

Guia Practica Nro 1



Nombre:

Paul Rodríguez

Carrera:

Big Data

Materia:

Marcos de referencia

Ciclo:

M3A

Fecha:

24/04/2024

Periodo:

Abril - Agosto

GUÍA PRÁCTICA

HERRAMIENTAS BÁSICAS DE OBTENCIÓN Y VISUALIZACIÓN DE DATOS

1. Datos Generales

Carrera:	Tecnología Superior en Big Data
Período académico:	Abril – Agosto 2024
Asignatura:	Marcos de Referencia a la Big Data
Unidad N°:	1. Conceptos Básicos

Tema:	Herramientas básicas de obtención y visualización de datos
Ciclo-Paralelo:	M3A
Fecha de inicio de la Unidad:	15/04/2024
Fecha de fin de la Unidad	13/05/2024
Práctica N°:	1
Horas:	9
Docente:	Mgtr. Verónica Paulina Chimbo Coronel

1. Contenido

a. Fundamentos

La Extracción, Transformación y Carga (ETL) de datos es un proceso esencial en el mundo de la gestión de datos, permitiendo la migración de información desde diversas fuentes hacia un sistema de almacenamiento centralizado. En esta guía, exploraremos cómo llevar a cabo un proceso ETL específico utilizando Pentaho Data Integration (también conocido como Kettle) para extraer datos de un archivo Excel que contiene información sobre taxis y cargarlos en una base de datos MySQL.

1. Objetivos de la Guía

- El objetivo principal de esta guía es proporcionar una visión detallada y paso a paso del proceso ETL, desde la configuración de Pentaho Data Integration hasta la carga efectiva de datos en una base de datos MySQL. En particular, nos enfocaremos en la conversión de datos provenientes de un archivo Excel, donde cada columna representa una categoría específica de información sobre los taxis.

1. Evaluación del Aprendizaje

Rúbrica de Evaluación de la Guía Práctica	Criterios de Evaluación	Puntuación
Máxima	Base de Datos	3/3
	Extracción de datos	1.5/3
	Limpieza de datos	0,5/1
	Llenado de Datos	0,5/1
	Script	1,5/2
	Puntuación Total	10

1. Preparación previa, materiales, herramientas, equipos y software

- Pentaho Data Integration: Asegúrate de tener Pentaho Data Integration instalado en tu sistema.

- Base de Datos MySQL: Contar con una base de datos MySQL disponible para la carga de datos. Si aún no tienes MySQL instalado.
- Conocimientos Básicos de SQL: Será útil tener un entendimiento básico de SQL para configurar la conexión a la base de datos y realizar ajustes según sea necesario.

1. Procedimientos a emplear

En esta guía, abordaremos el proceso ETL para procesar datos de las Cooperativas de Taxis de Cuenca relacionados con la flota de taxis. Estos datos contienen información clave sobre los taxis, incluyendo detalles como la placa, clase, marca, año de fabricación, categoría y tipo de taxi. El objetivo es realizar la extracción, transformación y carga de estos datos en una base de datos MySQL.

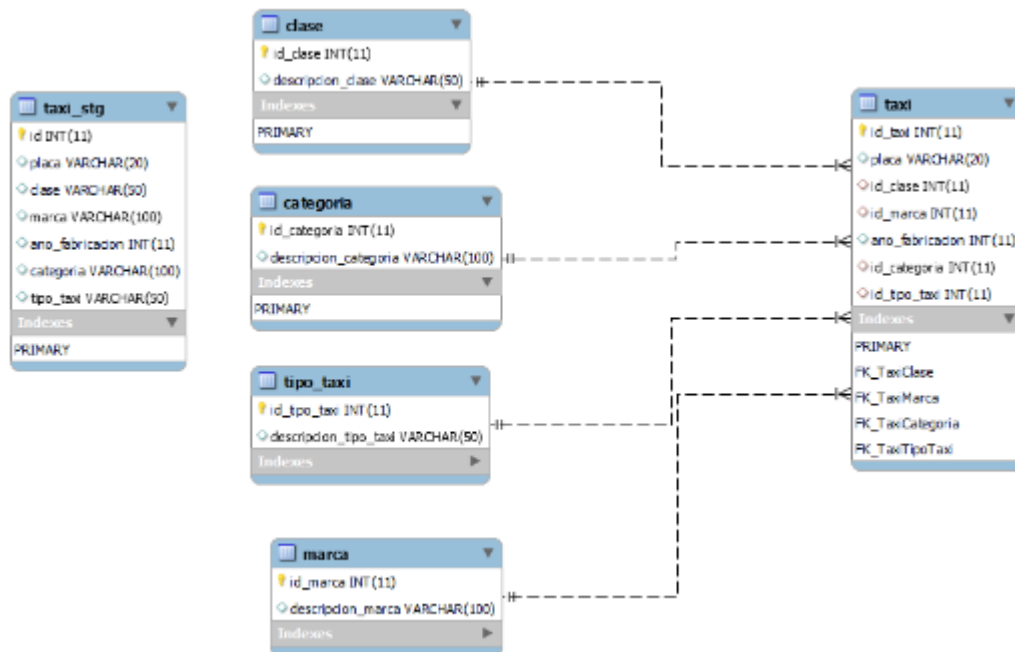
1. Descargarse los datos del siguiente repositorio:

https://docs.google.com/spreadsheets/d/1qDOA-ubd-2BNzR5qsxpa8z5pZ_JHKi3f/edit?usp=sharing&ouid=101015197702841688612&rtpof=true&sd=true

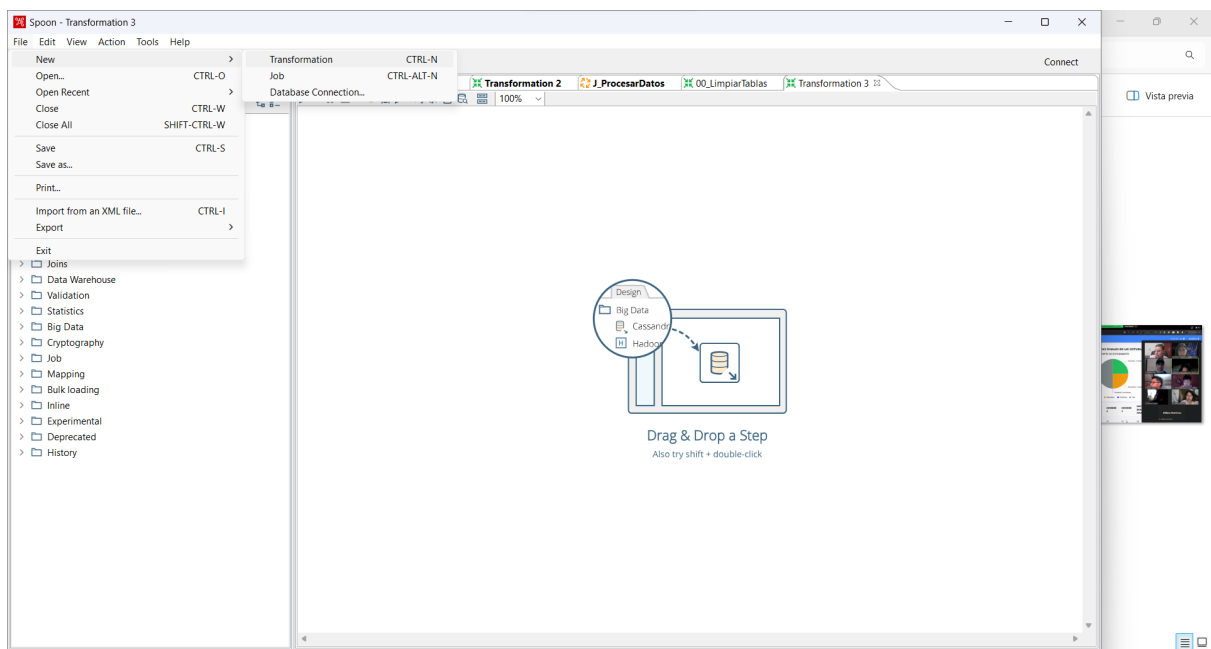
y colocar en una carpeta denominada Practicas donde se desarrollará la Practica.

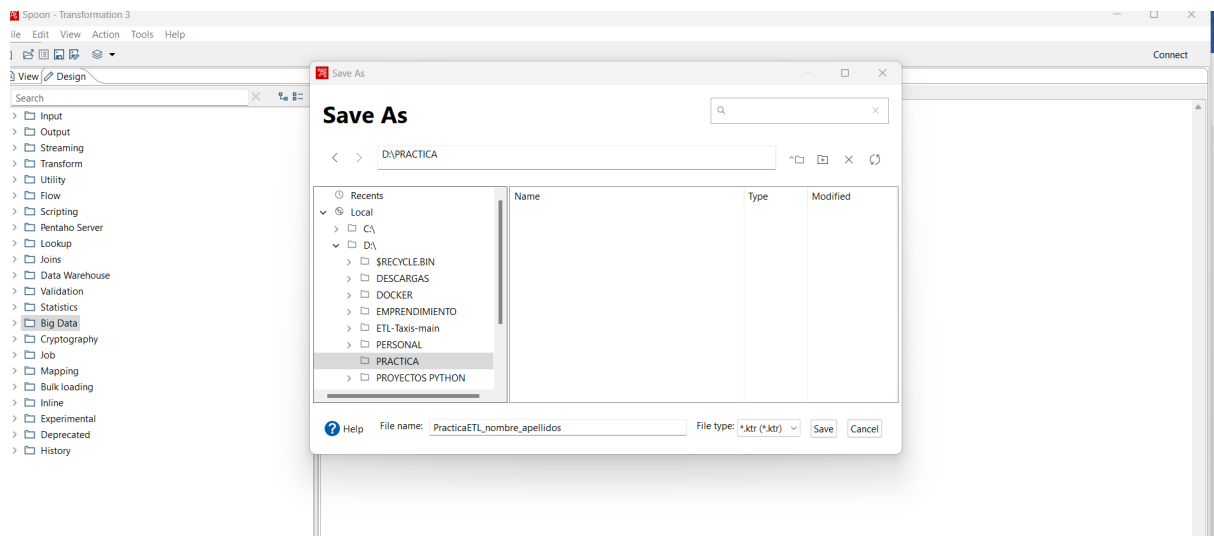
	A	B	C	D	E	F
1	Placa	Clase	Marca	Año	Categoría	Tipo taxi
2	AP-000024	Taxi Ejecutivo	Toyota	1990	Automovil	Sedan
3	AP-000033	Taxi Ejecutivo	Nissan	1990	Automovil	Sedan
4	AP-000035	Taxi Ejecutivo	Nissan	1990	Automovil	Sedan
5	AP-000049	Taxi Ejecutivo	Volvo	1990	Automovil	Sedan
6	AP-000074	Taxi Ejecutivo	Nissan	1990	Automovil	Sedan
7	AP-000080	Taxi Ejecutivo	Toyota	1989	Automovil	Noid
8	AP-000119	Taxi Ejecutivo	Toyota	1987	Automovil	Sedan
9	AP-000136	Taxi Ejecutivo	Nissan	1990	Automovil	Sedan
10	AP-000145	Taxi Ejecutivo	Toyota	1990	Automovil	Sedan
11	AP-000158	Taxi Ejecutivo	Lada	1985	Categoría no registrada	Sedan
12	AP-000160	Taxi Ejecutivo	Toyota	1990	Automovil	Noid
13	AP-000190	Taxi Ejecutivo	Zil	1980	Categoría no registrada	Noid
14	AP-000201	Taxi Ejecutivo	Toyota	1990	Automovil	Sedan
15	AP-000213	Taxi Ejecutivo	Toyota	1990	Automovil	Noid
16	AP-000216	Taxi Ejecutivo	Toyota	1990	Automovil	Sedan
17	AP-000229	Taxi Ejecutivo	Toyota	1986	Automovil	Noid

1. Crear la base de datos denominada bd_taxis_cuenca con la siguiente estructura

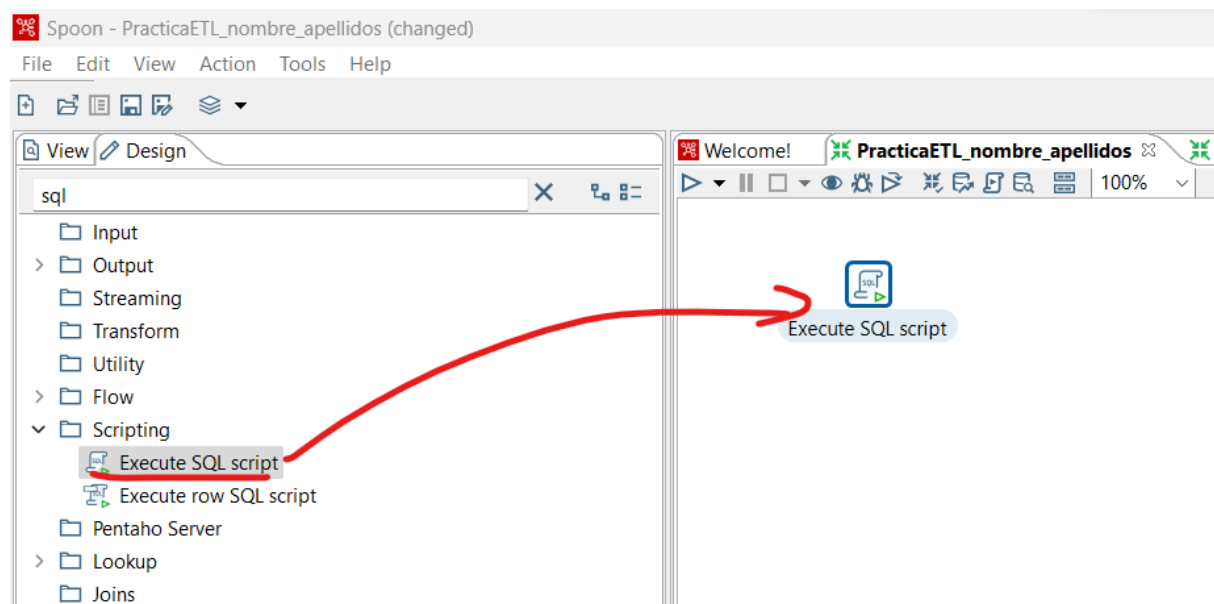


1. Crear un proyecto nuevo para ello ir a file, nuevo, transformación con el nombre PracticaETL_nombres_apellidos

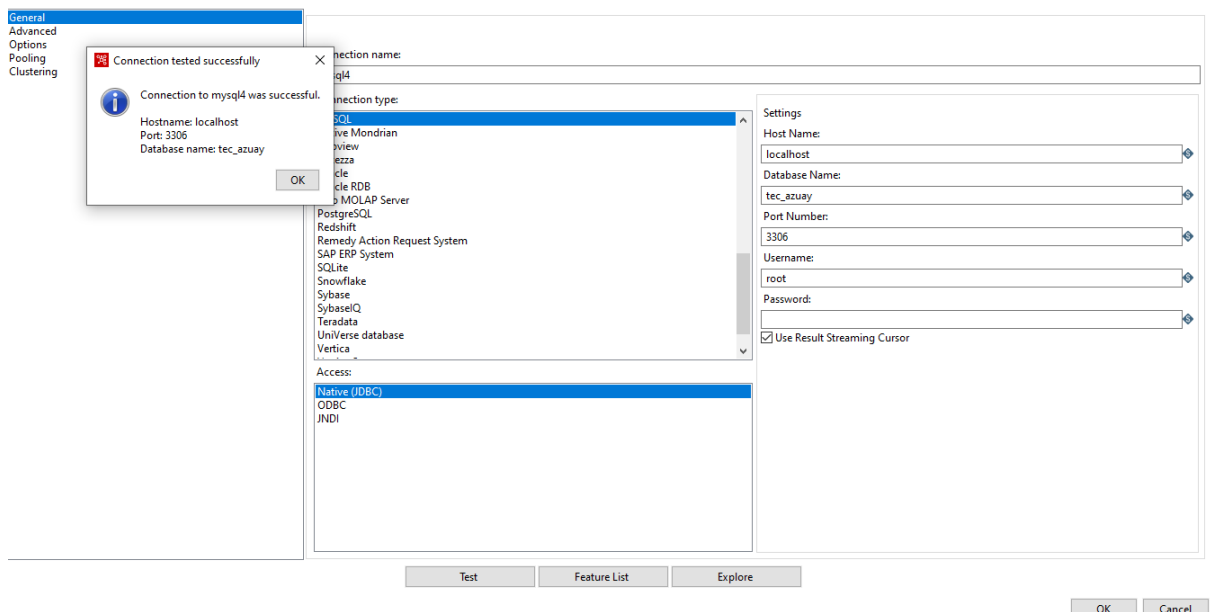
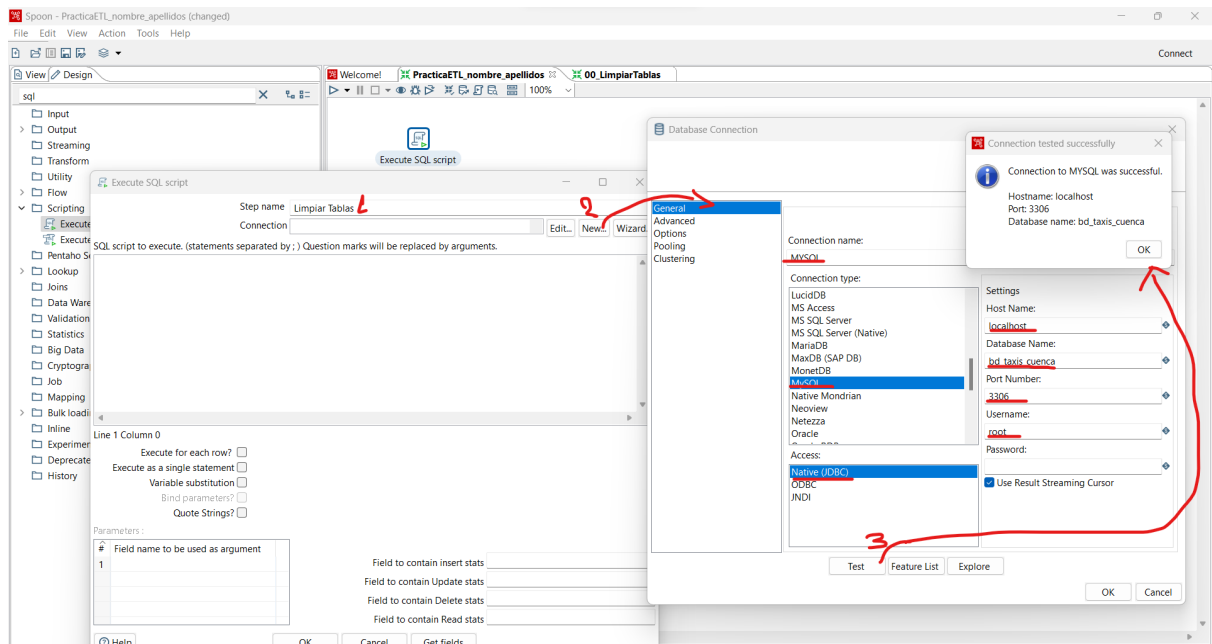




3.1. Agregar al área de trabajo el elemento Execute SQL script de la categoría Scripting para limpiar los datos de la base de datos.



Configurar el **Execute SQL script** dando doble click sobre el icono.



1. Renombrar el Execute con Limpiar Tablas
2. Crear la conexión a la base de datos y asegurarse de llenar todos los campos requeridos.
3. Realizar un test de conexión
4. Realizar las sentencias sql para eliminar los datos de las tablas y asignarles el **AUTO_INCREMENT** en 1

Execute SQL script

Step name:

Connection:

SQL script to execute. (statements separated by ;) Question marks will be replaced by arguments.

```

DELETE from taxi_stg;
ALTER TABLE taxi_stg AUTO_INCREMENT = 1;

DELETE from taxi;
ALTER TABLE taxi AUTO_INCREMENT = 1;

DELETE from clase;
ALTER TABLE clase AUTO_INCREMENT = 1;

DELETE from categoria;
ALTER TABLE categoria AUTO_INCREMENT = 1;

DELETE from marca;
ALTER TABLE marca AUTO_INCREMENT = 1;

DELETE from tipo_taxi;
ALTER TABLE tipo_taxi AUTO_INCREMENT = 1;

```

Line 17 Column 41

Execute for each row? ☐

Execute as a single statement ☐

Variable substitution ☐

Bind parameters? ☐

Quote Strings? ☐

Parameters :

#	Field name to be used as argument
1	

Field to contain insert stats

Field to contain Update stats

Field to contain Delete stats

Field to contain Read stats

Execute SQL script

Step name:

Connection:

SQL script to execute. (statements separated by ;) Question marks will be replaced by arguments.

```

DELETE FROM matricula_stg;
ALTER TABLE matricula_stg AUTO_INCREMENT = 1;

DELETE FROM matricula;
ALTER TABLE matricula AUTO_INCREMENT = 1;

DELETE FROM area;
ALTER TABLE area AUTO_INCREMENT = 1;

DELETE FROM campus;
ALTER TABLE campus AUTO_INCREMENT = 1;

DELETE FROM canton;
ALTER TABLE canton AUTO_INCREMENT = 1;

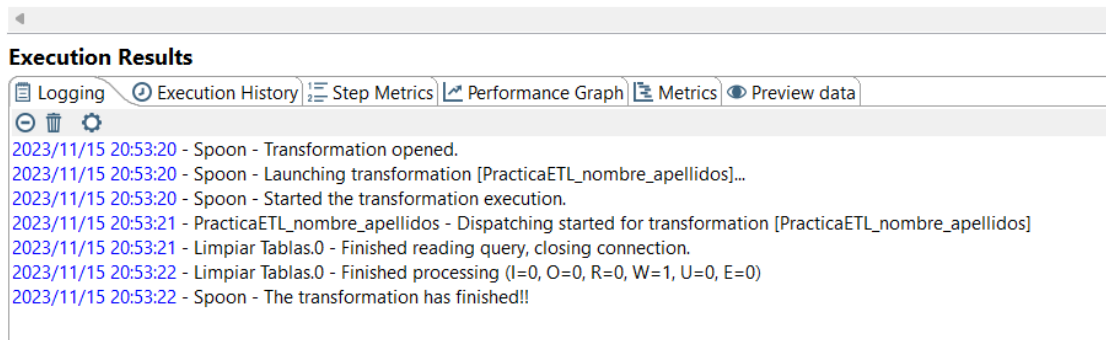
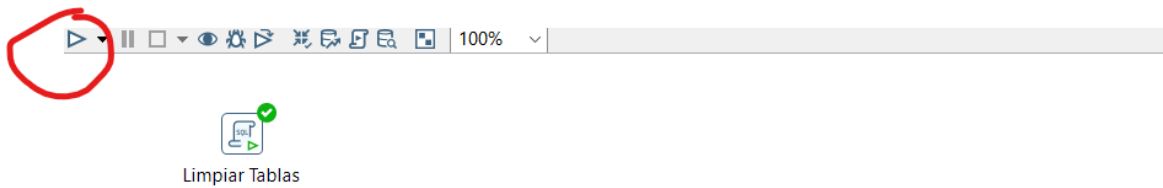
DELETE FROM carrera;
ALTER TABLE carrera AUTO_INCREMENT = 1;

DELETE FROM duracion;
ALTER TABLE duracion AUTO_INCREMENT = 1;

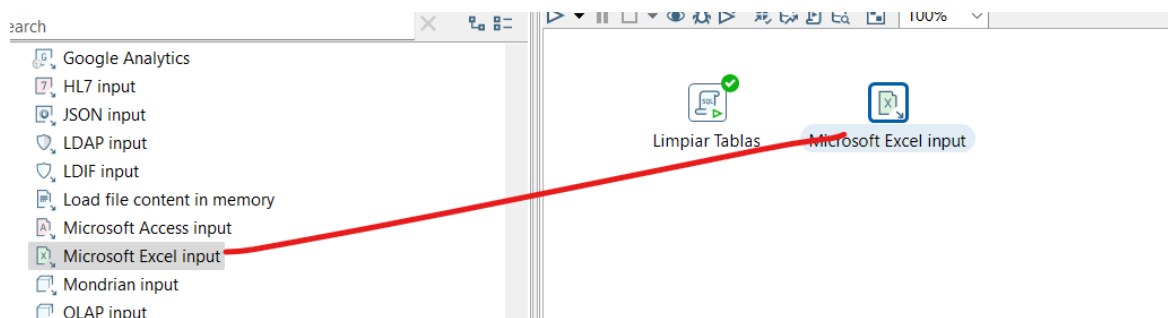
DELETE FROM institucion;

```

1. Comprobar que funcione correctamente la configuración dar click en ejecutar y verificar que se marque en color verde.



3.2. Cargar los datos del Excel Datos_taxi_Cuenca.xls para ello seleccionaremos de la categoría input e elemento Microsoft Excel input



1.- Dar click derecho sobre el icono y realizar las configuraciones, asignarle el nombre de Extracción de Datos del Dataset.

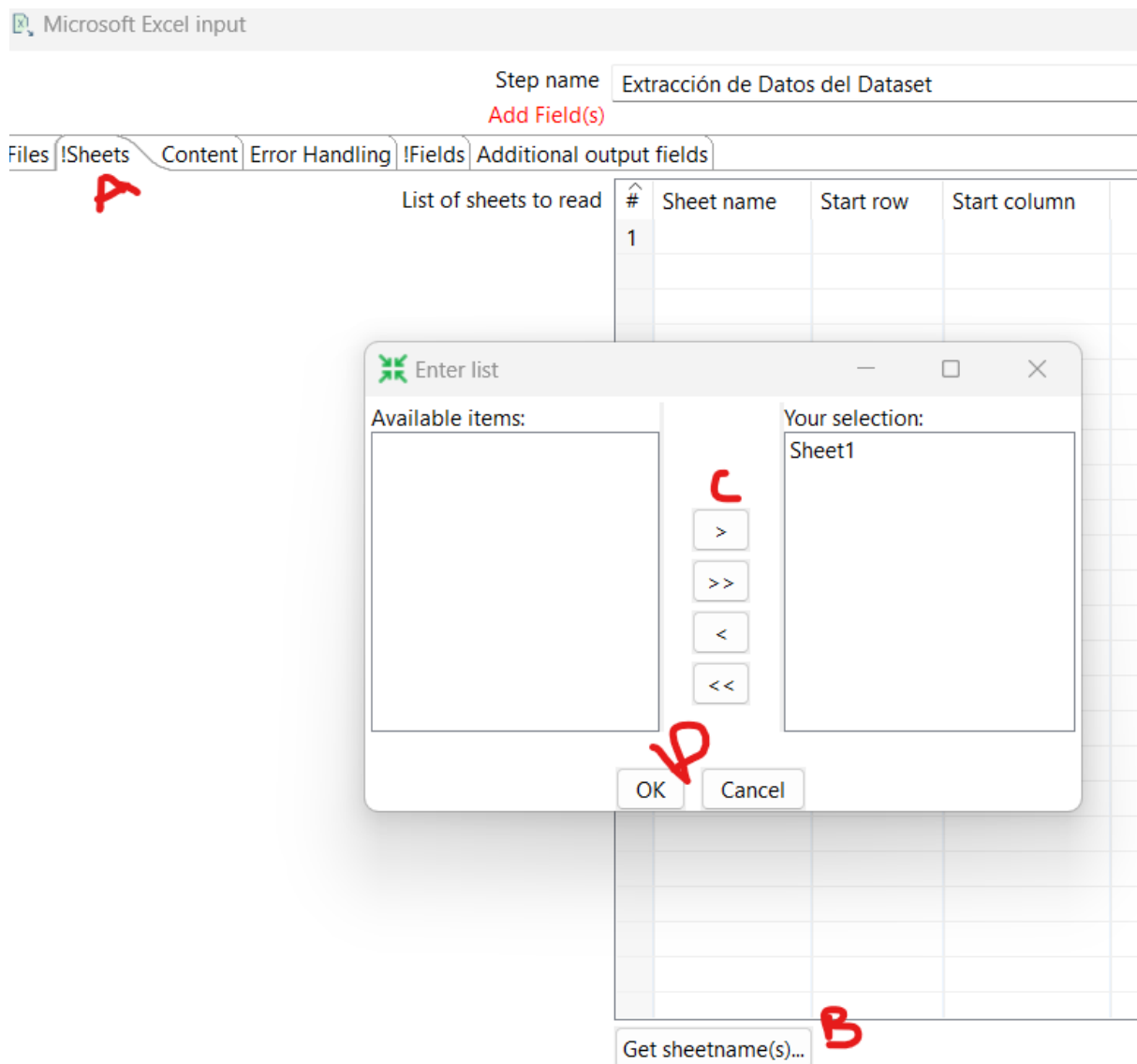
2.- Buscar el archivo de Excel de la data

3.- Agregar al panel de selección de files

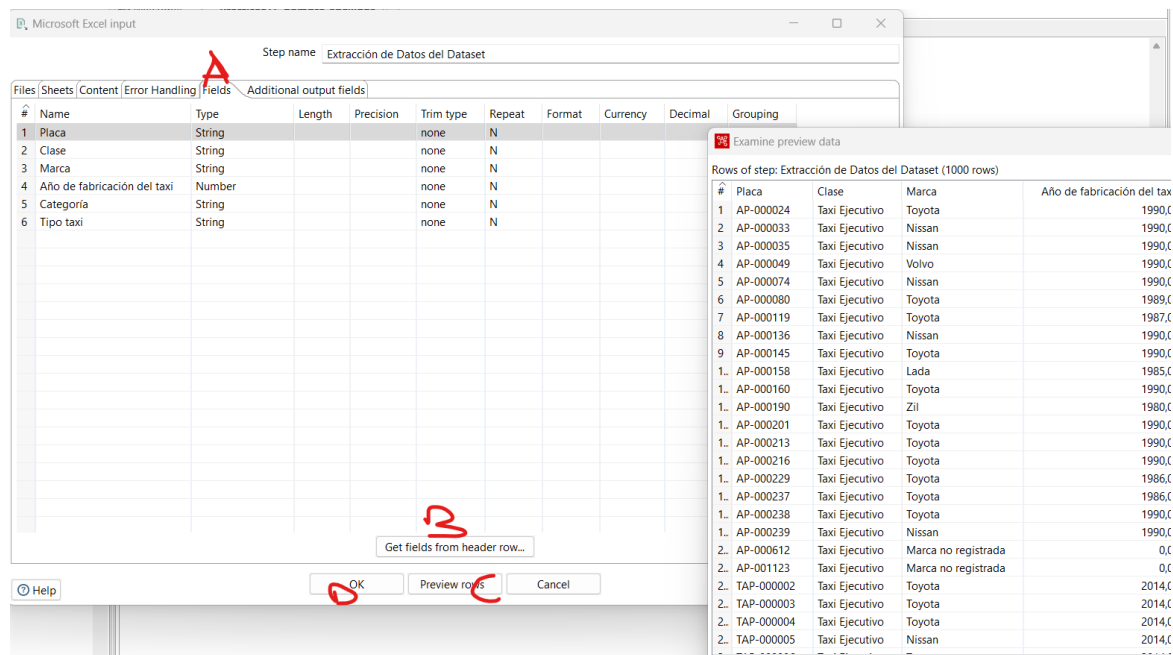
The screenshot shows the 'Microsoft Excel input' dialog box. The 'Step name' field is 'Extracción de Datos del Dataset' (annotated with a red '1'). The 'Add Field(s)' button is highlighted with a red '5'. The 'Files' tab is selected (annotated with a red '4'). The 'Spread sheet type (engine)' is 'Excel 97-2003 XLS (JXL)' (annotated with a red '5'). The 'File or directory' field is empty, with a 'Browse...' button next to it (annotated with a red '2'). The 'Regular Expression' and 'Exclude Regular Expression' fields are empty. The 'Password' field is empty. The 'Selected files' table has one row with the file 'D:/PRACTICA/Datos_Taxi_Cuenca.xls' (annotated with a red '3'). The 'Accept filenames from previous steps' section has a checkbox for 'Accept filenames from previous step' and two dropdown menus for 'Step to read filenames from' and 'Field in the input to use as filename'. The 'Show filename(s)...' button is at the bottom. The 'OK', 'Preview rows', and 'Cancel' buttons are at the bottom right.

#	File/Directory	Wildcard (RegExp)	Exclude wildcard	Required	Inc
1	D:/PRACTICA/Datos_Taxi_Cuenca.xls				N

1. Agregar los sheets



5.- Agregar los fields



6.- Conectar y probar

PracticaETL_nombre_apellidos

100%

Limpiar Tablas → Extracción de Datos del Dataset

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2023/11/15 21:30:58 - Spoon - Transformation opened.

2023/11/15 21:30:58 - Spoon - Launching transformation [PracticaETL_nombre_apellidos]...

2023/11/15 21:30:58 - Spoon - Started the transformation execution.

2023/11/15 21:30:58 - PracticaETL_nombre_apellidos - Dispatching started for transformation [PracticaETL_nombre_apellidos]

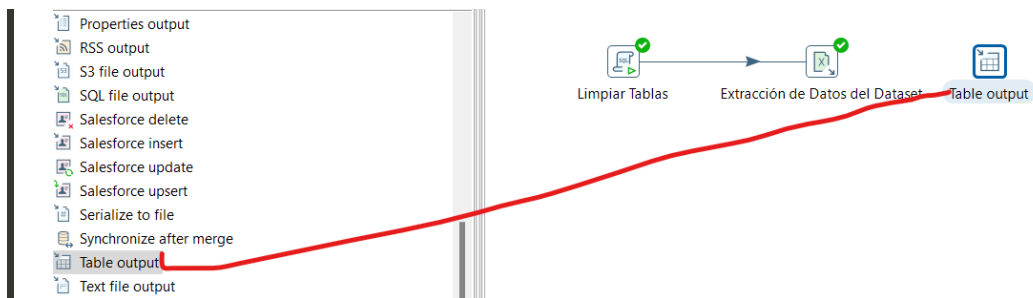
2023/11/15 21:30:58 - Limpiar Tablas.0 - Finished reading query, closing connection.

2023/11/15 21:30:58 - Limpiar Tablas.0 - Finished processing (I=0, O=0, R=0, W=1, U=0, E=0)

2023/11/15 21:30:58 - Extracción de Datos del Dataset.0 - Finished processing (I=11705, O=0, R=0, W=11705, U=0, E=0)

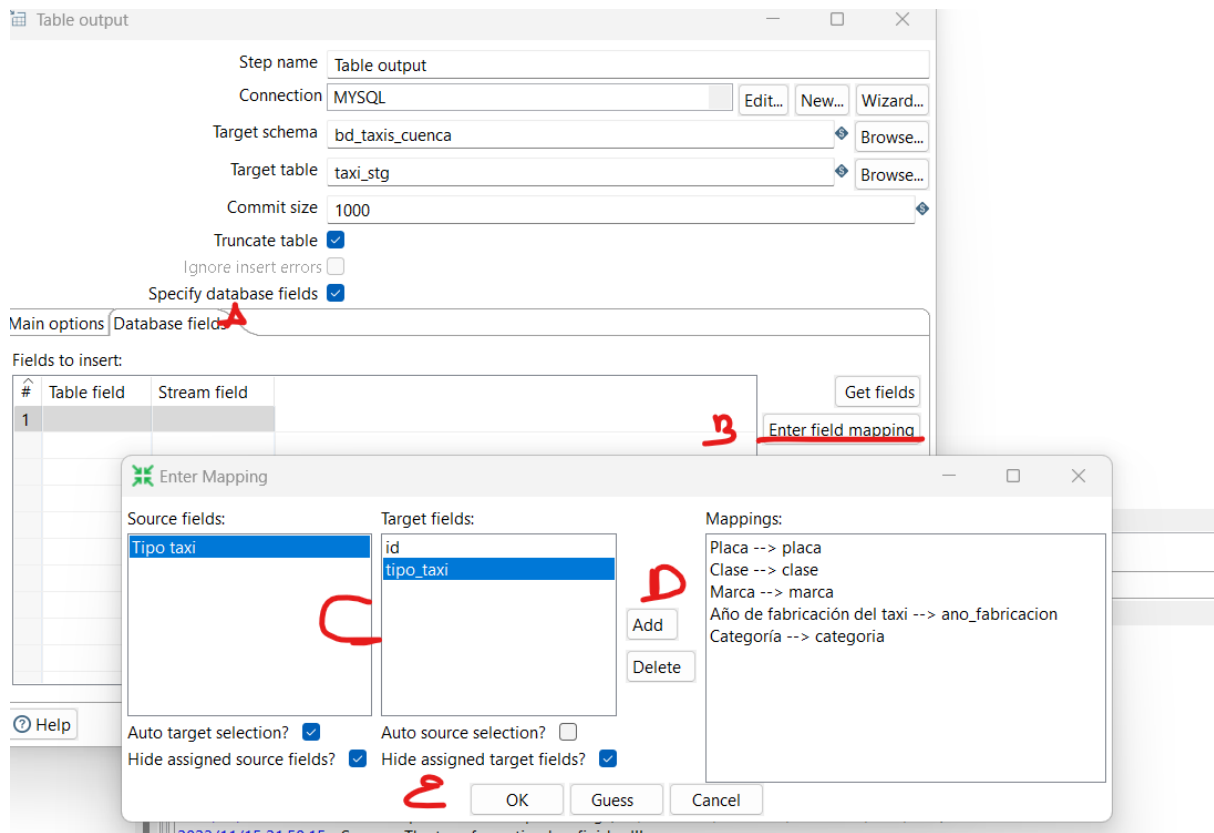
2023/11/15 21:30:58 - Spoon - The transformation has finished!!

3.3. Pasar los datos cargados del Excel a la tabla taxi_stg para ello utilizaremos el elemento Table output de la categoría Output.



1. Dar doble click y configurar los datos de la conexión, base de datos y la tabla donde se guardrán los datos.

1. Seleccionar los fields y relacionar



1. Una vez relacionada los fieds de la data con los de la tabla. Finalmente presionar ok

Table output

Step name: Table output

Connection: MYSQL [Edit... New... Wizard...]

Target schema: bd_taxis_cuenca [Browse...]

Target table: taxi_stg [Browse...]

Commit size: 1000

Truncate table: ☒

Ignore insert errors: ☐

Specify database fields: ☒

Main options Database fields

Fields to insert:

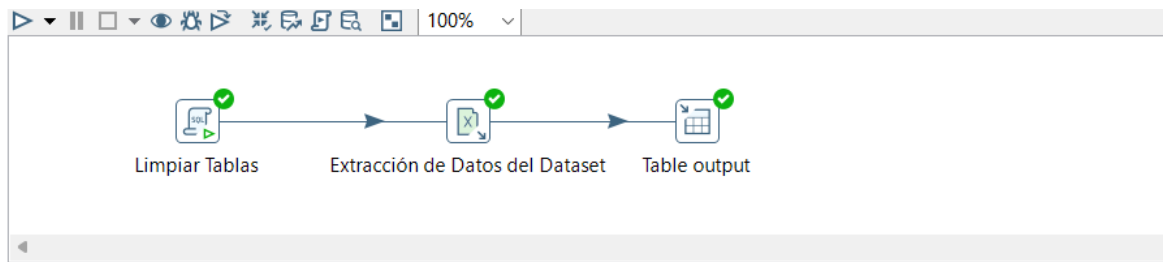
#	Table field	Stream field
1	placa	Placa
2	clase	Clase
3	marca	Marca
4	ano_fabricacion	Año de fabricación del taxi
5	categoria	Categoría
6	tipo_taxi	Tipo taxi

[Get fields]

[Enter field mapping]

[? Help] [OK] [Cancel] [SQL]

1. Comprobar enlazando los elementos y ejecutando para comprobar que el trabajo es exitoso se debe de hacer un select a la tabla en la herramienta visual de la base de datos en nuestro caso estamos utilizando MysqlWorkbench.



Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2023/11/15 21:57:10 - Spoon - Transformation opened.
 2023/11/15 21:57:10 - Spoon - Launching transformation [PracticaETL_nombre_apellidos]...
 2023/11/15 21:57:10 - Spoon - Started the transformation execution.
 2023/11/15 21:57:10 - PracticaETL_nombre_apellidos - Dispatching started for transformation [PracticaETL_nombre_apellidos]
 2023/11/15 21:57:10 - Table output.0 - Connected to database [MYSQL] (commit=1000)
 2023/11/15 21:57:10 - Limpiar Tablas.0 - Finished reading query, closing connection.
 2023/11/15 21:57:10 - Limpiar Tablas.0 - Finished processing (I=0, O=0, R=0, W=1, U=0, E=0)
 2023/11/15 21:57:10 - Extracción de Datos del Dataset.0 - Finished processing (I=11705, O=0, R=0, W=11705, U=0, E=0)
 2023/11/15 21:57:12 - Table output.0 - Finished processing (I=0, O=11705, R=11705, W=11705, U=0, E=0)
 2023/11/15 21:57:12 - Spoon - The transformation has finished!!



Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

!024/05/06 17:32:04 - Spoon - Launching transformation [data_transformada_exe]...
 !024/05/06 17:32:04 - Spoon - Started the transformation execution.
 !024/05/06 17:32:05 - data_transformada_exe - Dispatching started for transformation [data_transformada_exe]
 !024/05/06 17:32:05 - Table output.0 - Connected to database [mysql4] (commit=1000)
 !024/05/06 17:32:07 - Execute SQL script.0 - Finished reading query, closing connection.
 !024/05/06 17:32:07 - Execute SQL script.0 - Finished processing (I=0, O=0, R=0, W=1, U=0, E=0)
 !024/05/06 17:32:09 - Microsoft Excel input.0 - Finished processing (I=94, O=0, R=0, W=94, U=0, E=0)
 !024/05/06 17:32:09 - Table output.0 - Finished processing (I=0, O=94, R=94, W=94, U=0, E=0)
 !024/05/06 17:32:09 - Spoon - The transformation has finished!!

Verificación en la base de datos

Navigator: taxi_stg

SCHEMAS

Filter objects

- bd_proyecto_1
 - bd_taxis_cuenca
 - Tables
 - categoria
 - clase
 - marca
 - taxi
 - taxi_stg
 - tipo_taxi
 - Views
 - Stored Procedures
 - Functions
 - bdcalificaciones
 - bdfacturacion
 - bdpizza
 - bdpollos
 - control_empleados
 - datos
 - datos_taxi
 - db-project-management-s

Administration Schemas

Information

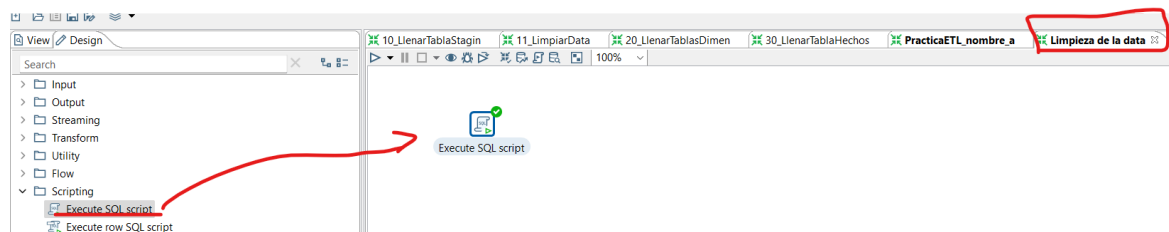
Limit to 1000 rows

1 • SELECT * FROM bd_taxis_cuenca.taxi_stg;

Result Grid

	id	placa	clase	marca	ano_fabricacion	categoria	tipo_taxi
1	AP-000024		Taxi Ejecutivo	Toyota	1990	Automovil	Sedan
2	AP-000033		Taxi Ejecutivo	Nissan	1990	Automovil	Sedan
3	AP-000035		Taxi Ejecutivo	Nissan	1990	Automovil	Sedan
4	AP-000049		Taxi Ejecutivo	Volvo	1990	Automovil	Sedan
5	AP-000074		Taxi Ejecutivo	Nissan	1990	Automovil	Sedan
6	AP-000080		Taxi Ejecutivo	Toyota	1989	Automovil	Noid
7	AP-000119		Taxi Ejecutivo	Toyota	1987	Automovil	Sedan
8	AP-000136		Taxi Ejecutivo	Nissan	1990	Automovil	Sedan
9	AP-000145		Taxi Ejecutivo	Toyota	1990	Automovil	Sedan
10	AP-000158		Taxi Ejecutivo	Lada	1985	Categoria no registrada	Sedan

3.4. Limpiar información inválida para ello se debe de crear un archivo nuevo con nombre Limpieza de la data y seleccionar el elemento Execute SQL script de la categoría Scripting.



Configurar el elemento para ello dar doble click renombrar el elemento, crear la conexión a la base de datos y agregar las sentencias sql donde se elimine los valores inválidos como: donde el año de fabricación sea 0, la marca tenga como valor Marca no registrada, en categoría Categoría no registrada y en tipo de taxi Noid.

Execute SQL script

Step name: Limpiar información inválida **A**

Connection: MYSQL **B** Edit... New... Wizard...

SQL script to execute. (statements separated by ;) Question marks will be replaced by arguments.

```
delete from taxi_stg where ano_fabricacion = 0;
delete from taxi_stg where marca like '%Marca%no%registrada%';
delete from taxi_stg where categoria like '%Categoria%no%registrada%';
delete from taxi_stg where tipo_taxi = 'Noid';
```

Line 5 Column 0

Execute for each row? ☐

Execute as a single statement ☐

Variable substitution ☐

Bind parameters? ☐

Quote Strings? ☐

Parameters :

#	Field name to be used as argument
1	

Field to contain insert stats

Field to contain Update stats


Field to contain Delete stats

Field to contain Read stats

Help OK Cancel Get fields

Ejecutar la tarea.

100%

 Limpiar información inválida

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2023/11/16 12:36:30 - Spoon - Transformation opened.

2023/11/16 12:36:30 - Spoon - Launching transformation [Limpieza de la data]...

2023/11/16 12:36:30 - Spoon - Started the transformation execution.

2023/11/16 12:36:30 - Limpieza de la data - Dispatching started for transformation [Limpieza de la data]

2023/11/16 12:36:30 - Limpiar información inválida.0 - Finished reading query, closing connection.

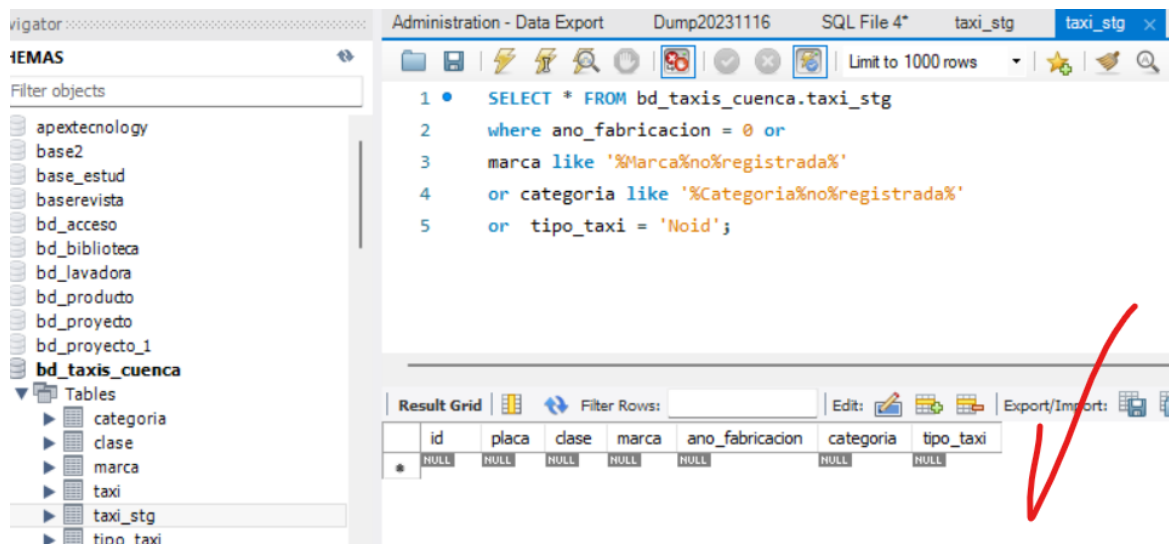
2023/11/16 12:36:30 - Limpiar información inválida.0 - Finished processing (I=0, O=0, R=0, W=1, U=0, E=0)

2023/11/16 12:36:30 - Spoon - The transformation has finished!!

Para verificar realizar lo siguiente en el gestor de base de datos ejecutar la siguiente sentencia:

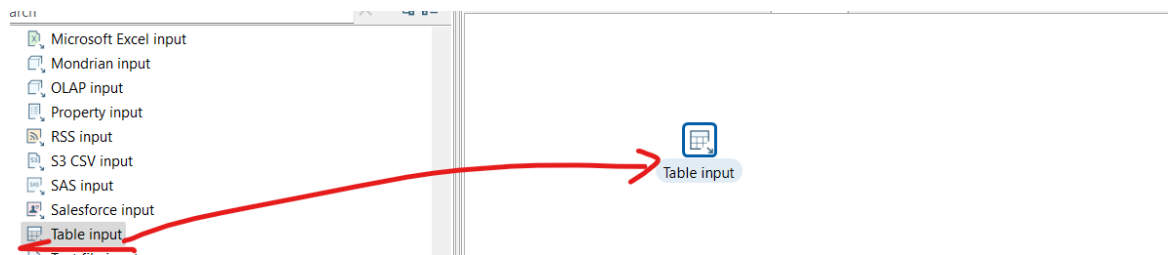
```
SELECT * FROM bd_taxis_cuenca.taxi_stg  
where ano_fabricacion = 0 or  
marca like '%Marca%no%registrada%'  
or categoria like '%Categoria%no%registrada%'  
or tipo_taxi = 'Noid';
```

Si se ejecuto correctamente debe de salir

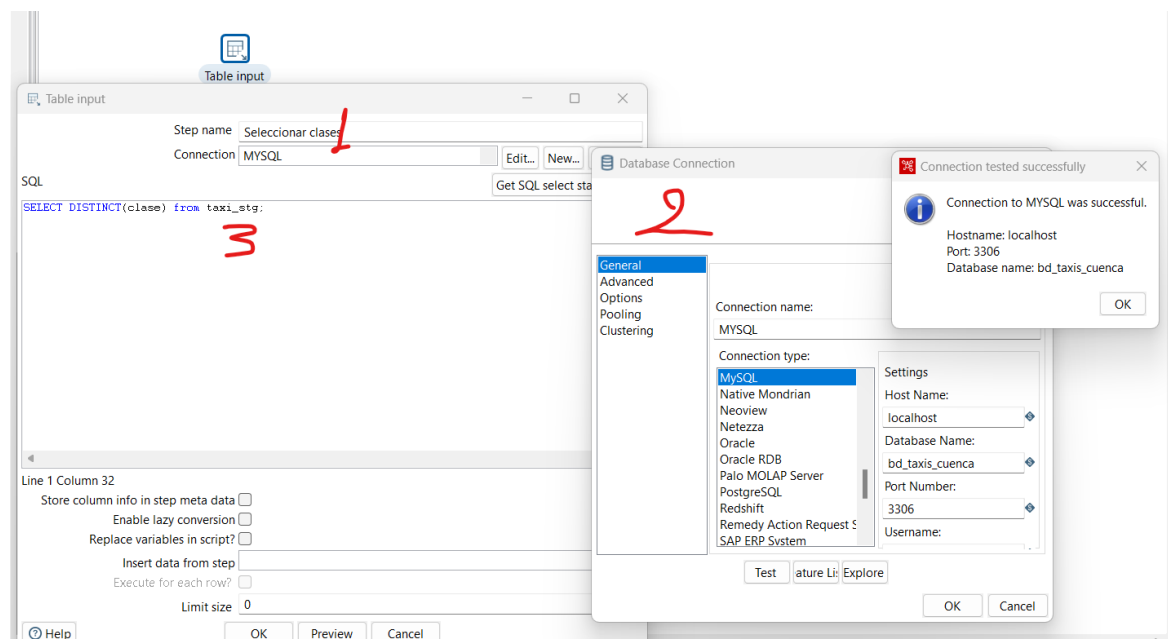


3.5. Llenar los datos en las tablas de la base de datos para ello crear un nuevo archivo con el nombre Llenar tablas de la base de datos y realizar los siguientes pasos:

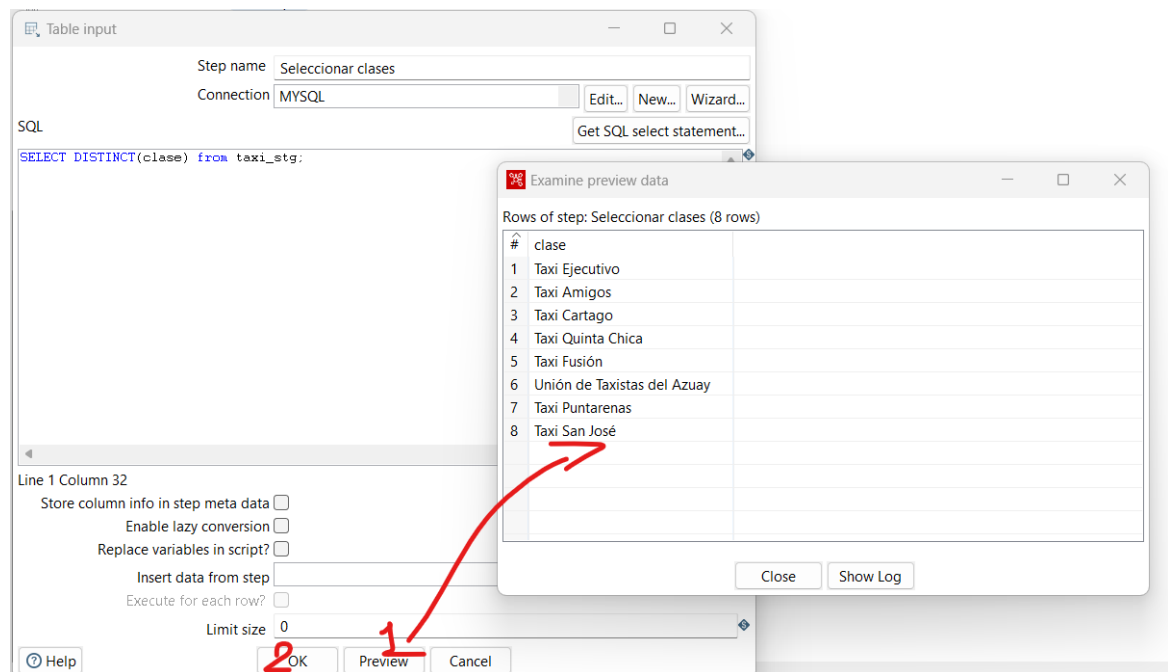
Agregar un elemento Table input de la categoría input al área de trabajo



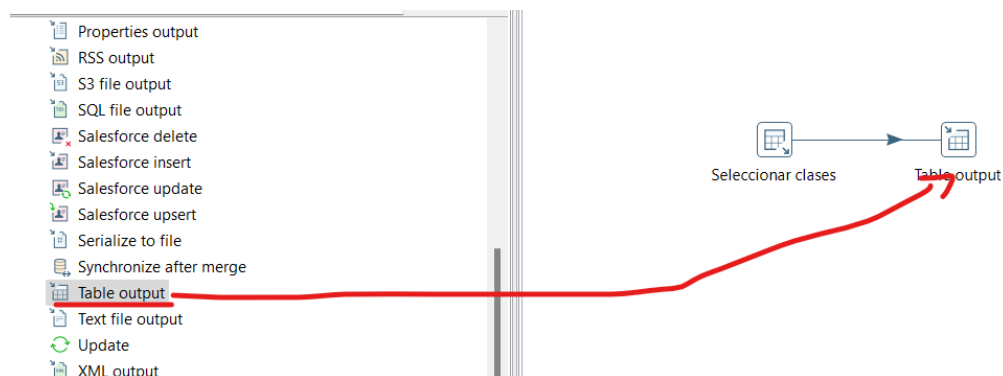
Dar doble click sobre el elemento, renombrar con el nombre Seleccionar clases, verificar la conexión a la base de datos y agregar la siguiente sentencia sql.



Verificar que funcione correctamente que funcione la sentencia sql para seleccionar los valore únicos de clases de taxis.



Luego crear las clases en la tabla clase para ello agregar al área de trabajo un Table output y conectar al elemento anterior.



Configurar el Table output llenar los campos requeridos renombrar con el nombre Llenar clases y verificar los demás campos así como la conexión a la base de datos.

Table output

Step name: Llenar clases ^L

Connection: MYSQL ² Edit... New... Wizard...

Target schema: bd_taxi_cuenca ³ Browse...

Target table: clase ⁴ Browse...

Commit size: 1000

Truncate table: ☒ ⁵

Ignore insert errors: ☐

Specify database fields: ☒

Main options Database fields

Partition data over tables: ☐

Partitioning field:

Partition data per month: ☒

Partition data per day: ☐

Use batch update for inserts: ☒

Is the name of the table defined: ☐

Field that contains name of table:

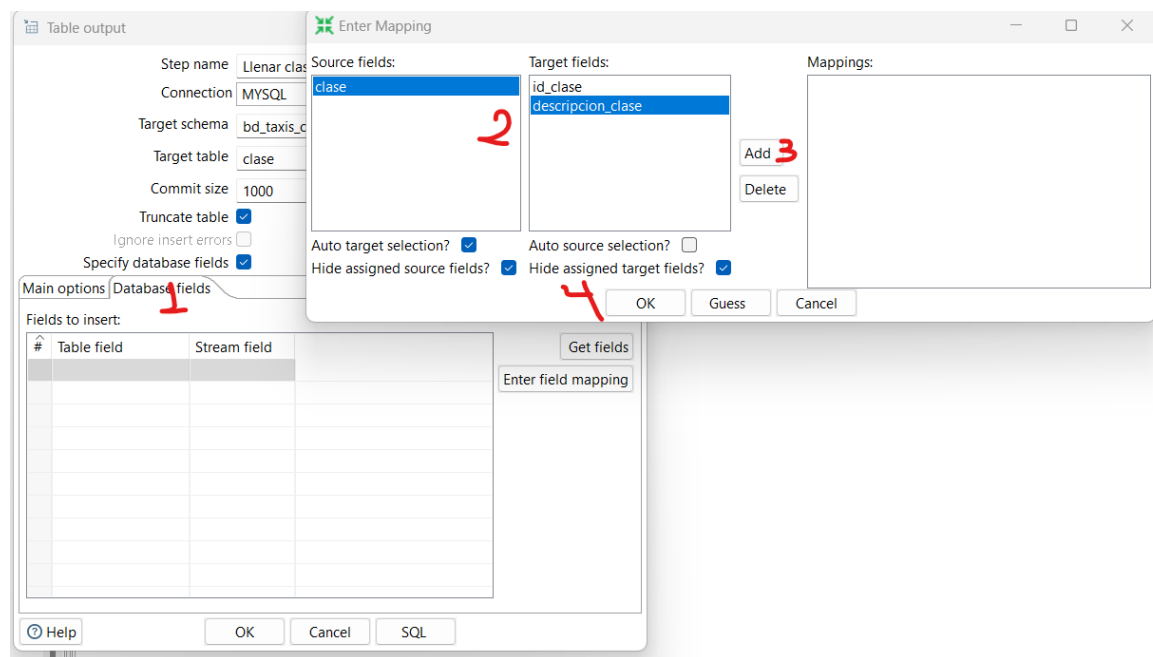
Store the tablename field: ☒

Return auto-generated key: ☐

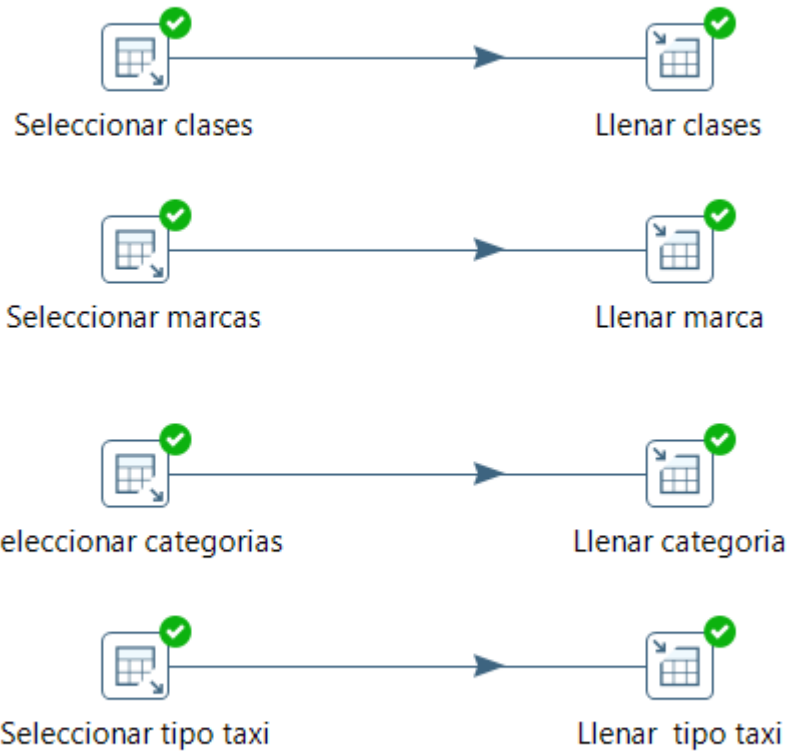
Name of auto-generated key:

Help OK Cancel SQL

Mapear los campos como lo muestra la siguiente imagen



Verificar que se haya cargado el campo a guardar y finalizar.



Ejecutar y comprobar que se hayan llenado las tablas en la base de datos.

PracticaETL_nombre_a Limpieza de la data Llenar tablas de la

Seleccionar clases Llenar clases

Seleccionar marcas Llenar marca

Seleccionar categorias Llenar categoria

Seleccionar tipo taxi Llenar tipo taxi

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics

2023/11/16 13:52:56 - Seleccionar marcas.0 - Finished reading query, closing connection

2023/11/16 13:52:56 - Seleccionar categorias.0 - Finished processing (I=7, O=0, R=0, W=4)

2023/11/16 13:52:56 - Seleccionar marcas.0 - Finished processing (I=46, O=0, R=0, W=4)

2023/11/16 13:52:56 - Llenar tipo taxi.0 - Finished processing (I=0, O=4, R=4, W=4, U=0)

2023/11/16 13:52:56 - Llenar categoria.0 - Finished processing (I=0, O=7, R=7, W=7, U=0)

2023/11/16 13:52:56 - Llenar clases.0 - Finished processing (I=0, O=8, R=8, W=8, U=0, E=0)

2023/11/16 13:52:56 - Llenar marca.0 - Finished processing (I=0, O=46, R=46, W=46, U=0)

2023/11/16 13:52:56 - Spoon - The transformation has finished!!

Navigation

Filter objects

apextechnology
base2
base_estud
baserevista
bd_acceso
bd_biblioteca
bd_lavadora
bd_producto
bd_proyecto
bd_proyecto_1
bd_taxis_cuenca

Table: clase

Columns:

id_clase
descripcion_clase

Result Grid

id_tipo_taxi	descripcion_tipo_taxi
1	Sedan
2	Microbus
3	Rural
4	Adaptado
NULL	NULL

tipo_taxi 1 x

Output

Action Output

#	Time	Action
28	13:53:11	SELECT * FROM bd_taxis_cu

data_transformada_ex Pasando datos Iner join Trabajo final Insercion da datos t Insercion datos tabl Insercion de datos t

Institucion Insercion_datos_I

Area Insercion_datos_A

Campus Insercion_datos_C

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics

2024/05/06 17:34:00 - Insercion_datos_C.0 - Finished processing (I=0, O=1, R=1, W=1, U=0)

2024/05/06 17:34:00 - Insercion_datos_C.0 - Finished processing (I=0, O=1, R=1, W=1, U=0)

2024/05/06 17:34:00 - Insercion_datos_N.F.0 - Finished processing (I=0, O=3, R=3, W=3, U=0)

2024/05/06 17:34:00 - Insercion_datos_I.0 - Finished processing (I=0, O=2, R=2, W=2, U=0)

2024/05/06 17:34:00 - Insercion_datos_C.R.0 - Finished processing (I=0, O=20, R=20, W=20, U=0)

2024/05/06 17:34:00 - Insercion_datos_P.0 - Finished processing (I=0, O=6, R=6, W=6, U=0)

2024/05/06 17:34:00 - Insercion_datos_A.0 - Finished processing (I=0, O=5, R=5, W=5, U=0)

2024/05/06 17:34:00 - Insercion_datos_T.F.0 - Finished processing (I=0, O=2, R=2, W=2, U=0)

2024/05/06 17:34:00 - Spoon - The transformation has finished!!

New Connection

Filter tables in tec_azuay

tec_azuay

area
campus
canton
carrera
duracion
institucion
matricula
matricula_stg
modalidad
nivel_formacion
periodo
provincia
tipo_financiamiento
tipo_institucion

Build complex queries using dr...

Query 1 x

CREMENT PRIMARY KEY,

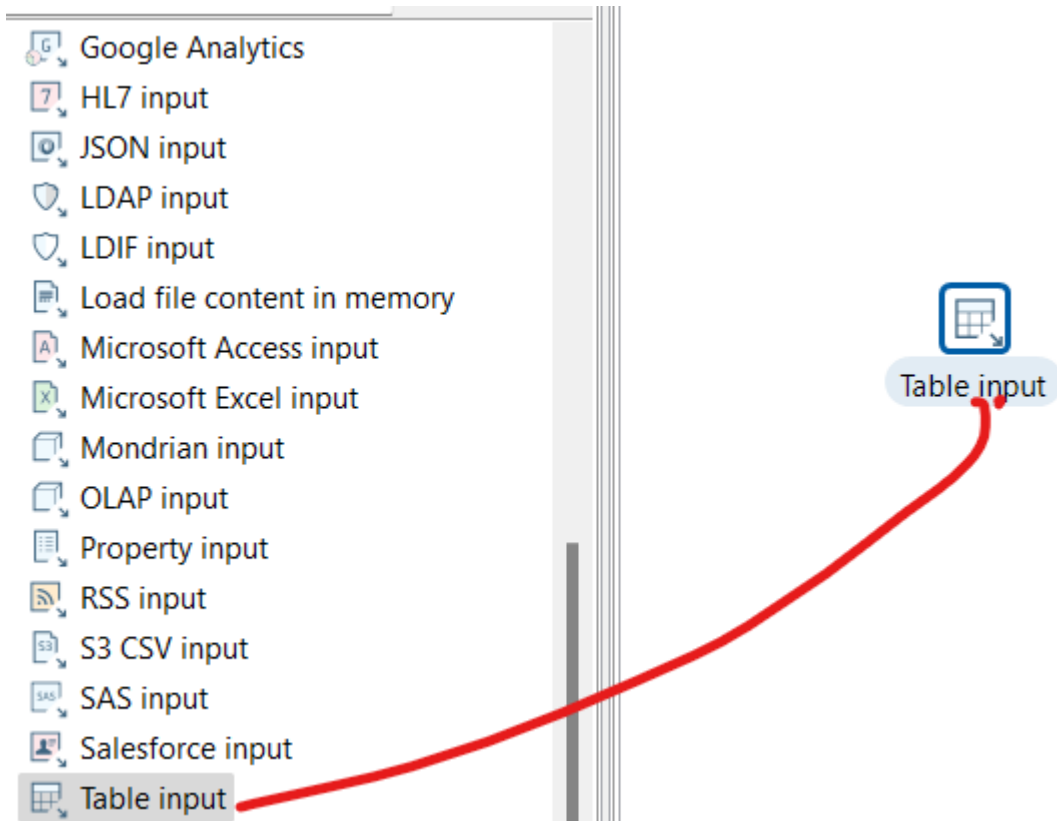
Table

id_area	descripcion_a
1	SERVICIOS
2	SALUD Y BIENE
3	ADMINISTRACIÓ
4	TECNOLOGÍAS D
5	INGENIERÍA, I

Database: t.

3.4. Cargar los dato de taxi para ello crear un nuevo archivo con el nombre Llenar Tabla Taxi y realizar las siguientes acciones:

Arrastrar al área de trabajo le elemento table input en el cual seleccionaremos los datos a guardar en la tabla taxi.



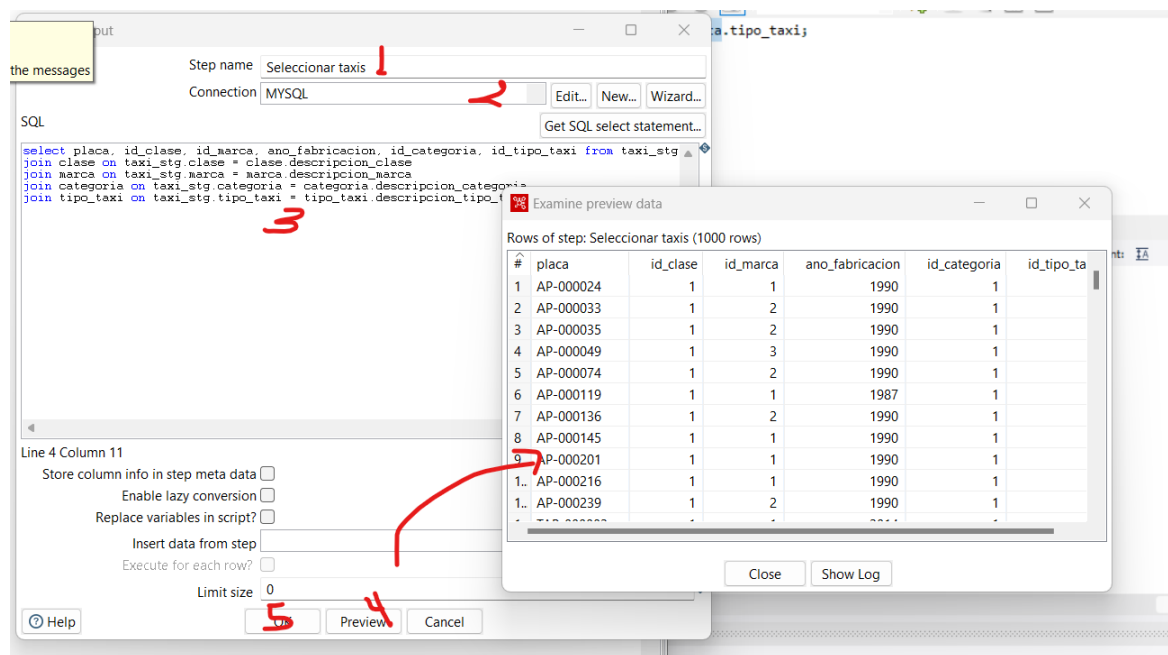
Renombrar con el nombre Seleccionar taxis, crear la conexión a la base de datos y agregar la siguiente sentencia.

```
select placa, id_clase, id_marca, ano_fabricacion,  
id_categoria, id_tipo_taxi from taxi_stg  
join clase on taxi_stg.clase = clase.descripcion_clase  
join marca on taxi_stg.marca = marca.descripcion_marca
```

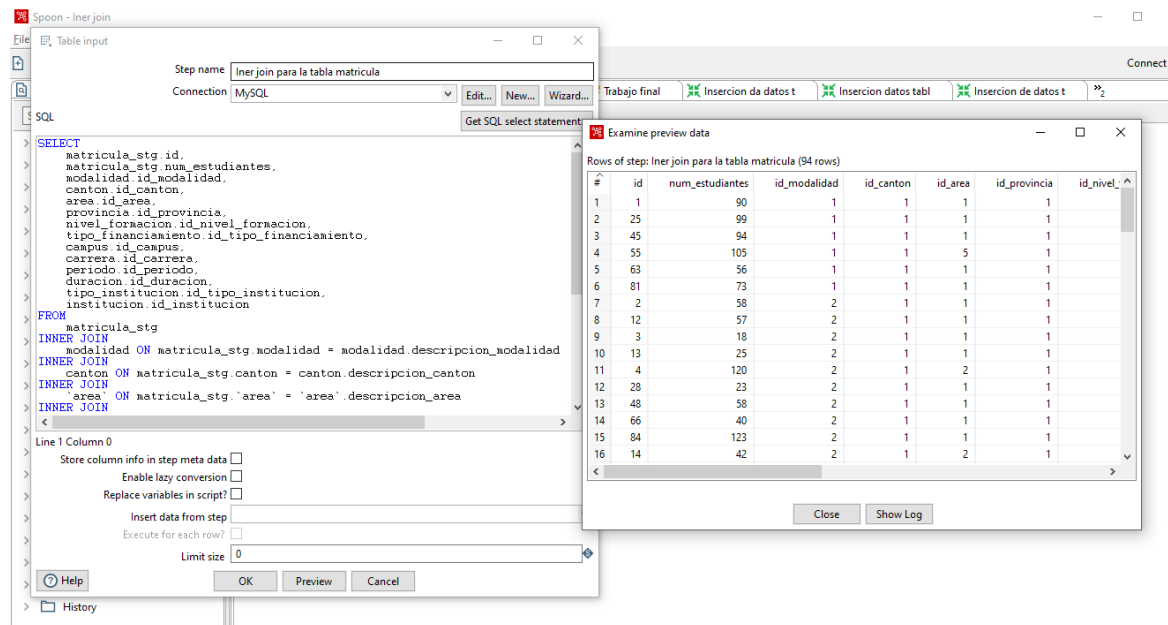
join categoria on taxi_stg.categoria =
categoria.descripcion_categoria

join tipo_taxi on taxi_stg.tipo_taxi =
tipo_taxi.descripcion_tipo_taxi

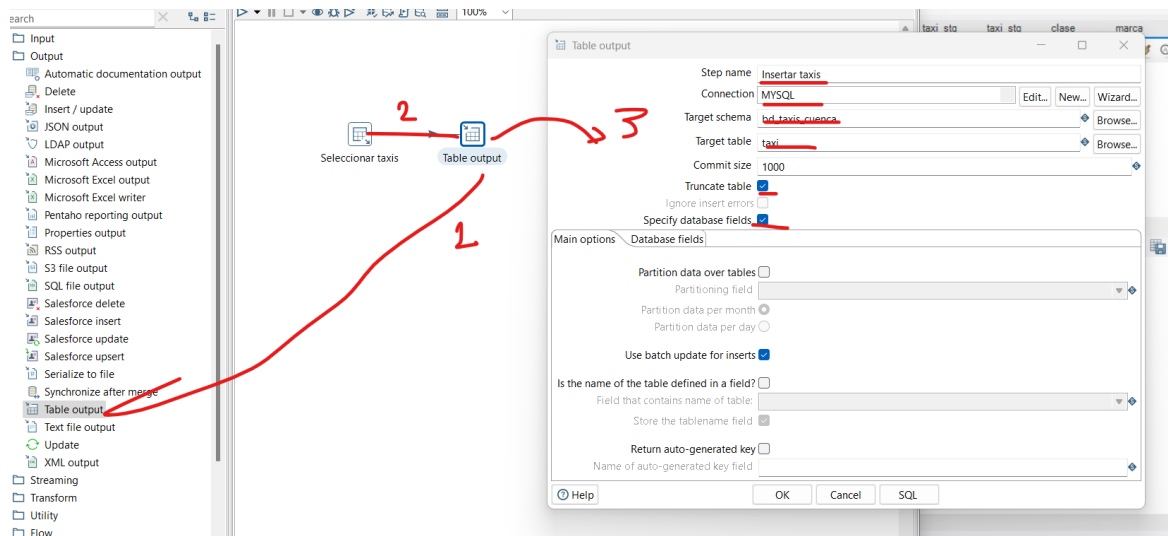
**Luego finalmente colocar un Preview de los datos y dar
click en ok.**



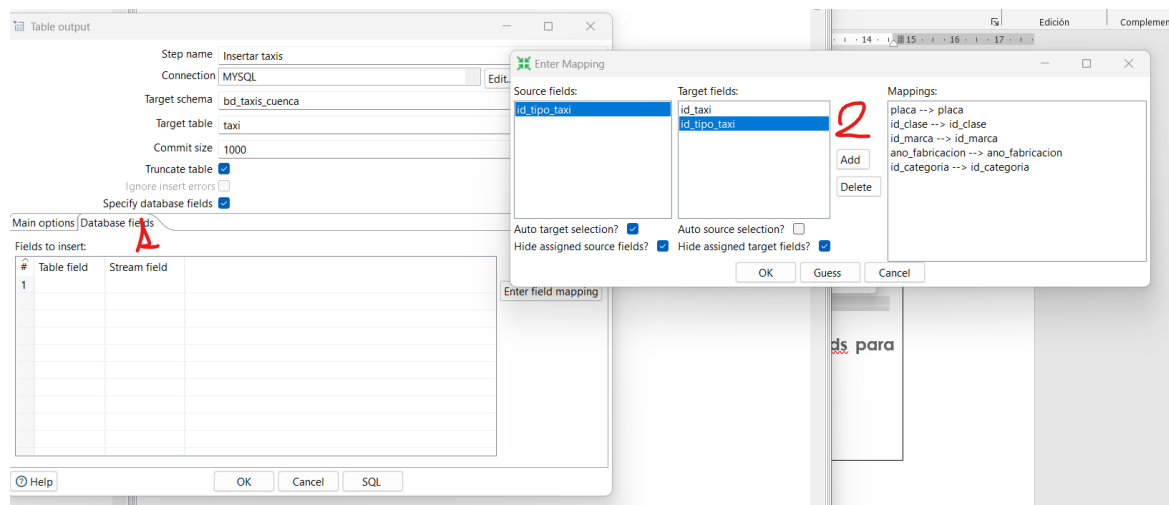
A su vez en la tabla matricula



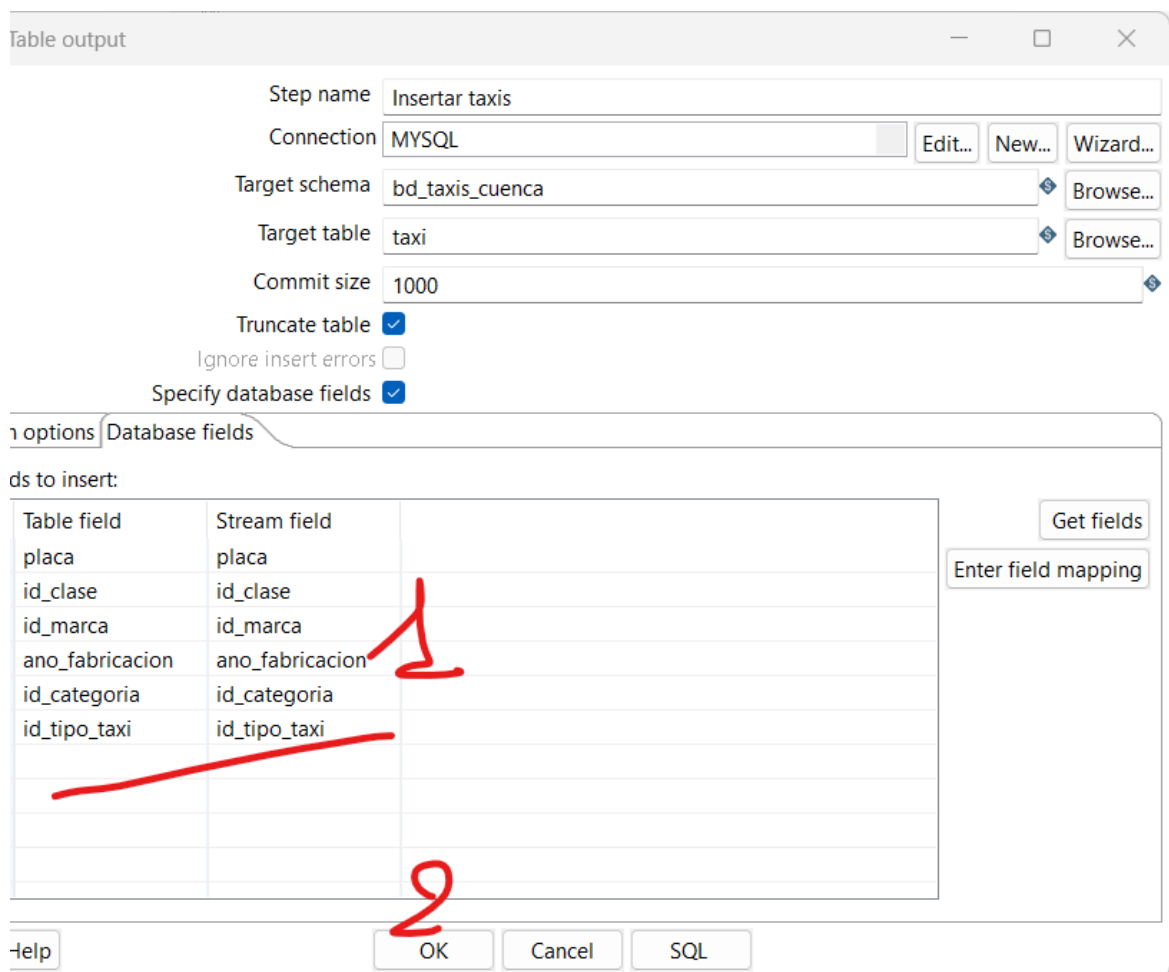
**El siguiente paso final es cargar los datos en la tabla taxi
para ello arrastrar el elemento Table output y cargar los
datos en la tabla taxi.**



**Realizadas las configuraciones pertinentes seleccionar lo
fields para ello se debe de mapear los datos:**



Verificar que se hayan agregado los fields a insertar y guardar la configuración.



Finalmente ejecutar y verificar que se hayan guardado los datos en la tabla realizando un select sobre taxi.

The screenshot shows two parts of the workflow. On the left, an Apache Airflow DAG with two tasks: 'Seleccionar taxis' and 'Insertar taxis', connected by an arrow. On the right, the DBeaver interface showing a SQL query: `SELECT * FROM bd_taxis_cuenca.taxi;` with a red checkmark. Below the query, a 'Result Grid' displays data from the 'taxi' table.

id_taxi	placa	id_clase	id_marca	ano_fabricacion	id_categoria	id_tipo_taxi
331	TA-000261	2	12	2015	3	3
332	TA-000262	2	5	2004	1	1
333	TA-000263	2	5	2005	1	1
334	TA-000265	2	5	2005	1	1
335	TA-000266	2	5	2008	1	1
336	TA-000267	2	1	2009	1	1
337	TA-000268	2	15	2009	1	1
338	TA-000269	2	7	2009	1	3

The screenshot shows three parts of the workflow. On the left, an Apache Airflow DAG with two tasks: 'Iner join para la tabla matricula' and 'Insercion_Matricula'. In the center, the DBeaver SQL script window showing a script to drop and create tables. On the right, the 'Table Data' view for the 'matricula' table.

```

1 DROP DATABASE taxis;
2 DROP TABLE IF EXISTS taxi;
3 DROP TABLE IF EXISTS categoria;
4 DROP TABLE IF EXISTS marca;
5 DROP TABLE IF EXISTS clase;
6 DROP TABLE IF EXISTS tipo_taxi;
7
8 CREATE TABLE tipo_taxi(
9     id_tipo_taxi INT AUTO_INCREMENT PRIMARY KEY,
10    fech_matricula DATE,
11    placas VARCHAR(10)
12 );
13

```

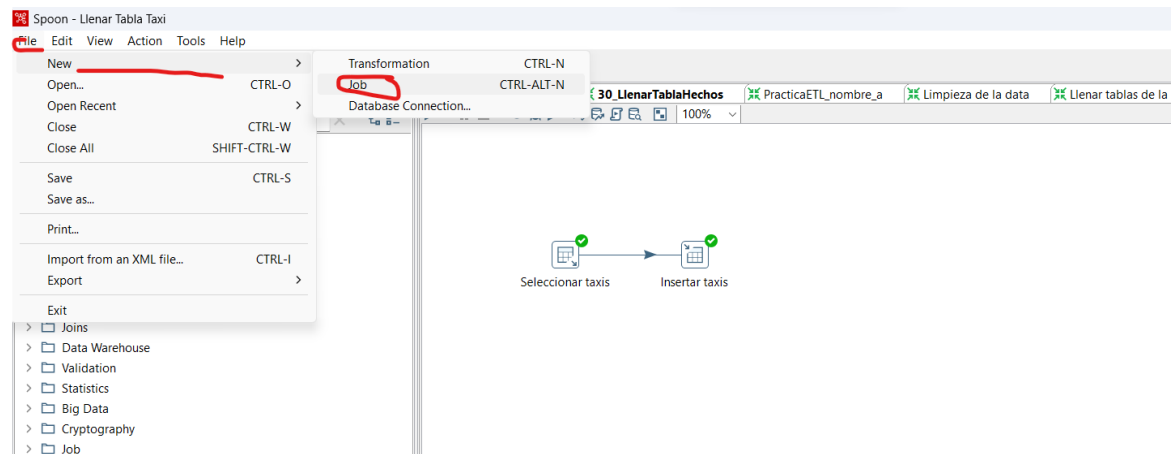
id_matricula	num_estudiantes	id_institucion	id_tipo
1	90	1	1
2	99	1	1
3	94	1	1
4	105	1	1
5	56	1	1

Execution Results

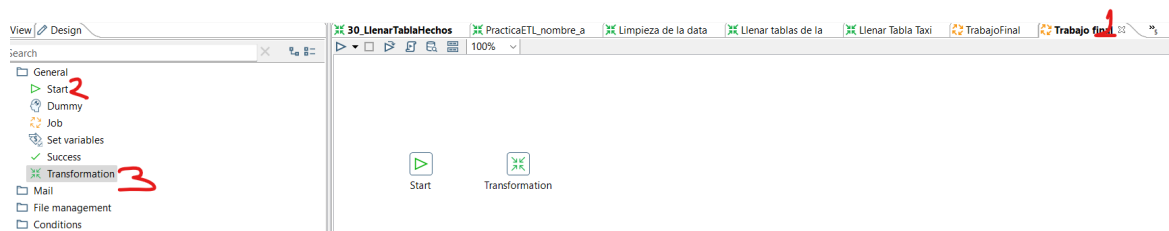
- 2024/05/06 17:36:56 - Spoon - transformation operator.
- 2024/05/06 17:36:56 - Spoon - Launching transformation [Iner join]...
- 2024/05/06 17:36:56 - Spoon - Started the transformation execution.

Para no estar ejecutando uno a uno los archivos de transformación se puede adjuntarlos en un archivo de trabajo para ello.

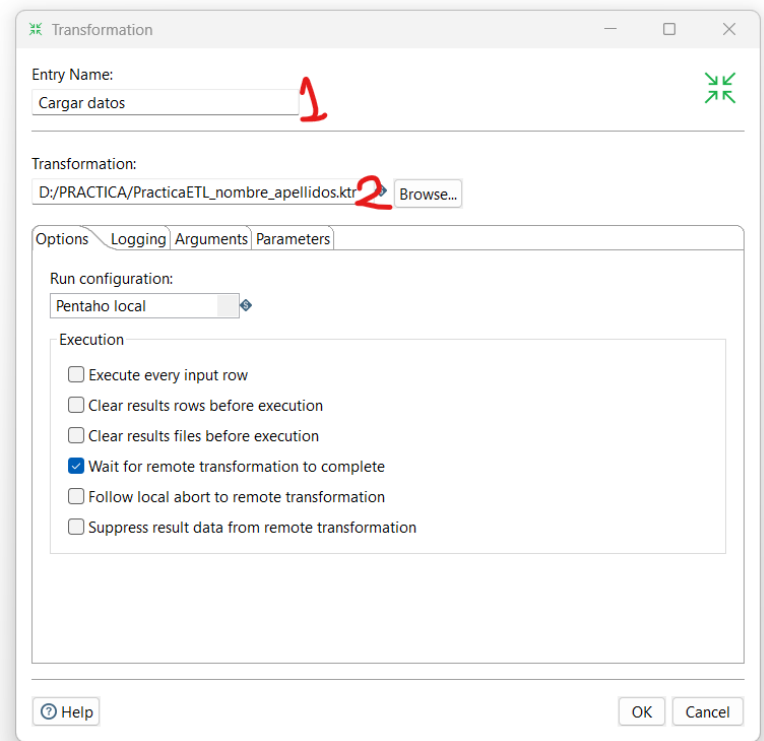
Crear un archivo de trabajo



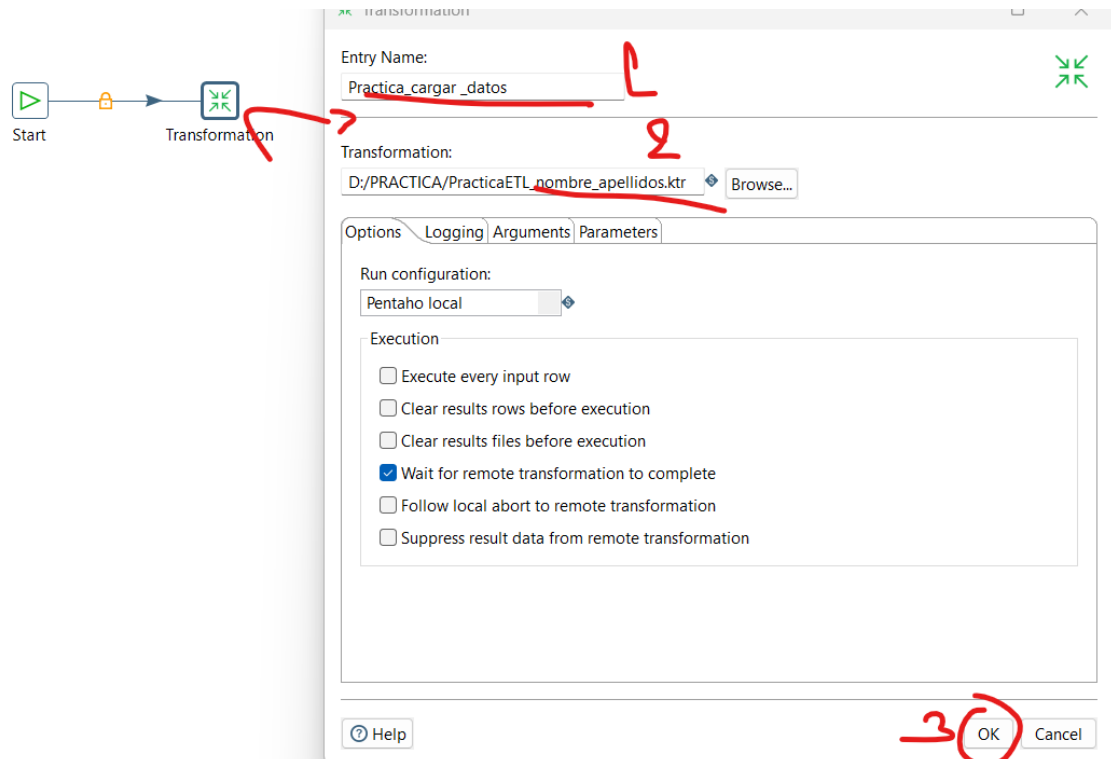
Guardarlo con el nombre Trabajo Final



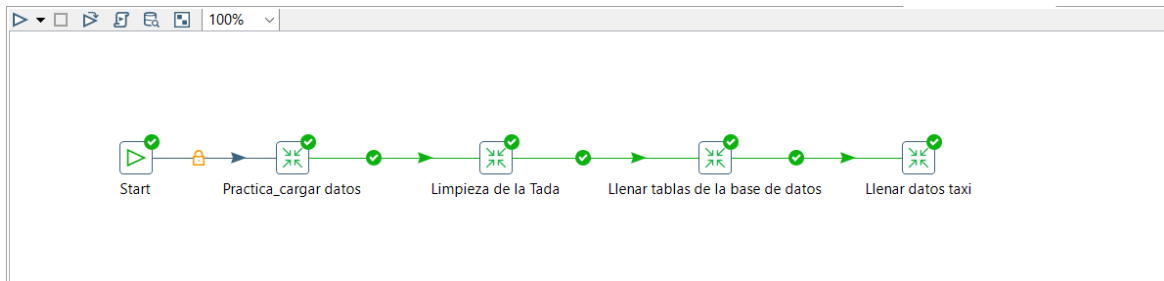
Agregar al área de trabajo los elementos start y configurar el transformation para realizar la secuencia de ejecución.



Renombrar con nombre Cargar datos y hacer referencia al primer archivo creado en la sección 3. De la misma manera agregar mas transformaciones y hacer referencia a los archivos creados en el mismo orden de creación.



Finalmente ejecutar y verificar los datos



1. Normas de Seguridad

Las normas de seguridad se han tomado del reglamento general de seguridad para el uso de los talleres, aulas y laboratorios del Instituto Superior Universitario Tecnológico del Azuay.

El estudiante, al ingresar a los talleres o laboratorios, está sujeto a este reglamento; y, tendrá la supervisión del profesor y del personal técnico; será responsable de:

- a) Usar los Equipos de Protección Personal(EPP) de acuerdo con lo establecido en la "Matriz de equipos de protección individual (EPP`s) requeridos para el ingreso de estudiantes y profesores a los laboratorios y talleres del INSTITUTO";
- b) Al inicio de cada práctica, recibir y revisar el material y herramientas requeridas para la Práctica, serán responsables de su buen uso.
- c) La operación de los equipos por los estudiantes deberá ser con el conocimiento de su funcionamiento y bajo las directrices del profesor o personal técnico del laboratorio o taller; bajo ninguna circunstancia el estudiante podrá trabajar solo y sin vigilancia;
- d) Seguir las instrucciones dadas por el docente o el personal técnico de apoyo;
- e) Al término de la práctica, entregar limpio tanto el material como su área de trabajo;
- f) Informar inmediatamente al profesor o personal técnico de apoyo, cualquier desperfecto que se localice en los equipos e instalaciones.

1. Resultados esperados

- El estudiante realiza paso a paso del proceso ETL, desde la configuración de Pentaho Data Integration hasta la carga efectiva de datos en una base de datos MySQL.

1. Bibliografía

Descripción en norma APA

- Borland, B. (Año de publicación). Pentaho Analytics for MongoDB. Editorial.
- Ramazzina, S. (Año de publicación). Pentaho Business Analytics Cookbook. Editorial.
- Casters, M., Bouman, R., & van Dongen, J. (Año de publicación). Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration. Editorial.
- Roldán, M. C. (Año de publicación). Pentaho Data Integration Beginner's Guide - Second Edition. Editorial.

1. Firmas de Responsabilidad

ESTUDIANTE	DOCENTE	DIRECTORA DE CARRERA
Nombre:	Nombre: Mgtr. Verónica Chimbo	Nombre: Mgtr. Verónica Chimbo
Firma	Firma	Firma
Fecha: ()	Fecha: (11/04/2024)	Fecha: ()