

DATA WAREHOUSE

JOBSHEET 1

Nama : My Babby Findia R.S
No. Absen : 16
Kelas : SIB-2B

TUGAS 1

Analisa lah data tersebut!

1. Berapa jumlah kolom pada data tersebut?
 - Terdapat 7 kolom yang ada dalam data berdasarkan baris pertama yang berisi header
 - a. Customer
 - b. Product
 - c. Region
 - d. Date
 - e. Item unit price
 - f. No. Items
 - e. Total Sale
2. Apa arti atau isi data dari setiap kolom yang ada?
 - a. Customer = Nama atau ID Pelanggan
 - b. Product = Produk mobil yang dibeli, seperti Yaris AT, Etios AT, Avanza AT, dll.
 - c. Region = Wilayah atau lokasi penjualan, seperti MLG, SBY, KDR, dll.
 - d. Date = Tanggal transaksi dalam format MM/DD/YYYY
 - e. Item unit price = Harga satuan per unit mobil dalam format angka desimal
 - f. No. Items = Jumlah unit mobil yang dibeli dalam transaksi tersebut.
 - e. Total Sale = Total penjualan yang dihitung dari item unit price x No. Items.
3. Adakah data yang memiliki nilai null / data yang tidak lengkap?
 - Iya terdapat beberapa baris dengan data yang tidak lengkap null values
 - ❖ Pada baris ke-9 (**Customer 2, Avanza AT,,,168.95,15,2534.25**) → Kolom **Region** dan **Date** kosong.
 - ❖ Pada baris ke-12 (**Customer 1,,,,340.95,7,2386.65**) → Kolom **Product, Region, dan Date** kosong.
 - ❖ Pada baris ke-20 (**Customer 1,Avanza AT,,8/12/2020,168.95,14,675.8**) → Kolom **Region** kosong.
4. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

➤ **Kolom date :**

Beberapa baris memiliki format tanggal yang tidak lengkap atau kosong (null), seperti baris ke-9 dan ke-12.

Kolom Region :

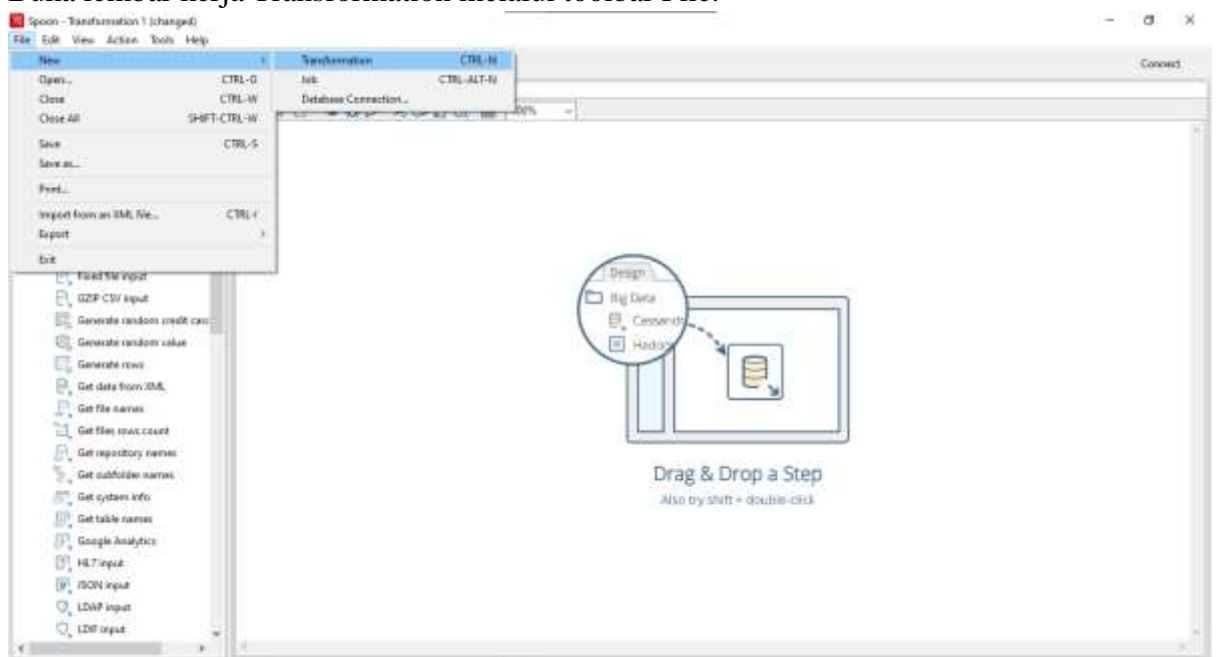
Ada baris yang seharusnya berisi kode wilayah tetapi kosong, seperti baris ke-9, ke-12, dan ke-20.

Kolom Item unit price & Total Sale:

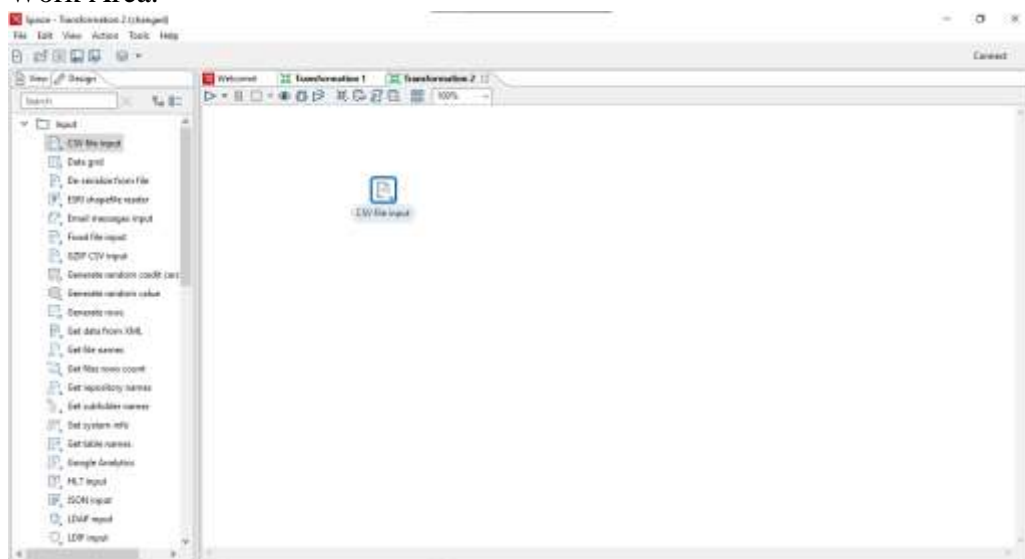
Data umumnya berupa angka desimal, tetapi ada kemungkinan kesalahan pemisah desimal (harusnya menggunakan titik . tetapi bisa saja ada yang memakai koma , tergantung format regional).

A. PENGAMBILAN DATA (Extract)

1. Buka lembar kerja Transformation melalui toolbar File.



2. Cari lah objek CSV file input pada Design Area. Drag and drop objek tersebut menuju Work Area.



- Double-click pada objek CSV file input hingga muncul jendela konfigurasinya.
- Ubah nama Step name menjadi Input Data CSV, (penamaan digunakan untuk identifikasi saja, selanjutnya penamaan boleh menyesuaikan).

CSV file input

Step name: Input Data CSV

Filename: Browse...

Delimiter: , Insert TAB

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									

Help OK Get Fields Preview Cancel

- Pilih lokasi file CarDistributionSales.csv melalui Browse pada filename.

CSV file input

Step name: Input Data CSV

Filename: E:\SEMESTER 4\DATA WAREHOUSE\CarDistributionSales.csv Browse...

Delimiter: , Insert TAB

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									

Help OK Get Fields Preview Cancel

- Jika data telah dipilih pastikan pada bagian Delimiter sesuai dengan kondisi pemisah data pada file yang dipilih. (Pada kesempatan ini adalah comma " , ").

Delimiter: , Insert TAB

- Tekan Get Fields untuk mengambil judul pada setiap kolom, pastikan nama-nama kolom telah sesuai

CSV file input

Step name: Input Data CSV

Filename: E:\SEMESTER 4\DATA WAREHOUSE\CarDistributionSales.csv

Delimiter: ,

Enclosure: "

NO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result: ☐

The row number field name (optional):

Running in parallel? ☐

Now line possible in fields? ☐

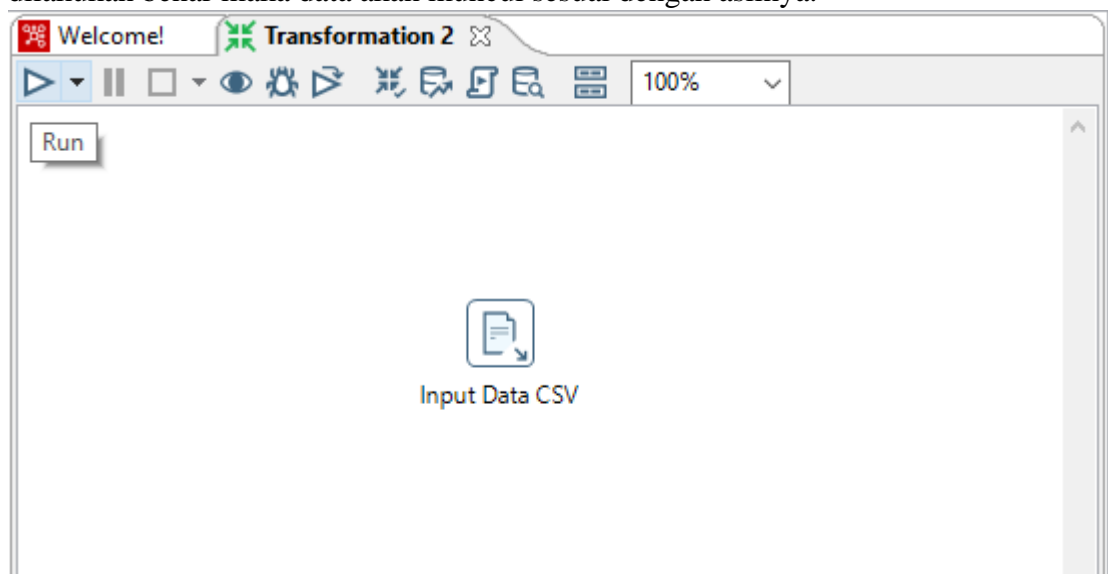
Format: mixed

File encoding:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Item type
1	Customer	String		10		Rp	.	.	none
2	Product	String		0		Rp	.	.	none
3	Region	String		3		Rp	.	.	none
4	Date	Date	MM/dd/yyyy			Rp	.	.	none
5	Item unit price	Number	#,##	12	0	Rp	.	.	none
6	No. Items	Integer	#	15	0	Rp	.	.	none
7	Total Sale	Number	#,##	15	0	Rp	.	.	none

Help OK Get Fields Preview Cancel

8. Tutup jendela konfigurasi tersebut.
9. Tekan tombol "Run" pada pojok kanan atas Area Kerja, maka akan muncul Execution Result Area. Pilih tab Preview Data pada Execution Result Area. Jika proses yang dilakukan benar maka data akan muncul sesuai dengan aslinya.



Execution Results							
<div> <div>Logging</div> <div>Execution History</div> <div>Step Metrics</div> <div>Performance Graph</div> <div>Metrics</div> <div>Preview data</div> </div>							
<div> <div>First rows</div> <div>Last rows</div> <div>Off</div> </div>							
#	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
1	Customer 1	Yaris AT	MLG	Wed Mar 13 00:00:00 ICT 2019	340.95	6	2045.7
2	Customer 1	Etios AT	SBY	Tue Apr 16 00:00:00 ICT 2019	799.95	1	799.95
3	Customer 3	Avanza AT	MLG	Tue Jun 18 00:00:00 ICT 2019	168.95	7	1182.65
4	Customer 3	<null>	MLG	Wed Jul 24 00:00:00 ICT 2019	168.95	2	337.9
5	Customer 1	Avanza AT	SBY	Wed Aug 21 00:00:00 ICT 2019	168.95	9	1520.55
6	Customer 2	Etios AT	KDR	Sat Sep 07 00:00:00 ICT 2019	799.95	7	5599.65
7	Customer 3	Etios AT	MLG	Sat Sep 07 00:00:00 ICT 2019	799.95	8	6399.6
8	Customer 2	Avanza AT	<null>	<null>	168.95	15	2534.25
9	Customer 1	Avanza AT	SBY	Sat Sep 14 00:00:00 ICT 2019	168.95	15	2534.25
10	Customer 1	Etios AT	SBY	Wed Sep 25 00:00:00 ICT 2019	799.95	2	1599.9
11	Customer 1	Avanza AT	SBY	Tue Jan 21 00:00:00 ICT 2020	168.95	14	2365.3
12	Customer 1	Yaris AT	MLG	Sat Feb 01 00:00:00 ICT 2020	340.95	2	681.9
13	Customer 1	<null>	<null>	<null>	340.95	7	2386.65
14	Customer 1	Etios AT	SBY	Sat May 30 00:00:00 ICT 2020	799.95	9	7199.55
15	Customer 3	Avanza AT	MLG	Sun Jun 07 00:00:00 ICT 2020	168.95	10	1689.5
16	Customer 3	Yaris AT	MLG	Tue Jun 09 00:00:00 ICT 2020	340.95	4	1363.8
17	Customer 2	<null>	KDR	Fri Aug 07 00:00:00 ICT 2020	799.95	14	11199.3
18	Customer 2	Avanza AT	KDR	Sat Aug 08 00:00:00 ICT 2020	168.95	12	2027.4
19	Customer 1	Avanza AT	<null>	Wed Aug 12 00:00:00 ICT 2020	168.95	4	675.8

TUGAS 2

- Apakah data hasil eksekusi sesuai dengan data aslinya?
 - Iya, data hasil eksekusi sesuai dengan data aslinya.
- PDI Spoon melakukan proses extract, perhatikan pada Tab Logging di Execution Results Area, langkah-langkah apa sajakah yang dilakukan PDI Spoon untuk melakukan extract data?
 - a. Memulai proses ekstraksi
 - b. Membuka koneksi ke sumber data
 - c. Mengeksekusi query atau membaca file
 - d. Memproses dan mengekstrak data
- Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E ?

Nama	Processor	RAM	Storage	VGA	CPU	Waktu Eksekusi
Babby	Intel(R) Core(TM) i3-6006U CPU @ 2.00GHz 1.99 GHz	4,00 GB	99,4 GB	NVIDIA GeForce 920MX Intel(R) HD Graphics 520	Intel(R) Core(TM) i3-6006U CPU @ 2.00GHz 1.99 GHz	0.01s
Veve	AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx 2.10 GHz	8,00 GB	1 TB	AMD Radeon(TM) Vega 8 Graphics	AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx	0.01s
Putera	Intel(R) Core(TM) i5-	16 GB	256 GB	Intel (R) HD Graphics 5500	Intel(R) Core(TM) i5-53000 CPU 2.30GHz	0.1s

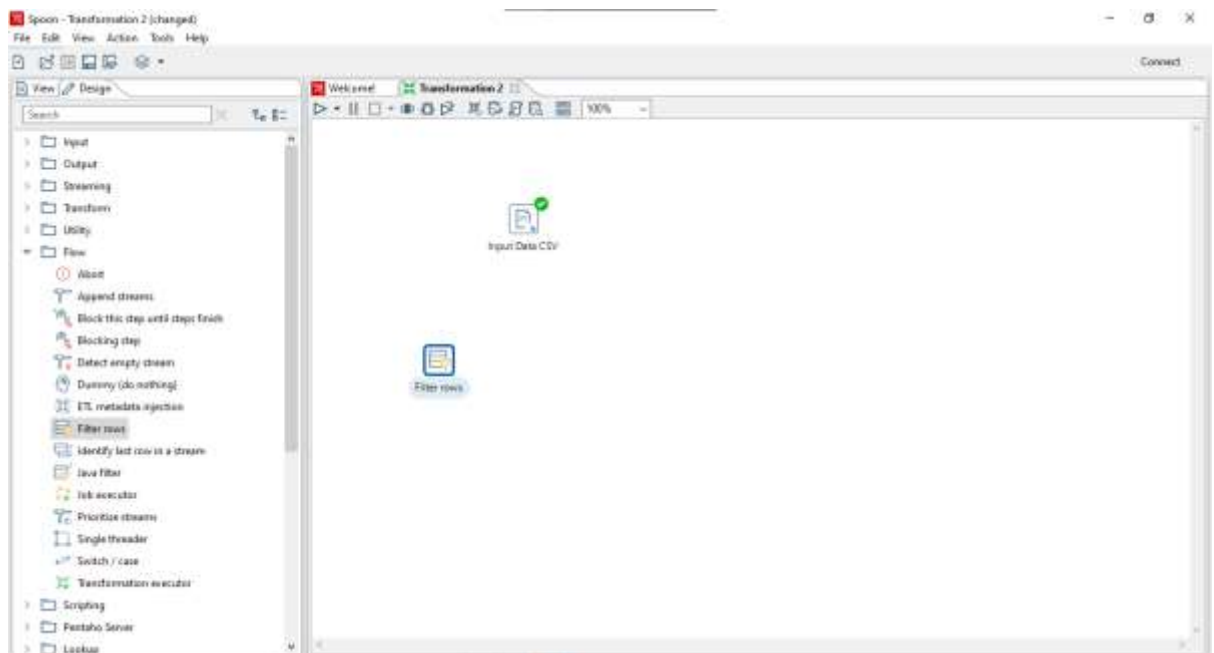
	53000 CPU 2.30GHz					
Dhevina	12th Gen Intel(R) Core(TM) i5 - 1235U 1.30 GHz	8 GB	152 GB	Inter(R) UHD Graphics 12th	Intel(R) Core(TM) i5 - 1235U	0.0s

2024/02/17 06:51:48 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)

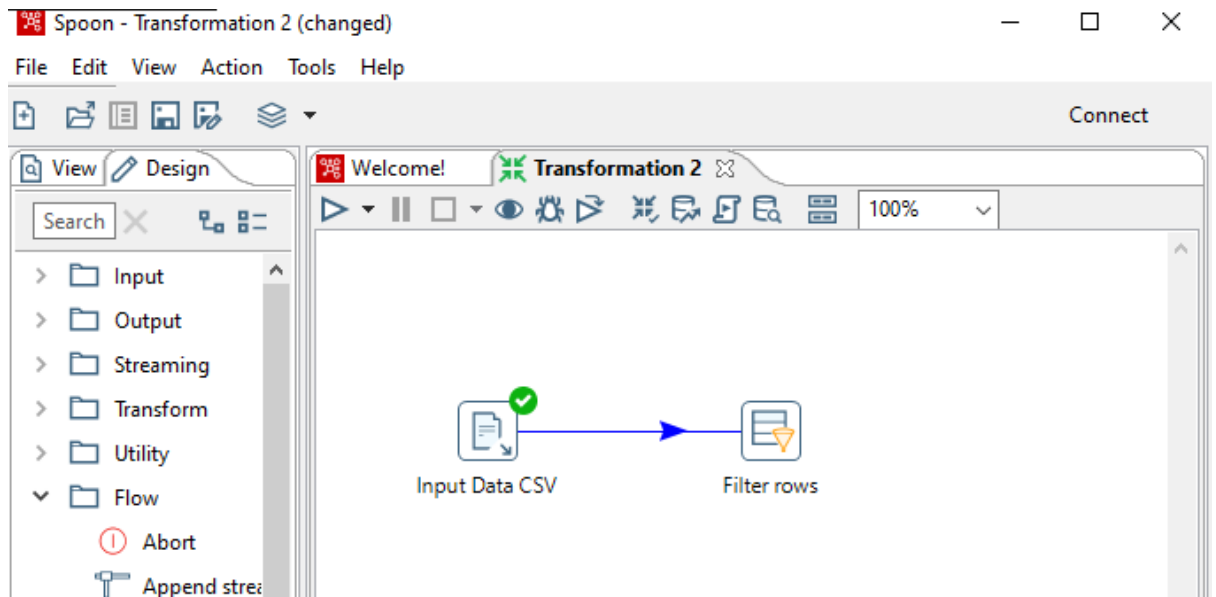
- Log ini menunjukkan bahwa 20 baris data telah dibaca, 19 baris berhasil diproses, tidak ada data yang ditolak, diperbarui, atau mengalami kesalahan.
- 4. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract data ini? Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU). Bandingkan dalam bentuk tabel.

B. Filter Data (Transform) dan Pengemasan data (Load)

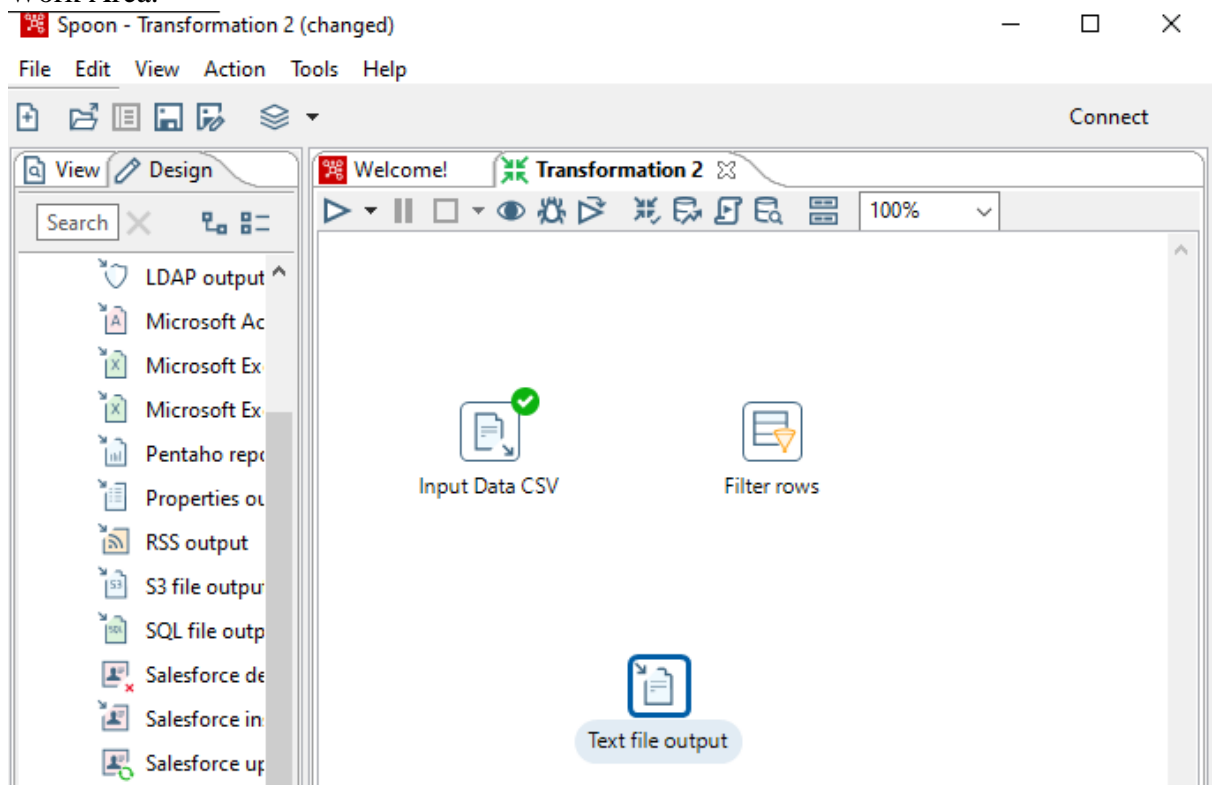
1. Carilah objek Filter rows pada Design Area. Drag and drop objek tersebut pada Work Area.



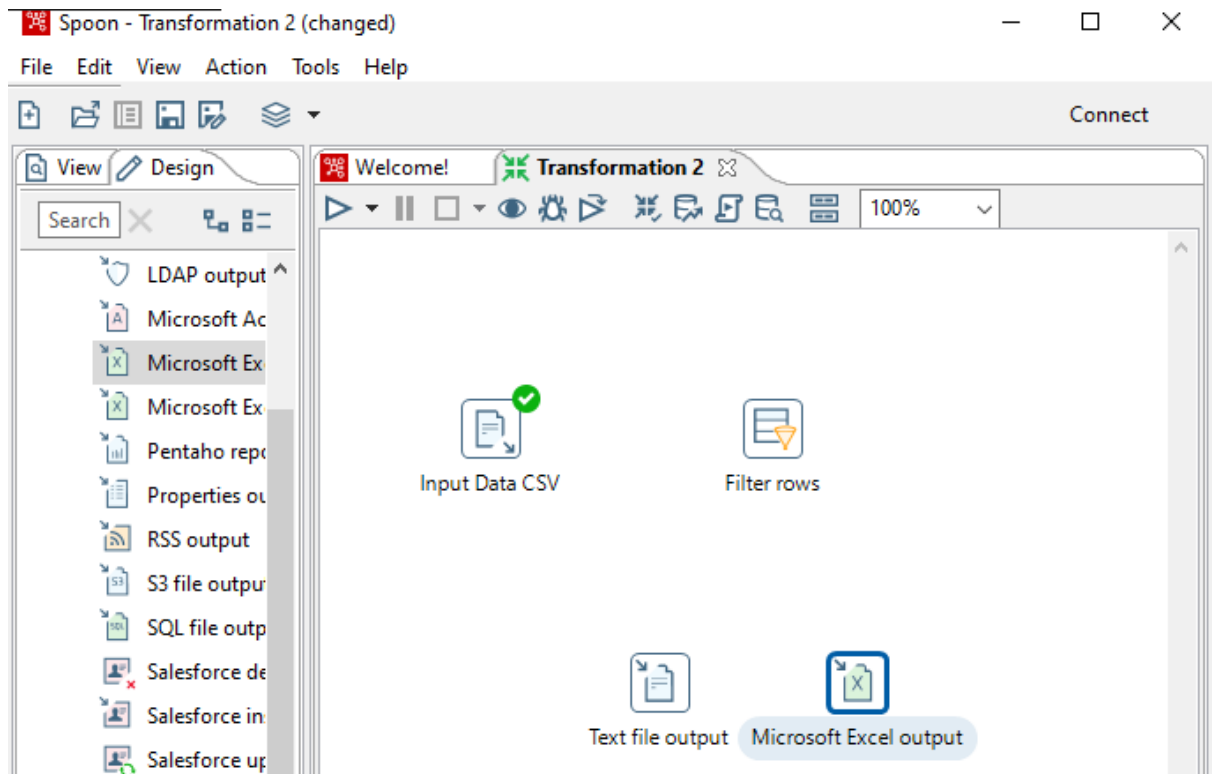
2. Hubungkan output pada Input Data CSV menuju Filter rows. Pilih koneksi Main output of step saat menghubungkan kedua objek tersebut.



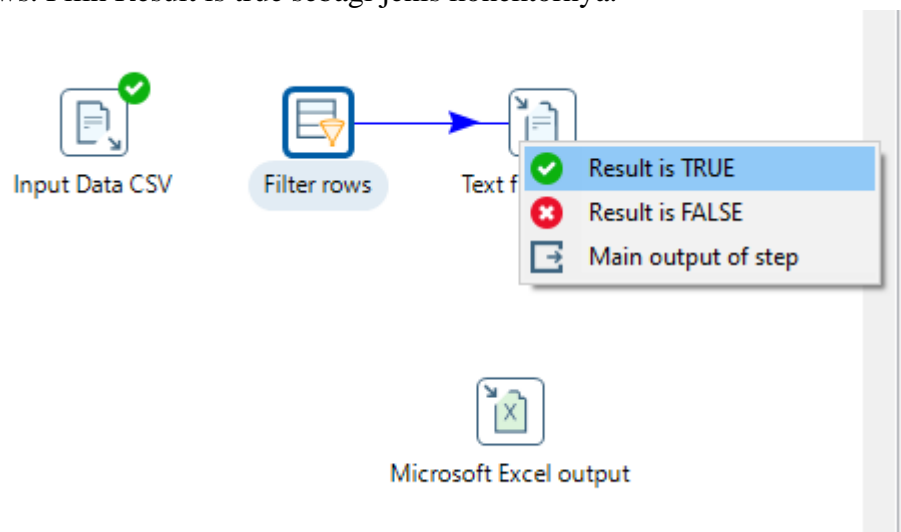
- Carilah objek Text file output pada Design Area. Drag and drop objek tersebut pada Work Area.



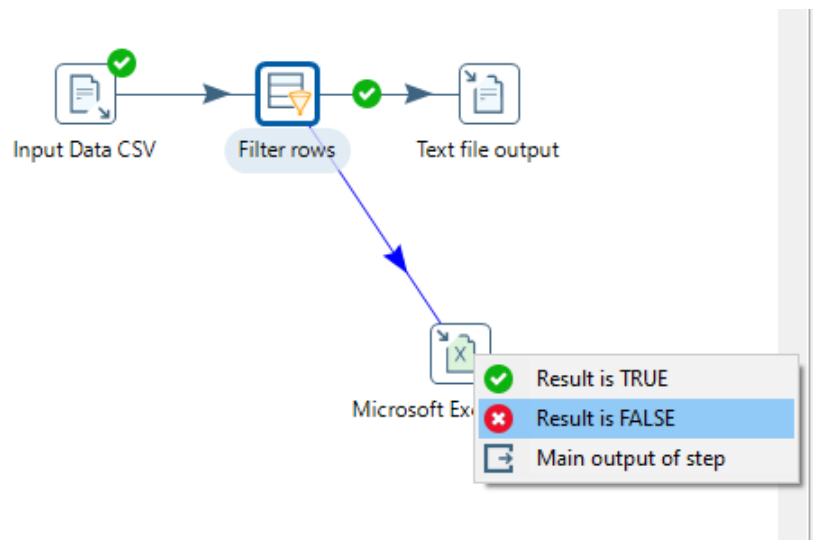
- Carilah objek Microsoft excel output pada Design Area. Drag and drop objek tersebut pada Work Area.



5. Hubungkan filter rows dengan text file output menggunakan konektor output dari filter rows. Pilih Result is true sebagai jenis konektornya.



6. Hubungkan filter rows dengan microsoft excel output menggunakan konektor output dari filter rows. Pilih Result is false sebagai jenis konektornya.



7. Double click pada filter rows hingga muncul jendela konfigurasi. ubah step name menjadi Filter data.
8. Cek bagian send 'true' data to step menjadi Text file output dan cek bagian Send 'false' data to step menjadi Microsoft Excel output.
9. Klik <Field> pada bagian condition

Filter rows

Step name:

Send 'true' data to step:

Send 'false' data to step:

The condition:

=

Help OK Cancel

10. Pilih field yang diinginkan untuk digunakan sebagai kondisi. Pada kesempatan ini pilih Customer, kemudian klik OK.

Filter rows

Step name:

Send 'true' data to step:

Send 'false' data to step:

The condition:

=

11. Pada bagian value pilih IS NOT NULL kemudian tekan OK , hal ini dilakukan karena data yang dipilih adalah data yang tidak Null

Filter rows

Step name:

Send 'true' data to step:

Send 'false' data to step:

The condition:

IS NOT NULL

12. Tekan tombol + untuk menambah kondisi.
13. Pilih statement yang masih null untuk menambah kondisi.
14. Ulangi proses nomor 9,10,11 dengan mengganti semua field yang ada. hingga statement menjadi seperti pada gambar dibawah ini.
15. Gunakan operator AND untuk menghubungkan
16. Jika semua field sudah masuk dalam kondisi maka tekan OK.

Filter rows

Step name: Filter data

Send 'true' data to step: Text file output

Send 'false' data to step: Microsoft Excel output

The condition:

```
(
    AND
    (
        AND
        (
            AND
            (
                AND
                (
                    No. Items IS NOT NULL
                    Total Sale IS NOT NULL
                )
            )
        )
    )
)
```

Help OK Cancel

17. Double-click pada text file output, pilih lokasi file untuk menyimpan hasil output file dengan menekan tombol browse. beri nama outputSalesCarDistribution.

18. Ubah extension menjadi csv.

Text file output

Step name: Text file output

File Content Fields

Filename: E:/SEMESTER 4/DATA WAREHOUSE/TUGAS/outputSalesCarDistribution Browse...

Pass output to servlet ☐

Create Parent folder ☒

Do not create file at start ☐

Accept file name from field? ☐

File name field:

Extension: csv

Include step nr in filename? ☐

Include partition nr in filename? ☐

Include date in filename? ☐

Include time in filename? ☐

Specify Date time format:

Date time format:

Show filename(s)...

Add filenames to result ☒

Help OK Cancel

Text file output

Step name: Text file output

File Content Fields

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim Type	Null
1	Customer	String		10					none	
2	Product	String		9					none	
3	Region	String		3					none	
4	Date	Date	MM/dd/yyyy						none	
5	Item unit price	Number	#,##	15	0	Rp	-	,	none	
6	No. Items	Integer	#	15	0	Rp	-	,	none	
7	Total Sale	Number	#,##	15	0	Rp	-	,	none	

Get Fields Minimal width

Help OK Cancel

19. Pada tab Fields, tekan tombol Get fields untuk mengambil field data.
20. Tekan ok untuk kembali ke work area
21. Hal yang sama dilakukan pada konfigurasi Microsoft Excel output, Double-click pada objek Microsoft Excel output hingga muncul jendela konfigurasi.
22. Pilih lokasi file untuk menyimpan file output dengan menekan tombol browse. Beri nama failedSalesCarDistribution. untuk extension tetap dengan extension xls.
23. Pada tab fields tekan tombol Get fields untuk mengambil field data.
24. Tekoan ok untuk kembali ke work area.

Microsoft Excel output

Step name: Microsoft Excel output

File Content Custom Fields

Filename: E:/SEMESTER 4/DATA WAREHOUSE/TL [Browse...](#)

Create Parent folder ☐

Do not create file at start ☐

Extension: xls

Include stepnr in filename? ☐

Include date in filename? ☐

Include time in filename? ☐

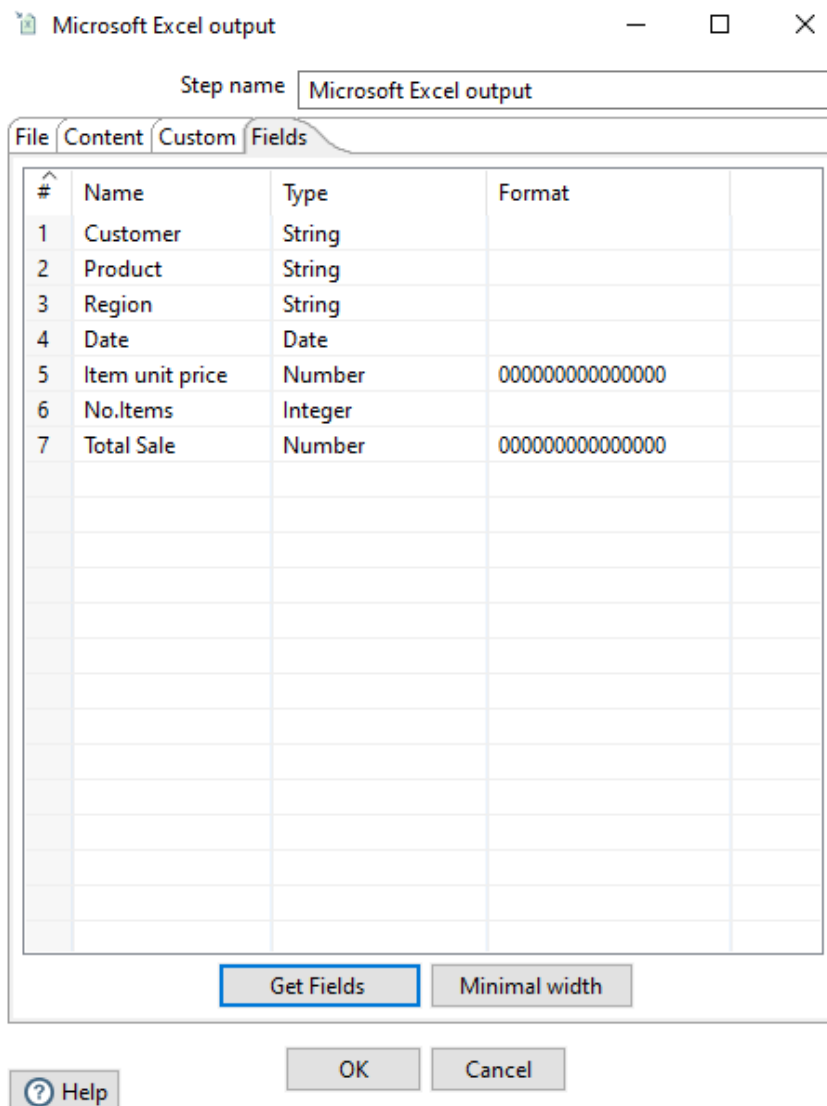
Specify Date time format ☐

Date time format:

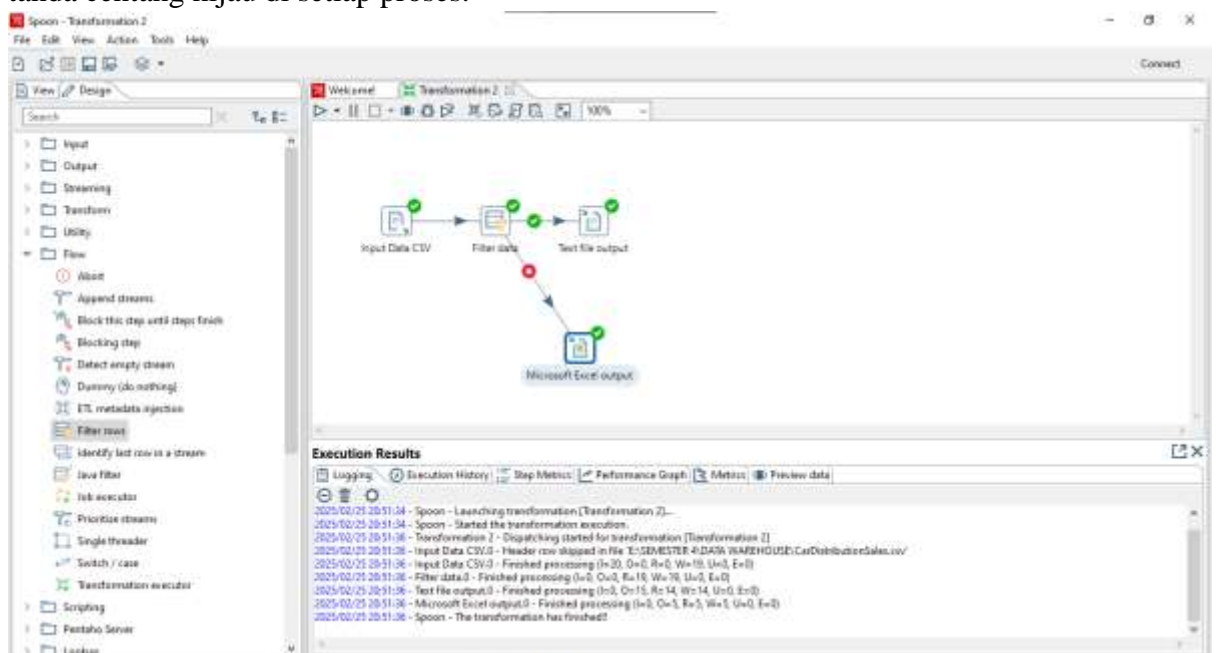
[Show filename\(s\)...](#)

Add filenames to result ☒

[Help](#) [OK](#) [Cancel](#)



25. Tekan tombol Run pada pojok kiri atas Work Area. jika berhasil maka akan muncul tanda centang hijau di setiap proses.



Execution Results							
<div> Logging Execution History Step Metrics Performance Graph Metrics Preview data </div> <div> <input checked="" type="radio"/> First rows <input type="radio"/> Last rows <input type="radio"/> Off </div>							
#	Customer	Product	Region	Date	Item unit price	No. Items	Total Sale
1	Customer 3	<null>	MLG	Wed Jul 24 00:00:00 ICT 2019	168.95	2	337.9
2	Customer 2	Avanza AT	<null>	<null>	168.95	15	2534.25
3	Customer 1	<null>	<null>	<null>	340.95	7	2386.65
4	Customer 2	<null>	KDR	Fri Aug 07 00:00:00 ICT 2020	799.95	14	11199.3
5	Customer 1	Avanza AT	<null>	Wed Aug 12 00:00:00 ICT 2020	168.95	4	675.8

Execution Results							
<div> Logging Execution History Step Metrics Performance Graph Metrics Preview data </div> <div> <input checked="" type="radio"/> First rows <input type="radio"/> Last rows <input type="radio"/> Off </div>							
#	Customer	Product	Region	Date	Item unit price	No. Items	Total Sale
1	Customer 1	Yaris AT	MLG	Wed Mar 13 00:00:00 ICT 2019	340.95	6	2045.7
2	Customer 1	Erios AT	SBY	Tue Apr 16 00:00:00 ICT 2019	799.95	1	799.95
3	Customer 3	Avanza AT	MLG	Tue Jun 18 00:00:00 ICT 2019	168.95	7	1182.65
4	Customer 1	Avanza AT	SBY	Wed Aug 21 00:00:00 ICT 2019	168.95	9	1520.55
5	Customer 2	Erios AT	KDR	Sat Sep 07 00:00:00 ICT 2019	799.95	7	5599.65
6	Customer 3	Erios AT	MLG	Sat Sep 07 00:00:00 ICT 2019	799.95	8	6399.6
7	Customer 1	Avanza AT	SBY	Sat Sep 14 00:00:00 ICT 2019	168.95	15	2534.25

TUGAS 3

1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?
 - Dalam konteks PDI Spoon, data yang diekstrak ke CSV biasanya hanya berisi teks mentah tanpa pemformatan tambahan, sedangkan jika diekspor ke Excel, data bisa memiliki format yang lebih kompleks.
2. Jelaskan apa yang terjadi pada proses Filter rows!
 - Proses Filter Rows dalam Pentaho Data Integration (PDI) Spoon adalah langkah yang digunakan untuk menyaring data berdasarkan kondisi tertentu. Berikut yang terjadi dalam proses ini:
 - a. Membaca data input dari sumber (misalnya, file CSV atau database).
 - b. Menerapkan kondisi filter, misalnya:
 1. Menyaring hanya data dengan tanggal tertentu.
 2. Menyaring data berdasarkan jumlah transaksi lebih dari 1000.
 3. Menghapus baris yang mengandung nilai kosong atau NULL.
3. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract Transfer Load data ini? Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU). bandingkan dalam bentuk tabel

Nama	Processor	RAM	Storage	VGA	CPU	Waktu Eksekusi	Jenis Perangkat
Babby	Intel(R) Core(TM) i3-6006U CPU @ 2.00GHz 1.99 GHz	4,00 GB	99,4 GB	NVIDIA GeForce 920MX Intel(R) HD Graphics 520	Intel(R) Core(TM) i3-6006U CPU @ 2.00GHz 1.99 GHz	0.01	Laptop
Veve	AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx 2.10 GHz	8,00 GB	1 TB	AMD Radeon(TM) Vega 8 Graphics	AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx	0.01s	Laptop
Putera	Intel(R) Core(TM) i5-53000 CPU 2.30GHz	16 GB	256 GB	Intel (R) HD Graphics 5500	Intel(R) Core(TM) i5-53000 CPU 2.30GHz	0.1s	Laptop
Dhevina	12th Gen Intel(R) Core(TM) i5 - 1235U 1.30 GHz	8 GB	152 GB	Inter(R) UHD Graphics 12th	Intel(R) Core(TM) i5 - 1235U	0.0s	Laptop

STUDI KASUS

KASUS:

Pak Pascanowo seorang ketua RW di Ibu Kota Nusabangsa (IKN). Terdapat 20 orang dari negara Konoha yang akan pindah dan berdomisili di IKN. Tata letak pemukiman pada IKN telah dikondisikan agar berkelompok disesuaikan dengan lokasi pekerjaannya agar akses terjangkau.

Sebagai contoh, mahasiswa akan dikumpulkan dan berdomisili di wilayah edukasi yang akses dekat dengan kampus-kampus perguruan tinggi. Sedangkan dokter akan berdomisili dekat dengan rumah sakit atau layanan kesehatan dan PNS akan berdomisili dekat dengan perkantoran dan layanan publik. Selain itu, orang yang bukan dari ketiga kategori tersebut bebas memilih domisili atau tempat tinggalnya di IKN.

Pak Pascanowo akan mendapatkan data dari pemerintah pusat tentang Masyarakat Konoha yang akan pindah ke IKN. dari data tersebut Pak Pascanowo akan membagi menjadi 4 kelompok,

Data masyarakat berprofesi mahasiswa akan diberikan ke Pak Emir Makarena selaku ketua RT wilayah pendidikan.

Data masyarakat berprofesi PNS akan diberikan ke Pak Yasana Laili selaku ketua RT wilayah pemerintahan

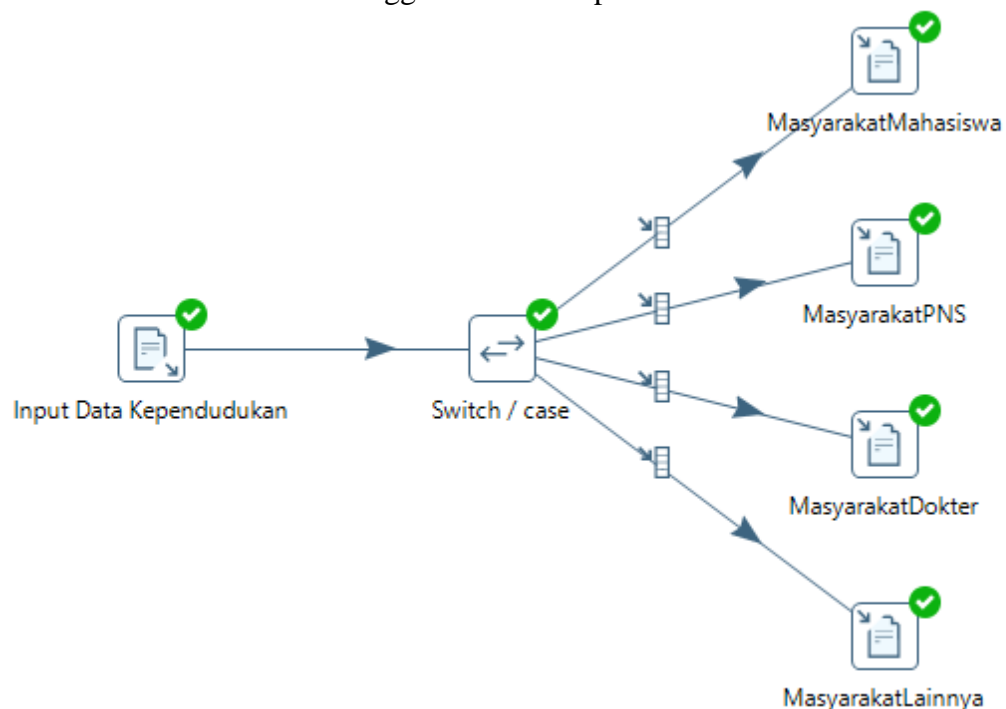
Data masyarakat berprofesi Dokter akan diberikan ke Pak Budi Pekerti selaku ketua RT wilayah kesehatan

data-data tersebut akan digunakan untuk analisa di masing-masing wilayah dan juga Dashboard untuk emmbantu mengambil keputusan.

Pak Pascanowo mendapatkan data penduduk yang akan pindah ke wilayahnya melalui link: <https://github.com/dik4rizky/datasources/blob/82821b8ec341176f979956c84b79720d4aead012/dataDummyKependudukan.csv>

Bantulah Pak Pascanowo untuk memishkan data tersebut menjadi 4 file:
MasyarakatMahasiswa, MasyarakatPNS, MasyarakatDokter, MasyarakatLainnya.

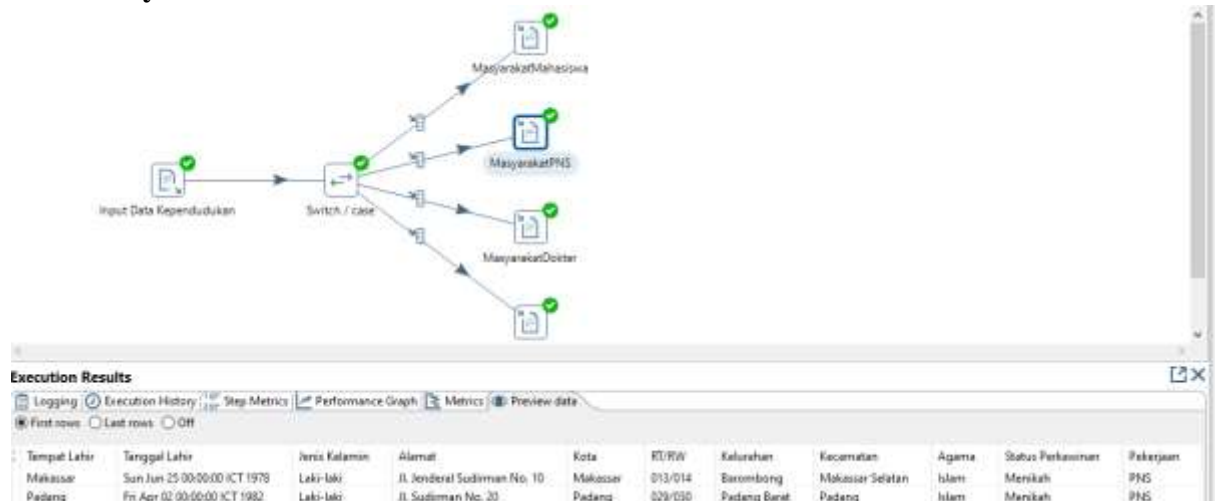
Dikarenakan proses tersebut akan repetisi setiap ada masyarakat Konoha yang berpindah ke IKN maka buatlah sistem tersebut menggunakan PDI Spoon.



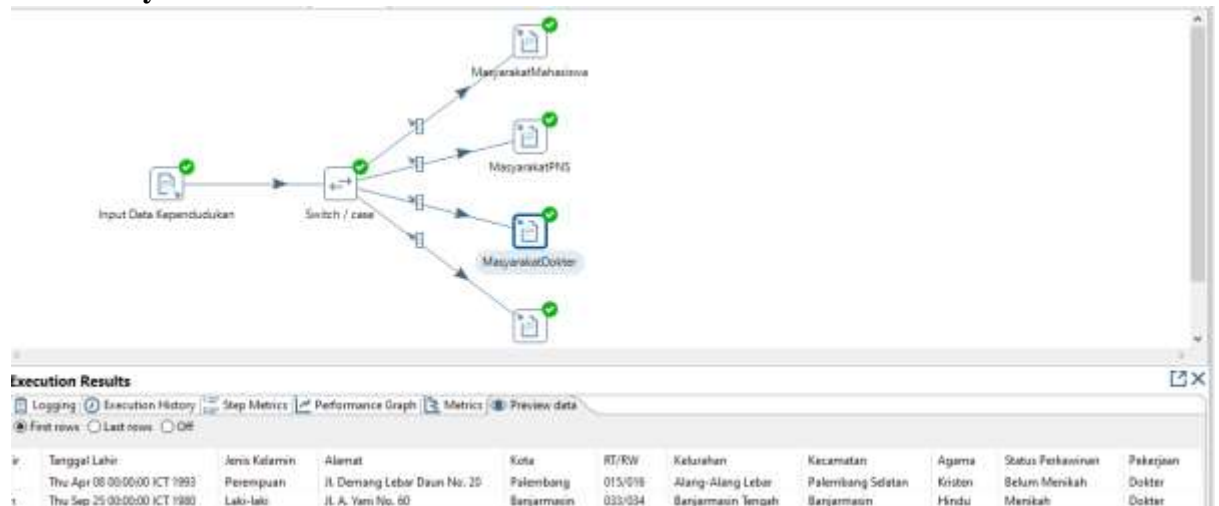
Data MasyarakatMahasiswa

K	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
1	Bandung	Tue May 15 00:00:00 ICT 1990	Perempuan	Jl. Cendrawasih No. 45	Bandung	003/004	Cikutra	Bandung Wetan	Kristen	Belum Menikah	Mahasiswa
2	Semarang	Sun Mar 05 00:00:00 ICT 1995	Laki-laki	Jl. Pemuda No. 21	Semarang	008/010	Pandananan	Semarang Tengah	Budha	Belum Menikah	Mahasiswa
3	Jayapura	Sat Oct 23 00:00:00 ICT 1993	Perempuan	Jl. Problemati No. 80	Jayapura	038/040	Kota Jayapura	Jayapura	Kristen	Belum Menikah	Mahasiswa

Data MasyarakatPNS



Data MasyarakatDokter



Data MasyarakatLainnya

