INTEGRACION DE DATOS-CURSO 23/24 PENTAHO PDI (KETTLE)

TAREFA D

Nesta tarefa crearemos una nova versión do noso ETL para probar a actualización SCD-2 de Kimball

Parte 1: Nova táboa

- 1. Selecciona unha das dimensións (agás a temporal) do teu proceso ETL.
- 2. Crea unha **nova táboa** no teu datamart asociada á dimensión elixida, coas seguintes condicións:
 - O nome das nova táboa debe ser igual que o nome da táboa orixinal, con "extensión" _scd2. (Exemplo: se a táboa orixinal se chamaba dim_axente, crea agora unha nova táboa chamada dim_axente_scd2)
 - A táboa debe ter as mesmas columnas que a orixinal, e as seguintes columnas extra:

valido_desde timestamp without time zone
valido_ate timestamp without time zone
num_version numeric(3,0)
version actual boolean

Parte 2: Nova transformación dimensional

Crea unha **nova transformación** para cargar **a nova táboa dimensional**, cumprindo as seguintes condicións:

- A transformación debe incluír un parámetro de entrada que estableza o momento da carga (dia, mes, ano, en función da dimensión temporal escollida por vós)
- A transformación debe actualizar a nova táboa dimensional indicada na Parte 1.
- A transformación debe implementar unha actualización SCD-2 (ou híbrida, con combinación de SCD-1 e SCD-2):
 - 1. Utiliza a tarefa *DimensionLookupUpdate* de Pentaho en vez da parella *CombinationLookupUpdate* + *Update* (Figura 1)
 - Escolle, para cada atributo da dimensión, a actualización adecuada (insert, update, punch trough). Como mínimo un dos atributos da dimensión debe dar lugar a unha actualización insert, creando novas versións da información das entidades almacenadas (ex. o atributo categoría de axentes, Figura 2).
 - Utiliza o campo 'Extended description' das propiedades da transformación (Figura
 para explicar o porqué das túas eleccións. Unha xustificación inadecuada, ou inexistente, reducirá a túa nota final.
 - 4. Aplica a túa elección para os ditos atributos na configuración de DimensionLookupUpdate.

Figura 1: Nova transformación



Figura 2: Tres versións diferentes da información do mesmo axente na BD

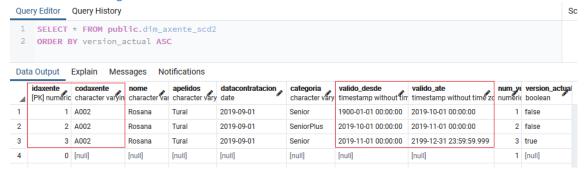
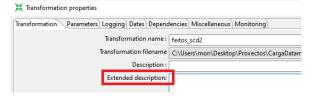


Figura 3: O campo "Extended description" das propiedades da transformación



Información adicional

Dispoñedes de información e exemplos sobre o uso das tarefas dispoñibles en Pentaho PDI no seguinte enderezo: <u>Pentaho Data Integration Steps - Pentaho Data Integration - Pentaho Community Wiki (atlassian.net)</u>

Tamén podedes atopar unha explicación do uso de Dimensión Lookup Update (DLU) aquí:

https://www.joyofdata.de/blog/handling-scd-type-2-using-dimension-lookup-update-step-in-pentaho-kettle/

Revisa tamén as diapositivas "Pentaho lookups" e "ETL: carga" dispoñibles no Moodle.