

Programa Ciencia de los Datos

Curso Minería de datos e Inteligencia de Negocios

Febrero 2020 Prof. Lorena Zúñiga

Tarea #2 Modelado dimensional

Valor 10%

Objetivo

- Identificar dimensiones y medidas a partir de un caso de negocio
- Aplicar principios básicos de modelado dimensional
- Diseñar de forma general un proceso de ETL

Descripción

Utilizando la información sobre las fuentes de datos y el contexto de negocios que se presenta a continuación, debe especificar un modelo dimensional indicando: dimensiones identificadas y posibles atributos, medidas (tabla de hechos) y granularidad.

Se debe identificar también las fuentes de datos (orígenes) desde donde se tomará la información para poblar el modelo dimensional (tablas de dimensión y hechos).

Las fuentes de datos disponibles

La empresa de *retail* actualmente tiene una base de datos llamada Compras, donde se registran las transacciones de las compras que el negocio hace para reponer el inventario de sus diferentes productos.

Una compra puede contener un conjunto de productos, con su respectiva cantidad y costo por unidad. Cada producto tiene asociado un proveedor “titular”, el cual casi siempre supe el producto solicitado. Sin embargo, si fuera necesario, se podría comprar el producto a otro proveedor.

Por su parte, las tiendas de la cadena también tienen proveedores locales para algunos productos propios de la zona en la cual se ubican. En este caso, cuando se hace un pedido se utiliza una hoja de cálculo para registrar la información de la compra que hizo la tienda. No hay políticas uniformes definidas para productos locales, por lo que cada gerente de tienda los solicita cuando siente que es necesario.

A continuación se muestra un ejemplo del contenido de las tablas disponibles en la base de datos de Compras:

Tabla Proveedores

ProvId	Nombre	Email	Teléfono	Descuento
P2092	Interplace	compras@interplace.com	4440-1231	0.06
P3321	Connexion	pedidos@connexion.com	4320-1142	0.11
P4220	EcoPlantas	orders@ecoplantas.com	2213-2234	0.03
P4280	Milano	serviciocliente@milano.com	4463-2215	0.00
P4494	VegeCorp	compras@vegecorp.com	2321-0432	0.07

Tabla Productos

ProdId	Descripcion	ProvId	Cantidad	Precio	FechaProxEnvio
P0005	Albahaca hidropónica	P4220	8	800.00	25/3/2019
P0008	Salsa de tomate	P4280	70	1500.00	01/04/2019
P0010	Apio picado	P4494	10	300.00	

Tabla Compras

Comprald	FechaCompra	Provd	FormaPago	FechaEntrega
C222404	02/04/2019	P4494	Credito	02/05/2019
C234587	02/04/2019	S5095332	Contado	02/16/2019
C324995	02/05/2019	S3399214	Contado	02/07/2019

Tabla ComprasDetalle

Comprald	ProdId	Cantidad	CostoUnidad
C222404	P0010	10	150.00
C222404	P0020	5	85.00
C234587	P0005	10	500.00

Un ejemplo del contenido de la hoja de cálculo para productos locales se muestra a continuación:

CodProd	Desc	Prov	Cant	Inventario	Precio	FechaCompra	Total
PL1	Souvenir 1	S-mart	10	1	200.00	11-Ene-2019	2000.00
PL2	Souvenir 2	xpress	15	4	1500	12-Ene-2019	22500.00
PL3	Souvenir 3	fast	20	0	500	18-Mar-2019	10000.00

¿Qué requiere el negocio?

En este caso la empresa espera que el *datawarehouse* le permita hacer un rastreo de los niveles de inventario a lo largo del tiempo, así como el valor del mismo.

Tenga en cuenta que el nivel de inventario es una medida no aditiva, que no puede ser agregada en el tiempo, es decir, sumar las cantidades de inventario o sumar el valor del inventario en un período de tiempo no tiene sentido.

Para calcular el valor del inventario, la empresa utiliza el precio unitario de cada producto y se multiplica por el número de unidades disponibles.

Algunas de las medidas de interés para el negocio son:

- El promedio del valor del inventario por trimestre
- La diferencia entre niveles de inventario de un mes a otro
- La diferencia entre niveles de inventario en un mismo mes pero en años diferentes
- El nivel mínimo de inventario por mes y año
- El nivel máximo de inventario por mes y año

Se le solicita:

1. Identificar dimensiones y sus respectivos atributos
2. Identificar las medidas por analizar
3. Identificar la granularidad del diseño dimensional
4. Diseñar un esquema estrella que incluya las tablas identificadas en los puntos 1 y 2
5. Para cada atributo identificar la fuente de datos (tabla y atributo origen)

Forma de trabajo: individual o en parejas

Forma de entrega: enviar documento con la solución a través del TECDigital

Fecha de entrega: hasta el domingo 01 de marzo de 2020, a las 11:55 p.m.