



Conceptos de Almacenes de Datos

Tema 2



Indice

- Arquitectura de un Data Warehouse
- Herramientas de Microsoft SQL Server
- El Modelo Multidimensional
- Operaciones OLAP



OLAP vs OLTP

- Sistemas de bases de datos tradicionales: diseñados y afinados para apoyar la operación del día a día:
 - garantizar, acceso simultáneo rápido a los datos
 - procesamiento de transacciones y control de concurrencia
 - Centrados en la consistencia de los datos de actualización en línea
 - Conocido como bases de datos operacionales o de procesamiento de transacciones en línea (OLTP)
- Características de los datos de las bases de datos OLTP:
 - Datos detallados
 - No incluya datos históricos
 - Altamente normalizada
 - Bajo rendimiento en consultas complejas
- El análisis de datos requiere un nuevo paradigma: el procesamiento analítico en línea (OLAP)
 - consulta típica OLTP: los pedidos pendientes para el cliente c1
 - consulta típica OLAP: cantidad total de ventas por producto y por cliente



OLAP vs OLTP

■ Características OLAP

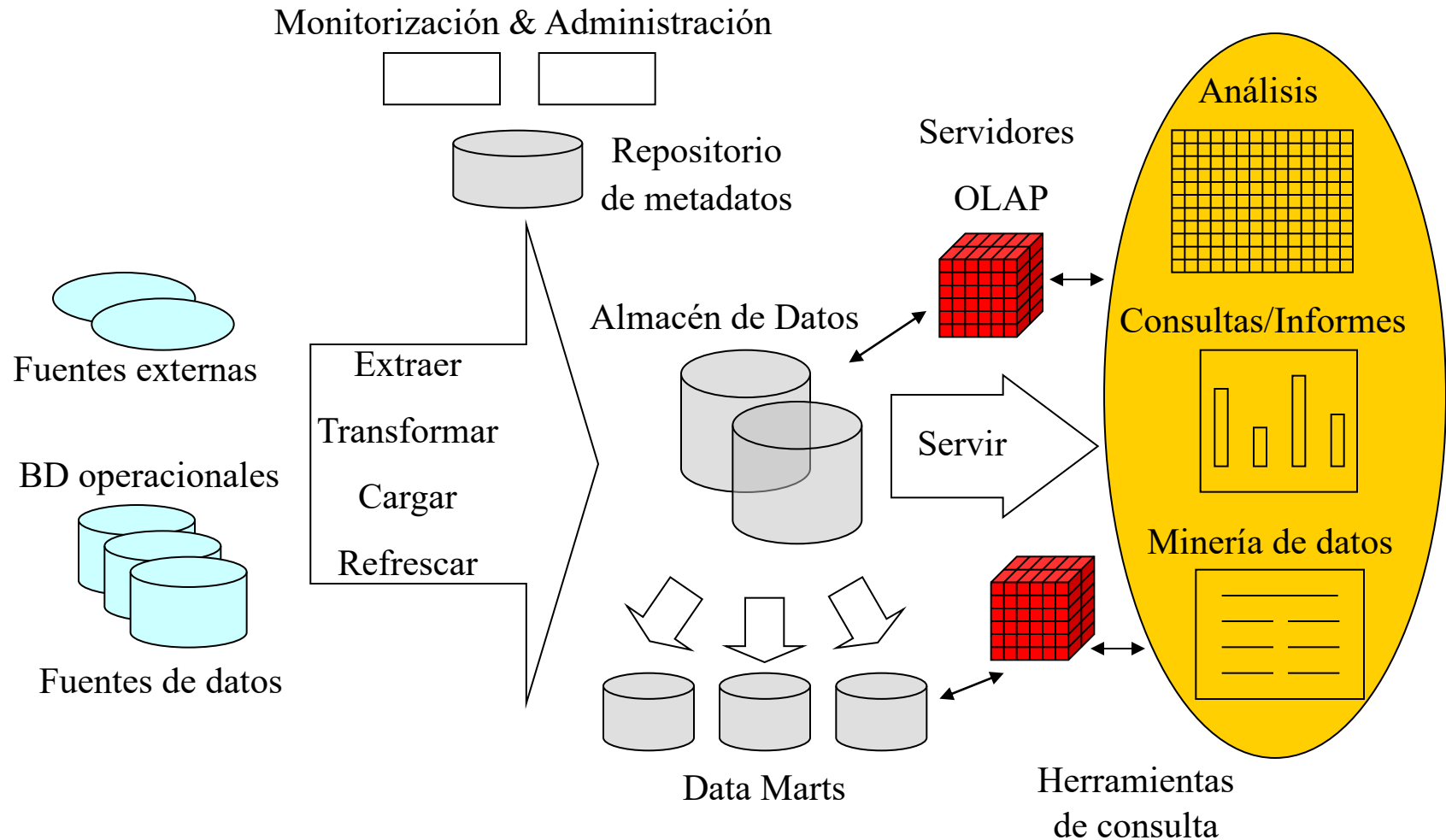
- Paradigma OLTP centrado en las transacciones, OLAP se centrado en consultas analíticas
- La normalización no es buena para consultas analíticas, la reconstrucción de datos requiere un alto número de joins
- Las bases de datos OLAP soportan una carga pesada de consultas
- Las técnicas de indexación OLTP no son eficientes en OLAP: orientada acceder a pocos registros
 - Las consultas OLAP normalmente incluyen la agregación
- La necesidad de un modelo de base de datos diferente para apoyar OLAP era clara: dio lugar a los almacenes de datos
- Almacén de datos: (Por lo general) grandes repositorios que consolidan datos de diversas fuentes (internas y externas a la organización), se actualizan offline, sigue el modelo de datos multidimensional, diseñado y optimizado para apoyar las consultas OLAP de manera eficiente.



Indice

- **Arquitectura de un Data Warehouse**
 - Esquema de una arquitectura de DW
 - Fuentes de datos: procesos y herramientas
 - El almacén de datos (DW)
 - Metadatos
 - Servidor del almacén de datos
 - Herramientas de consultas
 - Aplicaciones de DW en el mercado
- Herramientas de Microsoft SQL Server
- El Modelo Multidimensional
- Operaciones OLAP

Esquema de una arquitectura de DW





Esquema de una arquitectura de DW

- Arquitectura de tres capas
 - Servidor del repositorio o base de datos del almacén de datos
 - Casi siempre un SGBD Relacional
 - Servidores OLAP
 - Relational OLAP (ROLAP)
 - Extiende SGBD relacionales para permitir operaciones MD
 - Multidimensional OLAP (MOLAP)
 - Directamente implementa el modelo MD en vectores



Esquema de una arquitectura de DW

- Arquitectura de tres capas (II)
 - Clientes → Herramientas
 - Informes y consultas
 - OLAP (On-Line Analytical Processing)
 - Data Mining

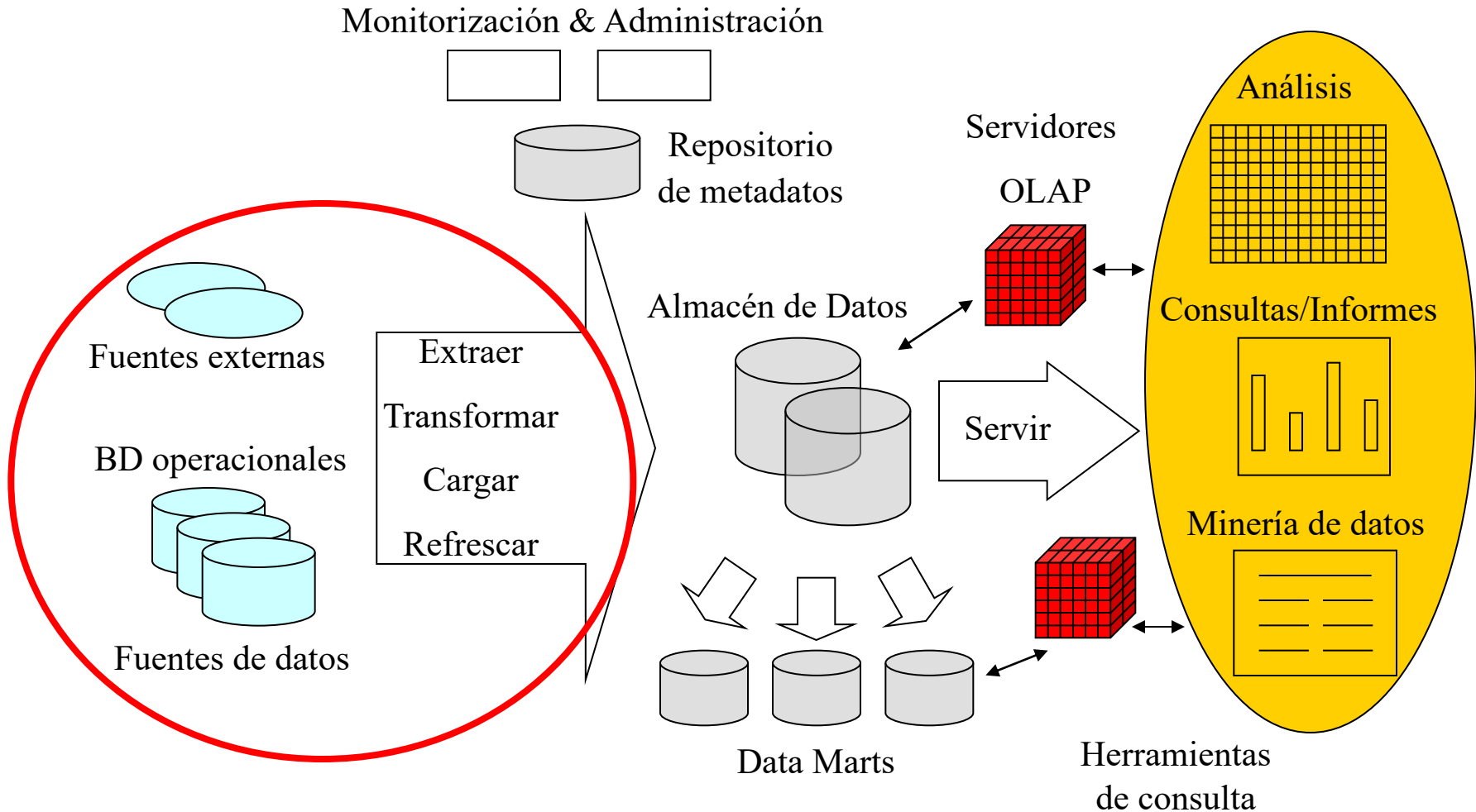


Esquema de una arquitectura de DW

- ¿ Por qué Data Warehouse separado ?
 - Rendimiento (Performance)
 - BD operacionales enfocadas y optimizadas a transacciones
 - Consultas OLAP complejas → Ralentización del servidor
 - Métodos de implementación, accesos, etc. distintos
 - Funcionalidad
 - Datos no existentes → históricos
 - Datos consolidados (agregados, sumados, resumidos, etc.)
 - De distintas fuentes
 - Calidad de datos
 - Diferentes fuentes → representaciones distintas, etc...

Esquema de una arquitectura de DW

Fuentes de datos: procesos y herramientas





Esquema de una arquitectura de DW

Fuentes de datos: procesos y herramientas

■ Fuentes de datos

- Fuentes de datos operacionales de la empresa
- Bases de datos externas (públicas o privadas)
- Ficheros planos
- Datos en formato tradicional:
 - documentos, facturas, albaranes, etc.
- Internet → Cantidad ingente de datos



Esquema de una arquitectura de DW

Fuentes de datos: procesos y herramientas

- Procesos para alimentar de datos el almacén (ETL)
 - Extracción (Extraction)
 - Limpieza (Cleaning) y Transformación (Transformation)
 - Carga (Loading)
 - Refresco



Esquema de una arquitectura de DW

Fuentes de datos: procesos y herramientas

- Procesos para alimentar de datos el almacén (ETL)
 - Extracción (Extraction)
 - Limpieza (Cleaning) y Transformación (Transformation)
 - Carga (Loading)
 - Refresco

- Extracción
 - Procesos que recogen los datos necesarios del almacén



Esquema de una arquitectura de DW

Fuentes de datos: procesos y herramientas

- Limpieza (Cleaning)
 - Fundamental que los datos del almacén sean correctos
 - Decisiones estratégicas
 - Muchas fuentes de datos → alta probabilidad de error y anomalías
 - Longitud de campos inconsistentes
 - Descripción inconsistente (¿¿¿¿¿ Qué es dirección ????)



Esquema de una arquitectura de DW

Fuentes de datos: procesos y herramientas

- Valores incoherentes
 - Universidad de Alicante
 - Univ. Alicante
 - U. de Alicante
 - Valores nulos para algunos campos
 - ¿¿ Violación de reglas de integridad ?? (ETL)
-
- Detectar herramientas que trabajen bien acopladas



Esquema de una arquitectura de DW

Fuentes de datos: procesos y herramientas

- Las herramientas de limpieza se pueden clasificar en 3 tipos:
 - Data migration (Migración de datos)
 - Proceso de mover datos de un sistema a otro
 - Se necesitan funciones de preparación de datos para poder ser cargados en el destino
 - Permiten reglas de transformaciones simples
 - Ej. Reemplazar "género" por "sexo"



Esquema de una arquitectura de DW

Fuentes de datos: procesos y herramientas

- Data Scrubbing (Limpieza de datos)
 - Proceso de corregir (o eliminar) registros inexactos en las fuentes de origen
 - Detectar “dirty data” (incorrectos, irrelevantes o incompletos)
 - Utilizan conocimiento específico del dominio
 - Ej. Direcciones postales
 - Asegurar precisión y consistencia con otros conjuntos de datos en el sistema
- Data Auditing tools (Auditoría de datos)
 - Proceso para evaluar la calidad y utilidad de los datos de entrada para el dominio objetivo
 - Escanean datos para descubrir reglas y relaciones y lanzar señales de violaciones si es necesario



Esquema de una arquitectura de DW

Fuentes de datos: procesos y herramientas

■ Carga (Loading)

- Una vez que los datos se extraen, limpian y transforman → CARGAR
- Se puede necesitar más pre-proceso antes de carga
 - Comprobar reglas de integridad de nuevo
 - Ordenar datos
 - Calcular datos agregados
 - Construir Tablas derivadas y virtuales e intermedias
 - Construir índices
 - Cantidad de datos a cargar
 - Calcular tiempo → Muchas veces por la noche
 - ¿¿ Y si multinacional ??



Esquema de una arquitectura de DW

Fuentes de datos: procesos y herramientas

- Técnicas de carga

- Cargas secuenciales

- Mucho tiempo (~ 100 días/ TB)
 - Reemplazar antigua tabla con la nueva después de transacción
 - Utiliza comprobaciones periódicas
 - Normalmente, comenzar después de fallo

- Procesos por lotes (batch)

- El administrador monitoriza el proceso
 - Procesos cortos con uso secuencial de I/O
 - Índices y derivados



Esquema de una arquitectura de DW

Fuentes de datos: procesos y herramientas

- Procesamiento paralelo y técnicas incrementales...
 - IMPORTANTE !!!!!
 - Sólo carga las actualizaciones → no toda la tabla
 - Cuando Committ !!!
 - Reemplazar el antiguo estado con los nuevos datos
 - Mientras carga, el DW puede ser consultado
 - Utilizar Comprobaciones periodicas de inconsistencia de estados
 - Auditorias



Esquema de una arquitectura de DW

Fuentes de datos: procesos y herramientas

■ Refresco

- Propagar actualizaciones sobre las fuentes de datos hacia los datos base y agregados del almacén
- ¿ Cuándo ?
 - Sobre cada actualización → muy caro
 - Sólo si datos muy actualizados. Ej. bolsa
- ¿ Cómo ? → Política
- Normalmente se hace periódicamente
- Se define una política en función de cada caso
 - No olvidar que SGBD ofrecen servidores para replicar datos



Esquema de una arquitectura de DW

Fuentes de datos: procesos y herramientas

■ Refresco...

■ Fundamentalmente 2 técnicas:

■ Extracción entera de las bases de datos

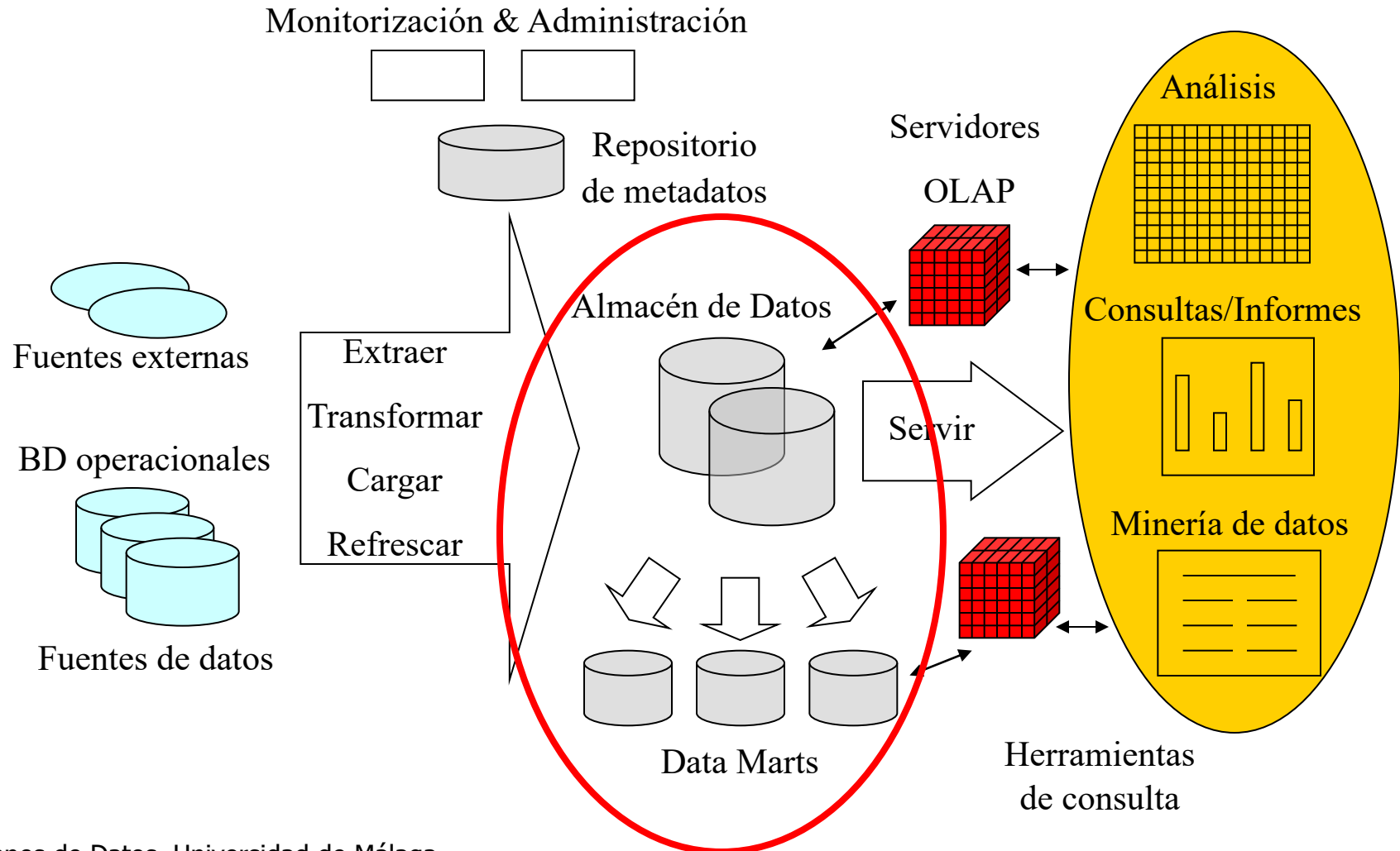
- Leen tablas o bases de datos enteras → caras
- A veces, única elección para ficheros o sistemas heredados

■ Técnicas incrementales

- Detectan y propagan cambios → servidores para replicación
 - Imágenes (snapshots) y triggers
Ej. Oracle
 - Transporte de transacciones (transaction shipping)
Ej. Sybase
 - Otras: IBM data replicator

Esquema de una arquitectura de DW

El almacén de datos (DW)





El almacén de datos: primera aproximación

- Definición según W. Inmon
 - “Una colección de datos **orientados por tema**, **integrados**, **variables en el tiempo** y **no volátiles** que se emplea como apoyo a la toma de decisiones estratégicas”



Esquema de una arquitectura de DW

El almacén de datos (DW)

- Datos orientados por tema
 - Los sistemas OLTP están optimizados para las transacciones
 - NORMALIZAR
 - Muchas transacciones con pocos datos
 - Ej. datos de clientes en varias tablas
 - ¿ Almacén de datos normalizado ? : PROBLEMAS
 - ¿ Directivo es capaz de tener visión de todas las tablas y relaciones ?
 - Pocas transacciones que incluyen muchos datos
 - Operación MAS cara en BD: unión de tablas



Esquema de una arquitectura de DW

El almacén de datos (DW)

- Los datos están orientados por tema
 - En un solo lugar (digamos tabla) datos referentes a un concepto que es el objeto de estudio
 - Ej. Tabla para clientes
 - Ventas
 - Compras
 - Vehículos, etc.



Esquema de una arquitectura de DW

El almacén de datos (DW)

- Integrados

- Están coherentemente agrupados a partir de datos de las fuentes de datos
- También hay datos derivados
- Para ello: procesos de limpieza y transformación
- Hay errores difíciles de detectar: iiii Cuidado !!!!!
 - Ej. Código producto válido → Exhaustivo análisis de datos



Esquema de una arquitectura de DW

El almacén de datos (DW)

- Integrados,.....

- Problemas de incoherencia: resumir en 4 tipos

- Descripción

- J. A. Rodríguez
 - Jose A. Rodriguez

- Codificación

- Varón "V", Hembra "H"; en otra BD Varón "H", Hembra "M"

- Unidades

- Estatura: 1,70 mts; 170 cm

- Formato

- Número de teléfono como cadena de caracteres (965- 90 34 00)
 - Número de teléfono como entero (965903400)



Esquema de una arquitectura de DW

El almacén de datos (DW)

- No volátiles

- En sistemas OLTP se pueden modificar datos (ej. tuplas)
 - Unidades de pedido 200; si cliente modifica, se cambia.
- En DW nunca se modifican (salvo excepciones: ver dimensiones que cambian lentamente en tema IV), se añaden nuevos datos para el análisis
 - Un pedido con una fecha de 200 uds.
 - Un pedido con otra fecha de 150 uds.



Esquema de una arquitectura de DW

El almacén de datos (DW)

- Variables en el tiempo

- No volatilidad → Dimensión básica: El TIEMPO

- Datos analizados en función del tiempo

- Ej. Anterior del pedido

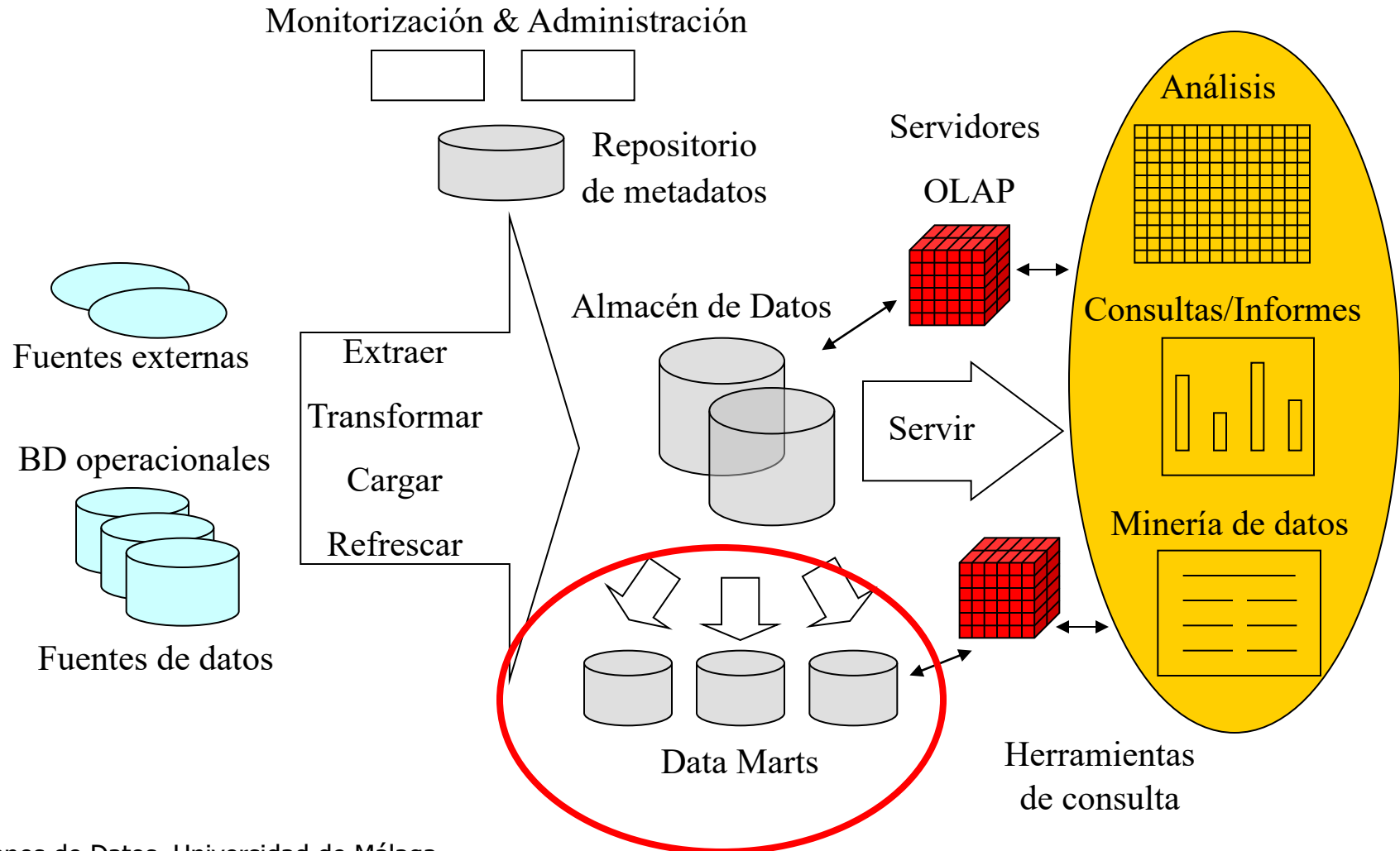
- ¿ Por qué un cliente ha variado la cantidad de su pedido en una semana ?

- ¿ Por qué han cambiado los gustos de un cliente en varios años ?

- Etc.

Esquema de una arquitectura de DW

El almacén de datos (DW)





Esquema de una arquitectura de DW

El almacén de datos (DW)

■ Data Marts

- Es como una vista del almacén de datos
- Se definen para satisfacer las necesidades de un departamento o sección dentro de una empresa
- Normalmente, en la práctica, suelen contener más cantidad de información agrupada que en detalle



Esquema de una arquitectura de DW

El almacén de datos (DW)

- Data Marts.....

- Para su construcción se pueden seguir dos aproximaciones:
 - (I) Definir primero el almacén de datos y, a partir de él, definir los data marts, ó
 - (II) Definir primero los data marts departamentales y, posteriormente integrarlos en un almacén de datos global para la organización
- Nota: Si la envergadura de la empresa es considerable y, la experiencia en construir DW poca, es aconsejable seguir la aproximación 2



Esquema de una arquitectura de DW

El almacén de datos (DW)

■ Resumen: Data warehouse vs. Data Marts

■ Enterprise DW

- Información sobre “temas” de toda organización
- Requiere complejo modelado de negocio
- Puede llevar AÑOS para construir e implementar

■ Data Mart

- Departamental → sub-temas
 - Ej. Marketing data mart, Clientes, productos, ventas !!!!!
- Más rápido agregar
- OJO !!! Integración con DW puede ser compleja



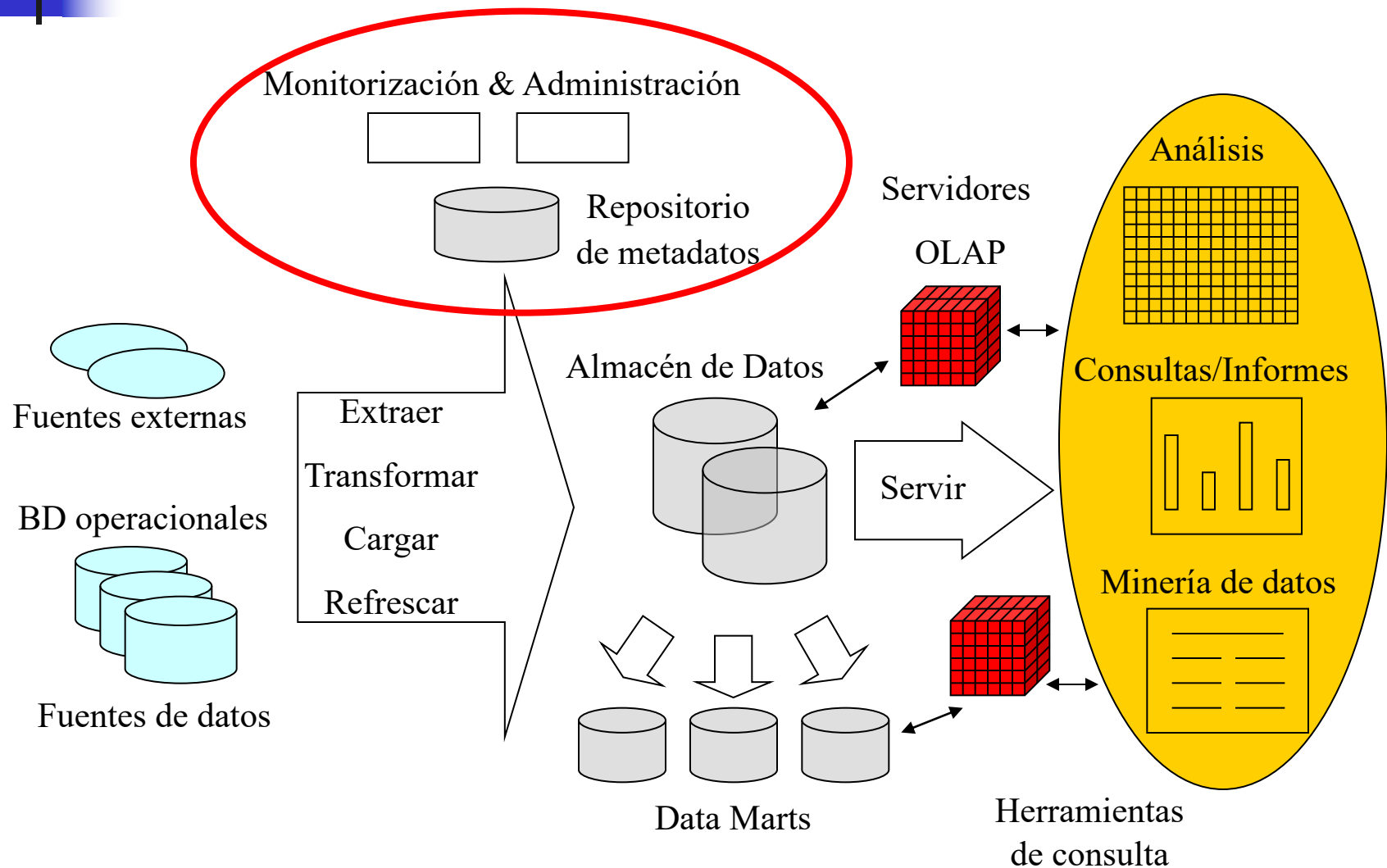
Esquema de una arquitectura de DW

El almacén de datos (DW)

- Virtual Data Warehouse
 - Vistas sobre bases de datos operacionales
 - Materializan algunas vistas agregadas → consultas eficientes
 - Mayor facilidad en la construcción
 - Requieren exceso de capacidad del servidor operacional
 - Muchas empresas dicen tener DW cuando el administrador ha creado un VIRTUAL DW

Esquema de una arquitectura de DW

Los metadatos





Esquema de una arquitectura de DW

Los metadatos

- Los metadatos

- Son datos sobre datos

- Qué dato se guarda (ej. clientes)
- Dónde se guarda (tabla clientes)
- Campos de la tabla
- Con qué datos de las fuentes se corresponden
- Niveles de agregación
- Procesos de carga → ¿ Cuándo se actualizan ?
 - ¿ Cuándo fue la última actualización?
- Patrón de dato válido (Ej. *Apellido1 Apellido2, Nombre*)



Esquema de una arquitectura de DW

Los metadatos

- Los metadatos ...
 - Son datos sobre datos ...
 - Reglas de transformación
 - ¿Cuándo se incorporan al almacén de datos?
 - Y muchos más...



Esquema de una arquitectura de DW

Los metadatos

■ Tipos

- Administrative metadata (Toda la información necesaria para el DW)
 - Fuentes de datos y contenidos
 - Descripciones del gateway
 - Esquema del data warehouse, vistas y datos agregados
 - Dimensiones de análisis con sus jerarquías
 - Consultas e informes predefinidos
 - Localización y contenido de los Data Marts
 - Diseño físico → particionamiento de datos



Esquema de una arquitectura de DW

Los metadatos

- Tipos ...

- Administrative metadata (Toda la información necesaria para el DW) ...
 - Información de los procesos ETL
 - Reglas de refresco
 - Ficheros y grupos de usuarios
 - Seguridad → autorización, control de acceso, etc.



Esquema de una arquitectura de DW

Los metadatos

- Tipos...

- Business metadata

- Información y términos de negocio
 - Políticas de posesión de datos
 - Políticas de permiso de datos por usuarios (seguridad)

- Operational metadata

- Obtiene información recogida durante el proceso del “almacén de datos”
 - Datos migrados y secuencia de transformaciones aplicadas



Esquema de una arquitectura de DW

Los metadatos

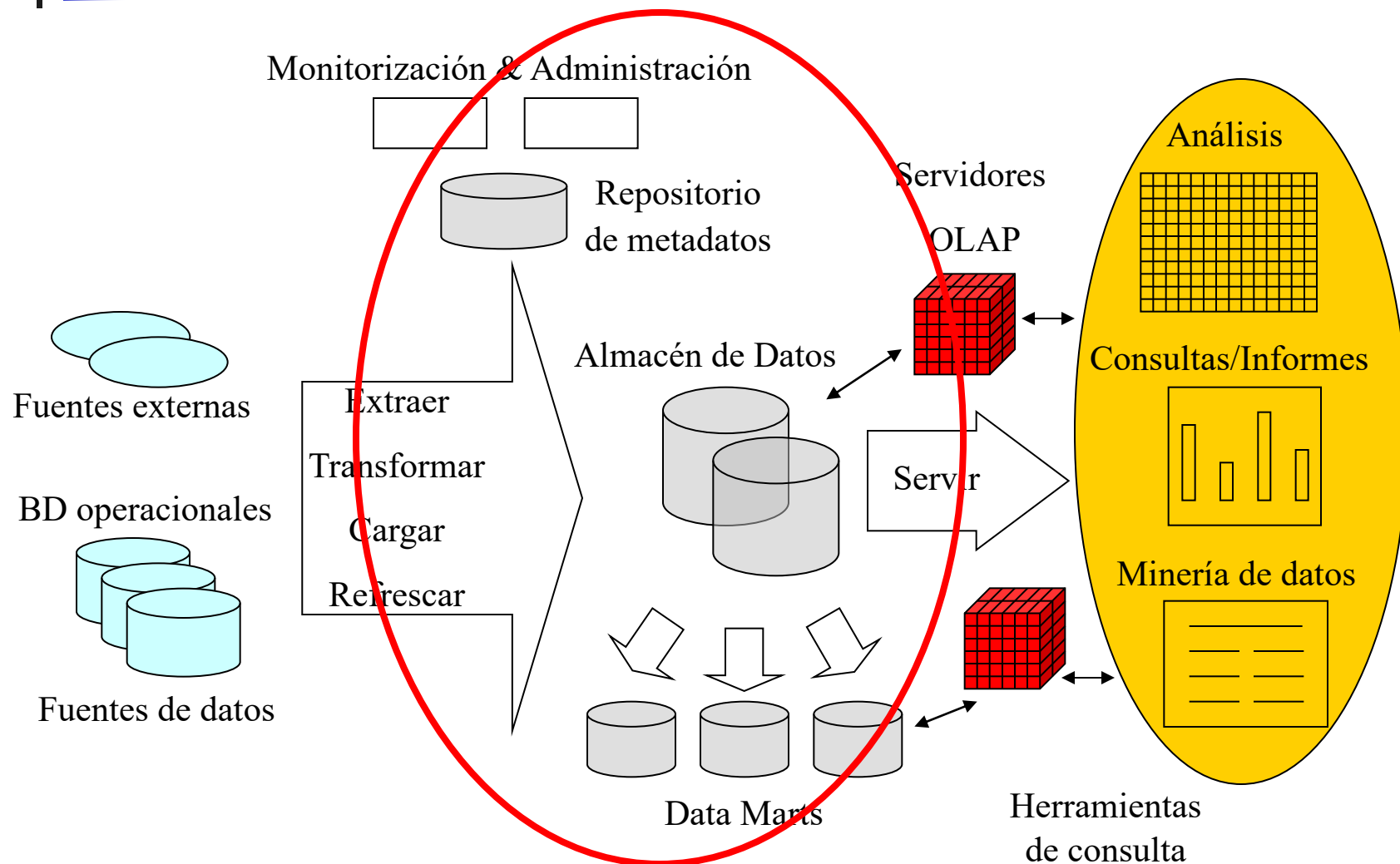
- Tipos....

- Operational metadata...

- Situación de datos: activos, archivados, eliminados ????
 - Información de monitorización
 - Estadísticas de uso de los DW
 - Informes de error
 - Auditoria

Esquema de una arquitectura de DW

Servidor del almacén de datos





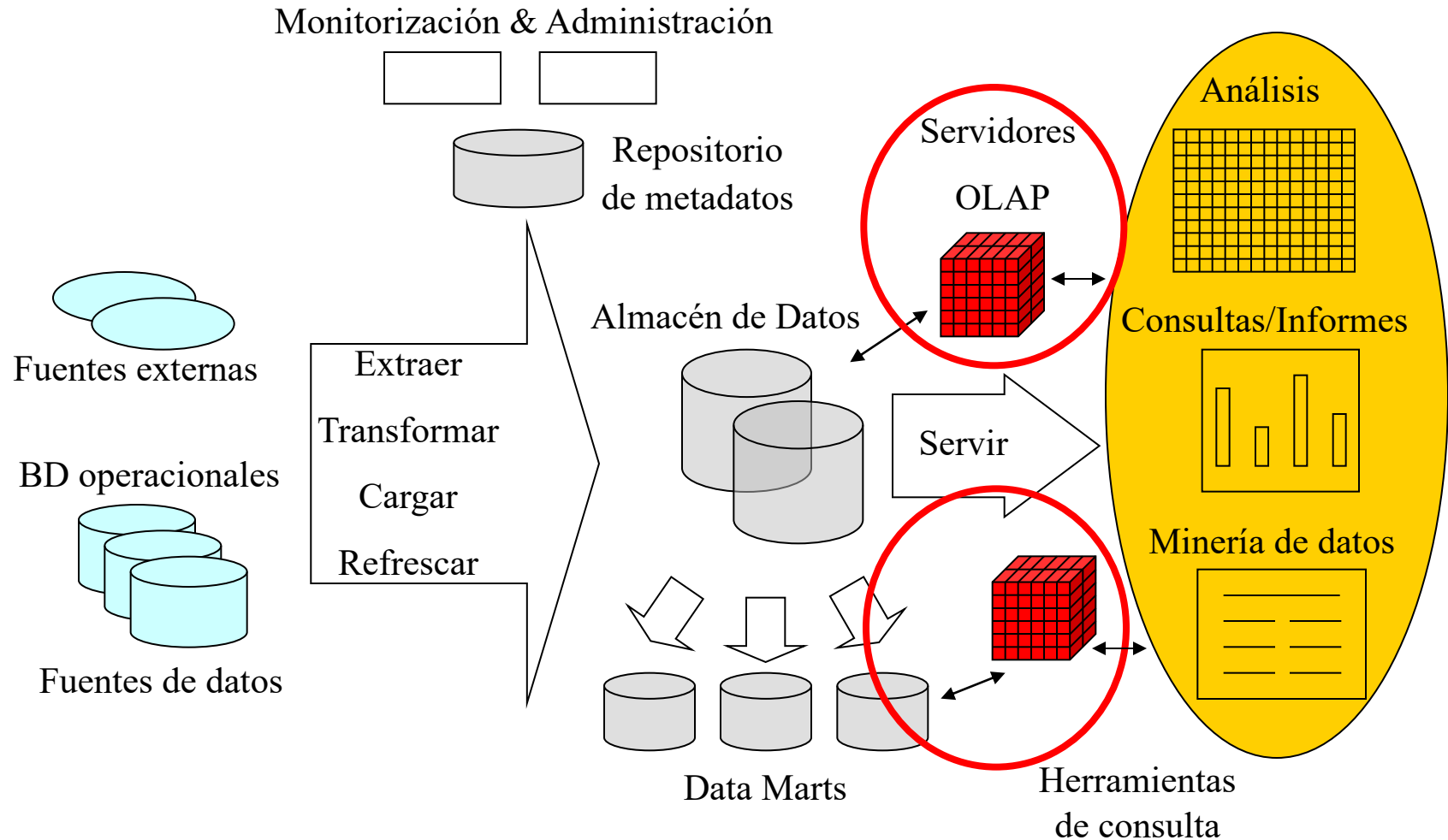
Esquema de una arquitectura de DW

Servidor del almacén de datos

- El servidor es un SGBD que se encarga de
 - Gestionar el repositorio propio del almacén de datos
 - Coordinar los procesos ETL que alimentan el DW
 - Procesan las consultas lanzadas sobre el almacén y devuelven los datos
 - Generalmente son servidores relacionales

Esquema de una arquitectura de DW

Servidores de consultas: OLAP





Esquema de una arquitectura de DW

Servidores de consultas: OLAP

- El servidor de consultas
 - En la mayoría de las arquitecturas se utiliza un servidor distinto al del almacén de datos
 - Rendimiento y mantenimiento
 - La mayoría de herramientas funcionan con esta arquitectura
 - Ejplo. MicroStrategy
 - Dos tecnologías ampliamente utilizadas
 - ROLAP
 - MOLAP



Esquema de una arquitectura de DW

Servidores de consultas: OLAP

■ Servidores ROLAP

- Utilizan tecnología Relacional (Relational OLAP)
- Utilizan extensiones del SQL estándar para soportar el acceso multidimensional a los datos
- Métodos de implementación adecuados para representar los datos multidimensionales en tecnología relacional
- Ventaja: Basado en un estándar (SQL)
- Algunos de los más extendidos
 - **Oracle** (Oracle 9i/10g/11g/12c)
 - IBM (DB2 y Business solutions)
 - Adquirió Informix Dynamic Server como servidor del Almacén de Datos
 - Adquirió Redbrick para gestionar procesos ETL
 - Microsoft SQL Server Analysis Server (modo tabular)



Esquema de una arquitectura de DW

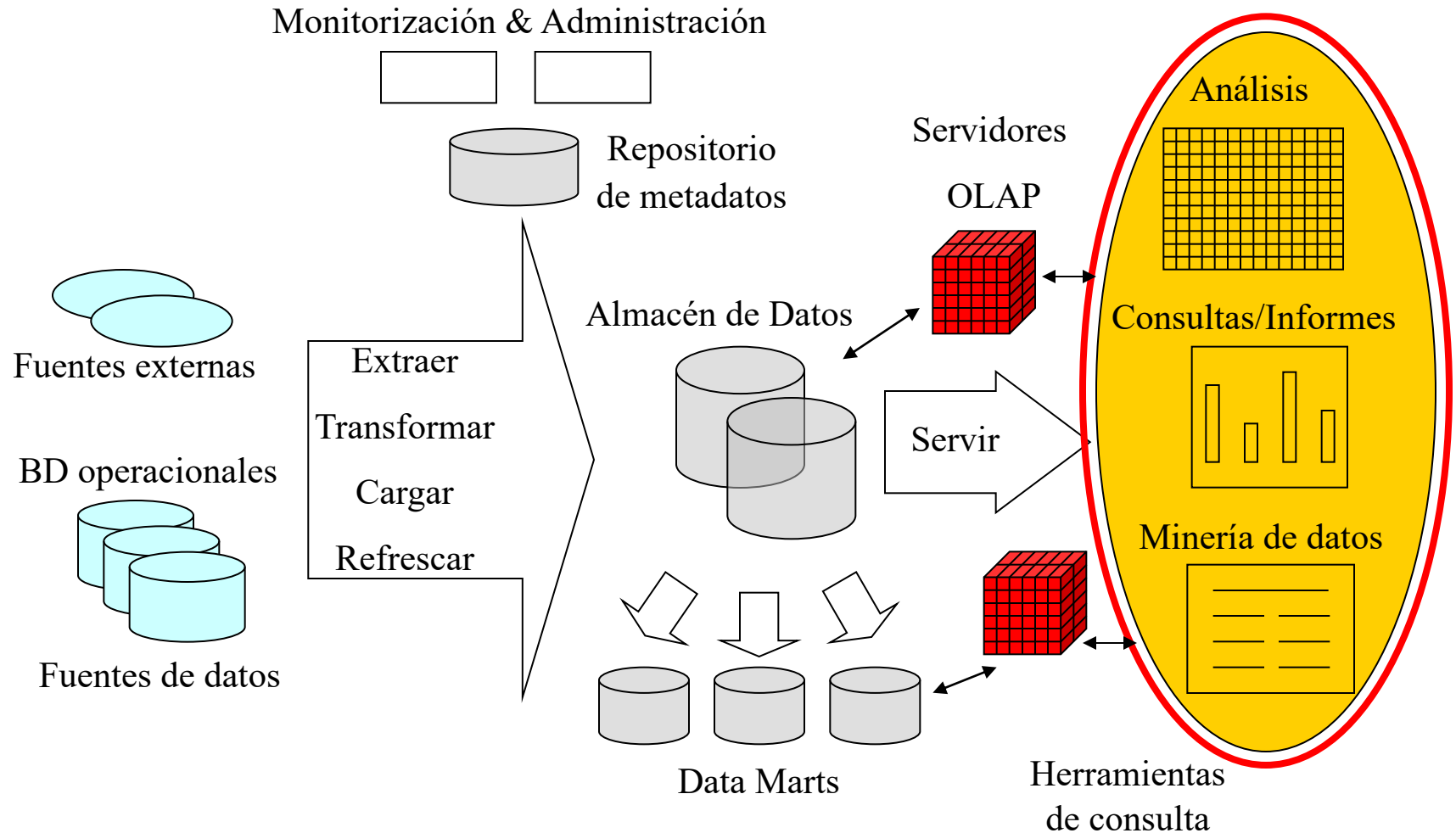
Servidores de consultas: OLAP

■ Servidores MOLAP

- Utilizan tecnología Multidimensional (Multidimensional OLAP)
- Los datos están almacenados directamente en matrices
- Operaciones de consulta están implementadas directamente sobre estas matrices
 - No están basados en SQL estándar
- Ventaja: Suelen ser más rápidos que los ROLAP
- Inconveniente: no basados en un estándar
- Algunos de los más famosos
 - Essabse (Arbor), Accumate (Kenan)
 - Microsoft SQL Server Analysis Server (modo multidimensional)

Esquema de una arquitectura de DW

Herramientas de consulta





Esquema de una arquitectura de DW

Herramientas de consulta

- Generación de informes

- Consultas ad-hoc e informes

- Permiten acceso a los datos base
 - Informe se construye con “point-and-click”
 - Ej. MS Access, Platinum Pro reports

- Entornos consulta/informes de directivos

- Muestra datos base en términos de negocio
 - Utiliza consultas predefinidas/almacenadas
 - Soporte limitado para consultas ad-hoc
 - Informes se pueden presentar como documentos
 - Ej. Busines Objects, SAS/EIS



Esquema de una arquitectura de DW

Herramientas de consulta

- OLAP Query Tools

- Objetivo: Dar soporte a consultas ad-hoc para el analista del negocio
 - Analistas del negocio conocen hojas de cálculo
- Extensión de hojas de cálculo para análisis del DW
 - Enriquecido semánticamente con términos del negocio
 - Análisis multidimensional de los datos
 - Hechos
 - Dimensiones
 - Ej. Oracle (OLAP, Discoverer, etc.), IBM Business Solutions ...



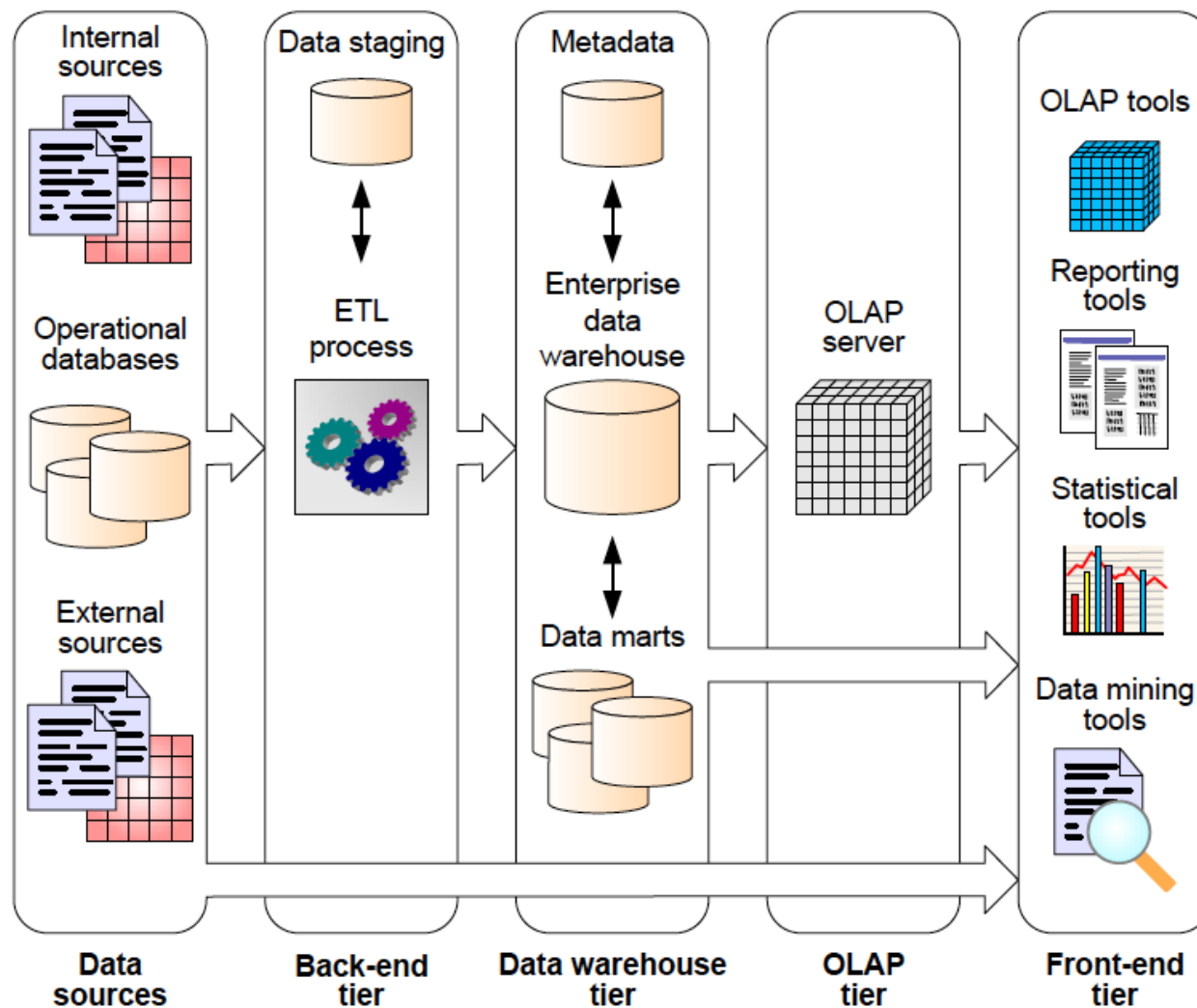
Esquema de una arquitectura de DW

Herramientas de consulta

- Data Mining (Minería de datos)
 - Descubre tendencias y patrones (minería interpretativa)
 - Crean modelos y hacen predicciones (minería predictiva)
 - Utilizan los datos existentes para detectar tendencias y crear modelos
 - Aplican los patrones y modelos a datos nuevos
 - Importante → Adaptación de modelos según nuevos datos

Esquema de una arquitectura de DW

Representación





Indice

- Arquitectura de un Data Warehouse
- Herramientas de Microsoft SQL Server
- El Modelo Multidimensional
- Operaciones OLAP



Herramientas de SQL Server

- SQL Server: Plataforma integrada para construir aplicaciones analíticas
- Componentes principales:
 - Analysis Services: Herramienta OLAP capacidades analíticas y de minería de datos
 - Permite definir, consultar, modificar y gestionar bases de datos OLAP
 - MDX (Multi-Dimensional eXpressions) lenguaje para recuperar datos
 - Acceso a datos OLAP mediante herramientas cliente (Excel u otros)
 - Proporciona algoritmos de minería de datos. DMX (Data Mining eXtensions) lenguaje para crear y consultar modelos de minería de datos
 - Integration Services: Servicios ETL
 - Reporting Services: Define, genera, almacena y gestiona informes
 - Desde distintas fuentes, incluidos cubos OLAP



Herramientas de SQL Server

- SQL Server 2016:
 - SQL Server Data Tools, plataforma de desarrollo integrada con Microsoft Visual Studio
 - Soporta proyectos de Analysis Services, Reporting Services, e Integration Services
 - SQL Server Management Studio proporciona Gestión integrada de todos los componentes SQL Server
- Business Intelligence Semantic Model (BISM):
 - Multidimensional. Aplicaciones BI y grandes volúmenes de datos. Lenguaje: MDX
 - Tabular. Más fácil de entender y más rápido de construir. Lenguaje: DAX

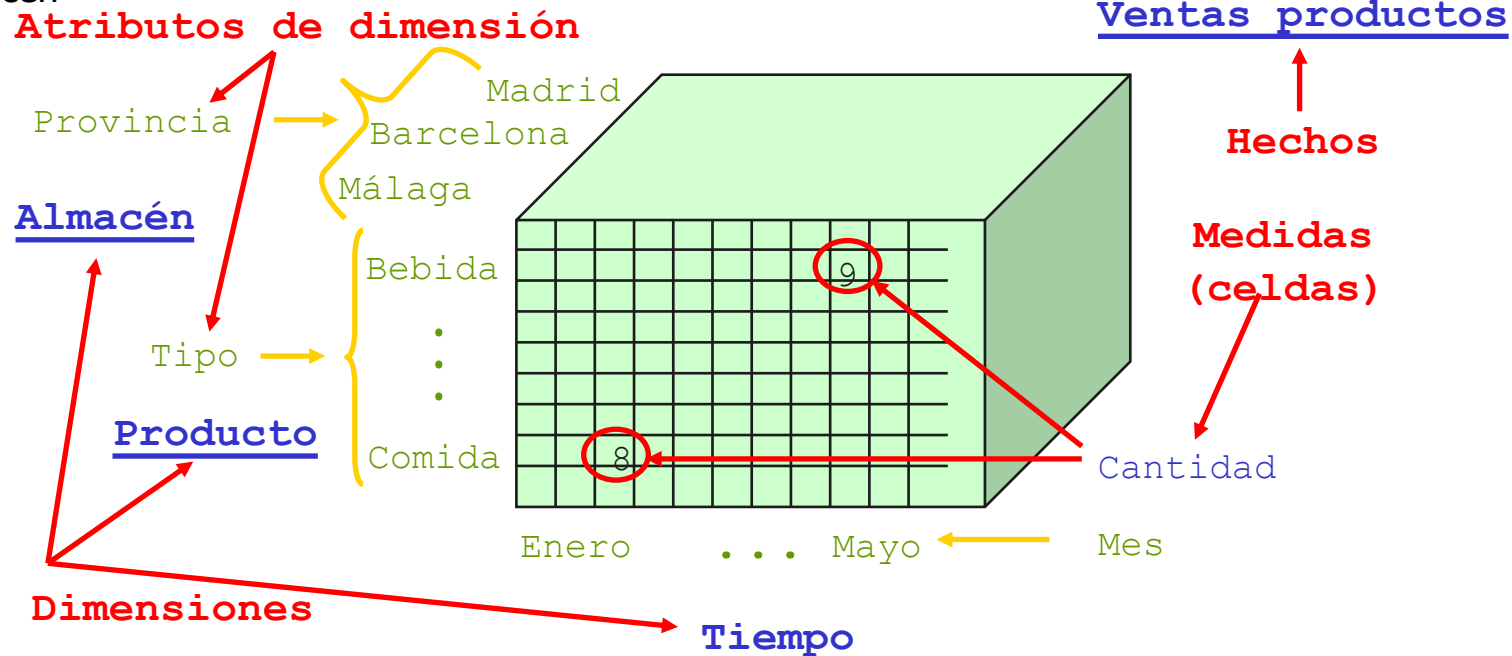


Indice

- Arquitectura de un Data Warehouse
- Herramientas de Microsoft SQL Server
- El Modelo Multidimensional
- Operaciones OLAP

Modelo Multidimensional

- Vistas de datos en un espacio n-dimensional: Cubo de datos
- Se compone de dimensiones y hechos
 - Ejemplo: Cubo tridimensional para los datos de ventas con las dimensiones: Producto, Tiempo y Almacén



- Los atributos describen las dimensiones. Producto tiene Tipo, Precio y Código.
- Las celdas o hechos tienen datos numéricos: medidas
- En el ejemplo, cada dato del cubo representa Cantidad de unidades vendidas por Provincia, Tipo de Producto y Mes



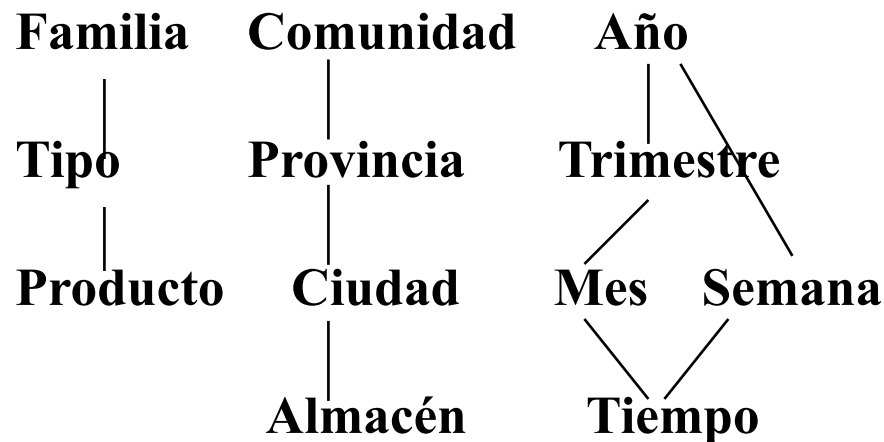
Modelo Multidimensional

- Granularidad. Nivel de detalle al que se representan las medidas para cada dimensión del cubo.
 - Por ejemplo: Cifras de Ventas agregadas por Tipo, Trimestre y Ciudad.
- Las instancias de una dimensión se denominan Miembros.
 - Por ejemplo: Comida, Bebida son miembros de Producto en la granularidad Tipo
- Un cubo contiene varias medidas. P.e. Cantidad indica la cantidad de ventas total.
- Un cubo puede ser disperso (lo normal) o denso.
 - P.e. No todos los clientes han pedido productos de todos los tipos en todos los meses
- Jerarquías. Permite ver los datos en distintas granularidades
 - Define una sucesión de relaciones que van desde los detallados de nivel inferior, a los de nivel más alto
 - El nivel inferior se llama hijo (child) y el nivel más alto se llama el padre
 - La estructura jerárquica de una dimensión se llama el esquema de la dimensión
 - Una instancia de la dimensión comprende todos los miembros de todos los niveles de la dimensión
 - Podemos querer cifras de ventas con una granularidad más fina (un día), o con una granularidad más gruesa (País)



Modelo Multidimensional

- Dimensiones: Producto, Almacén, Tiempo. Caminos de jerarquía por los que agregar





Medidas (atributos del hecho)

- Atómicos

- Ej. Cantidad vendida, precio, etc.

- Derivados

- Utilizan una fórmula para calcularlos
- Ej. $\text{Precio_total} = \text{precio} * \text{cantidad_vendida}$

- Aditividad

- Conjunto de operadores de agregación que se pueden aplicar para agregar los valores de medidas a lo largo de las jerarquías
- Es aditiva → SUM sobre todas las dimensiones
- Semi-aditiva → SUM sólo sobre algunas dimensiones. Inventario → Tiempo
- No aditiva → SUM sobre ninguna dimensión. Precio, Coste por unidad



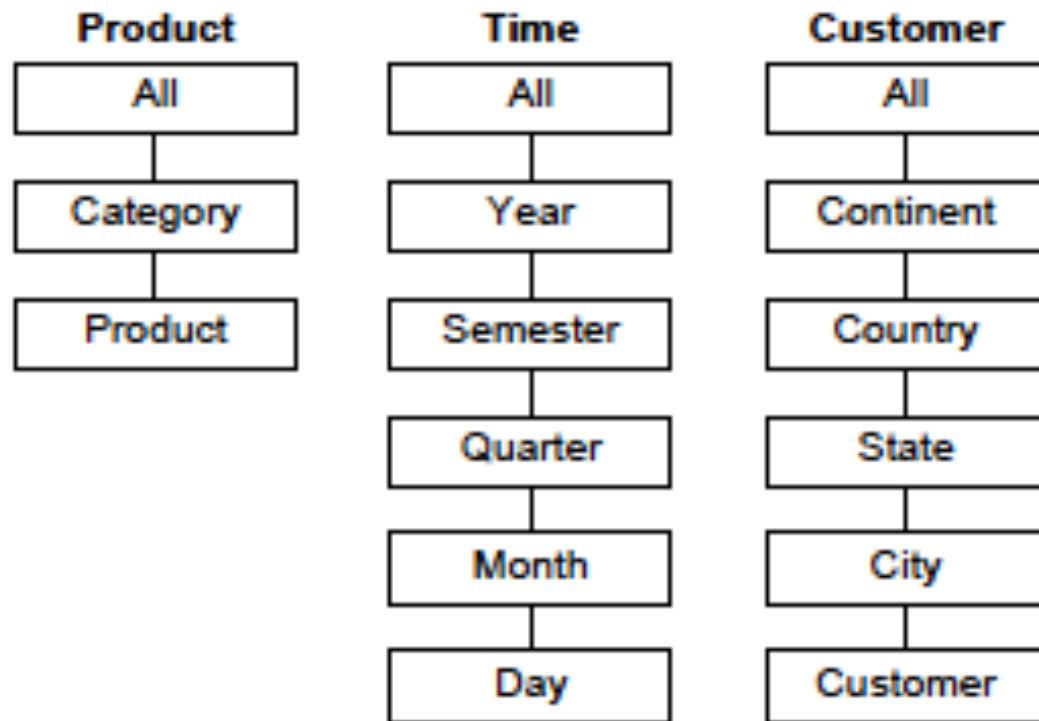
Indice

- Arquitectura de un Data Warehouse
- Herramientas de Microsoft SQL Server
- El Modelo Multidimensional
- Operaciones OLAP



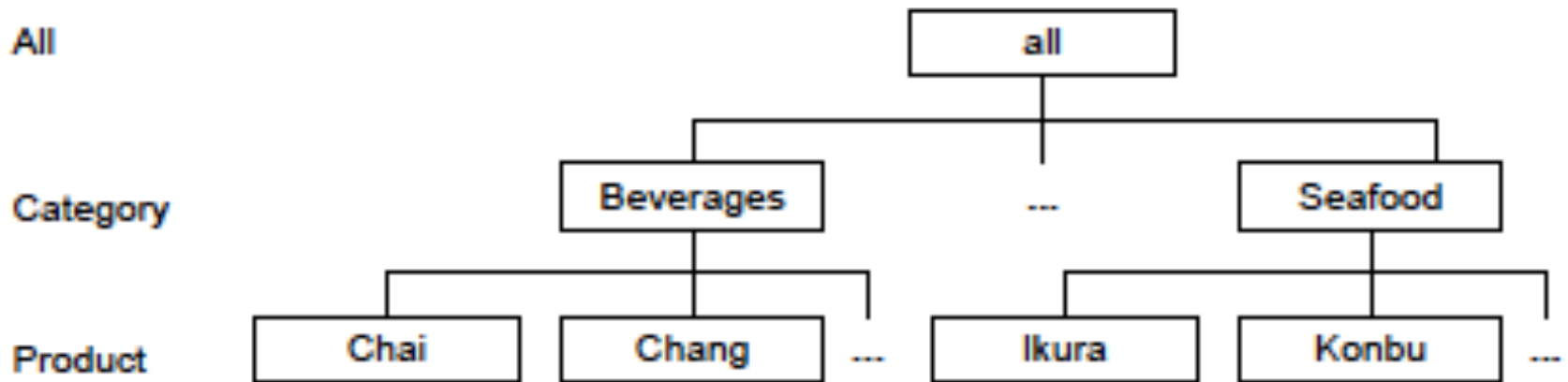
Sobre jerarquías (I)

- Jerarquías de las dimensiones Product, Time and Customer



Sobre jerarquías (II)

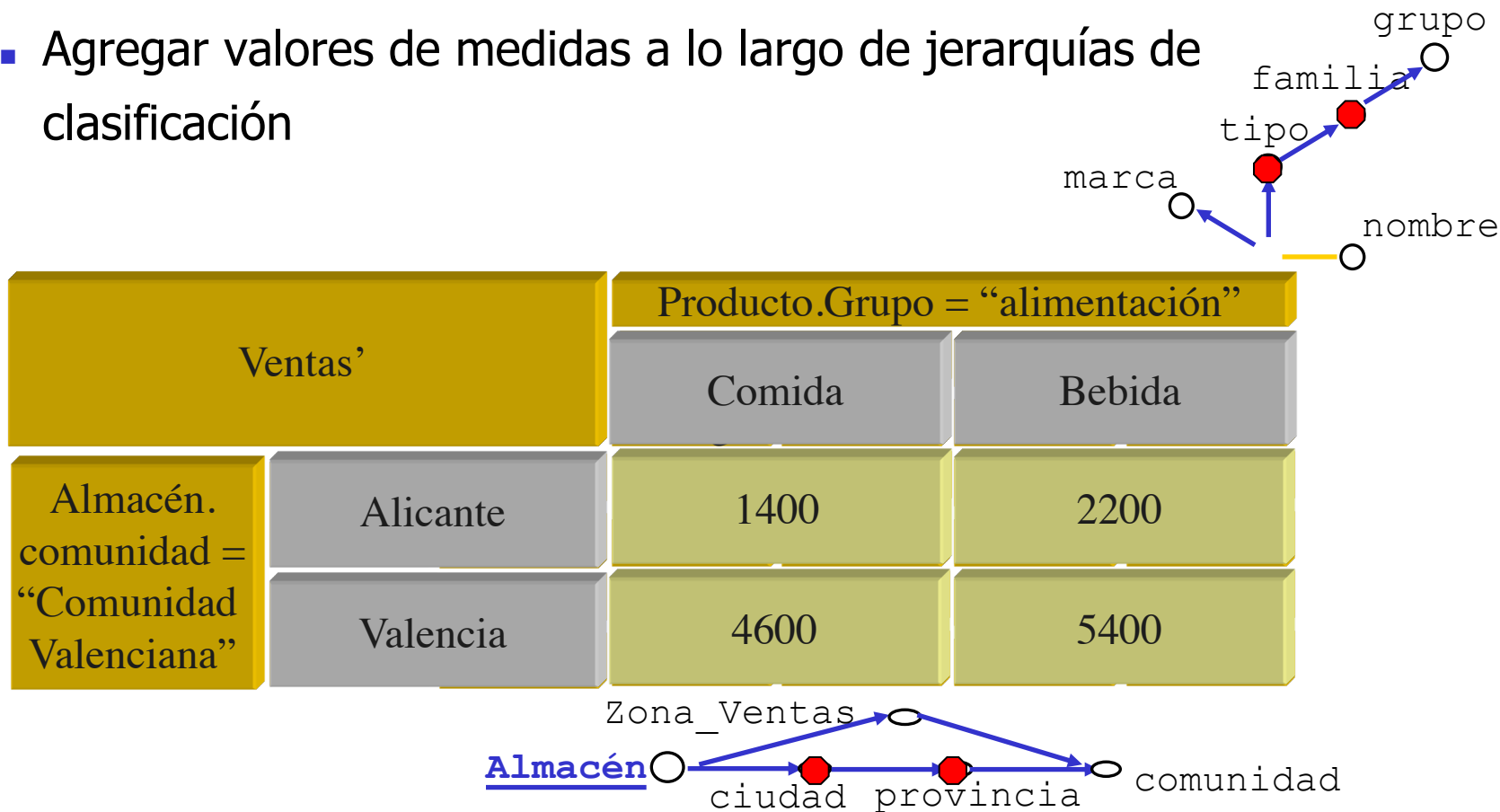
- Miembros de una jerarquía Producto -> Categoría



Operaciones OLAP

■ Roll-up

- Agregar valores de medidas a lo largo de jerarquías de clasificación



Operaciones OLAP

■ Roll-up

ROLLUP(Sales, Customer -> Country, SUM(Quantity))

Time (Quarter)	Customer (City)	Product (Category)			
		Produce	Seafood	Beverages	Condiments
Q1	Köln	21	10	18	35
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35
Q2	Köln	27	14	11	30
	Berlin	30	14	23	17
	Lyon	32	12	20	18
	Paris	31	10	33	18
Q3	Köln	26	12	35	32
	Berlin	32	10	33	18
	Lyon	31	10	33	18
	Paris	31	10	33	18
Q4	Köln	14	20	47	31
	Berlin	31	10	33	18
	Lyon	31	10	33	18
	Paris	31	10	33	18

Original cube

Roll-up to the Country level

ROLLUP(CubeName, (Dimension -> Level)*, AggFunction(Measure)*)

Roll-up extendido: las dimensiones no indicadas se hace un roll-up hacia "all"

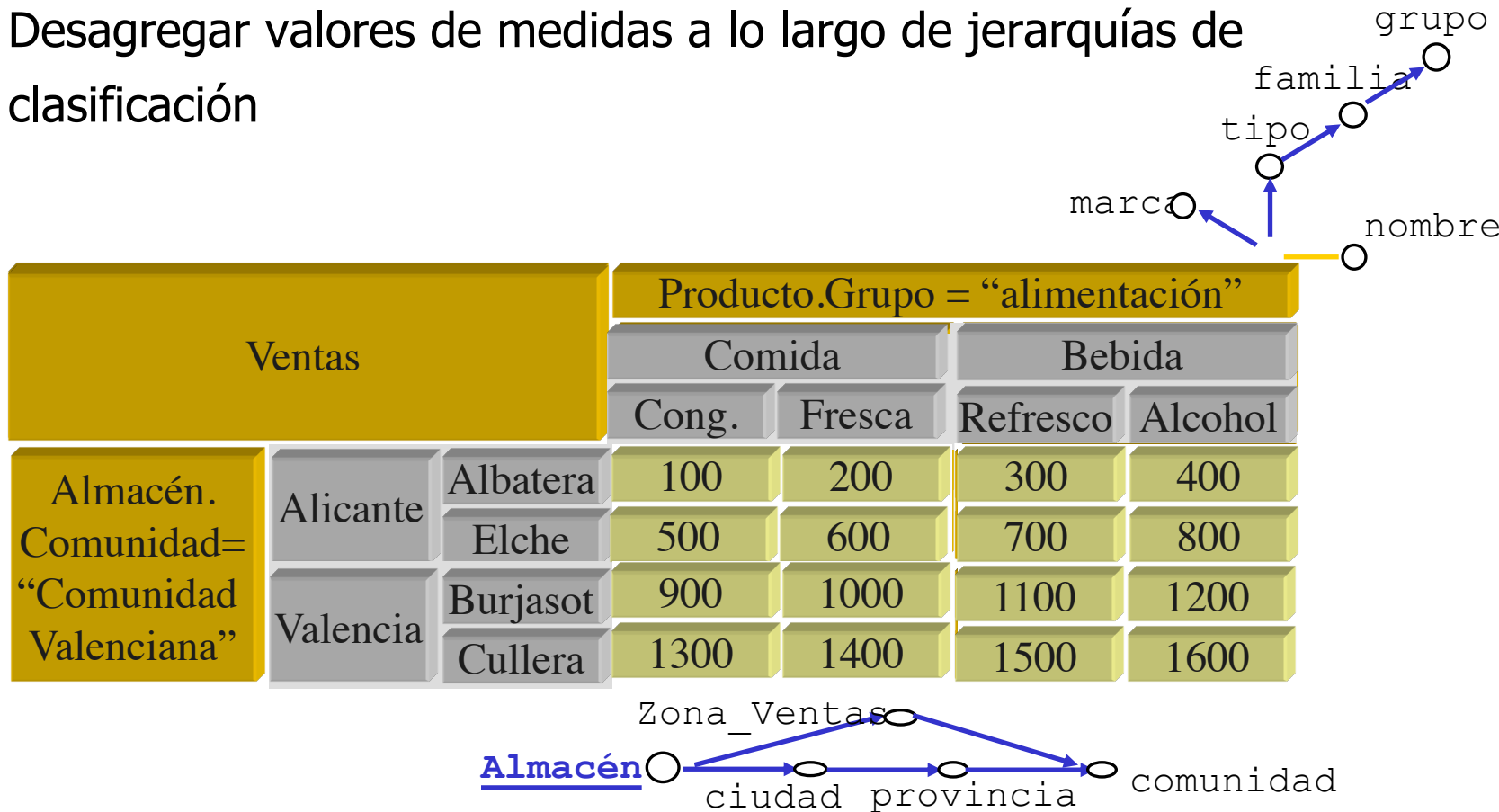
¿Qué hace?

ROLLUP*(Sales, Customer -> Country, SUM(Quantity))

Operaciones OLAP

■ Drill-down

- Desagregar valores de medidas a lo largo de jerarquías de clasificación



Operaciones OLAP

■ Drill-down

DRILLDOWN(Sales, Time -> Month)

Time (Quarter)	Customer (City)	Product (Category)			
		Produce		Seafood	
		Beverages	Condiments		
	Köln	24	18	28	14
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35
Q1		21	10	18	35
Q2		27	14	11	30
Q3		26	12	35	32
Q4		14	20	47	31

Original cube

Time (Quarter)	Customer (City)	Product (Category)			
		Produce		Seafood	
		Beverages	Condiments		
	Köln	8	6	9	5
	Berlin	10	8	11	8
	Lyon	4	7	8	14
	Paris	7	2	6	20
Jan		7	2	6	20
Feb		8	4	8	8
Mar		6	4	4	7
...	
Dec		4	4	16	7

Drill-down to the Month level

DRILLDOWN(CubeName, (Dimension -> Level)*)

Operaciones OLAP

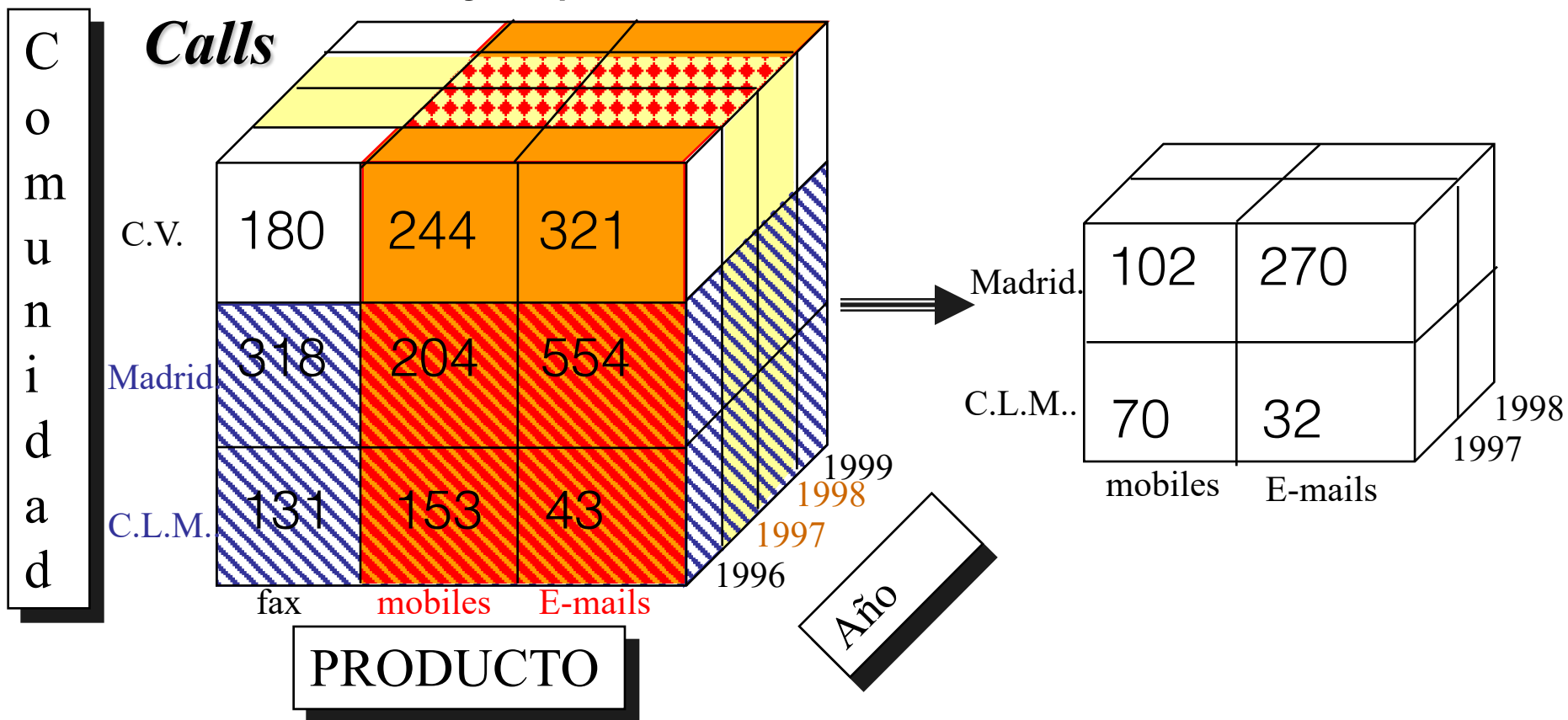
- Slice-dice
 - Definir restricciones sobre niveles de jerarquías
 - Ej. Analizar datos donde el año sea 1999, la ciudad "Alicante" y el tipo de Producto sea "Comida"

Ventas'			Producto.Grupo = "Alimentación"	
			Comida	
			Congelada	Fresca
Almacén. Comunidad = "Comunidad Valenciana"	Alicante	Albatera	100	200
		Elche	500	600

Operaciones OLAP

■ Slice-dice (cont.)

■ Otro ejemplo



Operaciones OLAP

■ Slice-dice (cont.)

SLICE(CubeName,
Dimension, Level = Value)

DICE(CubeName, condición_booleana)

DICE(Sales, (Customer.City = 'Paris'
OR Customer.City = 'Lyon') AND
(Time.Quarter = 'Q1' OR
Time.Quarter = 'Q2'))

SLICE(Sales, Customer, City = 'Paris')

Product (Category)	Customer (City)	Time (Quarter)			
		Q1	Q2	Q3	Q4
	Paris	21	27	26	14
	Lyon	12	14	11	13
	Berlin	33	28	35	32
	Köln	24	23	25	18
		Product (Category)			
		Produce	Seafood	Beverages	Condiments

Time (Quarter)	Product (Category)			
	Produce	Seafood	Beverages	Condiments
Q1	21	10	18	35
Q2	27	14	11	30
Q3	26	12	35	32
Q4	14	20	47	31

Pivot

Slice on City='Paris'

Time (Quarter)	Customer (City)	Product (Category)			
		Produce	Seafood	Beverages	Condiments
	Lyon	12	20	24	33
Q1	Paris	21	10	18	35
		35	14		
Q2		27	14	11	30
		30	20		

Dice on City='Paris' or 'Lyon' and Quarter='Q1' or 'Q2'

Operaciones OLAP

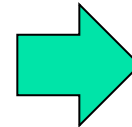
■ Pivoting

PIVOT(CubeName, (Dimension -> Axis)*)

- Reorientar la vista multidimensional de los datos, es decir, cambiar la distribución de filas/columnas
- Algunos autores consideran también el intercambio de medidas y hechos como pivoting (kimball, 1996) (Inmon, 1996)

	Customer (City)	Köln	24	18	28	14
		Berlin	33	25	23	25
		Lyon	12	20	24	33
		Paris	21	10	18	35
Time (Quarter)	Q1	21	10	18	35	35
	Q2	27	14	11	30	30
	Q3	26	12	35	32	32
	Q4	14	20	47	31	31
			Produce	Seafood		
			Beverages	Condiments		
			Product (Category)			

Original cube



	Product (Category)	Seafood	35	30	32	31
		Condiments	18	11	35	47
		Produce	10	14	12	20
		Beverages	21	27	26	14
Customer (City)	Paris	21	27	26	14	14
	Lyon	12	14	11	13	13
	Berlin	33	28	35	32	32
	Köln	24	23	25	18	18
			Q1	Q2	Q3	Q4
			Time (Quarter)			

Pivot

PIVOT(Sales, Time -> X, Customer -> Y, Product -> Z)



Operaciones OLAP

■ Drill-across

- Dados dos cubos se construye un nuevo cubo con las medidas de ambos en cada celda
- Consultar medidas de varios hechos en el mismo cubo
 - Ej. Que en la tabla MD analizáramos el ratio de ventas respecto de compras.
 - $1000 / 400$
- OJO: será necesario tener nombres de cubos diferentes y medidas diferentes -> uso de RENAME

Operaciones OLAP: Drill-across y Rename

RENAME(CubeName, (SchemaElement -> NewName)*)

DRILLACROSS(CubeName1, CubeName2, [Condition])

Time (Quarter)	Customer (City)	Product (Category)			
		Produce		Seafood	
		Beverages	Condiments	Beverages	Condiments
Q1	Köln	21	10	18	35
Q2	Köln	27	14	11	30
Q3	Köln	26	12	35	32
Q4	Köln	14	20	47	31
Q1	Berlin	33	25	23	25
Q2	Berlin	30	12	10	29
Q3	Berlin	28	11	31	28
Q4	Berlin	12	22	45	29
Q1	Lyon	12	20	24	33
Q2	Lyon	14	18	22	28
Q3	Lyon	19	12	31	28
Q4	Lyon	19	12	31	28
Q1	Paris	21	10	18	35
Q2	Paris	30	12	10	29
Q3	Paris	28	11	31	28
Q4	Paris	12	22	45	29

Original cube

Time (Quarter)	Customer (City)	Product (Category)			
		Produce		Seafood	
		Beverages	Condiments	Beverages	Condiments
Q1	Köln	19	12	31	28
Q2	Köln	30	12	10	29
Q3	Köln	28	11	31	28
Q4	Köln	12	22	45	29
Q1	Berlin	30	22	21	26
Q2	Berlin	29	16	21	19
Q3	Berlin	28	14	20	19
Q4	Berlin	29	12	31	19
Q1	Lyon	14	18	22	28
Q2	Lyon	14	18	22	28
Q3	Lyon	19	12	31	28
Q4	Lyon	19	12	31	28
Q1	Paris	19	12	31	28
Q2	Paris	30	12	10	29
Q3	Paris	28	11	31	28
Q4	Paris	12	22	45	29

Cube for 2011

Time (Quarter)	Customer (City)	Product (Category)			
		Produce		Seafood	
		Beverages	Condiments	Beverages	Condiments
Q1	Köln	19	12	31	28
Q2	Köln	30	12	10	29
Q3	Köln	28	11	31	28
Q4	Köln	12	22	45	29
Q1	Berlin	30	22	21	26
Q2	Berlin	29	16	21	19
Q3	Berlin	28	14	20	19
Q4	Berlin	29	12	31	19
Q1	Lyon	14	18	22	28
Q2	Lyon	14	18	22	28
Q3	Lyon	19	12	31	28
Q4	Lyon	19	12	31	28
Q1	Paris	19	12	31	28
Q2	Paris	30	12	10	29
Q3	Paris	28	11	31	28
Q4	Paris	12	22	45	29

Drill-across operator

RENAME(Sales, Sales -> Sales2012, Quantity -> Quantity2012)
Sales2011-2012 <- DRILLACROSS(Sales2011, Sales2012)



Operaciones OLAP: AddMeasure, DropMeasure

- Es posible calcular nuevas medidas a partir de las existentes con ADDMEASURE

```
ADDMEASURE(CubeName, (NewMeasure = Expression)* )
```

- Estas medidas se agregarán a las existentes en el hecho
- Si queremos borrar alguna medida que no nos interese utilizaremos DROPMEASURE

```
DROPMEASURE(CubeName, Measure* )
```

Operaciones OLAP: AddMeasure, DropMeasure

- Podemos añadir el porcentaje de cambio en las ventas entre los años 2011 y 2012:

```
ADDMEASURE(Sales2011-2012, PercChange = (Quantity2011-Quantity2012)/Quantity2012)
```

Drill-across operator

Time (Quarter)	Customer (City)	Product (Category)							
		Produce		Seafood		Beverages		Condiments	
Q1	Köln	20	24	22	18	24	28	16	14
	Berlin	30	33	22	25	21	26	26	14
	Lyon	14	18	22	23	28	25	18	14
	Paris	19	12	31	24	33	28	18	14
Q2	Köln	12	10	18	35	29	30	16	14
	Berlin	30	27	12	14	10	11	29	30
	Lyon	28	26	11	12	31	35	28	32
	Paris	12	14	10	11	29	30	16	14
Q3	Köln	28	32	12	10	31	33	29	31
	Berlin	28	32	12	10	31	33	29	31
	Lyon	28	32	12	10	31	33	29	31
	Paris	28	32	12	10	31	33	29	31
Q4	Köln	12	14	22	20	45	47	29	31
	Berlin	12	14	22	20	45	47	29	31
	Lyon	12	14	22	20	45	47	29	31
	Paris	12	14	22	20	45	47	29	31

Percentage change

Time (Quarter)	Customer (City)	Product (Category)							
		Produce		Seafood		Beverages		Condiments	
Q1	Köln	20	-18	17	-13	10	14	10	-4
	Berlin	10	14	10	-4	10	14	10	-4
	Lyon	-14	11	9	18	10	14	10	-4
	Paris	11	-17	-42	25	18	14	10	-4
Q2	Köln	11	-17	-42	25	18	14	10	-4
	Berlin	11	-17	-42	25	18	14	10	-4
	Lyon	11	-17	-42	25	18	14	10	-4
	Paris	11	-17	-42	25	18	14	10	-4
Q3	Köln	-10	17	10	3	3	-13	0	-11
	Berlin	-10	17	10	3	3	-13	0	-11
	Lyon	-10	17	10	3	3	-13	0	-11
	Paris	-10	17	10	3	3	-13	0	-11
Q4	Köln	-7	9	13	14	14	-14	6	-5
	Berlin	-7	9	13	14	14	-14	6	-5
	Lyon	-7	9	13	14	14	-14	6	-5
	Paris	-7	9	13	14	14	-14	6	-5

Si queremos borrar las medidas anteriores:

```
DROPMEASURE(Sales2011-2012, Quantity2011, Quantity2012)
```

Operaciones OLAP: Funciones de agregación

- Las funciones de agregación en OLAP pueden ser:
 - Acumulativo: calcular el valor de la medida de una celda partir de otras celdas; algunos ejemplos son SUM, COUNT y AVG
 - Filtrado: Filtra los miembros de una dimensión que aparecen en el resultado; algunos ejemplos son MIN y MAX
 - Las funciones de filtrado no sólo calculan el valor agregado, sino también los miembros de la dimensión que pertenecen al resultado (e.g. el valor máximo y el elemento que proporciona ese máximo)
- Para agregar medidas de un cubo en **la granularidad actual** sin realizar un roll-up:

AggFunction(CubeName, Measure) [BY Dimension*]

- Ejemplo: Total de ventas por trimestre y ciudad:

SUM(Sales, Quantity) BY Time, Customer

Time (Quarter)	Q1	84	89	106	84
	Q2	82	77	93	79
	Q3	105	72	65	88
	Q4	112	61	96	102
			Lyon		Köln
		Paris		Berlin	
		Customer (City)			



Operaciones OLAP: Funciones de agregación

- Agregación:
 - Ejemplo "Cantidad **total** global": produce una sola celda cuyas coordenadas para las tres o n dimensiones son iguales a *a//*
`SUM(Sales, Quantity)`
 - Agregación sin cambiar granularidad:
 - Máximas de ventas por trimestre y ciudad
`MAX(Sales, Quantity) BY Time, Customer`
 - produciendo un cubo en el que sólo las celdas que contienen el máximo por trimestre y ciudad tendrán valores, los demás serán **null**
 - Los dos primeros máximos de ventas por producto y de ciudad
`MAX(Sales, Quantity, 2) BY Time, Customer`
 - Promedio de una ventana de tres meses para las ventas:
`ADDMEASURE(Sales, MovAvg = AVG(Quantity) OVER Time 2 CELLS PRECEDING)`
 - Suma de año hasta la fecha:

`ADDMEASURE(Sales, YTDQuantity = SUM(Quantity) OVER Time ALL CELLS PRECEDING)`

- La ventana contiene la celda actual y todas las anteriores (indicado por ALL CELLS PRECEDING)

Operaciones OLAP: Funciones de agregación

Original

Customer (City)	Time (Quarter)	Product (Category)			
		Produce	Beverages	Condiments	Seafood
Köln	Q1	21	10	18	35
Berlin	Q1	33	25	23	25
Lyon	Q1	12	20	24	33
Paris	Q1	21	10	18	35
Köln	Q2	27	14	11	30
Berlin	Q2	33	25	23	25
Lyon	Q2	12	20	24	33
Paris	Q2	21	10	18	35
Köln	Q3	26	12	35	32
Berlin	Q3	33	25	23	25
Lyon	Q3	12	20	24	33
Paris	Q3	21	10	18	35
Köln	Q4	14	20	47	31
Berlin	Q4	33	25	23	25
Lyon	Q4	12	20	24	33
Paris	Q4	21	10	18	35

Customer (City)	Time (Quarter)	Product (Category)			
		Produce	Beverages	Condiments	Seafood
Köln	Q1	35	35	35	23
Berlin	Q1	33	25	23	25
Lyon	Q1	12	20	24	33
Paris	Q1	21	10	18	35
Köln	Q2	30	30	30	23
Berlin	Q2	33	25	23	25
Lyon	Q2	12	20	24	33
Paris	Q2	21	10	18	35
Köln	Q3	35	35	35	23
Berlin	Q3	33	25	23	25
Lyon	Q3	12	20	24	33
Paris	Q3	21	10	18	35
Köln	Q4	47	47	47	31
Berlin	Q4	33	25	23	25
Lyon	Q4	12	20	24	33
Paris	Q4	21	10	18	35

Maximum sales by quarter and city

Customer (City)	Time (Quarter)	Product (Category)			
		Produce	Beverages	Condiments	Seafood
Köln	Q1	35	35	35	23
Berlin	Q1	33	25	23	25
Lyon	Q1	12	20	24	33
Paris	Q1	21	10	18	35
Köln	Q2	30	30	30	23
Berlin	Q2	33	25	23	25
Lyon	Q2	12	20	24	33
Paris	Q2	21	10	18	35
Köln	Q3	35	35	35	23
Berlin	Q3	33	25	23	25
Lyon	Q3	12	20	24	33
Paris	Q3	21	10	18	35
Köln	Q4	47	47	47	31
Berlin	Q4	33	25	23	25
Lyon	Q4	12	20	24	33
Paris	Q4	21	10	18	35

Top two sales by quarter and city

Customer (City)	Time (Month)	Product (Category)			
		Produce	Beverages	Condiments	Seafood
Köln	Jan	7	2	6	20
Berlin	Jan	10	8	11	8
Lyon	Jan	4	7	8	14
Paris	Jan	7	2	6	20
Köln	Feb	8	4	8	8
Berlin	Feb	10	8	11	8
Lyon	Feb	4	7	8	14
Paris	Feb	7	2	6	20
Köln	Mar	6	4	4	7
Berlin	Mar	10	8	11	8
Lyon	Mar	4	7	8	14
Paris	Mar	7	2	6	20
Köln
Berlin
Lyon
Paris
Köln	Dec	4	4	16	7
Berlin	Dec	10	8	11	8
Lyon	Dec	4	7	8	14
Paris	Dec	7	2	6	20

Customer (City)	Time (Quarter)	Product (Category)			
		Produce	Beverages	Condiments	Seafood
Köln	Jan	7	2	6	20
Berlin	Jan	10	8	11	8
Lyon	Jan	4	7	8	14
Paris	Jan	7	2	6	20
Köln	Feb	7.5	3	7	14
Berlin	Feb	10	8	11	8
Lyon	Feb	4	7	8	14
Paris	Feb	7	2	6	20
Köln	Mar	7	3.3	6	11.6
Berlin	Mar	10	8	11	8
Lyon	Mar	4	7	8	14
Paris	Mar	7	2	6	20
Köln
Berlin
Lyon
Paris
Köln	Dec	4.5	6	12	9
Berlin	Dec	10	8	11	8
Lyon	Dec	4	7	8	14
Paris	Dec	7	2	6	20

Three-months moving average

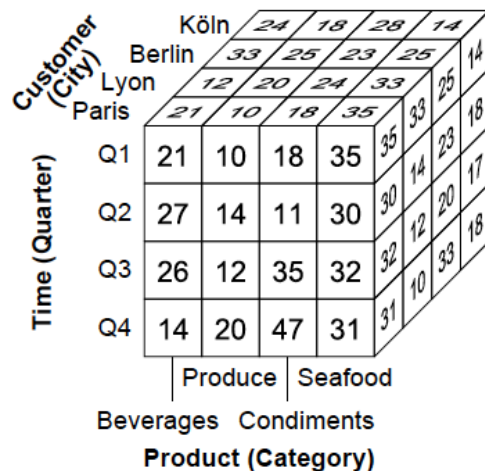
Customer (City)	Time (Quarter)	Product (Category)			
		Produce	Beverages	Condiments	Seafood
Köln	Jan	7	2	6	20
Berlin	Jan	10	8	11	8
Lyon	Jan	4	7	8	14
Paris	Jan	7	2	6	20
Köln	Feb	15	6	14	26
Berlin	Feb	10	8	11	8
Lyon	Feb	4	7	8	14
Paris	Feb	7	2	6	20
Köln	Mar	21	10	18	33
Berlin	Mar	10	8	11	8
Lyon	Mar	4	7	8	14
Paris	Mar	7	2	6	20
Köln
Berlin
Lyon
Paris
Köln	Dec	89	98	86	75
Berlin	Dec	10	8	11	8
Lyon	Dec	4	7	8	14
Paris	Dec	7	2	6	20

Year-to-date sum

Operaciones OLAP: Porcentajes e índices

- Para calcular porcentajes es necesario indicar el tipo de ordenación, utilizando la operación TOPPERCENT
- *"Mostrar las ventas por ciudad y categoría ordenadas por trimestre hasta que se cubra el 70% del total de ventas"*

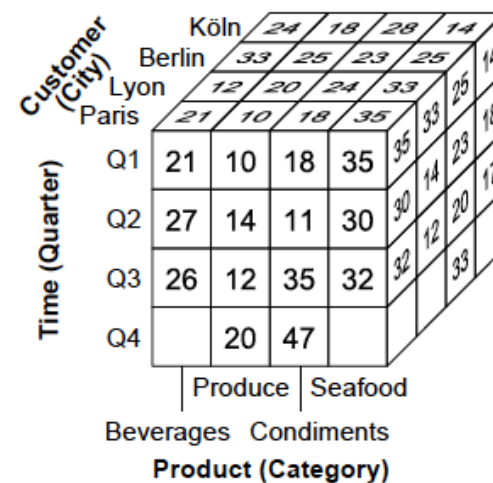
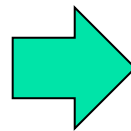
TOPPERCENT(Sales, Quantity,70) BY City, Category ORDER BY Quarter ASC



A 3D cube diagram representing sales data. The vertical axis is labeled 'Time (Quarter)' with values Q1, Q2, Q3, and Q4. The horizontal axis is labeled 'Product (Category)' with values Produce, Seafood, Beverages, and Condiments. The depth axis is labeled 'Customer (City)' with values Köln, Berlin, Lyon, and Paris. The data values are as follows:

Time (Quarter)	Customer (City)	Produce	Seafood	Beverages	Condiments
Q1	Köln	21	10	18	35
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35
Q2	Köln	27	14	11	30
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35
Q3	Köln	26	12	35	32
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35
Q4	Köln	14	20	47	31
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35

Original cube



A 3D cube diagram representing the top 70% of sales data per quarter. The vertical axis is labeled 'Time (Quarter)' with values Q1, Q2, Q3, and Q4. The horizontal axis is labeled 'Product (Category)' with values Produce, Seafood, Beverages, and Condiments. The depth axis is labeled 'Customer (City)' with values Köln, Berlin, Lyon, and Paris. The data values are as follows:

Time (Quarter)	Customer (City)	Produce	Seafood	Beverages	Condiments
Q1	Köln	21	10	18	35
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35
Q2	Köln	27	14	11	30
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35
Q3	Köln	26	12	35	32
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35
Q4	Köln		20	47	
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35

Top 70% per quarter

Operaciones OLAP: Porcentajes e índices

- Para calcular porcentajes es necesario indicar el tipo de ordenación, utilizando la operación TOPPERCENT
- *"Mostrar las ventas por ciudad y categoría ordenadas por cantidad hasta que se cubra el el 70% del total de ventas"*

TOPPERCENT(Sales, Quantity,70) BY City, Category ORDER BY Quantity DESC

Time (Quarter)	Customer (City)	Köln				14					
		Berlin					14				
		Lyon						14			
		Paris							14		
										14	
Q1	21	10	18	35	35	14					23
Q2	27	14	11	30	30	12	20				17
Q3	26	12	35	32	32	10	33	18			
Q4	14	20	47	31	31						
		Produce		Seafood							
		Beverages		Condiments							
		Product (Category)									

Original cube

Time (Quarter)	Customer (City)	Köln				25	16
		Berlin					
		Lyon					
		Paris					
Q1	21			35	35	14	17
Q2	27	14				12	16
Q3	26	12	35	32	32	33	
Q4		20	47	31	31		
		Produce		Seafood			
		Beverages		Condiments			
Product (Category)							

Operaciones OLAP: Porcentajes e índices

- La operación RANK permite generar un índice o ranking tras especificar el orden de las celdas
- *"Mostrar el ranking de trimestres (referidas a las ventas) por ciudad y categoría y ordenadas descendentemente por cantidad"*

RANK(Sales, Time) BY City, Category ORDER BY Quantity DESC

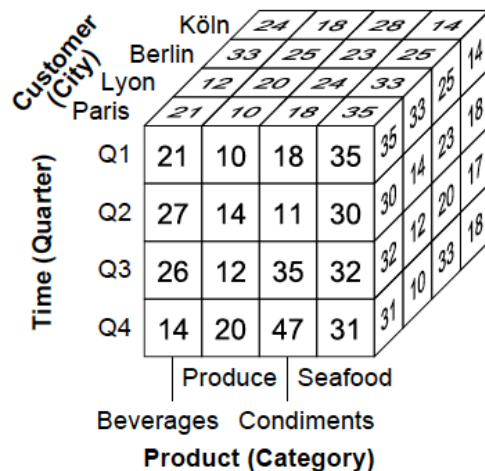


Diagram illustrating the original OLAP cube structure. The dimensions are Customer (City), Time (Quarter), and Product (Category). The data is organized into a 4x4x4 grid.

Time (Quarter)	Customer (City)	Product (Category)			
		Produce	Seafood	Beverages	Condiments
Q1	Köln	21	10	18	35
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35
Q2	Köln	27	14	11	30
	Berlin	30	14	20	17
	Lyon	32	12	33	18
	Paris	31	10	33	18
Q3	Köln	26	12	35	32
	Berlin	32	12	33	18
	Lyon	31	10	33	18
	Paris	31	10	33	18
Q4	Köln	14	20	47	31
	Berlin	31	10	33	18
	Lyon	31	10	33	18
	Paris	31	10	33	18

Original cube

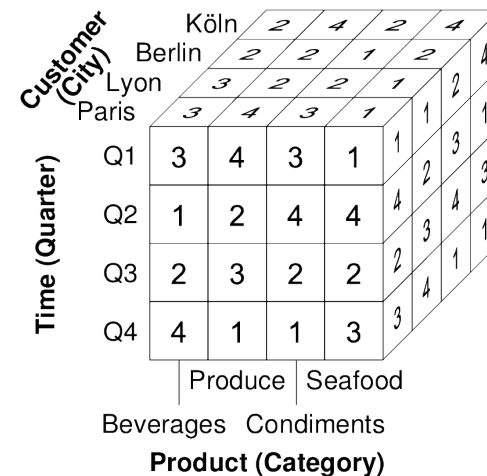
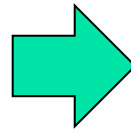


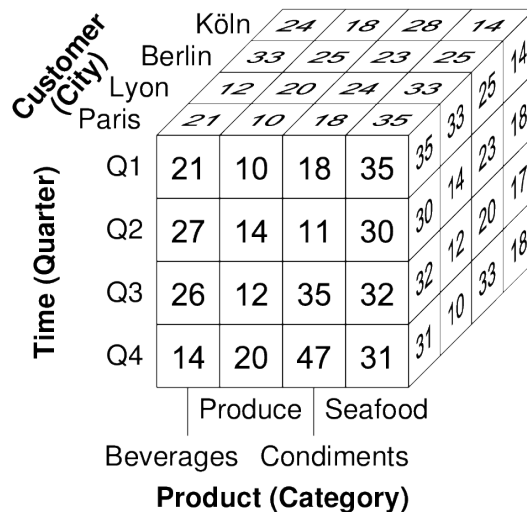
Diagram illustrating the result of the RANK operation. The dimensions are Customer (City), Time (Quarter), and Product (Category). The data is organized into a 4x4x4 grid, where the values represent the rank of the quantity for each cell.

Time (Quarter)	Customer (City)	Product (Category)			
		Produce	Seafood	Beverages	Condiments
Q1	Köln	3	4	3	1
	Berlin	2	2	1	2
	Lyon	3	2	2	1
	Paris	3	4	3	1
Q2	Köln	1	2	4	4
	Berlin	4	4	4	4
	Lyon	3	2	2	1
	Paris	3	4	3	1
Q3	Köln	2	3	2	2
	Berlin	4	4	4	4
	Lyon	3	2	2	1
	Paris	3	4	3	1
Q4	Köln	4	1	1	3
	Berlin	4	4	4	4
	Lyon	3	2	2	1
	Paris	3	4	3	1

Operaciones OLAP: Unión y Diferencia

- La operación **Union** fusiona dos cubos que tienen el mismo esquema, pero instancias inconexas.
- Ejemplo: Si *SalesSpain* es un cubo con el mismo esquema que el cubo original, pero que contiene sólo las ventas a clientes españoles, que pueden llevar a cabo:

UNION(Sales, SalesSpain)

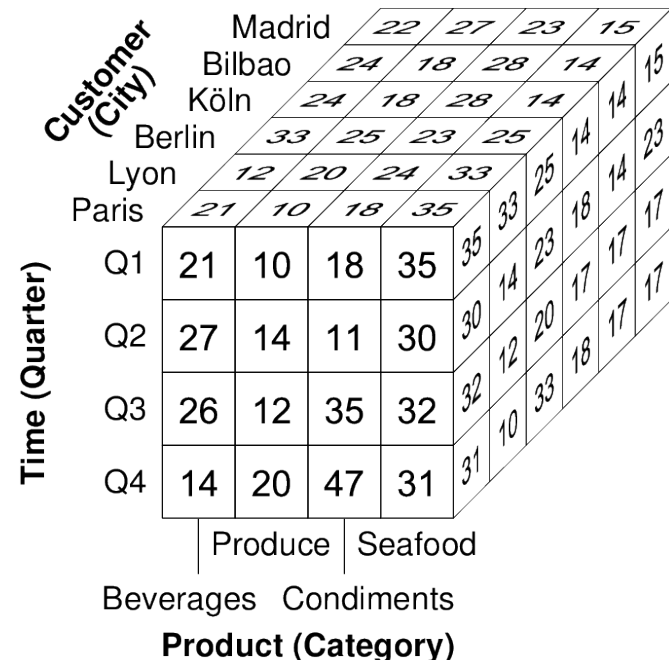


A 3D cube diagram representing sales data. The vertical axis is labeled 'Time (Quarter)' with values Q1, Q2, Q3, and Q4. The horizontal axis is labeled 'Customer (City)' with values Köln, Berlin, Lyon, and Paris. The depth axis is labeled 'Product (Category)' with values Produce, Seafood, Beverages, and Condiments. The cube contains numerical data for each combination of these dimensions.

Time (Quarter)	Customer (City)	Produce	Seafood	Beverages	Condiments
Q1	Köln	24	18	28	14
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35
Q2	Köln	27	14	11	30
	Berlin	32	12	35	32
	Lyon	14	20	47	31
	Paris	21	10	18	35
Q3	Köln	26	12	35	32
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35
Q4	Köln	14	20	47	31
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35



Cubo *SalesSpain*



A 3D cube diagram representing SalesSpain data. The vertical axis is labeled 'Time (Quarter)' with values Q1, Q2, Q3, and Q4. The horizontal axis is labeled 'Customer (City)' with values Madrid, Bilbao, Köln, Berlin, Lyon, and Paris. The depth axis is labeled 'Product (Category)' with values Produce, Seafood, Beverages, and Condiments. The cube contains numerical data for each combination of these dimensions.

Time (Quarter)	Customer (City)	Produce	Seafood	Beverages	Condiments
Q1	Madrid	22	27	23	15
	Bilbao	24	18	28	14
	Köln	24	18	28	14
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35
Q2	Madrid	27	14	11	30
	Bilbao	30	12	35	32
	Köln	32	10	33	31
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35
Q3	Madrid	26	12	35	32
	Bilbao	32	12	35	32
	Köln	32	10	33	31
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35
Q4	Madrid	14	20	47	31
	Bilbao	33	25	23	25
	Köln	32	10	33	31
	Berlin	33	25	23	25
	Lyon	12	20	24	33
	Paris	21	10	18	35

Operaciones OLAP: Unión y Diferencia

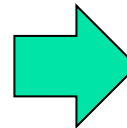
- **Difference** elimina las celdas en un cubo que pertenecen a otro; los dos cubos deben tener el mismo esquema
- Ejemplo: Dados los cubos *TopTwoSales* y la original, calcular un cubo con todas las medidas de ventas excepto las dos primeras ventas por trimestre y ciudad

DIFFERENCE(Sales, TopTwoSales)

Customer (City)	Time (Quarter)				Product (Category)			
	Q1				Q2			
	Q3				Q4			
	Q1				Q2			
	Q3				Q4			
Time (Quarter)	Q1				Q2			
	Q3				Q4			
	Q1				Q2			
	Q3				Q4			
Product (Category)	Q1				Q2			
	Q3				Q4			
	Q1				Q2			
	Q3				Q4			



Customer (City)	Time (Quarter)				Product (Category)			
	Q1				Q2			
	Q3				Q4			
	Q1				Q2			
	Q3				Q4			
Time (Quarter)	Q1				Q2			
	Q3				Q4			
	Q1				Q2			
	Q3				Q4			
Product (Category)	Q1				Q2			
	Q3				Q4			
	Q1				Q2			
	Q3				Q4			



Customer (City)	Time (Quarter)				Product (Category)			
	Q1				Q2			
	Q3				Q4			
	Q1				Q2			
	Q3				Q4			
Time (Quarter)	Q1				Q2			
	Q3				Q4			
	Q1				Q2			
	Q3				Q4			
Product (Category)	Q1				Q2			
	Q3				Q4			
	Q1				Q2			
	Q3				Q4			



Resumen Operaciones OLAP

Operator	Purpose
Add measure	Adds a new measure to a cube computed from other measures or dimensions.
Aggregation operators	Aggregates the cells of a cube, possibly after performing a grouping of cells.
Dice	Keeps the cells that satisfy a Boolean condition over dimension levels, attributes, and measures.
Difference	Removes the cells of a cube that are in another cube. Both cubes must have the same schema.
Drill-across	Merges two cubes that have the same schema and instances using a join condition.
Drill-down	Disaggregates measures along a dimension hierarchy to obtain data at a finer granularity. It is the opposite of the roll-up operation.
Drill-through	Shows data in the operational systems from which the cube was derived. This operation does not formally belong to the OLAP algebra since the result is not a cube.
Drop measure	Removes one or several measures from a cube.
Pivot	Rotates the axes of a cube to provide an alternative presentation of its data.
Recursive roll-up	Performs an iteration of roll-ups over a recursive hierarchy until the top level is reached.
Rename	Renames one or several schema elements of a cube.
Roll-up	Aggregates measures along a dimension hierarchy to obtain data at a coarser granularity. It is the opposite of the drill-down operation.
Roll-up*	Shorthand notation for a sequence of roll-up operations.
Slice	Removes a dimension by fixing a single value in a level of the dimension.
Sort	Orders the members of a dimension according to an expression.
Union	Combines the cells of two cubes that have the same schema but disjoint members.



Ejercicio I

- Un almacén de datos de un proveedor de teléfono consiste en 5 dimensiones, a saber: cliente emisor, cliente receptor, tiempo, tipo de llamada y programa de llamada. También incluye tres medidas: número de llamadas, duración y cuantía. Define las operaciones **OLAP** que hacen las siguientes consultas:
 1. Cuantía total percibida por cada programa de llamadas en 2012.
 2. La duración total de las llamadas realizadas por los clientes de Bruselas en 2012.
 3. Número total de llamadas realizadas por los clientes de fin de semana desde Bruselas a los clientes en Amberes en 2012.
 4. Duración total de las llamadas internacionales iniciadas por los clientes en Bélgica en 2012.
 5. Total recaudado por los clientes en Bruselas que están inscritos en el programa corporativo en 2012.



Ejercicio II

- Un almacén de datos de una compañía de trenes contiene información acerca de los trayectos de tren entre estaciones. Se compone de seis dimensiones, a saber: la estación de partida, la estación de llegada, el viaje, el tren, la hora de llegada y la hora de salida, y tres medidas, a saber, el número de pasajeros, la duración y el número de kilómetros. Define las operaciones **OLAP** que se lleva a cabo con el fin de responder a las consultas. **Proponer jerarquías de dimensión** cuando sea necesario.
1. Número total de kilómetros realizados por los trenes “Alstom” durante el año 2012 partiendo de estaciones de Francia o Bélgica.
 2. Duración total de viajes internacionales durante el año 2012, es decir, viajes partiendo de una estación situada en un país y llegando a una estación ubicado en otro país.
 3. Número total de viajes con origen o destino París durante Julio de 2012.
 4. Duración media de los trayectos de trenes en Bélgica en 2012.
 5. Para cada viaje, el número promedio de pasajeros por trayecto, es decir, tomar todos los trayectos de cada viaje, y el promedio del número de pasajeros.