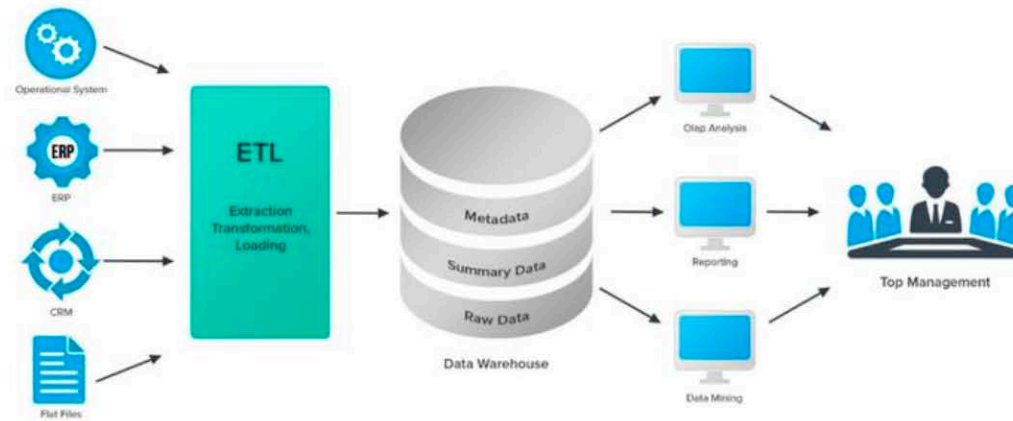


FACULTAD DE INGENIERÍA ESCUELA DE COMPUTACIÓN



Datawarehouse y Minería de Datos

Guía #5: Creación de cubos OLAP - Parte 1

OLAP como herramienta de análisis

Se entiende por **OLAP** (***Proceso Analítico en Línea***) a la metodología para organizar y consultar datos sobre una estructura multidimensional. A diferencia de las bases de datos relacionales, todas las potenciales consultas están calculadas de antemano, lo que proporciona una mayor agilidad y flexibilidad al usuario de negocio.

Los sistemas OLAP quedan representados gráficamente por la figura de un **cubo**; de esta manera es posible, mediante sus dimensiones, catalogar datos descriptivos y además mediante las medidas se informa de datos cuantitativos.

Cada pieza de este cubo contiene ***información específica y cuenta con movilidad***, favoreciendo el proceso analítico de consulta con agilidad.

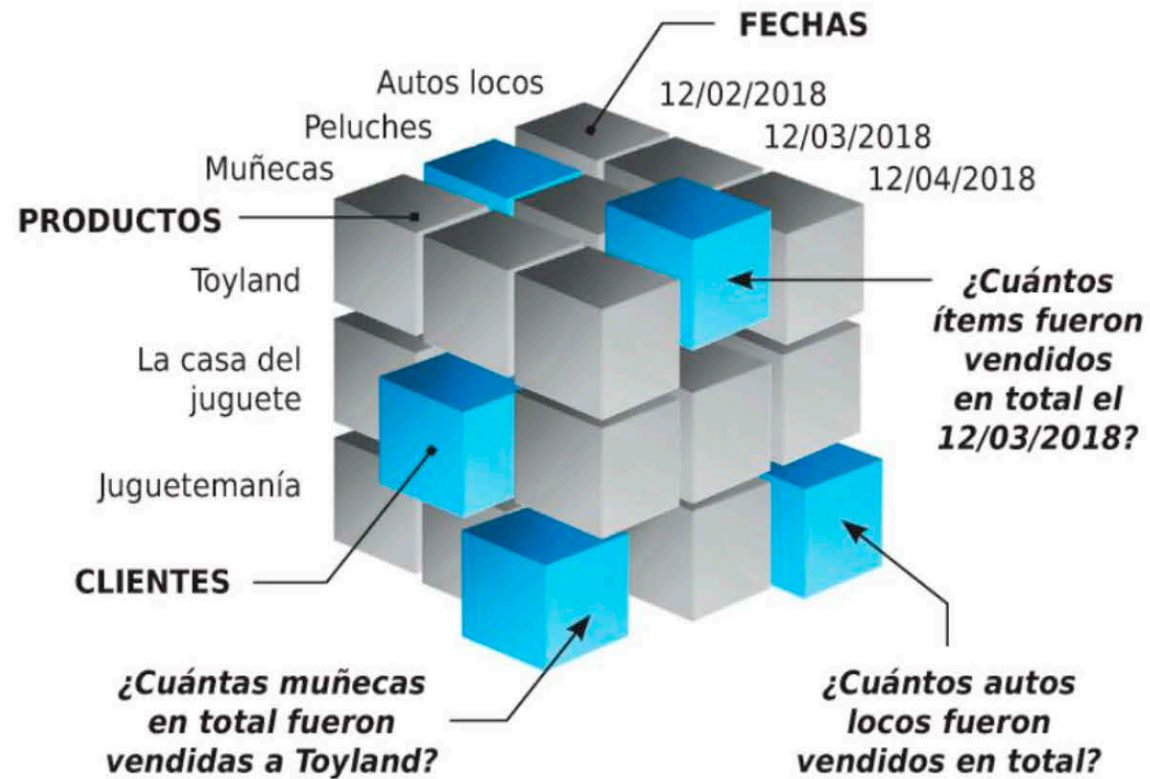
Una herramienta OLAP esta formada por un ***motor y un visor***. El motor es el concepto que acabamos de describir y el visor es una interfaz que permite consultar, manipular, reordenar y filtrar datos existentes en una estructura OLAP mediante una interfaz gráfica de usuario que dispone de funciones de consultas MDX (lenguaje de consultas sobre estructuras multidimensionales).

OLAP como herramienta de análisis

En función de la movilidad de las piezas consultadas se determinara si la técnica de análisis es:

- ***Drill Down (de la generalidad a la particularidad)***: es una de las herramientas más significativas y útiles del sistema OLAP para gerentes de una empresa, ya que permite al usuario desglosar cualquier dato, navegando en su búsqueda desde lo general a lo particular.
- ***Drill Up (un informe te llevar a otro informe relevante para los datos que se analizan)***: es una herramienta cuya técnica de análisis permite al usuario la comparativa entre diferentes informes, por lo que facilitaría un trabajo de comparativa de ventas entre diferentes comerciales.

Ejemplo de un Sistema OLAP aplicado a una juguetería



Elementos OLAP

OLAP permite el análisis multidimensional, lo que significa que la información esta estructurada en ejes (puntos de vistas de análisis) y celdas (valores que se están analizando).

En el contexto OLAP existen diferentes elementos:

- **Esquema:** es una colección de cubos, dimensiones, tablas de hecho, métricas y roles.
- **Cubo:** es una colección de dimensiones asociadas a una tabla de hecho. Un cubo virtual permite cruzar la información entre tablas de hecho a partir de sus dimensiones comunes.
- **Tabla de hecho:** representa un proceso de negocio que analizar.
- **Dimensión:** representa las diferentes perspectivas de análisis de un proceso de negocio.
- **Métrica:** representa el resultado de un proceso de negocio.
- **Jerarquía:** es un conjunto de miembros organizados en niveles. Desde el punto de vista de las bases de datos se puede entender como una ordenación de los atributos de una dimensión.

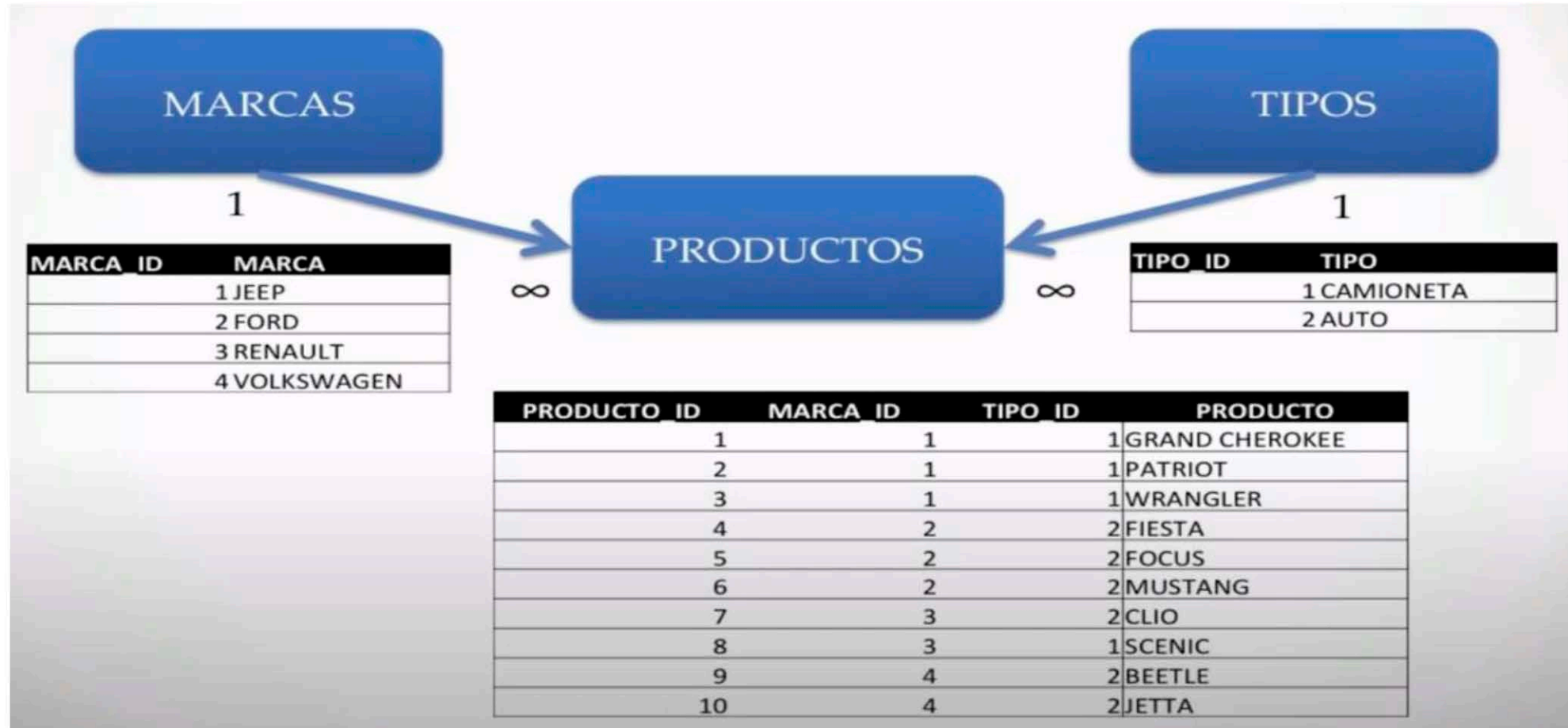
Elementos OLAP

- **Nivel:** es un grupo de miembros en una jerarquía que tienen los mismos atributos y nivel de profundidad en la jerarquía.
- **Granularidad:** Cuanto mayor nivel de detalle tenga la información sobre la que se desea trabajar, mayor será su grado de granularidad, por tanto, mayor será la cantidad de datos a analizar.
- **Miembro:** es un punto en la dimensión de un cubo que pertenece a un determinado nivel de la jerarquía. Las métricas (medidas) en OLAP se consideran un tipo especial de miembro que pertenece a su propio tipo de dimensión. Un miembro puede tener propiedades asociadas.
- **Roles:** permisos asociados a un grupo de usuarios.
- **MDX:** es una acrónimo de multidimensional expressions (aunque también conocido como multidimensional query expresion). Es un lenguaje de consulta de estructuras OLAP, fue creado por Microsoft (año 1997).

Elementos OLAP

- **Nivel:** es un grupo de miembros en una jerarquía que tienen los mismos atributos y nivel de profundidad en la jerarquía.
- **Granularidad:** Cuanto mayor nivel de detalle tenga la información sobre la que se desea trabajar, mayor será su grado de granularidad, por tanto, mayor será la cantidad de datos a analizar.
- **Miembro:** es un punto en la dimensión de un cubo que pertenece a un determinado nivel de la jerarquía. Las métricas (medidas) en OLAP se consideran un tipo especial de miembro que pertenece a su propio tipo de dimensión. Un miembro puede tener propiedades asociadas.
- **Roles:** permisos asociados a un grupo de usuarios.
- **MDX:** es una acrónimo de multidimensional expressions (aunque también conocido como multidimensional query expresion). Es un lenguaje de consulta de estructuras OLAP, fue creado por Microsoft (año 1997).

Modelo Relacional



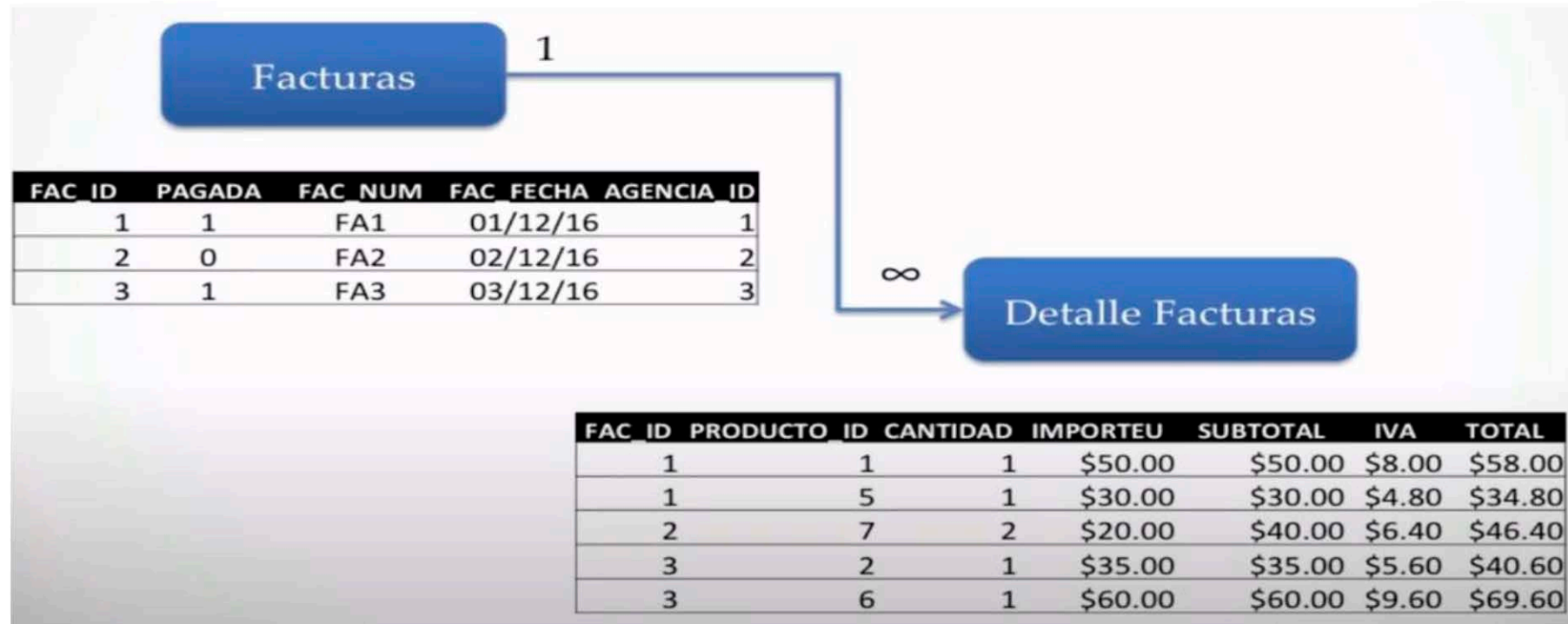
Dimensión Productos

dim_productos

PRODUCTO_ID	PRODUCTO	MARCA	TIPO
1	GRAND CHEROKEE	JEEP	CAMIONETA
2	PATRIOT	JEEP	CAMIONETA
3	WRANGLER	JEEP	CAMIONETA
4	FIESTA	FORD	AUTO
5	FOCUS	FORD	AUTO
6	MUSTANG	FORD	AUTO
7	CLIO	RENAULT	AUTO
8		RENAULT	CAMIONETA
9	BEETLE	VOLKSWAGEN	AUTO
10	JETTA	VOLKSWAGEN	AUTO

Creación de tablas de hechos

Estará conformada por las tablas: factura y detalle de facturas



Formación de tabla de hechos

Diferencias entre tabla de dimensión y tabla de hechos

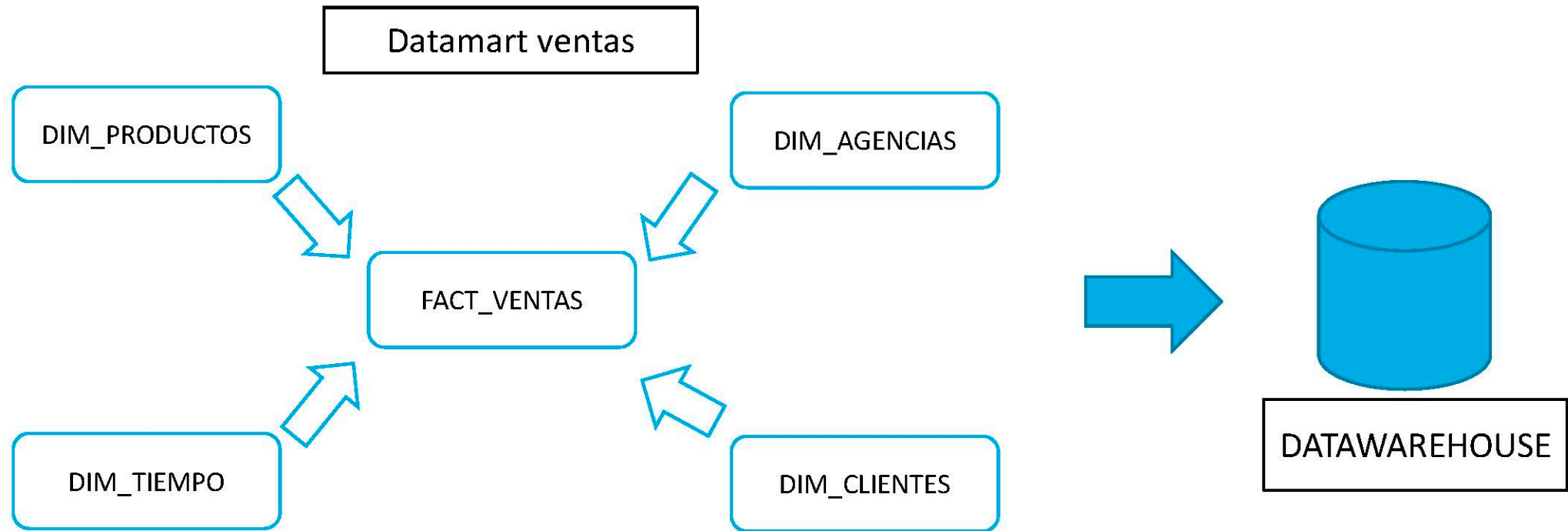
1. En las tablas de dimensiones encontramos valores descriptivos
2. En la tabla de hechos encontramos fechas y valores cuantitativos.

fact_ventas

Llaves foráneas de las tablas dimensiones				Valores cuantitativos						
FAC_ID	PAGADA	PRODUCTO_ID	AGENCIA_ID	FECHA	CANTIDAD	IMPORTEU	SUBTOTAL	IVA	TOTAL	
1	1	1	1	01/12/16	1	\$50.00	\$50.00	\$8.00	\$58.00	
1	1	5	1	01/12/16	1	\$30.00	\$30.00	\$4.80	\$34.80	
2	0	7	2	02/12/16	2	\$20.00	\$40.00	\$6.40	\$46.40	
3	1	2	3	03/12/16	1	\$35.00	\$35.00	\$5.60	\$40.60	
3	1	6	3	03/12/16	1	\$60.00	\$60.00	\$9.60	\$69.60	

Fechas de transacciones

Esquema de Datawarehouse



Comencemos con nuestra práctica

