



Information Technology Department - State Polytechnic of Malang
Jobsheet-1: Data Warehouse Sederhana dan Pengenalan PDI Spoon
Course: Data Warehouse
Lecturer: Vipkas Al Hadid Firdaus, ST., MT
07 March 2025

Name: Azkiya Putri Ihtiarilah
Nim: 2341760175

Tujuan Praktikum

Setelah melakukan praktikum ini, mahasiswa diharapkan dapat lebih mengenal apa itu data warehouse berdasarkan uji coba membuat data warehouse sederhana.

Studi Kasus

Pak Amir adalah Manager Sales Astro Mobil, yang bergerak di bidang distributor mobil wilayah Jawa Timur. Pak Amir meminta data penjualan dari beberapa cabang kepada supervisor. Data tersebut akan digunakan untuk membuat suatu Dashboard pengambilan keputusan. Dikarenakan proses permintaan ini dilakukan secara berulang setiap hari setelah jam kantor maka, Pak Amir akan membuat proses untuk menarik data yang ada pada file milik supervisor tersebut secara otomatis. Namun, kadang data tersebut tidak lengkap. Sehingga Pak Amir akan mengambil data yang lengkap saja dan mengembalikan data yang tidak lengkap.

Dari studi kasus tersebut maka, akan dilakukan:

1. Cek dan Analisa data penjualan.
2. Import data dari file (Extract)
3. Identifikasi data yang tidak lengkap (missing data) dan meletakkan pada file yang berbeda (Transform)
4. Memindahkan data yang sudah lengkap ke file dashboard. (Load)
5. Mengumpulkan data yang belum lengkap untuk dikembalikan.

Data penjualan tersebut dapat diakses dan di download melalui link berikut:

<https://github.com/dik4rizky/datasources/blob/7e3f70f516a1b107635141d09862a8c30fbb6812/CarDistributiorSales.csv>

TUGAS 1

| Customer | Product | Region | Date | Item unit price | No.Items | Total Sale |
|------------|-----------|--------|------------|-----------------|----------|------------|
| Customer 1 | Yaris AT | MLG | 3/13/2019 | 340.95 | 6 | 2045.7 |
| Customer 1 | Etios AT | SBY | 4/16/2019 | 799.95 | 1 | 799.95 |
| Customer 3 | Avanza AT | MLG | 6/18/2019 | 168.95 | 7 | 1182.65 |
| Customer 3 | | MLG | 7/24/2019 | 168.95 | 2 | 337.9 |
| Customer 1 | Avanza AT | SBY | 8/21/2019 | 168.95 | 9 | 1520.55 |
| Customer 2 | Etios AT | KDR | 09/07/2019 | 799.95 | 7 | 5599.65 |
| Customer 3 | Etios AT | MLG | 09/07/2019 | 799.95 | 8 | 6399.6 |
| Customer 2 | Avanza AT | | | 168.95 | 15 | 2534.25 |
| Customer 1 | Avanza AT | SBY | 9/14/2019 | 168.95 | 15 | 2534.25 |
| Customer 1 | Etios AT | SBY | 9/25/2019 | 799.95 | 2 | 1599.9 |
| Customer 1 | Avanza AT | SBY | 1/21/2020 | 168.95 | 14 | 2365.3 |
| Customer 1 | Yaris AT | MLG | 02/01/2020 | 340.95 | 2 | 681.9 |
| Customer 1 | | | | 340.95 | 7 | 2386.65 |
| Customer 1 | Etios AT | SBY | 5/30/2020 | 799.95 | 9 | 7199.55 |
| Customer 3 | Avanza AT | MLG | 06/07/2020 | 168.95 | 10 | 1689.5 |
| Customer 3 | Yaris AT | MLG | 06/09/2020 | 340.95 | 4 | 1363.8 |
| Customer 2 | | KDR | 08/07/2020 | 799.95 | 14 | 11199.3 |
| Customer 2 | Avanza AT | KDR | 08/08/2020 | 168.95 | 12 | 2027.4 |
| Customer 1 | Avanza AT | | 08/12/2020 | 168.95 | 4 | 675.8 |

Analisa lah data tersebut!

1. Berapa jumlah kolom pada data tersebut?

Dalam Data tersebut terdapat 7 kolom, yaitu:

- a) Customer
- b) Product
- c) Region
- d) Date
- e) Item unit price
- f) No. items
- g) Total Sale

2. Apa arti atau isi data dari setiap kolom yang ada?

Isi data dari masing masing kolom tersebut sebagai berikut:

- a) **Customer:** Berisi Data dari customer id
- b) **Product:** Nama produk mobil dalam transaksi
- c) **Region:** Wilayah transaksi mobil dilakukan
- d) **Date:** Tanggal transaksi terjadi
- e) **Item unit price:** Harga satuan dari mobil yang dibeli
- f) **No. items:** Jumlah unit mobil yang dibeli dalam transaksi
- g) **Total Sale:** Total penjualan dalam transaksi

3. Adakah data yang memiliki nilai null / data yang tidak lengkap?

Dalam data tersebut, ada beberapa data yang memiliki isi data kolom null.

- a) **Pada Baris 4** terdapat kolom kosong (Product)
- b) **Pada Baris 8** terdapat kolom kosong (Region, Date)
- c) **Pada Baris 13** terdapat kolom kosong (Product, Region, Date)
- d) **Pada Baris 17** terdapat kolom kosong (Product)
- e) **Pada Baris 19** terdapat kolom kosong (Region)

4. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

- **Kolom "Product"**

Ada beberapa baris yang kosong dalam kolom ini, misalnya pada baris dengan Customer 3 (7/24/2019) dan Customer 2 (8/7/2020).

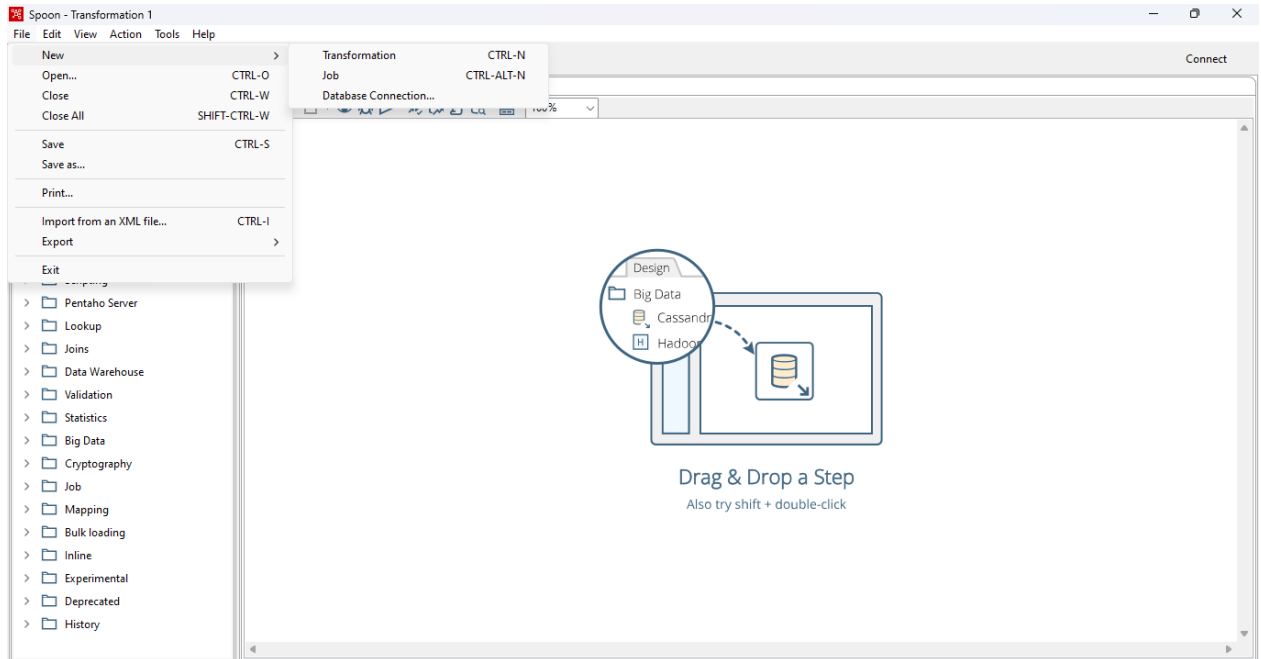
Seharusnya semua nilai dalam kolom ini adalah teks (string), tetapi nilai kosong dapat dianggap sebagai tipe yang berbeda (misalnya, NULL atau NaN dalam beberapa sistem).

- **Kolom "Region"**

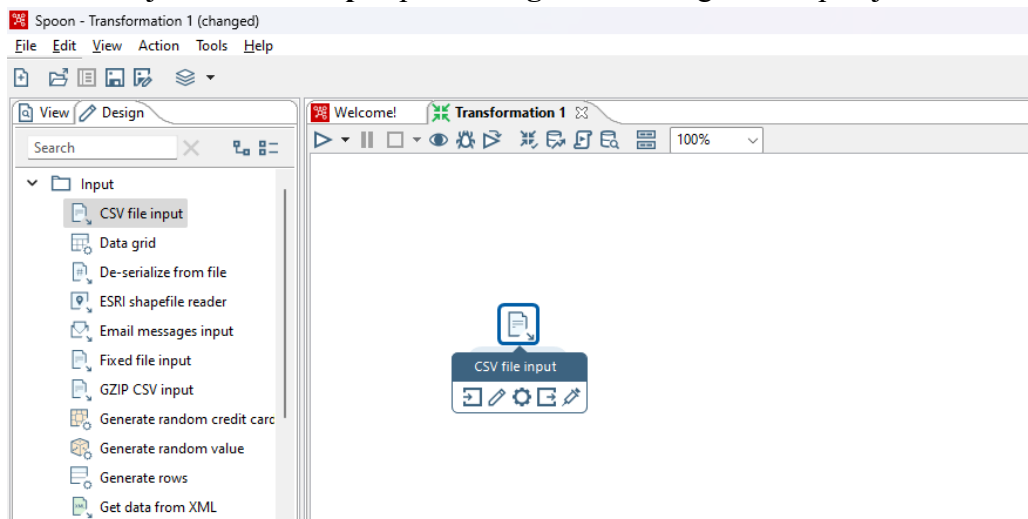
Sama seperti kolom "Product", ada beberapa baris yang kosong dalam kolom ini, misalnya pada Customer 2 (tanggal tidak tersedia). Seharusnya semua nilai dalam kolom ini adalah teks.

A. Pengambilan Data (Extract)

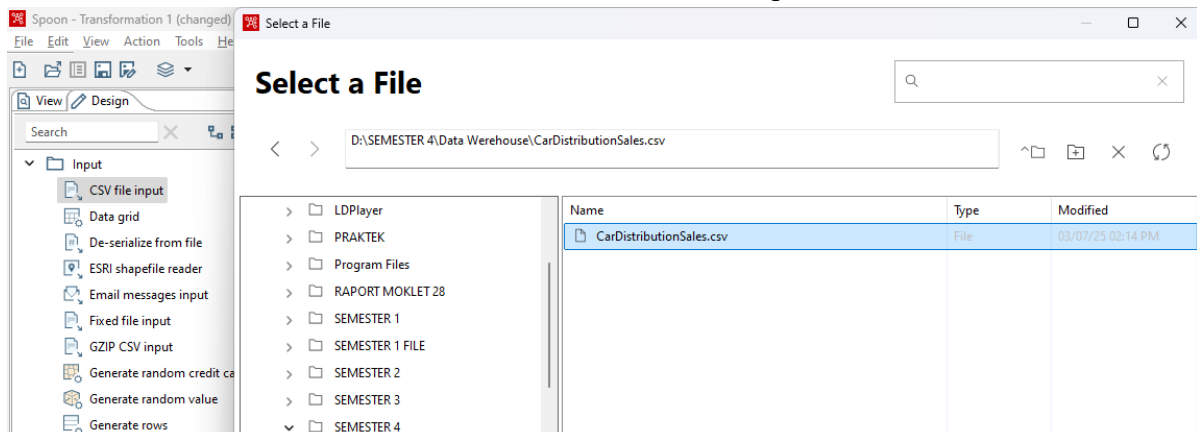
1. Buka lembar kerja **Transformation** melalui toolbar **File**.



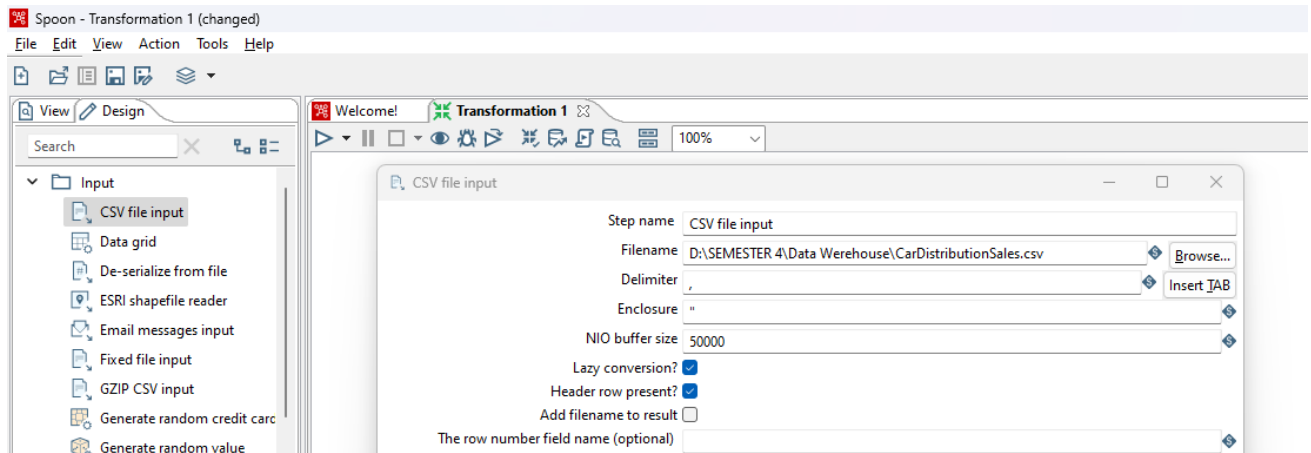
2. Cari lah objek **CSV file input** pada **Design Area**. Drag and drop objek tersebut menuju **Work Area**.



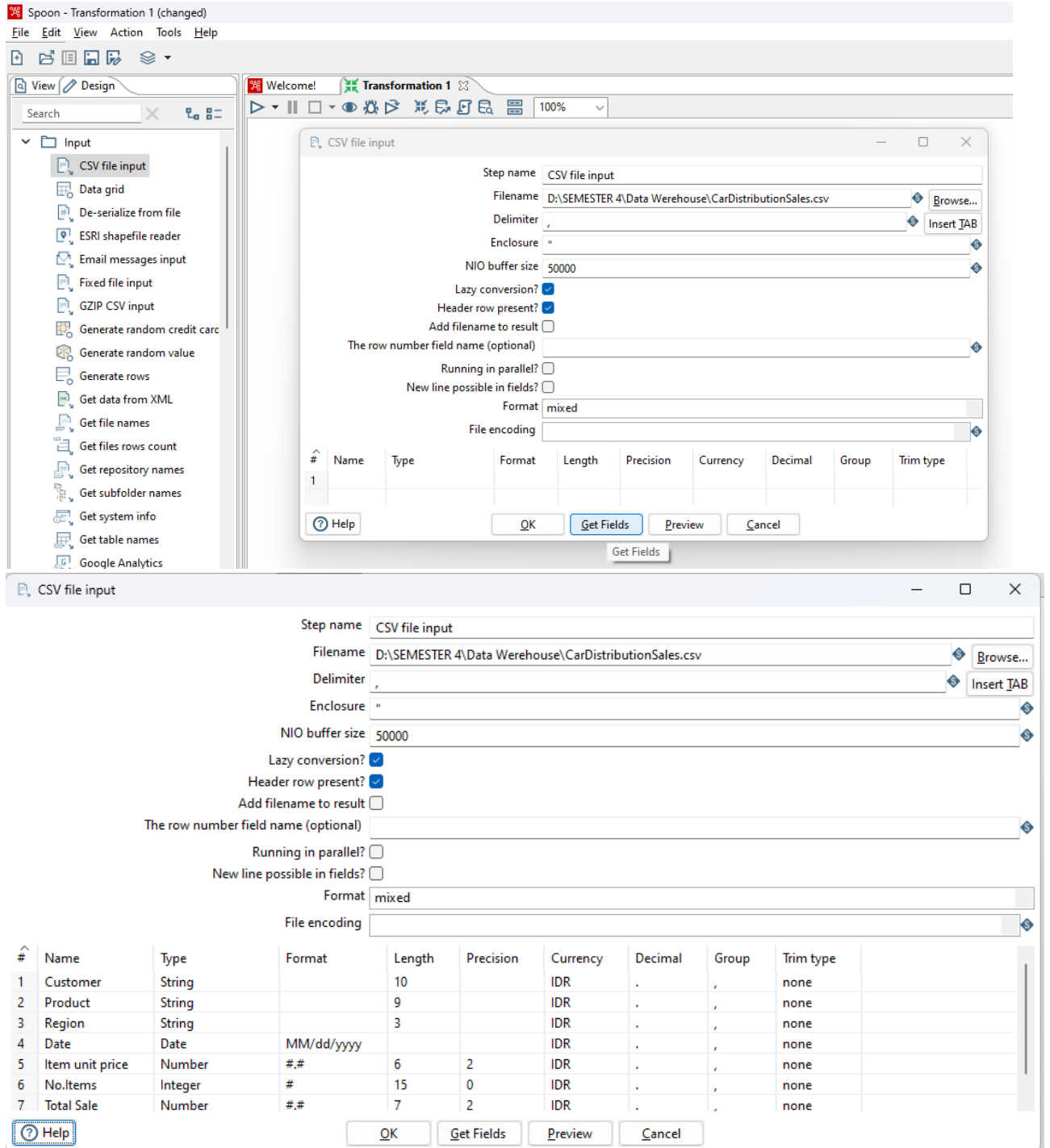
3. Double-click pada objek **CSV file input** hingga muncul jendela konfigurasinya.
4. Ubah nama **Step name** menjadi **Input Data CSV**, (penamaan digunakan untuk identifikasi saja, selanjutnya penamaan boleh menyesuaikan).
5. Pilih lokasi file **CarDistributionSales.csv** melalui **Browse** pada **filename**.



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

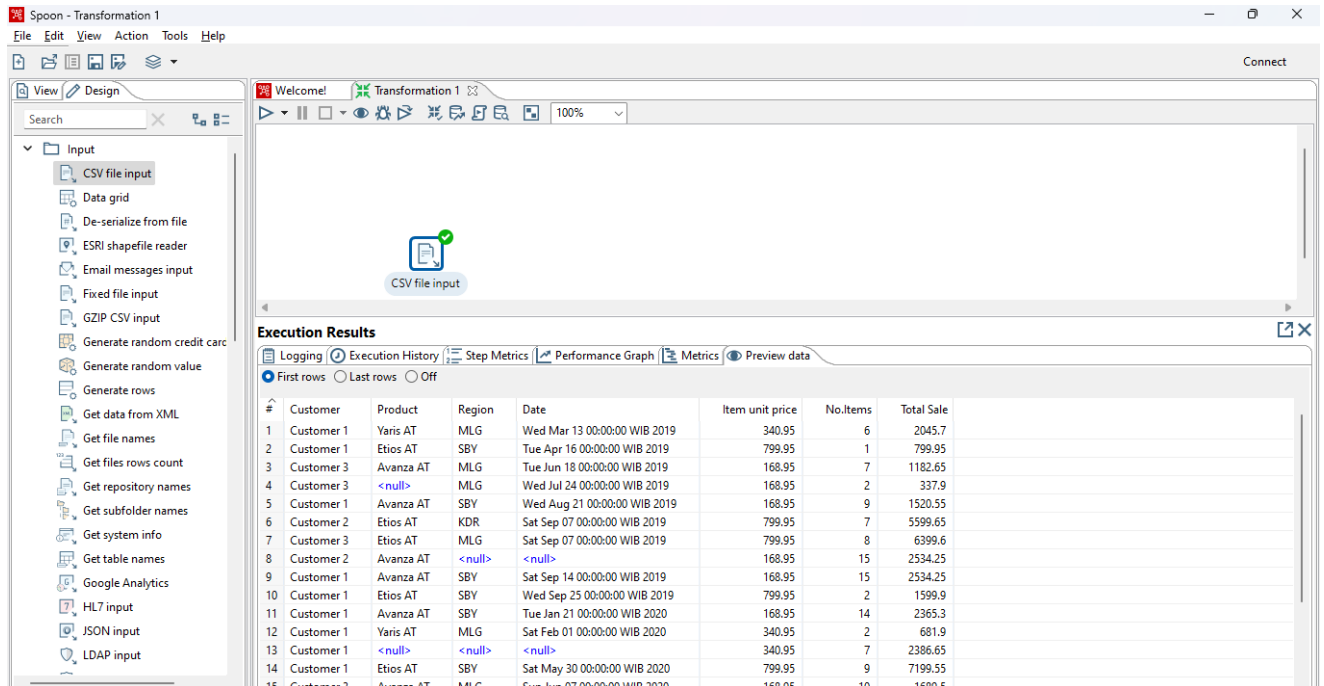


7. Tekan **Get Fields** untuk mengambil judul pada setiap kolom, pastikan nama-nama kolom telah sesuai



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.



TUGAS 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?
[Data hasil eksekusi tersebut, setelah saya cek hasilnya sesuai dengan data aslinya.](#)
2. 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?
[Berdasarkan perintah Paduan jobsheet diatas, Langkah Langkah yang diperlukan untuk melakukan extract data pada PDI Spoon:](#)
 1. Double-click pada objek CSV file input hingga muncul jendela konfigurasi.
 2. Mengubah Step Name menjadi Input Data CSV.
 3. Pilih Lokasi file CarDistributionSales.csv melalui Browse pada filename.
 4. Menyesuaikan delimiter (dalam kasus ini adalah koma ,)
 5. Tekan Get Fields untuk mengambil judul pada setiap kolom
 6. Tutup jendela konfigurasi tersebut.
 7. Tekan tombol "Run" pada pojok kanan atas Area Kerja, maka akan muncul Execution Result Area. Pilih tab Preview Data pada Execution Result Area.

3. Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E ?

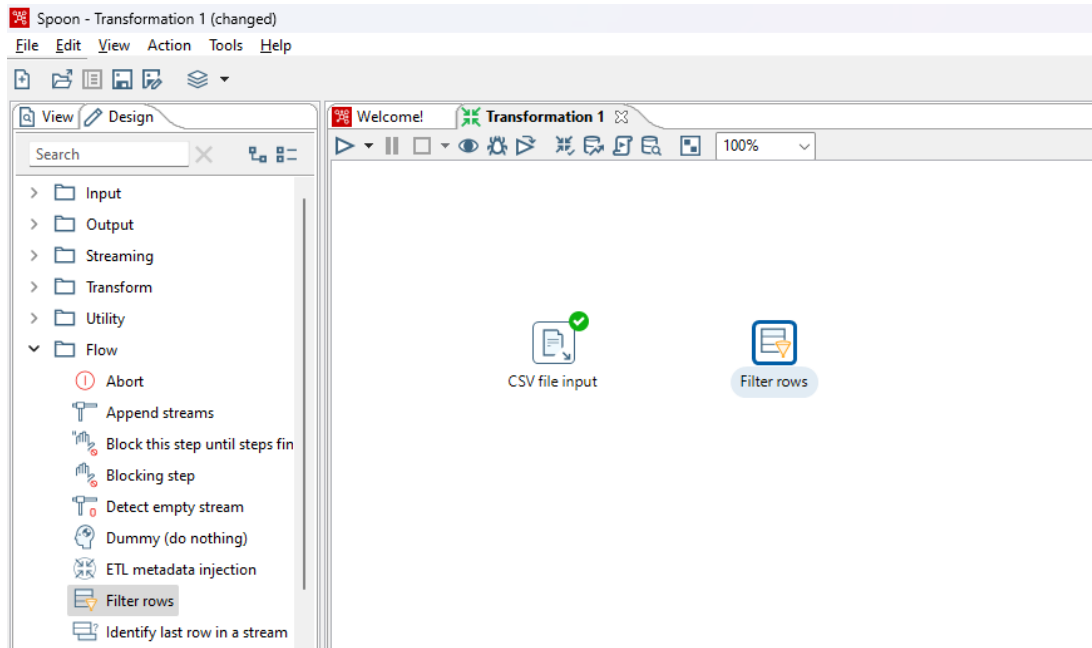
2024/02/17 06:51:48 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)

[Apa yang dimaksud dengan I,O,R,W,U,E](#)

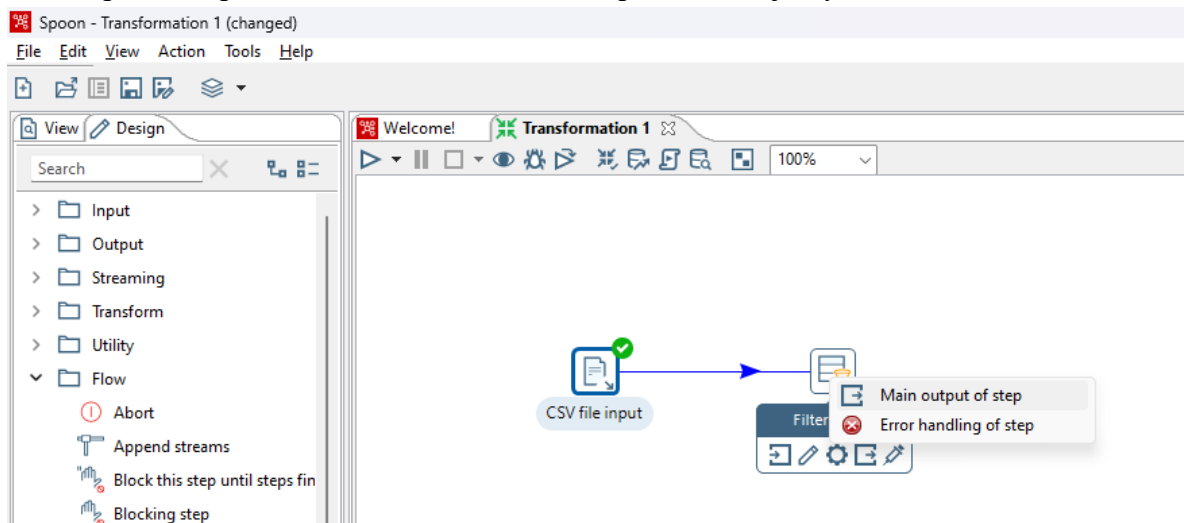
- I(Input): Jumlah data yang masuk ke dalam proses
 - O(Output) : Jumlah data yang keluar setelah proses
 - R(Read) : Jumlah data yang telah dibaca dari sumber
 - W(Write) : Jumlah data yang ditulis ke output
 - U(Update) : Jumlah data yang diperbarui
 - E(Error) : Jumlah data yang mengalami kesalahan dalam proses
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)

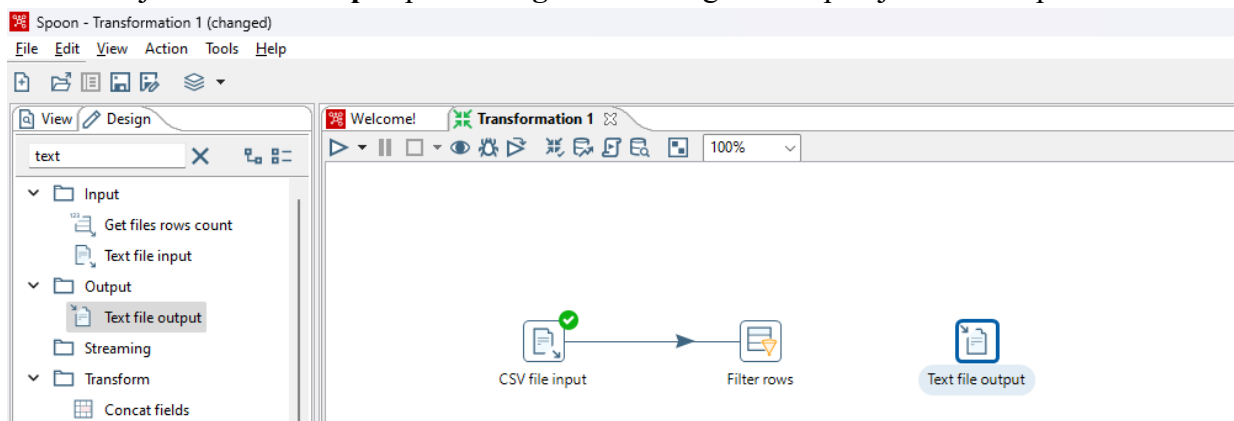
1. Carilah objek **Filter rows** pada **Design Area**. Drag and drop objek tersebut pada **Work Area**.



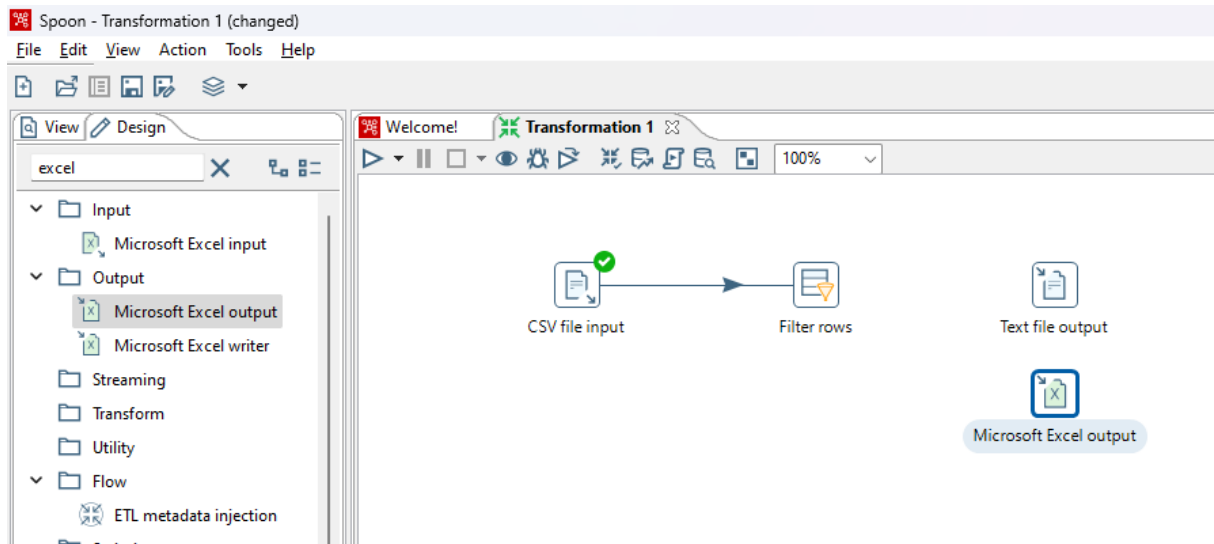
2. Hubungkan output pada **Input Data CSV** menuju **Filter rows**. Pilih koneksi **Main output of step** saat menghubungkan kedua objek tersebut.
Pada tahap ini Input Data CSV dihubungkan oleh konektor pada Filter rows memiliki makna bahwa setelah proses Input Data CSV dilakukan maka proses selanjutnya adalah Filter rows.



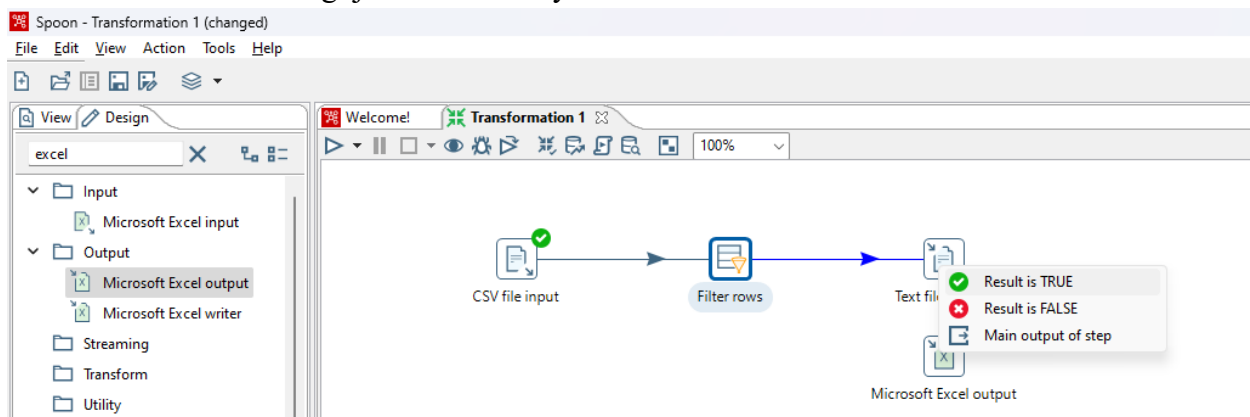
3. Carilah objek **Text file output** pada **Design Area**. Drag and drop objek tersebut pada **Work Area**.



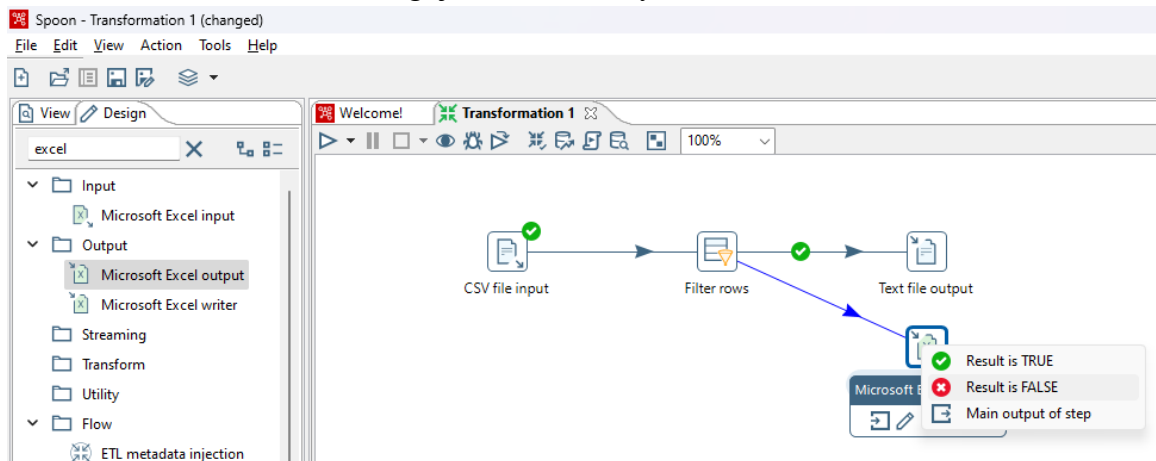
4. Carilah objek **Microsoft excel output** pada **Design Area**. Drag and drop objek tersebut pada **Work Area**.



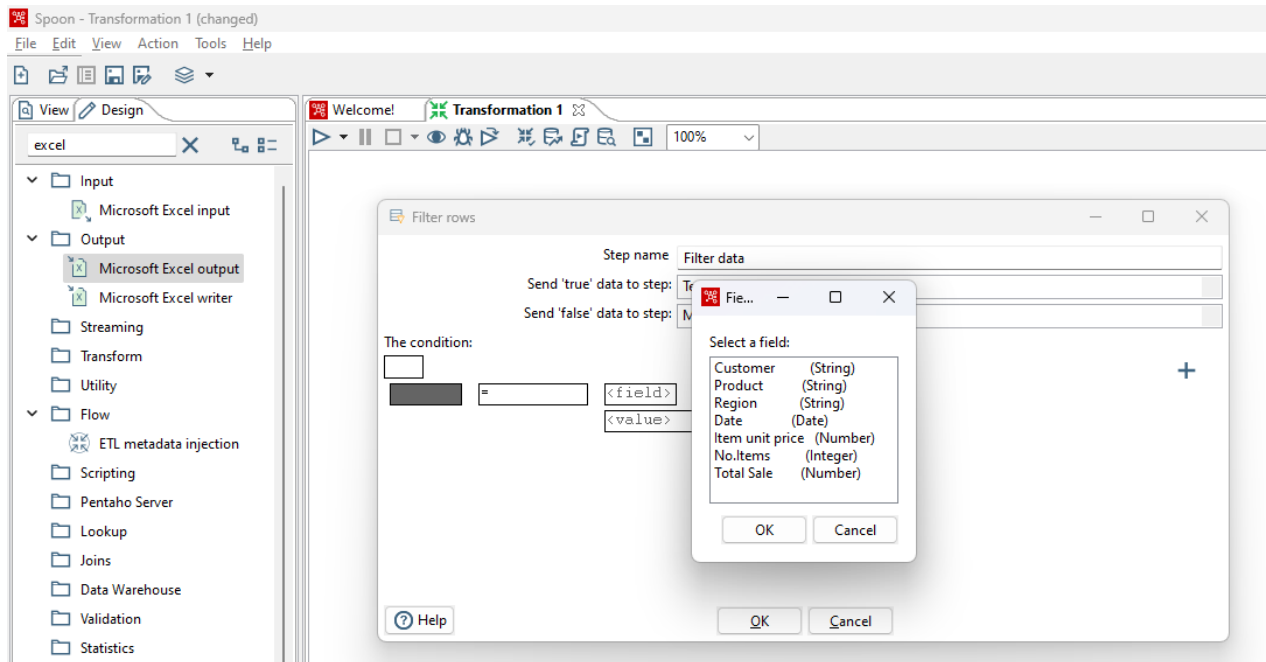
5. Hubungkan **filter rows** dengan **text file output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is true** sebagai jenis konektornya.



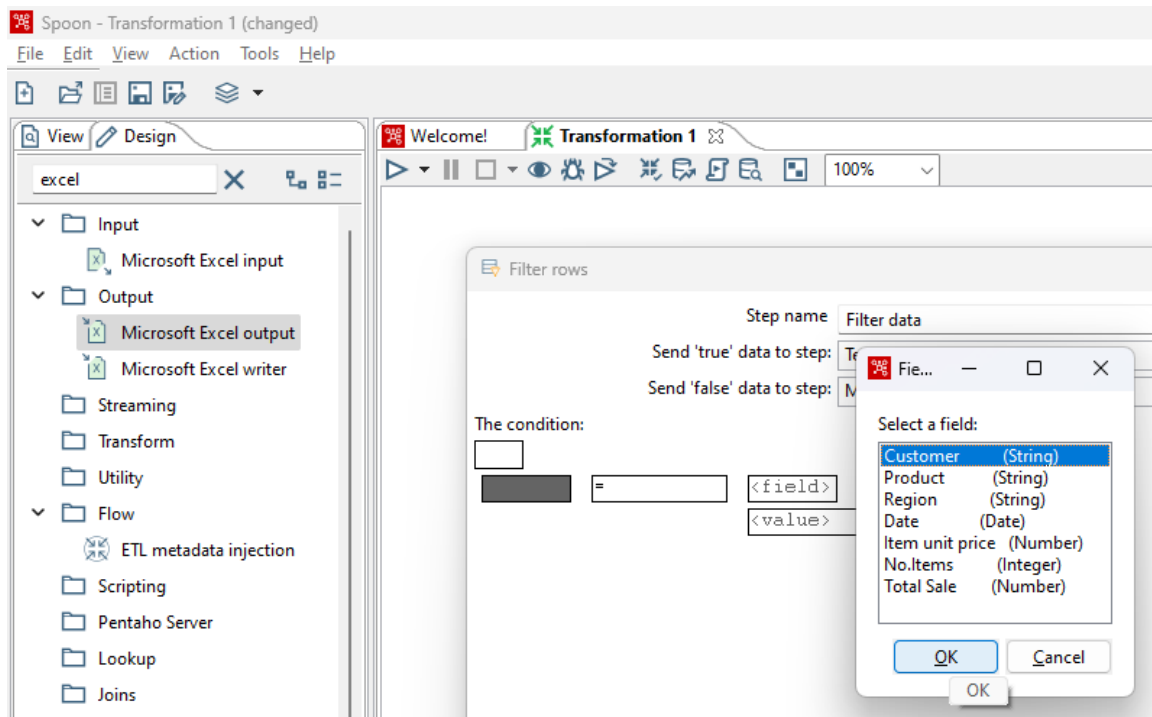
6. Hubungkan **filter rows** dengan **microsoft excel output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is false** sebagai jenis konektornya.



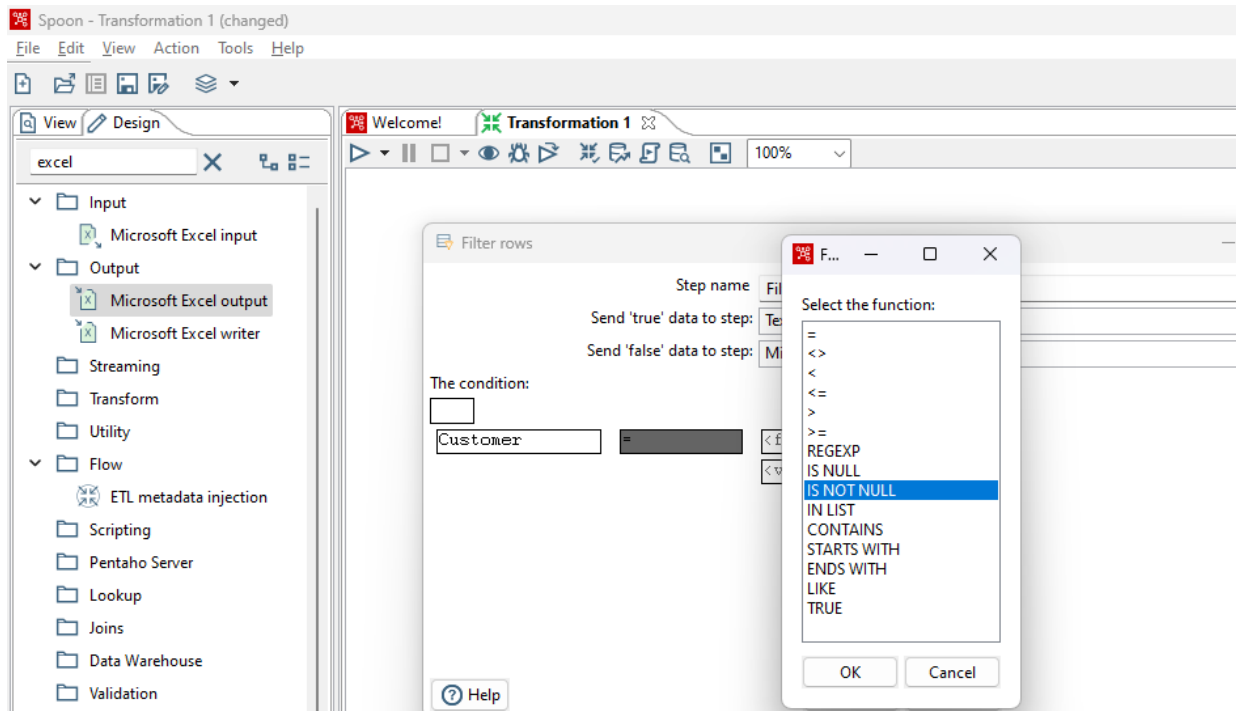
7. **Double click** pada **filter rows** hingga muncul jendela konfigurasi. ubah **step name** menjadi **Filter data**.
8. Cek bagian **send 'true' data to step** menjadi **Text file output** dan cek bagian **Send 'false' data to step** menjadi **Microsoft Excel output**.
9. Klik **<Field>** pada bagian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



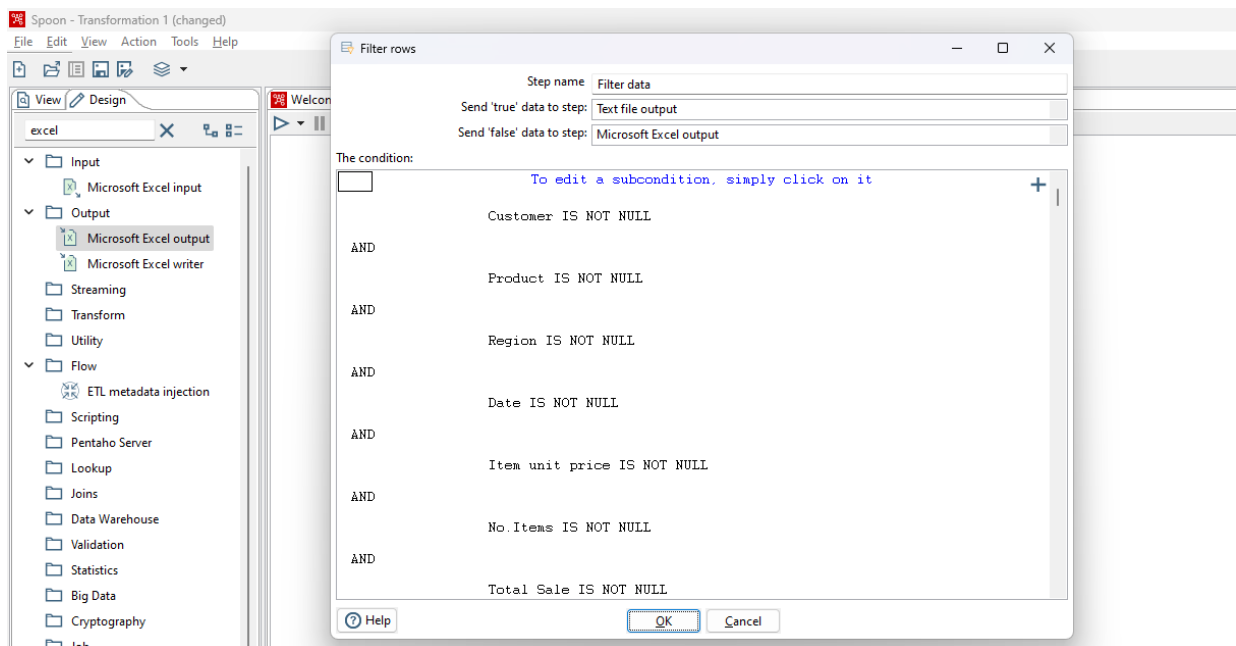
10. Pilih field yang diinginkan untuk digunakan sebagai kondisi. Pada kesempatan ini pilih **Customer**, kemudian klik **OK**.



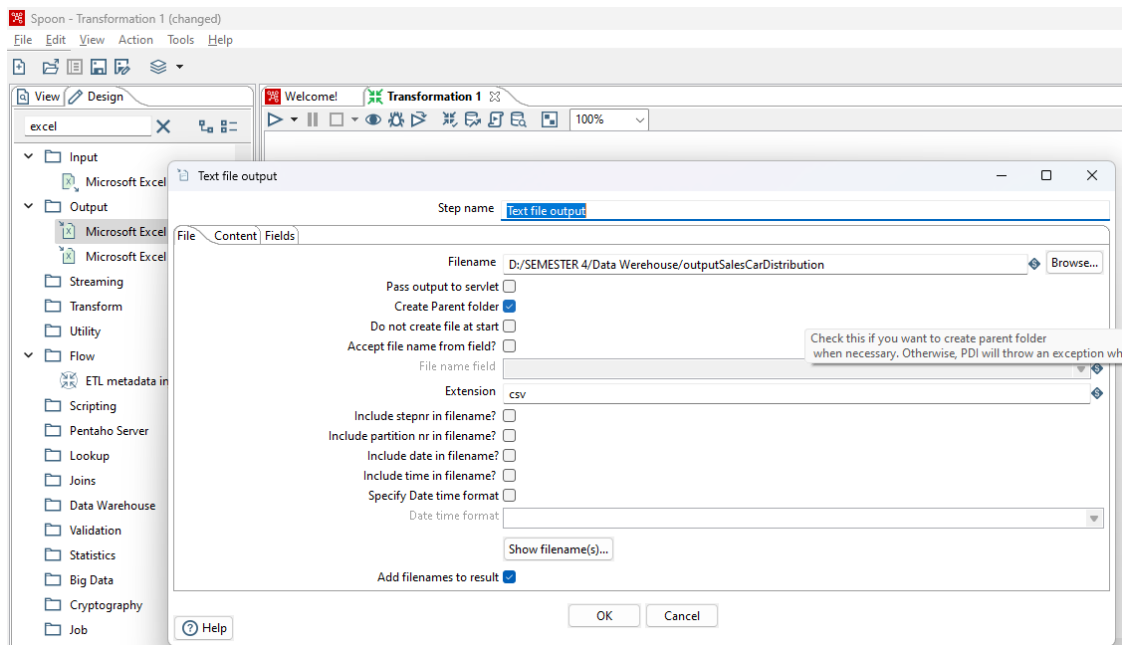
11. Pada bagian **value** pilih **IS NOT NULL** kemudian tekan **OK**, hal ini dilakukan karena data yang dipilih adalah data yang tidak Null



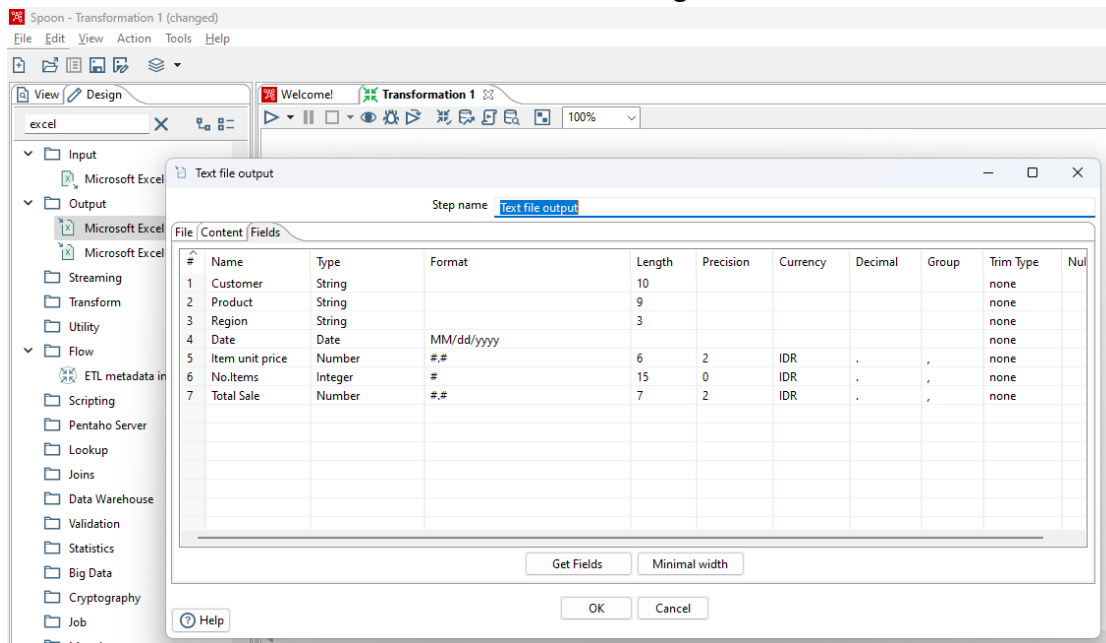
12. Tekan tombol + untuk menambah kondisi.
13. Pilih **statement** yang masih null untuk menambah kondisi.
14. Ulangi proses nomor 9,10,11 dengan mengganti semua **field** yang ada. hingga statement menjadi seperti pada gambar dibawah ini.
15. Gunakan operator **AND** untuk menghubungkan
16. Jika semua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



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**.



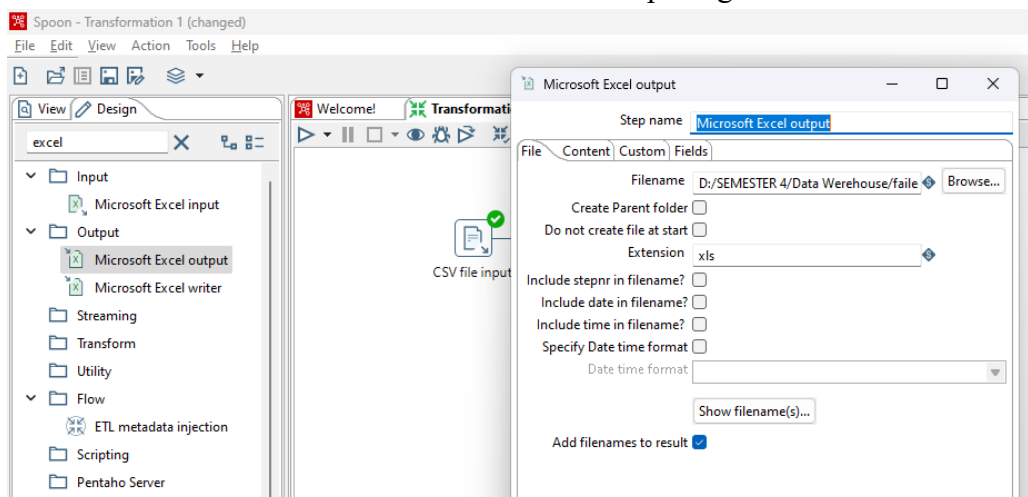
19. Pada tab **Fields**, tekan tombol **Get fields** untuk mengambil field data.



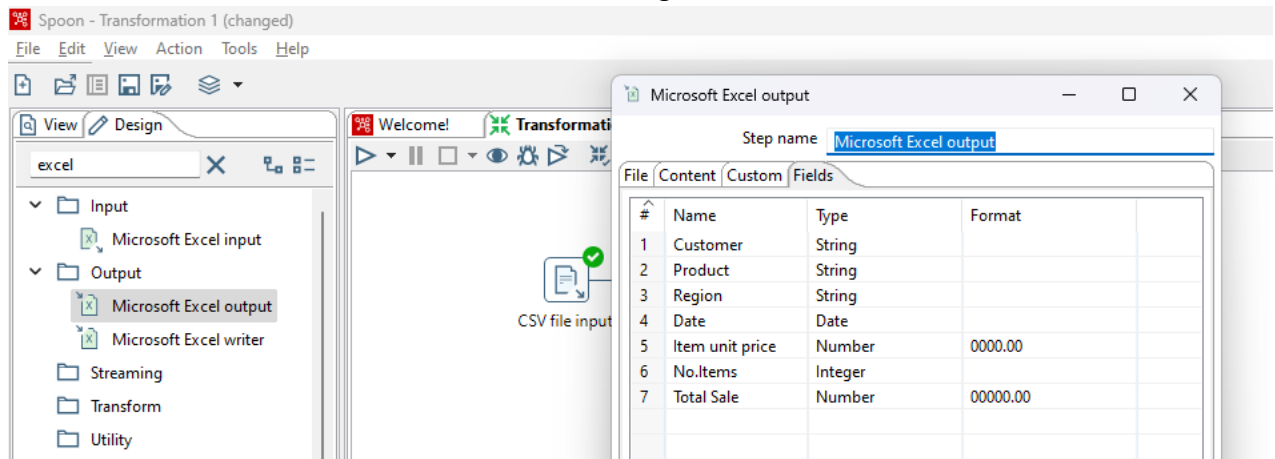
20. Tekan **ok** untuk kembali ke **work area**

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**.

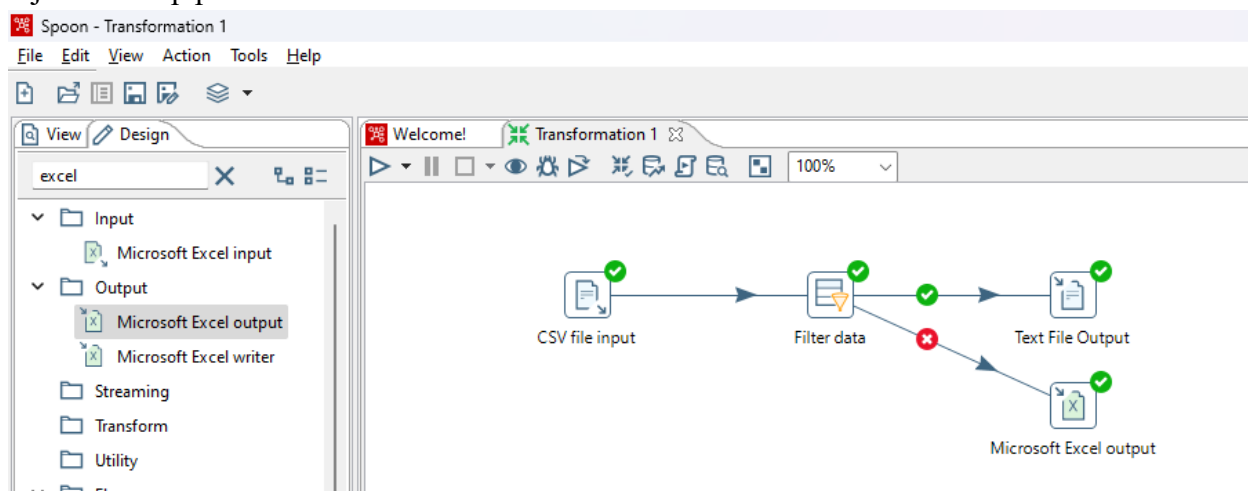


23. Pada tab **fields** tekan tombol **Get fields** untuk mengambil field data.



24. Tekan **ok** untuk kembali ke **work area**.

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



Langkah-langkah diatas merupakan proses melakukan filter data, dimana data yang lengkap akan diletakkan pada text file csv dan data yang belum lengkap akan diletakkan pada ms excel.

a. outputSalesCarDistribution.csv

| Customer | Product | Region | Date | Item unit price | No.Items | Total Sale |
|------------|-----------|--------|------------|-----------------|----------|------------|
| Customer 1 | Yaris AT | MLG | 3/13/2019 | 340.9 | 6 | 2045.7 |
| Customer 1 | Etios AT | SBY | 4/16/2019 | 800 | 1 | 800 |
| Customer 3 | Avanza AT | MLG | 6/18/2019 | 168.9 | 7 | 1182.7 |
| Customer 1 | Avanza AT | SBY | 8/21/2019 | 168.9 | 9 | 1520.5 |
| Customer 2 | Etios AT | KDR | 09/07/2019 | 800 | 7 | 5599.6 |
| Customer 3 | Etios AT | MLG | 09/07/2019 | 800 | 8 | 6399.6 |
| Customer 1 | Avanza AT | SBY | 9/14/2019 | 168.9 | 15 | 2534.2 |
| Customer 1 | Etios AT | SBY | 9/25/2019 | 800 | 2 | 1599.9 |
| Customer 1 | Avanza AT | SBY | 1/21/2020 | 168.9 | 14 | 2365.3 |
| Customer 1 | Yaris AT | MLG | 02/01/2020 | 340.9 | 2 | 681.9 |
| Customer 1 | Etios AT | SBY | 5/30/2020 | 800 | 9 | 7199.6 |
| Customer 3 | Avanza AT | MLG | 06/07/2020 | 168.9 | 10 | 1689.5 |
| Customer 3 | Yaris AT | MLG | 06/09/2020 | 340.9 | 4 | 1363.8 |
| Customer 2 | Avanza AT | KDR | 08/08/2020 | 168.9 | 12 | 2027.4 |

b. failedSalesCarDistribution.xls

| Customer | Product | Region | Date | Item unit price | No.Items | Total Sale |
|------------|-----------|--------|------------------|-----------------|----------|------------|
| Customer 3 | | MLG | 24/07/2019 00:00 | 0168,95 | 2,00 | 00337,90 |
| Customer 2 | Avanza AT | | | 0168,95 | 15,00 | 02534,25 |
| Customer 1 | | | | 0340,95 | 7,00 | 02386,65 |
| Customer 2 | | KDR | 07/08/2020 00:00 | 0799,95 | 14,00 | 11199,30 |
| Customer 1 | Avanza AT | | 12/08/2020 00:00 | 0168,95 | 4,00 | 00675,80 |

TUGAS 3

1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?
Pada file excel, data yang ditampilkan merupakan data yang pada salah satu atau lebih isi kolomnya kosong. Sedangkan data yang berada pada file csv, merupakan data data yang isi kolomnya lengkap.
2. Jelaskan apa yang terjadi pada proses **Filter rows**!
Filter Rows adalah proses menyaring baris dalam suatu dataset berdasarkan kondisi tertentu. Ini berguna untuk menampilkan hanya data yang relevan, sehingga analisis lebih fokus dan efisien.
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 table.
Tidak sampai memakan waktu satu menit.