



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 **OLTP** (sistemas de bases de datos tradicionales)
 - Datos detallados
 - **No** incluya datos **históricos**
 - Altamente normalizada
 - **Bajo rendimiento** en **consultas complejas**
 - Basados en **transacciones** (operaciones del día a día)
- Sistemas **OLAP** (almacenes de datos)
 - Centrados en **consultas analíticas**
 - **Gran rendimiento** en **consultas complejas**
 - **Respuesta a preguntas de negocio**

consulta típica OLAP: cantidad total de ventas por producto

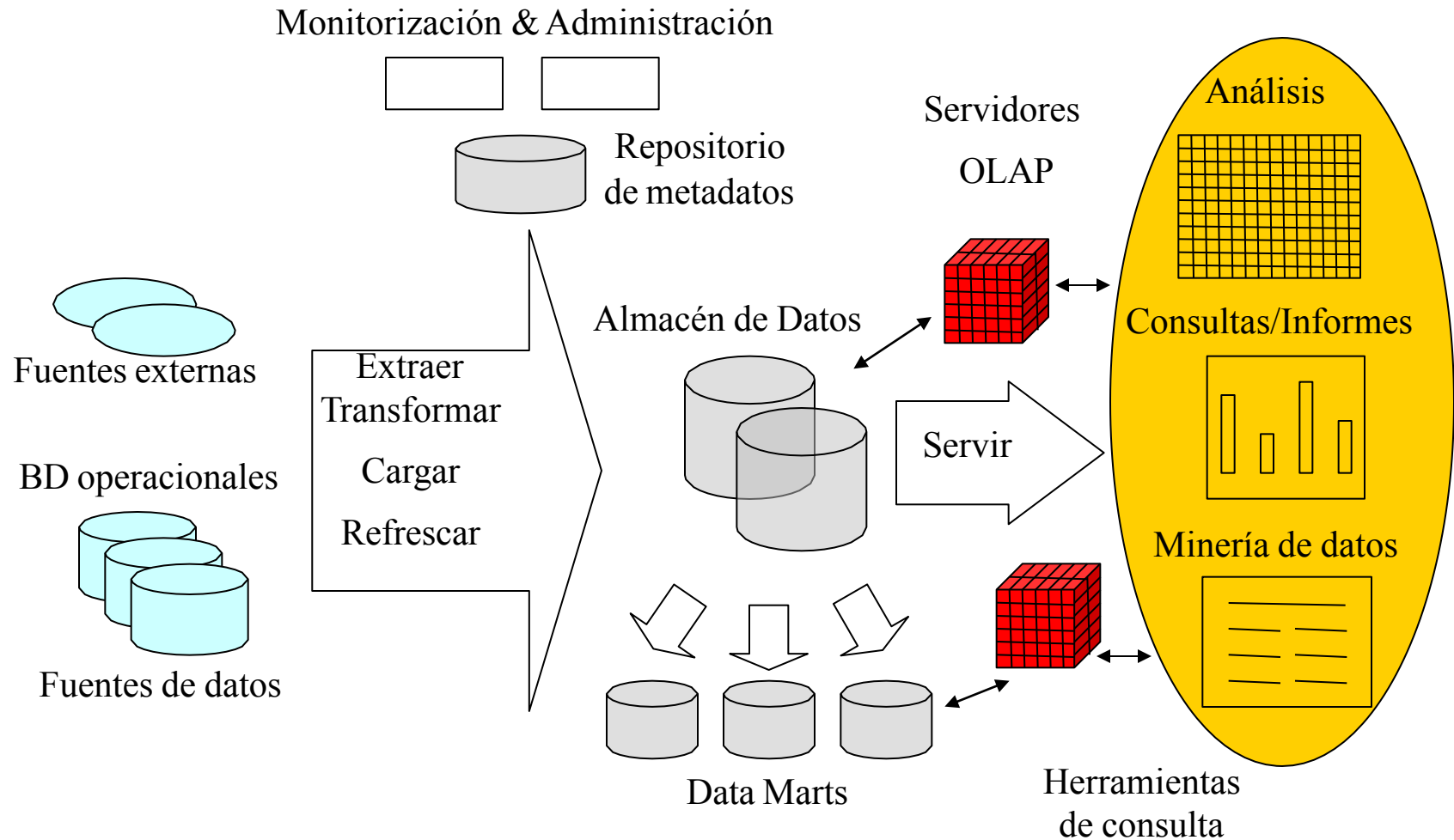
consulta típica OLTP: los pedidos pendientes para el cliente c1



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

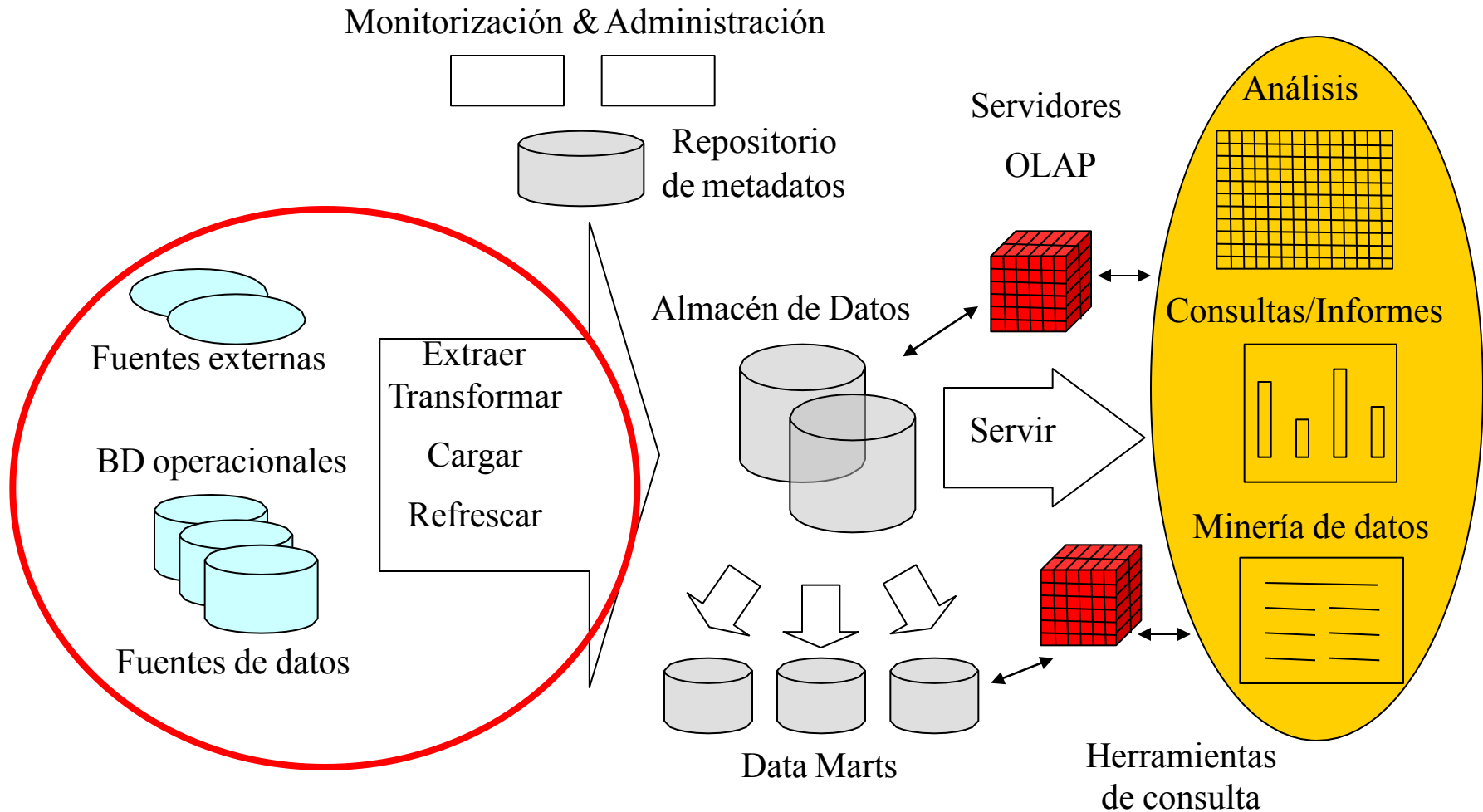


Esquema de una arquitectura de DW

- ¿ Por qué Data Warehouse separado ?
 - Rendimiento (Performance)
 - BD operacionales enfocadas y optimizadas a transacciones
 - Consultas OLAP complejas → Ralentización del servidor
 - Funcionalidad
 - Datos no existentes → históricos
 - Datos de distintas fuentes
 - Enfocados a consultas complejas (negocio)

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
 - 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 y transformación
 - Fundamental que los datos del almacén sean correctos
 - Decisiones estratégicas
 - Muchas fuentes de datos → alta probabilidad de error y anomalías
 - Longitud de campos inconsistentes
 - Valores incoherentes
 - Universidad de Alicante
 - Univ. Alicante



Esquema de una arquitectura de DW

Fuentes de datos: procesos y herramientas

- Herramientas

- 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



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
 - Ordenar datos
 - Revisar la integridad
 - Cantidad de datos a cargar
 - Calcular tiempos de carga



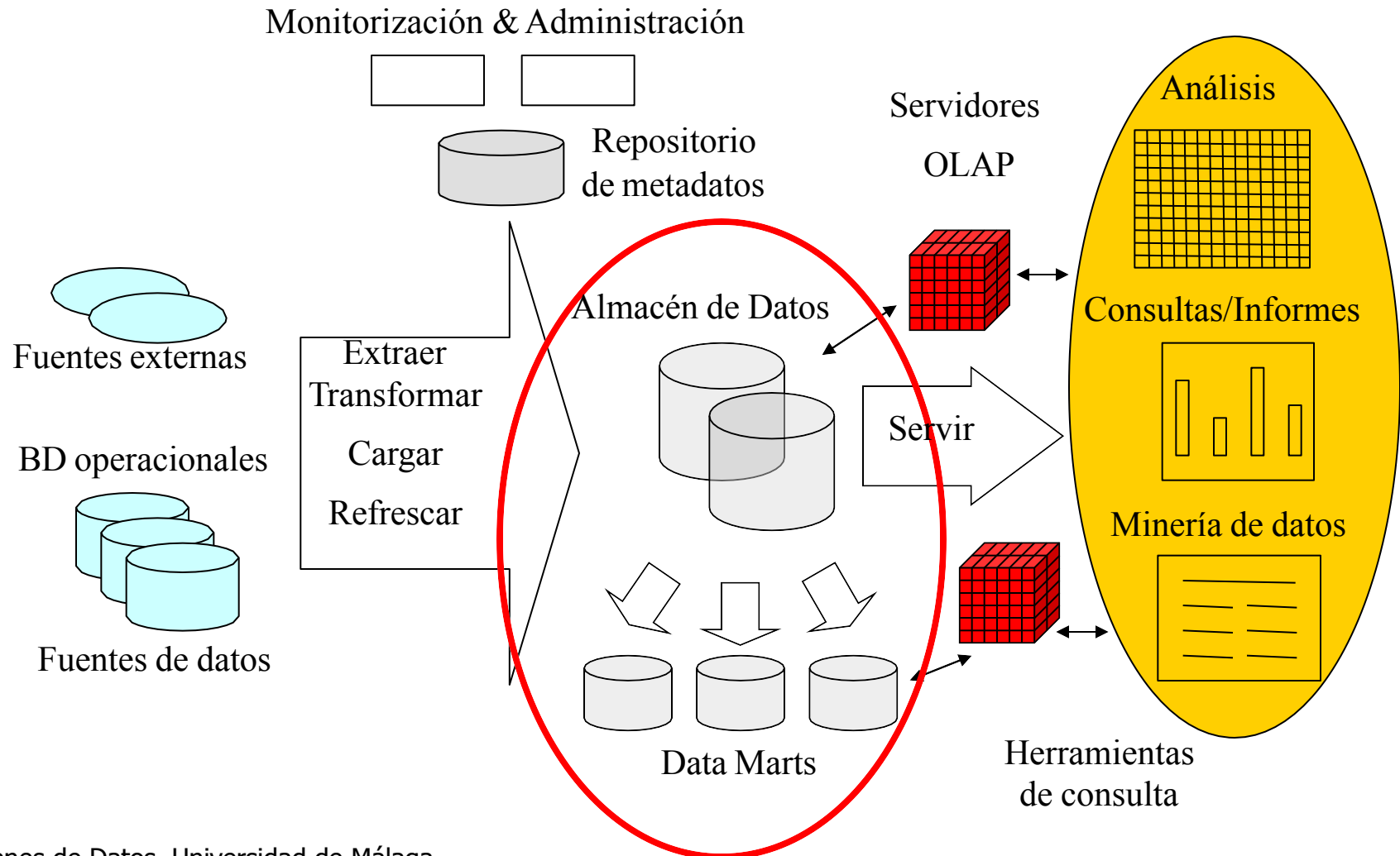
Esquema de una arquitectura de DW

Fuentes de datos: procesos y herramientas

- Técnicas de carga y refresco
 - Carga total
 - Primera vez
 - Fuentes de datos -> DW
 - Carga incremental (refresco)
 - Sólo datos a partir de la última fecha de carga
 - Cada actualización (bolsa, ...)

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 equipo de futbol en varias tablas (Estadio, ciudad, ...)
 - ¿ 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 (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
 - Para ello: procesos de limpieza y transformación
 - Hay errores difíciles de detectar: iiii Cuidado !!!!!
 - Datos correctamente integrados → 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, se añaden nuevos datos para el análisis
 - Una venta de un producto con una fecha en una provincia (200 uds)
 - Una venta de un producto con otra fecha en una provincia (300 uds)



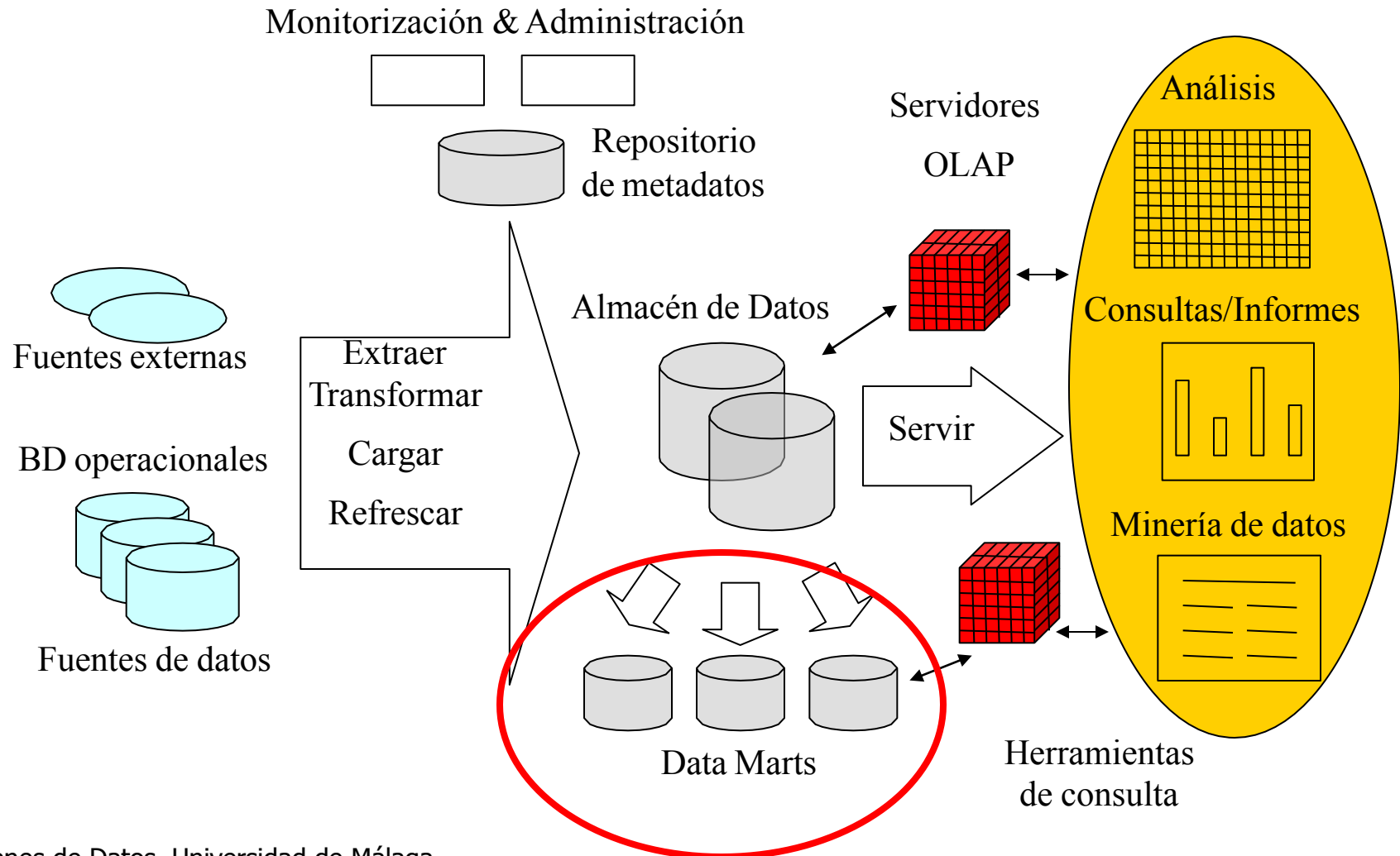
Esquema de una arquitectura de DW

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 ?
 - 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



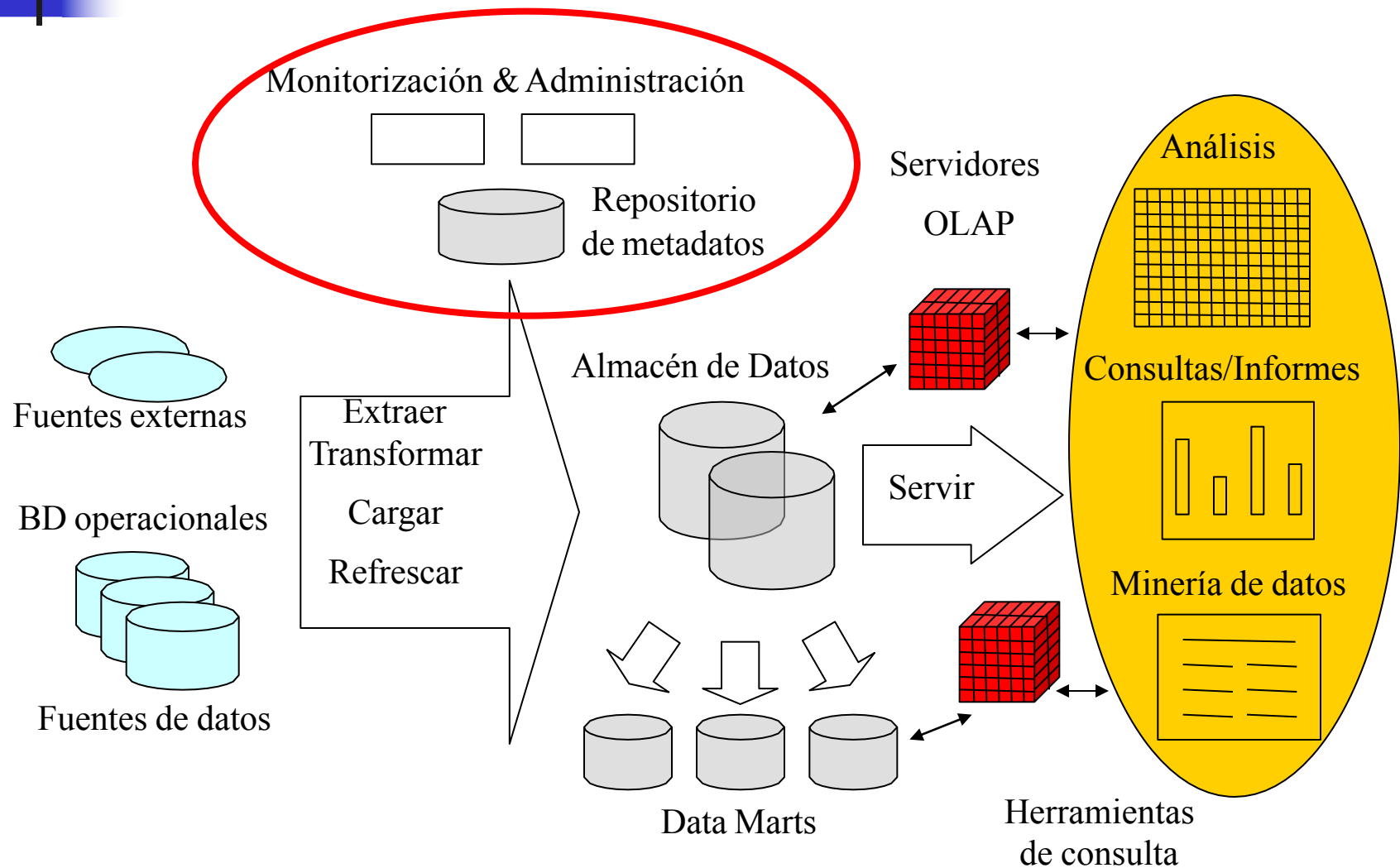
Esquema de una arquitectura de DW

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

Los metadatos





Esquema de una arquitectura de DW

Los metadatos

- Los metadatos
 - Son datos sobre datos
 - Qué dato se guarda (ej. Clientes, ventas)
 - Dónde se guarda (tabla clientes, ventas)
 - Campos de la tabla
 - Con qué datos de las fuentes se corresponden
 - 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) ...
 - Información de los procesos ETL
 - Seguridad → autorización, control de acceso, etc.



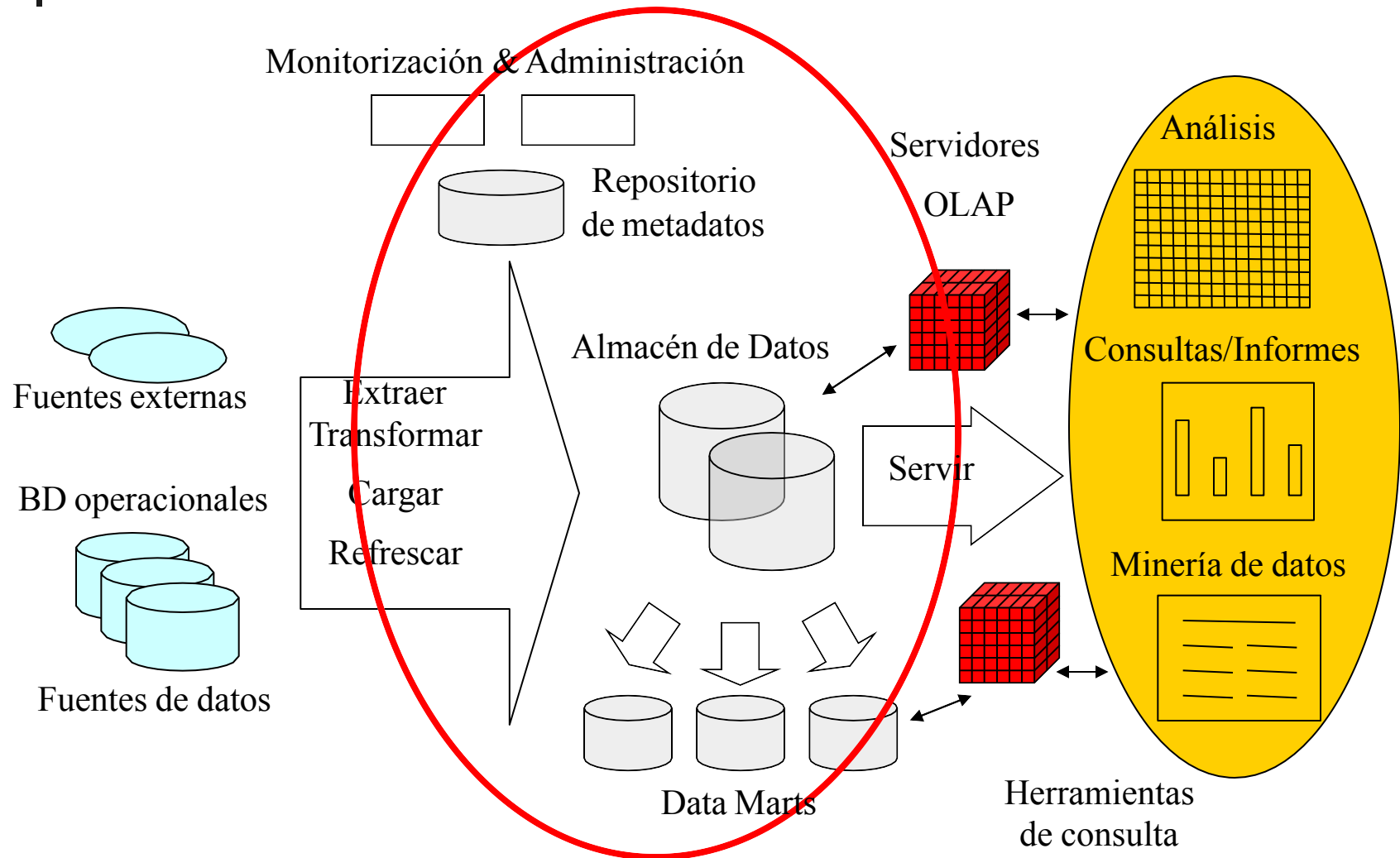
Esquema de una arquitectura de DW

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

Esquema de una arquitectura de DW

Servidor del almacén de datos





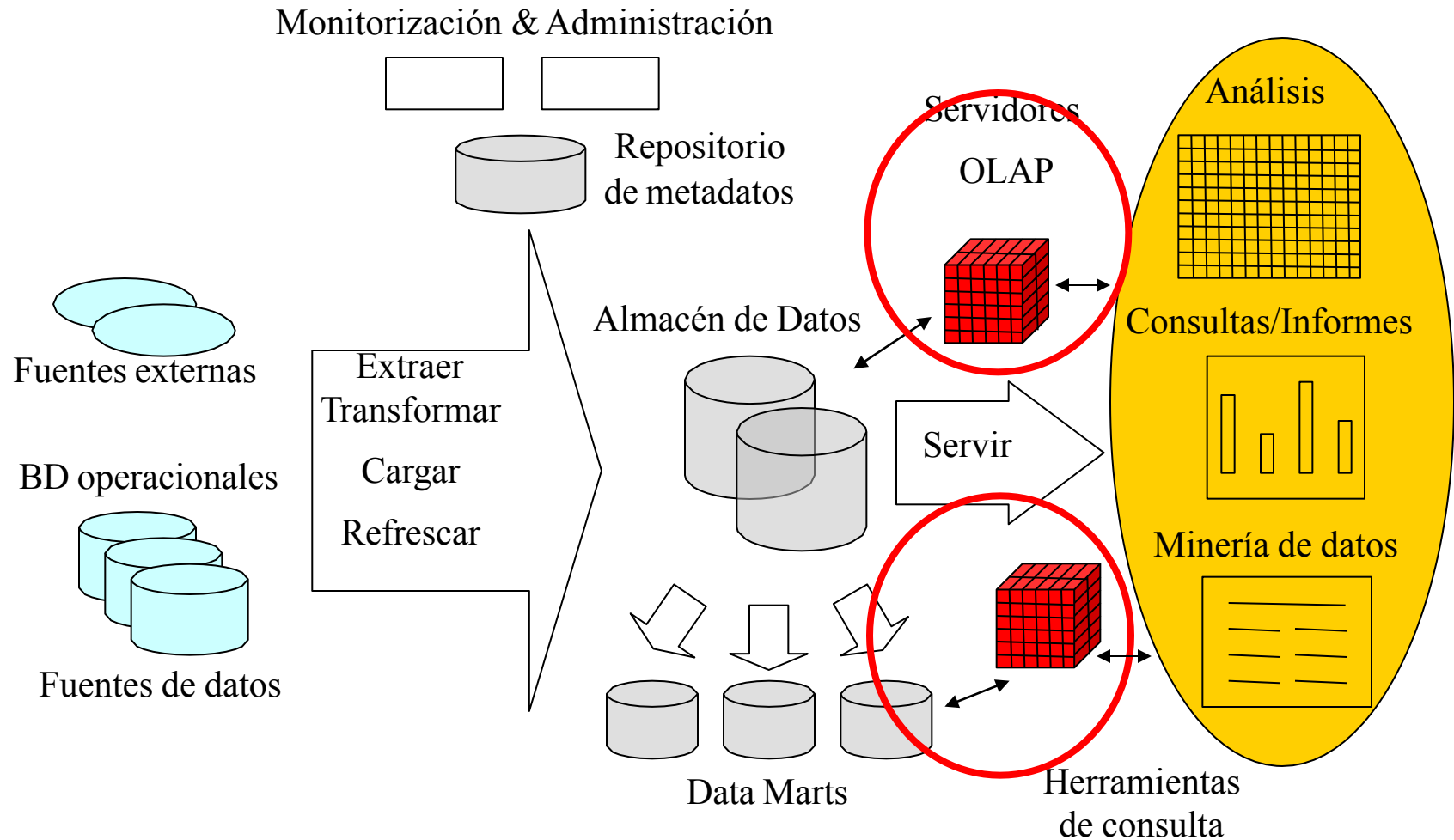
Esquema de una arquitectura de DW

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
 - 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)



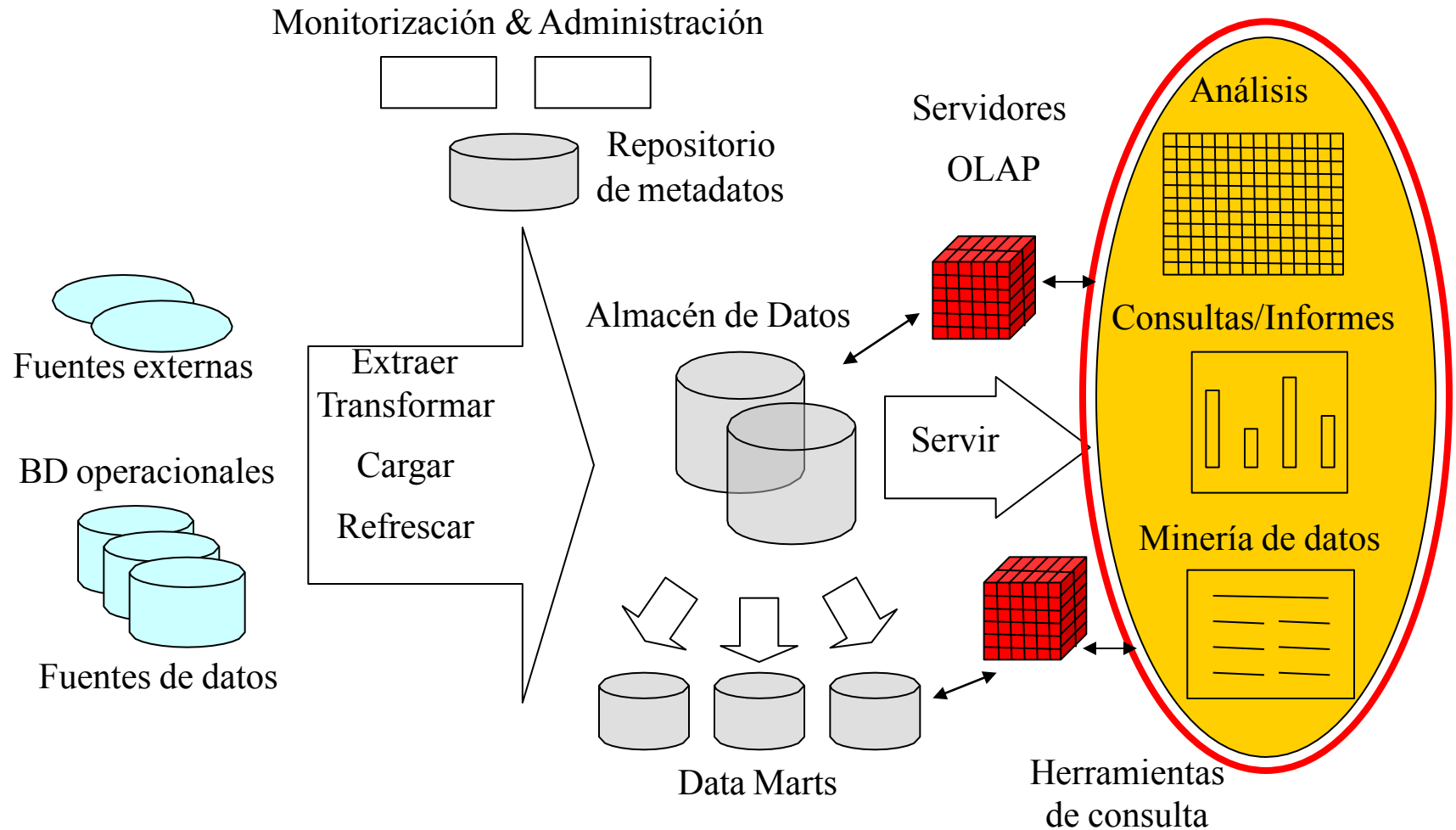
Esquema de una arquitectura de DW

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)

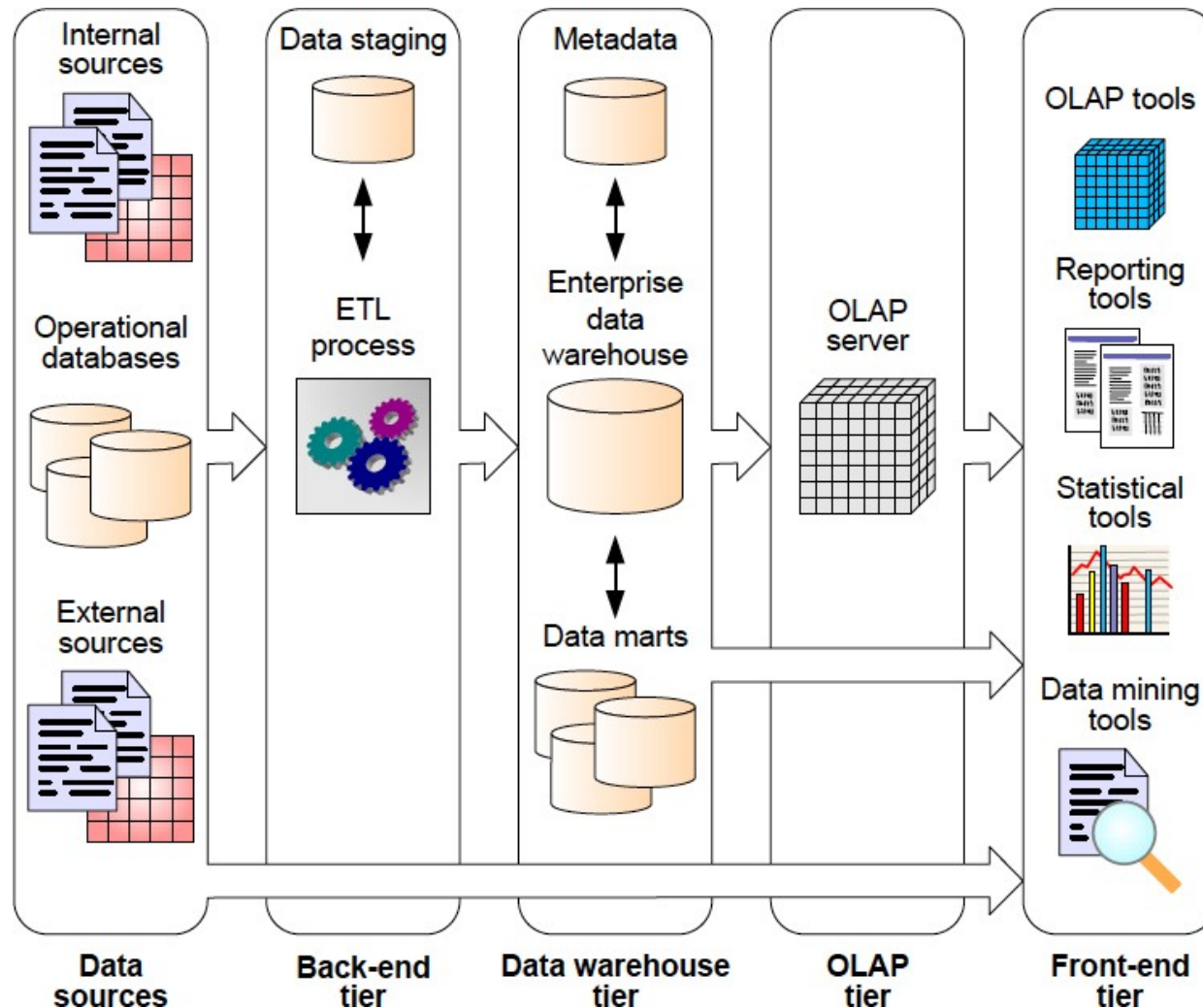
Esquema de una arquitectura de DW

Herramientas de consulta



Esquema de una arquitectura de DW

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



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



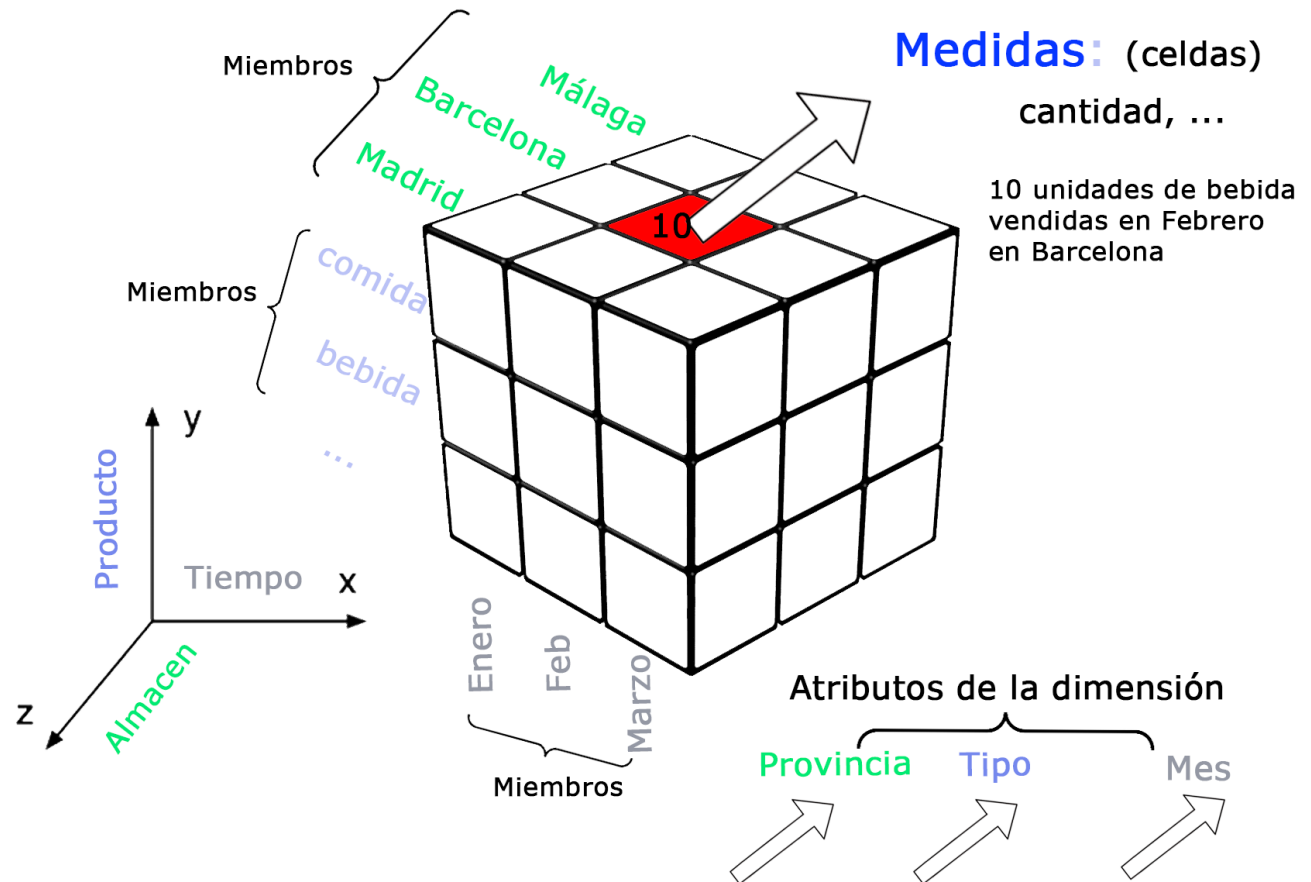
Indice

- 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

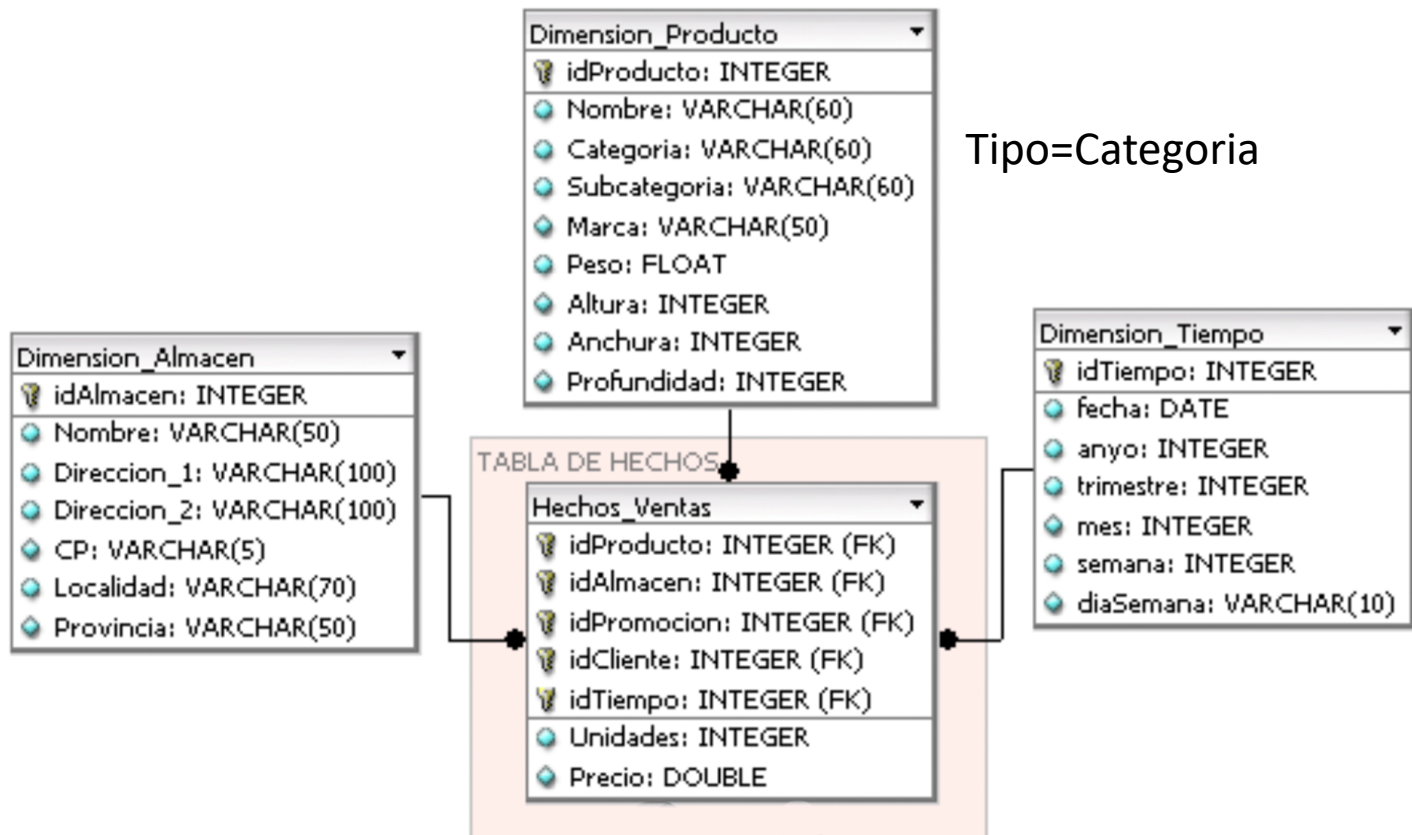
Hecho: Ventas



Dimensiones: Almacen, producto, Tiempo

Modelo Multidimensional

- Tablas de hechos y dimensiones





Modelo Multidimensional

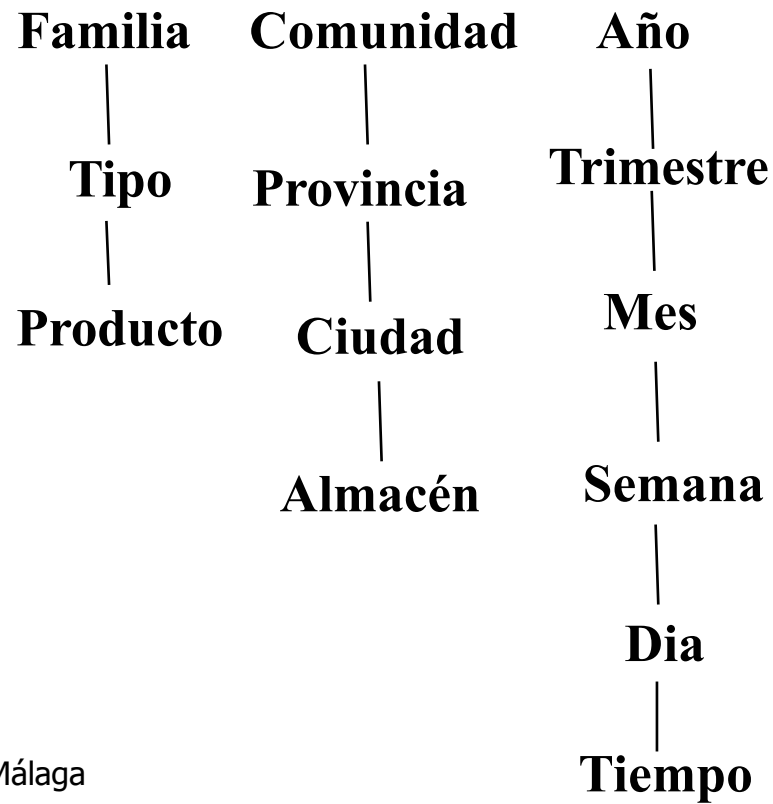
- Granularidad. Nivel de detalle al que se representan las medidas para cada dimensión del cubo.
 - Por ejemplo: Cifras de Ventas agregadas por Tipo, mes y provincia.
- Las instancias de una dimensión se denominan Miembros.
 - Por ejemplo: Comida, Bebida son miembros de Producto en la granularidad Tipo
- 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

Esquema de las dimensiones producto, almacen, y tiempo





Medidas (atributos del hecho)

- Atómicos
 - Ej. Cantidad vendida, precio, etc.
- Derivados
 - Utilizan una fórmula para calcularlos
 - Ej. $\text{Precio_total} = \text{precio} * \text{cantidad_vendida}$



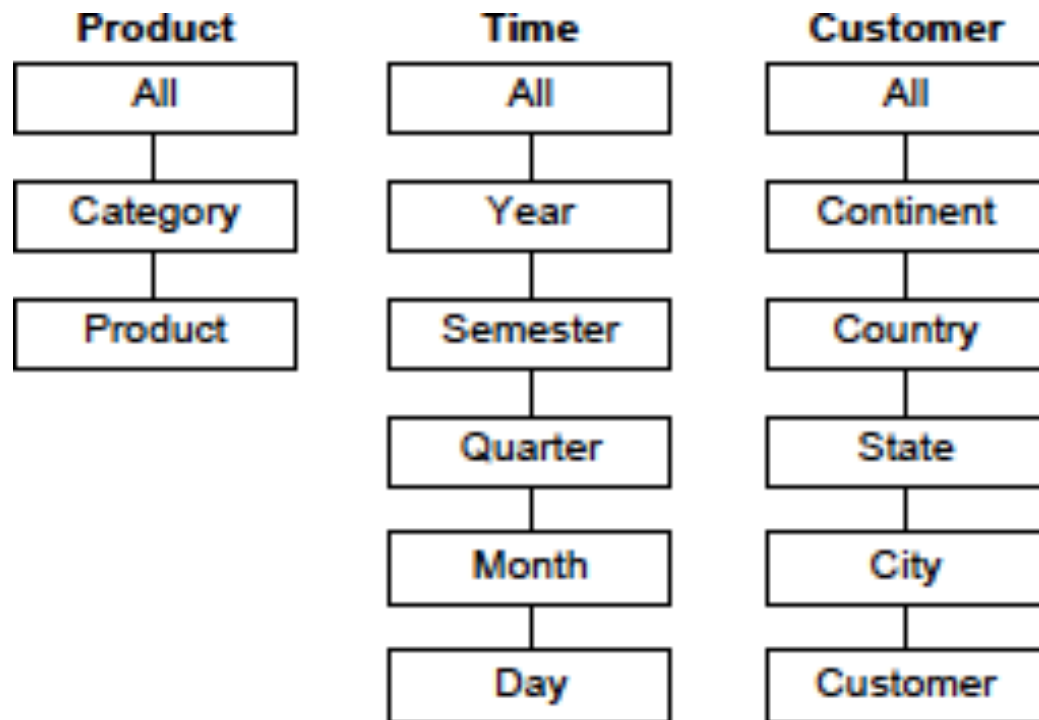
Indice

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



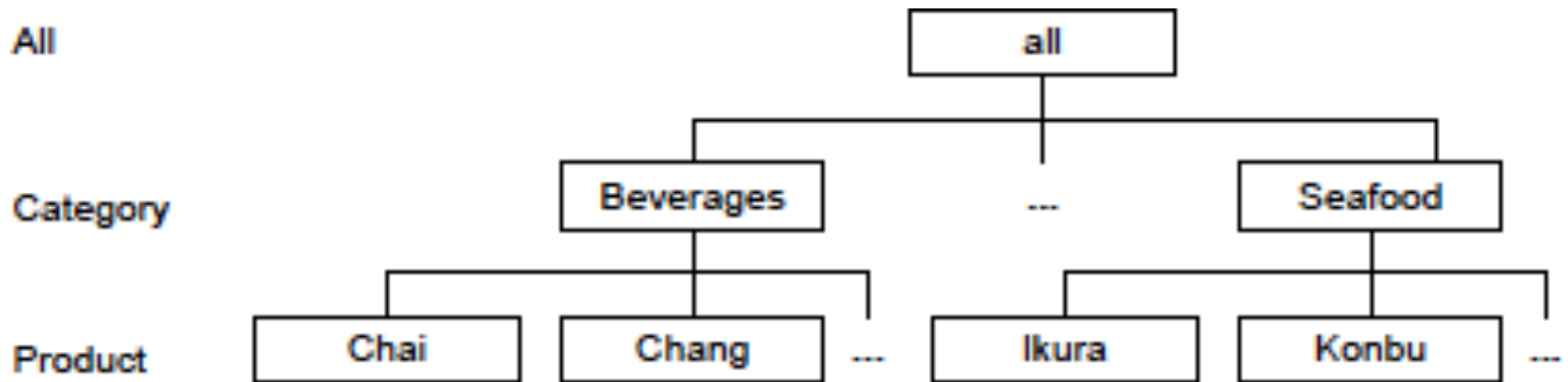
Sobre jerarquías (I)

- Jerarquías de las dimensiones Product, Time and Customer



Sobre jerarquías (II)

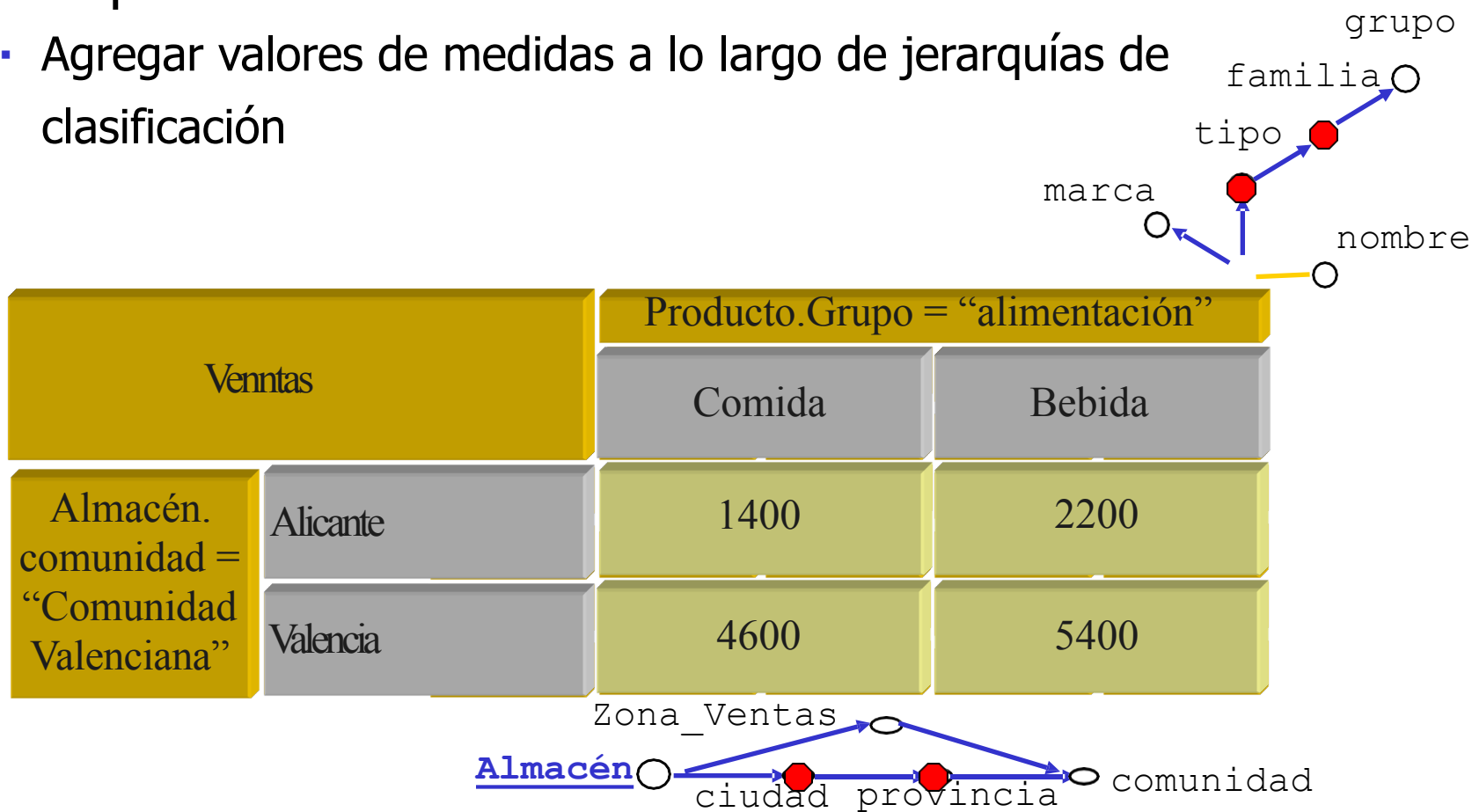
- Miembros de una jerarquía Producto -> Categoría



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))

Customer (City)

	Köln	24	18	28	14
Berlin	33	25	23	25	14
Lyon	12	20	24	33	25
Paris	21	10	18	35	33
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)				

Time (Quarter)

Original cube

Customer (Country)

	Germany	57	43	51	39
France	33	30	42	68	39
Q1	33	30	42	68	68
Q2	39	26	41	44	44
Q3	30	22	46	44	44
Q4	25	29	49	41	41
	Produce	Seafood	Beverages	Condiments	
	Product (Category)				

Time (Quarter)

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"

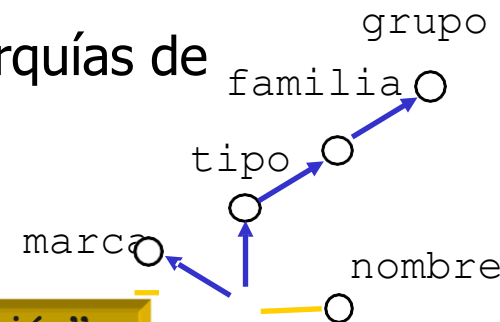
¿Qué hace?

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

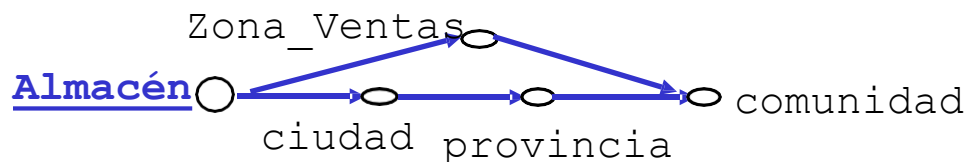
Operaciones OLAP

- Drill-down

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



Ventas			Producto.Grupo = "alimentación"			
			Comida		Bebida	
			Cong.	Fresca	Refresco	Alcohol
Almacén. Comunidad= "Comunidad Valenciana"	Alicante	Albatera	100	200	300	400
		Elche	500	600	700	800
	Valencia	Burjasot	900	1000	1100	1200
		Cullera	1300	1400	1500	1600



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.)

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'

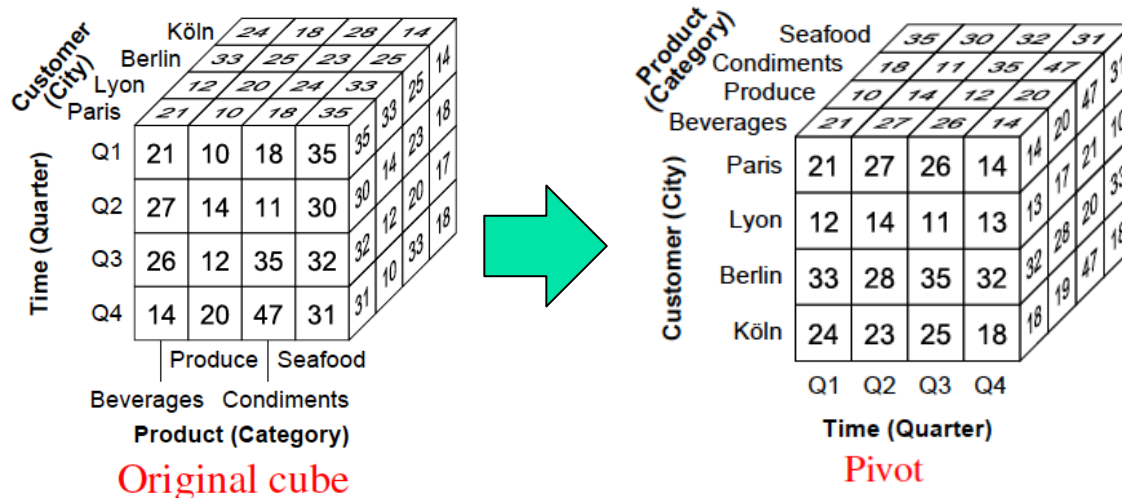
Time (Quarter)	Customer (City)	Product (Category)			
		Produce	Seafood	Beverages	Condiments
	Lyon	12	20	24	33
Q1	Paris	21	10	18	35
	Q2	27	14	11	30

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

Operaciones OLAP

■ Pivoting

- 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)



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])

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

Original cube

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

Cube for 2011

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

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)			
		Beverages	Condiments	Produce	Seafood
Q1	Köln	19	12	31	28
	Berlin	21	10	18	35
	Lyon	30	27	14	10
	Paris	28	26	11	31
Q2	Köln	28	26	11	31
	Berlin	32	30	14	12
	Lyon	29	31	10	20
	Paris	31	29	12	10
Q3	Köln	29	31	10	20
	Berlin	35	28	16	14
	Lyon	25	23	21	26
	Paris	25	23	21	26
Q4	Köln	25	23	21	26
	Berlin	18	14	10	4
	Lyon	18	14	10	4
	Paris	18	14	10	4

Percentage change

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

`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 los 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

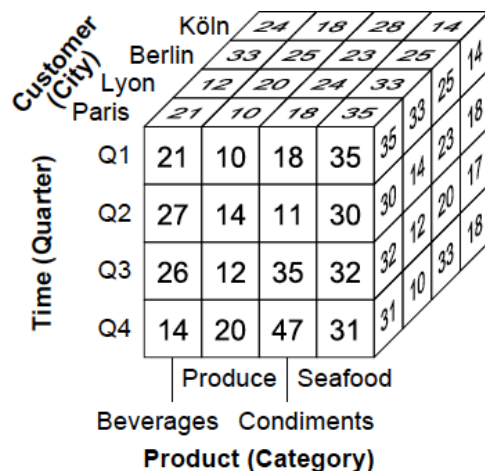
Customer (City)	Time (Quarter)	Product (Category)			
		Produce	Beverages	Condiments	Seafood
Köln	Jan	7	2	6	20
Berlin	Jan	10	8	11	8
Lyon	Jan	4	7	8	14
Paris	Jan	7	2	6	20
Köln	Feb	15	6	14	26
Berlin	Feb	10	8	11	8
Lyon	Feb	4	7	8	14
Paris	Feb	7	2	6	20
Köln	Mar	21	10	18	33
Berlin	Mar	10	8	11	8
Lyon	Mar	4	7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

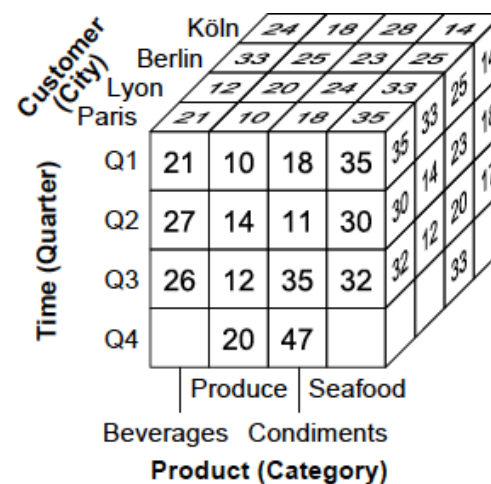
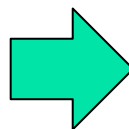
- Para calcular porcentajes es necesario indicar el tipo de ordenación, utilizando la operación TOPPERCENT
- “Mostrar las ventas por ciudad y categoría ordenadas por trimestre hasta que se cubra el 70% del total de ventas”

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



Original cube

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



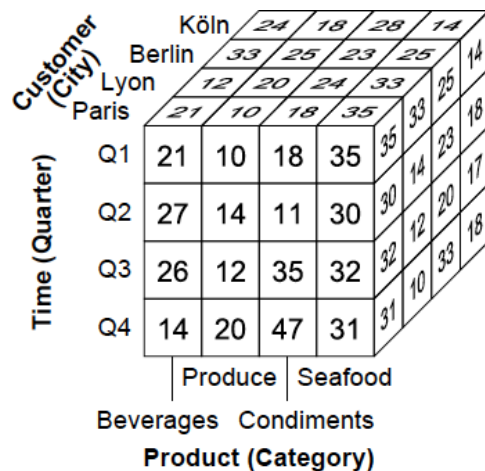
Top 70% per quarter

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

Operaciones OLAP: Porcentajes e índices

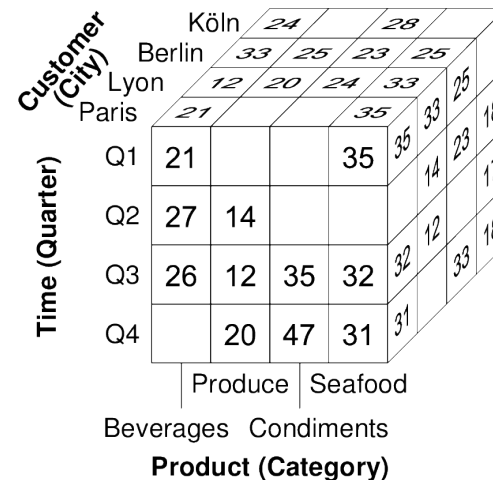
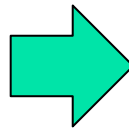
- Para calcular porcentajes es necesario indicar el tipo de ordenación, utilizando la operación TOPPERCENT
- "Mostrar las ventas por ciudad y categoría ordenadas por cantidad hasta que se cubra el 70% del total de ventas"

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



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	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



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

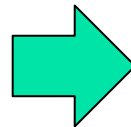
Operaciones OLAP: Porcentajes e índices

- 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

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

Original cube



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

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)

Customer (City)		Köln				Berlin				Lyon				Paris			
		24		18		28		14		33		25		14			
		33		25		23		25		33		25		18			
		12		20		24		33		35		33		18			
		21		10		18		35		35		33		25			
Time (Quarter)	Q1	21	10	18	35	35	14	23	17	18	14	25	18	14			
	Q2	27	14	11	30	30	14	20	17	18	14	20	18	14			
	Q3	26	12	35	32	32	10	33	18	14	20	18	14	14			
	Q4	14	20	47	31	31	10	33	18	14	20	18	14	14			
			Produce				Seafood				Beverages				Condiments		
		Product (Category)															



Cubo SalesSpain

		Customer (City)							
		Madrid	Bilbao	Köln	Berlin	Lyon	Paris		
Time (Quarter)	Q1	21	10	18	35	35	14	23	17
	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)							

Operaciones OLAP: Unión y Diferencia

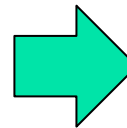
- **Difference** elimina las celdas en un cubo que pertenecen a otro; los dos cubos deben tener el mismo esquema
- Ejemplo: Dados los cubos TopTwoSales y la original, calcular un cubo con todas las medidas de ventas excepto las dos primeras ventas por trimestre y ciudad

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)	Produce				Seafood			
	Beverages				Condiments			
	Produce				Seafood			
	Beverages				Condiments			



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)	Produce				Seafood			
	Beverages				Condiments			
	Produce				Seafood			
	Beverages				Condiments			



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)	Produce				Seafood			
	Beverages				Condiments			
	Produce				Seafood			
	Beverages				Condiments			



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.