

Seminario de Actualización

Proceso ETL – Creación de dimensiones parte 1

Las tareas tienen un orden, como todo proceso ETL (Extracción – Transformación – Carga) deben fijarse las fuentes, las modificaciones y el destino de la información procesada.

En esta primer etapa generaremos las dimensiones pagos y productos.

Para estos ejercicios es necesario tener 2 bases de datos destinadas a tal fin, una en MySQL para el usar como fuente y otra en SqlServer para el DataWarehouse.

dimType:

1. Cargar manualmente una lista de tipos de pagos a través de la función *Añadir Filas Constantes*.
2. En la tabla generar 2 campos, typeld de tipo entero y detalle de tipo string.
3. Ingresar los siguientes tipos de pagos:
 - Efectivo
 - Débito
 - Crédito
 - Cheque
 - Billetera Virtual
 - Transferencia
4. Destinar esa tabla a una llamada dimType en el datawarehouse SqlServer con *clave subrogada csType*.

dimProduct:

1. Utilizar *Tabla – Entrada* para leer de MySQL toda la tabla productos.
2. Cargar las categorías del archivo xml con la función *Obtener datos XML* q incluye el id y el nombre de cada una.
3. Utilizando *Unión por clave*, unir la dos fuentes seleccionando en primer paso la tabla MySQL y segundo paso la lectura del archivo xml, y unir las por la categoryId.
4. Filtrar los productos que estén habilitados seleccionando aquellos que su campo *enabled* sea **1**.
5. Utilizando *Seleccionar / Renombrar Valores* eliminar las columnas:

additional_info

created_at

updated_at
categoryId

6. Almacenar los datos en el datawarehouse SqlServer en la tabla dimProduct con su la clave subrogada se denominará ***scProduct***.

dimUsers:

1. Utilizando *JSON input*, y establecer como fuente de origen el archivo Users.json, recuperar todos los campos y todos los registros.
2. Convertir en mayúsculas el apellido de los usuarios, mediante un paso *String operations* (operaciones de cadena).
3. Concatenar mediante un paso *Fórmula* al apellido previamente transformado en mayúsculas conjuntamente con el nombre, el campo calculado se llamará user.
4. Filtrar solos los registros cuyo campo active se encuentre en verdadero.
5. Quitar todos los campos, quedando únicamente el id de usuario, conjuntamente con la concatenación de apellido y nombre (user). Similar al paso 5 de dimProduct.
6. Volcar todo en un Data Warehouse tabla dimUsers, la clave subrogada se denominará ***csUser***.

dimDate:

1. Incorporar de la conexión a Mysql solo la columna created_at de la tabla movements.
2. Incluir una Entrada Excel, cuyo origen sea Movimientos.xlsx, en la cual se establezcan todos los campos de la única hoja que contiene el libro. Es muy importante establecer que el formato de la columna created_at sea de tipo timestamp.
3. A la entrada de Excel eliminarle todas las columnas excepto la created_at y colocarle una precisión de 9 a través del paso *Seleccionar / Renombrar Valores*.
4. Incluir una Unión ordenada mediante la cual se fusionen ambas extracciones.
5. Por medio de *Unique Rows*, establecer únicamente las filas diferentes.
6. Mediante un paso de tipo *calculadora*, crear tres nuevos campos calculados, de forma jerárquica, en los cuales se obtengan el año, el mes, y el día de created_at.
7. Volcar todo en un Data Warehouse tabla dimDate, la clave subrogada se denominará ***csUser***.

A modo de referencia el espacio de trabajo de pentajo debe resultar similar a la siguiente imagen.

