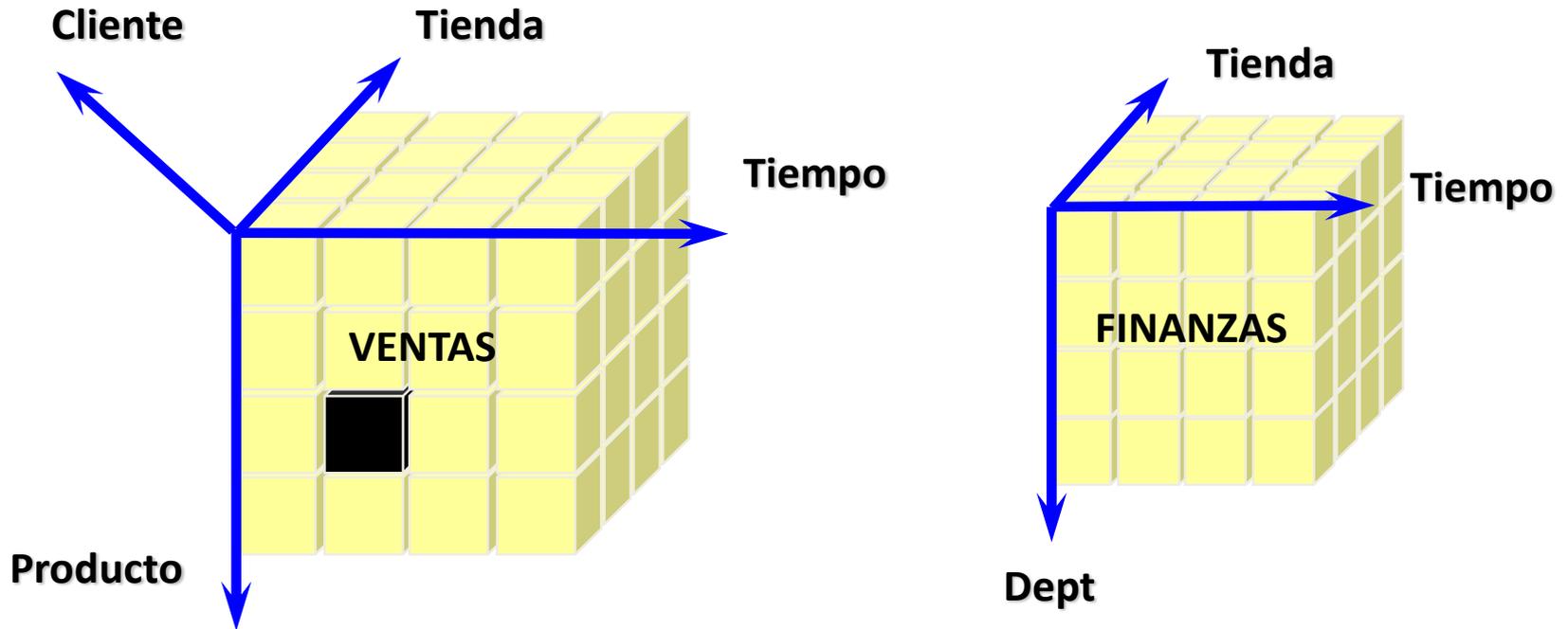
Un cubo n-D se llama un paralelepípedo.

Base de datos relacional

	Atributo 1 Nombre	Atributo 2 edad	Atributo 3 sexo	Atributo 4 No. Emp
Fila 1	Anderson	31	F	1001
Fila 2	Green	42	M	1007
Fila 3	Lee	22	M	1010
Fila 4	Ramos	32	F	1020

Tabla de empleados

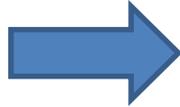
Modelo BD multidimensional



Los datos se encuentra en la intersección de las dimensiones

Dos dimensiones

City	Time	Total Revenue
Glasgow	Q1	29726
Glasgow	Q2	30443
Glasgow	Q3	30582
Glasgow	Q4	31390
London	Q1	43555
London	Q2	48244
London	Q3	56222
London	Q4	45632
Aberdeen	Q1	53210
Aberdeen	Q2	34567
Aberdeen	Q3	45677
Aberdeen	Q4	50056
.....
.....

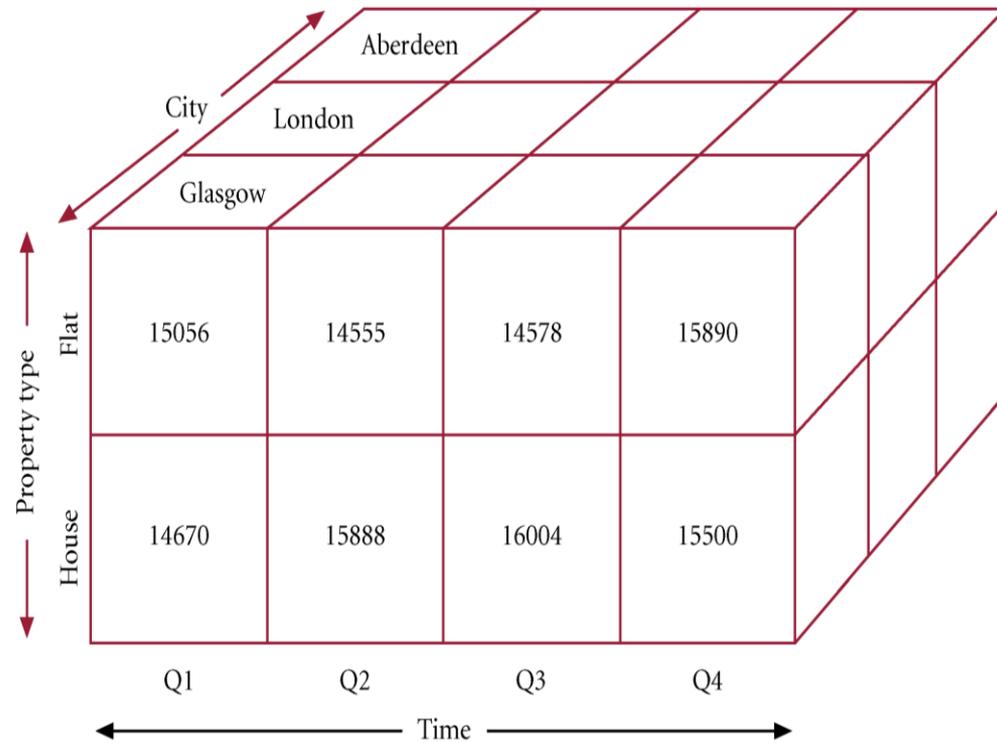


		City			
		Glasgow	London	Aberdeen
Time	City	Glasgow	London	Aberdeen
	Quarter	Q1	Q2	Q3	Q4
	Q1	29726	43555	53210
	Q2	30443	48244	34567
	Q3	30582	56222	45677
Q4	31390	45632	50056	

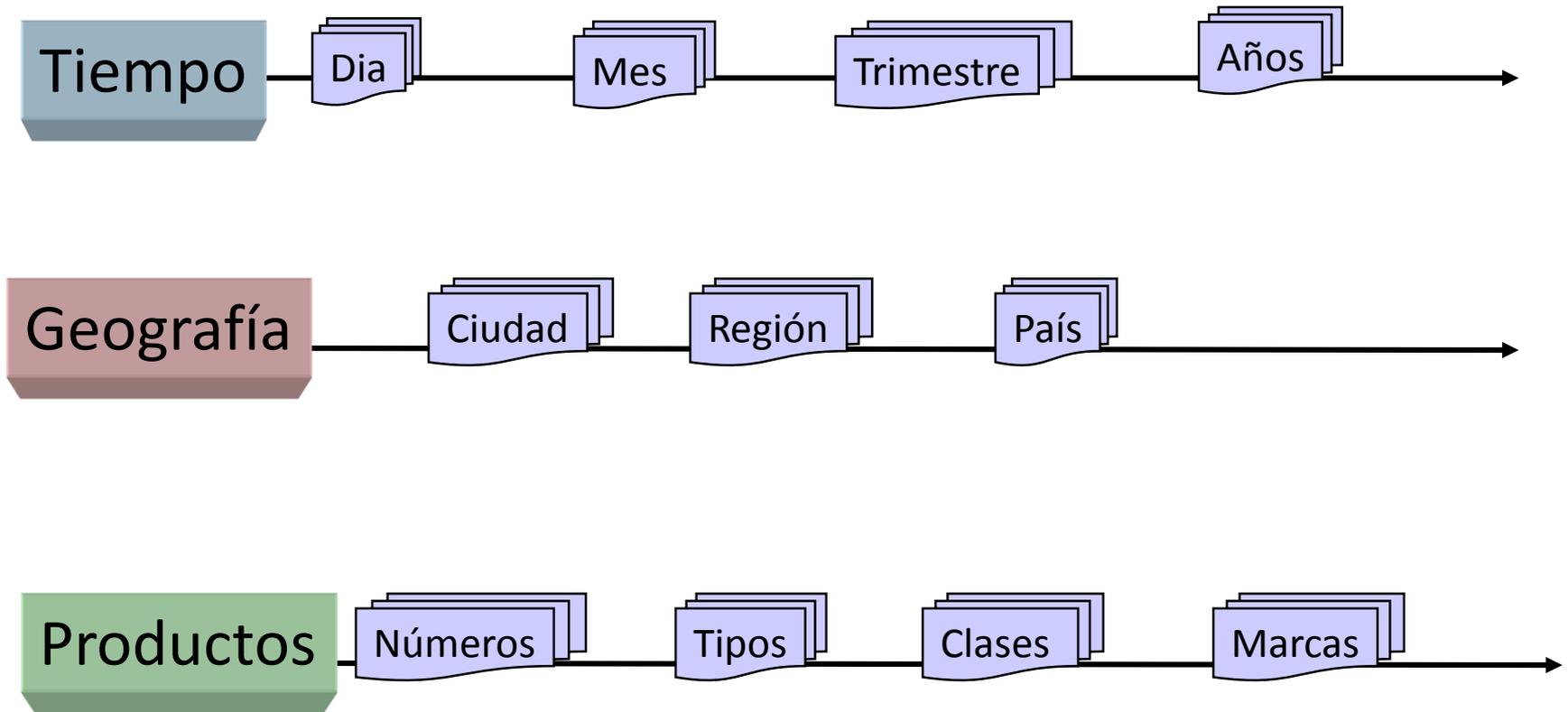
Tres dimensiones



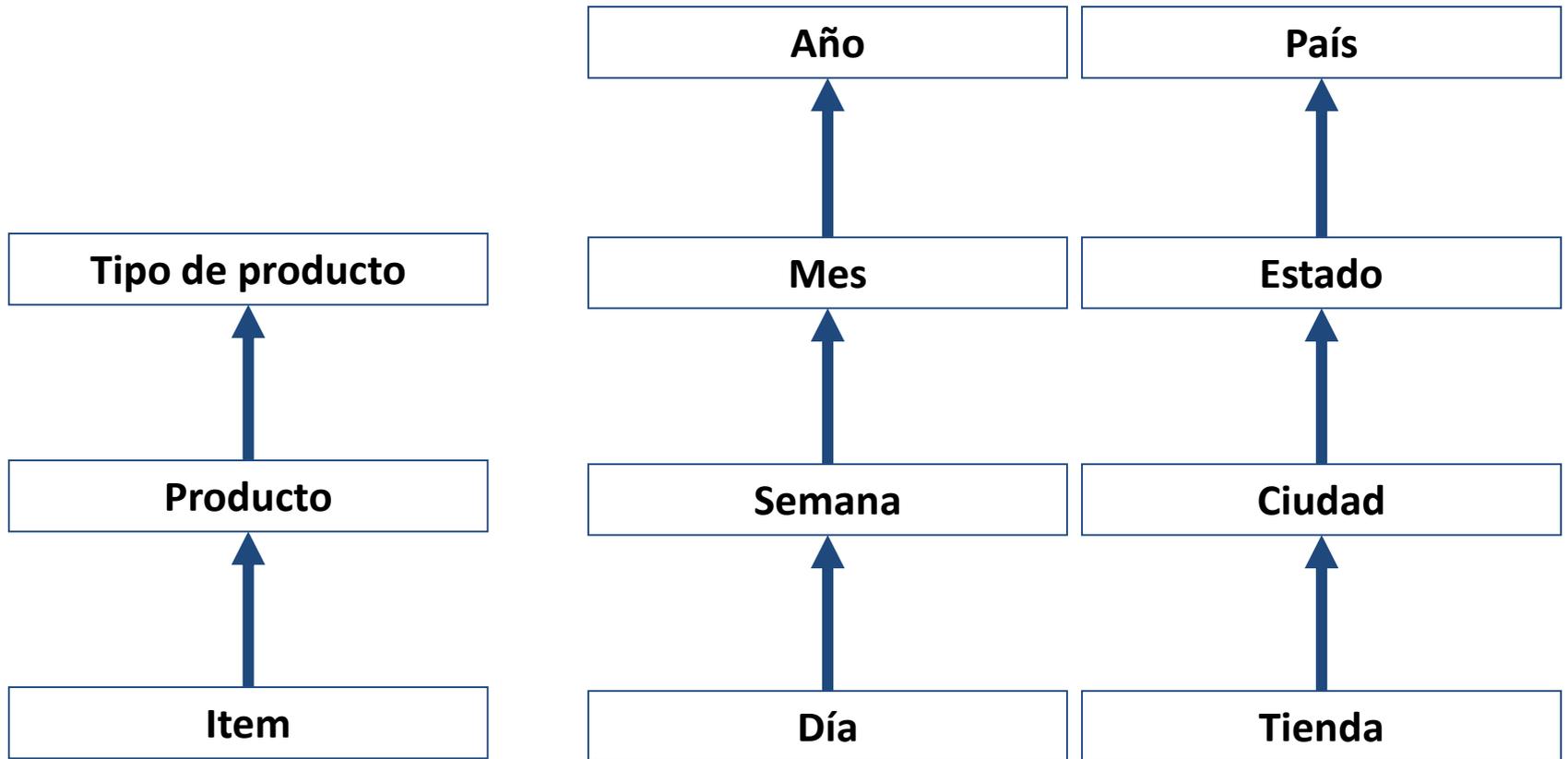
Property Type	City	Time	Total Revenue
Flat	Glasgow	Q1	15056
House	Glasgow	Q1	14670
Flat	Glasgow	Q2	14555
House	Glasgow	Q2	15888
Flat	Glasgow	Q3	14578
House	Glasgow	Q3	16004
Flat	Glasgow	Q4	15890
House	Glasgow	Q4	15500
Flat	London	Q1	19678
House	London	Q1	23877
Flat	London	Q2	19567
House	London	Q2	28677
.....
.....



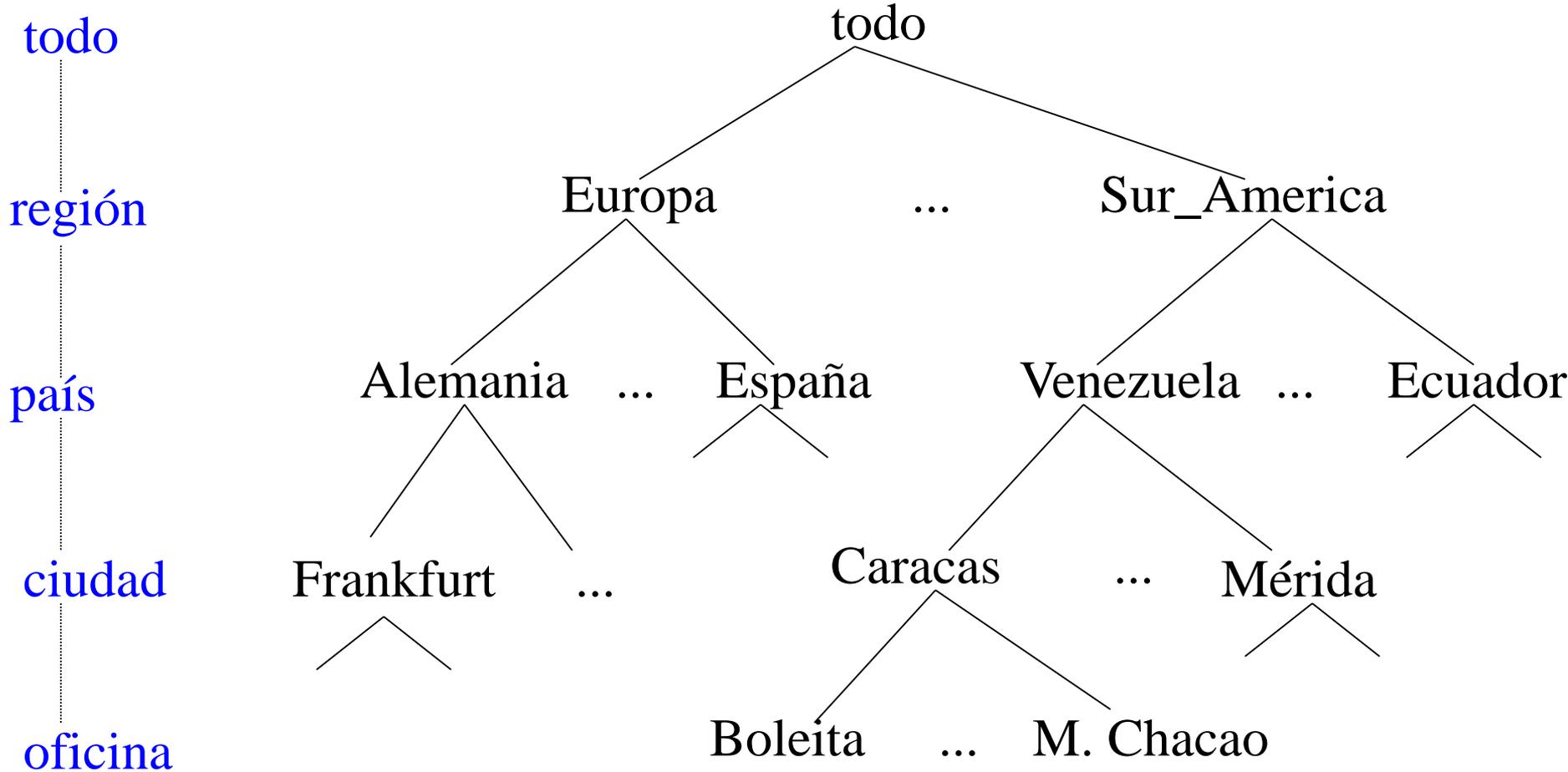
La granularidad de las dimensiones



Jerarquía Dimensional



Jerarquía Dimensional (localidad)



Jerarquía Dimensional

- jerarquía de esquema

día < mes < cuatrimestre < año

- Se pueden agrupar las jerarquías

{día 1 al 10}

{días} < 30

Las multidimensiones

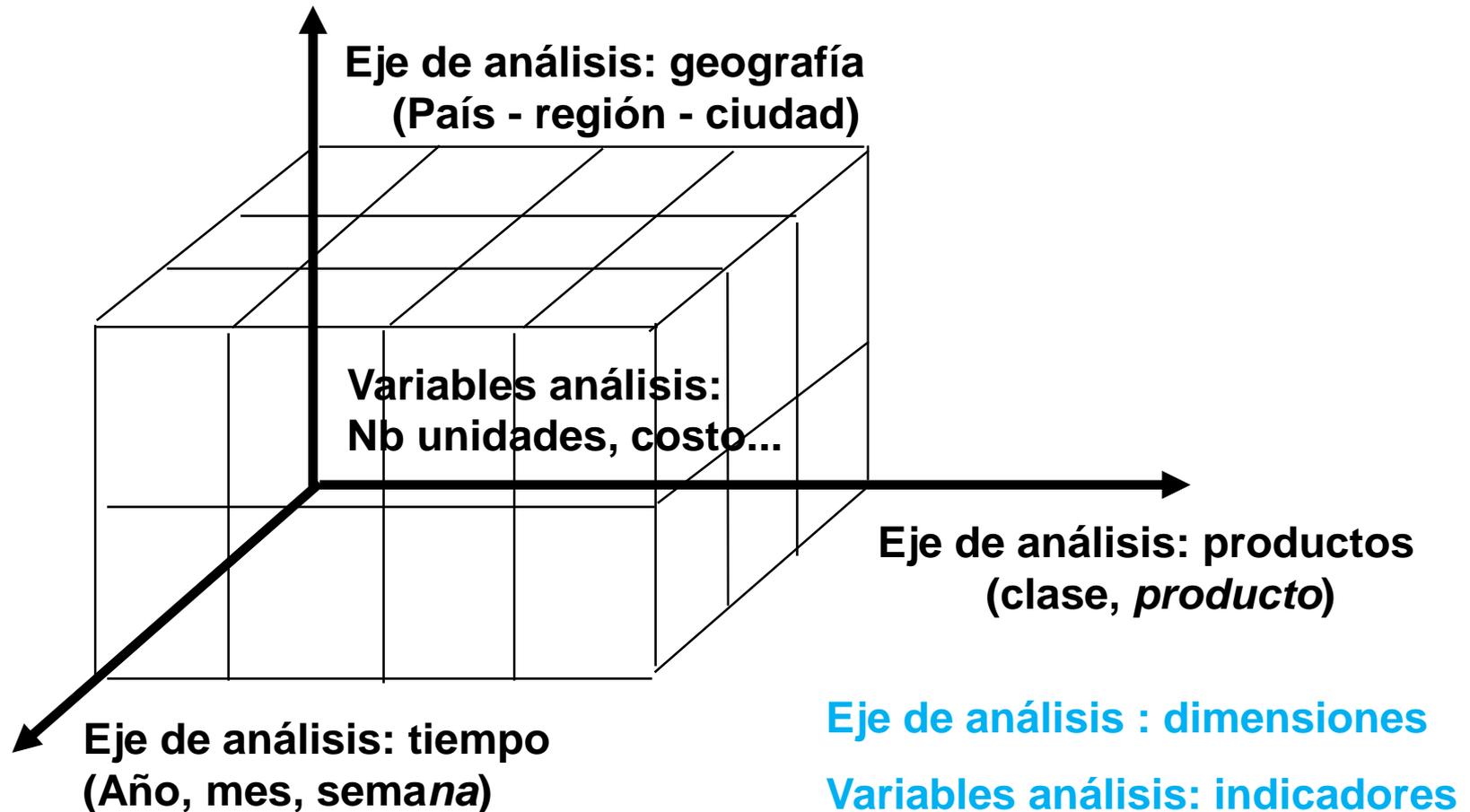
- **Dimensiones:**

- Tiempo
- Geografía
- Productos
- Clientes
- Canales de ventas.....

- **Indicadores:**

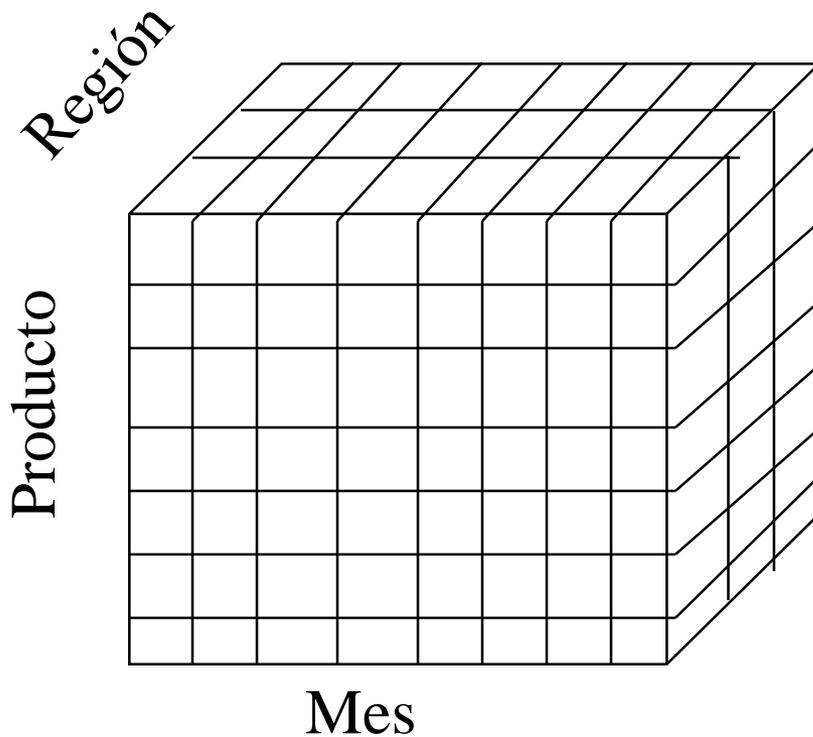
- Número de unidades vendidas
- Costo

Cubo de dato, dimensiones e indicadores



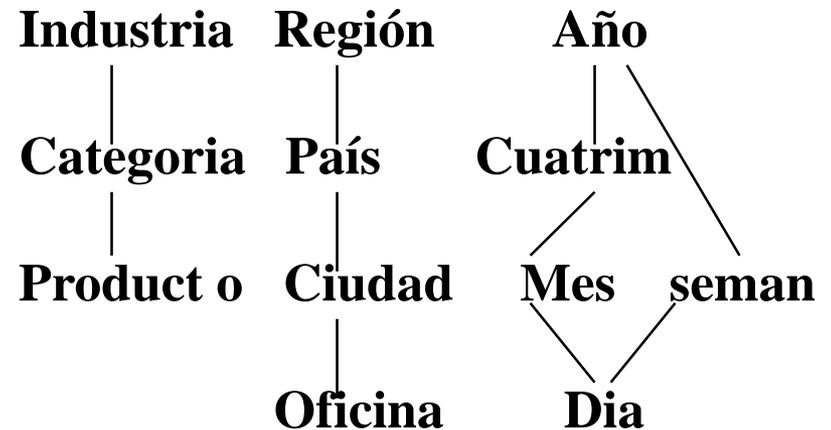
Datos Multidimensionales e e indicadores

El volumen de ventas en función del producto, el mes, y el área



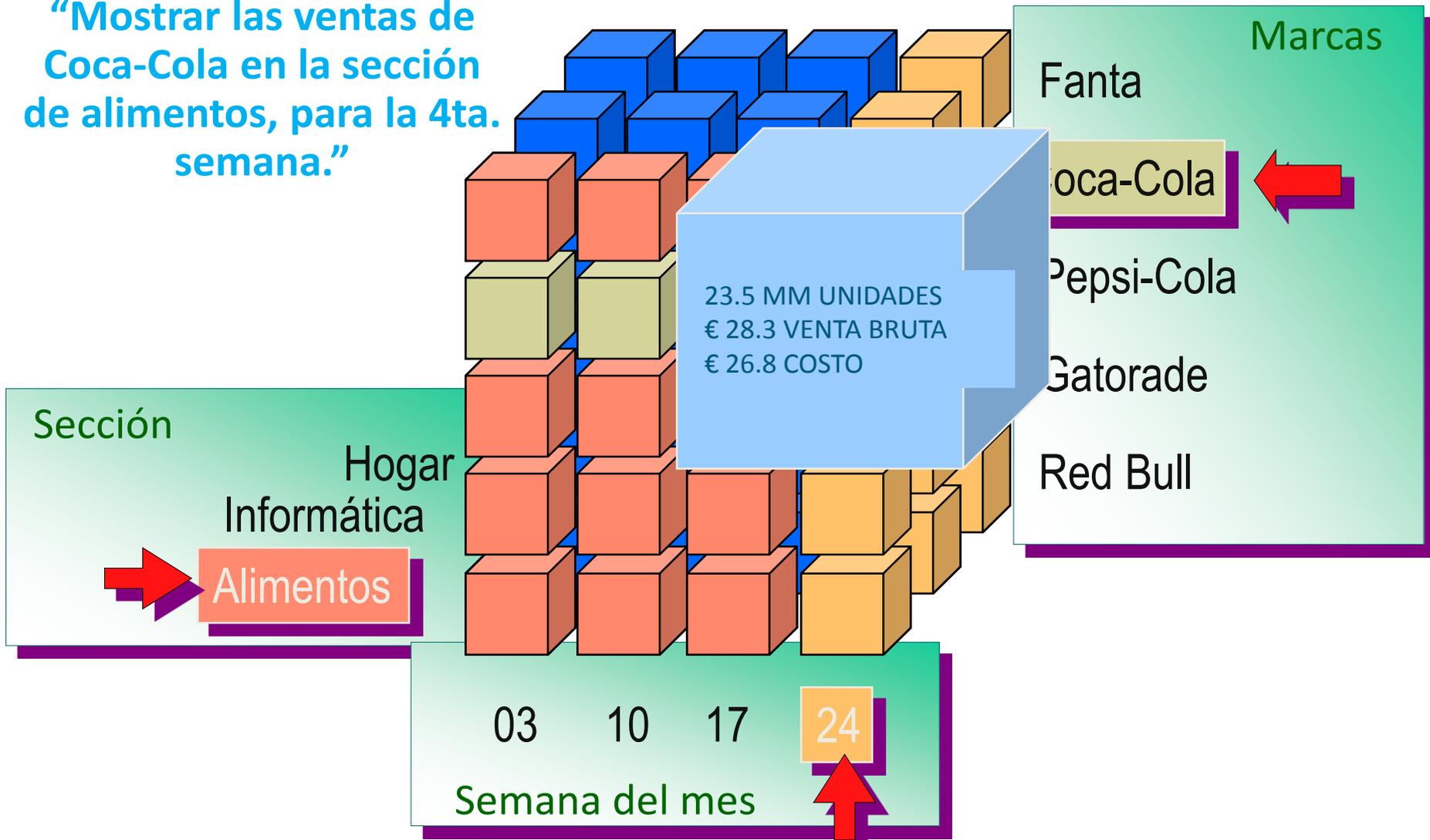
Dimensiones: Producto, Localidad, Tiempo

Caminos jerarquicos

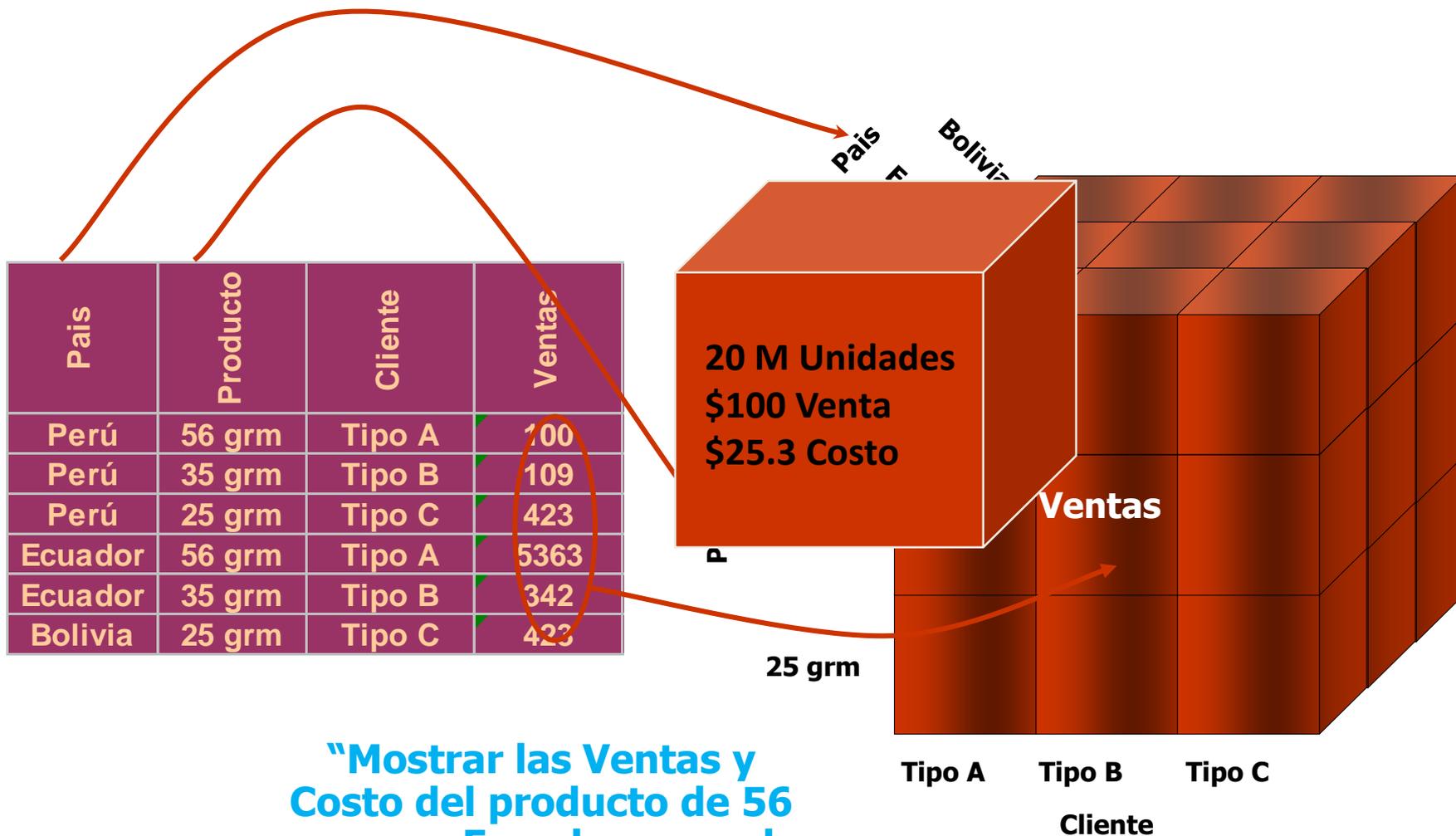


Cubo Multidimensional e e indicadores

“Mostrar las ventas de Coca-Cola en la sección de alimentos, para la 4ta. semana.”

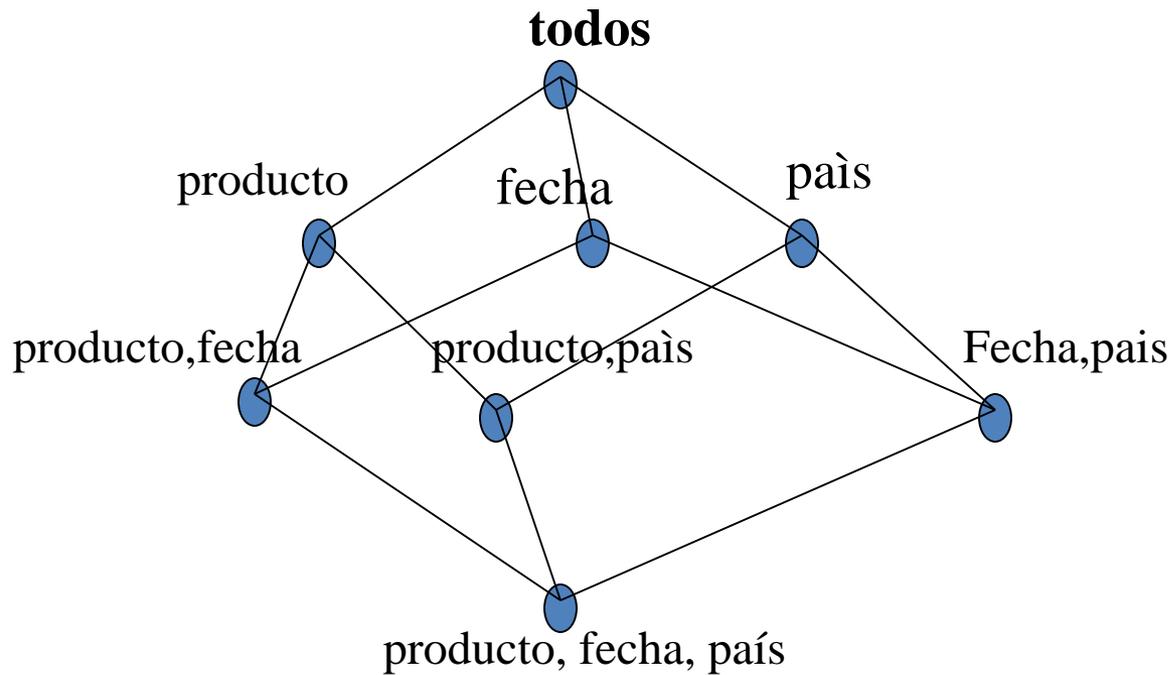


Cubo Multidimensional e indicadores



“Mostrar las Ventas y Costo del producto de 56 gm en Ecuador para el Tipo de Cliente A”

Cuboides correspondientes al Cubo



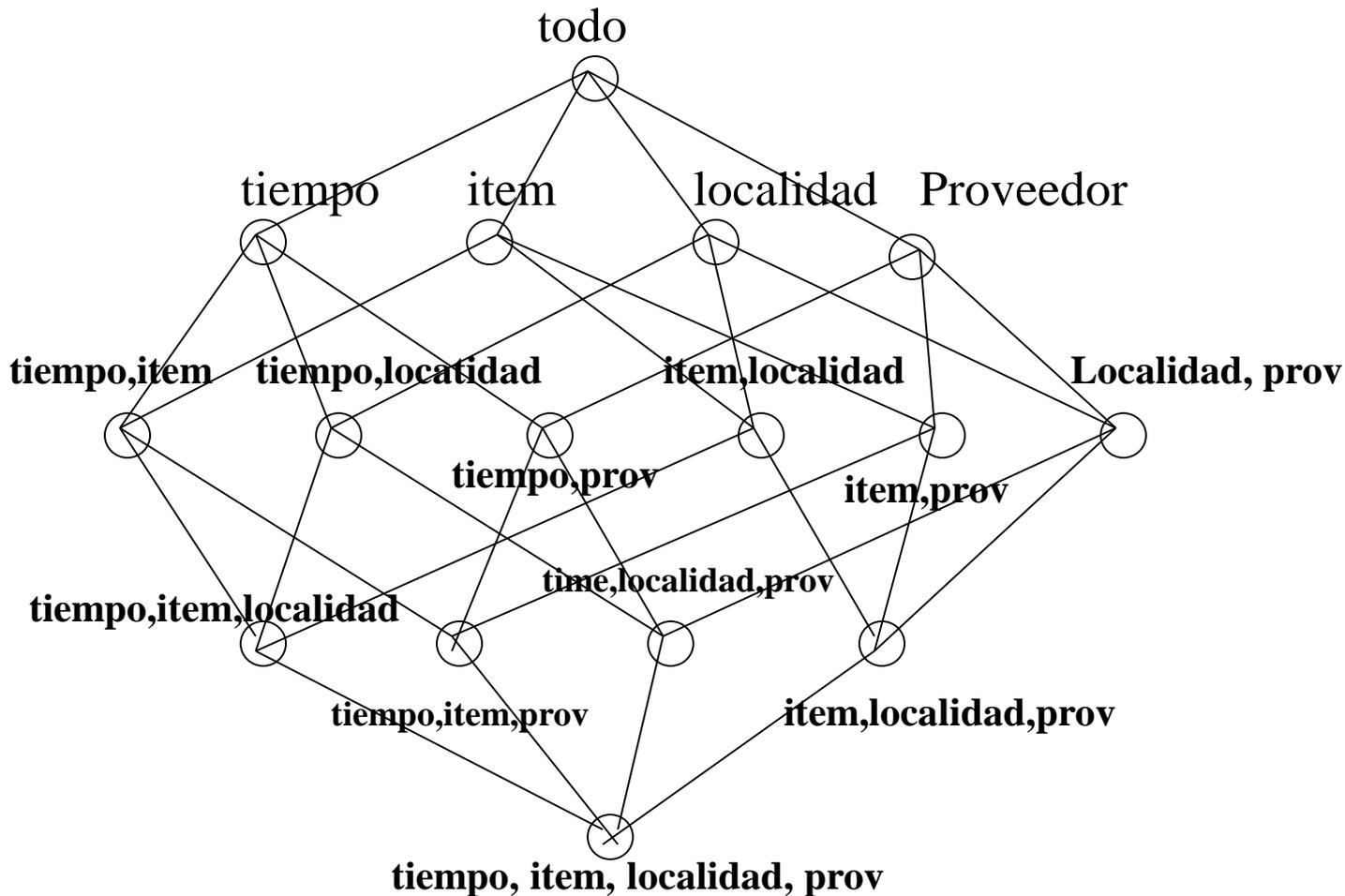
0-D cuboide

1-D cuboides

2-D cuboides

3-D cuboides

Cuboides correspondientes al Cubo



0-D cuboide

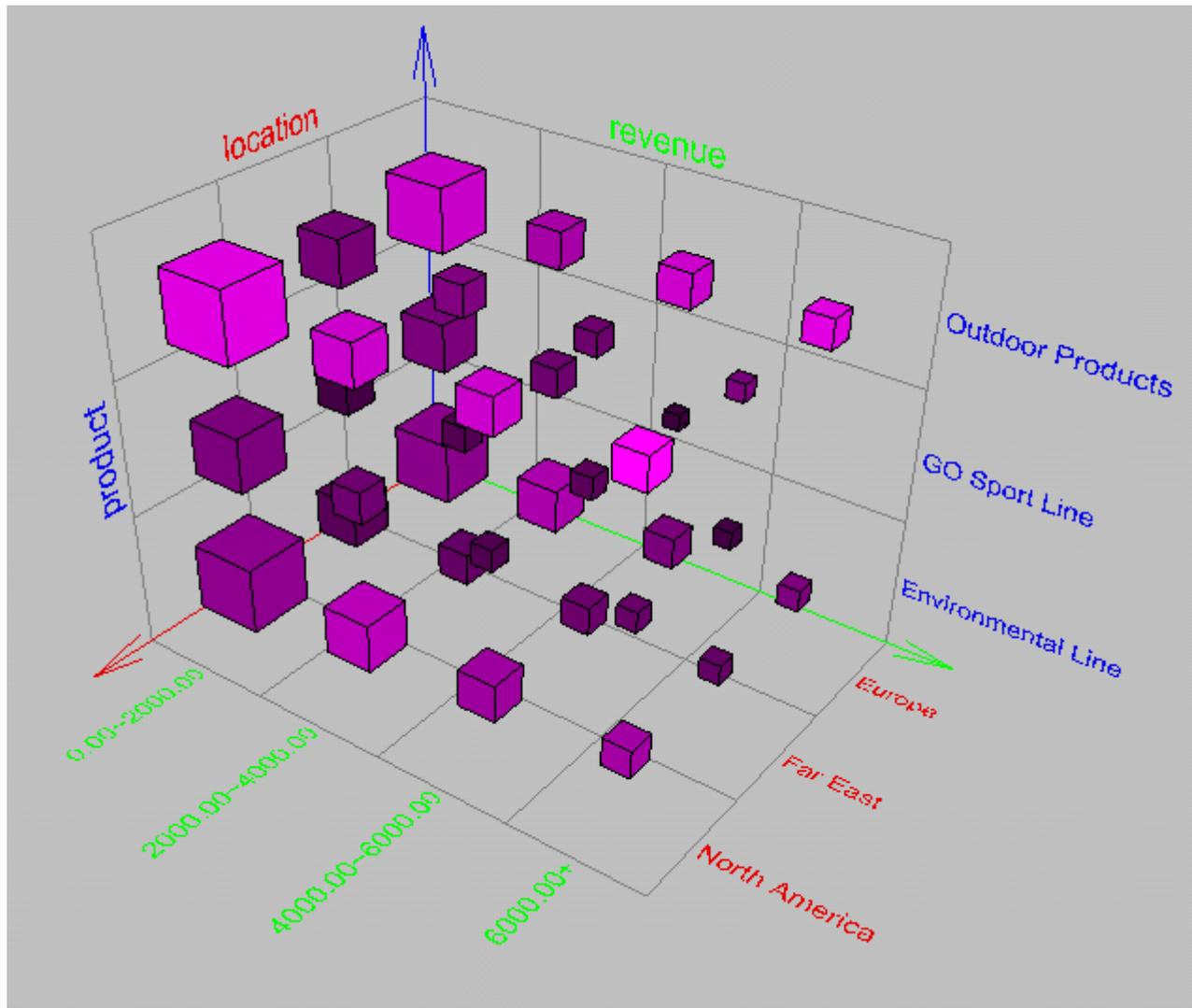
1-D cuboides

2-D cuboides

3-D cuboides

4-D cuboides

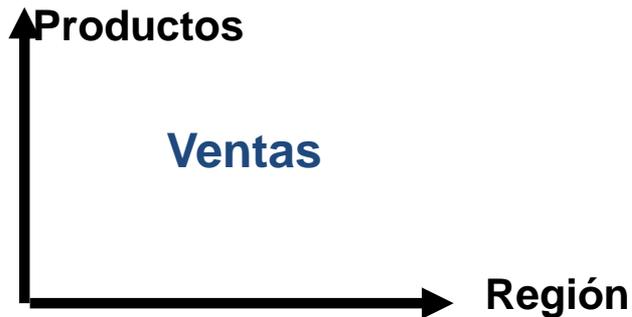
Navegar por un cubo de datos



- Visualización
- OLAP

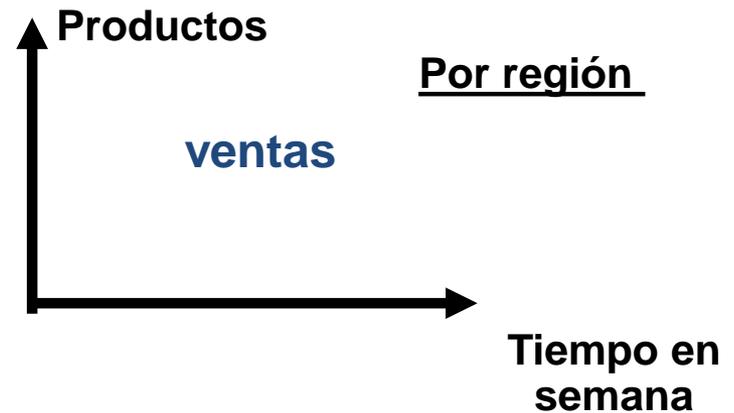
Navegar por un cubo de datos

2 dimensiones



3 dimensiones

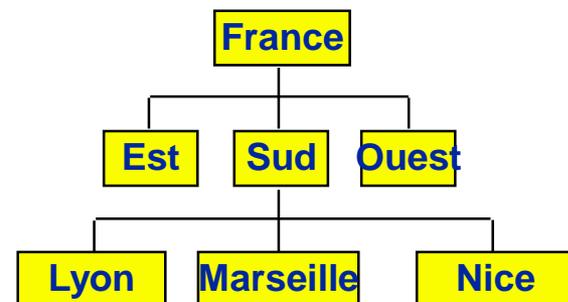
Copo de nieve



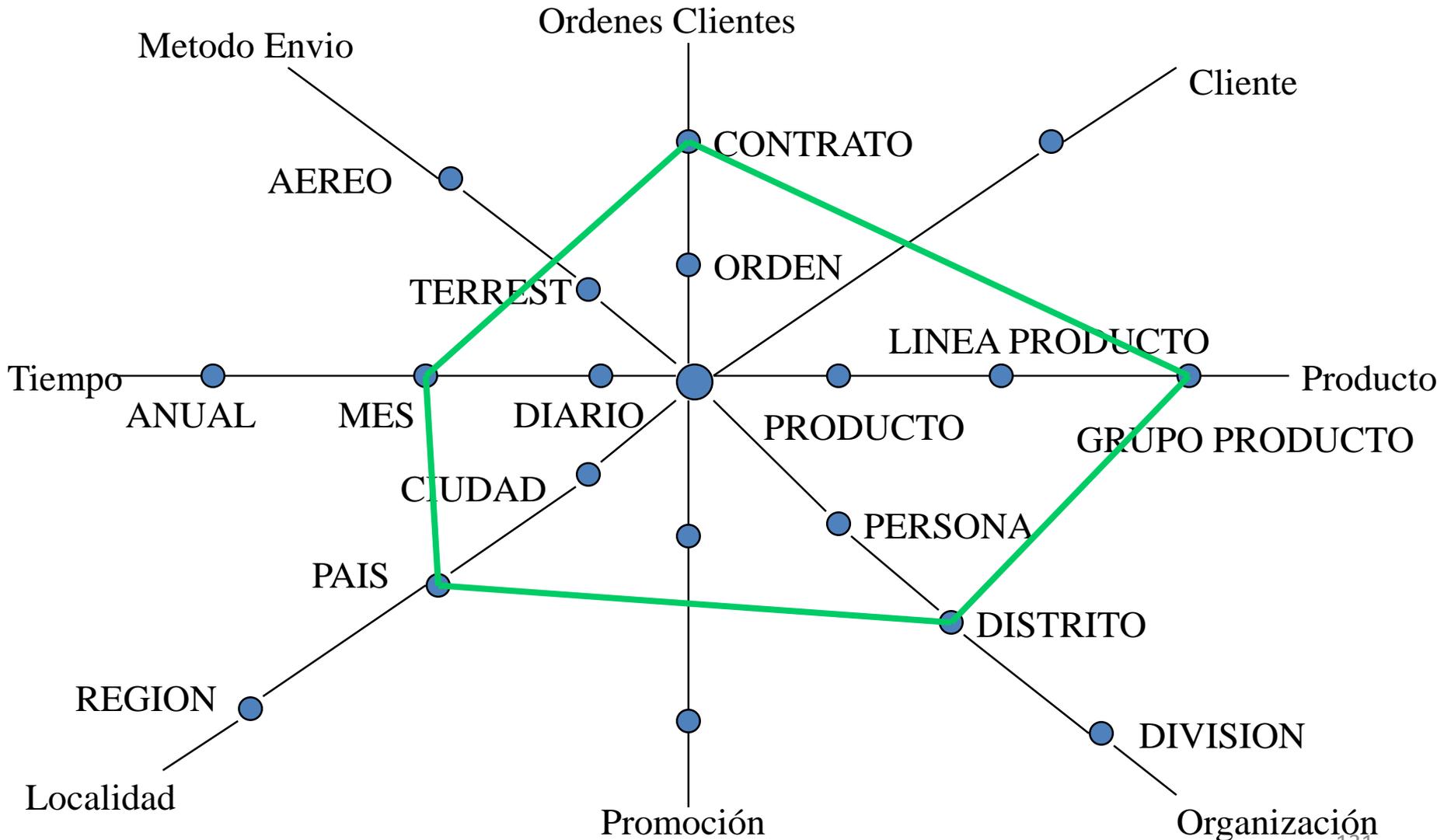
1 dimensión



Zoom a una dimensión



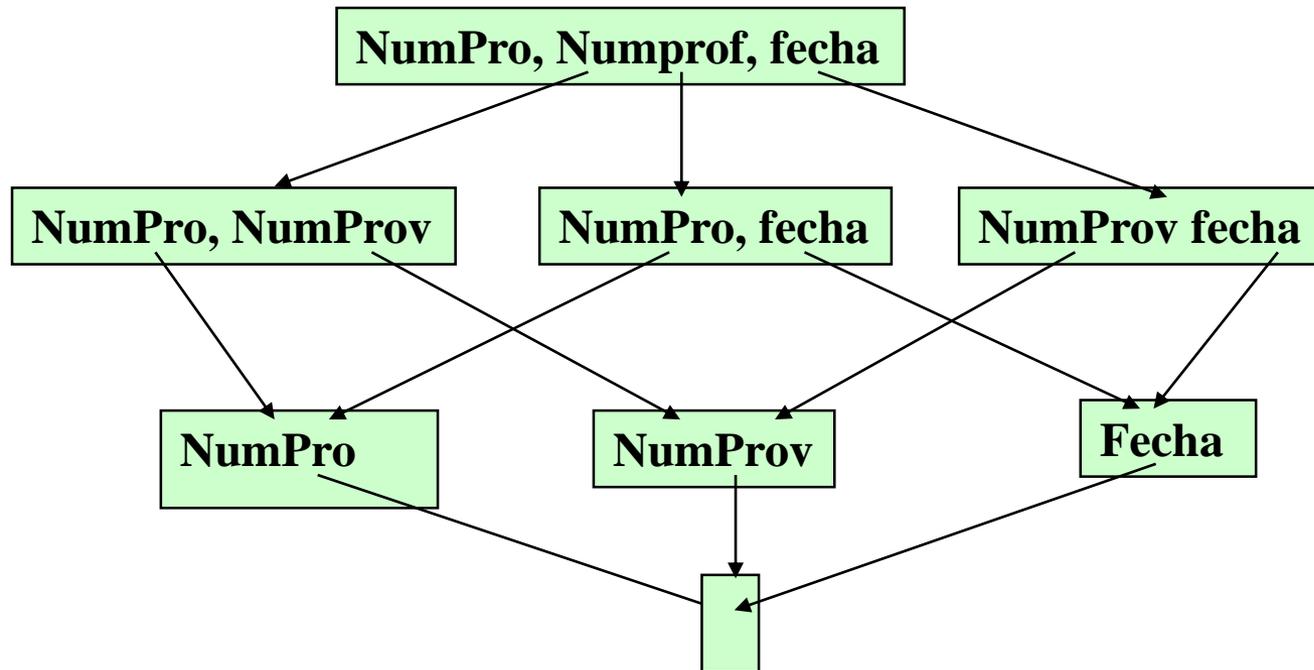
Modelo de consulta



Navegar por un cubo de datos

- Un cubo se puede rotar, agrupar, etc.

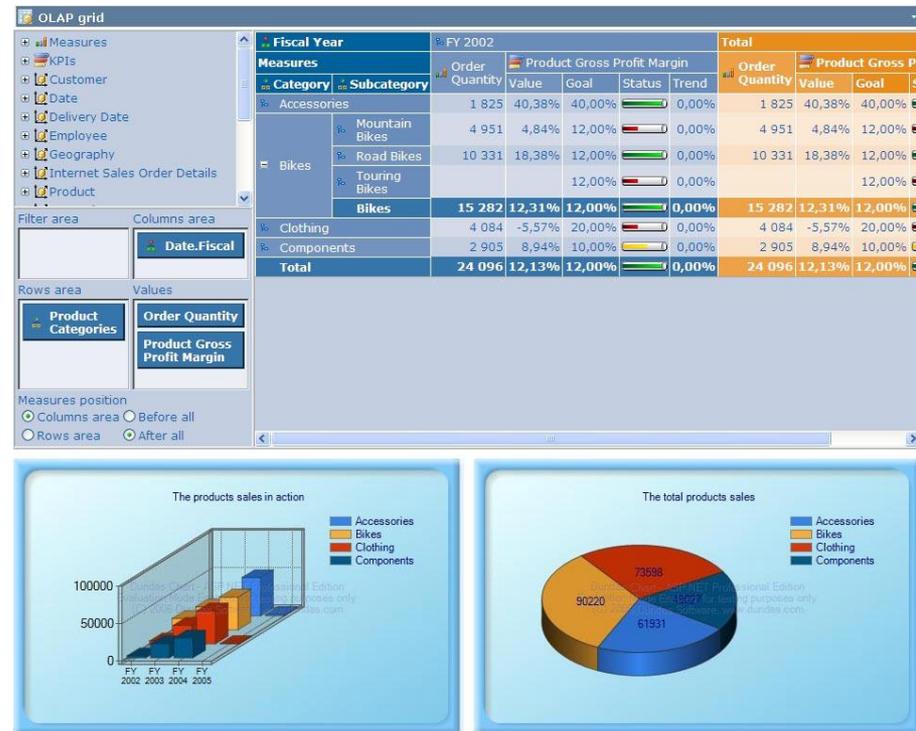
Se obtienen retículas de puntos de vista



Herramientas para explotación del Datawarehouse

Análisis multidimensional (OLAP online analytical processing)

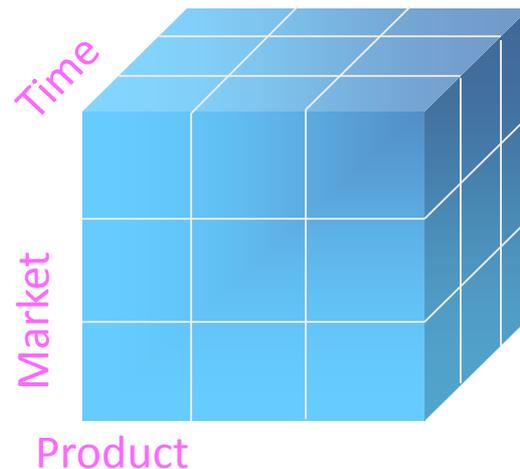
Facilitan el análisis de datos a través de dimensiones y jerarquías, utilizando consultas rápidas predefinidas



On-Line Analytical Processing (OLAP)

Idea básica: los usuarios deben poder manipular los modelos de datos organizacionales a través de muchas dimensiones para comprender que se está ocurriendo.

- Los datos utilizados en OLAP deberían estar en la forma de un cubo multidimensional.



Herramienta Multidimensional Especializada

- **Beneficios:**
 - Acceso rápido a grandes volúmenes de datos
 - Bibliotecas extensas de funciones complejas de análisis
 - Capacidades de modelado y predicción
 - Puede acceder a las estructuras de bases de datos multidimensionales y relacionales

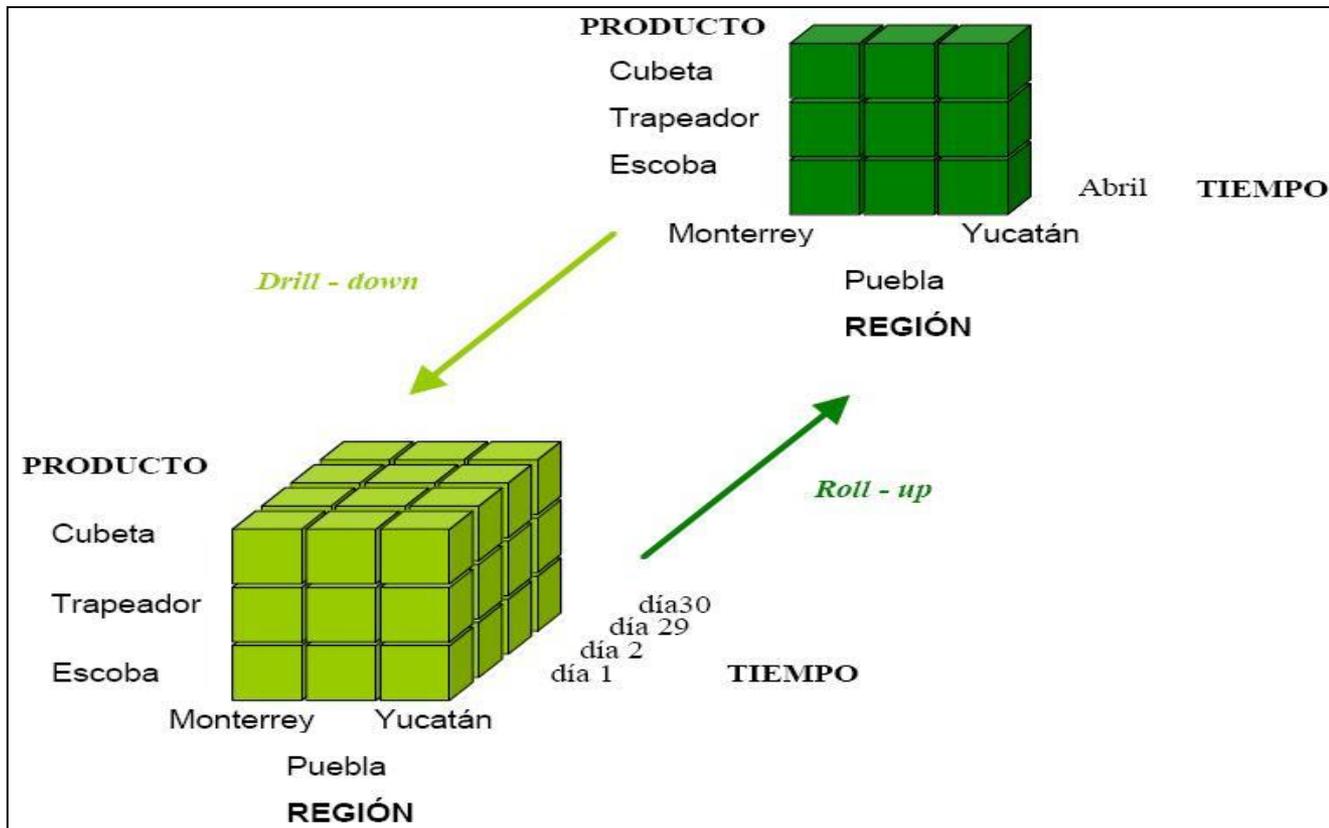
Arquitecturas OLAP

- **OLAP Relacional (ROLAP)**
 - Usa un esquema relacional para manejar la navegación y administrar los datos consolidados
 - Incluye agregación
 - Gran escalabilidad
- **OLAP Multidimensional (MOLAP)**
 - Almacenamiento con técnicas multidimensionales
 - Acceso rápido a datos pre-calculados previamente
- **OLAP Híbrido (HOLAP)**
 - Bajo nivel MOLAP, Alto nivel ROLAP
- **Motores de BD especializados**
 - Manejan consultas especializadas (como las de SQL) con esquemas estrella o copo de nieve

Operaciones clásicas OLAP

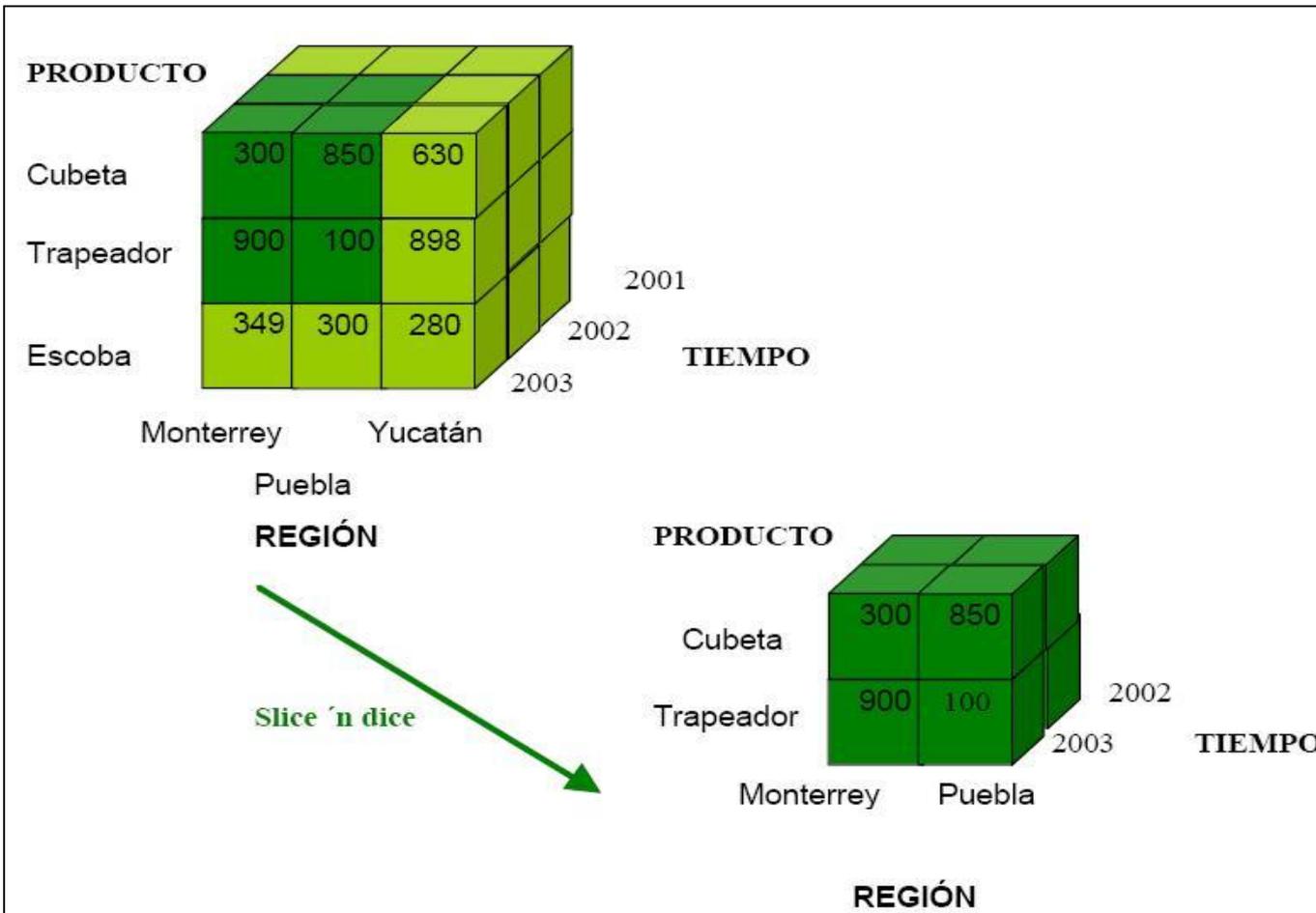
Roll up (drill-up): agrega medidas que van de un nivel N_i a un nivel mas general N_j de una dimensión.

Drill down (roll down): es la operación inversa. A partir de un nivel superior este operador permitir bajar de nivel.



Operaciones clásicas OLAP

Slice and dice: permite restringir los valores asociados a una o varias dimensiones del cubo, es decir, toma un subconjunto de dimensiones y de niveles seleccionados del DW.



Otras operaciones

drill across

navegar a través de más de una tabla de hechos

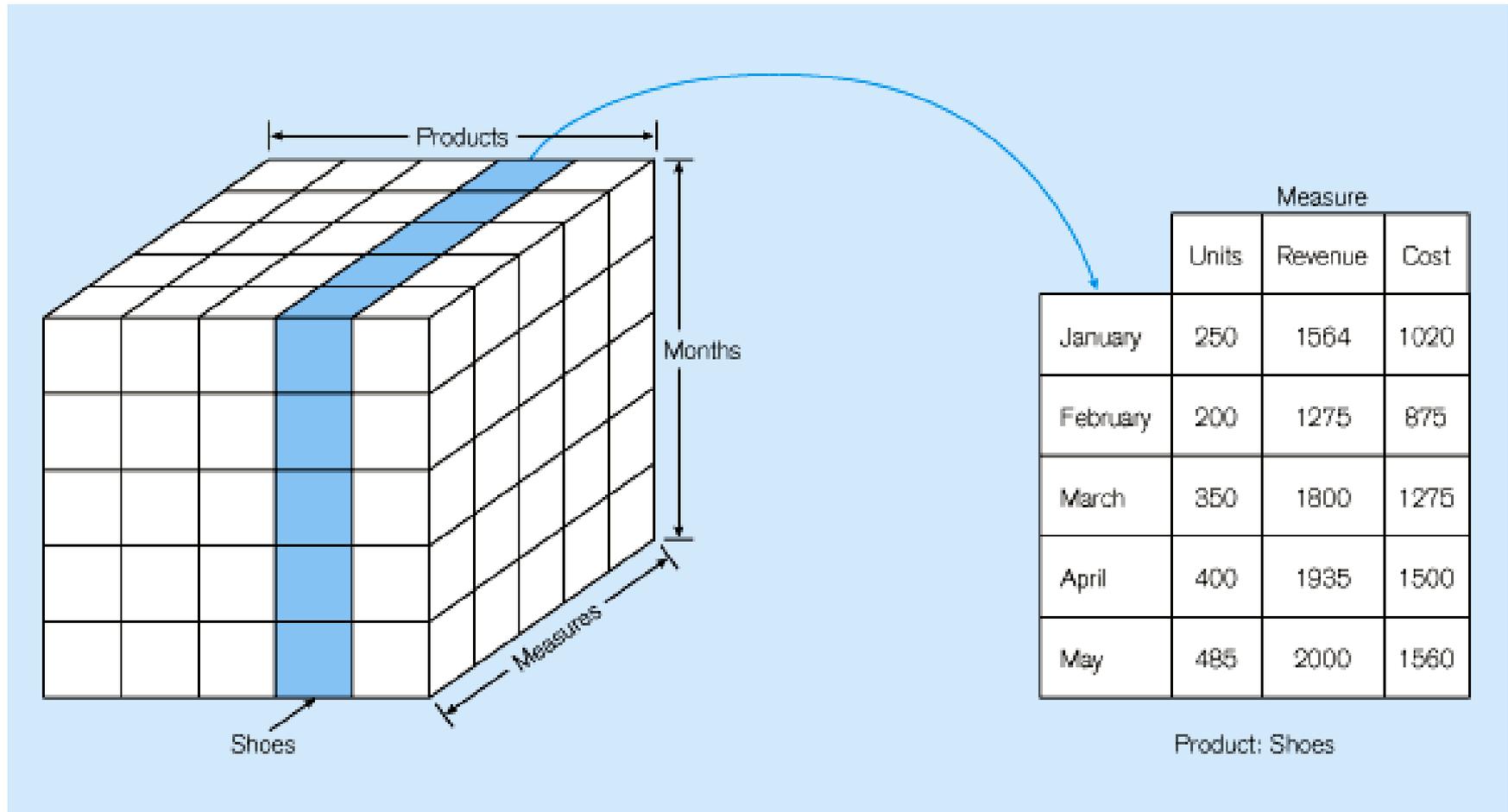
drill through

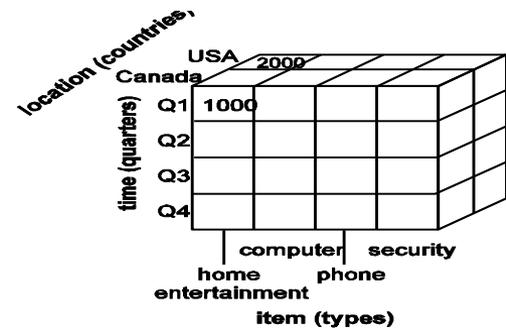
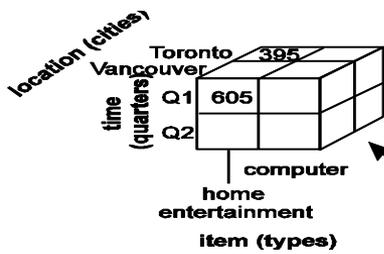
navegar a través del nivel inferior del cubo a tablas relacionales

Pivote (rotar)

Rotar el cubo

Cortando/rebanando un cubo de datos



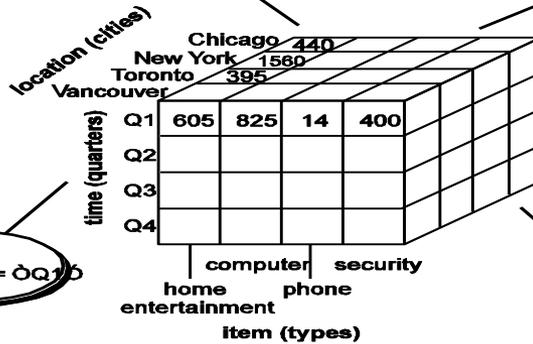


Slice and dice

dice for (location = 'Toronto' or 'Vancouver') and (time = 'Q1' or 'Q2') and (item = 'home entertainment' or 'computer')

roll-up on location (from cities to countries)

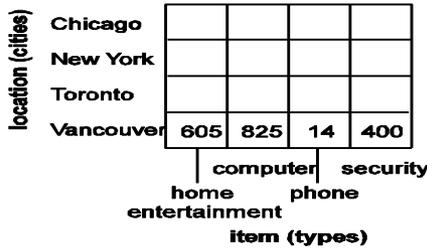
Roll up



slice for time = 'Q1'

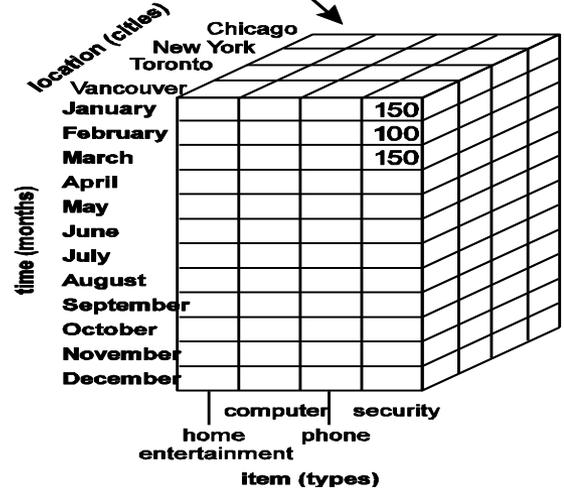
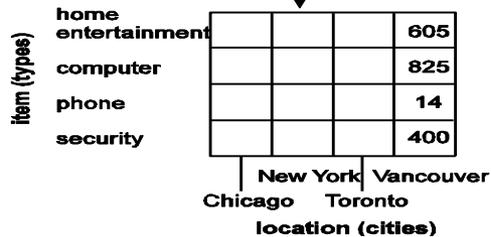
Drill down

drill-down on time (from quarters to months)



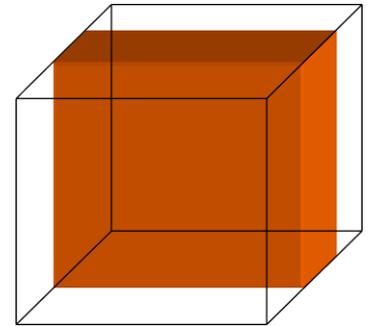
Pivote (rotar):

pivot

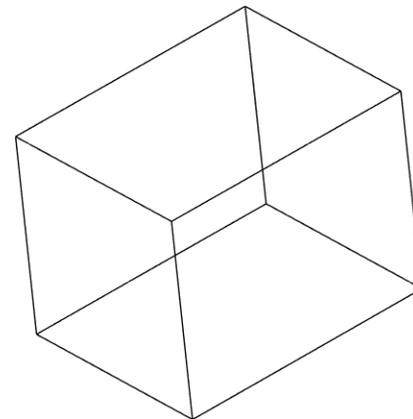


Resumen Operaciones clásicas OLAP

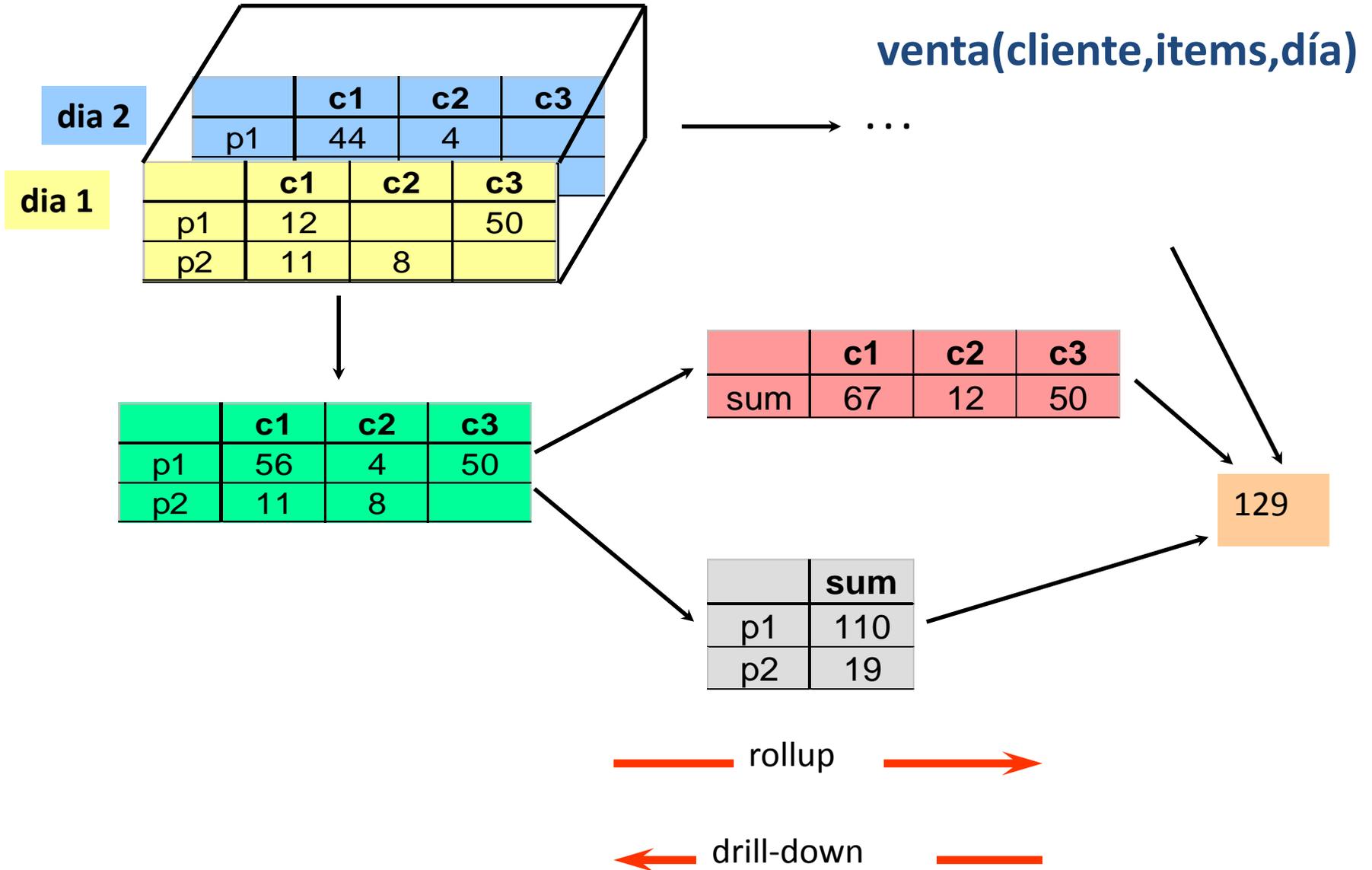
- *Rollup*: decrease nivel de detalle
- *Drill-down*: aumenta nivel de detalle
- *Slice-and-dice*: selección y proyección



- *Pivot*: re-orienta vista multidimensional

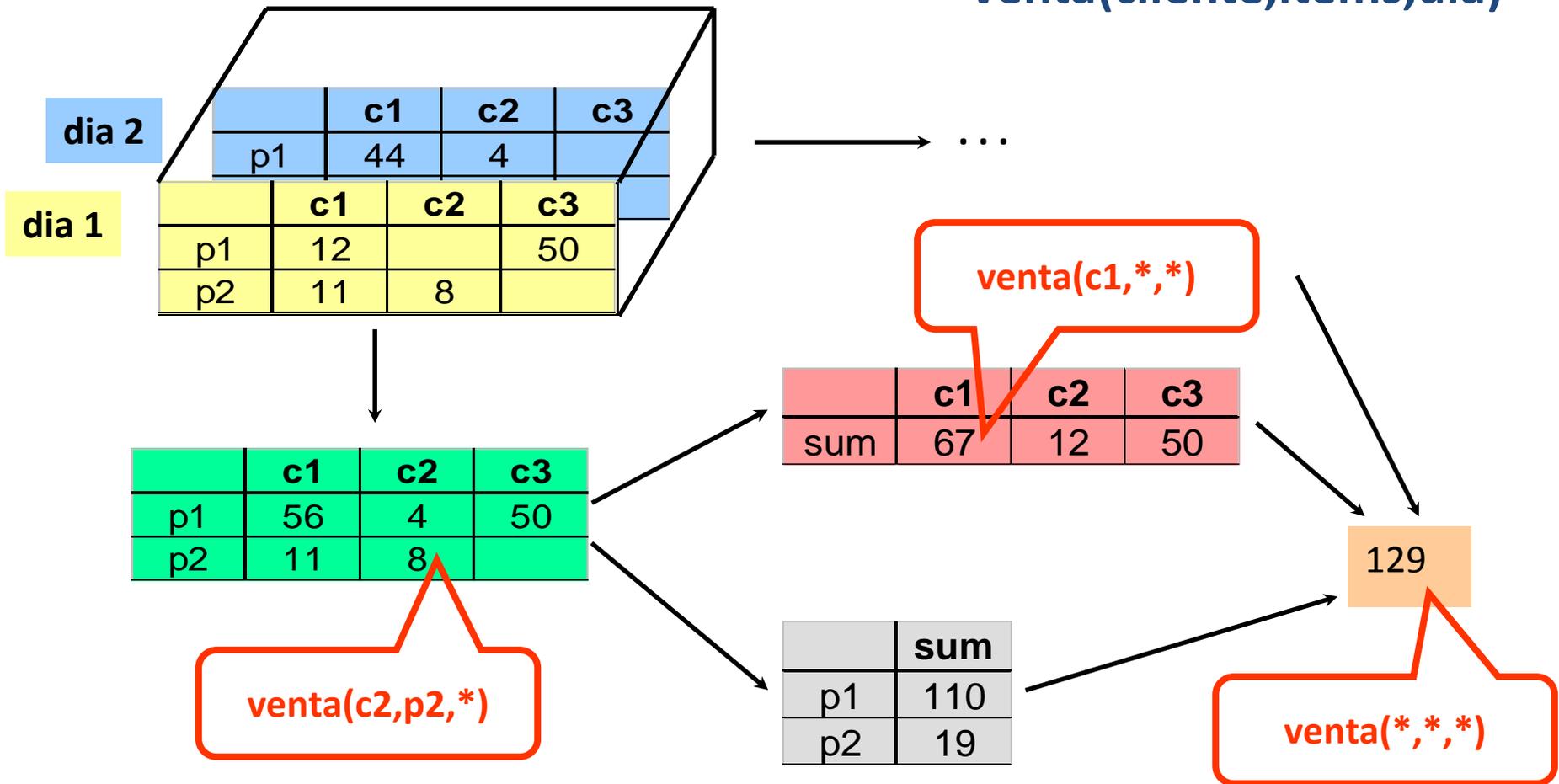


Agregación en Cubos



Agregación en Cubos

venta(cliente,item,día)



Cara de Cubos

venta(cliente,items,día)

		c1	c2	c3	*
dia 2	*				
	p1	56	4	50	110
	p2	11	8		19
	*				129
dia 1		c1	c2	c3	*
	p1	12		50	62
	p2	11	8		19
	*	23	8	50	81

venta(*,p2,*)

Agregación usando jerarquía

venta(cliente,items,día)

		c1	c2	c3
Dia 2	p1	44	4	
Dia 1	p1	12		50
	p2	11	8	

cliente
|
region
|
país

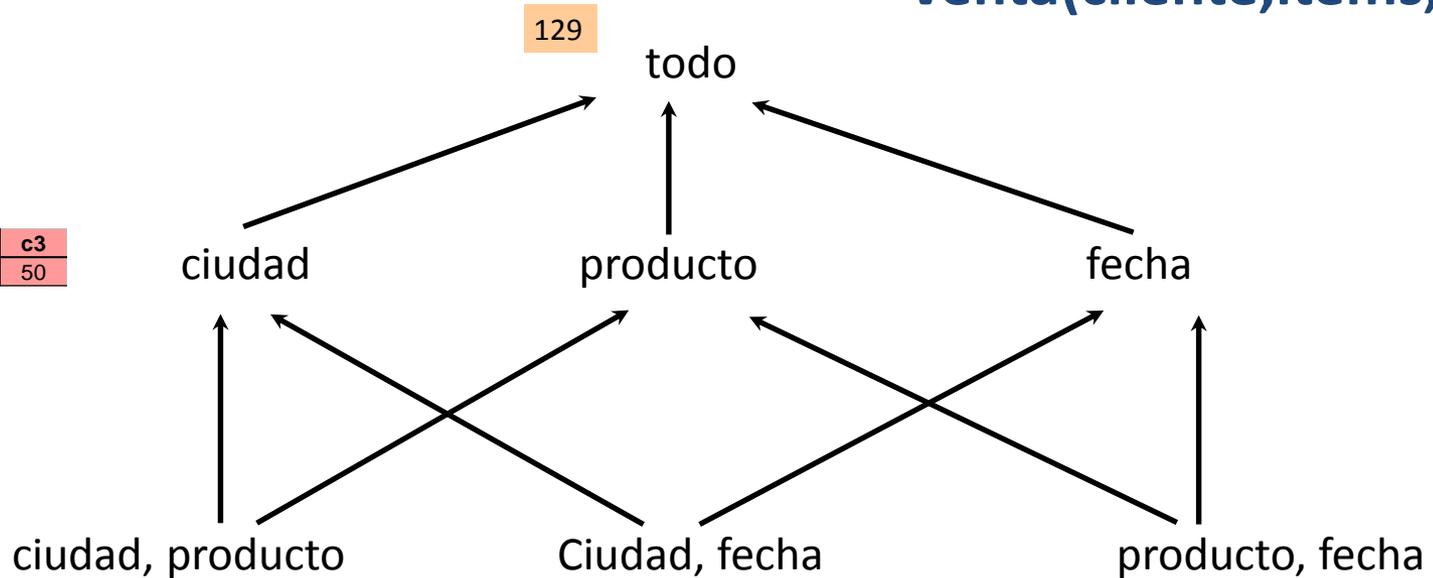
	region A	region B
p1	56	54
p2	11	8

(cliente c1 en Region A;
cliente c2, c3 en Region B)

Agregación en Cubos

venta(cliente,items,día)

	c1	c2	c3
p1	67	12	50



	c1	c2	c3
p1	56	4	50
p2	11	8	

	c1	c2	c3
day 2			
p1	44	4	
day 1			
p1	12		50
p2	11	8	

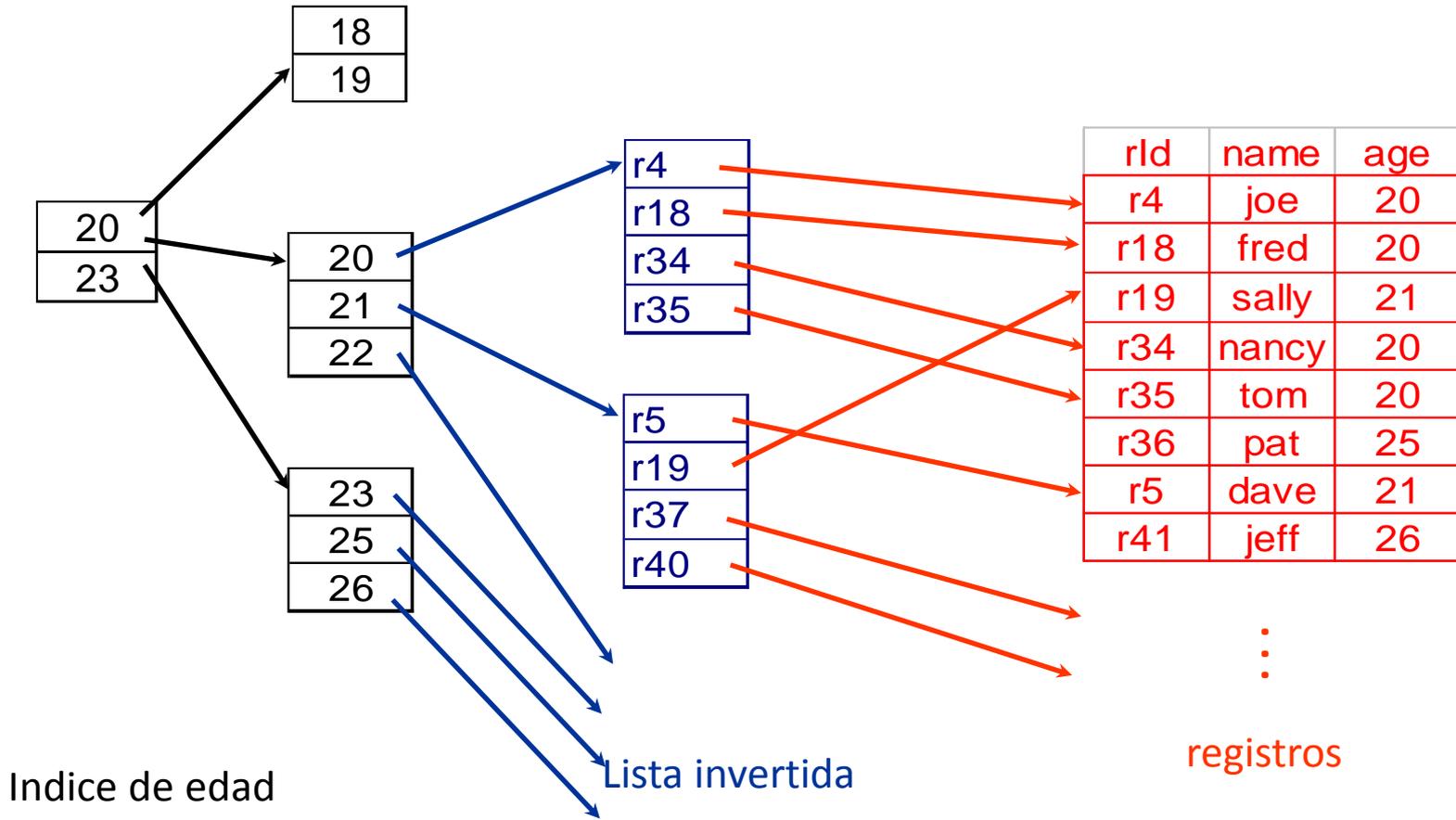
ciudad, producto, fecha

Algoritmos especiales para Calcular a diferentes niveles

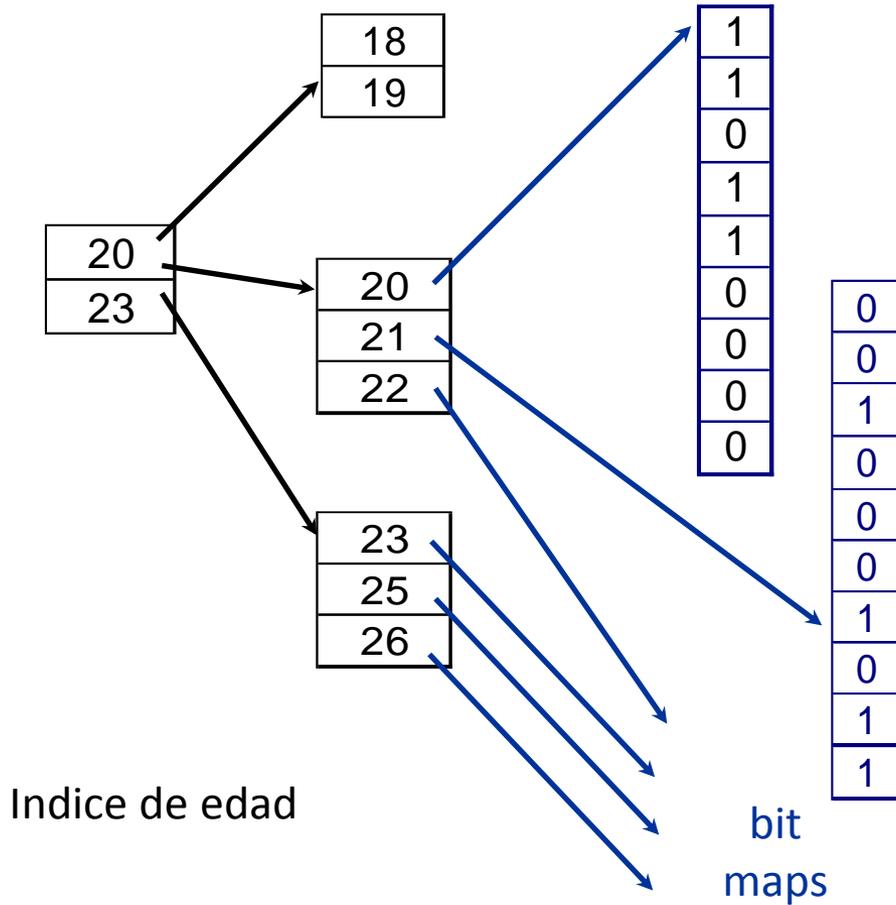
Estructuras Índices

- **Métodos de acceso tradicional**
 - Árboles B, tablas hash, ...
- **Popular en Warehouses**
 - Listas invertidas
 - Índices de mapa de bits (bit map)
 - Índices de unión (join indexes)
 - Índices de texto

Listas invertidas



Bit Maps

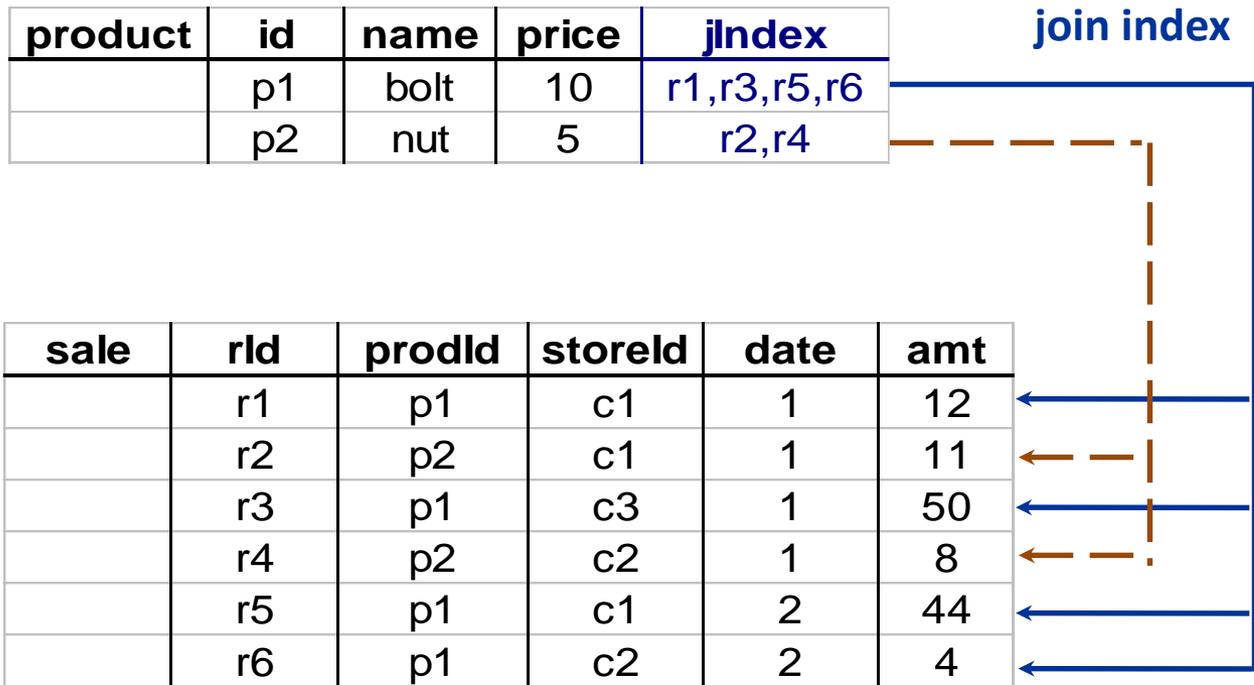


id	name	age
1	joe	20
2	fred	20
3	sally	21
4	nancy	20
5	tom	20
6	pat	25
7	dave	21
8	jeff	26

⋮

registros

Join Indexes



- Relaciona los valores de las dimensiones de un esquema en estrella a filas de la tabla de hechos.
- Por ejemplo, la tabla de hecho ventas y la de dimensión producto
- Un índice join en producto guarda para cada producto una lista de los IDs de las tuplas que registran las ventas de ese producto

Extension de SQL

- **ROLLUP:**

- SELECT <column list>
- FROM <table...>
- GROUP BY
ROLLUP(column_list);

Hace $n+1$ agregaciones en una columna,

- **CUBO:**

- SELECT <column list>
- FROM <table...>
- GROUP BY
CUBE(column_list);

Crea n combinaciones, n será el número de columnas del grupo

Ejemplo CUBO

Animaux

Animal	Lieu	Quantite
Chat	Paris	18
Chat	Naples	9
Chat	-	27
Chien	Paris	12
Chien	Naples	5
Chien	Rome	14
Chien	-	31
Tortue	Naples	1
Tortue	Rome	4
Tortue	-	5
-	-	63
-	Paris	30
-	Naples	15
-	Rome	18

Animal	Lieu	Quantite
Chien	Paris	12
Chat	Paris	18
Tortue	Rome	4
Chien	Rome	14
Chat	Naples	9
Chien	Naples	5
Tortue	Naples	1

**SELECT Animal, Lieu,
SUM(Quantite) as Quantite
FROM Animaux
GROUP BY ROLLUP Animal**

Ejemplo CUBO

SELECT item, city, year, SUM (amount)

FROM SALES

CUBE BY item, city, year

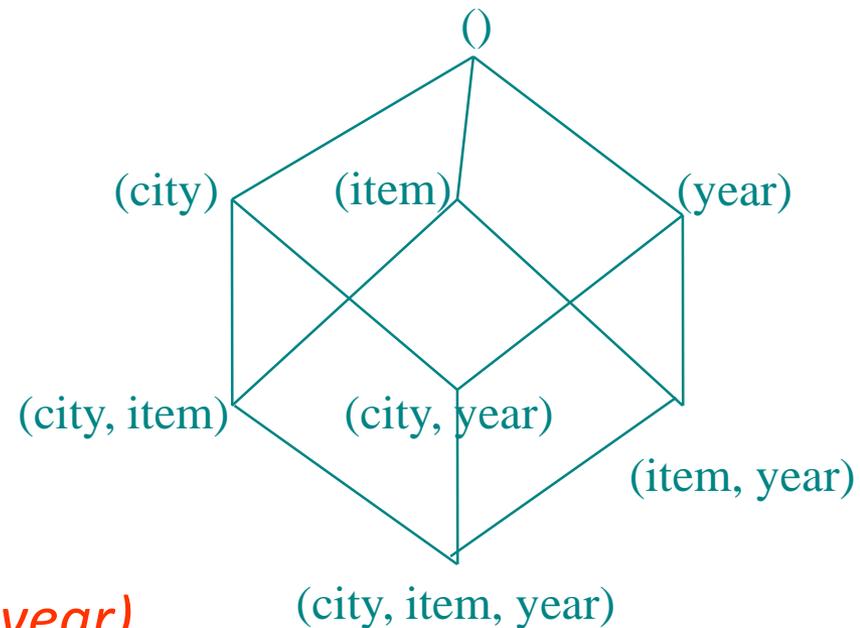
- Se debe calcular

(item, city, year),

(item, city), (item, year), (city, year),

(item), (city), (year)

()



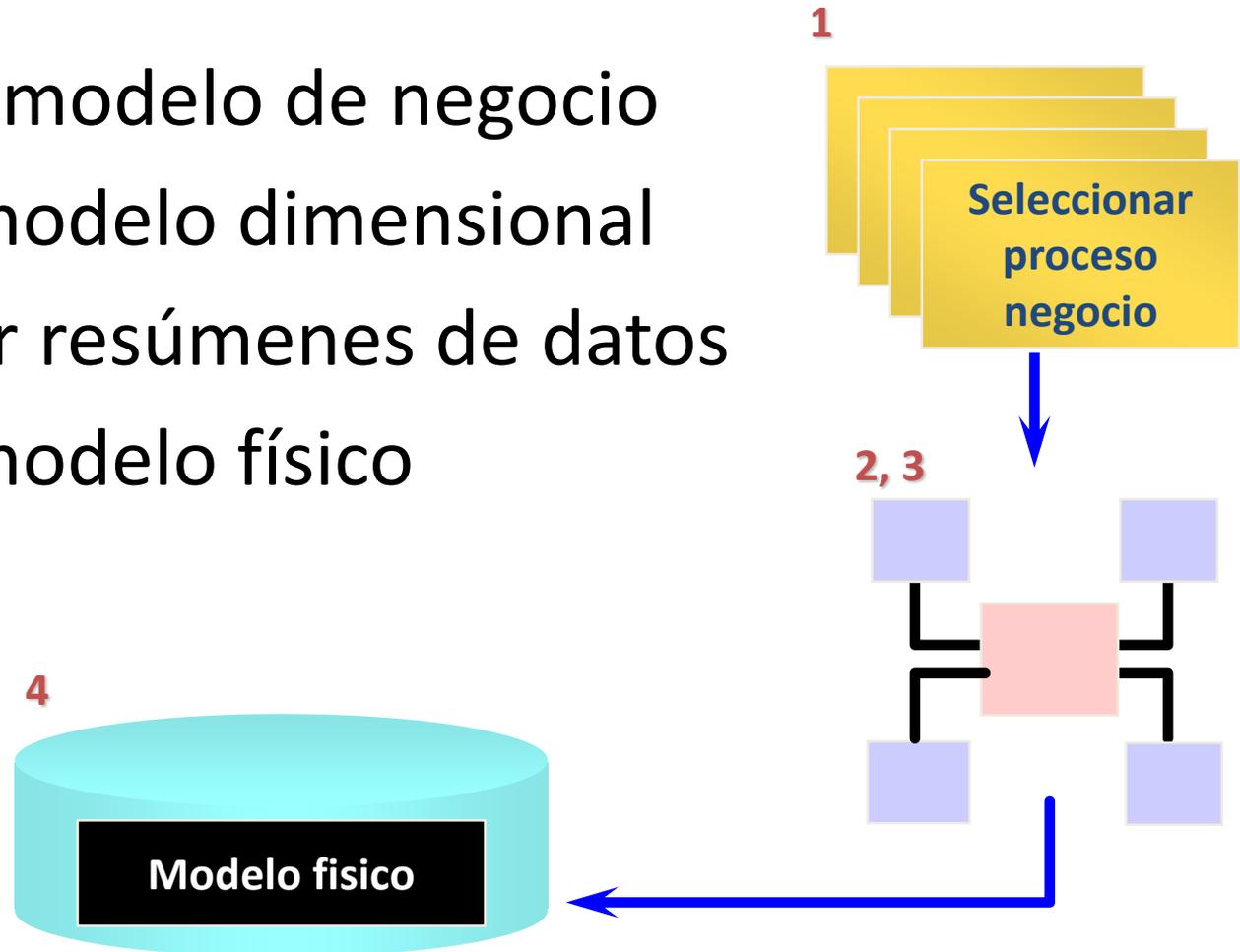
Extension de SQL

- Agrupar todos los subconjuntos de {item, región, mes}, encontrar el precio máximo en 1997 de cada grupo, y el total de ventas entre todas las tuplas de precios máximos

```
select item, región, mes, max(precio), sum(R.ventas)
from compras
where año = 1997
cube by item, region, mes: R
such that R.precio = max(precio)
```

Modelado en Data Warehouse

1. Definir el modelo de negocio
2. Crear el modelo dimensional
3. Identificar resúmenes de datos
4. Crear el modelo físico



Crear Modelo Dimensional

- Seleccionar una entidad para comenzar a armar tabla de hechos
- Determinar granularidad
- Identificar claves operacionales para tabla de hechos
- Buscar jerarquías
- Añadir dimensiones
- Caracterizar los atributos de las dimensiones

Granularidad (unidad de análisis)

Determina lo que representa cada registro de la tabla de hechos: el nivel de detalles.

- Ejemplos
 - Puntos en el tiempo
 - Lineas en un documento
- Depende del proyecto de IN

Crear Modelo Dimensional

- Identificar tablas de hechos
 - Traducir medidas pregunta madre en tablas de hechos
 - Analizar las fuentes de datos para las medidas
 - Identificar tablas de dimensiones
- Enlazar tabla de hechos con las tablas de dimensiones
- Crear vistas para los usuarios (operaciones OLAP)

Identificar resúmenes de datos

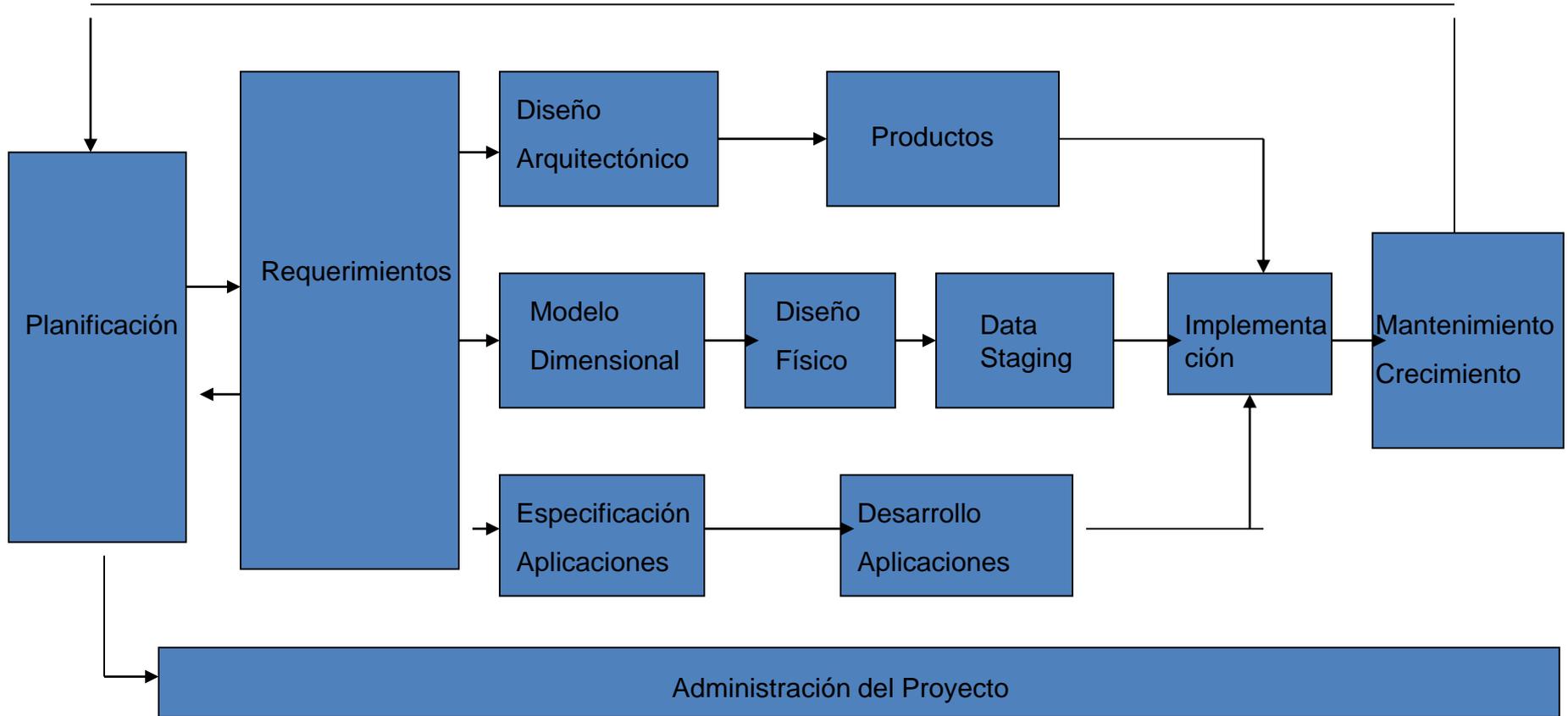
- Proporciona un acceso rápido a datos pre-calculados
- Reduce el uso de E/S, CPU y memoria
- Se calcula desde las fuentes de datos y otros resúmenes pre-calculados
- Por lo general, se guardan en las tablas de hechos

Identificar resúmenes de datos

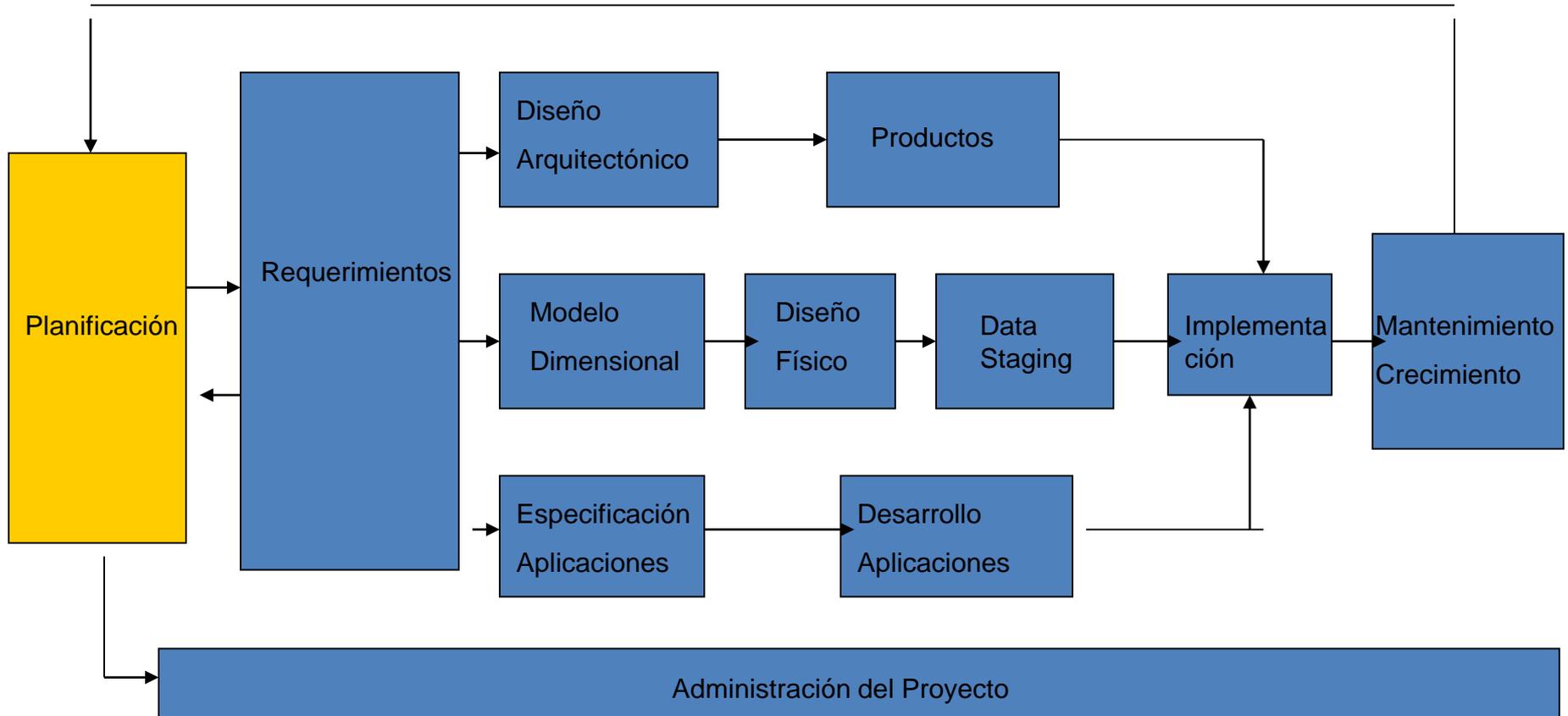
- Promedio
- Máximo
- Total
- Porcentaje

	Unidad	Venta	Tienda
Producto A			
Total			
Producto B			
Total			
Producto C			
Total			

Ciclo de vida



Planificación



Planificación

- Predisposición de la organización
- Alcance
- Justificación
- Aspectos humanos
- Plan del proyecto

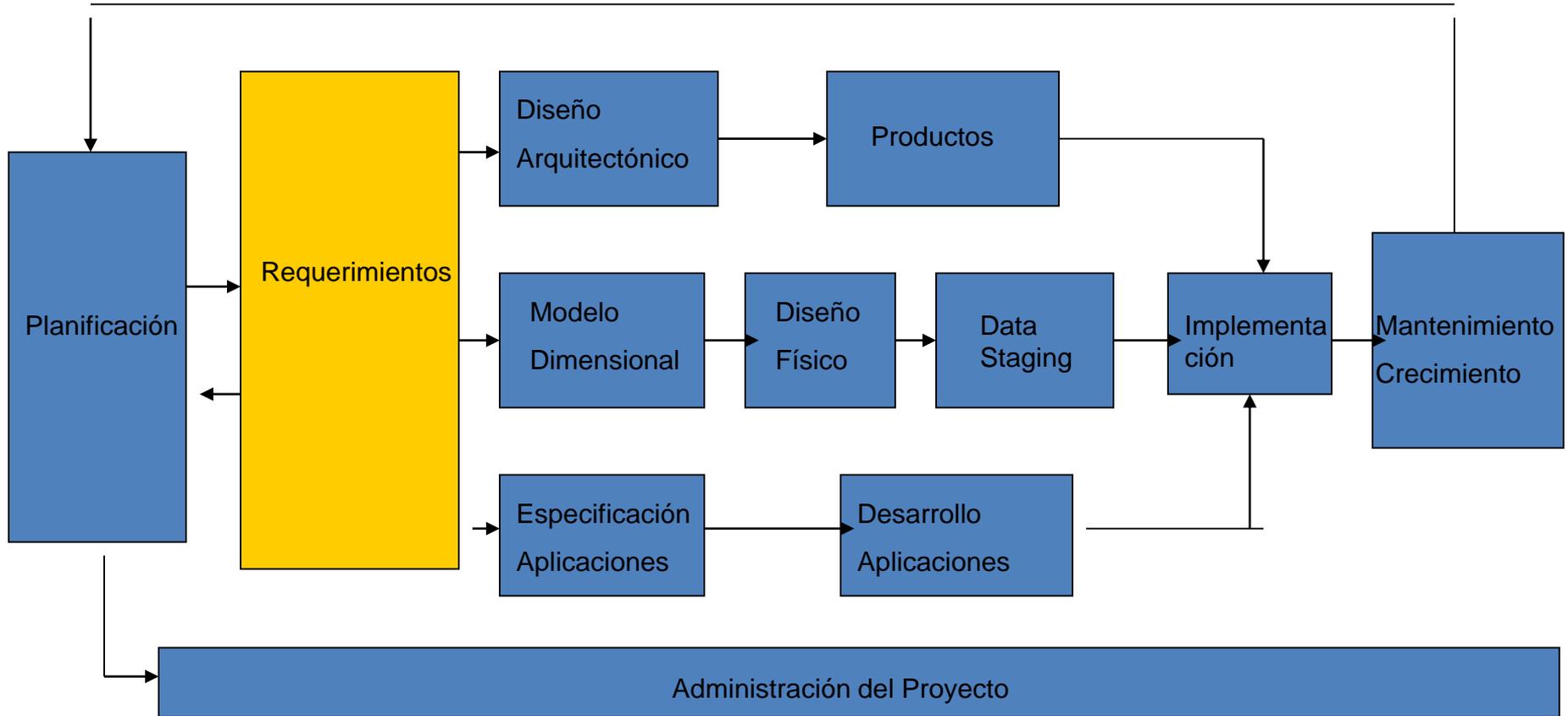
Predisposición de la organización

- Apoyo de la Gerencia (Sponsor)
- Motivación en la organización
- Participación de la gente vinculada al problema a analizar y de Sistemas
- Cultura de análisis de la información
- Factibilidad

Puntos clave

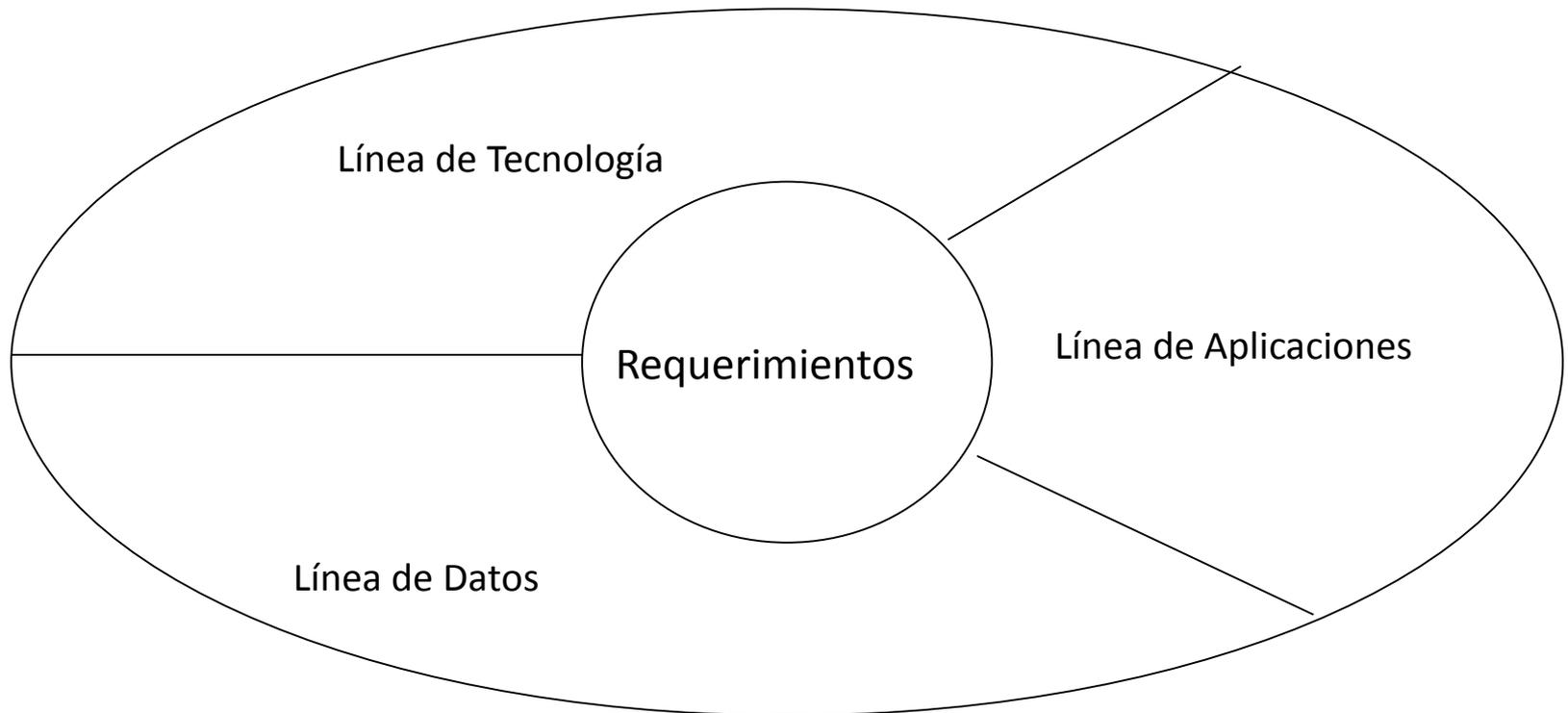
- Buscar un patrocinante bien ubicado
- Hacer un balance entre valor para la organización y manejabilidad
- Desarrollar cuidadosamente el plan del proyecto
- Ser un director de proyecto con capacidad de motivar, administrar y comunicar a todos los niveles

Requerimientos

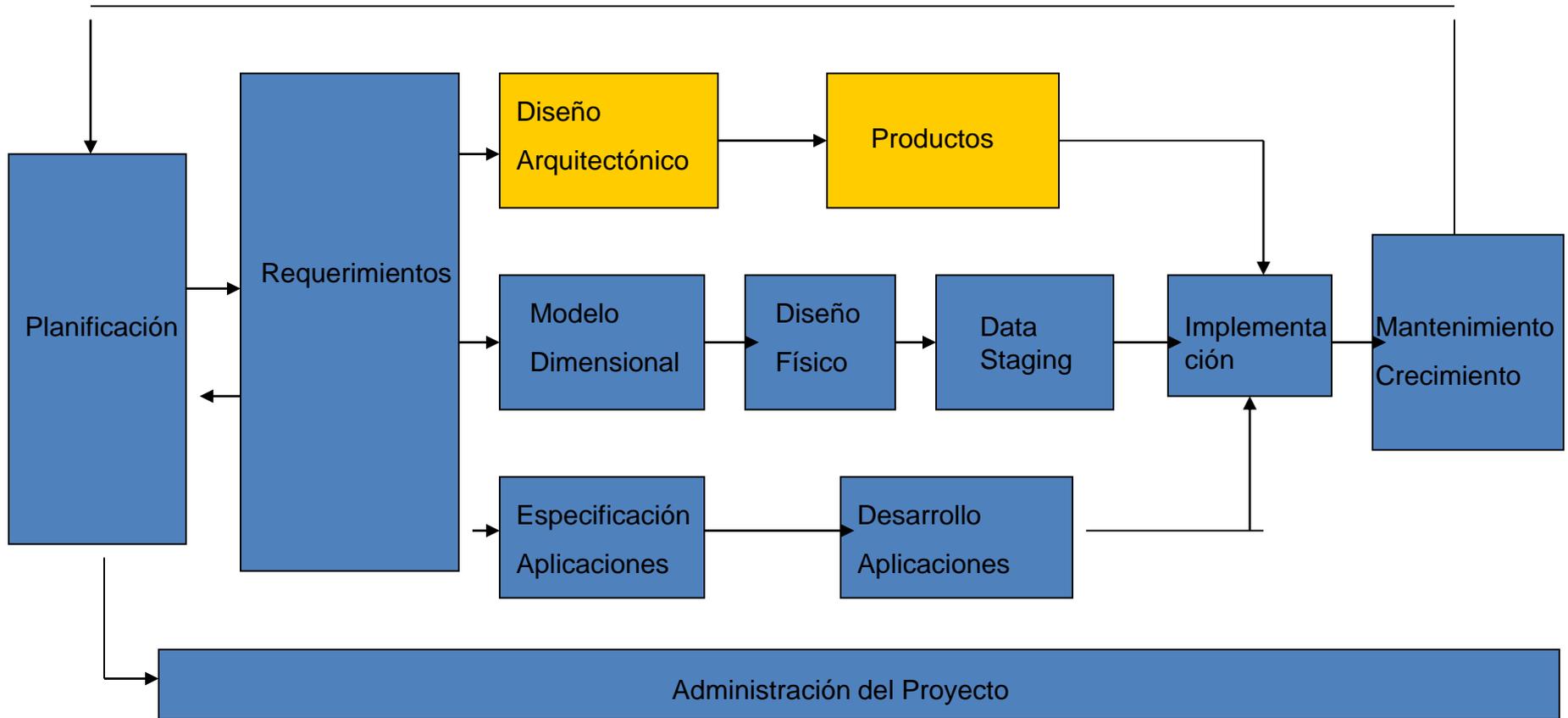


Desarrollo del proyecto

El desarrollo del proyecto se realiza en tres líneas



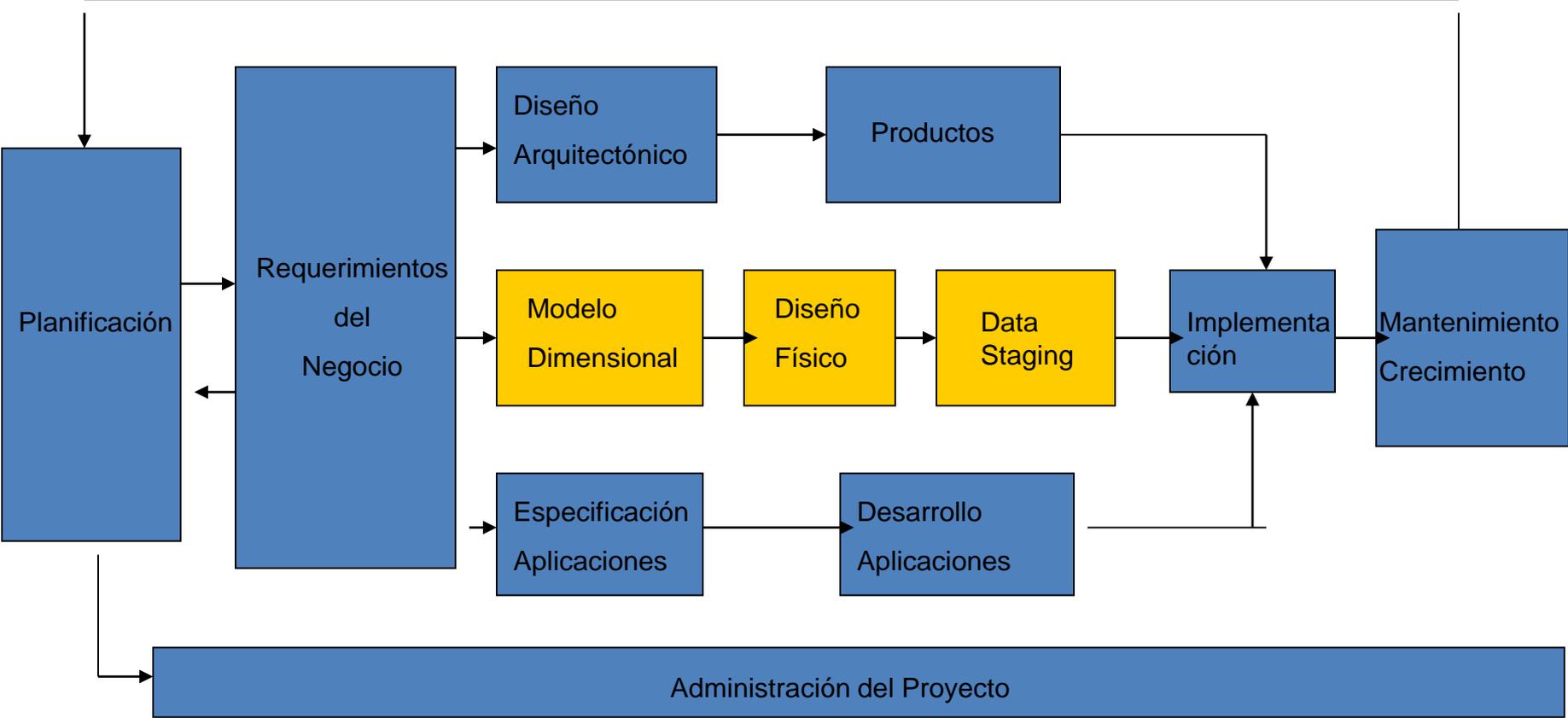
Línea de Tecnología



Selección de productos

- Interés Organizacional
- Requerimientos técnicos
- Criterios de selección
- Factores de ponderación
- Matriz de evaluación

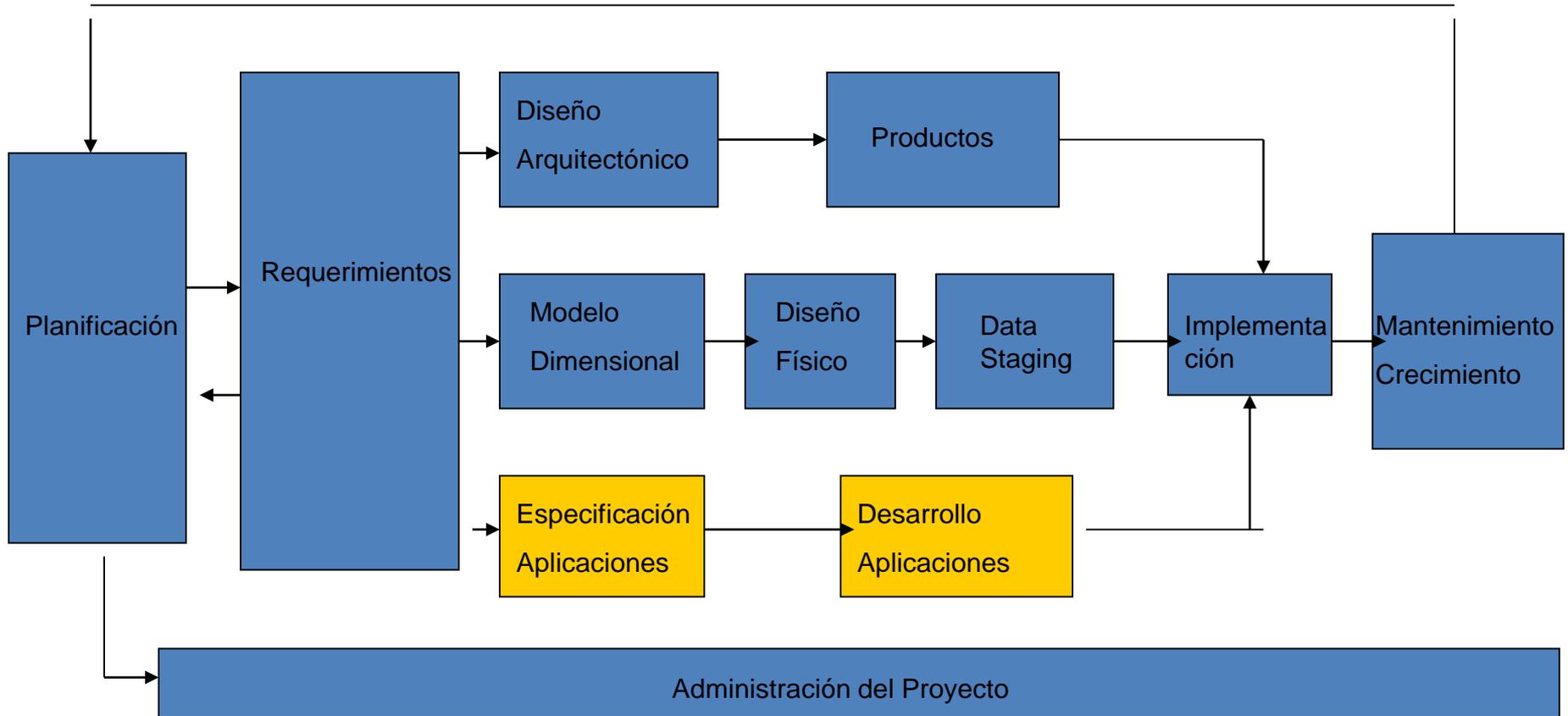
Línea de Datos



Línea de Datos

- Modelado dimensional
 - Tablas de hechos
 - Tablas de dimensión
 - Claves
- Diseño y desarrollo del ETL
 - Herramientas y técnicas
 - Organización de las tablas de dimensión
 - Organización de las tablas de hechos

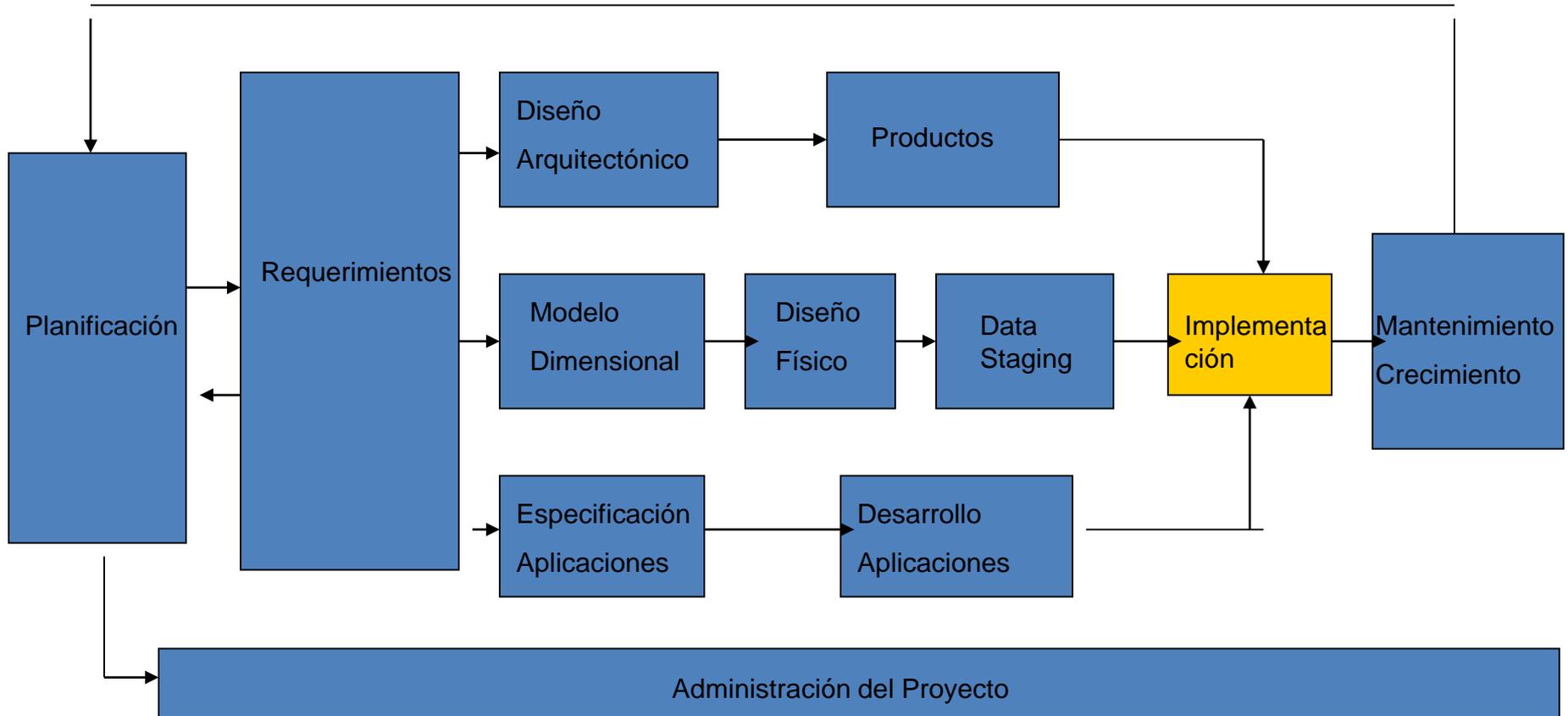
Línea de Aplicaciones



Línea de Aplicaciones

- Especificación y desarrollo de aplicaciones
 - Vías de acceso
 - Internet
 - Correo electrónico
 - Tableros de control
 - Personalización de herramientas

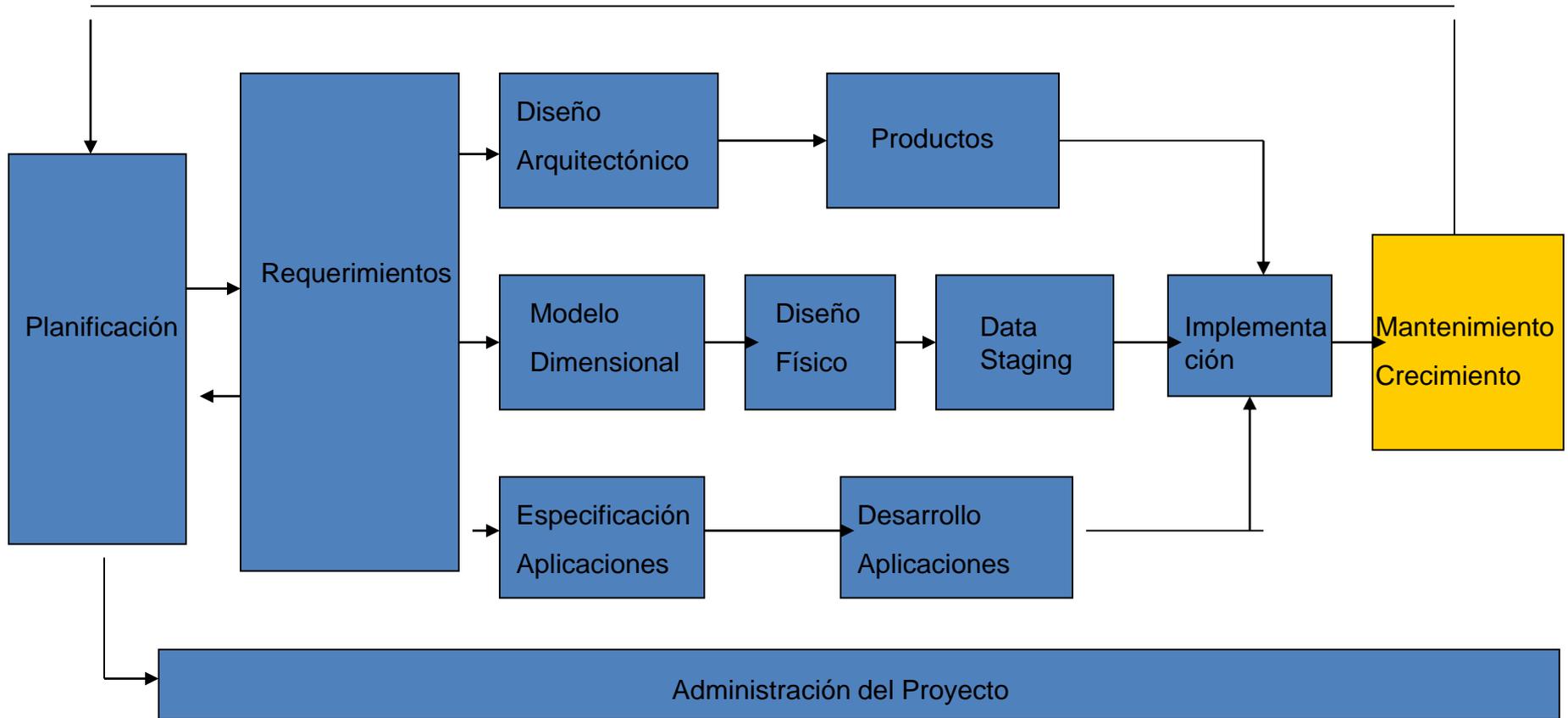
Implementación



Implementación

- Convergencia de las tres líneas
- La línea de datos es la que debería durar más tiempo
- Asegurarse de que el Data Warehouse esté en condiciones
- Educación: Gestión de conocimiento en la organización

Mantenimiento y Crecimiento



Mantenimiento y Crecimiento

- **Soporte a los usuarios**
 - Si no hay consultas, posiblemente no estén usando el Data Warehouse
 - Detectar áreas de datos o de aplicaciones no cubiertas
 - Calidad del Data Warehouse
- **Educación**
 - Cursos de gestión de conocimiento
 - Usuarios calificados

Mantenimiento y Crecimiento

- **Demandas de crecimiento**
 - Nuevos usuarios
 - Nuevos datos
 - Nuevas aplicaciones
 - Mejoras de las aplicaciones existentes
- **Revisión de las prioridades establecidas**
- **Identificar los productos objetos de gestión de conocimiento organizacional**



Modelado de Datos (algunos comentarios finales)

Jose Aguilar
CEMISID, Escuela de Sistemas
Facultad de Ingeniería
Universidad de Los Andes
Mérida, Venezuela

Expresiones multidimensionales (MDX)

- Es el acrónimo de MultiDimensional eXpressions
- Es un lenguaje de consulta para bases de datos multidimensionales sobre cubos OLAP
- Fue creado en 1997 por Microsoft. No es un lenguaje estándar, sin embargo diferentes fabricantes de herramientas OLAP lo han adoptado.
- Se utiliza para generar reportes para la toma de decisiones basados en datos históricos, usando la estructura o rotación del cubo

Expresiones multidimensionales (MDX)

- **Una consulta MDX es muy similar a una consulta SQL,**
 - devuelve un conjunto de celdas, que es resultado de tomar un subconjunto de las celdas del cubo original.
- **MDX utiliza en varias situaciones las jerarquías.**
 - Por ejemplo, si una dimensión se denomina región, esta puede contener países. Los países a su vez contienen provincias y las provincias ciudades.
 - Para manejar estos componentes MDX tiene funciones como Children (hijos), cousin (primos) y parents (padres).
- **Su cliente OLAP puede manipular el cubo de distintas formas:**
 - Rotarlo
 - Rebanarlo
 - Cortarlo

Expresiones multidimensionales (MDX)

Consulta MDX Básica:

- Sintaxis:

SELECT <especificación del eje y> on columns,

<especificación de eje x> on rows

FROM <especificación del cubo>

WHERE <especificación Slicer (rebanador)>

Expresiones multidimensionales (MDX)

```
CREATE CUBE Sales
( DIMENSION Time TYPE TIME,
    HIERARCHY [Fiscal],
        LEVEL [Fiscal Year] TYPE YEAR,
        LEVEL [Fiscal Qtr] TYPE QUARTER,
        LEVEL [Fiscal Month] TYPE MONTH OPTIONS (SORTBYKEY, UNIQUE_KEY),
    HIERARCHY [Calendar],
        LEVEL [Calendar Year] TYPE YEAR,
        LEVEL [Calendar Month] TYPE MONTH,
    DIMENSION Products,
        LEVEL [All Products] TYPE ALL,
        LEVEL Category,
        LEVEL [Sub Category],
        LEVEL [Product Name],
    DIMENSION Geography,
        LEVEL [Whole World] TYPE ALL,
        LEVEL Region,
        LEVEL Country,
        LEVEL City,
    MEASURE [Sales]
        FUNCTION SUM
        FORMAT 'Currency',
    MEASURE [Units Sold]
        FUNCTION SUM
        TYPE DBTYPE_UI4 )
```

Metadatos de un ejemplo

Nombre del cubo: CuboNW

Medidas: Total y Quantity.

Dimensiones:

- Products:

Jerarquias: Category name-Productname.

- Vw_ordenes2:

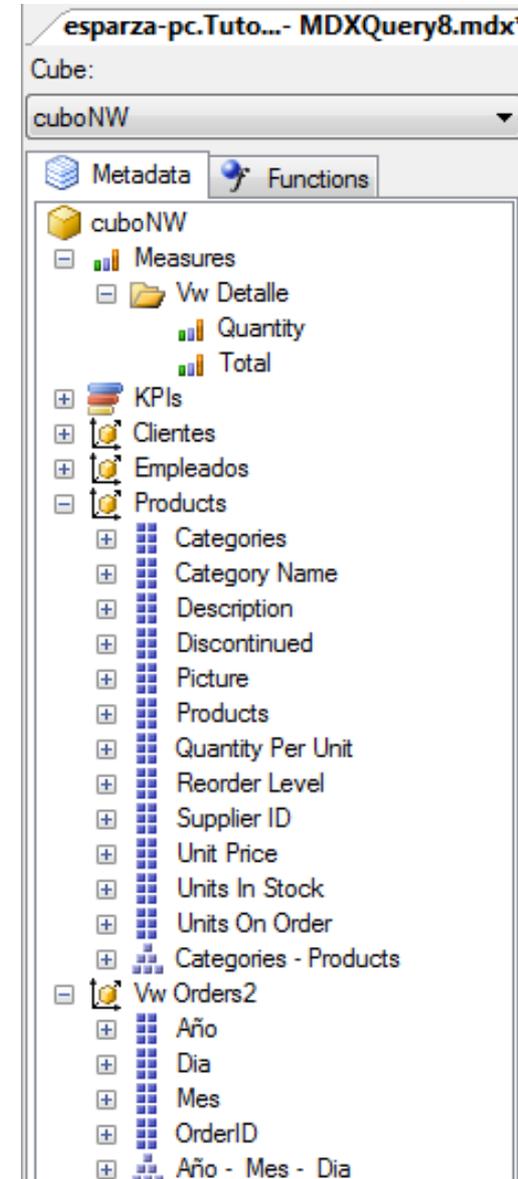
Jerarquias:Año-Mes-Dia

- Clientes.

Jerarquia: Country-Region-City-Company name

- Empleados.

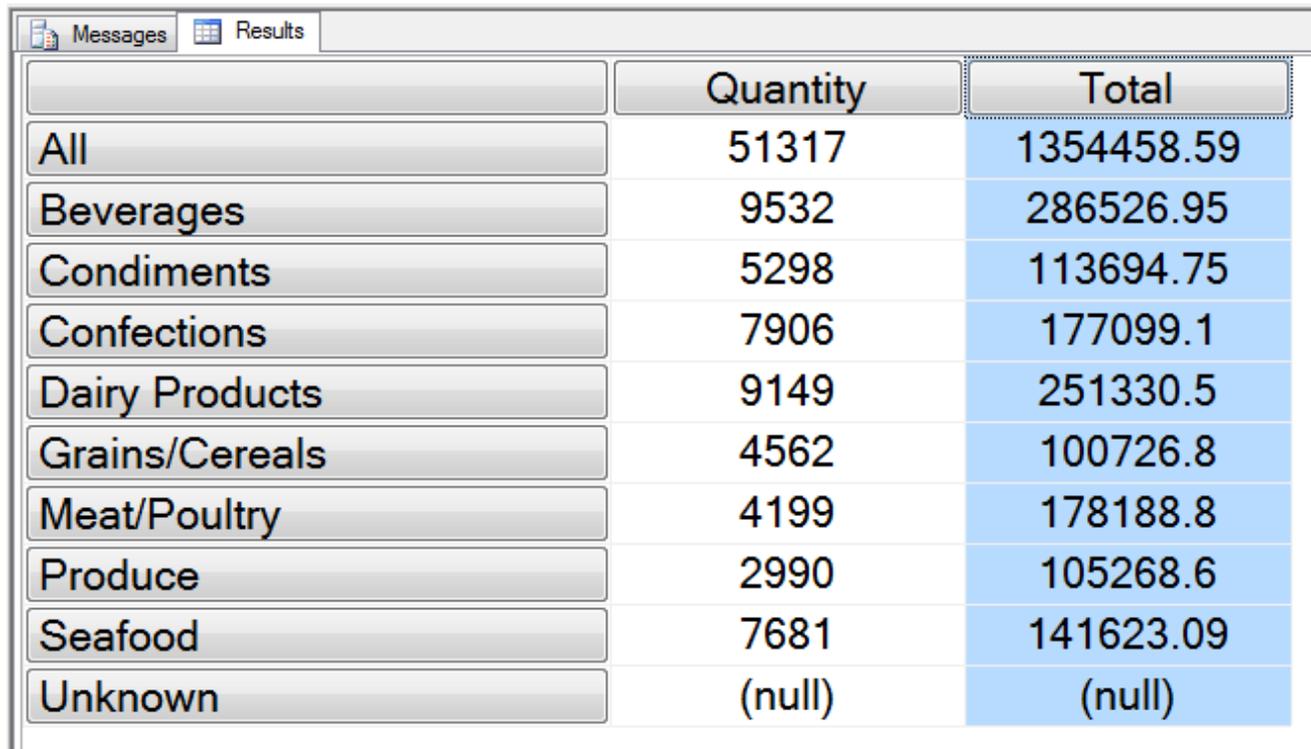
Jerarquia: Empcountry-Empregion-EmpCity-Lastname



Expresión con los nombres de las categorías y todas las medidas.

```
SELECT Measures.MEMBERS ON COLUMNS,  
[category name].MEMBERS ON ROWS  
FROM [cuboNW]
```

Totales y cantidades
para todos los
productos

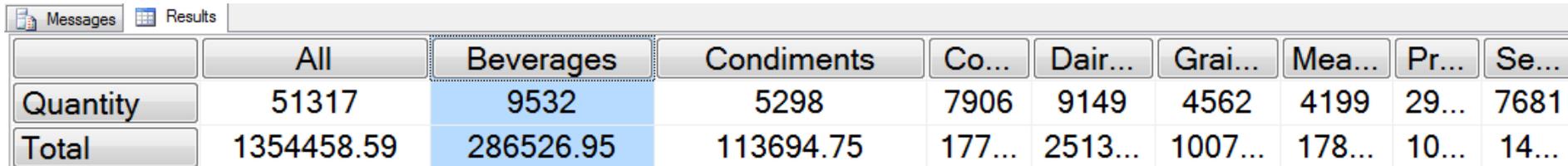


The screenshot shows a window titled 'Messages Results' containing a table with three columns: an empty header cell, 'Quantity', and 'Total'. The rows list various product categories with their corresponding quantities and total values.

	Quantity	Total
All	51317	1354458.59
Beverages	9532	286526.95
Condiments	5298	113694.75
Confections	7906	177099.1
Dairy Products	9149	251330.5
Grains/Cereals	4562	100726.8
Meat/Poultry	4199	178188.8
Produce	2990	105268.6
Seafood	7681	141623.09
Unknown	(null)	(null)

Cambio de ejes de los resultados

**SELECT [category name].MEMBERS ON COLUMNS,
Measures.MEMBERS ON ROWS
FROM [cuboNW]**



The screenshot shows a data grid with a 'Results' tab. The grid has a header row with categories: All, Beverages, Condiments, Co..., Dair..., Grai..., Mea..., Pr..., and Se... The first column contains 'Quantity' and 'Total'. The 'Beverages' column is highlighted in blue. The values for 'Quantity' are 51317, 9532, 5298, 7906, 9149, 4562, 4199, 29..., and 7681. The values for 'Total' are 1354458.59, 286526.95, 113694.75, 177..., 2513..., 1007..., 178..., 10..., and 14...

	All	Beverages	Condiments	Co...	Dair...	Grai...	Mea...	Pr...	Se...
Quantity	51317	9532	5298	7906	9149	4562	4199	29...	7681
Total	1354458.59	286526.95	113694.75	177...	2513...	1007...	178...	10...	14...

Ejemplos donde se tiene en uno de los renglones una dimensión y en las columnas medidas

Lista de elementos en ejes

Ahora vamos a combinar en ambos ejes dos dimensiones: año y las categorías

```
SELECT
{ [AÑO].members } ON COLUMNS,
[products].[category name].MEMBERS ON ROWS
FROM [cuboNW]
WHERE MEASURES.TOTAL
```

Es necesario especificarle en WHERE la medida que se desea ver (p. e TOTAL)

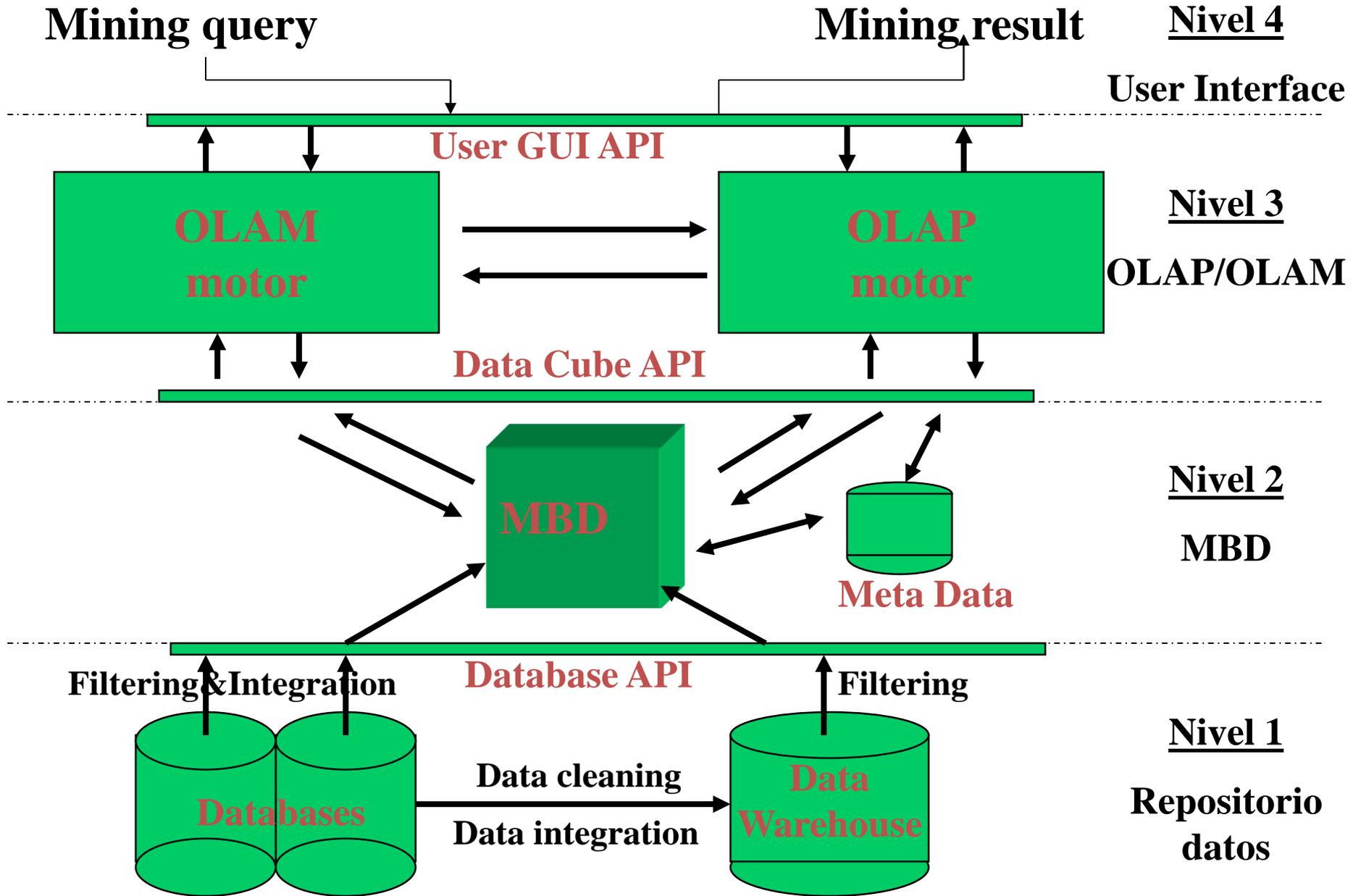
	All	1996	1997	1998
All	1354458.59	226298.5	658388.75	469771....
Beverages	286526.95	53879.2	110424	122223....
Condiments	113694.75	19458.3	59679	34557.45
Confections	177099.1	31511.6	87227.77	58359.73
Dairy Products	251330.5	44615.8	123910.8	82803.9
Grains/Cereals	100726.8	9817.6	60486.95	30422.25
Meat/Poultry	178188.8	30292.2	87621.03	60275.57
Produce	105268.6	15134.2	57718.55	32415.85

De procesamiento analítico en línea a Minería analítica en línea (OLAM)

¿Por qué la minería analítico en línea?

- Alta calidad de los datos en data warehouses
 - DW contiene, datos integrados, coherentes, limpios
- Estructura de procesamiento de la información disponible
 - Herramientas presentación de informes y OLAP , etc.
- Análisis exploratorio de datos basados en OLAP
 - Minería con drilling, dicing, pivoting, etc.
- Selección on-line de las funciones de minería de datos
 - Integración e intercambio de múltiples funciones, algoritmos y tareas.

Arquitectura OLAM



Modelos Perfilados de Visualización

Analistas

Usuarios Casuales

Análisis Estratégico

Consultas Analíticas

Reportes Interactivos

Tablero de Control

Herramienta de Consulta

Reportes Dinámicos

