



Data Warehouse

Praktikum

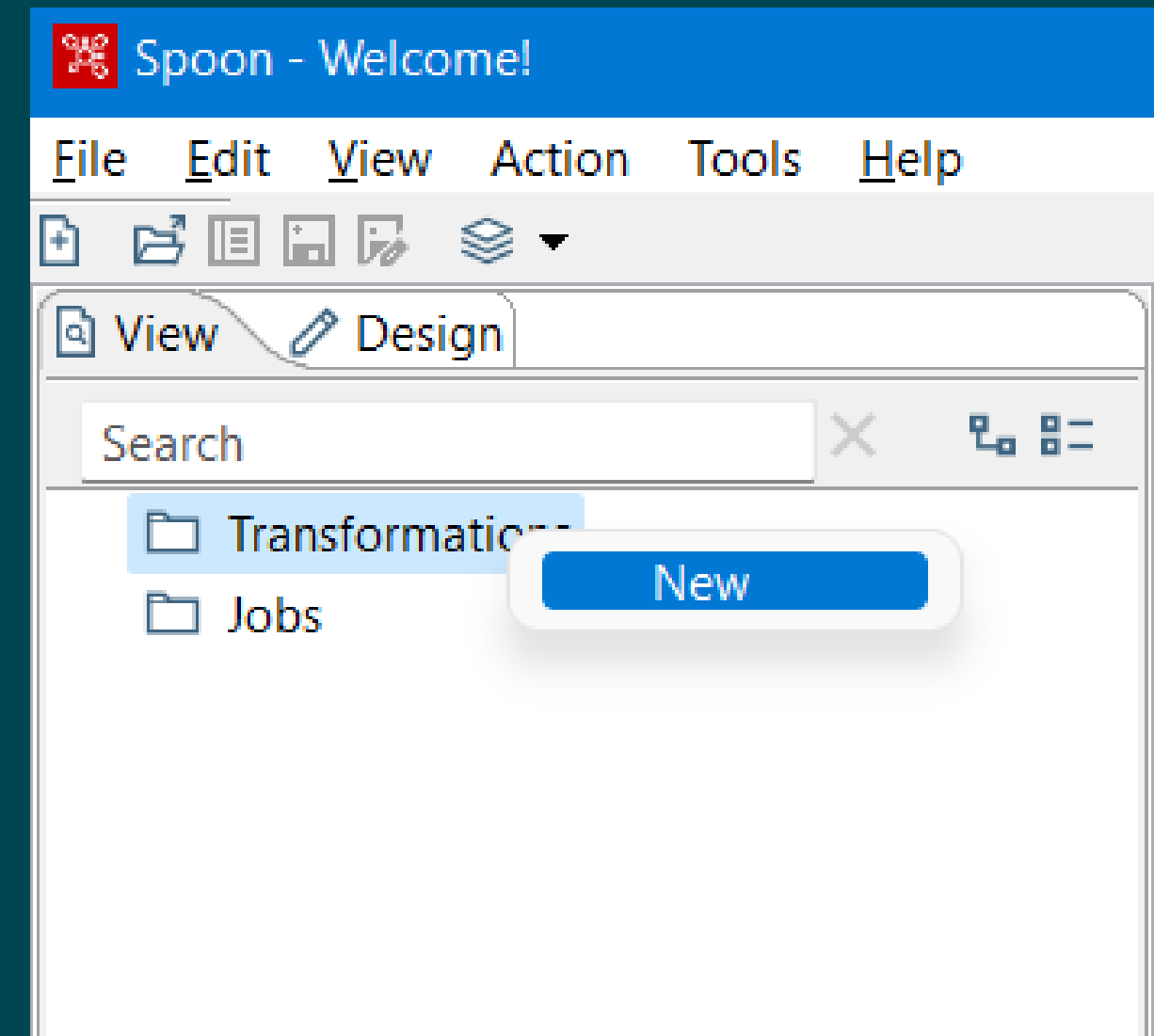
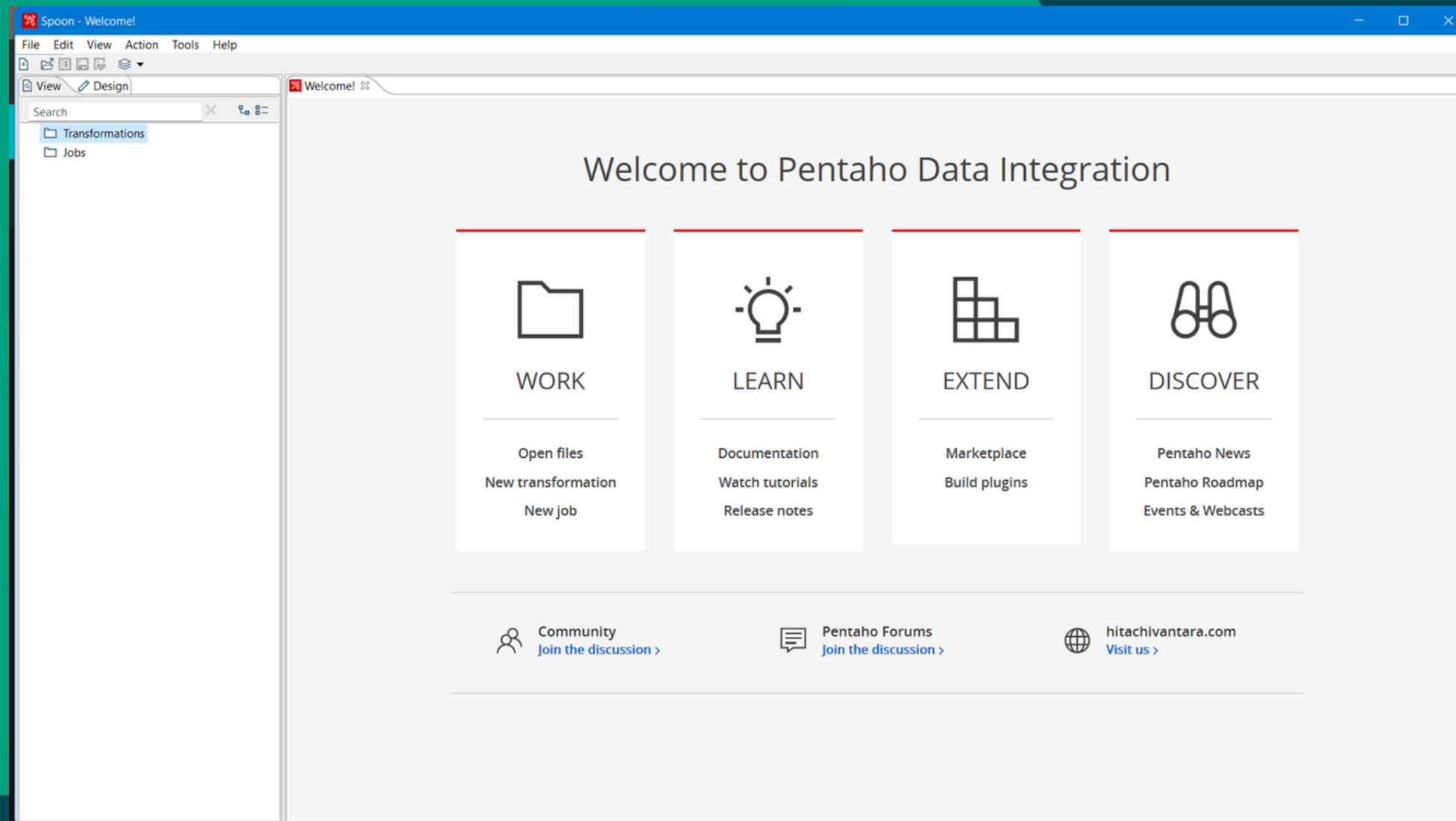
ETL

WEEK-13

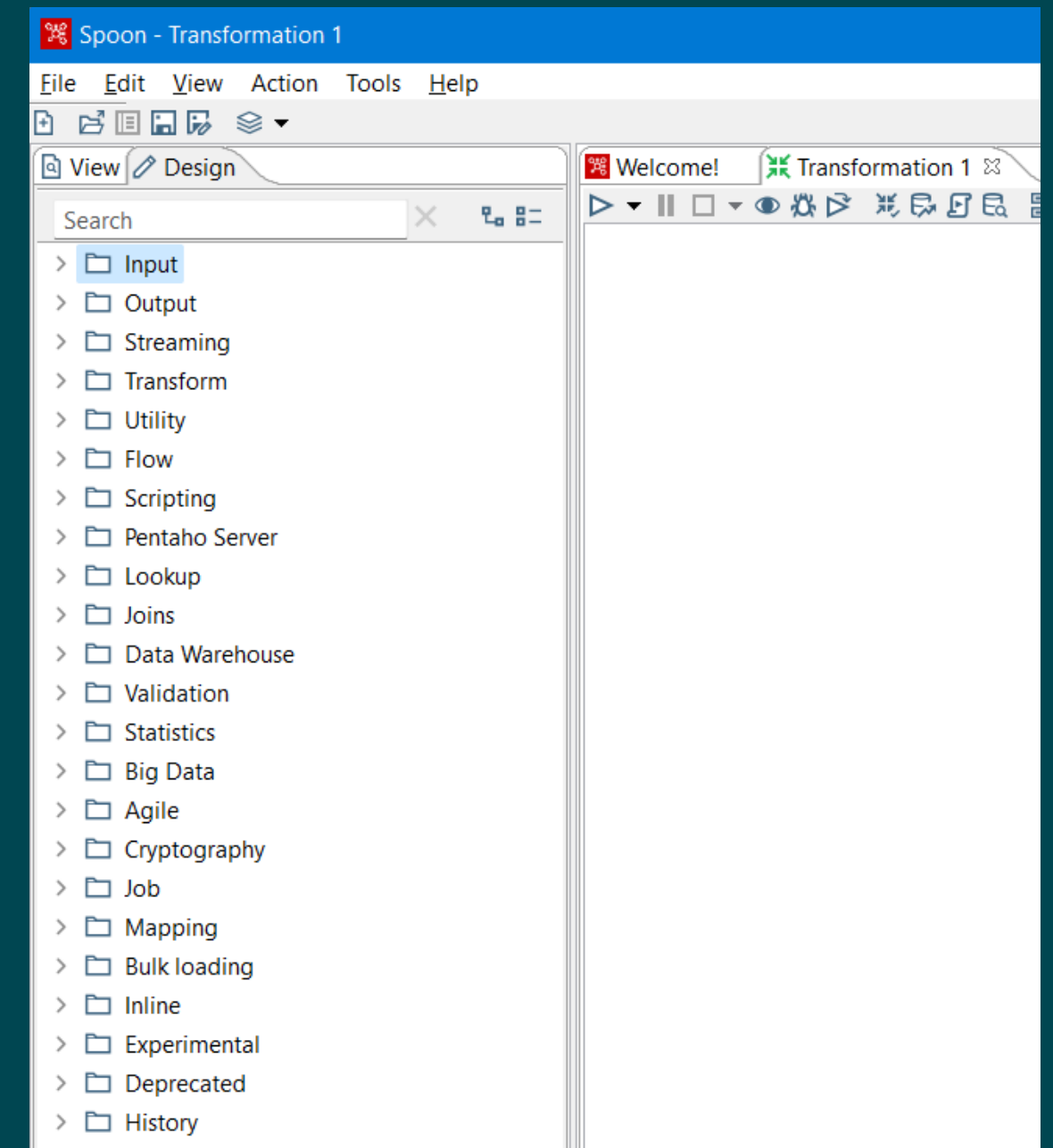
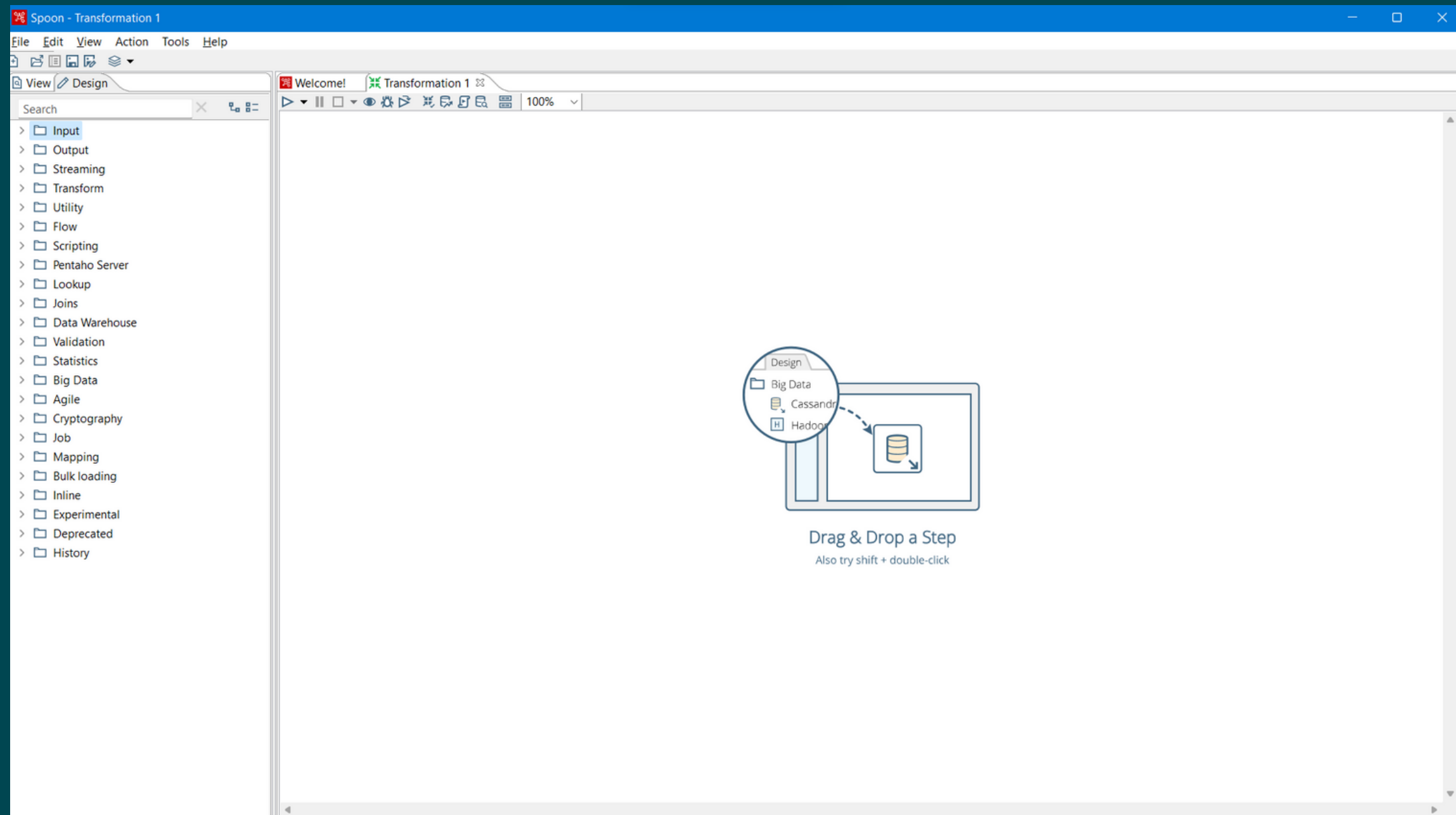
1. Download data CSV pada link berikut

https://drive.google.com/drive/folders/1p1Wq5BV53osxG5PlsETTh7sJa1_5hcu83?usp=share_link

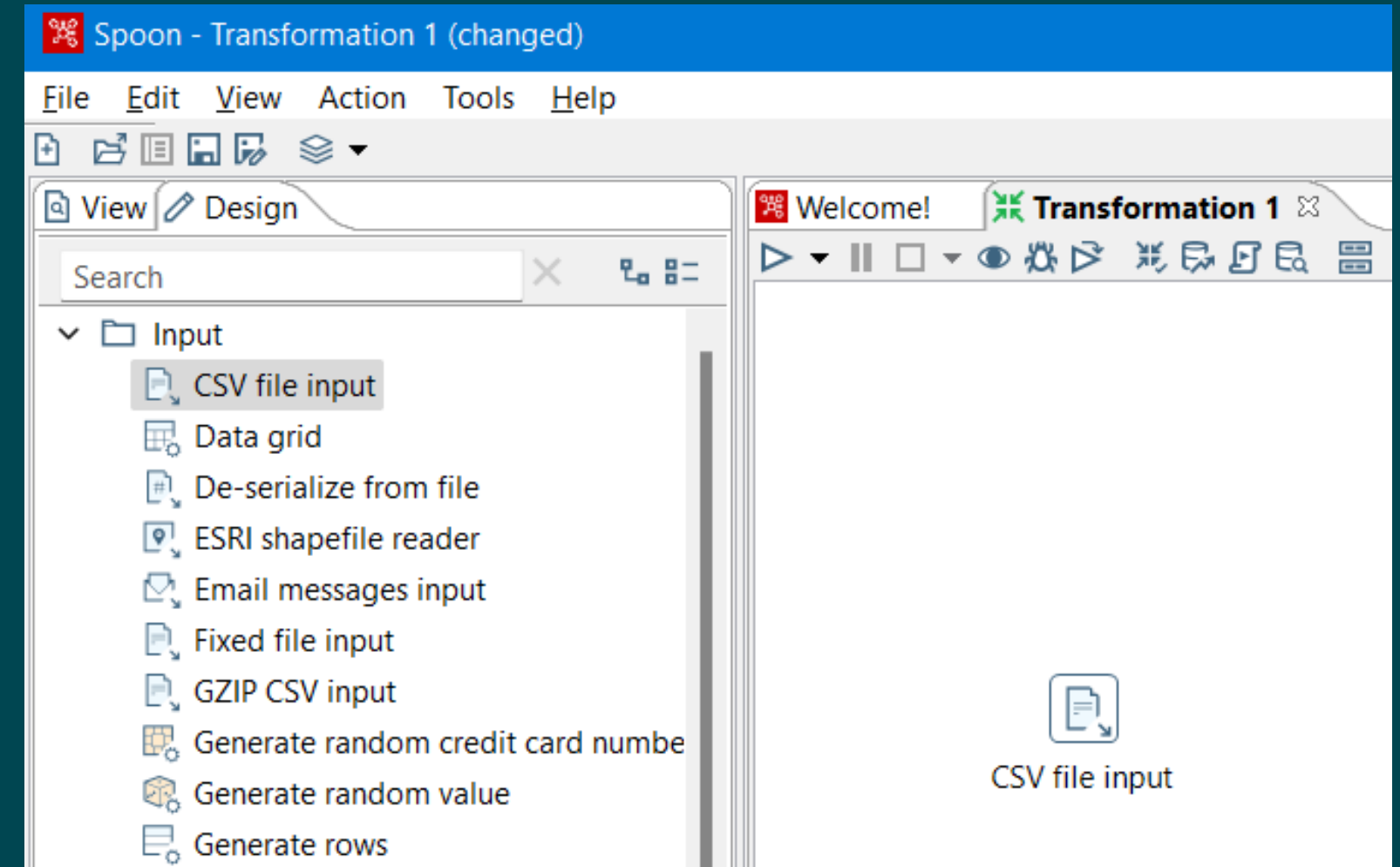
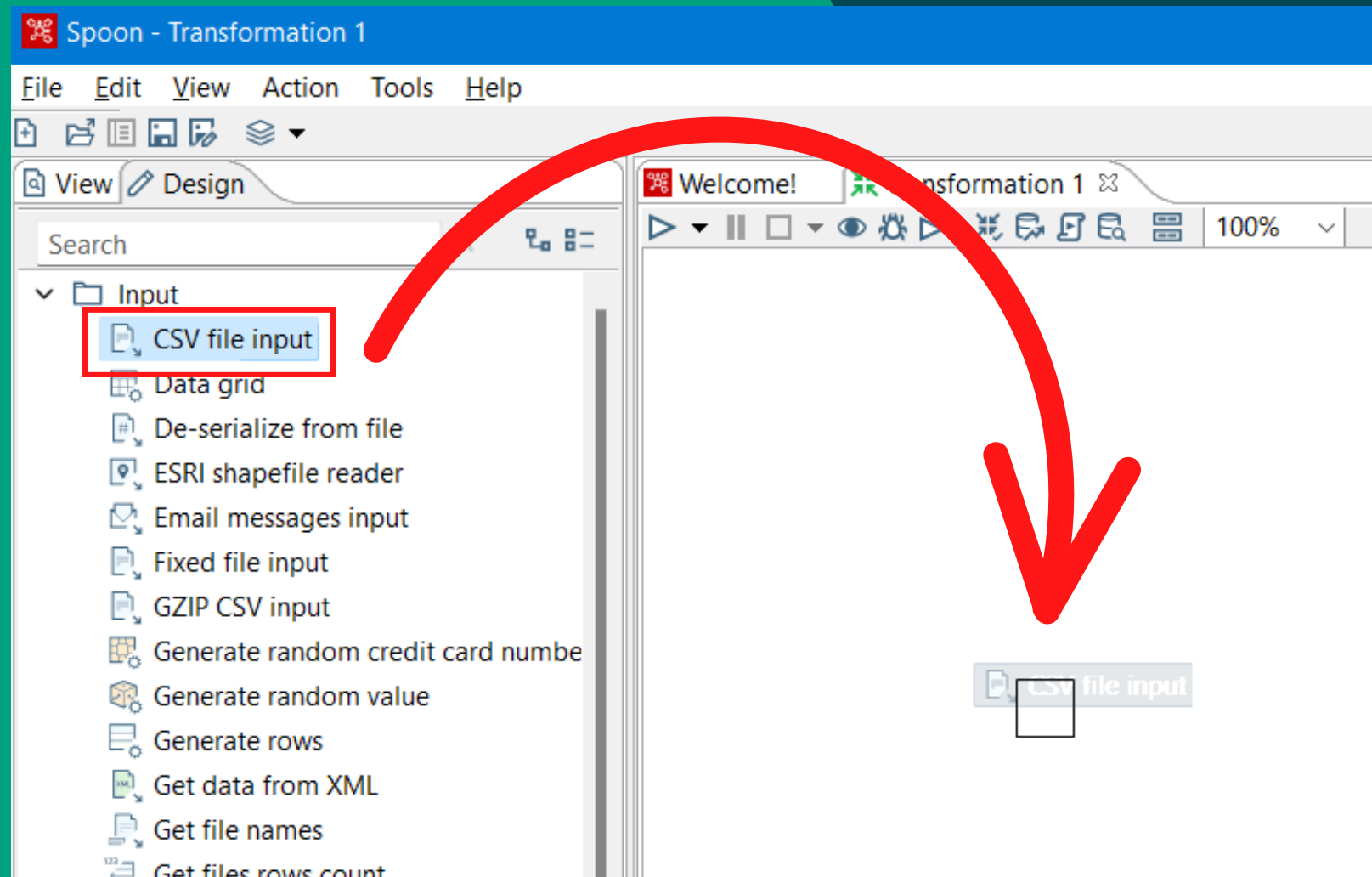
2. Buka apk pentaho kemudian pilih **Transformation**, klik kanan pilih **New**



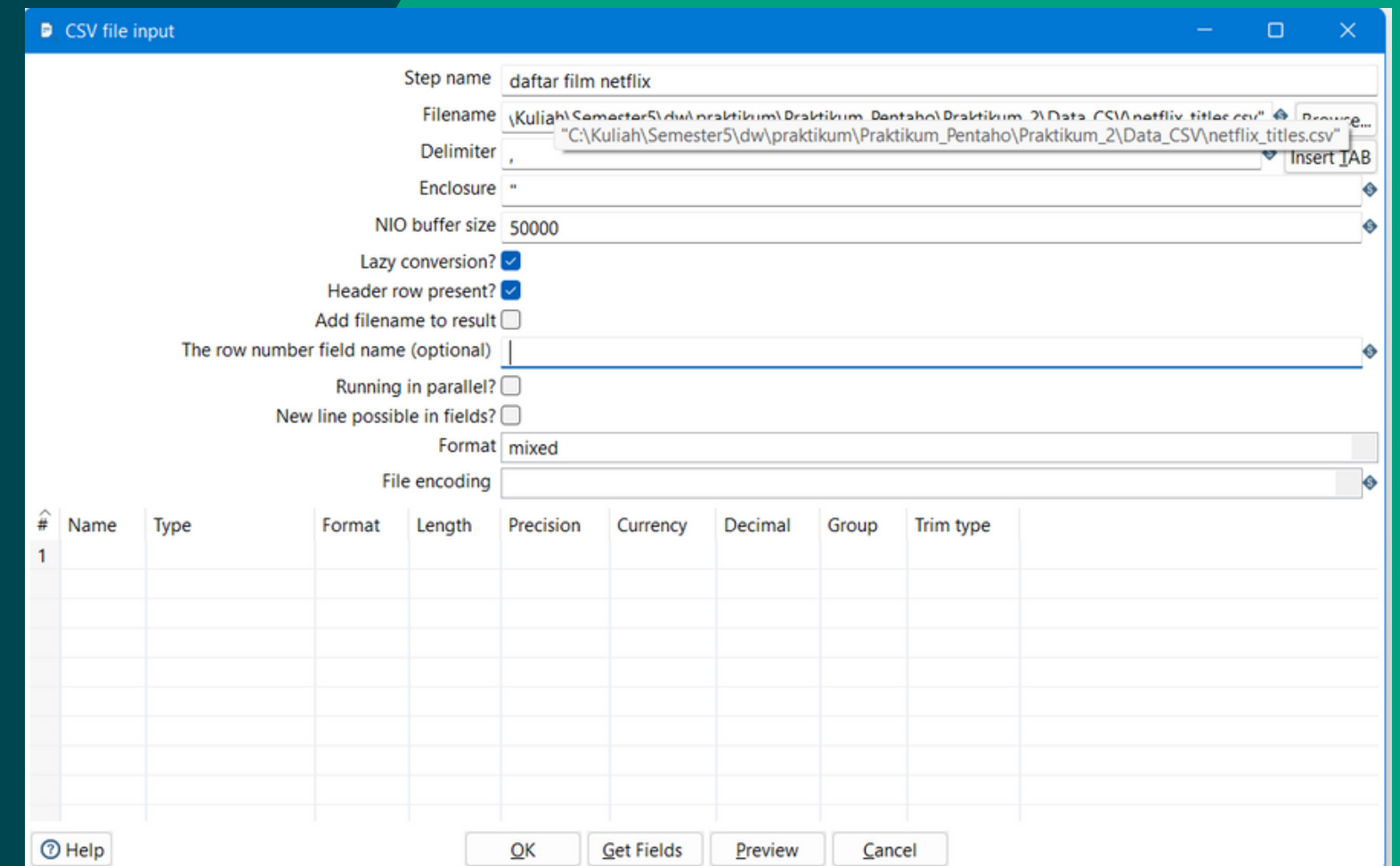
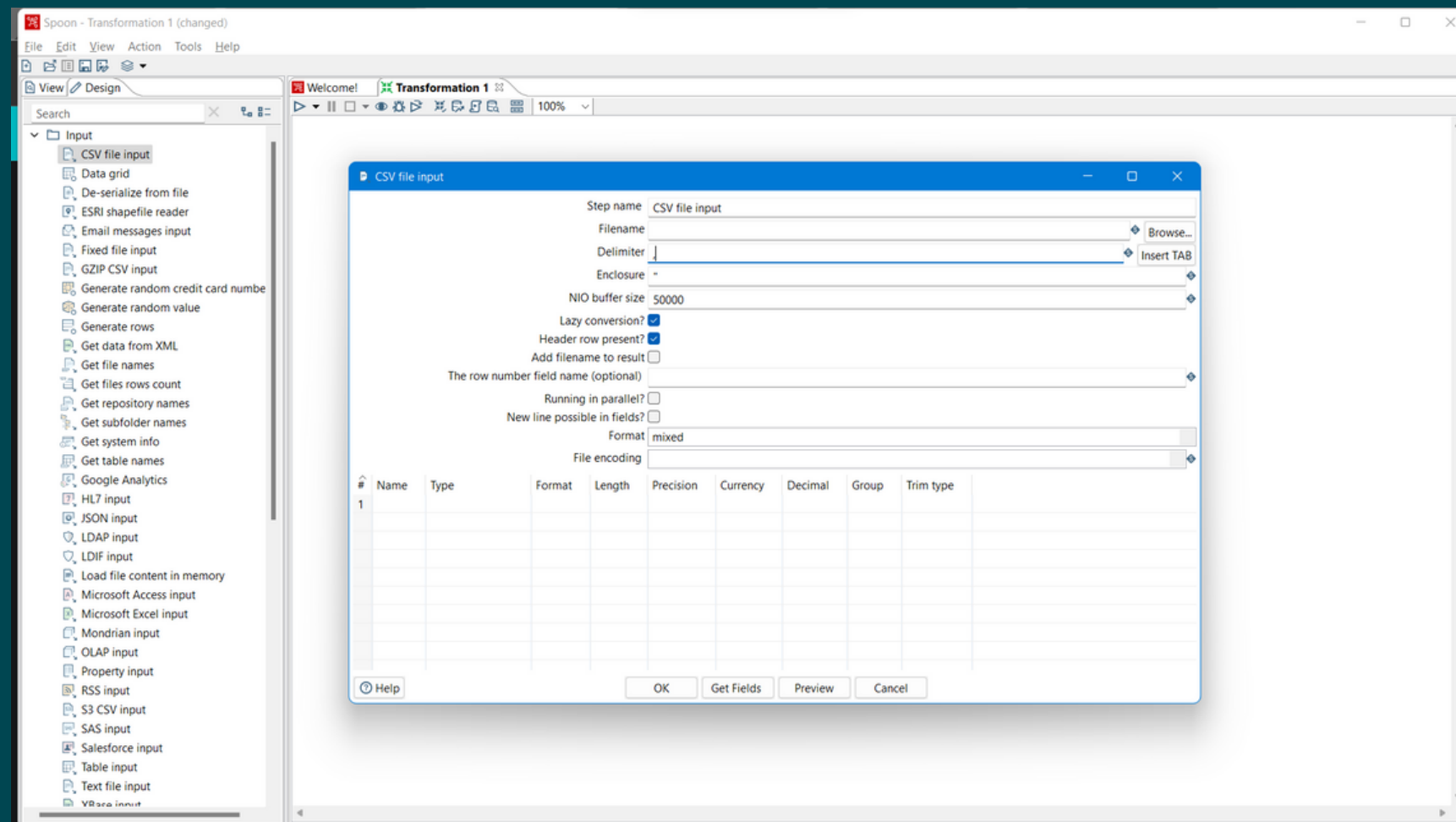
3. Kita akan melakukan tranformasi data dari file **CSV** ke **XML**. Pada PDI dapat dilihat pada bagian kiri ada panel yang menampilkan jenis Steps yang disediakan. Step adalah elemen penyusun transformasi, yang dapat berupa input, proses atau output.



4. karena input berformat csv dan outputnya XML, maka pilih direktori Input dan pilih step CSV file lalu drag



5. Sekarang kita akan mengkonfigurasi step ini, double klik step “CSV file input”. Akan muncul dialog seperti gambar dibawah. Isi step name dan klik tombol “**Browse**” untuk memilih file csv yang sebelumnya telah di download (Pilih File **netflix_titles.csv**) [link download](#)



Note : Jika tombol "**Browse**" tidak bisa di klik lakukan copy path secara manual di directory tempat menyimpan file csv tadi kemudian paste

5. Jika langkah - langkah tadi sudah berhasil dilakukan, klik tombol **Get Fields**

Step name: daftar film netflix

Filename: C:\Kuliah\Semester5\dw\praktikum\Praktikum_Pentaho\Praktikum_2\Data_CSV\netflix_titles.csv

Delimiter: ,

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

File encoding:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1									

Buttons: Help, OK, **Get Fields**, Preview, Cancel



Step name: Daftar Film Netflix

Filename: C:\Kuliah\Semester5\dw\praktikum\Praktikum_Pentaho\Praktikum_2\Data_CSV\netflix_titles.csv

Delimiter: ,

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

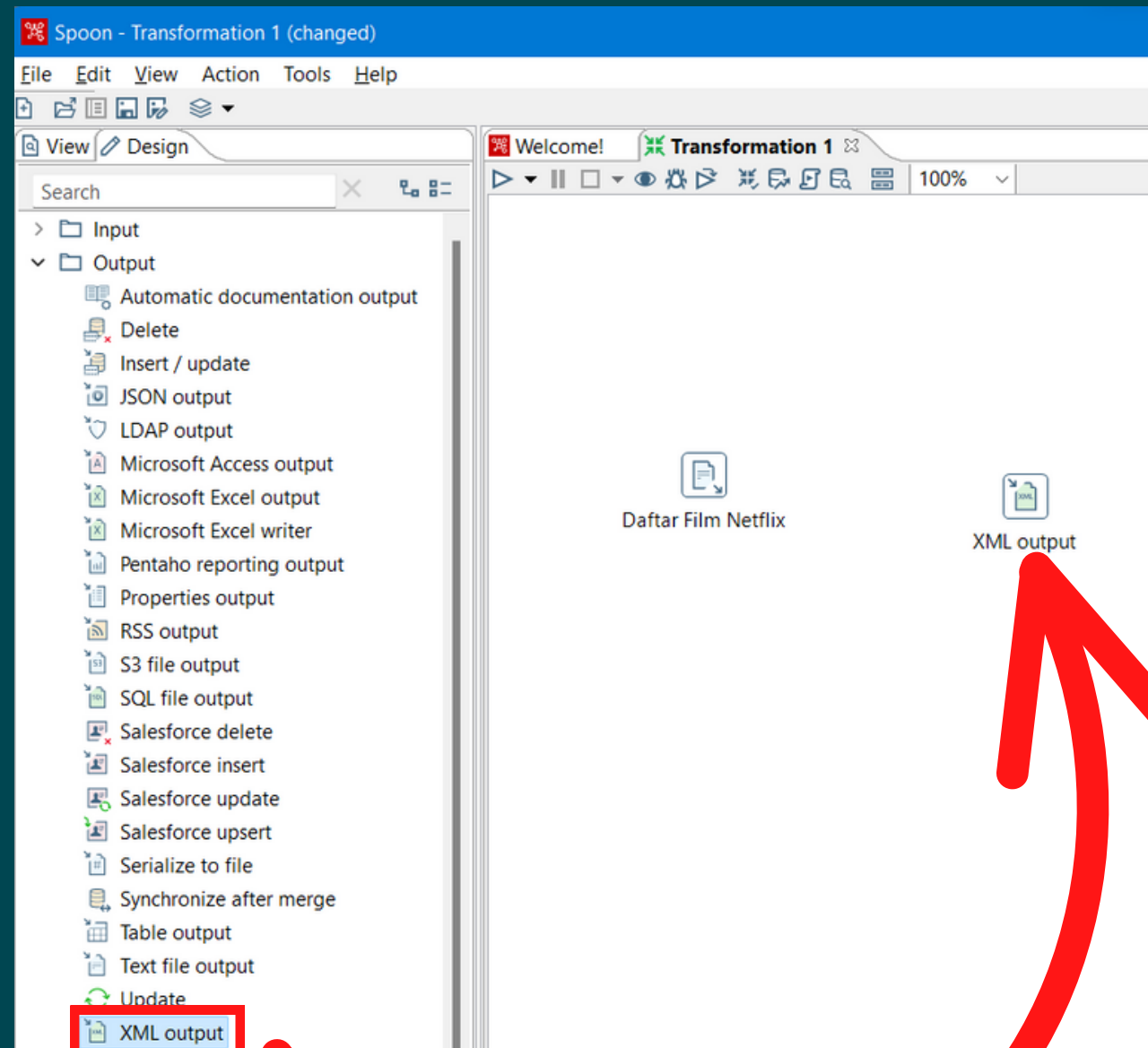
File encoding:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	show_id	String		4		Rp	,	.	none
2	type	String		7		Rp	,	.	none
3	title	String		58		Rp	,	.	none
4	director	String		54		Rp	,	.	none
5	cast	String		532		Rp	,	.	none
6	country	String		69		Rp	,	.	none
7	date_added	String		18		Rp	,	.	none
8	release_year	Integer	#	15	0	Rp	,	.	none
9	rating	String		5		Rp	,	.	none
10	duration	String		9		Rp	,	.	none
11	listed_in	String		72		Rp	,	.	none
12	description	String		157		Rp	,	.	none

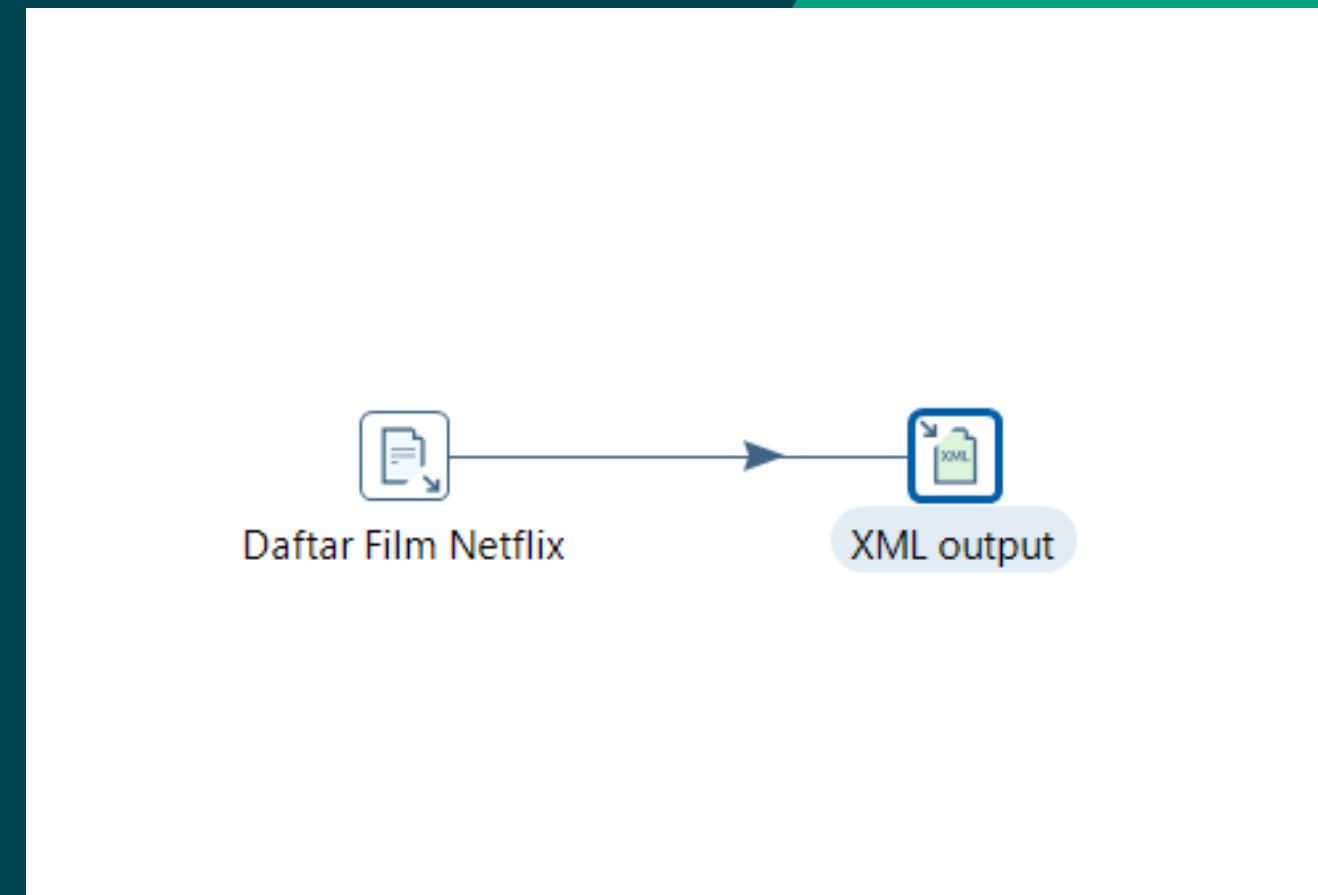
Buttons: Help, OK, Get Fields, Preview, Cancel

klik tombol **“Preview”** untuk melihat keluaran dari proses loading. Setelah selesai, tekan OK.

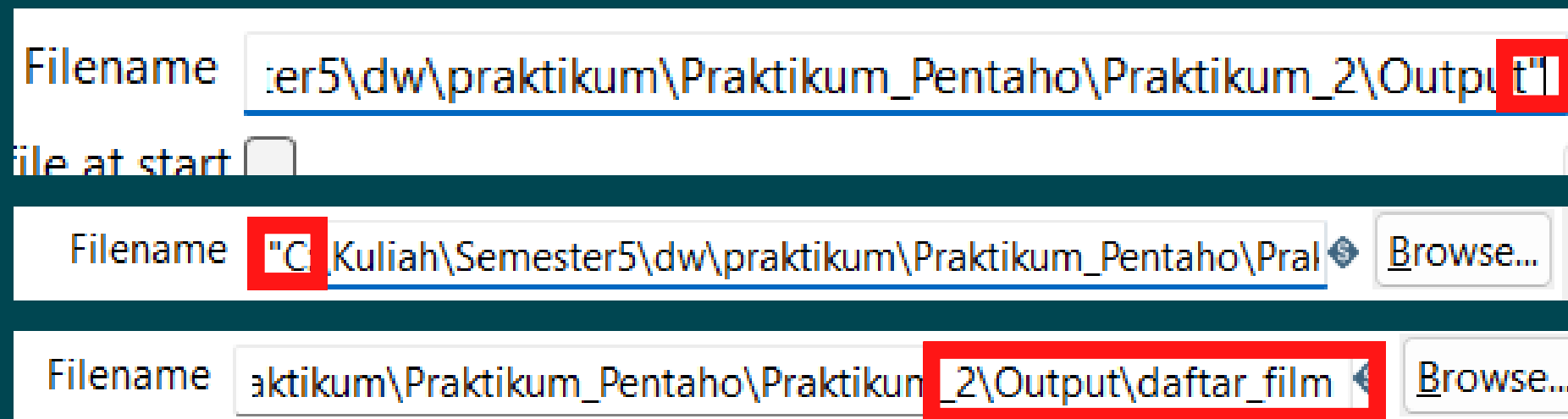
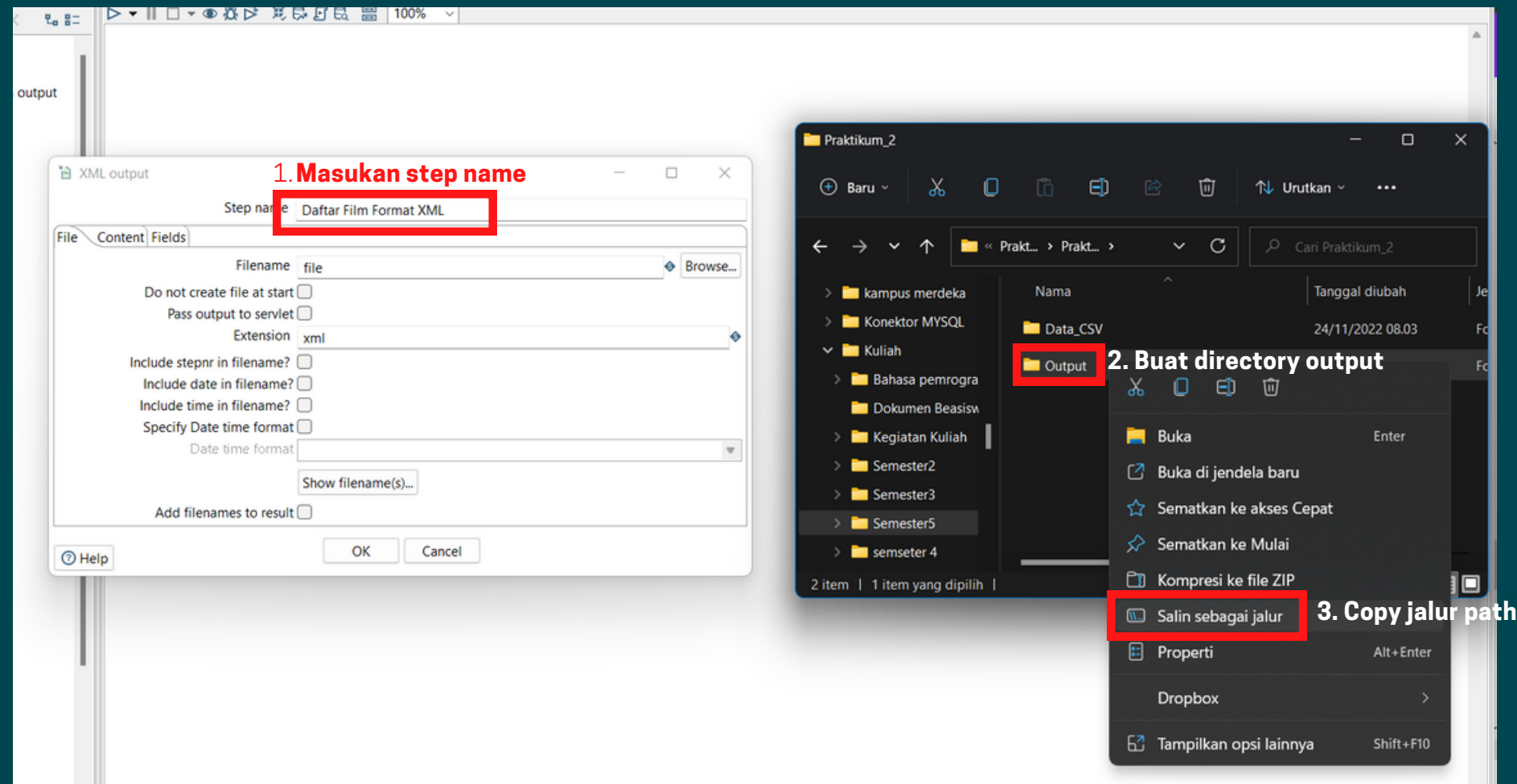
6. Selanjutnya kita akan menambahkan output XML. Pilih di panel Steps direktori output. Lalu pilih XML Output dan drag ke tab transformasi.



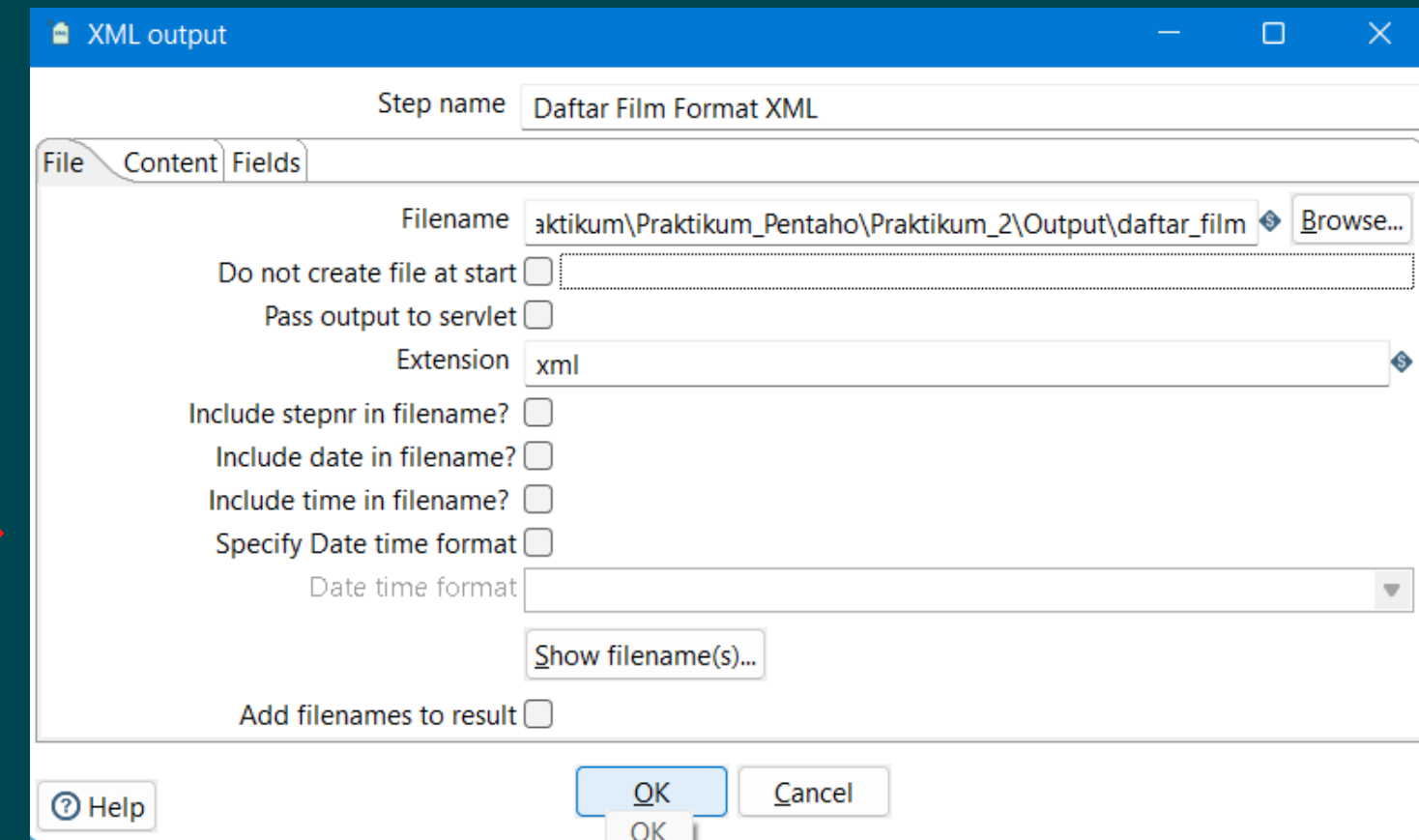
7. Selanjutnya kita akan menambahkan penghubung antara step input csv dan xml output. Dalam PDI, ini disebut Hop. Untuk menambahkan hop, klik csv input, tekan SHIFT dan drag ke xml output. Kemudian pilih “Main output of step” (gambar bawah).



8. Double klik XML output, masukan nama file output yang diinginkan

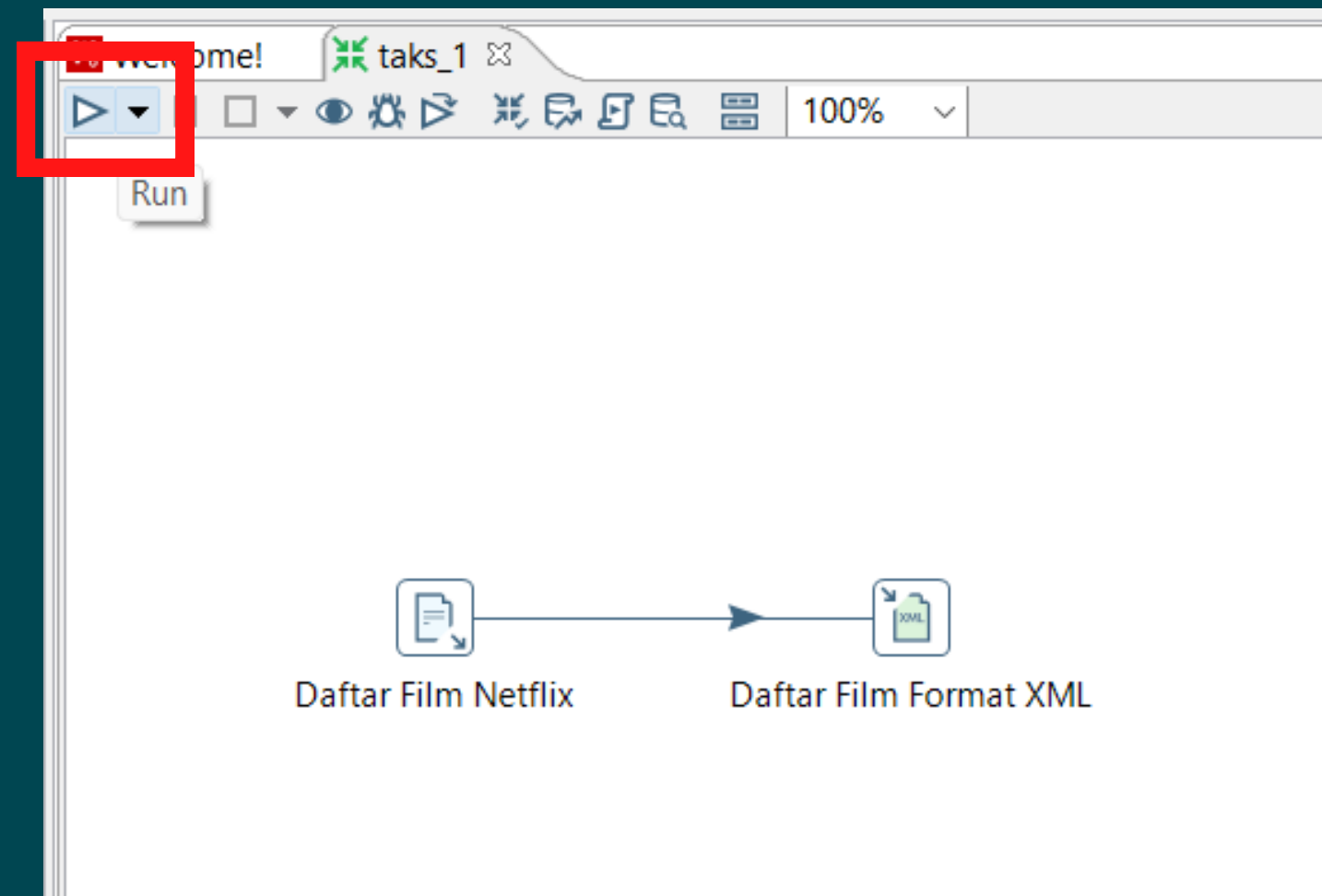
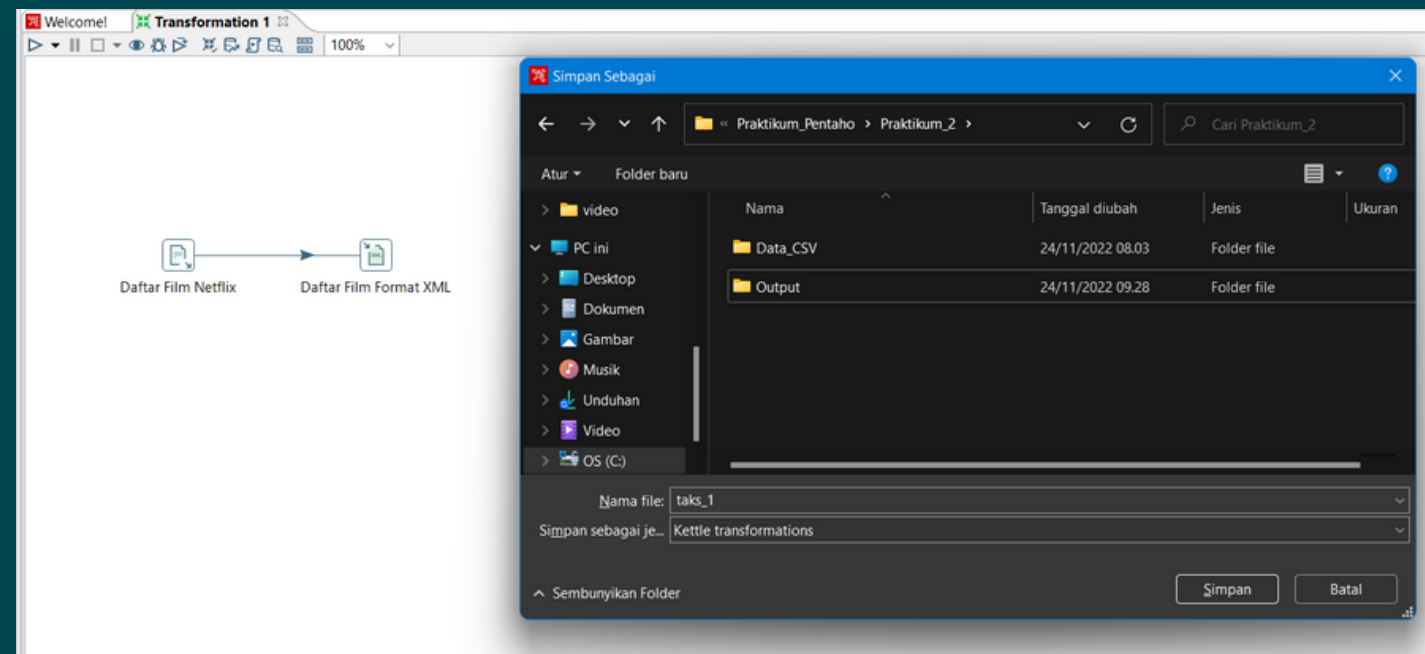


Paste jalur path tadi pada filename, **pastikan untuk menghapus tanda "(kutip)**. Kemudian tambahkan nama file outputnya.

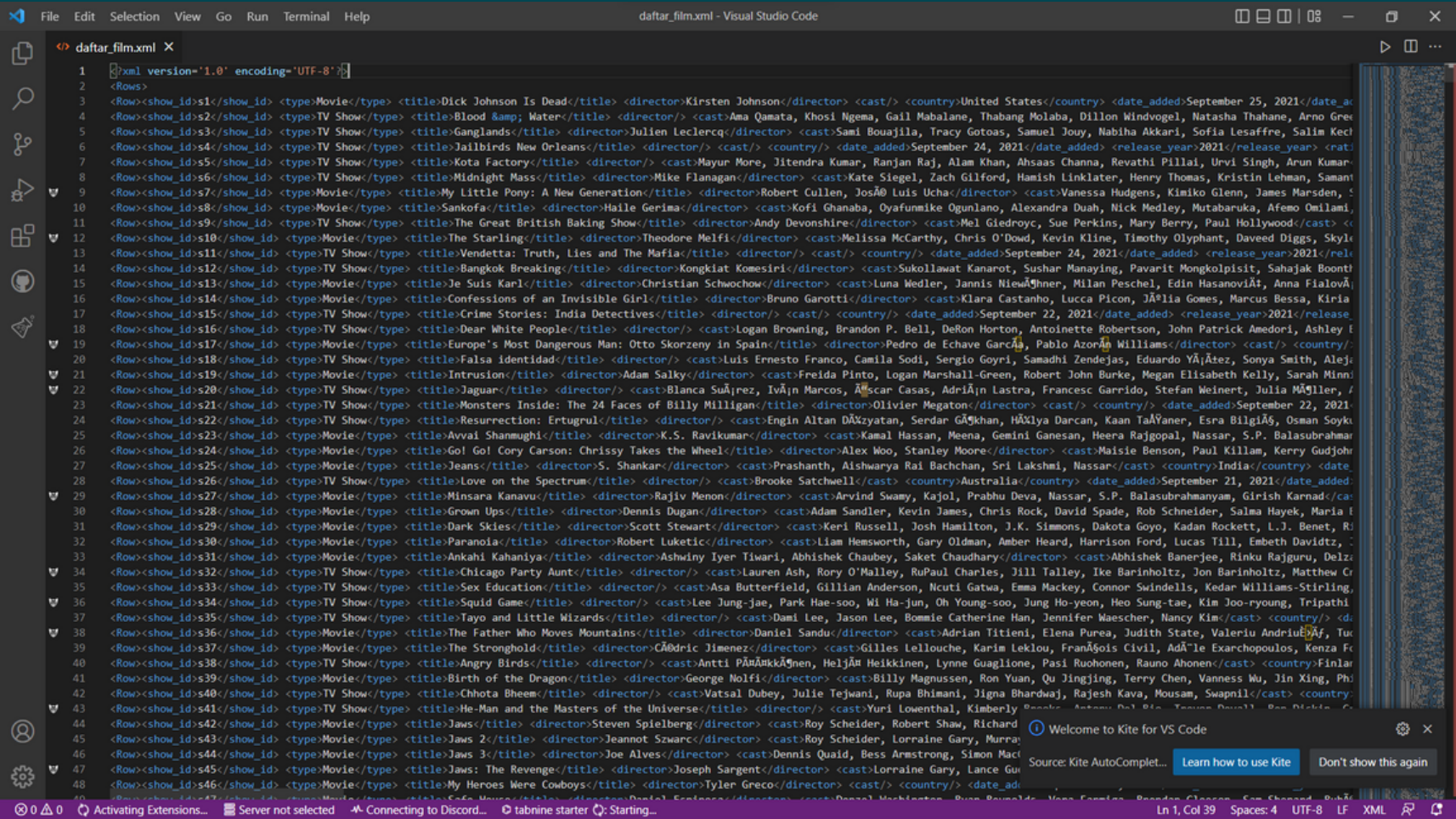
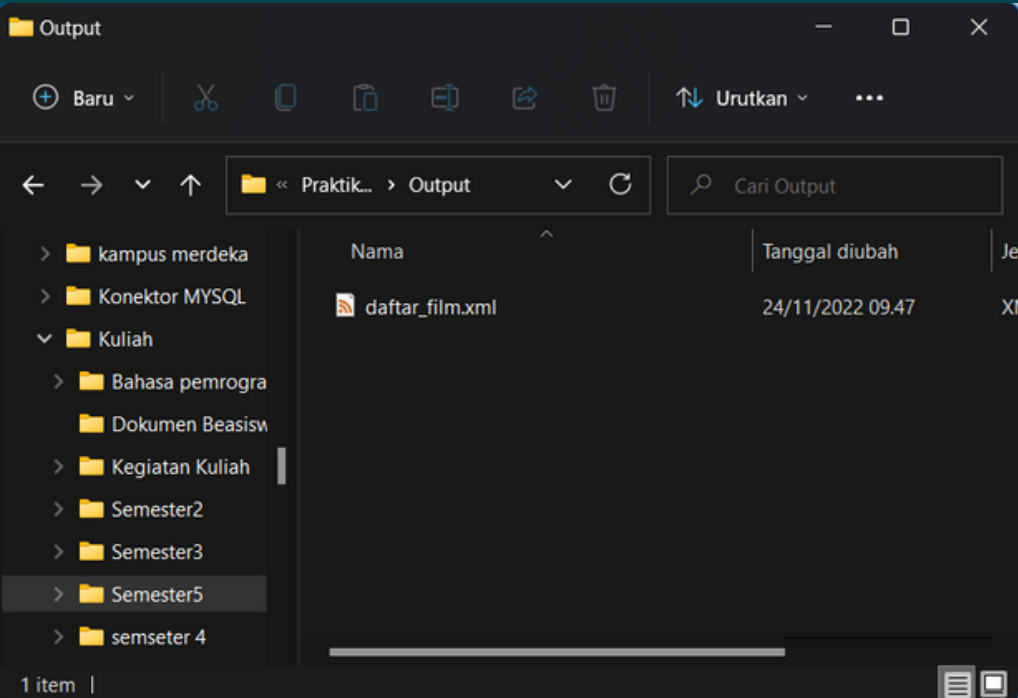
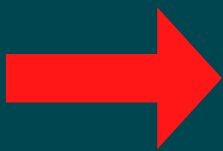
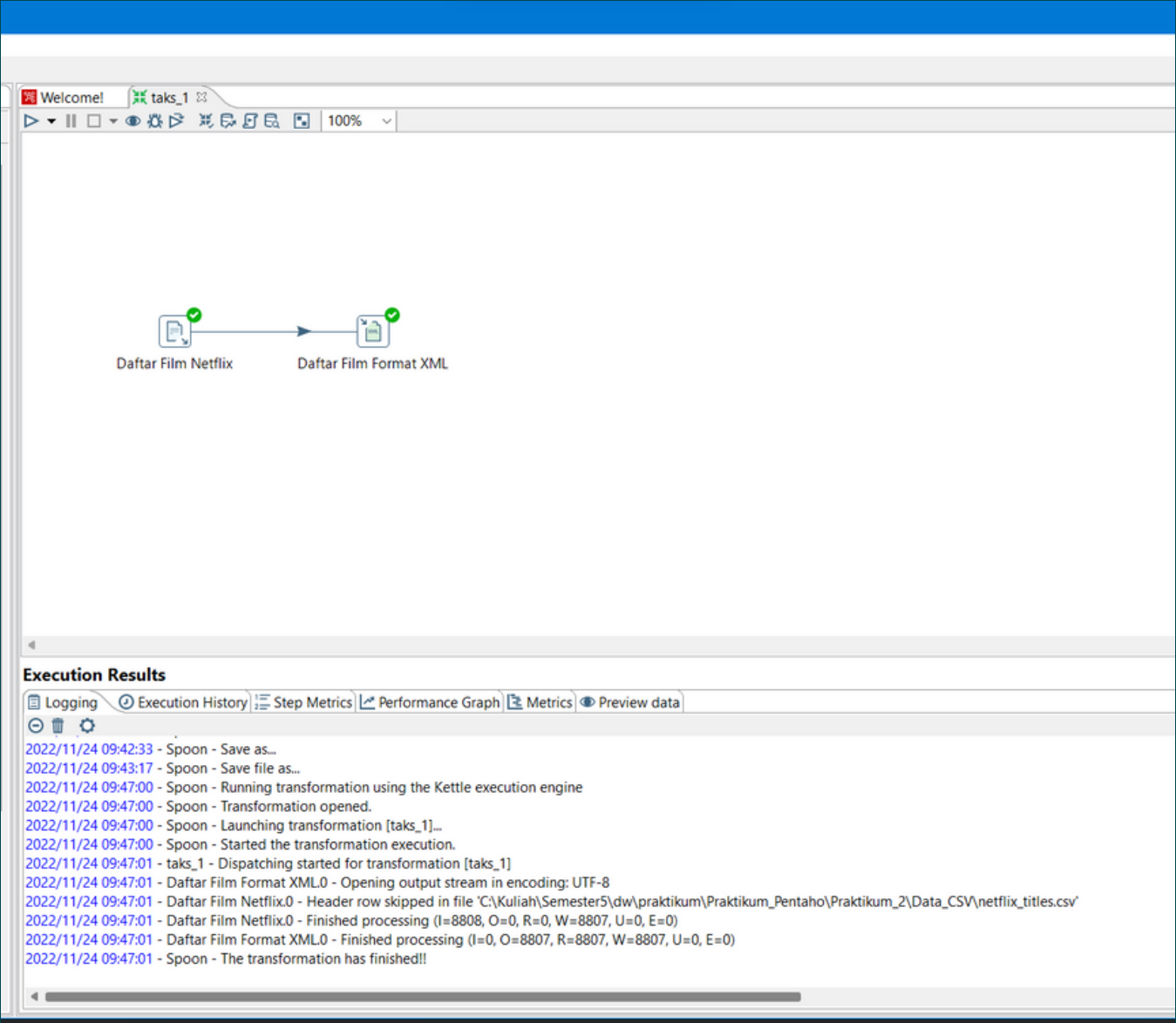


Langkah selanjutnya klik tombol **"ok"**

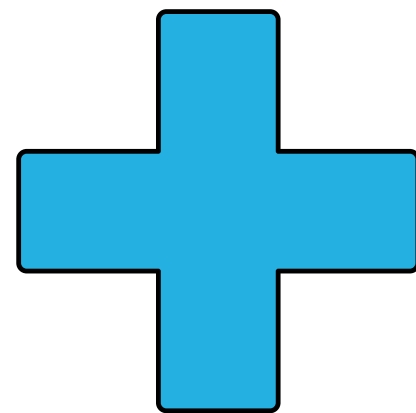
9. Sebelum mengeksekusi transformation, save transformation terlebih dahulu klik **ctrl + s** untuk melakukan save. Kemudian jalankan transformation dengan meklik icon "**play**"



10. Transformasi akan menghasilkan file XML out.xml, jika dilihat maka outputnya adalah sebagai berikut:



Konek to MYSQL



MYSQL

[Ke Halaman Agenda](#)

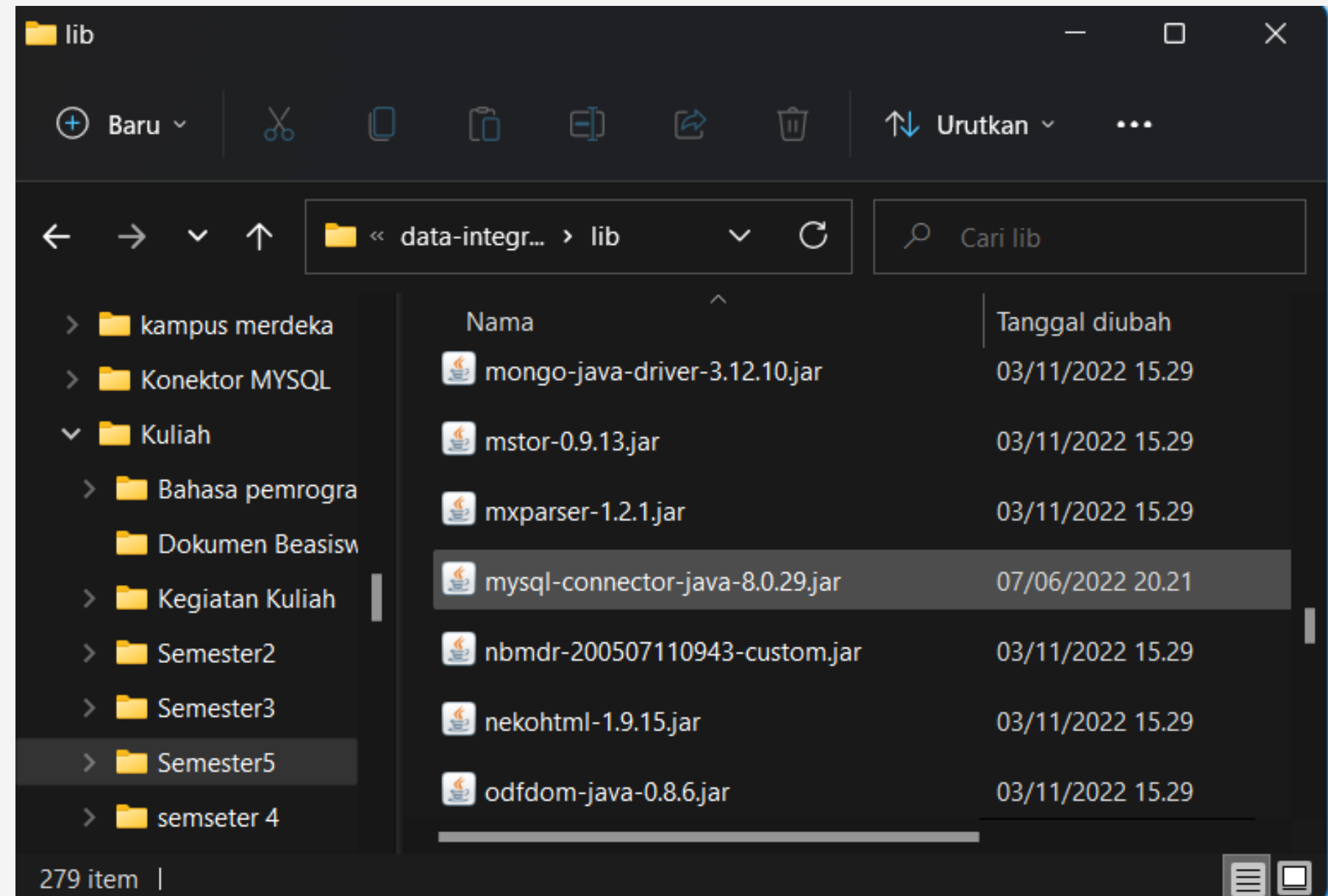
Connector MYSQL

Link Download

https://drive.google.com/file/d/1r0PNL-bsW1Q_3B2g40aEPD4NR4NeUD6d/view?usp=sharing

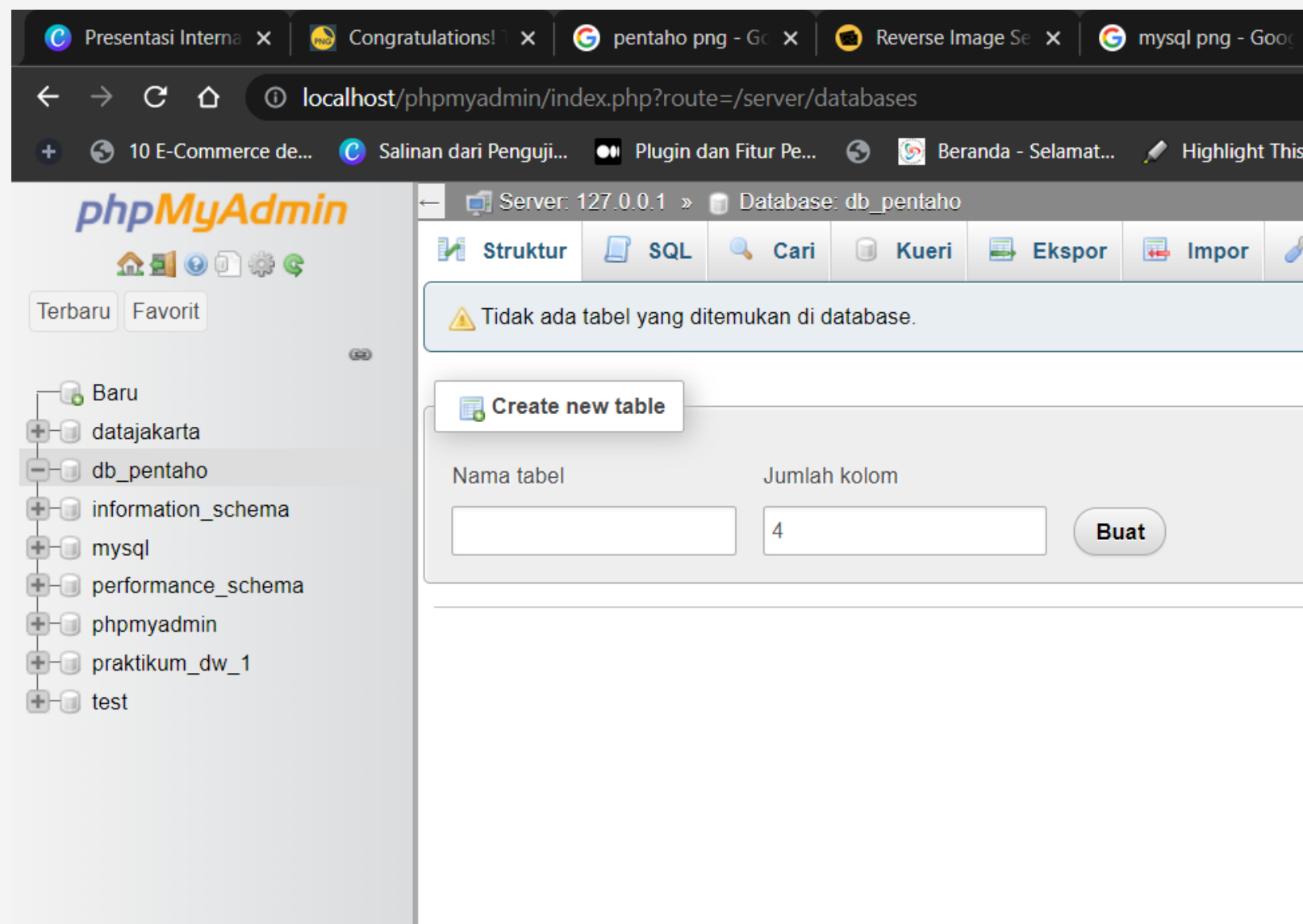
Pindahkan file mysql-connector-java-x.y.z-bin.jar ke direktori [pdi]\data-integration\lib (x,y,z adalah versi dari connector). Restart Spoon agar JDBC ini dapat digunakan.

[Ke Halaman Agenda](#)



Create Database

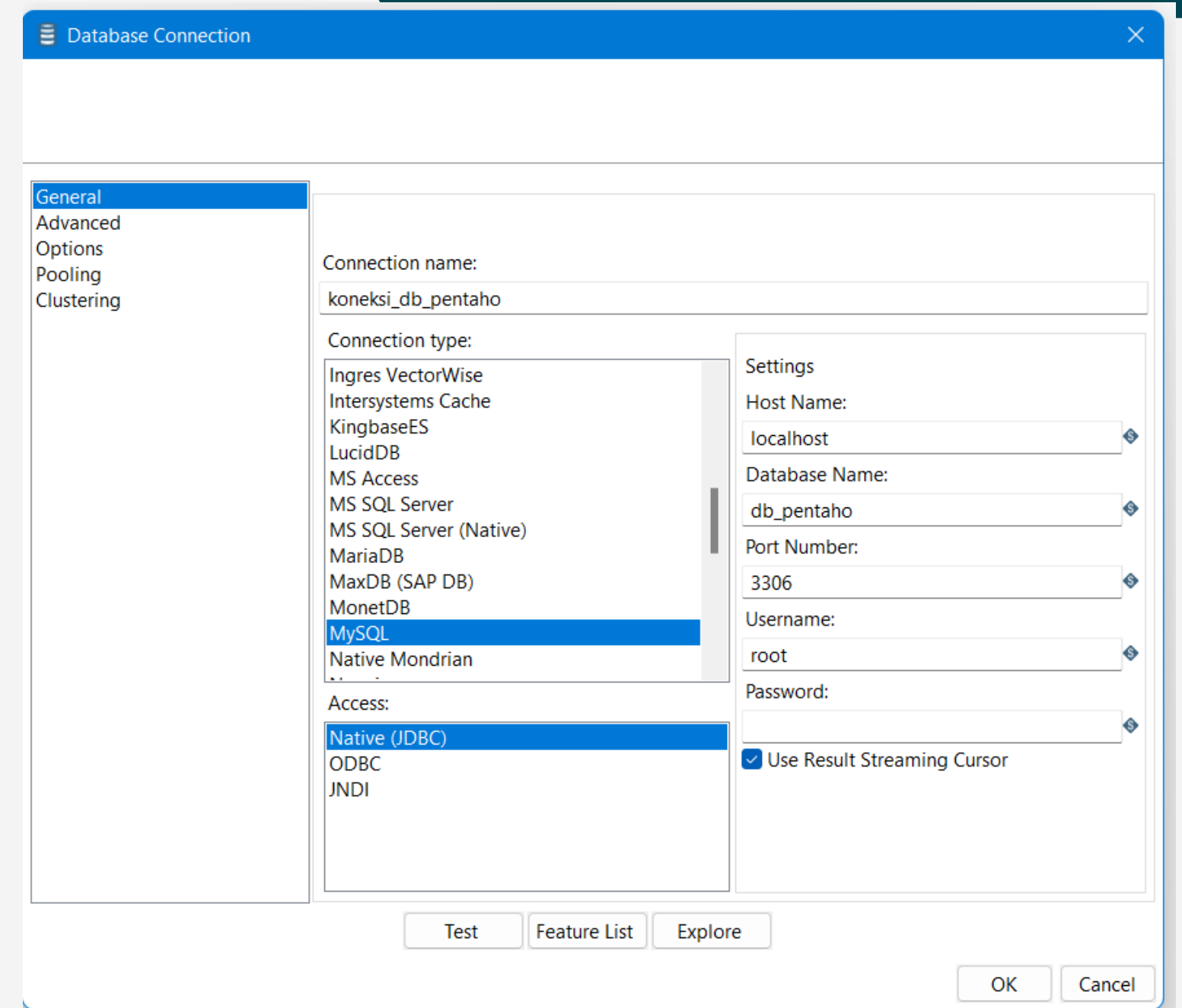
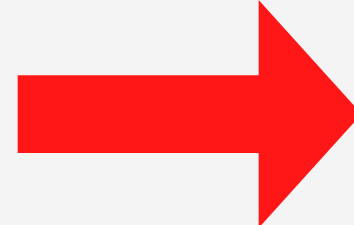
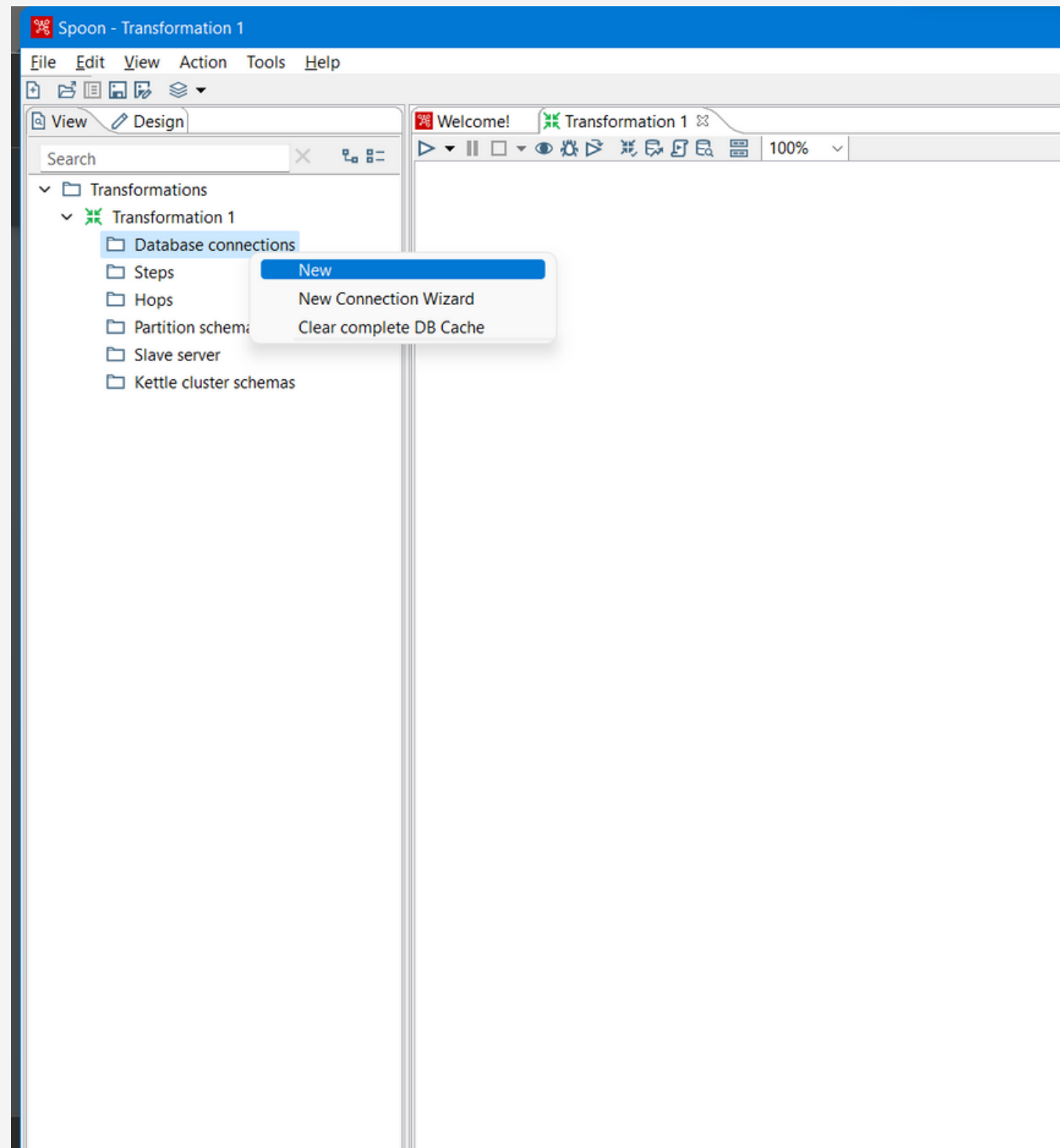
Untuk membuat database bisa menggunakan XAMPP/MYSQL Workbance/PotsgresSQL. Disini kita akan menggunakan XAMPP dengan PHPmyadmin-nya.



Buat Database db_pentaho

Connect to Database

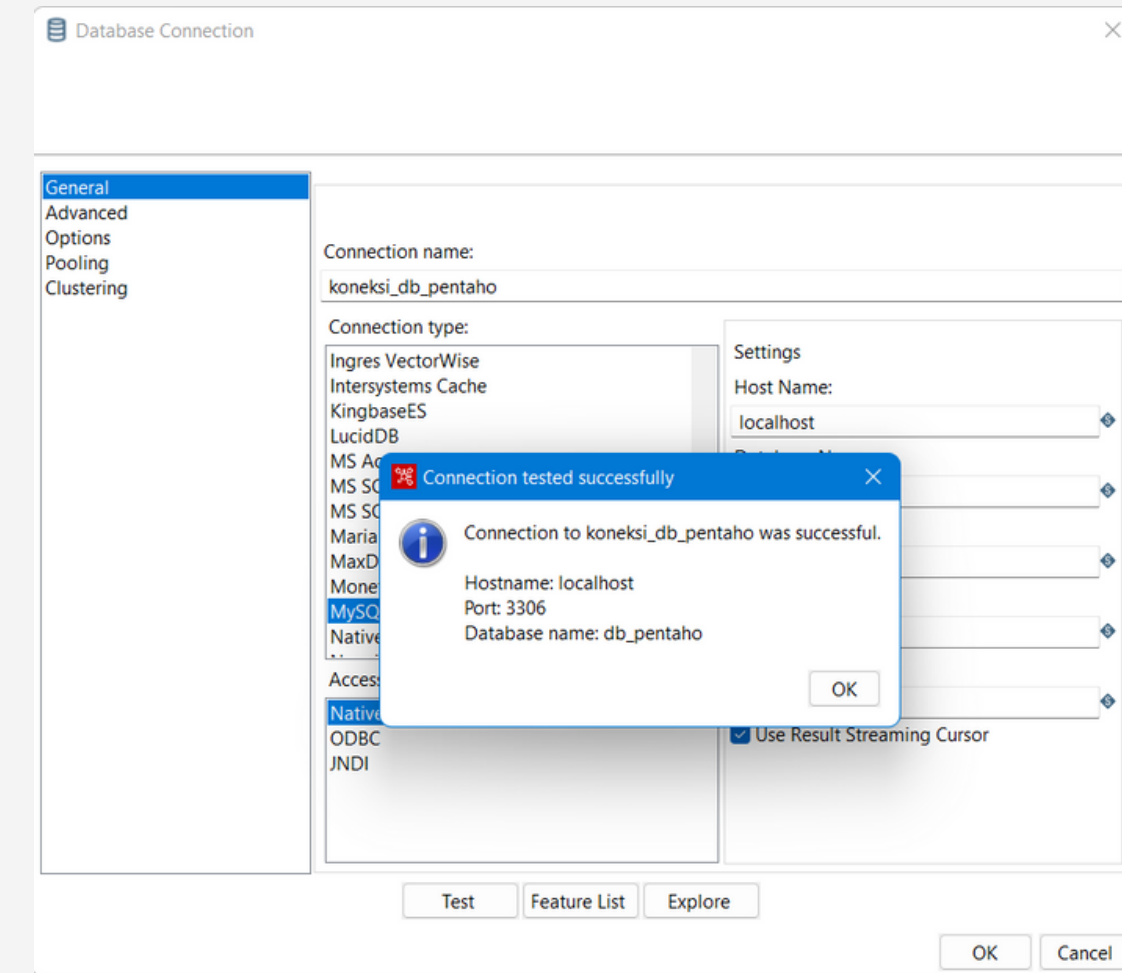
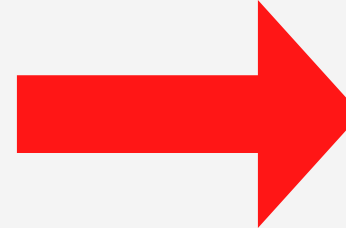
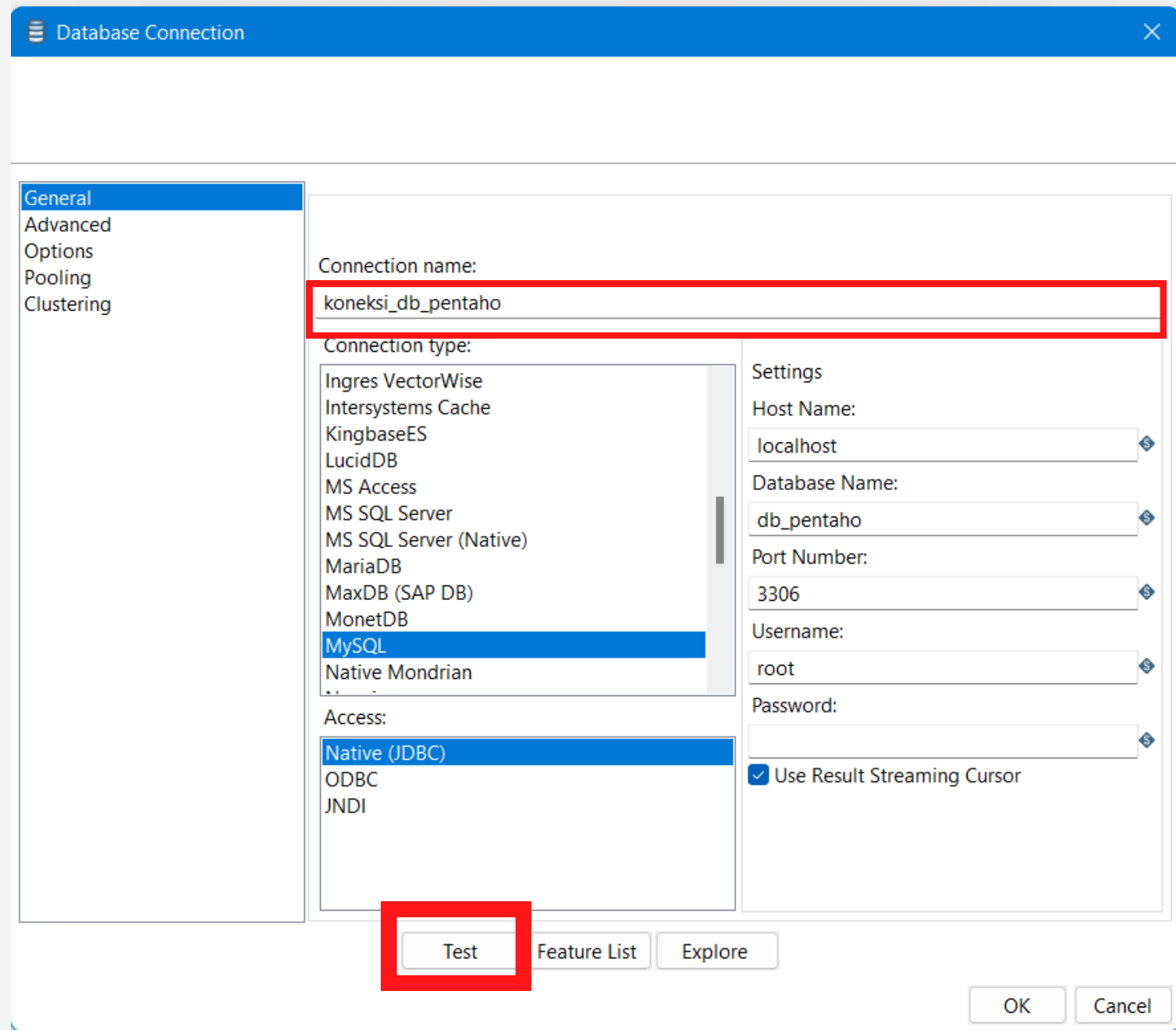
Kembali ke Spoon, buatlah transformasi baru lalu klik **"View"**, pilih database connections klik kanan lalu pilih new



Pilih MySQL sebagai connection type. Perhatikan juga tipe koneksi yang lain yang dapat digunakan. Isi nama host, nama database, port (biasanya tidak perlu diubah), username dan password. Jika menggunakan XAMPP, default username adalah 'root' dengan password dikosongkan.

Connect to Database

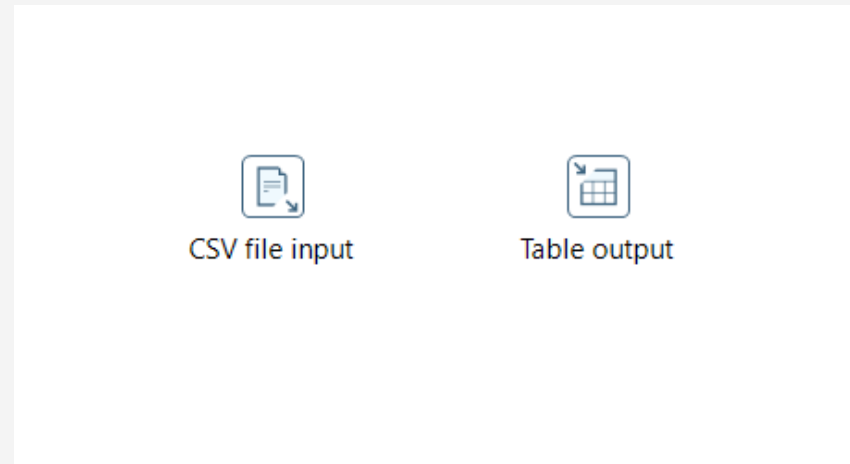
Untuk memastikan koneksi sudah berhasil, tekan tombol test.
Jangan lupa mengisi connection name



Akan muncul dialog seperti gambar diatas jika koneksi berhasil

Input & Table Output

Sama seperti langkah sebelumnya tambahkan step input csv.
Sedangkan untuk output, pilih Table output



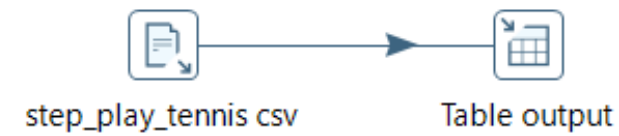
Sama seperti langkah sebelumnya Sekarang kita akan mengkonfigurasi step ini, double klik step “**CSV file input**”. Akan muncul dialog seperti gambar dibawah. Isi step name dan klik tombol “Browse” untuk memilih file csv yang sebelumnya telah di download (Pilih File **play_tennis.csv**) link download. Jika sudah lakukan **Get Fields**

The screenshot shows the 'CSV file input' configuration dialog. The 'Step name' is 'step_play_tennis csv'. The 'Filename' is 'C:\Kuliah\Semester5\dw\praktikum\Praktikum_Pentaho\Praktikum_2\Data_CSV\play_tennis.csv'. The 'Delimiter' is ',' and the 'Enclosure' is '"'. The 'NIO buffer size' is '50000'. The 'Lazy conversion?' checkbox is checked. The 'Header row present?' checkbox is checked. The 'Add filename to result' checkbox is unchecked. The 'The row number field name (optional)' field is empty. The 'Running in parallel?' checkbox is unchecked. The 'New line possible in fields?' checkbox is unchecked. The 'Format' is 'mixed'. The 'File encoding' is empty. Below the configuration fields is a table with 10 columns: #, Name, Type, Format, Length, Precision, Currency, Decimal, Group, and Trim type. The table contains 6 rows of data.

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	day	String		3		Rp	,	.	none
2	outlook	String		8		Rp	,	.	none
3	temp	String		4		Rp	,	.	none
4	humidity	String		6		Rp	,	.	none
5	wind	String		6		Rp	,	.	none
6	play	String		3		Rp	,	.	none

Input & Table Output

Hubungkan kedua step ini denga hop (shift-drag)



Double klik step “Table Output”, akan muncul dialog seperti di bawah, kemudian isi sesuai pada gambar dibawah :

The screenshot shows the 'Table output' dialog box with the following settings:

- Step name: table_play_tennis
- Connection: koneksi_db_pentaho
- Target schema: (empty)
- Target table: tbl_play_tennis
- Commit size: 1000
- Truncate table: ☒
- Ignore insert errors: ☐
- Specify database fields: ☐

The 'Database fields' tab is selected, showing the following options:

- 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)
- Store the tablename field: ☒
- Return auto-generated key: ☐
- Name of auto-generated key field: (empty)

Buttons at the bottom: ? Help, OK, Cancel, SQL.

Input & Table Output

Klik tombol "sql" kemudian pilih execute

Table output

Step name: table_play_tennis

Connection: koneksi_db_pentaho

Target schema:

Target table: tbl_play_tennis

Commit size: 1000

Truncate table: ☒

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 in a field?: ☐

Field that contains name of table:

Store the tablename field: ☒

Return auto-generated key: ☐

Name of auto-generated key field:

Help OK Cancel SQL



Table output

Step name: table_play_tennis

Connection: koneksi_db_pentaho

Target schema:

Target table: tbl_play_tennis

Commit size: 1000

Truncate table: ☒

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 in a field?: ☐

Field that contains name of table:

Store the tablename field: ☒

Return auto-generated key: ☐

Name of auto-generated key field:

Help OK Cancel SQL

Simple SQL editor

SQL statements, separated by semicolon :

```
CREATE TABLE tbl_play_tennis
(
  day VARCHAR(3)
  , outlook VARCHAR(8)
  , temp VARCHAR(4)
  , humidity VARCHAR(6)
  , wind VARCHAR(6)
  , play VARCHAR(3)
)
```

Line 1 column 0

Execute Clear cache Close

Results of the SQL state...

The SQL statements had the following results

SQL executed: CREATE TABLE tbl_play_tennis

```
(
  day VARCHAR(3)
  , outlook VARCHAR(8)
  , temp VARCHAR(4)
  , humidity VARCHAR(6)
  , wind VARCHAR(6)
  , play VARCHAR(3)
)
```

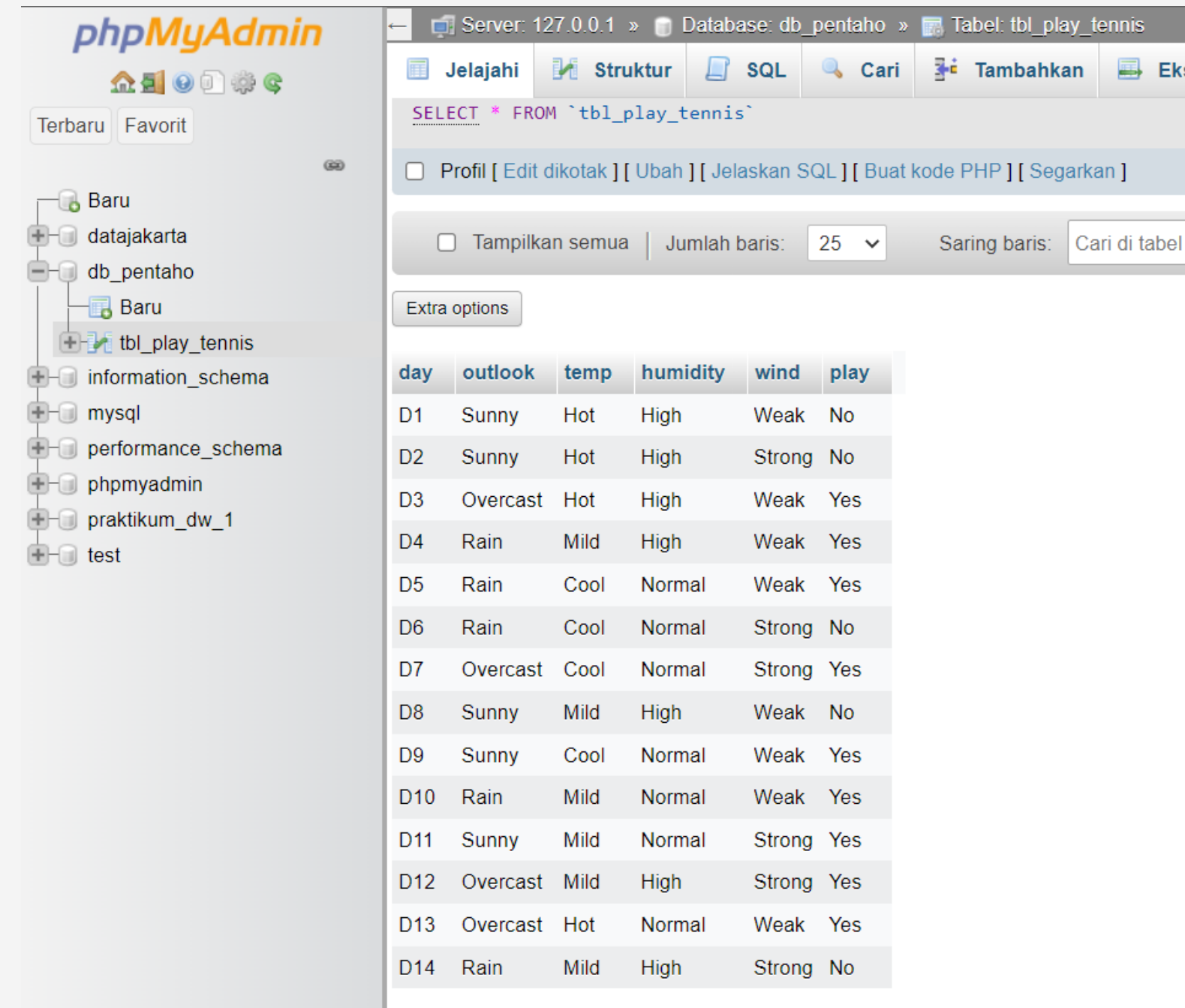
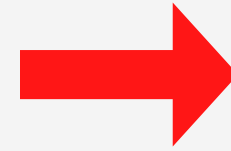
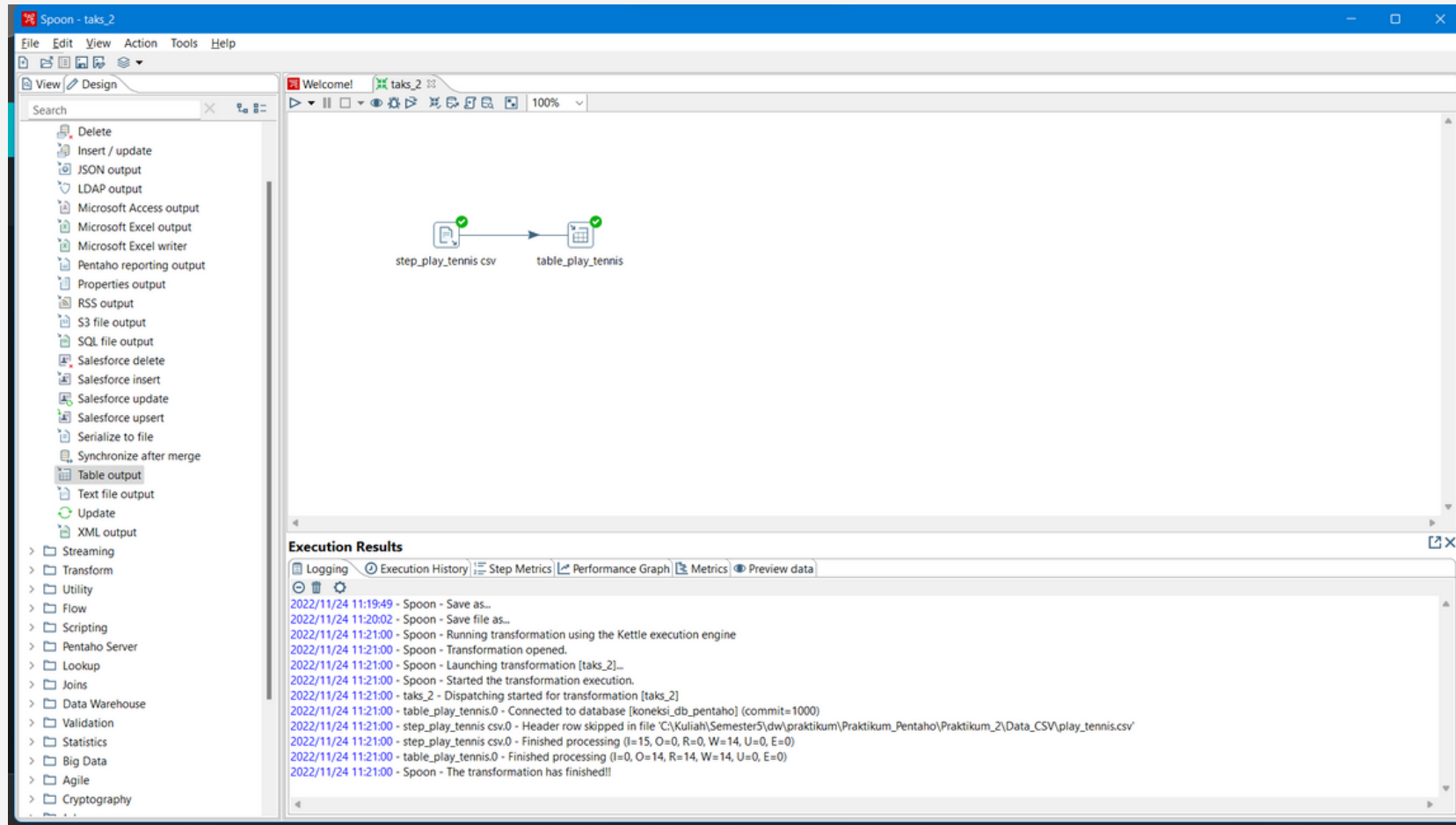
1 SQL statements executed

OK Cancel

Kemudian Klik OK

RUN

Sebelum melakukan run lakukan save transformation terlebih dahulu. Untuk mengecek apakah transformation berhasil cek pada database admin. jika terdapat tabel **tbl_play_tennis** artinya transformation berhasil



LATIHAN INDIVIDU

WWW.KAGGLE.COM