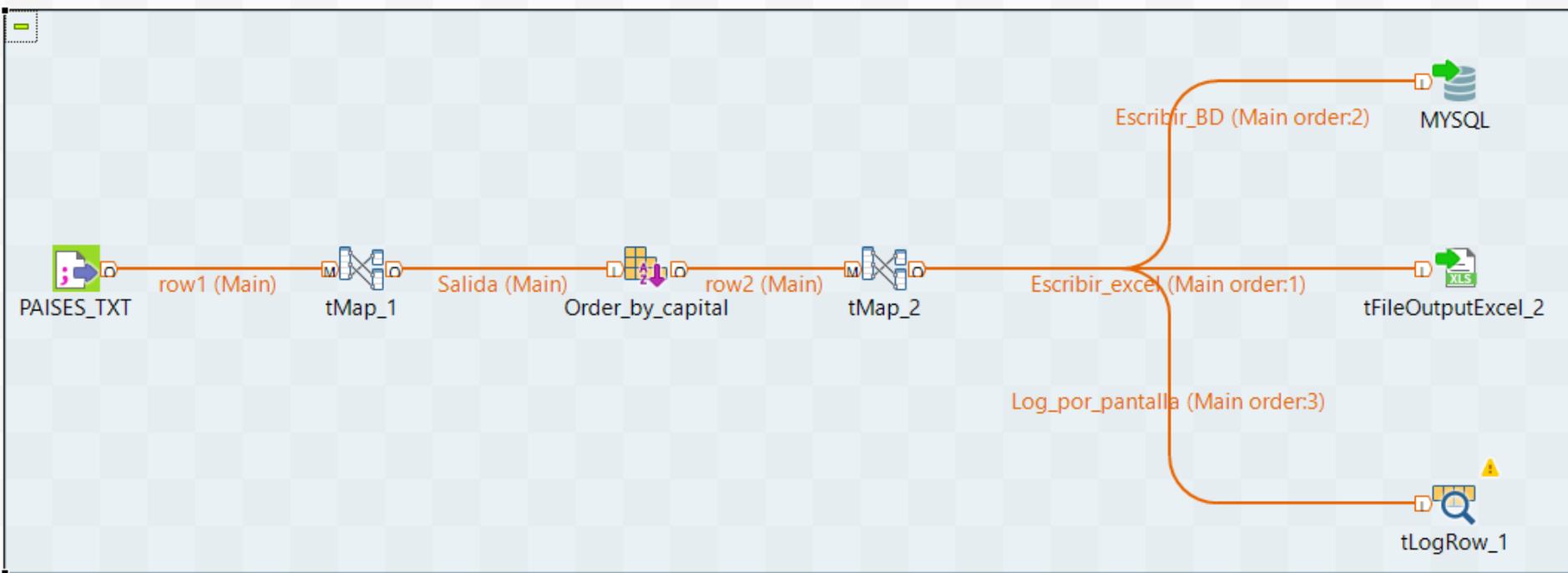




Ejemplo de ETL paso a paso con la herramienta Talend Open Source





© Copyright **Lituus Spa**

Este documento es propiedad de Lituus Spa y su contenido es confidencial. Este no puede ser reproducido, en su totalidad o parcialmente, ni mostrado a terceros, ni utilizado para otros propósitos que los que han originado su entrega, sin el previo permiso escrito de **Lituus Spa**.



```
Paises.txt: Bloc de notas
Archivo Edición Formato Ver Ayuda
Nombre comun;Forma de Gobierno;Capital;Continente;Estatus ONU
Emiratos Árabes Unidos;Monarquía constitucional electiva;Abu Dabi;Asia;Miembro
Nigeria;República presidencialista;Abuya;África;Miembro
Ghana;República presidencialista;Acra;África;Miembro
Etiopía;República parlamentaria;Adís Abeba;África;Miembro
Jordania;Monarquía constitucional;Amán;Asia;Miembro
Andorra;Monarquía constitucional;Andorra la Vieja;Europa;Miembro
Turquía;República parlamentaria;Ankara;Asia-Europa;Miembro
Madagascar;República semipresidencialista;Antananarivo;África;Miembro
Argelia;República semipresidencialista;Argel;África;Miembro
Turkmenistán;República presidencialista;Asjabad;Asia;Miembro
Eritrea;República unipartidista;Asmara;África;Miembro
Paraguay;República presidencialista;Asunción;América;Miembro
Grecia;República parlamentaria;Atenas;Europa;Miembro
Irak;República parlamentaria;Bagdad;Asia;Miembro
Azerbaiyán;República semipresidencialista;Bakú;Asia-Europa;Miembro
Mali / Malí;República semipresidencialista;Bamako;África;Miembro
Brunéi;Monarquía absoluta;Bandar Seri Begawan;Asia;Miembro
Tailandia;Monarquía constitucional;Bangkok;Asia;Miembro
República Centroafricana;República semipresidencialista;Bangui;África;Miembro
Gambia;República presidencialista;Banjul;África;Miembro
San Cristóbal y Nieves;Monarquía constitucional;Basseterre;América;Miembro
Líbano;República parlamentaria;Beirut;Asia;Miembro
Serbia;República parlamentaria;Belgrado;Europa;Miembro
Belice;Monarquía constitucional;Belmopán;América;Miembro
Alemania;República parlamentaria;Berlín;Europa;Miembro
Suiza;República parlamentaria;Berná;Europa;Miembro
Kirguistán;República parlamentaria;Biskek;Asia;Miembro
Guinea-Bisáu;República semipresidencialista;Bisáu;África;Miembro
Colombia;República presidencialista;Bogotá;América;Miembro
Brasil;República presidencialista;Brasilia;América;Miembro
Eslovaquia;República parlamentaria;Bratislava;Europa;Miembro
República del Congo;República semipresidencialista;Brazzaville;África;Miembro
Línea 1, columna 1 100% Windows (CRLF) ANSI
```



Crear un proyecto



Talend Open Studio for Data Integration

On connection: Local (Local) Manage Connections

Select an existing project

Local_Project - java

Create a new project: LITUUS Create

Import a demo project

Import an existing project

Always ask me at startup Finish

Talend Open Studio for Data Integration

On connection: Local (Local) Manage Connections

Select an existing project

LITUUS - java
Local_Project - java

Create a new project

Import a demo project

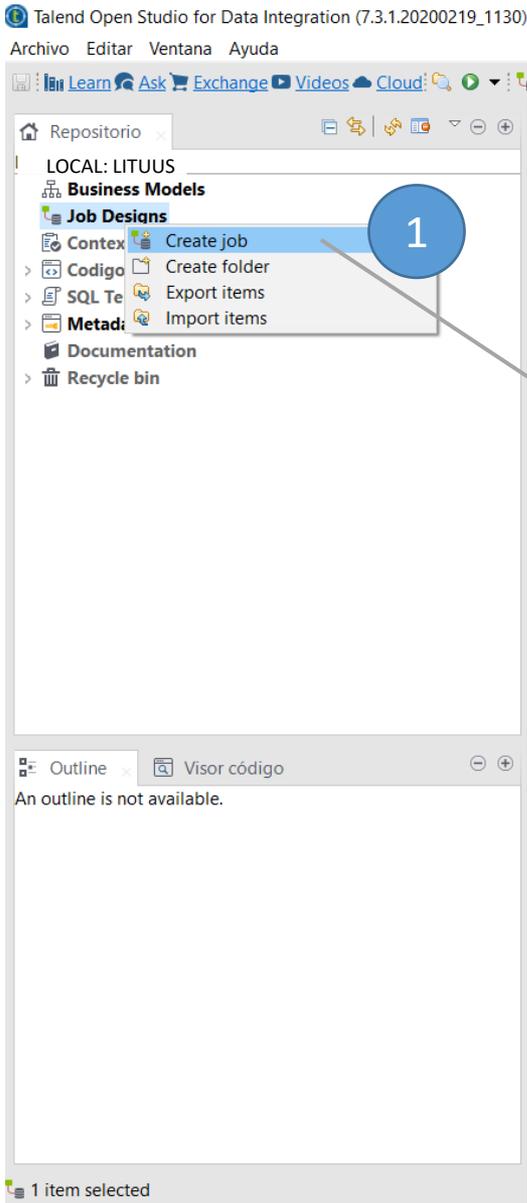
Import an existing project

Always ask me at startup Finish



Una vez dentro del proyecto, crearemos un Job o ETL al dar clic al botón secundario sobre Job Designs

Nota: Procure seguir una nomenclatura, por ejemplo **JB** para todos los objetos que sean del tipo Job



New job

Add a job in the repository

Nombre

Purpose

Descripción

Author

Bloqueador

Version

Estado

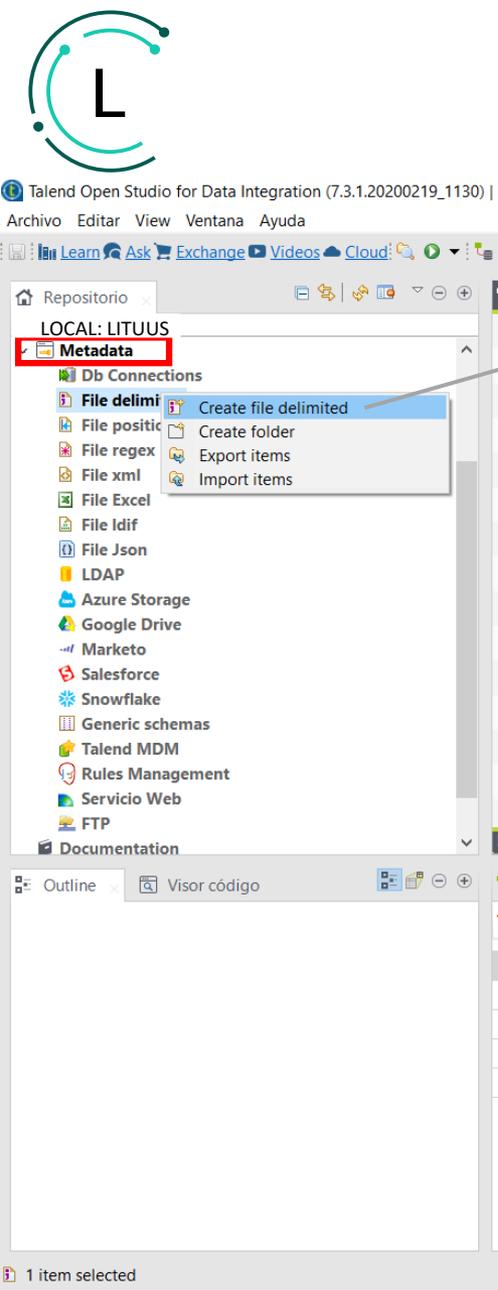
Path



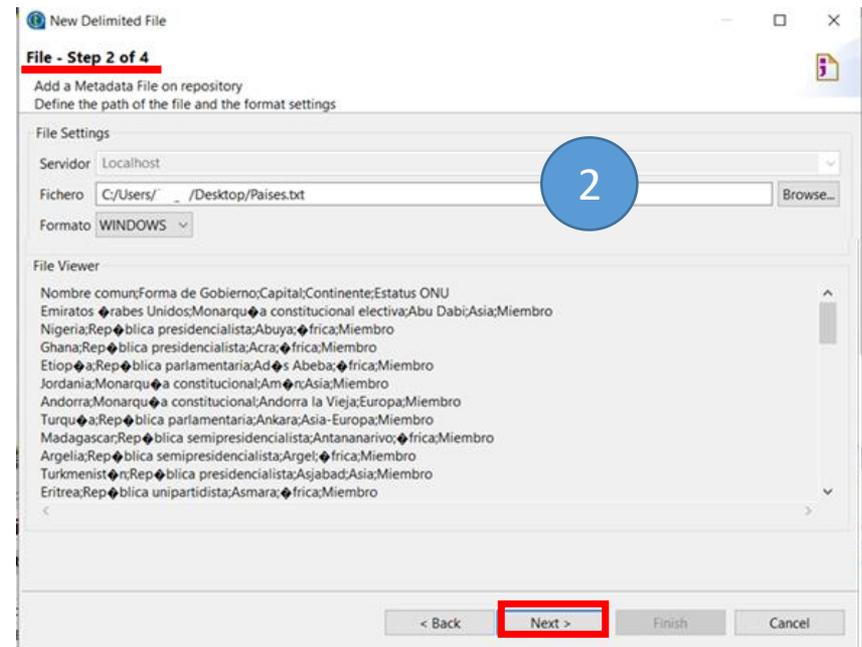
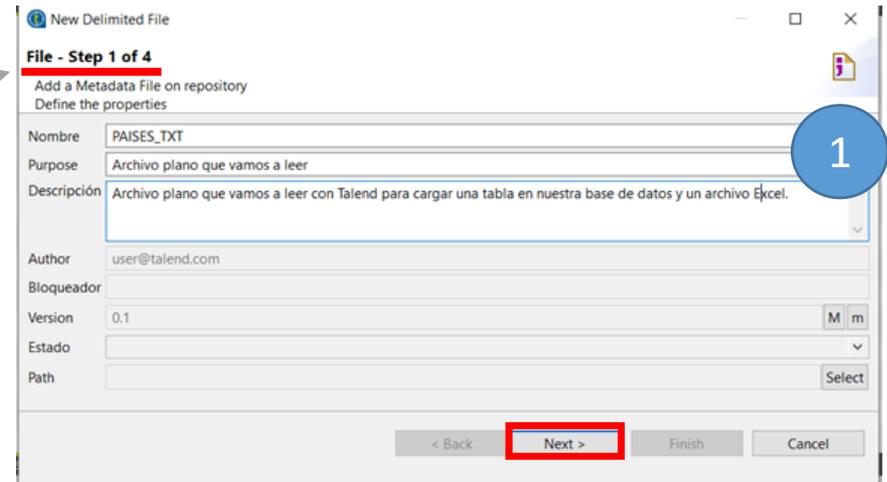
Una vez dentro del Job, se desplegará un canvas vacío que nos permitirá definir la lógica del mismo de forma gráfica

Canvas

Nombre	Valor



Para leer el archivo Paises.txt crearemos un objeto o metadato que referencie a dicho archivo en la sección:
Metadata -> File Delimited





3

New Delimited File

File - Step 3 of 4

Add a Metadata File on repository
Define the setting of the parse job

File Settings

Codificación: ISO-8859-1

Separador de Campo: Corresponding Character

Separador de Fila: Corresponding Character

Escape Char Settings

CSV: Delimited

Carácter de Escape: Vacío

Text Enclosure: Vacío

Rows To Skip

Encabezado: 1

Pie de Página:

Skip empty row:

Limit Of Rows

Limit:

Preview

Set heading row as column names: Refresh Preview

Nombre comun	Forma de Gobierno	Capital	Continente	Estatus ONU
Emiratos Árabes Unidos	Monarquía constitucional electiva	Abu Dabi	Asia	Miembro
Nigeria	República presidencialista	Abuya	África	Miembro
Ghana	República presidencialista	Acra	África	Miembro
Etiopía	República parlamentaria	Adís Abeba	África	Miembro

Exportar como contexto Revertir Contexto

< Back Next > Finish Cancel

Indicamos a la herramienta como debe leer el archivo (codificación, encabezado, separadores, etc.)

4

New Delimited File

File - Step 4 of 4

Add a Schema on repository
Define the Schema

Nombre: PAISES_TXT

Comentario:

Esquema

Click to update schema preview Guess

Description of the Schema

Columna	Clave	Tipo	N.	Date Pattern (Ctrl...	Longitud	Precision	Default	Com
Nombre_comun	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		26	0		
Forma_de_Gobierno	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		36	0		
Capital	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		20	0		
Continente	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		11	0		
Estatus_ONU	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		7	0		

< Back Next > Finish Cancel

Finalmente nos aseguramos que las columnas cuenten con un tipo de dato correcto

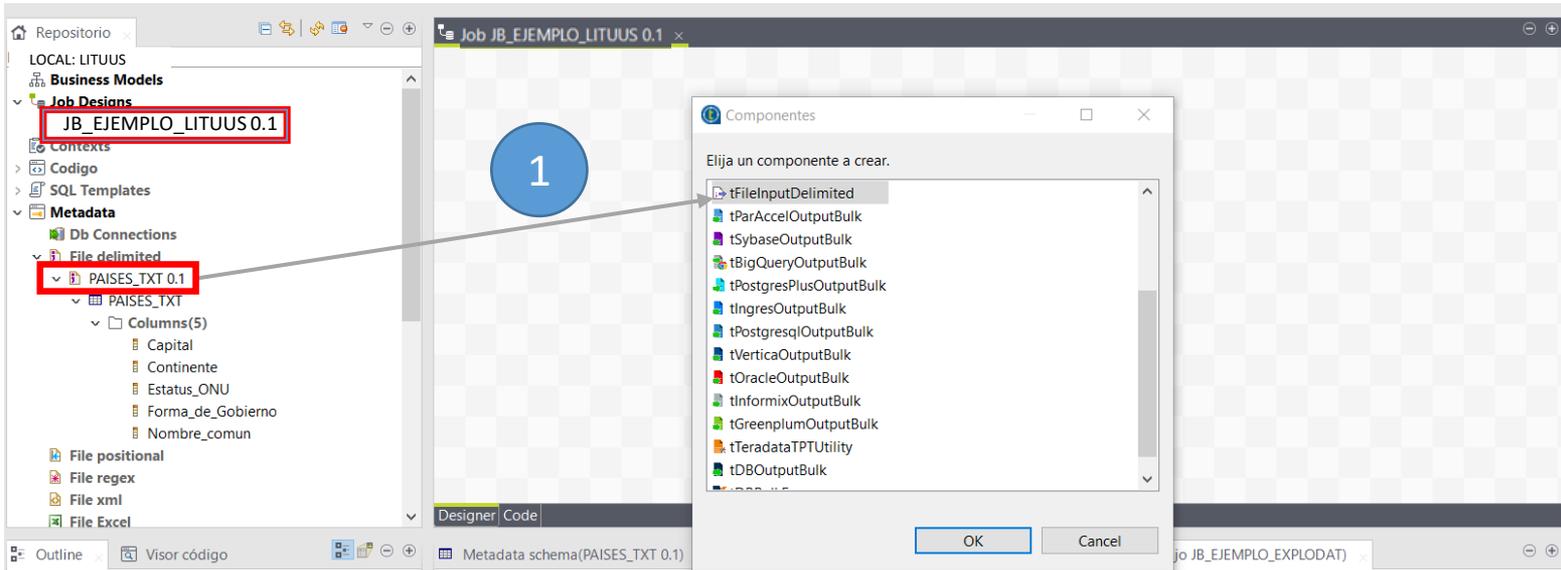
Repositorio

LOCAL: LITUUS

- Business Models
- Job Designs
 - JB_EJEMPLO_LITUUS 0.1
- Contexts
- Codigo
- SQL Templates
- Metadata**
 - Db Connections
 - File delimited**
 - PAISES_TXT 0.1**
 - File positional
 - File regex
 - File xml
 - File Excel
 - File Idif
 - File Json
 - LDAP
 - Azure Storage
 - Google Drive
 - Marketo
 - Salesforce

5

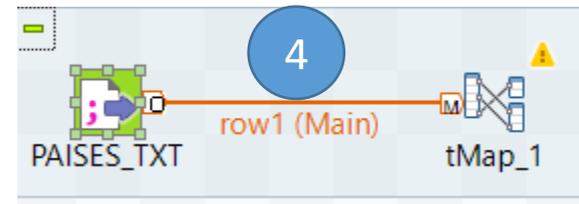
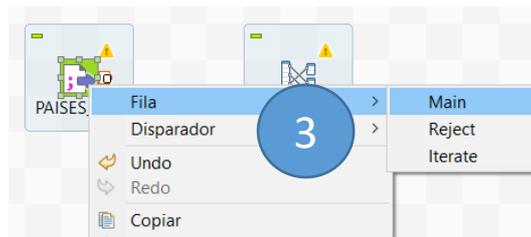
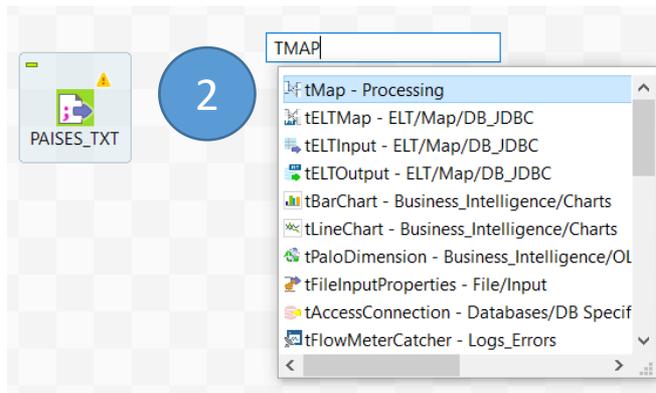
Una vez finalizado el proceso, nos quedará un objeto listo para ser consumido desde nuestros ETLs



2. Escribimos tMAP en el canvas.
Este objeto nos permite realizar
diversas transformaciones

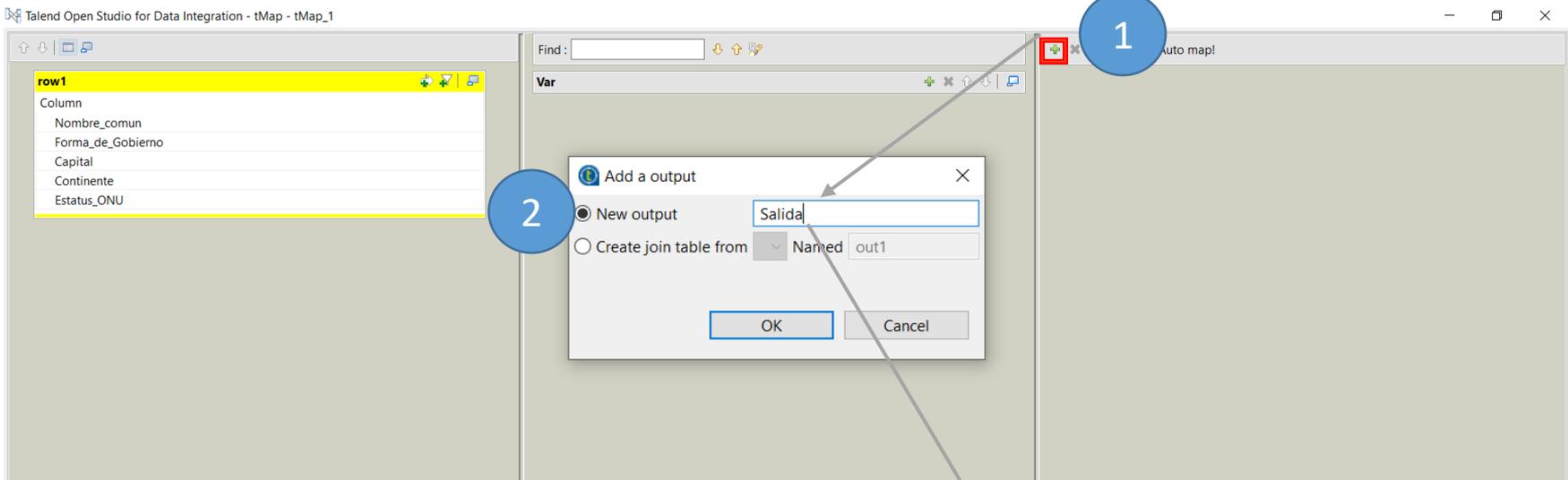
3. Clic secundario en
PAISES y escogemos la
opción: *Fila -> Main*

4. La línea generada la
conectamos con el tMap_1
Los datos pasarán al tMap

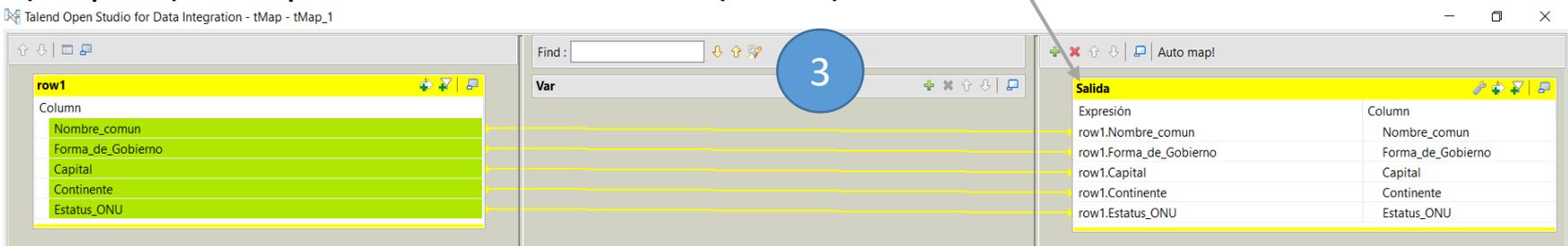




EL objeto tMap nos permite generar múltiples salidas con diversas transformaciones. Para este caso crearemos una salida directa del archivo TXT, la cual procesaremos con otro objeto en el paso siguiente. Más adelante incorporaremos otro tMap

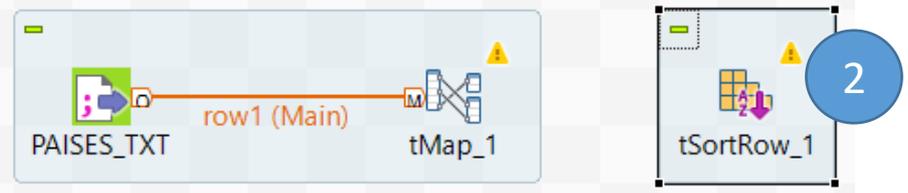
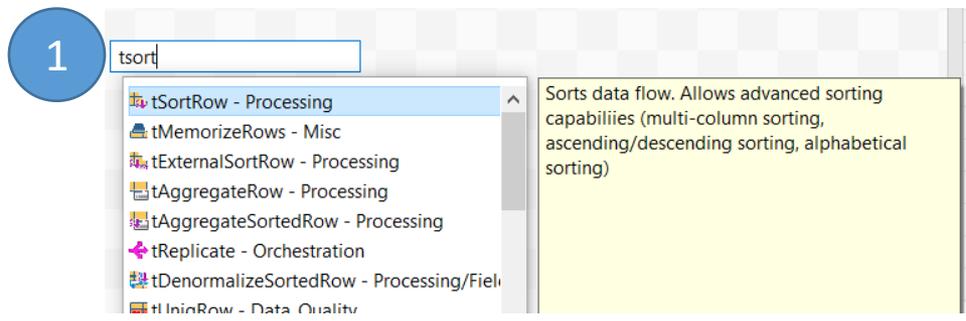


Para este ejemplo trabajaremos con todos los campos, por lo que los arrastraremos (mapear) de izquierda hacia la derecha (Salida)

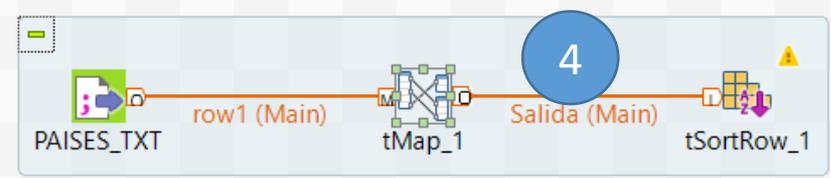
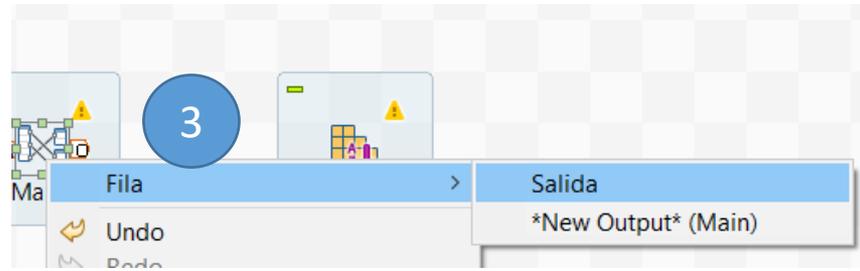




A continuación, aplicaremos un Order By a los registros entregados por el tMap.
Para ordenar los registros escribiremos tSort en el canvas e incorporaremos dicho objeto.



Similar al paso anterior, nos posicionamos en tMap y al dar clic derecho seleccionamos la opción Fila -> Salida





El objeto tSortRow_1 nos permite definir criterios de orden (Order By)

Criterio(s)	Columna del esquema	Orden numérico o alfabético?	Orden Ascendente o Descendente?
-------------	---------------------	------------------------------	---------------------------------

Debemos indicar el campo que usaremos para ordenar y en qué orden:

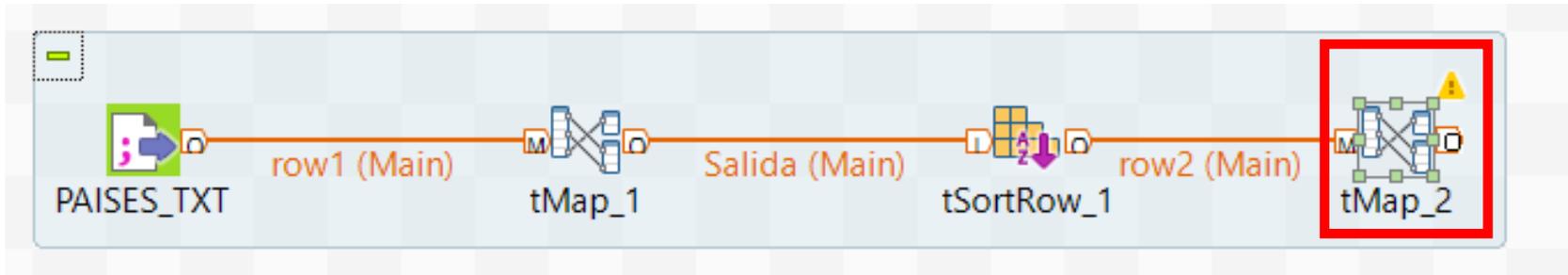
Criterio(s)	Columna del esquema	Orden numérico o alfabético?	Orden Ascendente o Descendente?
	Nombre_comun	Alfanumérico	Ascendente

Para este caso usamos el campo llamado Capital, quedando el siguiente orden:

Criterio(s)	Columna del esquema	Orden numérico o alfabético?	Orden Ascendente o Descendente?
	Capital	Alfanumérico	Ascendente



El objeto tMap nos permite realizar diversas transformaciones, para este ejemplo lo usaremos para agregar un nuevo campo del tipo secuencia.

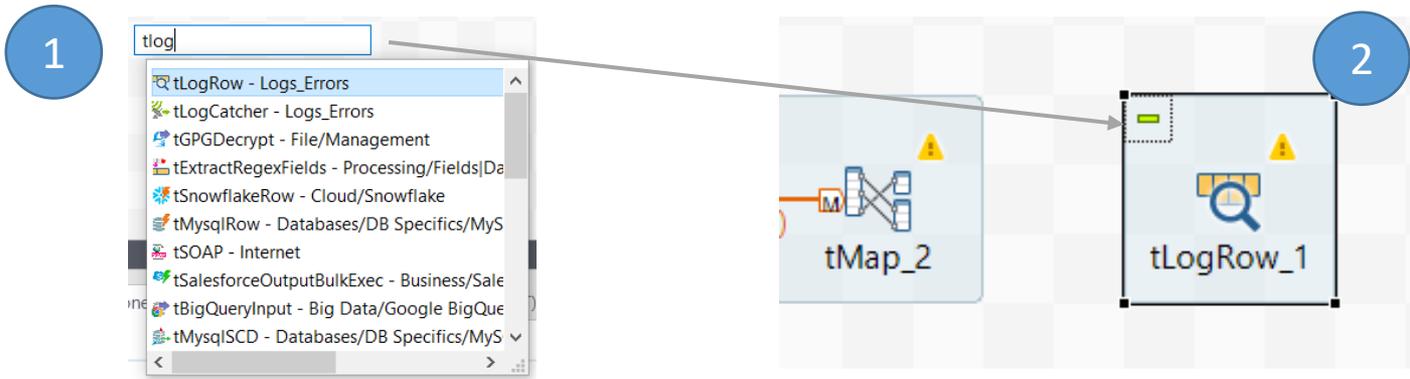


El campo se llamará ID cuyo correlativo será generado por la función sequence. La salida del tMap se llamará Log_por_pantalla

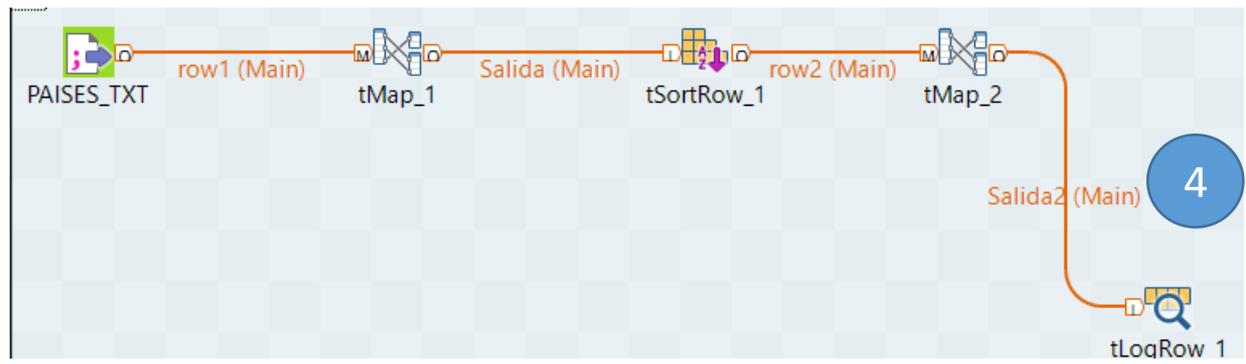
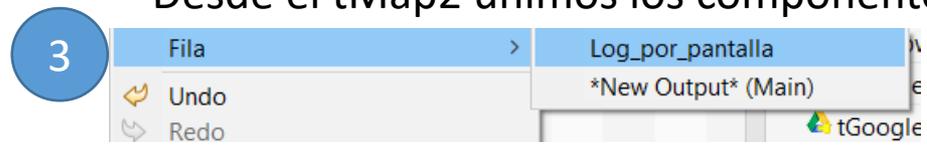
Log_por_pantalla	
Expresión	Column
Numeric.sequence("mySeq", 1, 1)	ID
row2.Nombre_comun	NOMBRE
row2.Forma_de_Gobierno	FORMA_GOBIERNO
row2.Capital	CAPITAL
row2.Continente	CONTINENTE
row2.Estatus_ONU	ESTATUS_ONU



Para poder mostrar por pantalla en formato log el resultado, usaremos el objeto tLogRow el cual lo incorporaremos al canvas simplemente escribiendo su nombre



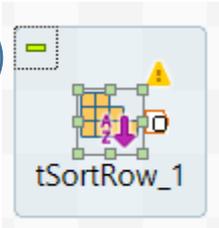
Desde el tMap2 unimos los componentes





Para renombrar un objeto a un nombre más descriptivo se debe seleccionar e ir a la pestaña Component -> View

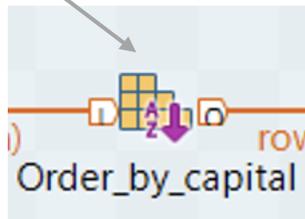
1



En la propiedad “Label format” ingresamos el nombre de interés. Por ejemplo, Order_by_capital

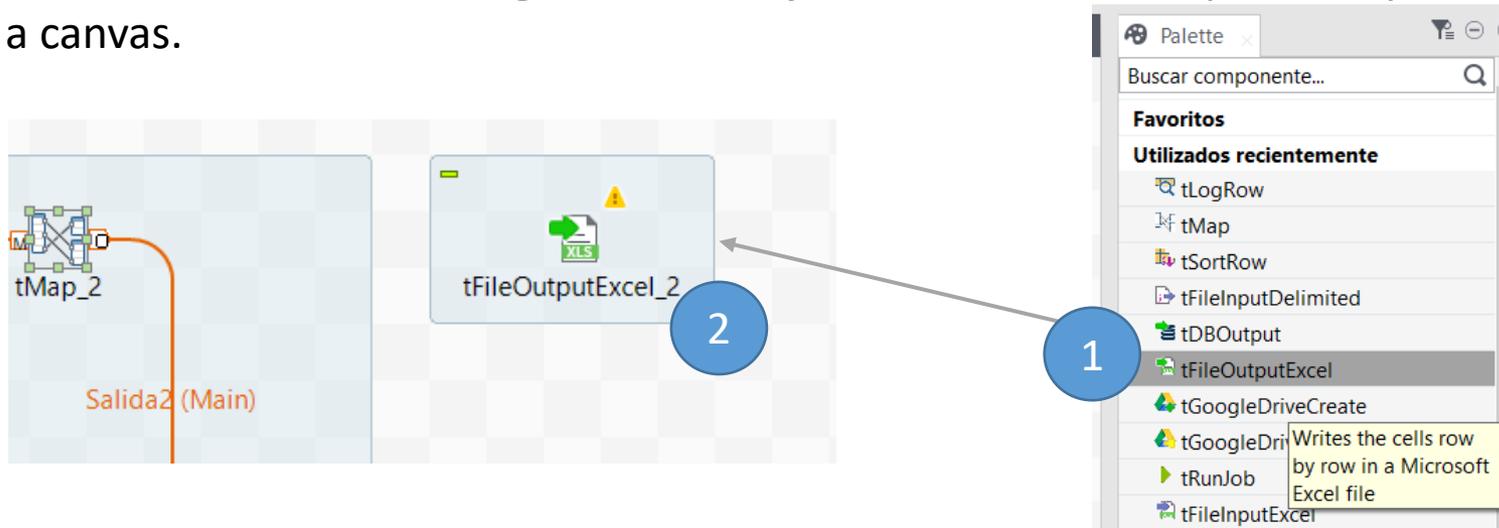
2

Property	Value
Basic settings	Label format Order_by_capital
Advanced settings	Hint format __UNIQUE_NAME__ __COMMENT__
Dynamic settings	Connection format row

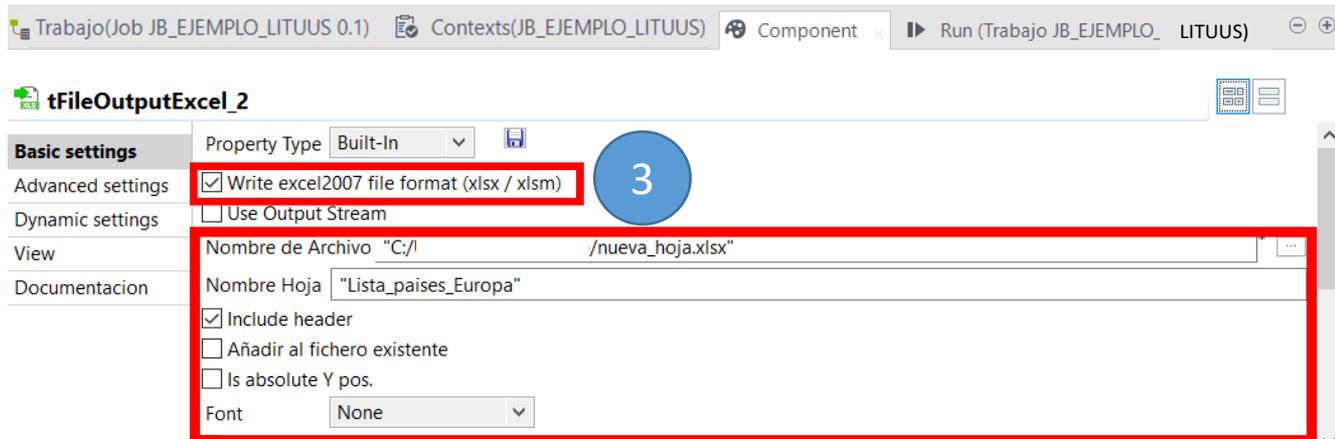




El objeto tMap puede tener múltiples salidas, para este caso, incorporaremos un Excel. De la paleta (menú lateral derecho) escogeremos el objeto llamado tFileOutputExcel y lo arrastraremos a canvas.

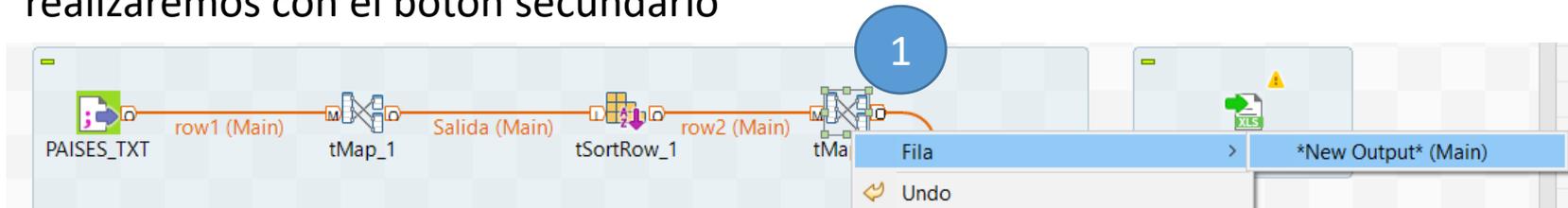


Seleccionamos el objeto y definimos sus propiedades u opciones de salida

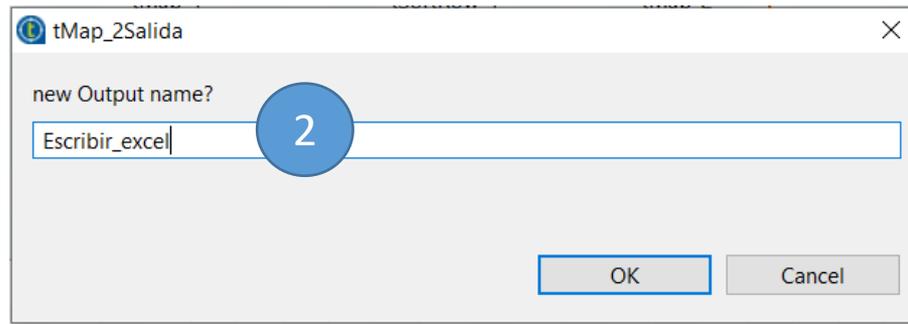




Desde el tMap2 crearemos una nueva salida, sin embargo, esta vez lo realizaremos con el botón secundario



Definimos un nombre de la salida: Escribir_excel

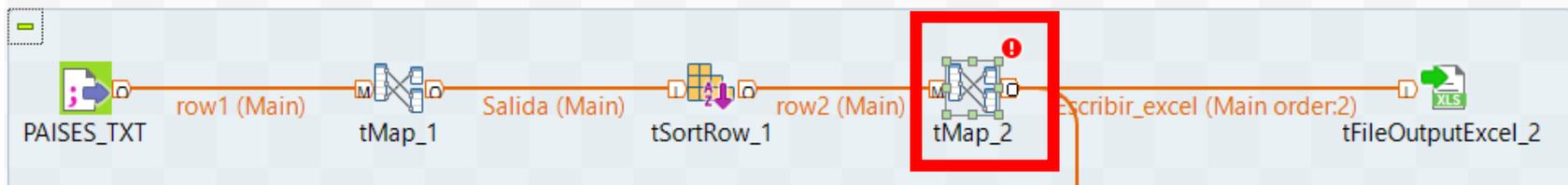


Conectamos la salida al excel





El objeto tMap nos marca un error debido a que no hemos mapeado que campos queremos pasar al archivo Excel.



Entramos al tMap2 y mapeamos los campos de interés a cargar en el Excel. Además aprovecharemos de agregar un filtro para que solo cargue países del continente europeo

Escribir_excel

```
row2.Continente.equals("Europa")
```

Expresión	Column
Numeric.sequence("mySeq", 1, 1)	ID
row2.Nombre_comun	Nombre_comun
row2.Forma_de_Gobierno	Forma_de_Gobierno
row2.Capital	Capital
row2.Continente	Continente
row2.Estatus_ONU	Estatus_ONU



Aprovechando que el tMap nos permite generar múltiples salidas en 1 solo paso, agregaremos como destino una tabla en la base de datos MySQL. Desde la paleta (menú lateral derecho) escogemos un componente de salida o destino y luego lo arrastramos al canvas.

3 Ingresamos los parámetros de nuestra base de daos

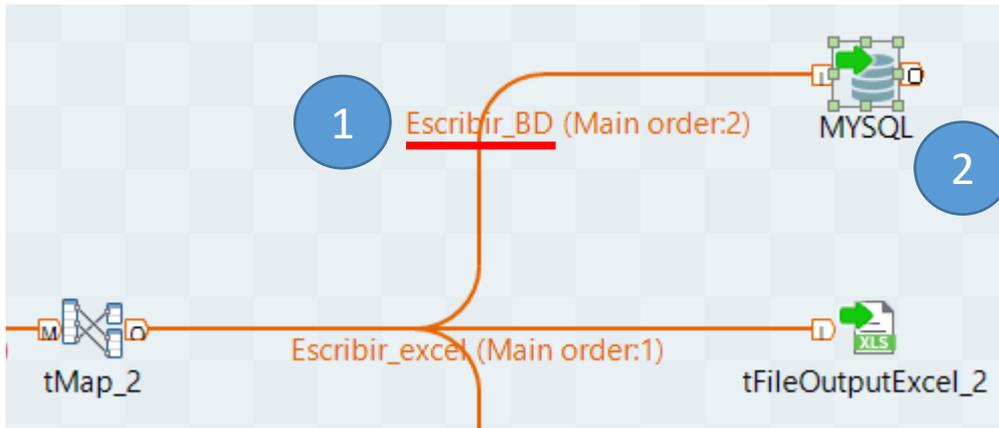
Property	Value
Database	MySQL
Property Type	Built-In
DB Version	Mysql 8
Host	"Ingreso_Host"
Puerto	"3306"
Database	"Ingreso_Base_de datos"
Nombre de usuario	"Ingreso_usuario"
Contraseña	*****
Table	"Ingreso_tabla"
Action on table	None
Action on data	Insert
Esquema	Built-In

1 tMySQLOutput

2 tDBOutput_1



Aprovechamos el tMap2 para generar una nueva salida: Escribir_BD



Note que hemos cambiado el nombre del componente destino por uno más fácil de identificar: MYSQL

Expresión	Column
Numeric.sequence("mySeq", 1, 1)	ID
row2.Nombre_comun	NOMBRE
row2.Capital	CAPITAL

Columna	Cl...	Tipo	<input checked="" type="checkbox"/> N.	Date Pattern (Ctrl+...)	Longitud	Precision	Default	Comentari...
ID	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>					
NOMBRE	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		26	0		
CAPITAL	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>					

Dentro del tMap indicaremos el mapeo, que para este caso contará con menos campos.



Nuestro Job/ETL esta listo para ser ejecutado.

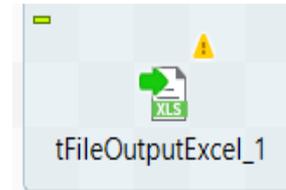
The screenshot displays the Lituus ETL Job Designer interface. The main workspace shows a workflow diagram for 'Job JB_EJEMPLO_EXPLODAT 0.1'. The workflow starts with a 'PAISES_TXT' source, followed by 'row1 (Main)', 'tMap_1', 'Salida (Main)', 'Order_by_capital', 'row2 (Main)', and 'tMap_2'. From 'tMap_2', the data is split into three parallel paths: 'Escribir_BD (Main order:2)' leading to a 'MYSQL' target, 'Escribir_excel (Main order:1)' leading to a 'tFileOutputExcel_2' target, and 'Log_por_pantalla (Main order:3)' leading to a 'tLogRow_1' target.

Below the workspace, the 'Trabajo JB_EJEMPLO_LITUUS' execution control panel is visible. It includes a 'Run' button highlighted with a red box, along with 'Kill' and 'Clear' buttons. The central execution area contains the text 'Presionamos el botón Run'. On the left, there are tabs for 'Basic Run', 'Debug Run', 'Advanced settings', 'Target Exec', and 'Memory Run'. On the right, there is a table with columns 'Nombre' and 'Valor'.

Nombre	Valor



Podemos ver los resultados en Talend o desde su Gestor de BD preferido



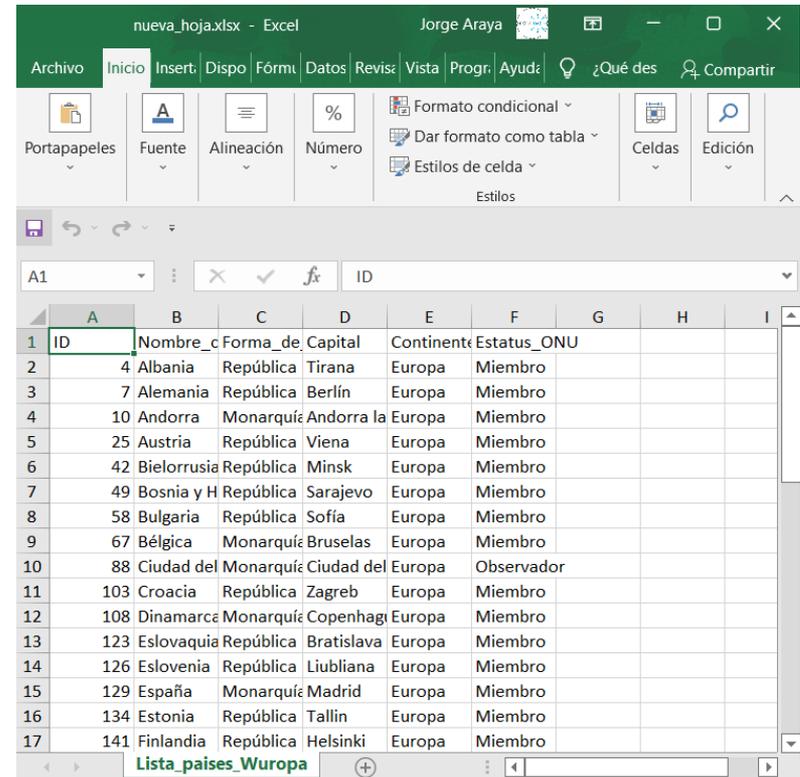
Verificamos el Excel tanto en formato como nombre de las hojas

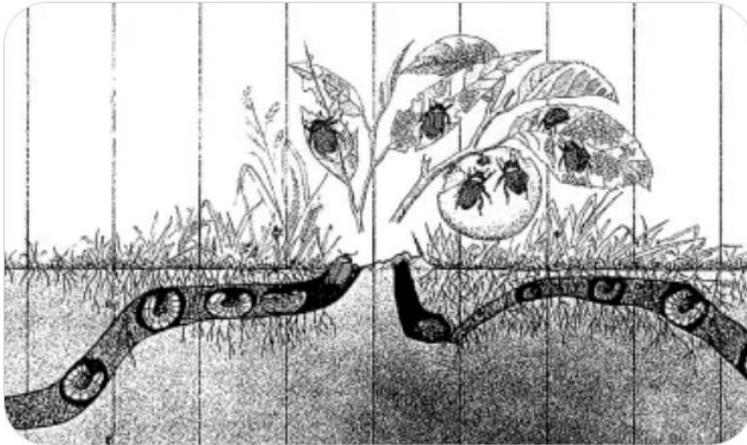
```
Mostrando filas 0 - 24 (total de 189, La consulta tardó 0,0000 segundos)
SELECT * FROM `PAISES`
```

1 > >> | Mostrar todo | Número de filas: 25

+ Opciones

ID	NOMBRE	CAPITAL
2	Afganistán	Kabul
5	Albania	Tirana
8	Alemania	Berlín
11	Andorra	Andorra la Vieja
13	Angola	Luanda
15	Arabia Saudita / Arabia Saudí	Riad
17	Argelia	Argel
19	Argentina	Buenos Aires
21	Armenia	Ereván
23	Australia	Canberra
26	Austria	Viena

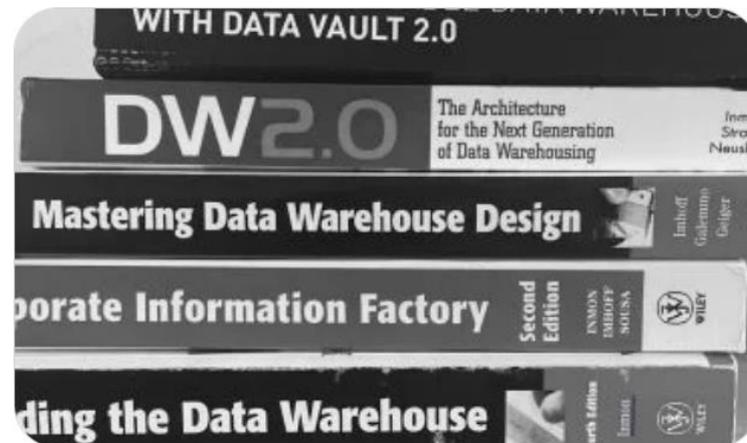




¿Qué es la visualización de datos cuantitativos y por qué es importante?

11/03/2021 Lenin González

Intentaré convencerte de por qué debemos aprender visualización de datos cuantitativos. Verás ejemplos, un poco de historia y cuatro recomendaciones prácticas que puedes aplicar desde hoy mismo.



¿Qué es un Data Warehouse y Para Qué Sirve?

18/01/2021 Lenin González

En la medida que las empresas requieren ser más competitivas suelen incorporar Software que les permita gestionar sus procesos de negocios de forma tal que se vuelvan predecibles y optimizables. Cada proceso de negocio tiene su complejidad y vida propia, por lo que sin quererlo ni beberlo, las organizaciones comienzan adquirir más Software o Módulos encargados de administrar sus procesos de forma atómica, durable y concurrente. Suele suceder que dichos sistemas/módulos no usan las mismas tecnologías, puede que no sean del mismo proveedor o sus arquitecturas sean incompatibles. Es Aquí cuando un Data Warehouse puede aportar credibilidad en la toma de decisiones al integrar en una sola fuente de la verdad nuestros dato



Presentado por:

- Lenin González S.
- Director General, CEO
- +569 3072 9405
- lgonzalez@lituus.cl

