

MANUAL DE CREACIÓN DE TRASFORMACIONES EN SPOON

Autor:

Ing. Alex Eduardo Vargas Soria. (alexeduvargas12@yahoo.es)

Contenido

1. Introducción.....	1
2. Objetivos.....	1
3. Contenido	1

1. Introducción

Spoon es el diseñador gráfico de transformaciones y trabajos del sistema de ETLs de Pentaho Data Integration (PDI), también conocido como Kettle (acrónimo recursivo: "Kettle Extraction, Transformation, Transportation, and Load Environment "). Está diseñado para ayudar en los procesos ETLs, que incluyen la Extracción, Transformación, Transporte y Carga de datos.

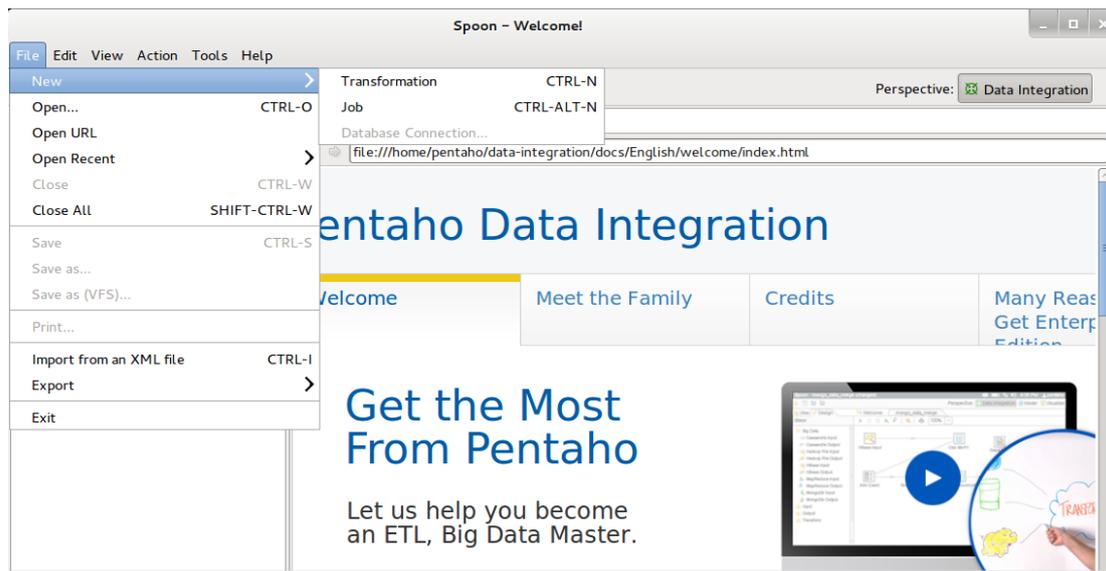
Spoon es una Interfaz Gráfica de Usuario (GUI), que permite diseñar transformaciones y trabajos que se pueden ejecutar con las herramientas de Kettle (Pan y Kitchen).

2. Objetivos

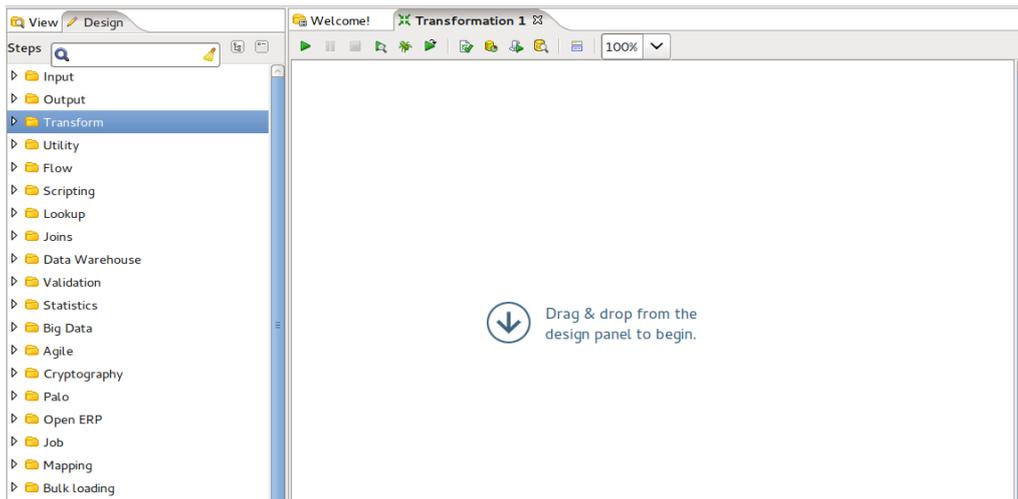
- Proporcionar una guía que permita al usuario que acceda a este manual mayor facilidad al momento de crear transformaciones en Pentaho.
- Documentar el proceso de creación de transformaciones con el fin de generar información.

3. Contenido

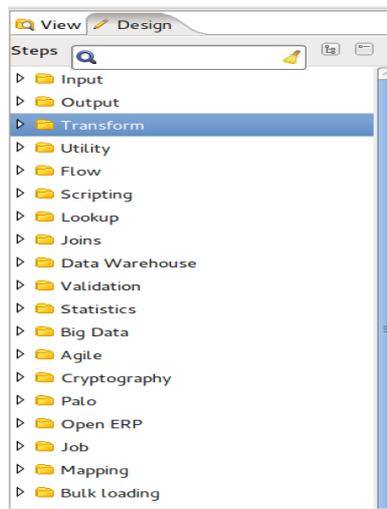
- Primero debemos crear una transformación, para esto es necesario realizar los siguientes pasos, o a la vez presionar las teclas ctrl + n:
 - ✓ File
 - ✓ New
 - ✓ Transformation



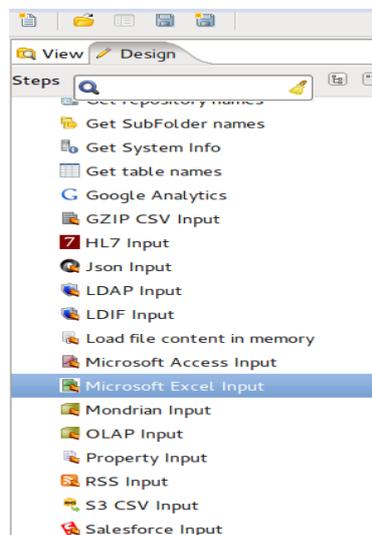
- Se muestra la nueva transformación creada

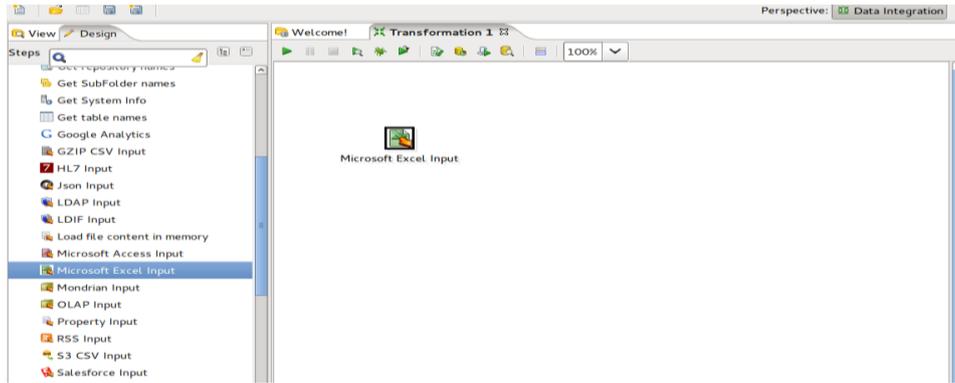


- Para ingresar algún tipo de información se utiliza las herramientas contenidas en la carpeta "Input".

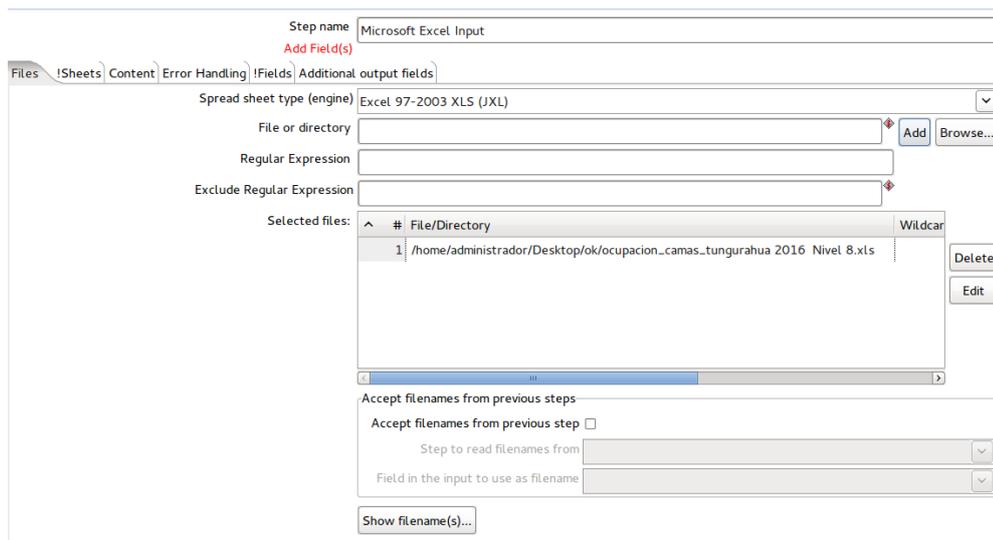


- Para este ejemplo utilizaremos un archivo de Microsoft Excel

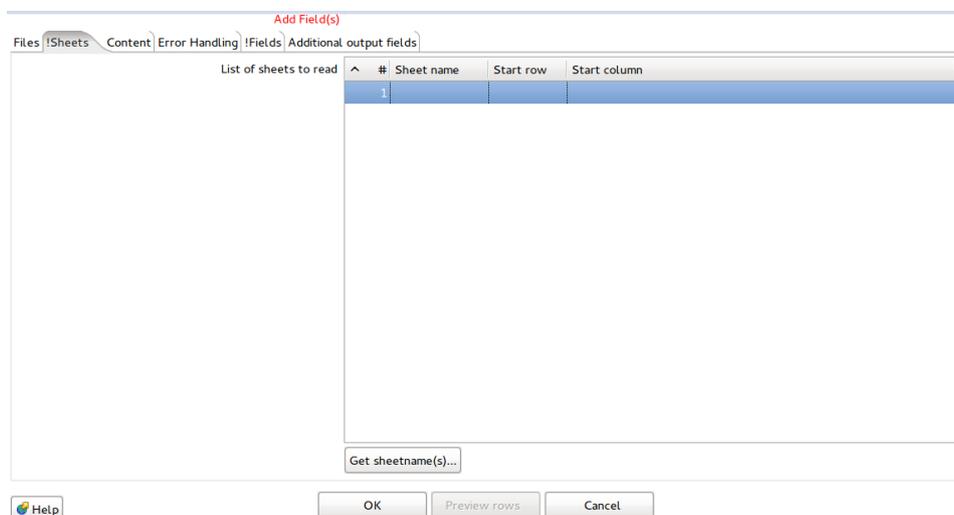




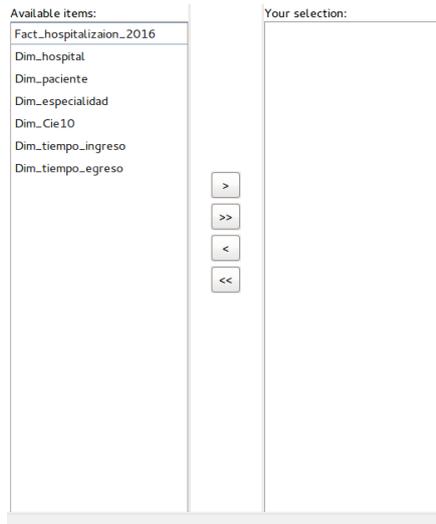
- Configuraciones necesarias del Microsoft Excel Input
 - ✓ En files se añade el archivo en Excel a ingresar.
 - ✓ Una vez seleccionado se presiona “Add” para agregarlo



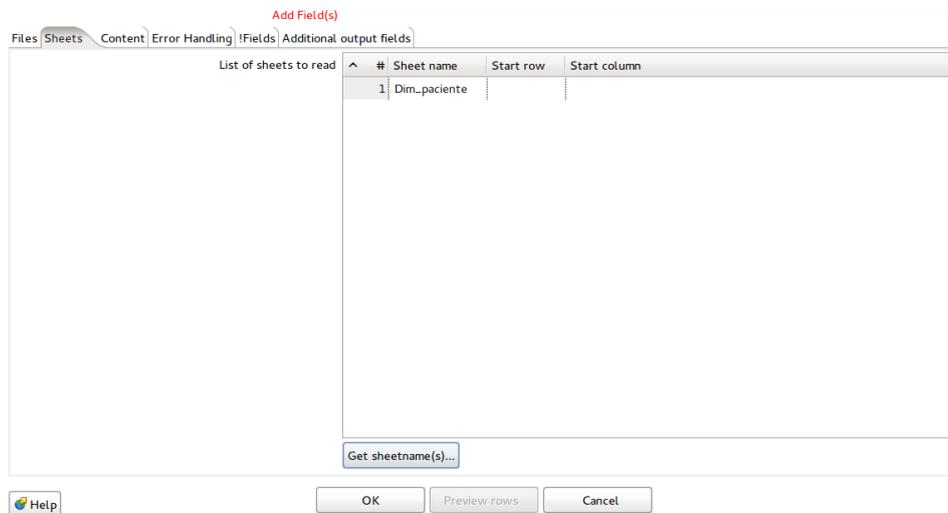
- En la pestaña !Sheets seleccionamos la hoja del libro de Excel a utilizar presionamos “Get sheetname(s)”.



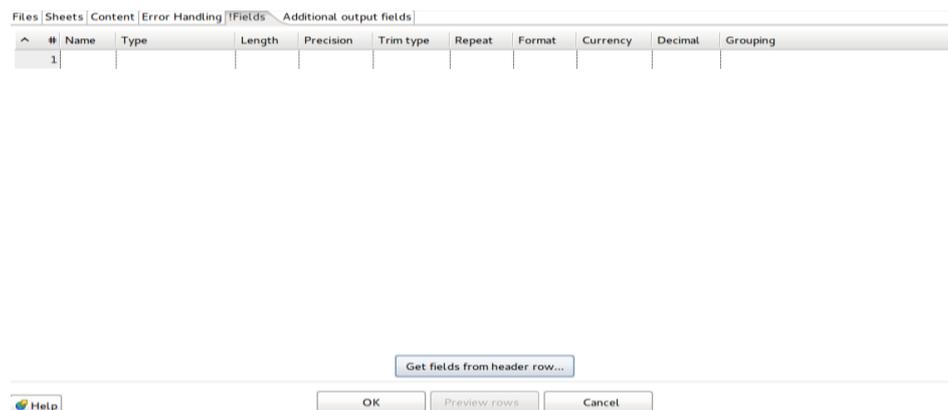
- Aparecen todas las hojas contenidas en el archivo, seleccionamos la que contiene la información requerida y presionamos “ok”.



- Se puede visualizar en nombre de la hoja seleccionada



- En la pestaña !Fields presionamos “Get fields from header row” para obtener la información contenida en el archivo de Excel que corresponde a una hoja.



- Dentro de la hoja está contenida toda esta información, en ocasiones es necesario modificar el tipo de información para que se ajuste a nuestra información.

Files Sheets Content Error Handling Fields Additional output fields

#	Name	Type	Length	Precision	Trim type	Repeat	Format	Currency	Decimal	Grouping
1	ID_PACIENTE	Integer			none	N	#			
2	sexo	String			none	N				
3	edad	Integer			none	N	#			
4	edad_anios	Integer			none	N	#			
5	edad_descripcion	String			none	N				
6	etnia	String			none	N				
7	discapacidad	String			none	N				
8	nacional_extranjero	String			none	N				
9	nacionalidad_paciente	String			none	N				
10	nombre_pais	String			none	N				
11	cod_pais	String			none	N				
12	provincia_residencia	String			none	N				
13	canton_residencia	String			none	N				
14	parroquia_residencia	String			none	N				
15	Causa	String			none	N				

Get fields from header row...

Help OK Preview rows Cancel

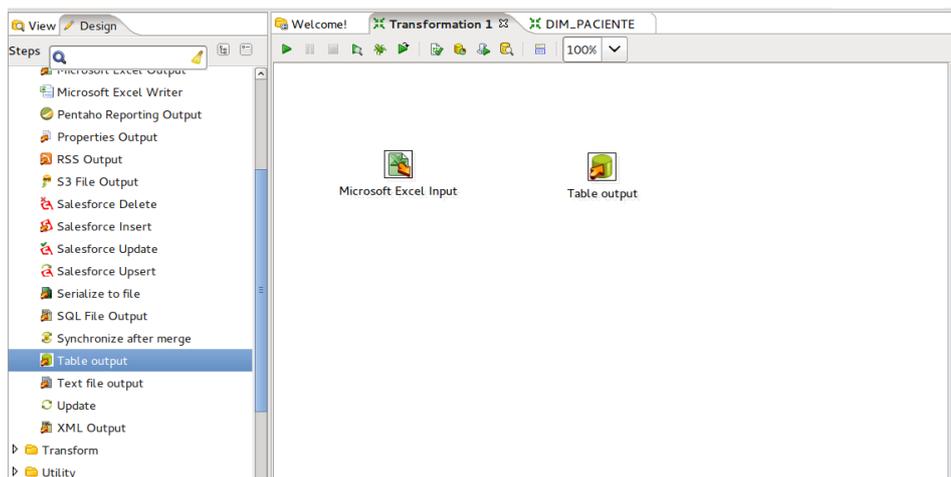
- Al presionar “Preview rows” se puede pre visualizar los datos a cargar y finalmente presionamos “ok”.

Examine preview data

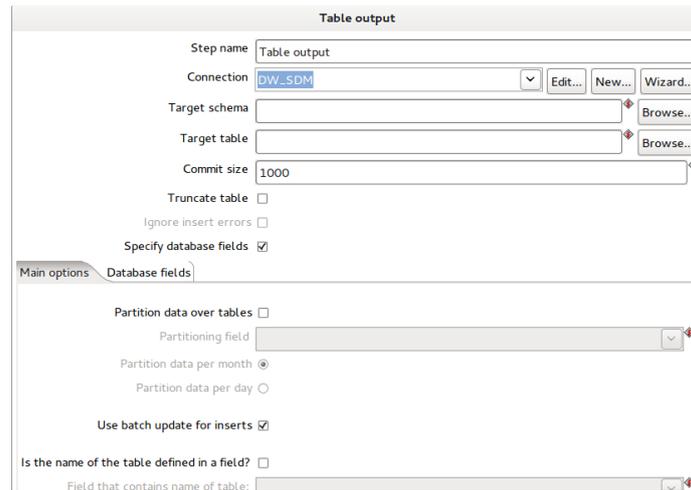
Rows of step: Microsoft Excel Input (1000 rows)

#	ID_PACIENTE	sexo	edad	edad_anios	edad_descripcion	etnia	discapacidad	nacional_extranjero	nacionalidad_p
1	1	Hombre	59	59	Anios (1 a 115 años de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
2	2	Hombre	22	22	Anios (1 a 115 años de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
3	3	Hombre	71	71	Anios (1 a 115 años de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
4	4	Hombre	6	0	Dias (1 a 28 días de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
5	5	Hombre	3	3	Anios (1 a 115 años de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
6	6	Hombre	7	7	Anios (1 a 115 años de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
7	7	Hombre	20	20	Anios (1 a 115 años de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
8	8	Hombre	1	0	Meses (1 a 11 meses de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
9	9	Hombre	64	64	Anios (1 a 115 años de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
10	10	Hombre	53	53	Anios (1 a 115 años de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
11	11	Hombre	53	53	Anios (1 a 115 años de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
12	12	Hombre	3	0	Dias (1 a 28 días de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
13	13	Hombre	91	91	Anios (1 a 115 años de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
14	14	Hombre	1	1	Anios (1 a 115 años de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
15	15	Hombre	8	8	Anios (1 a 115 años de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
16	16	Hombre	22	0	Horas (1 a 23 horas de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
17	17	Hombre	19	19	Anios (1 a 115 años de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
18	18	Hombre	2	0	Dias (1 a 28 días de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR
19	19	Hombre	81	81	Anios (1 a 115 años de edad)	Mestizo/a	Ninguna	Ecuatoriano/a	ECUADOR

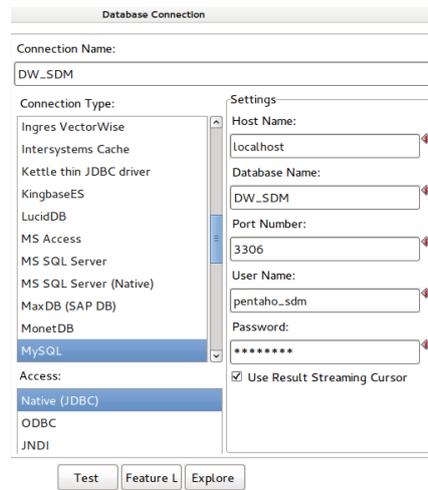
- Para almacenar la información en una base de datos se requiere la herramienta Table output.



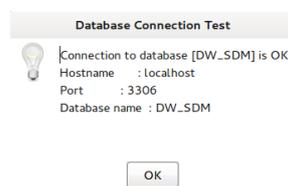
- Se debe establecer una conexión con la base de datos a cargar la información.



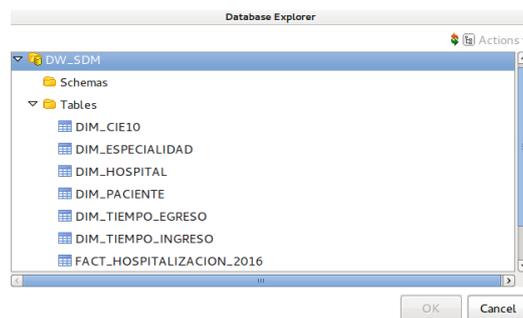
- los parámetros de configuración en este ejemplo son los siguientes



- Test de conexión exitoso.



- En la opción Target Tablet utilizaremos el buscador para seleccionar la tabla correspondiente a la base de datos donde cargaremos la información.



- Para este ejemplo seleccionamos DIM_PACIENTE

Table output

Step name: Table output

Connection: DW_SDM

Target schema: [Empty]

Target table: DIM_PACIENTE

Commit size: 1000

Truncate table:

Ignore insert errors:

Specify database fields:

Main options: Database fields

Partition data over tables:

Partitioning field: [Empty]

Partition data per month:

Partition data per day:

Use batch update for inserts:

Is the name of the table defined in a field?:

Field that contains name of table: [Empty]

- En la pestaña Database fields presionamos “get fields” y seleccionamos según corresponda y presionamos “ok”.

Truncate table:

Ignore insert errors:

Specify database fields:

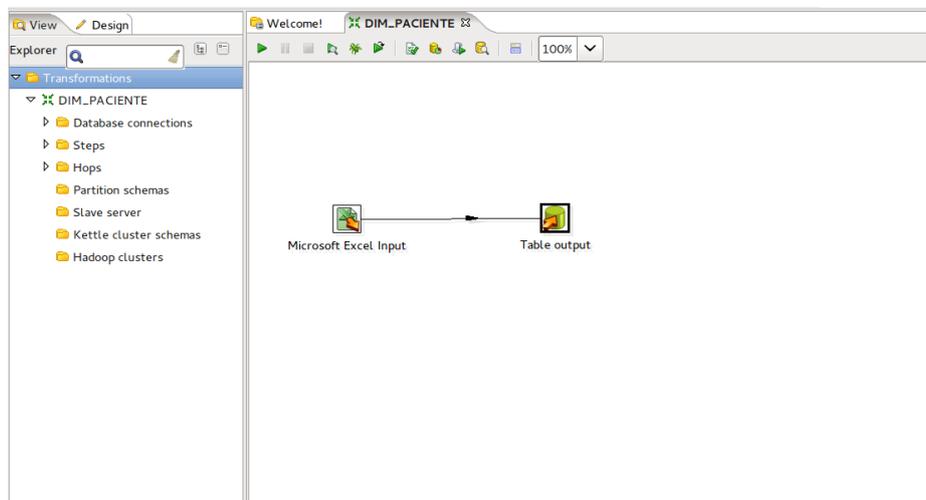
Main options: Database fields

Fields to insert:

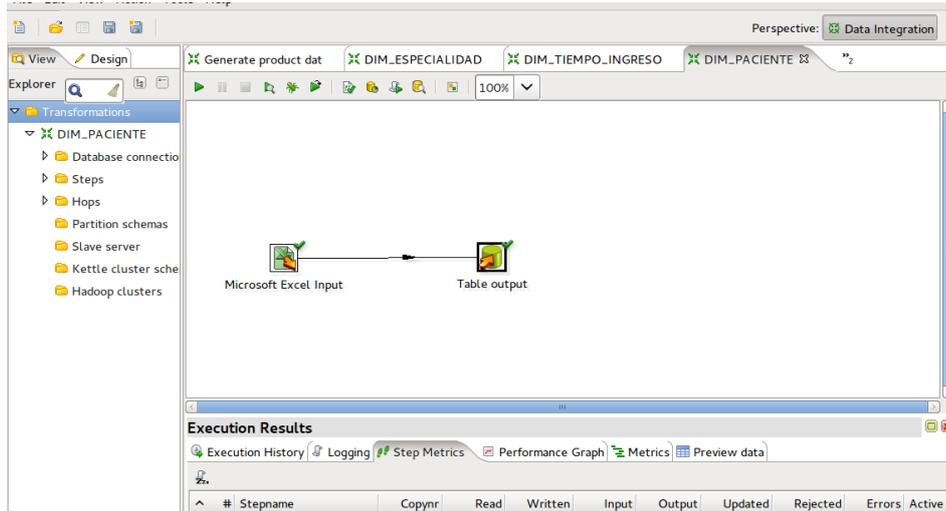
#	Table field	Stream field
3	EDAD	edad
4	EDAD_ANIOS	edad_anios
5	EDAD_DESCR	edad_descripcion
6	ETNIA_PAC...	etnia
7	DISCAPACIA	discapacidad
8	NACIONALID	nacional_extranjero
9	NACIONALID	nacionalidad_paciente
10	PAIS_PAC...	nombre_pais
11	COD_PAIS_P	cod_pais
12	PARROQUIA	provincia_residencia
13	CANTON_PA	canton_residencia
14	PARROQUIA	parroquia_residencia
15	CAUSA	Causa

Buttons: Get fields, Enter field mapping, OK, Cancel, SQL

- Una vez configurados los parámetros, creamos un salto como se muestra en la figura.



- Finalmente ejecutamos la transformación para cargar la información en la base de datos.



- Se verifica que la tabla está cargada en la tabla DIM_PACIENTE

