

UNIVERSIDAD NACIONAL DE LA MATANZA

INTELIGENCIA DE NEGOCIOS

Modelo Dimensional – Parte 2

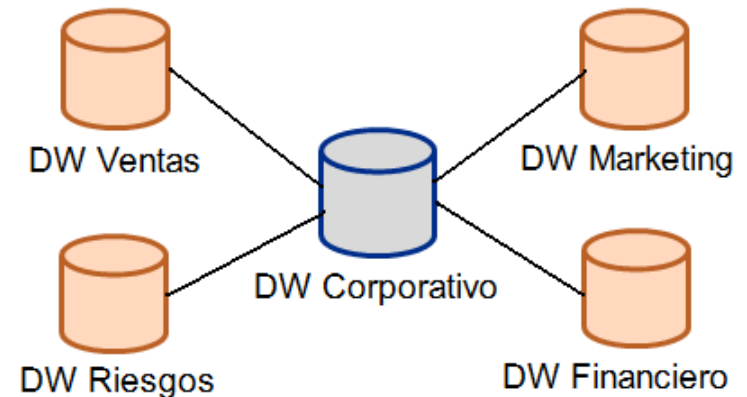
Profesor: Mg. Diego Basso

Curso 2017

DATA MART



- Subconjunto de los datos del DW con el objetivo de responder a un determinado análisis, función o necesidad, con un conjunto específico de usuarios.
- Cubre las necesidades de un grupo de trabajo o de una determinada área en una organización.
- El alcance es lo que diferencia un Data Mart de un Data Warehouse.
 - Independiente: Se genera a través de un proceso de ETL
 - Dependiente : Se toma del Data Warehouse



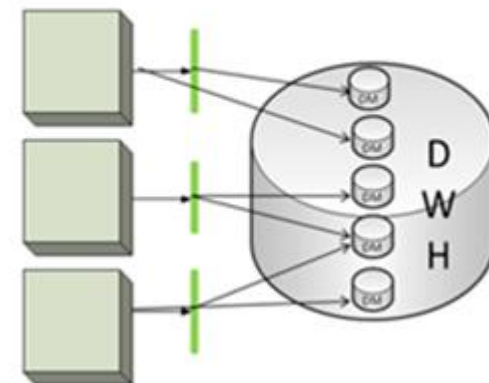
PARADIGMAS DE CONSTRUCCIÓN



○ Ralph Kimball

- Construir los Data Mart de acuerdo a las prioridades del negocio. El DW es la unión de todos los Data Mart de la organización.
- Ventajas
 - ❑ Soluciones más pequeñas y más fáciles de construir
 - ❑ Soluciones más fáciles de entender
- Desventajas
 - ❑ El mantenimiento es más complejo
 - ❑ Pueden existir problemas de sincronización

Bottom Up Ralph Kimball

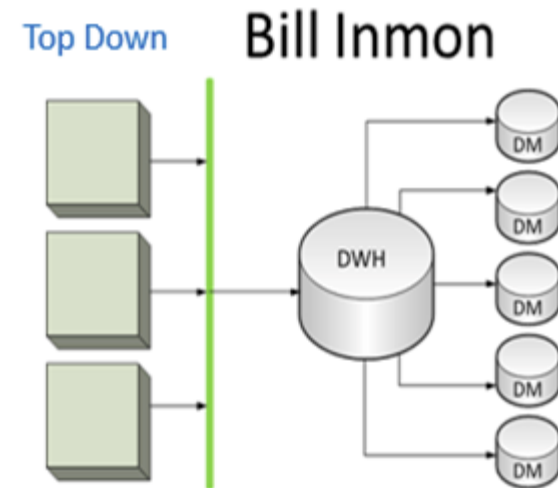


PARADIGMAS DE CONSTRUCCIÓN



○ Bill Inmon

- Construir el DW centralizado y luego extraer la información de negocio para cada Data Mart.
- Ventaja
 - ❑ Estructura del datos del DW en 3FN, normalizada y limpia.
- Desventajas
 - ❑ Mayor tiempo de construcción
 - ❑ Costo muy alto



DIMENSIÓN CONFORMADA



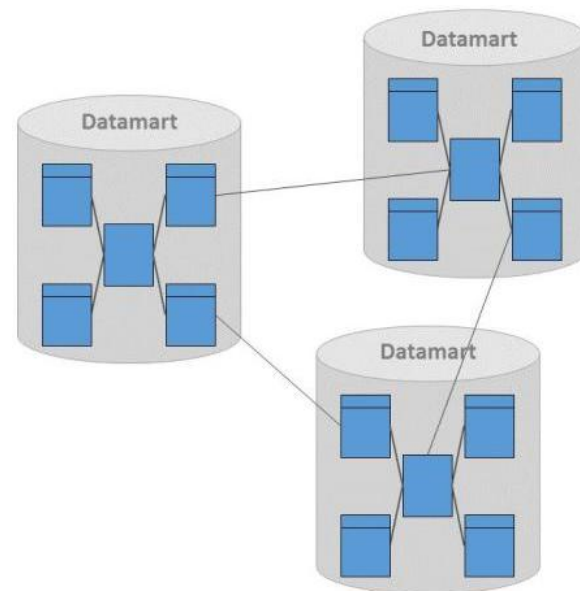
- Es una dimensión que tiene el mismo significado para todos los Data Mart que se armen.
 - Ejemplo: La **dimensión cliente** podría ser compartida por varios Data Mart (ventas, compras, pedidos, cobranzas, etc.)
- Ventajas
 - Evita duplicar dimensiones para distintos Data Mart.
 - Menor costo de mantenimiento. Los datos de la dimensión están organizados de la misma manera en todos los cubos.
- Desventajas
 - Deben emplearse del mismo modo en todos los cubos que usan estas dimensiones.
 - Un cambio implica modificar la dimensión en todos los cubos.



DIMENSIÓN CONFORMADA



- Es responsabilidad del equipo que diseña el DW establecer, publicar y mantener las dimensiones conformadas.
- Cualquier modificación o cambio en una dimensión conformada debe ser válida para todos los cubos que la empleen.
 - El DW debe funcionar como un todo integrado.



MÉTODO DE LA MATRIZ (BUS MATRIX)



- Identificar todos los Data Mart y dimensiones.
- Armar una matriz e interceptar los **hechos** (fact) con las **dimensiones** que correspondan.
- Las dimensiones que se relacionen con varios hechos probablemente sean conformadas.

	Tiempo	Llamada	Documento	Geografía	Centro de Distribución
Reclamos	X	X		X	
Facturación	X		X	X	X
....					
Stock				X	X

- El nivel mínimo de granularidad de cada dimensión conformada debe ser el mismo.



SISTEMAS OLAP



- On Line Analytical Process
- Es una **tecnología** para el acceso y análisis de datos en línea, en el que se usan **herramientas analíticas**, que facilitan el **análisis de la información** del negocio.
- Es el proceso de almacenar y administrar datos sobre la base de las variables del negocio (dimensiones), para permitir a los profesionales de negocios visualizarlos y analizarlos para entender cuál es su significado.



HERRAMIENTAS DE EXPLOTACIÓN DW



- Las herramientas OLAP presentan al usuario una visión multidimensional de los datos.



HERRAMIENTAS DE EXPLOTACIÓN DW



- El usuario formula **consultas** a la herramienta OLAP seleccionando **atributos** del modelo multidimensional, sin conocer la estructura interna del DW.
- La herramienta OLAP genera la correspondiente consulta y la envía al gestor de consultas del sistema.
 - Mediante una sentencia **SELECT**



HERRAMIENTAS DE EXPLOTACIÓN DW



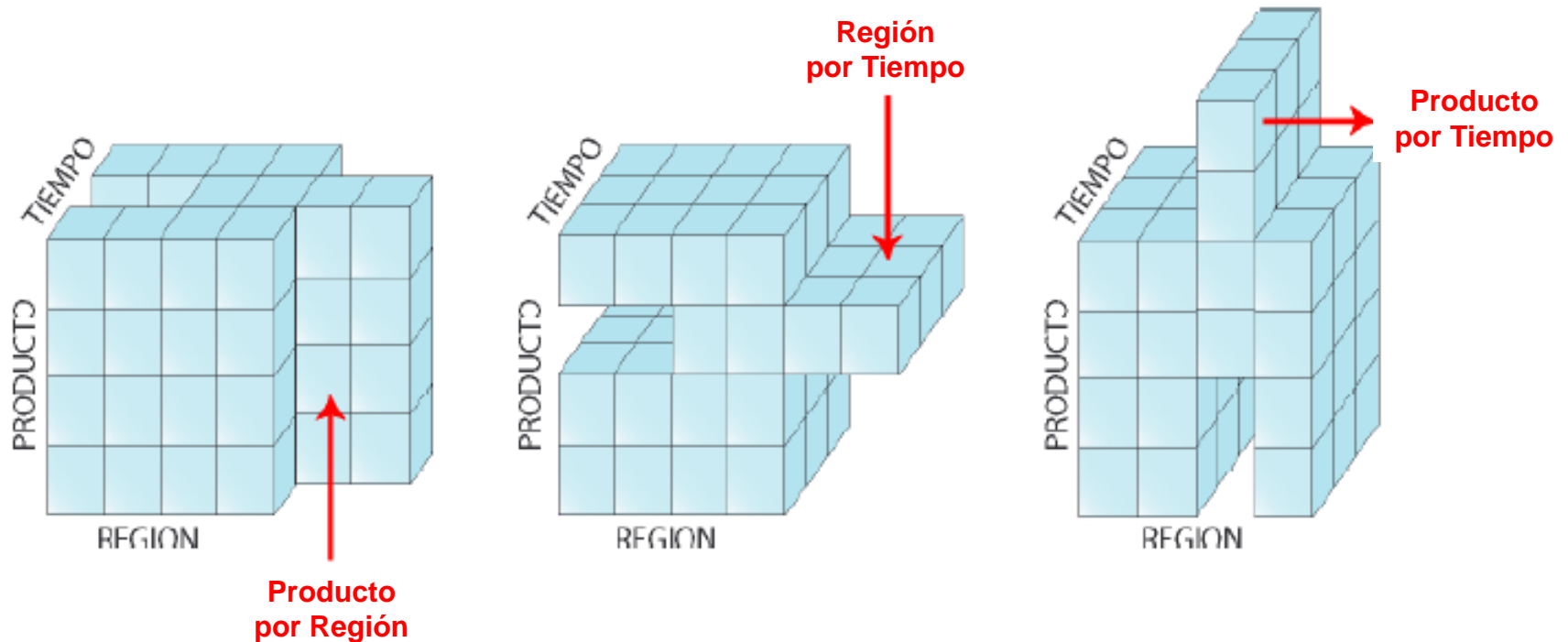
- Una consulta a un DW consiste en obtener **medidas** sobre los hechos, parametrizados por **atributos** de las **dimensiones** y restringidas por **condiciones** aplicadas sobre las dimensiones.
 - Importe total de **ventas** durante el **año 2015** de los productos del **área de limpieza**, por **trimestre** y **categoría**.
 - *Medida*: Importe
 - *Restricciones*: Productos del área de limpieza, ventas durante el año 2015.
 - *Parámetros de la consulta*: Por categoría de producto y por trimestre.



HERRAMIENTAS DE EXPLOTACIÓN DW



- Se pueden tomar rebanadas del cubo para responder a una diversidad de preguntas.

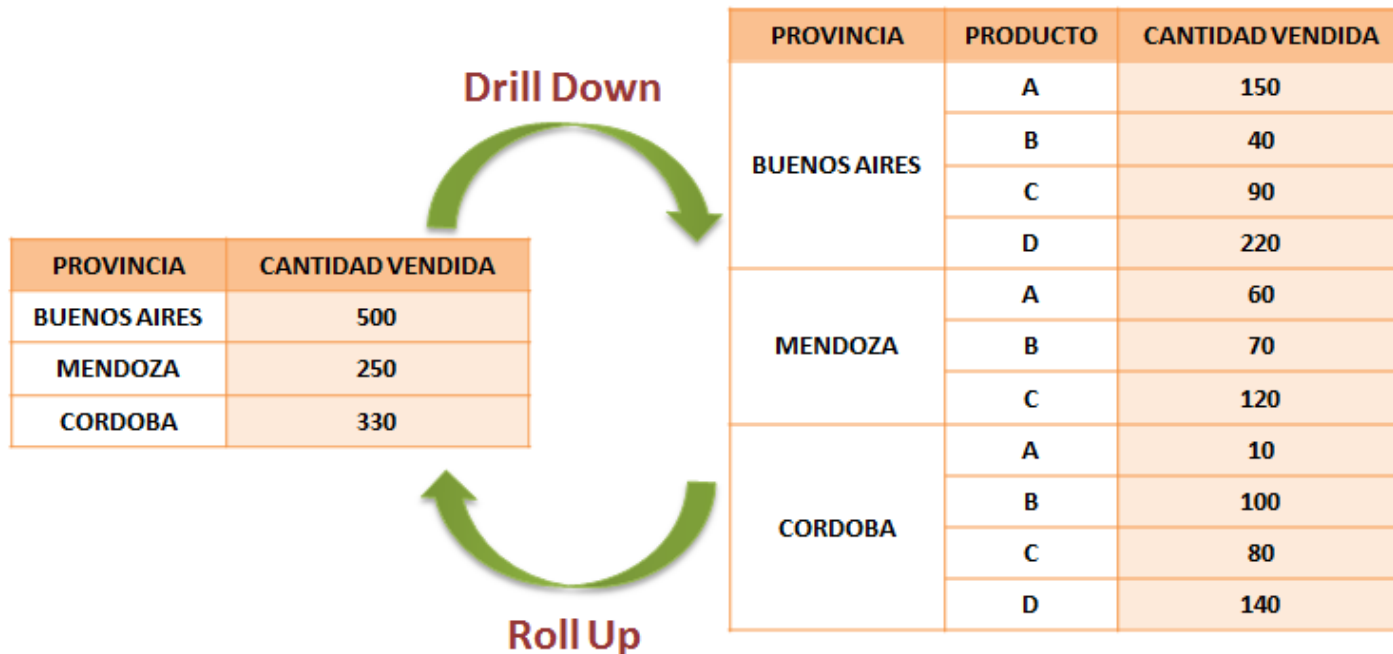




OPERACIONES OLAP

Relacionadas con la granularidad

- **Drill down (roll down):** Mayor nivel de detalle
 - De datos sumados (o agregados) a datos desagregados
- **Roll up (Drill up):** Menor nivel de detalle
 - Mayor consolidación de datos



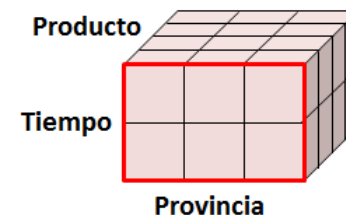
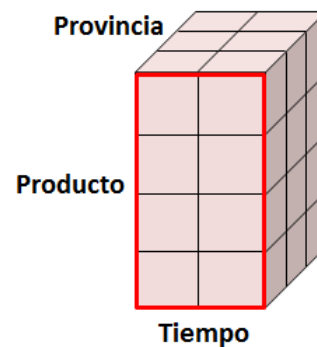
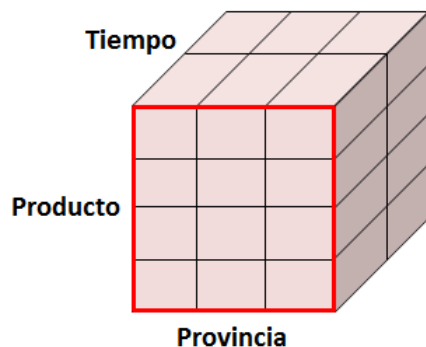


OPERACIONES OLAP

Navegación por las dimensiones

○ Slice and dice

- Selección de valores específicos para una o varias dimensiones, cortes o subconjuntos de las vistas.
- Cambio en la orientación y forma de visualización del cubo.
- Se puede agrupar la información de una manera, analizarla, agruparla de otra y realizar otro análisis.
 - Provincia = Mendoza (filtro estático)
 - Los 2 productos más vendidos del año 2015 (filtro dinámico)
- El usuario posee diferentes perspectivas para el análisis.





OPERACIONES OLAP

- Slice and dice

VENTAS

Todos los Clientes

Año	Argentina		Brasil		Total
	Alim.	Bebidas	Alim.	Bebidas	
2006	2400	2200	11000	5000	20600
2007	4500	3200	12000	6000	25700
2008	5600	2900	10000	5500	24000
Total	12500	8300	33000	16500	70300

Slice

VENTAS

Cientes Grupo 1

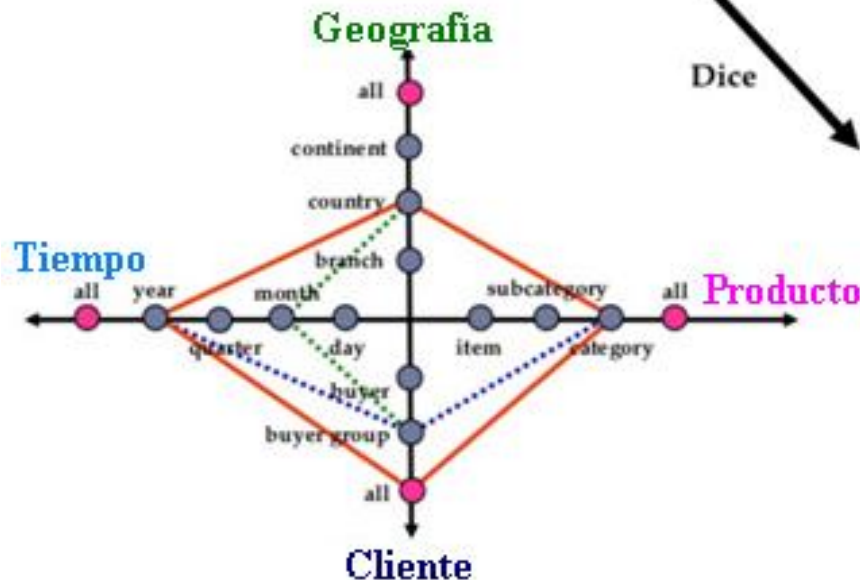
Año	Argentina		Brasil		Total
	Alim.	Bebidas	Alim.	Bebidas	
2006	1100	1500	7000	2000	11600
2007	1200	1200	8000	3000	13400
2008	3000	1200	5000	1800	11000
Total	5300	3900	20000	6800	36000

Dice

VENTAS

Cientes Grupo 1 en Enero

Año	Argentina		Brasil		Total
	Alim.	Bebidas	Alim.	Bebidas	
2006	200	200	800	300	1500
2007	100	100	900	400	1500
2008	300	200	400	100	1000
Total	600	500	2100	800	4000



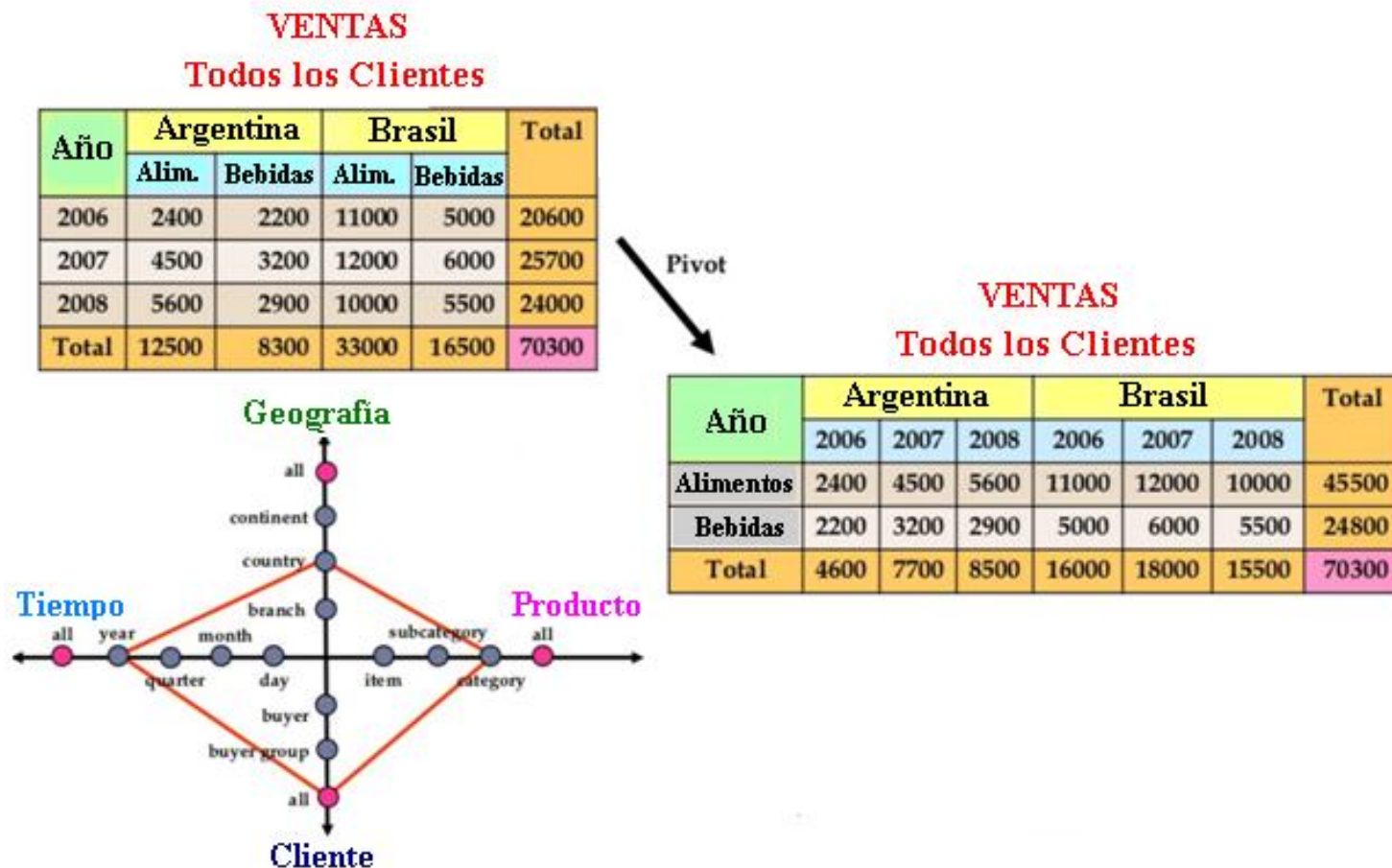


OPERACIONES OLAP

Navegación por las dimensiones

○ Pivot

- Reorientación de dimensiones.



OTRAS OPERACIONES OLAP



- **Drill across:** Conexión con otras tablas de hechos
 - A través de al menos una dimensión en común del cubo.
 - Ejemplo: Información de Ventas: [fecha/sucursal/producto](#)
 - Al analizar el cubo se observan ventas sospechosas en la sucursal XX el día DD/MM/YYYY.
 - Se quiere ver quiénes eran las personas que trabajaron en esa sucursal y en esa fecha.
 - Puede existir también información de Asistencia de empleados: [fecha/sucursal/empleado](#)





OTRAS OPERACIONES OLAP

○ Drill across

VENTAS
Todos los Clientes

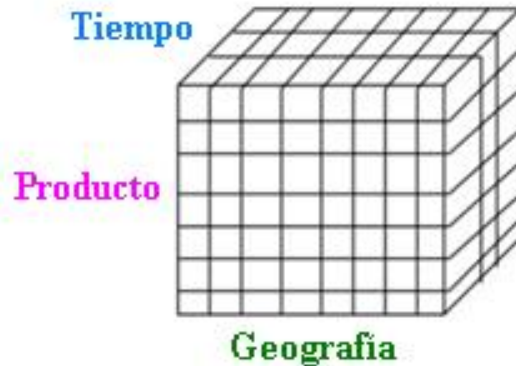
Año	Argentina		Brasil		Total
	Alim.	Bebidas	Alim.	Bebidas	
2006	2400	2200	11000	5000	20600
2007	4500	3200	12000	6000	25700
2008	5600	2900	10000	5500	24000
Total	12500	8300	33000	16500	70300

Drill Across →

COMPRAS
Todos los Clientes

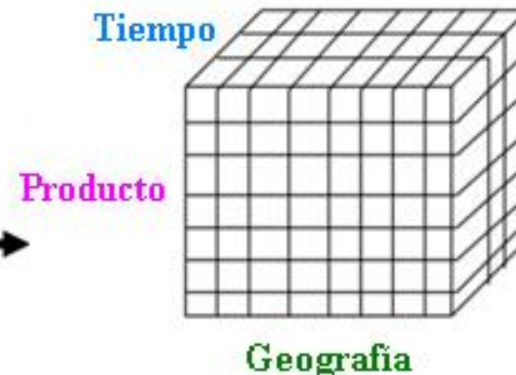
Time	Argentina		Brasil		Total
	Alim.	Bebidas	Alim.	Bebidas	
2006	1500	1500	6000	3000	12000
2007	3500	2500	8000	4000	18000
2008	4000	1500	5000	3000	13500
Total	9000	5500	19000	10000	43500

VENTAS



Drill Across →

COMPRAS



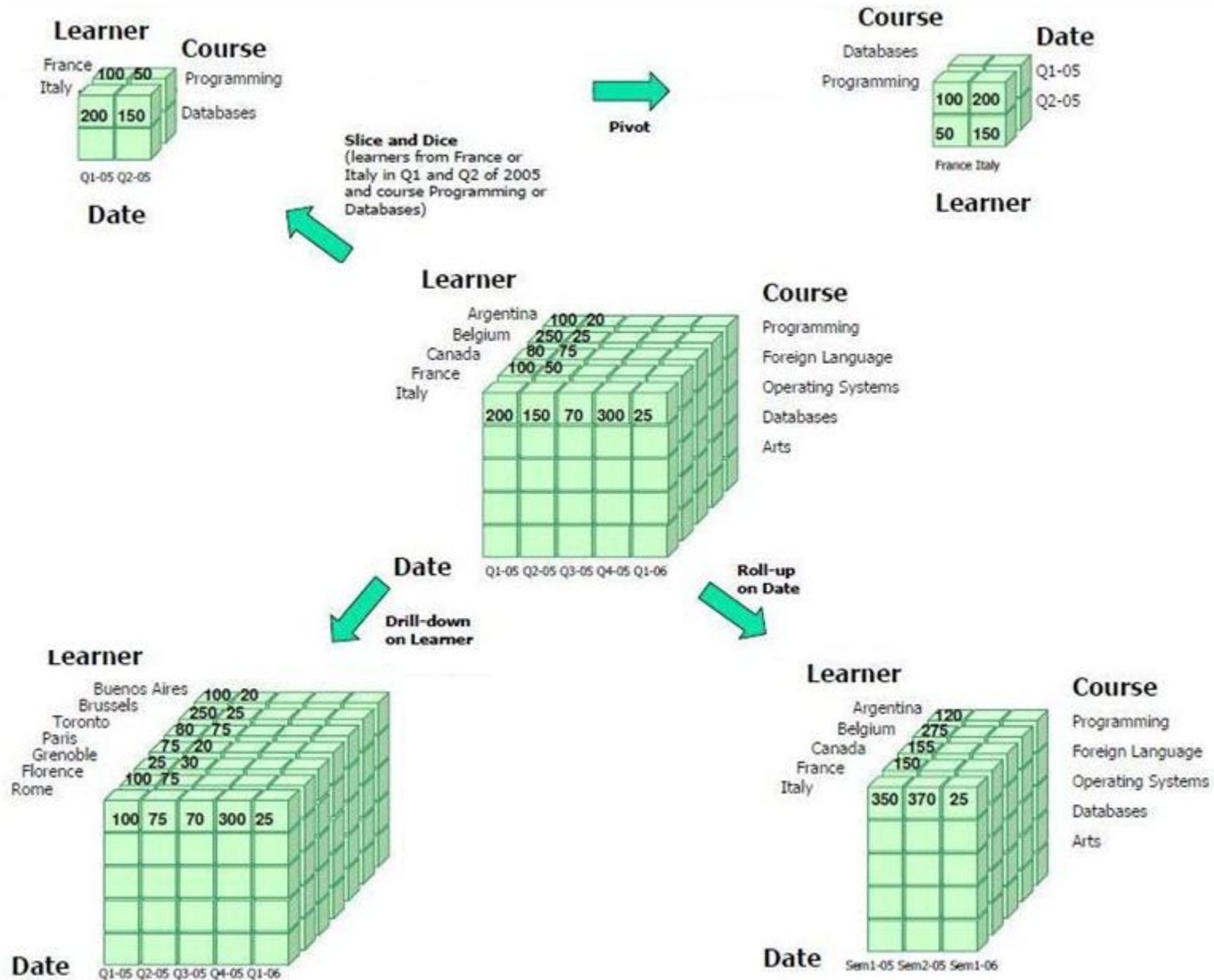
OTRAS OPERACIONES OLAP



- **Drill Through/Drill Out:** Conexión con tablas externas.
 - Desde el nivel más bajo de detalle del cubo se genera una consulta SQL a las tablas relacionales de donde se sacaron los datos, para llegar al nivel atómico.
 - Operación costosa. No debería ser lo habitual sino la excepción.
 - Si es habitual, está mal definido el cubo y hay que agregarle dimensiones para llegar al nivel atómico.



RESUMEN OPERACIONES OLAP

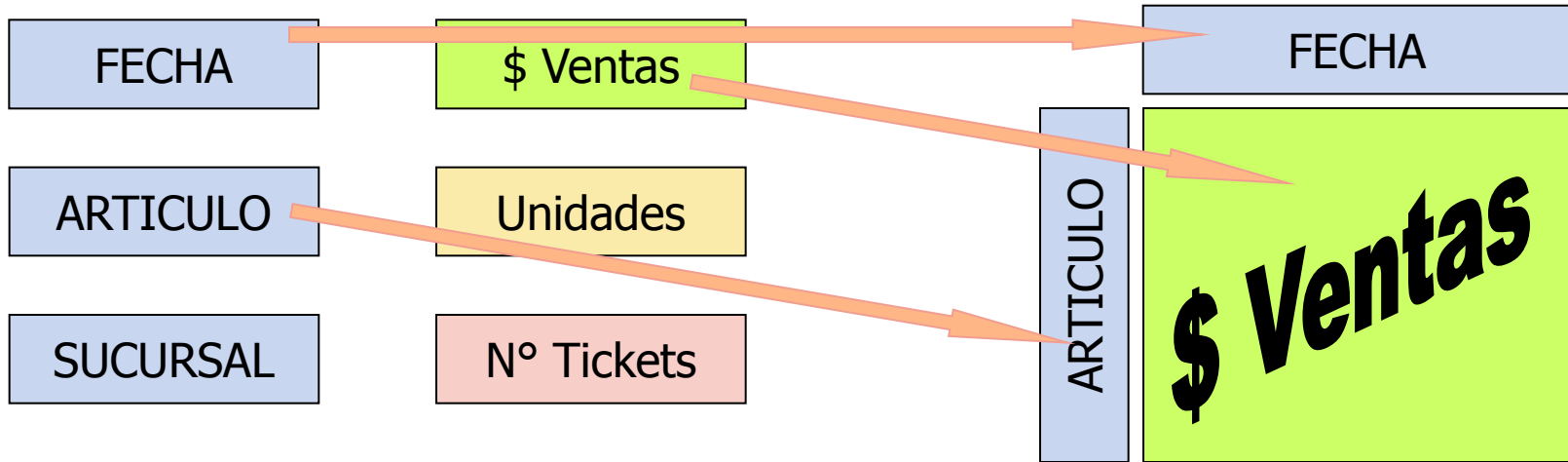


ANÁLISIS OLAP



DIMENSIONES

MEDIDAS



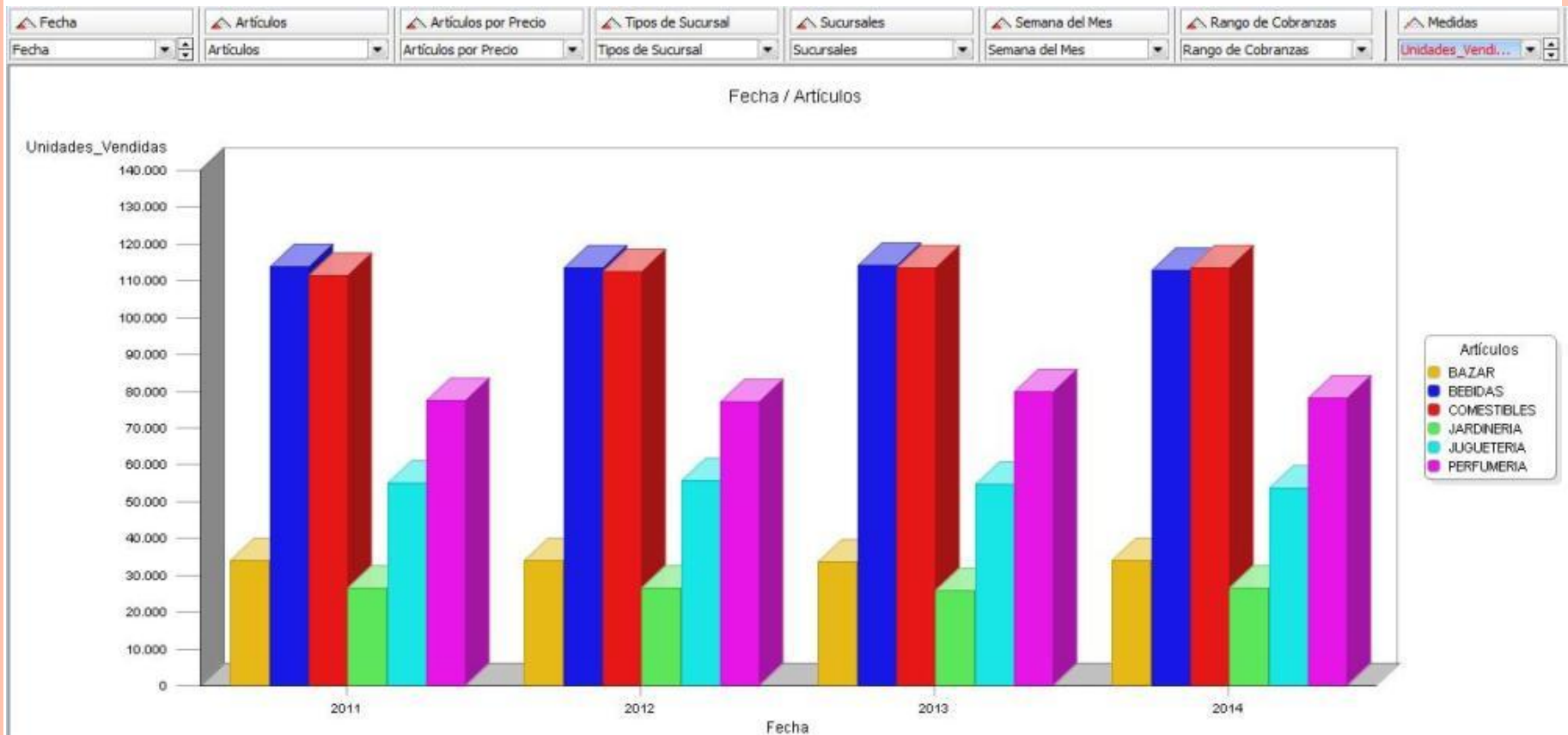
- Se elige la dimensión para las filas.
- Se elige la dimensión para las columnas.
- Se elige la medida a representar.



ANÁLISIS OLAP



- Unidades vendidas por rubro de art. entre 2011 y 2014

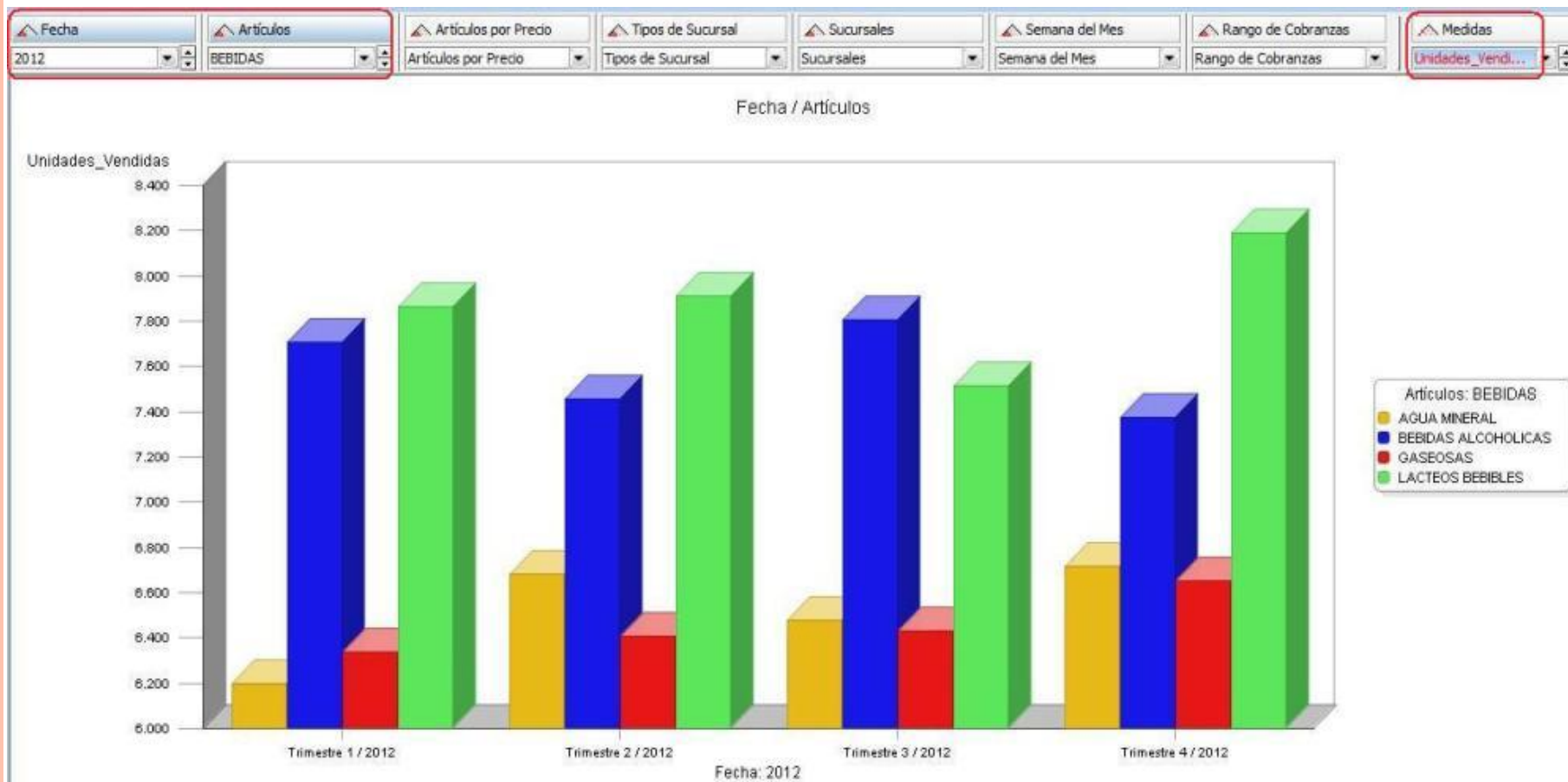


ANÁLISIS OLAP



○ Unidades vendidas por artículos entre 2011 y 2014

□ *Drill Down:* Año **2012** y artículos de **Bebidas**



ANÁLISIS OLAP



- Ventas del 1º trimestre de 2014 por rubro de artículo

Fecha	Artículos	Artículos por Precio	Tipos de Sucursal
Trimestre 1 / 2014	Artículos	Artículos por Precio	Tipos de Sucursal

Fecha -> 2014 -> Trimestre 1 / 2014 / Articulos

Artículos	Fecha		
	Mes		
	Rubro	ene / 2014	feb / 2014
BAZAR		43.826,56	39.569,17
BEBIDAS		34.415,11	29.392,49
COMESTIBLES		48.717,89	43.336,45
JARDINERIA		4.709,52	3.828,19
JUGUETERIA		64.339,10	58.069,97
PERFUMERIA		41.744,60	36.401,13





ANÁLISIS OLAP

- Ventas del 1° trimestre de 2014 por rubro de artículo
 - Agregamos Sucursales

Fecha

Trimestre 1 / 2014

Artículos

Artículos

Artículos por Precio

Artículos por Precio

Tipos de Sucursal

Tipos de Sucursal

Sucursales

Sucursales

Fecha -> 2014 -> Trimestre 1 / 2014 / Artículos, Sucursales

Artículos	Sucursales	Fecha		
		Mes		
Rubro	Ciudad	ene / 2014	feb / 2014	mar / 2014
BAZAR	Avellaneda	7.415,04	7.888,42	8.392,17
	Capital	14.510,94	12.763,32	13.740,90
	Córdoba	3.132,55	3.346,92	3.519,12
	Hurlingham	2.964,38	2.882,54	3.815,02
	Mar del Plata	8.054,26	7.210,76	6.806,98
	Rosario	4.411,15	2.419,55	4.071,08
	San Isidro	3.338,24	3.057,69	3.960,72
BEBIDAS	Avellaneda	5.422,68	5.279,61	6.360,58
	Capital	10.423,60	10.192,80	11.529,99
	Córdoba	3.255,62	2.615,63	2.626,01
	Hurlingham	2.486,82	2.400,79	2.819,45
	Mar del Plata	7.710,47	4.922,80	7.293,84
	Rosario	2.740,09	2.293,69	2.499,22
	San Isidro	2.375,84	1.687,17	3.152,86
COMESTIB...	Avellaneda	8.610,71	6.506,59	6.844,95
	Capital	17.413,63	15.611,91	14.760,34
	Córdoba	3.859,65	2.936,63	4.023,49
	Hurlingham	3.378,07	3.547,02	2.930,89
	Mar del Plata	8.481,91	8.426,26	9.584,42
	Rosario	3.441,46	3.198,11	3.684,04
	San Isidro	3.532,45	3.109,92	2.941,90



ANÁLISIS OLAP

- Ventas de art. de perfumería por nivel de precios y año.
 - Operación Slice and Dice

Fecha	Artículos	Artículos por Precio	Tipos de Sucursal	Sucursales	Semana del Mes
Fecha	PERFUMERIA	Artículos por Precio	Tipos de Sucursal	Sucursales	Semana del Mes

Artículos por Precio, Fecha / Artículos -> PERFUMERIA

Artículos Familia	Artículos por Precio Nivel de Precio				Fecha Año			
	Alto				Medio			
	2011	2012	2013	2014	2011	2012	2013	2014
CHAMPU	111.628,20	135.624,72	159.727,95	175.594,85	49.200,90	60.669,84	70.363,76	78.889,27
DESODORANTES PERSONALES	0,00	0,00	0,00	0,00	47.683,30	56.486,52	65.602,04	76.536,06
JABONES	0,00	0,00	0,00	0,00	10.384,40	11.871,60	13.562,64	15.930,59
PERFUMES	43.101,90	52.090,44	60.121,85	68.696,93	23.260,50	27.064,80	29.767,50	36.095,62

- Operación Pivot

Fecha	Artículos	Artículos por Precio	Tipos de Sucursal
Fecha	PERFUMERIA	Artículos por Precio	Tipos de Sucursal

Fecha / Artículos -> PERFUMERIA, Artículos por Precio

Artículos Familia	Artículos por Precio Nivel de Precio	Fecha Año			
		2011	2012	2013	2014
CHAMPU	Alto	111.628,20	135.624,72	159.727,95	175.594,85
	Medio	49.200,90	60.669,84	70.363,76	78.889,27
DESODORANTES PERSONALES	Medio	47.683,30	56.486,52	65.602,03	76.536,05
	Bajo	10.384,40	11.871,60	13.562,64	15.930,59
JABONES	Bajo	23.546,30	27.570,60	32.465,07	35.130,91
	Alto	43.101,90	52.090,44	60.121,85	68.696,93
PERFUMES	Alto	43.101,90	52.090,44	60.121,85	68.696,93
	Medio	23.260,50	27.064,80	29.767,50	36.095,62



TIPOS DE ALMACENAMIENTO OLAP

¿Dónde se guardan los datos?

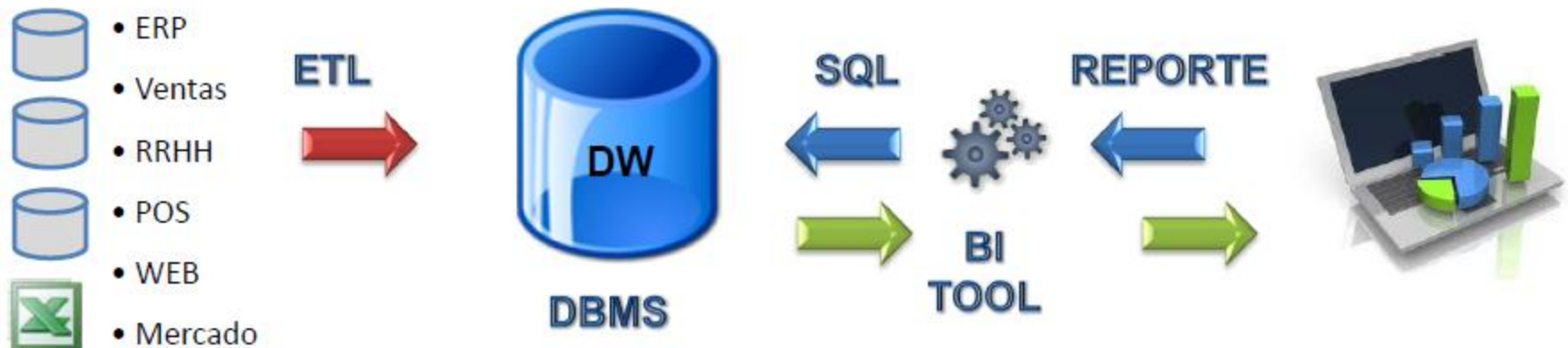
- La base de datos OLAP puede tener distintas implementaciones de acuerdo a la tecnología.
 - ROLAP
 - MOLAP
 - HOLAP





ROLAP (OLAP RELACIONAL)

- Los datos del DW se guardan en BBDD relacionales.
- Se explota toda la información desde el mismo DW.
 - Las consultas multidimensionales generan SQL's al DW
 - Se pueden tener tablas con datos precalculados que aceleran las respuestas a las consultas.
- Escalable
 - Se pueden agregar nuevas tablas o nuevas columnas a tablas existentes.





ROLAP (OLAP RELACIONAL)

○ Ventajas

- Seguridad e integridad de la base de datos.
- Requiere menos espacio ya que no necesita replicar los datos.
- Los datos se almacenan en la base de datos, con lo cual no hay límite de crecimiento.

○ Desventajas

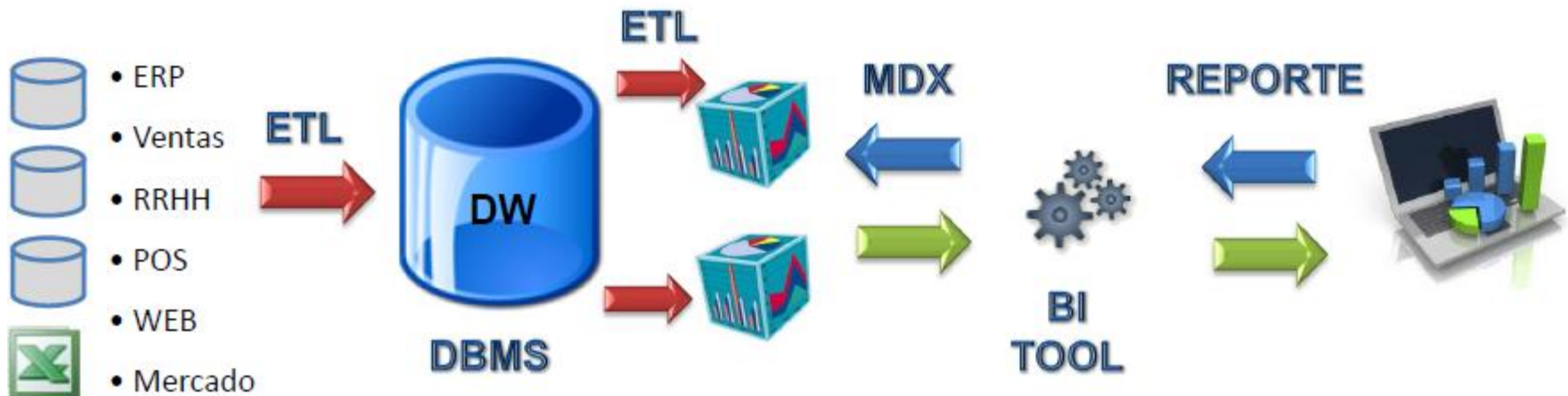
- La respuesta de las consultas es más lenta debido a que siempre tiene que ir a la base relacional a buscar los datos.
- Se debe mantener siempre conexión a la base de datos.
- Los cálculos están limitados a las funciones de la base de datos.





MOLAP (OLAP MULTIDIMENSIONAL)

- Los datos se guardan en una BBDD propietaria con estructura multidimensional.
 - Incluye datos precalculados y metadatos de las jerarquías.
- Todos los datos se cargan directamente a los cubos sin tener un DW y desde allí se consultan los datos.
- Buen tiempo de respuesta para las consultas.





VENTAJAS MOLAP

- Mayor velocidad de acceso a los datos agregados y precalculados.
 - Se guardan en el servidor OLAP en un formato multidimensional sin necesidad de acceder al DW.
- Usa técnicas de compresión para guardar datos en el servidor OLAP, ocupando menos espacio que en una base relacional.
 - Actualmente, las BBDD usan técnicas de compresión.
- El modelo de almacenamiento en vectores/matrices proporciona indexación de tablas hash.
- Estructura muy optimizada para maximizar el rendimiento de las consultas.





DESVENTAJAS MOLAP

- La carga de datos en estructuras multidimensionales puede durar mucho, sobre todo en grandes volúmenes de datos.
 - Esto se puede evitar realizando procesamiento incremental.
- Dificultad y complejidad para analizar modelos con más de 10 dimensiones.
- MOLAP guarda una copia de los datos relacionales en el servidor OLAP, por lo que se necesita invertir en espacio adicional.
- No puede acceder a datos que no están en el cubo.





HOLAP (OLAP HÍBRIDO)

- Utiliza una combinación de MOLAP y ROLAP.
- Los datos con menor nivel de detalle se almacenan en estructuras ROLAP.
- Los datos agregados y pre-calculados por diferentes dimensiones se almacenan en los cubos MOLAP.
- En una consulta, el motor OLAP primero busca si tiene los datos en el cubo. Si los datos no están en el cubo, dispara la consulta sobre la BBDD relacional.

