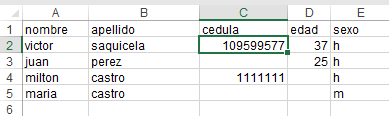
**Interciclo Minería de Datos**

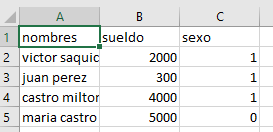
Nombre: Freddy L. Abad L.

Fuentes de Datos

1. Archivo Excel - Nombre de Archivo Excel.xlsx



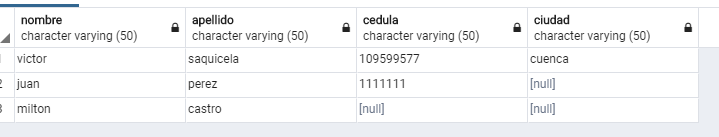
1. Archivo CSV (MSDOS) - Nombre de Archivo: archivocomas.csv



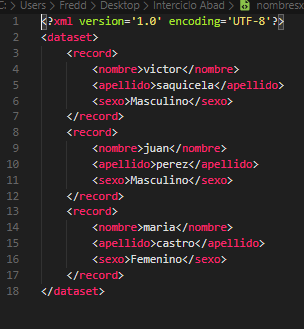
1. Base de Datos **PostgreSQL**

Nombre Base de Datos: Interciclo\_dw

Nombre Tabla: personas

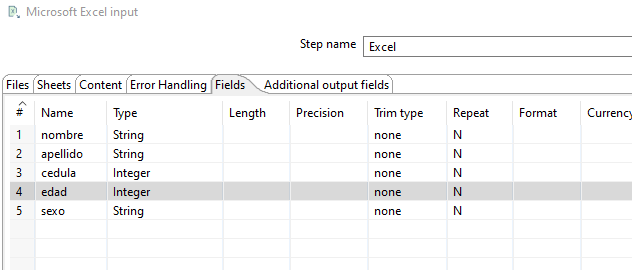


1. XML - Nombre de Archivo: nombresxml.xml

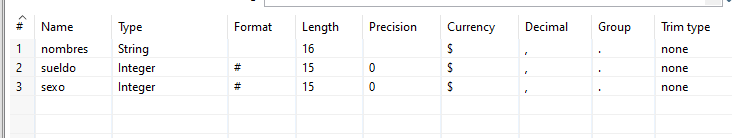


Configuración de Tipos de Datos - Lectura de Archivo

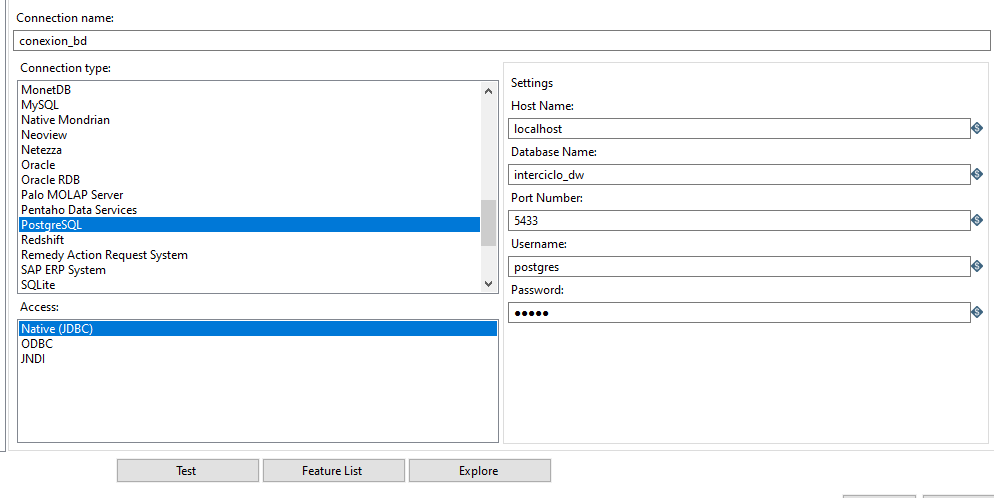
Excel

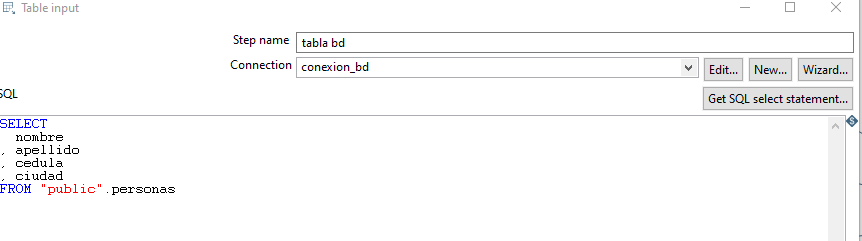


CSV

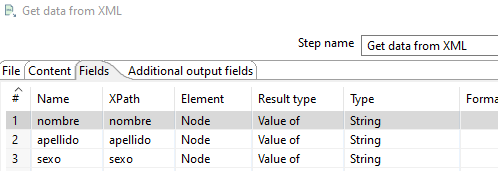


BD

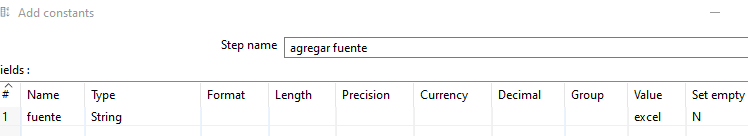




XML

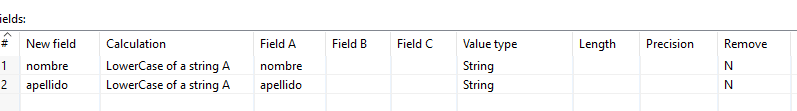


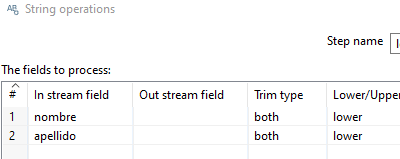
Despues del ingreso de los archivos, en lso 4 casos se define un operador tipo addconstant, definiendo que tipo de fuente es (excel, csv, postgres, xml)



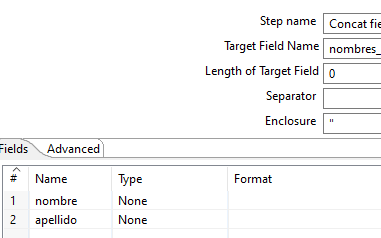
Se estandariza en minusculas todos los nombres mediante el operador calculator o string operation



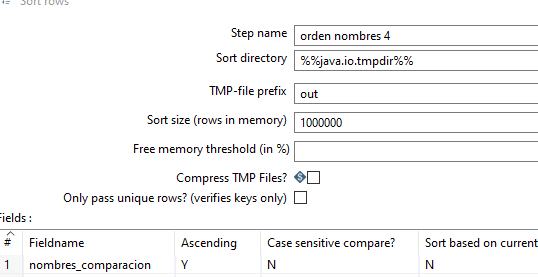




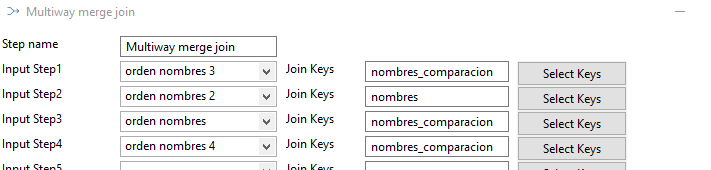
En el caso de tipo de archiv xml, excel y base de datos, donde los registros de nombre y apellido separados se procede a hacer un paso tipo concat field para obtener un solo campo a comparar



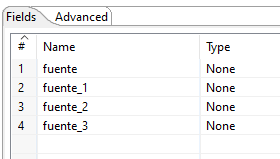
Se ordena los valores a comparar en el join mediante el Sort Row



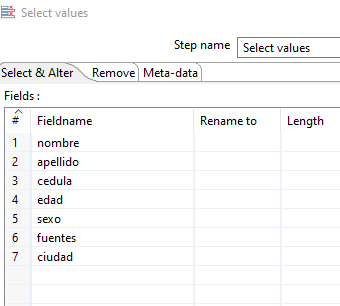
Se una un multiple merge join en un tipo Full Outer Join



Con el resultado de estos se une las salidas de las fuentes que permitiran saber que fuentes de datos nomas interviniern en cada registro

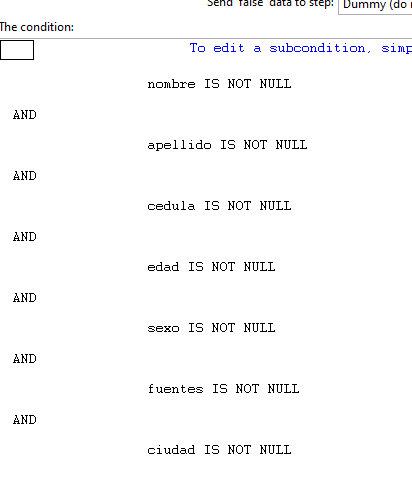


Al ser un full outter join, se unen todas las variables implicadas, lo que genera variables innecesarias, para seleccionar las necesarias se usa sele rows

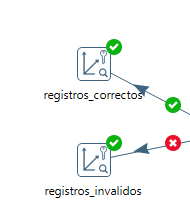


Se filtra los datos para obtener las 2 bases de datos con datos llenos y varios respectivamente, mediante el operador filter rows con la restricción de no ser nulo

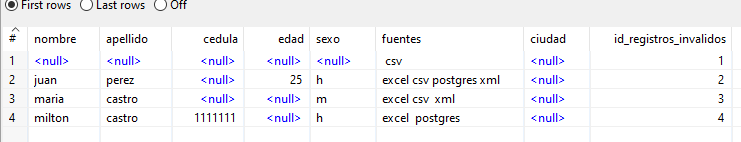




Finalmente se registra los datos resultantes nuevas tabla definidas por el operador combination look/update



Obteniendo finalmente la tabla de datos con vacíos

 y la tabla con datos llenos