PRACTICA CON HERRAMIENTA SPOON

SAMIL SANCHEZ ACEVEDO

TALLER

REALIZAR ACTIVIDAD ETL CON SPOON EL DPI DE PENTAHO

**Fundacion universitaria UNINPAHU**

**Bogota 8 Marzo de 2019**

Tabla de contenido

[Introducción 3](#_Toc5736460)

[Desarrollo del Taller 4](#_Toc5736461)

[Carga de datos 4](#_Toc5736462)

[Filtrar 5](#_Toc5736463)

[Cargar Datos en Archivo de Excel 6](#_Toc5736464)

[Carga de datos complementaria 7](#_Toc5736465)

[Corregir los datos faltantes 8](#_Toc5736466)

[Limpiar Diseño 9](#_Toc5736467)

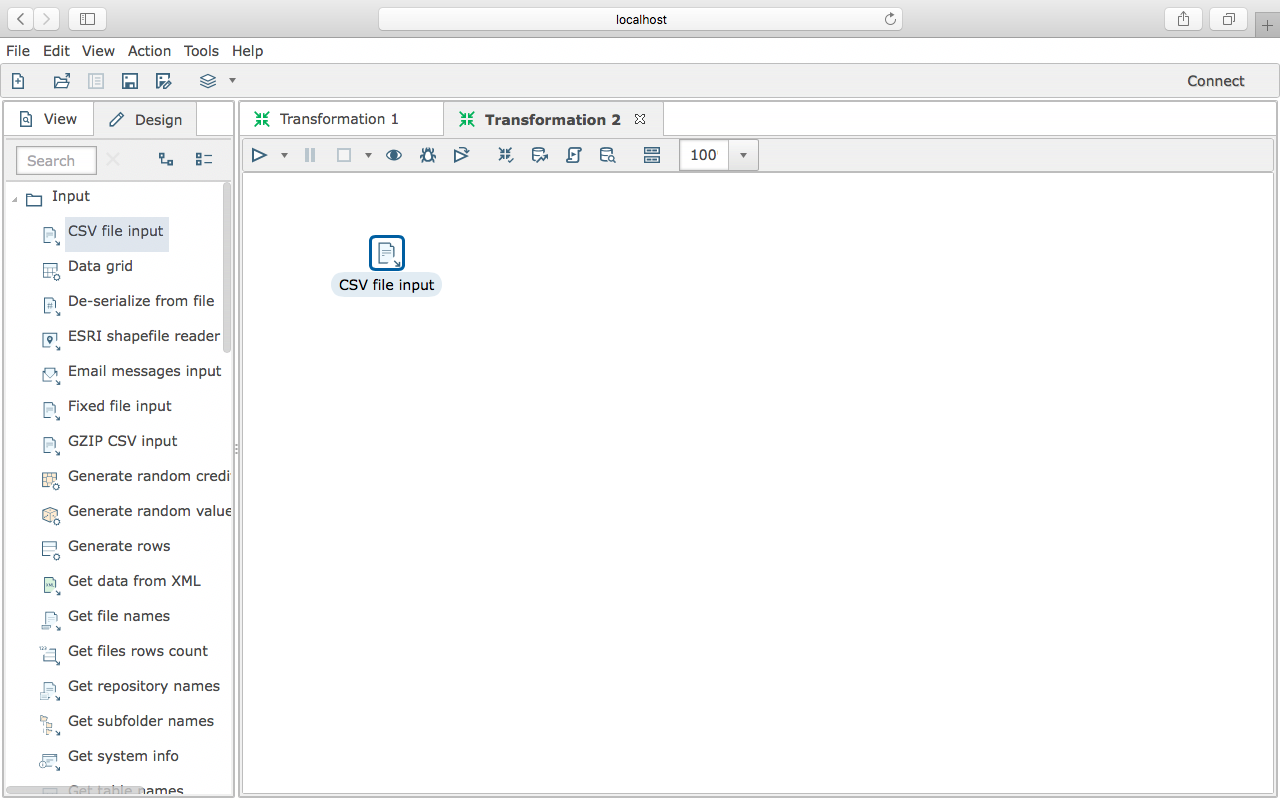
**Introducción**

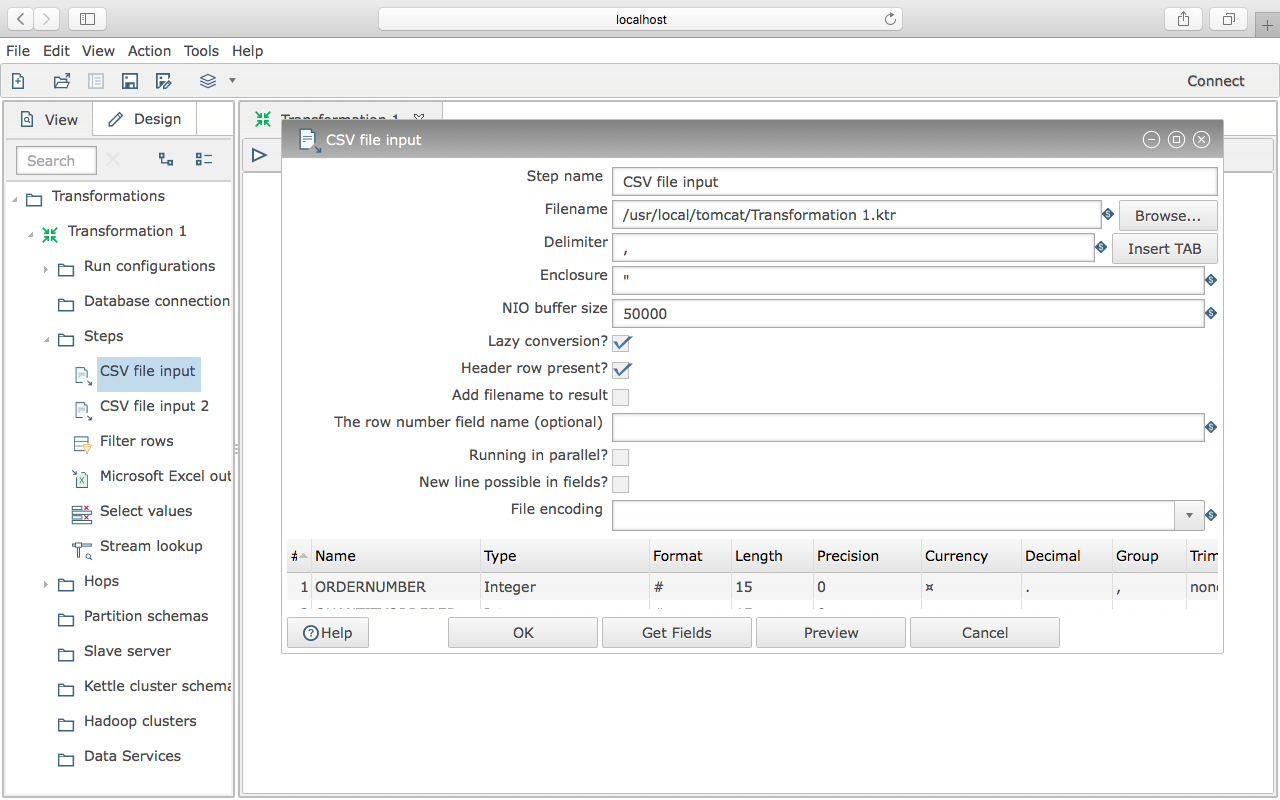
Según la página oficial de pentaho, las transformaciones (Transformations) se utilizan para describir los flujos de datos para ETL, como la lectura desde un origen, la transformación de datos y la carga en una ubicación de destino. Los trabajos (Jobs) se utilizan para coordinar actividades de ETL, como definir el flujo y las dependencias para las transformaciones de orden que deben ejecutarse, o prepararse para la ejecución mediante la verificación de condiciones como "¿Está disponible mi archivo de origen?" O "¿Existe una tabla en mi base de datos?"

# Desarrollo del Taller

## Carga de datos

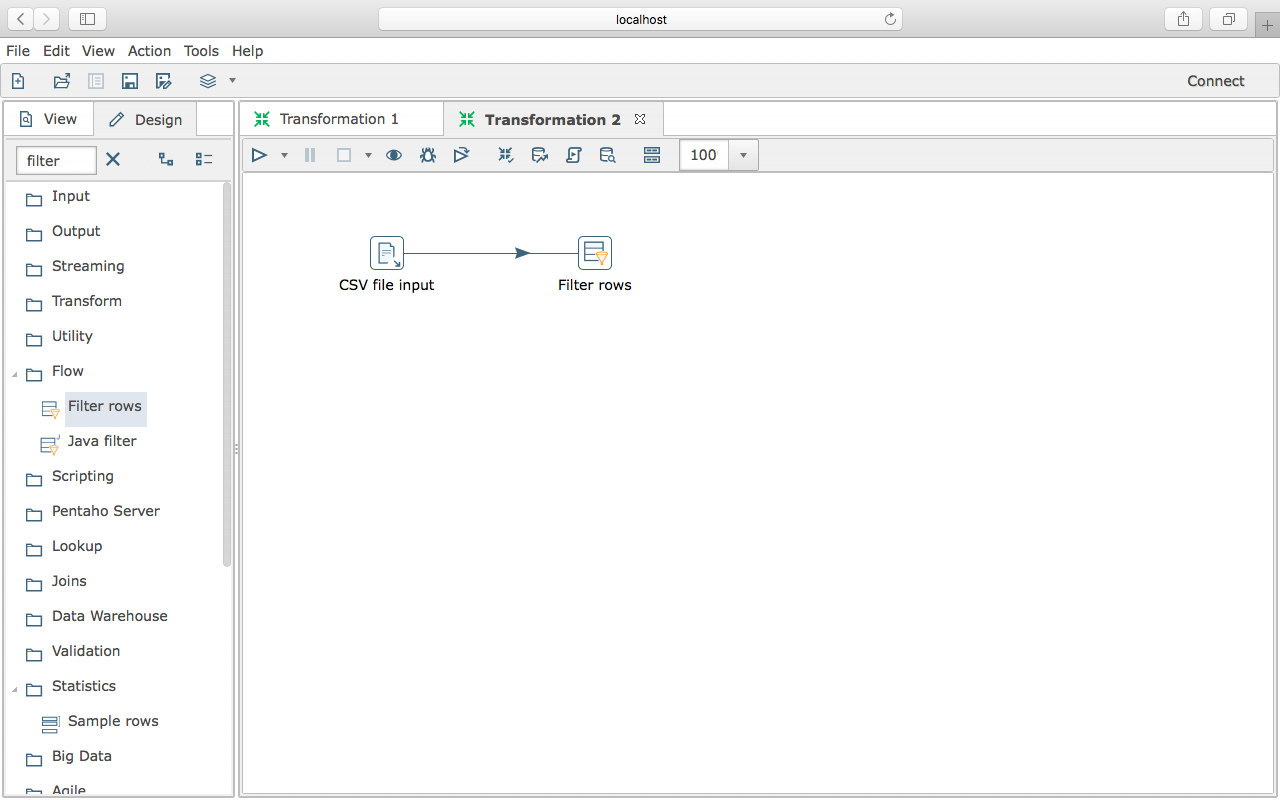
Primero agregamos un CVS file Input donde se cargaran los datos iniciales, arrastramos el icono a nuestra area de trabajo y luego configuramos la ruta donde esta almacenada la informacion:



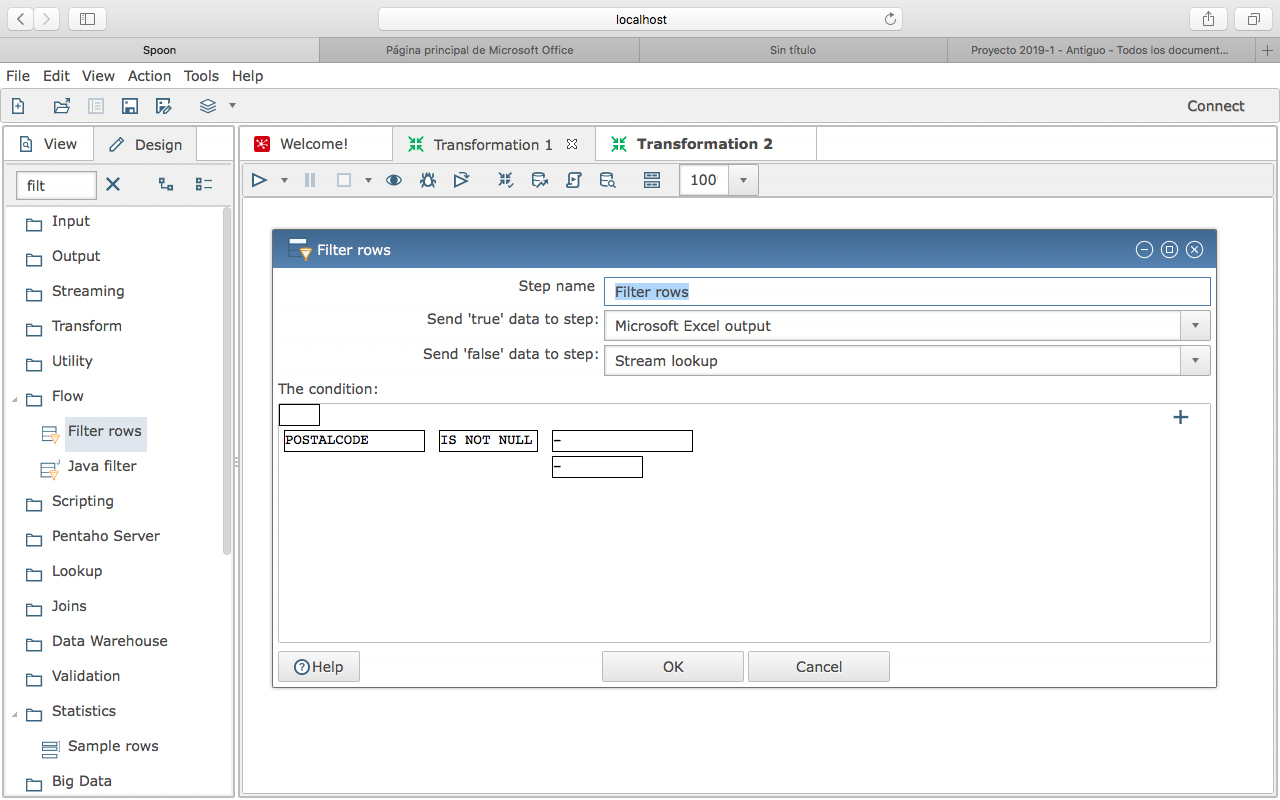


## Filtrar

Continuamso filtrando los registros con códigos postales que faltan en la tabla inicial, para ello utilizamos Filter Rows, traemos el icono a nuestra area de trabajo y los unimos con un salto o paso

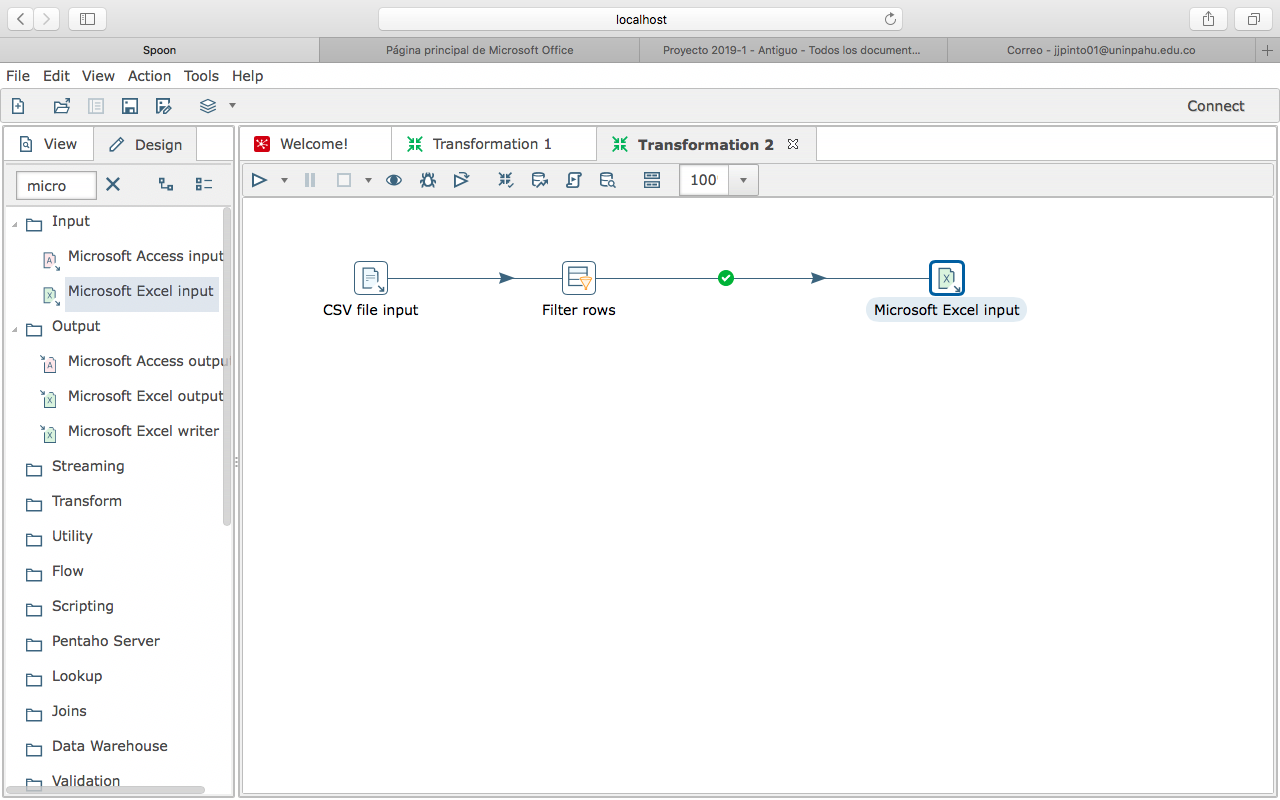


Agregamos los parametros del filtro que seran la condicion de buscar en la columna de codigo postal los campos que contienen datos, la condicion en verdadero pasara las registros al archivo de excel que crearemos mas adelante, los registros en negativo para este caso son los que estan vacios en el campo Codigo Postal pasaran a un Stream lookup que crearemos mas adelante.

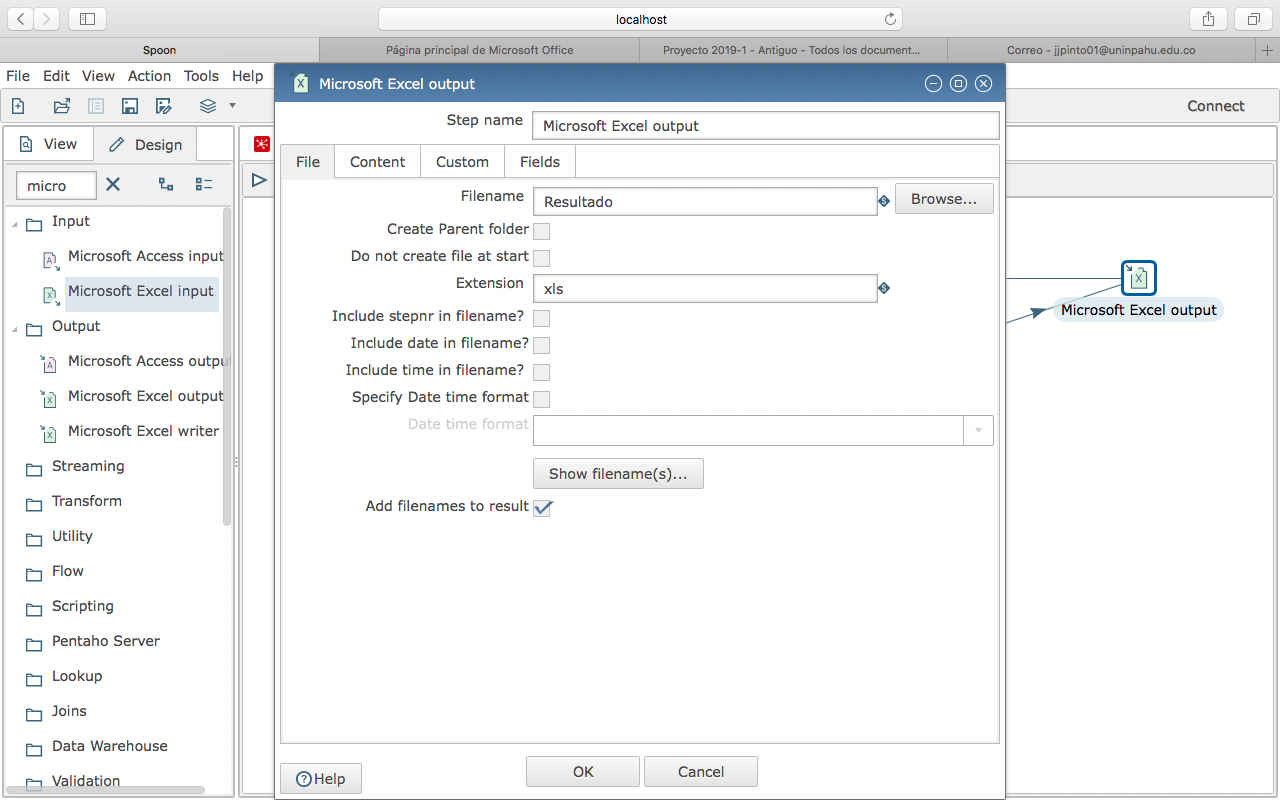


## Cargar Datos en Archivo de Excel

Para generar como salida final de la transformacion un archivo de excel agregamos un objeto Microsoft Excel output con un paso desde Filter Rows para que envie la informacion desde el filtro al archivo destino de excel.

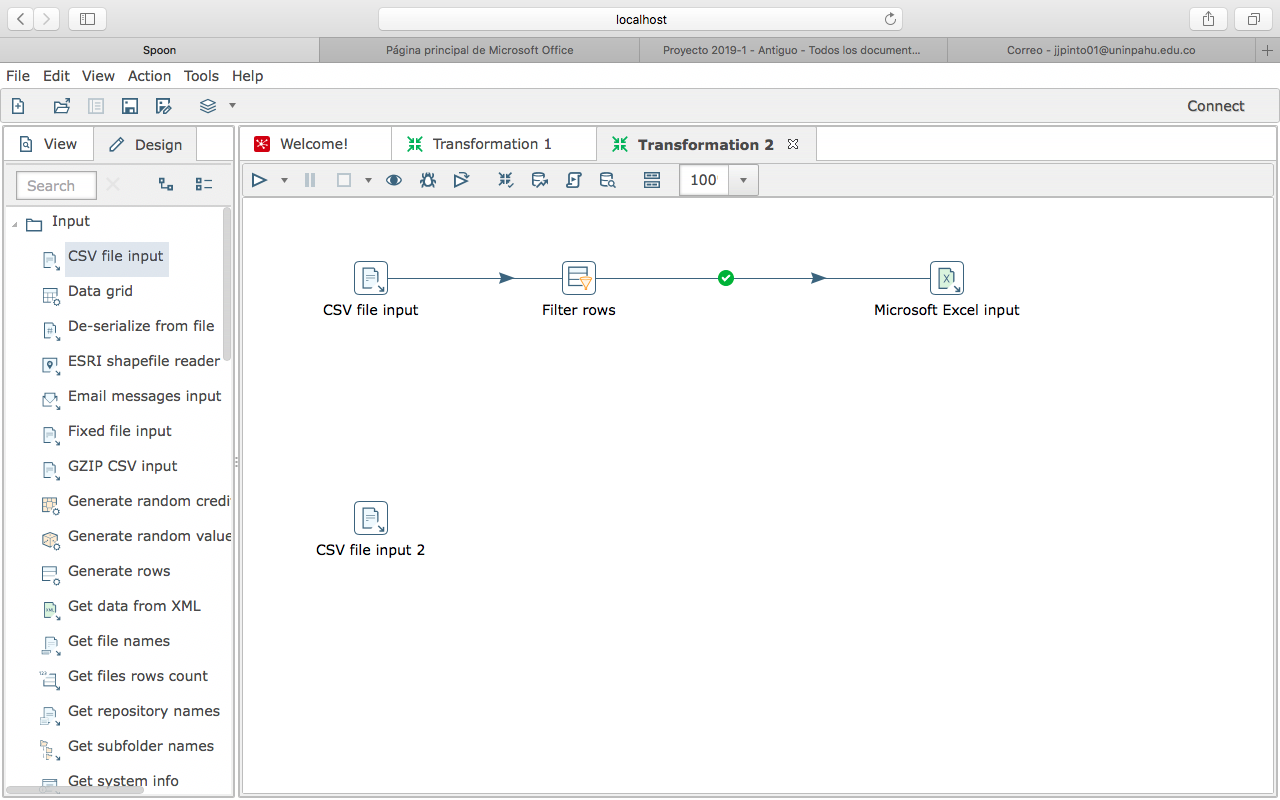


Configuramos los parametros basicos como nombre, tipo de archivo, entre otros:

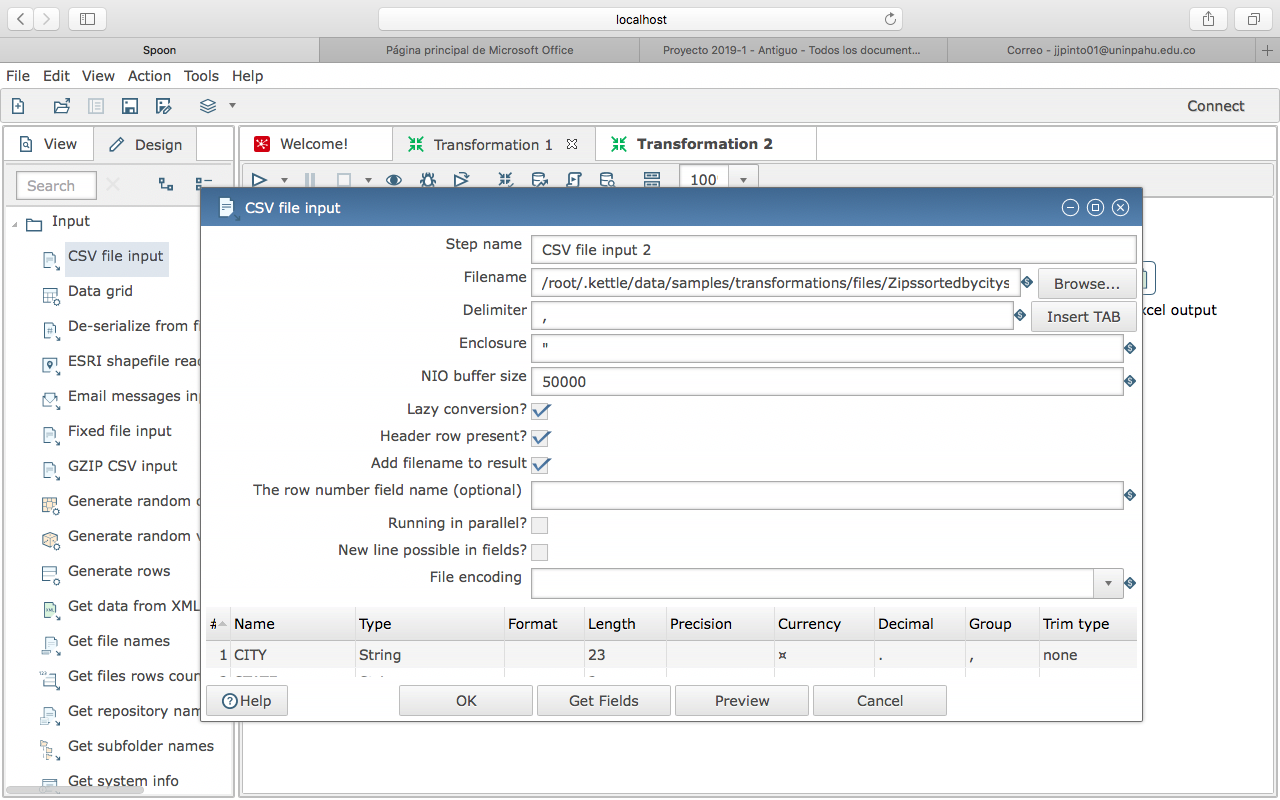


## Carga de datos complementaria

Como tenemos algunos registros sin codigo postal que resultaron del filtro nos apoyaremos con un archivo de datos que contiene ciudades, estados y códigos postales que ahora utilizará para buscar los códigos postales de todos los registros en los que faltaban. La carga es igual el primer archivo CVS que hicimos al inicio solo que cambiamos la ruta donde se encuentra.

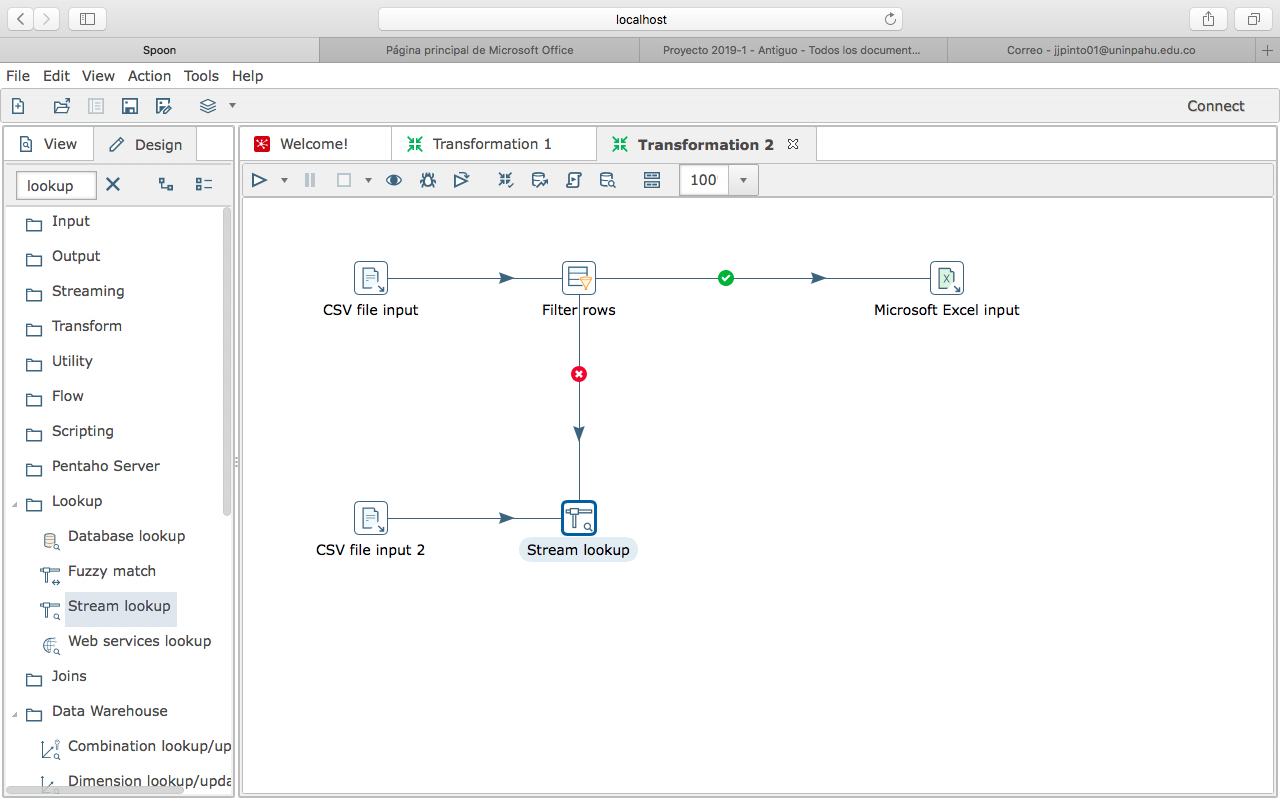


Agregamos la ruta donde se encuentra el archivo.

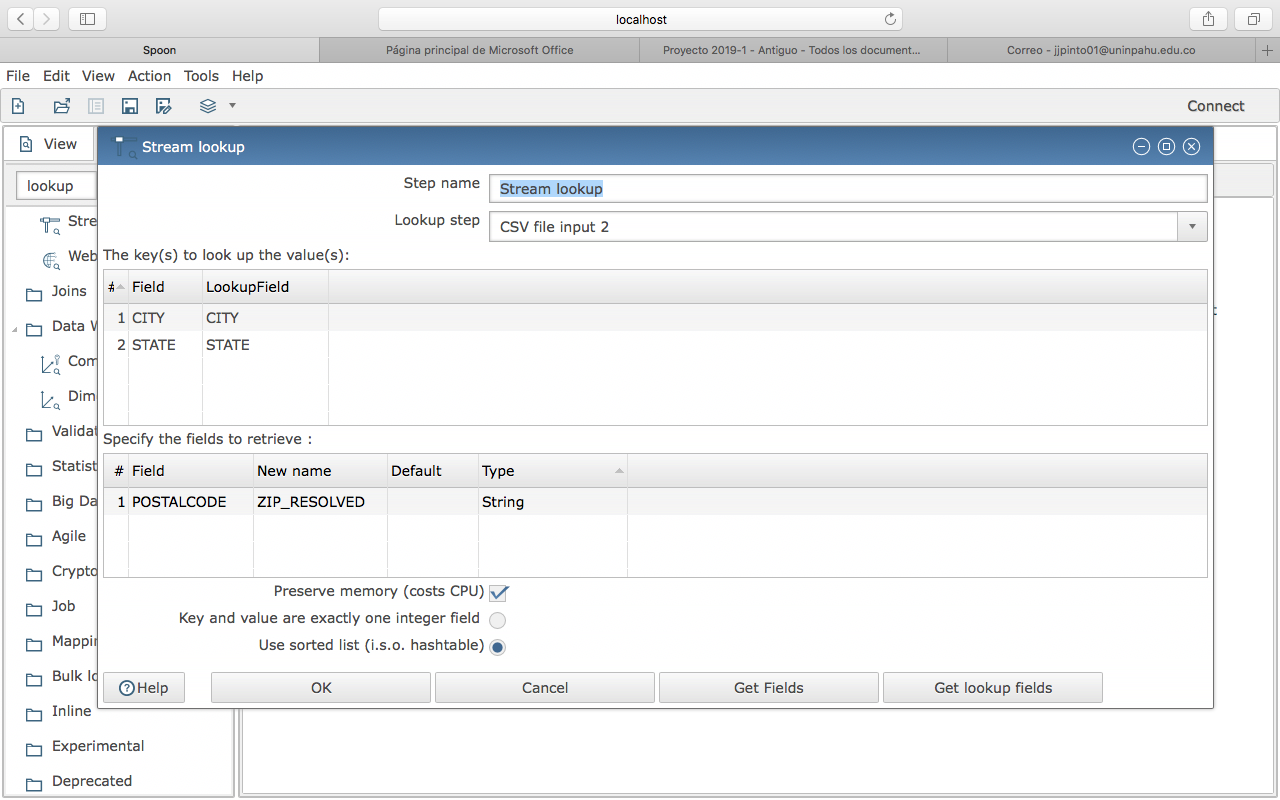


## Corregir los datos faltantes

Para agregar el codigo postal faltante agregamos Stream lookup que tomara los datos que faltan y los consultara en el segundo Stream lookup CVS que agregamos para tal fin, este lo unimos con el paso False del Filter Rows y un segundo paso desde el CVS dos para que pueda consultar por ciudad y traer el codigo postal que corresponda



Agregamos la configuracion anteriormente mencionada.

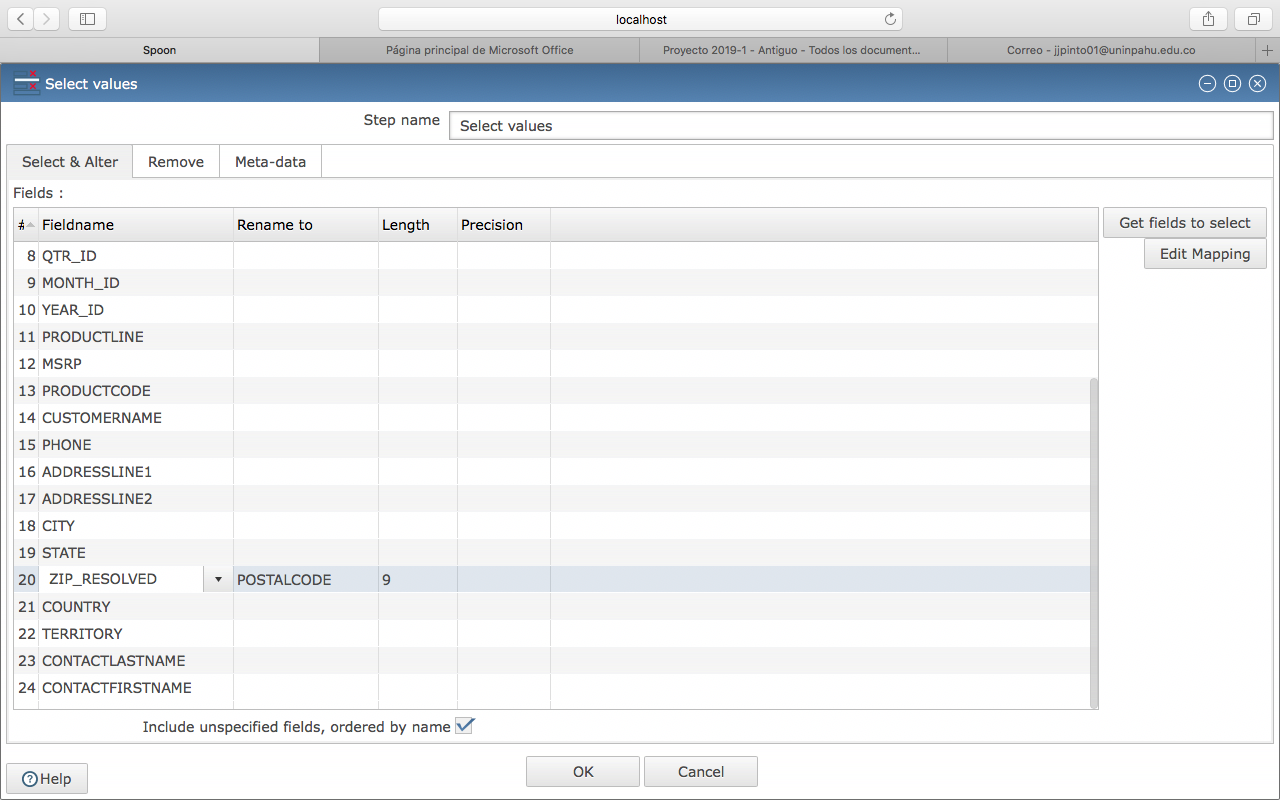


## Limpiar Diseño

Para finalizar realizamos una limpiesa en que coincidad los datos optenidos en cuanto a diseño y formato, eliminaremos los campos innecesarios.

Ingresamos a la configuracion para recuperar los campos, luego el campo ZIP\_RESOLVED lo subimo de nivel con Control y flecha arriba para moverlo justo debajo de POSTALCODE, luego tomamos codigo POSTALCODE y lo eliminamos.

Como se ha formateado el campo lo configuramos con longitud 9 y en la columna renombrar escribimos POSTALCODE para que tome el campo y asi coincida con la fuente.



**Con esto ya tendremos nuestra transformación**

