**Jobsheet 1: Data Warehouse Sederhana dan Pengenalan PDI Spoon**

**Studi Kasus**

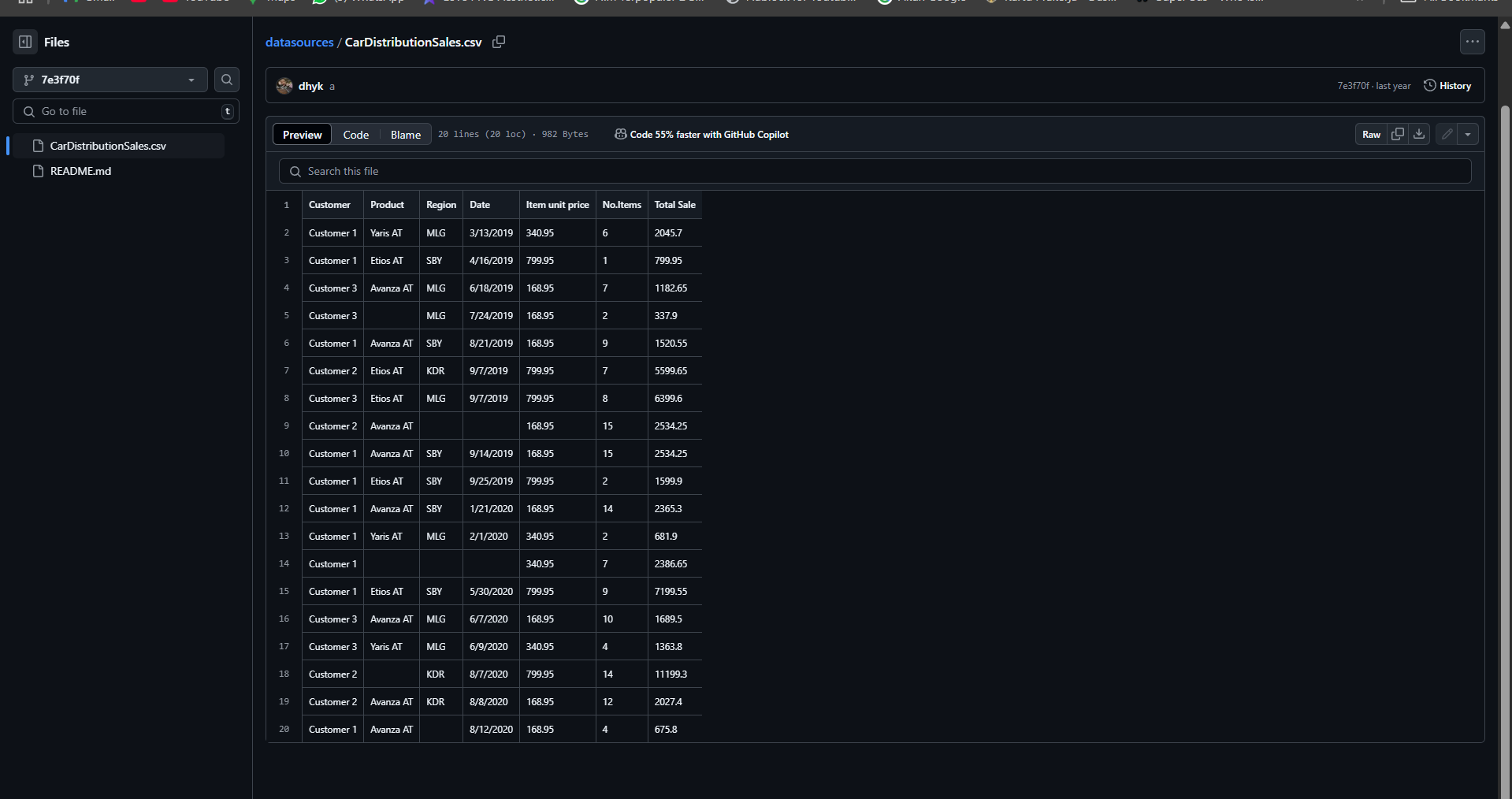
Pak Amir adalah Manager Sales Astro Mobil, yang bergerak di bidang distributor mobil wilayah Jawa Timur. Pak Amir meminta data penjualan dari beberapa cabang kepada supervisor. Data tersebut akan digunakan untuk membuat suatu Dashboard pengambilan keputusan. Dikarenakan proses permintaan ini dilakukan 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/CarDistributionSales.csv>



**TUGAS 1**

Analisa lah data tersebut!

1. Berapa jumlah kolom pada data tersebut?

Terdapat 7 Kolom yaitu :

* 1. **Customer** (Nama pelanggan)
  2. **Product** (Nama produk)
  3. **Region** (Wilayah penjualan)
  4. **Date** (Tanggal transaksi)
  5. **Item unit price** (Harga satuan produk)
  6. **No.Items** (Jumlah produk yang dibeli)
  7. **Total Sale** (Total harga penjualan)

1. Apa arti atau isi data dari setiap kolom yang ada?
   1. **Customer** (Nama pelanggan)

Nama Pembeli

* 1. **Product** (Nama produk)

Nama Product

* 1. **Region** (Wilayah penjualan)

Kode Wilayah Penjualan ex: MLG = Malang, SBY = Surabaya

* 1. **Date** (Tanggal transaksi)

Tanggal Melakukan Transaksi

* 1. **Item unit price** (Harga satuan produk)

Harga Per-Unit

* 1. **No.Items** (Jumlah produk yang dibeli)

Keterangan Jumlah barang yang di beli dijabarkan dengan angka (Int)

* 1. **Total Sale** (Total harga penjualan)

Total harga penjualan yang sudah di beli

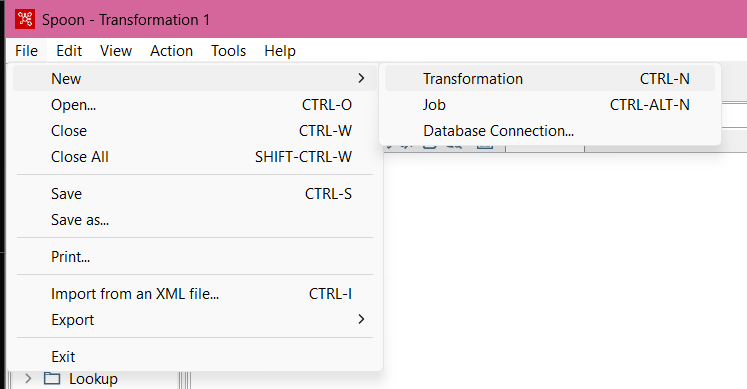
1. Adakah data yang memiliki nilai null / data yang tidak lengkap?

Ya ada, beberapa data dari customer

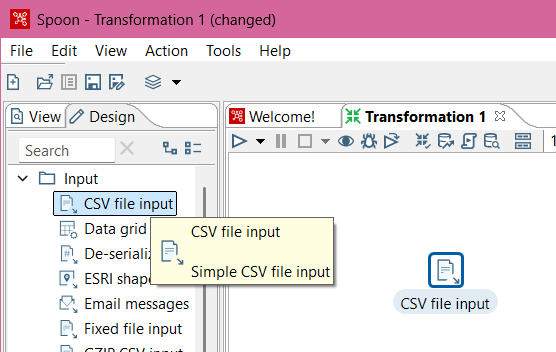
1. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

Tidak tidak ada

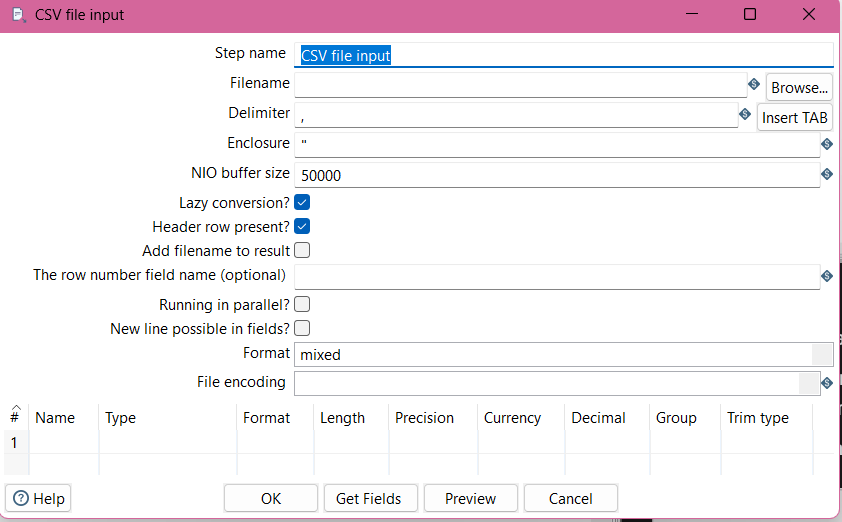
1. **Pengambilan Data (Extract)**
2. **Buka lembar kerja Transformation melalui toolbar File.**

****

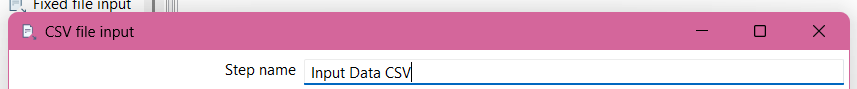
1. **Cari lah objek** CSV file input **pada** Design Area**. Drag and drop objek tersebut menuju** Work Area.

****

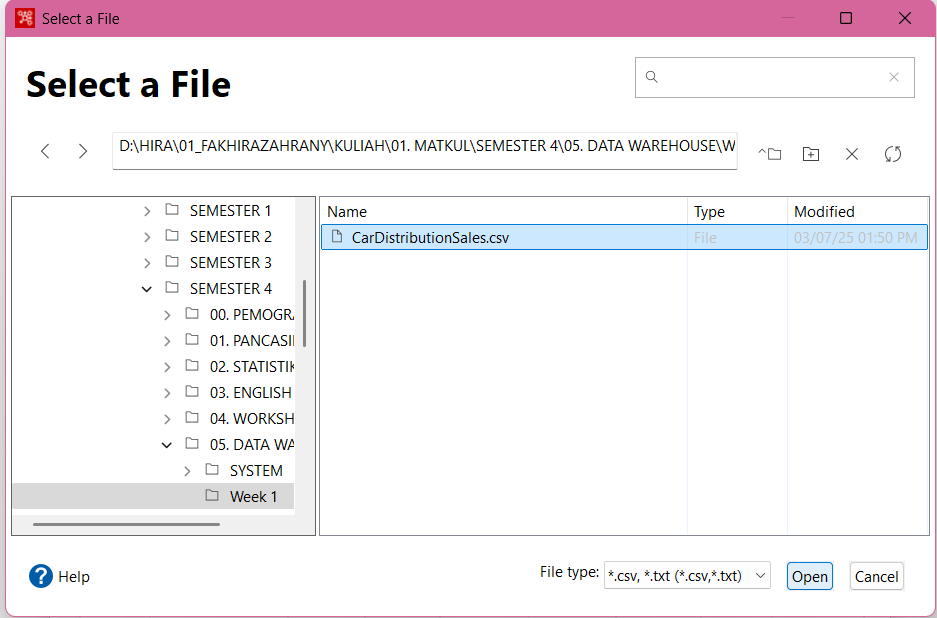
1. **Double-click pada objek CSV file input hingga muncul jendela konfigurasinya.**

****

1. **Ubah nama** Step name **menjadi** Input Data CSV**, (penamaan digunakan untuk identifikasi saja, selanjutnya penamaan boleh menyesuaikan).**

****

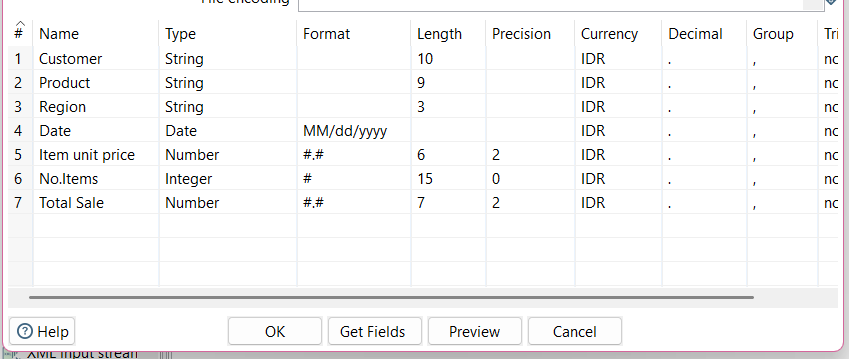
1. **Pilih lokasi file** CarDistributionSales.csv **melalui** Browse **pada** filename**.**

****

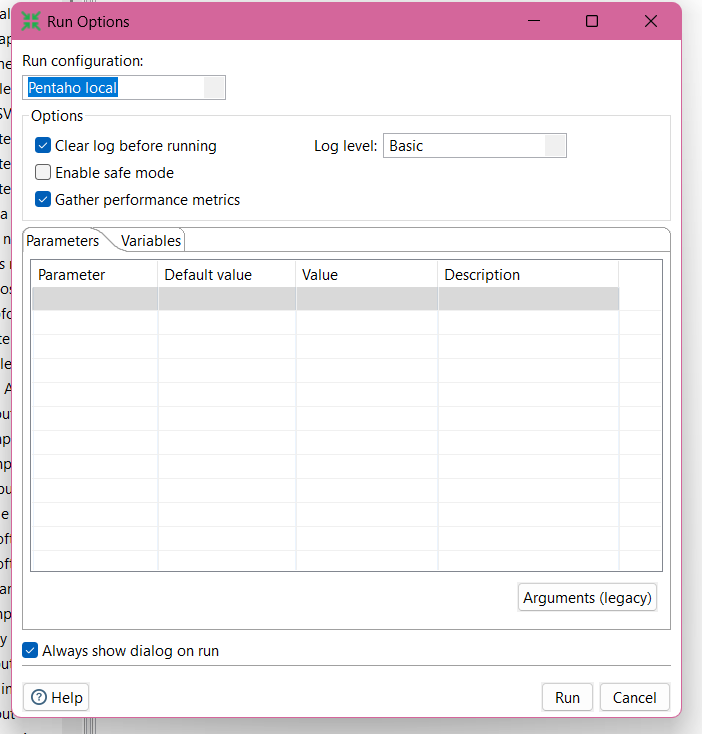
1. **Jika data telah dipilih pastikan pada bagian Delimiter sesuai dengan kondisi pemisah data pada file yang dipilih. (Pada kesempatan ini adalah comma " , " ).**

****

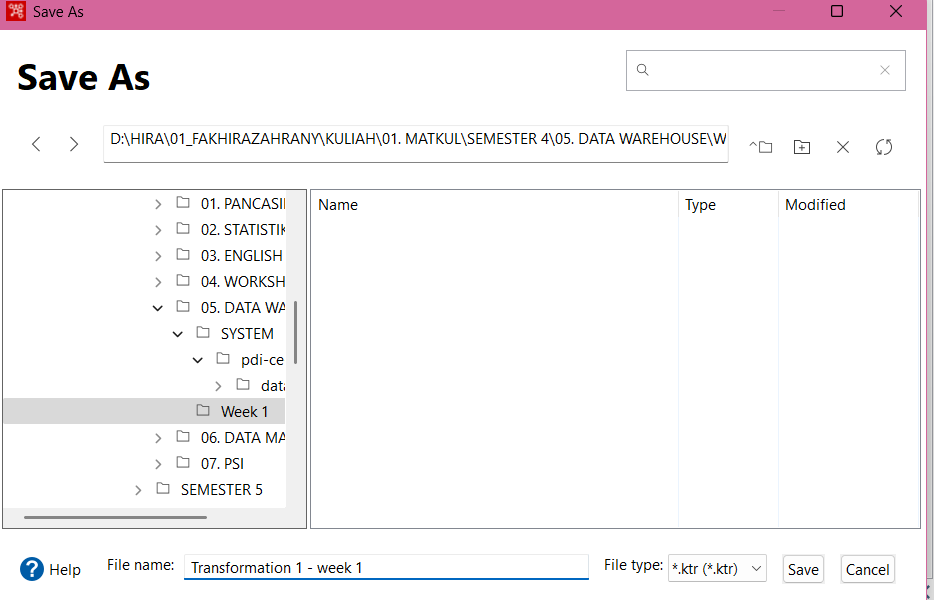
1. **Tekan Get Fields untuk mengambil judul pada setiap kolom, pastikan nama-nama kolom telah sesuai**

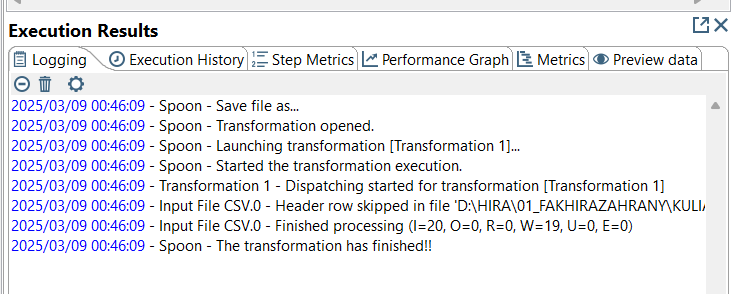
****

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

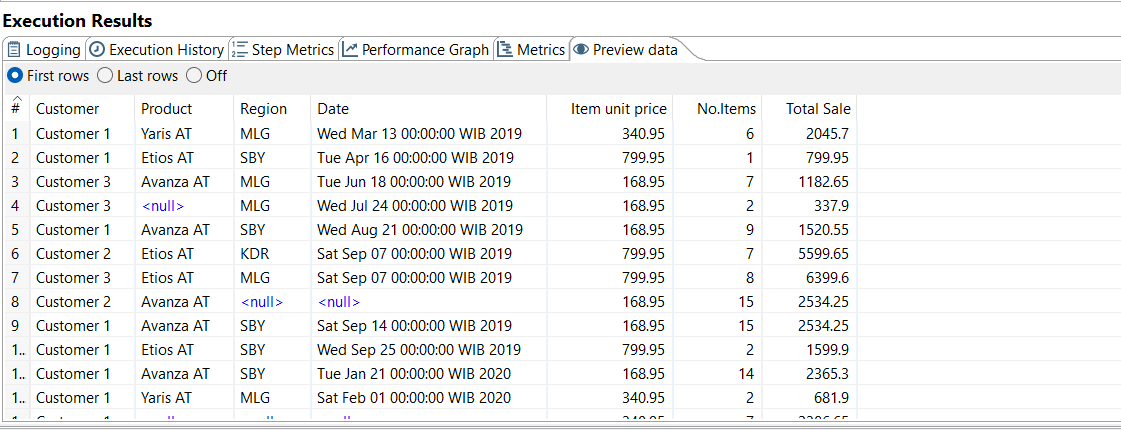
****

**Save the transformation First on local D:**

****

****

**Then see the Execution Result, and prefew the data**

****

****

**Look a like from the original file**

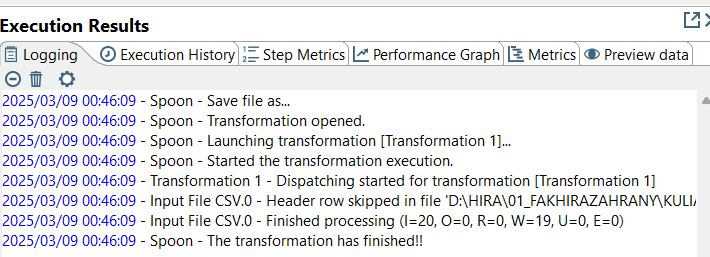
**Jika langkah-langkah diatas telah berhasil, maka proses extract telah berhasil dilakukan.**

**TUGAS 2**

1. **Apakah data hasil eksekusi sesuai dengan data aslinya?**

**Ya, sesuai**

1. **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?**

****

* 1. Spoon – Save File as = save the transformation on local Directory as .ktr
  2. Spoon - Transformation opened = open the CSV file
  3. Spoon - Launching transformation [Transformation 1]
  4. Spoon - Started the transformation execution.
  5. Transformation 1 - Dispatching started for transformation [Transformation 1]
  6. Input File CSV.0 - Header row skipped in file 'D:\HIRA\01\_FAKHIRAZAHRANY\KULIAH\01. MATKUL\SEMESTER 4\05. DATA WAREHOUSE\Week 1\CarDistributionSales.csv'
  7. Input File CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)
  8. Spoon - The transformation has finished!!

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

****

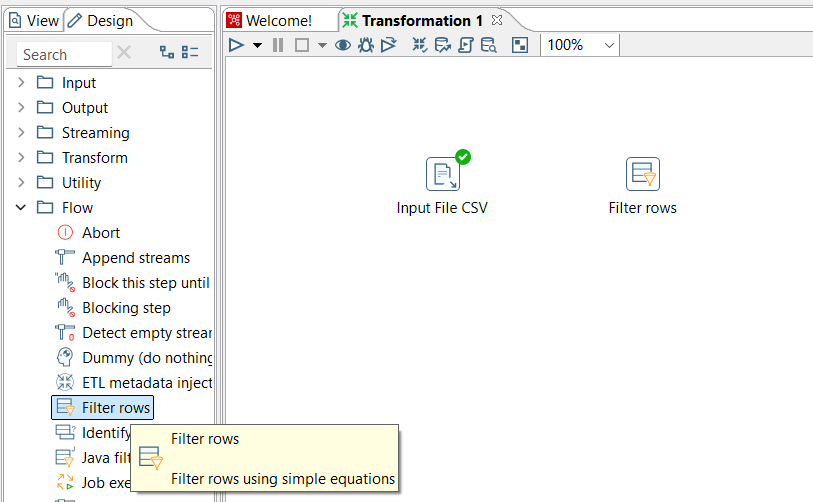
* 1. **I (Input) = 20 → Jumlah record yang dibaca dari file input (CSV).**
  2. **O (Output) = 0 → Jumlah record yang ditulis ke output (misalnya database atau file lain).**
  3. **R (Read) = 0 → Jumlah record yang dibaca ulang dalam transformasi.**
  4. **W (Written) = 19 → Jumlah record yang berhasil ditulis ke target.**
  5. **U (Updated) = 0 → Jumlah record yang diperbarui (jika ada operasi update).**
  6. **E (Error) = 0 → Jumlah record yang mengalami error selama proses eksekusi.**

1. **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.**

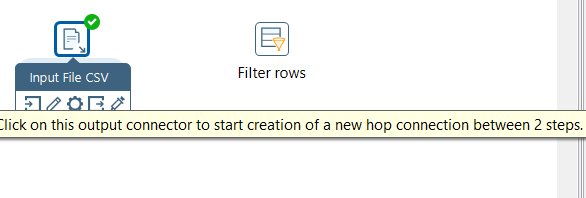
Based on my logging record, the execution need 1 seccond to di the execution

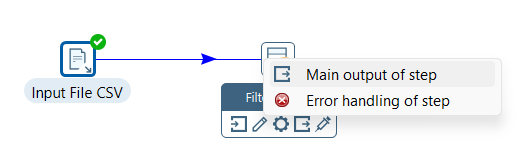
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | Execution Time | Processor | RAM | Storage | VGA | CPU Usage |
| Fakhira zahra | 1 Second | Intel® Core™ i7-13650HX (13th Gen, 20 Cores) | 16GB | SSD 477GB | Intel® UHD Graphics + NVIDIA GeForce RTX 4050 | 7% |
| Hanin Maryam | 0 seconds | Intel®  Core™ i7-12650H | 16GB | SSD  512 GB | GeForce RTX 2050 | 13% |
| Shabrina Rahmadhini |  |  |  |  |  |  |
| Sabrina Qotrunnada |  |  |  |  |  |  |

1. **Filter Data (Transform) dan Pengemasan data (Load)**
2. Carilah objek Filter rows pada Design Area. Drag and drop objek tersebut pada Work Area. Design – Flow – Filtter Row

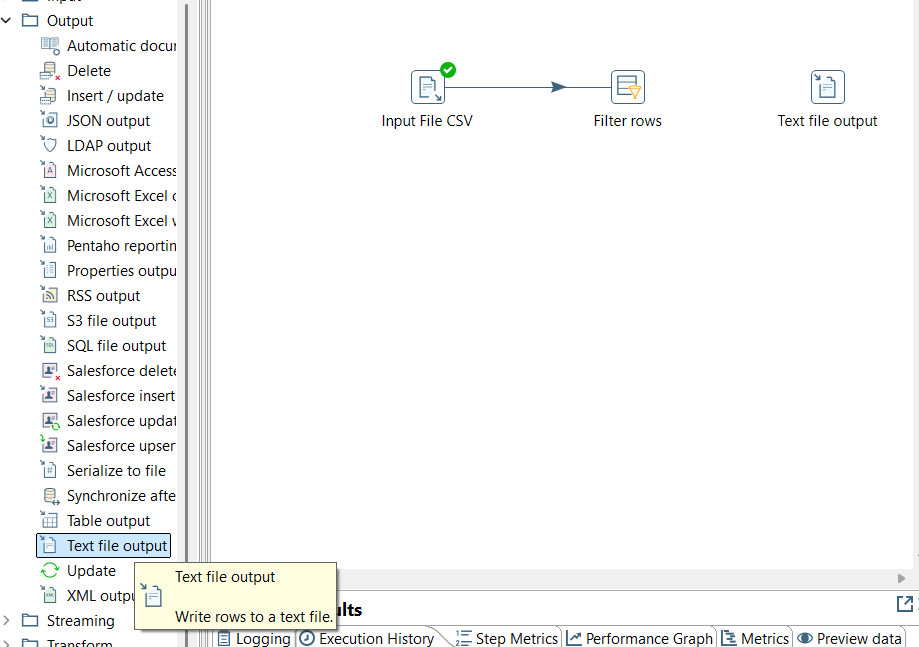


1. Hubungkan output pada **Input Data CSV** menuju **Filter rows**. Pilih koneksi **Main output of step** saat menghubungkan kedua objek tersebut. Pada tahap ini Input Data CSV dihubungkan oleh konektor pada Filter rows memiliki makna bahwa setelah proses Input Data CSV dilakukan maka proses selanjutnya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

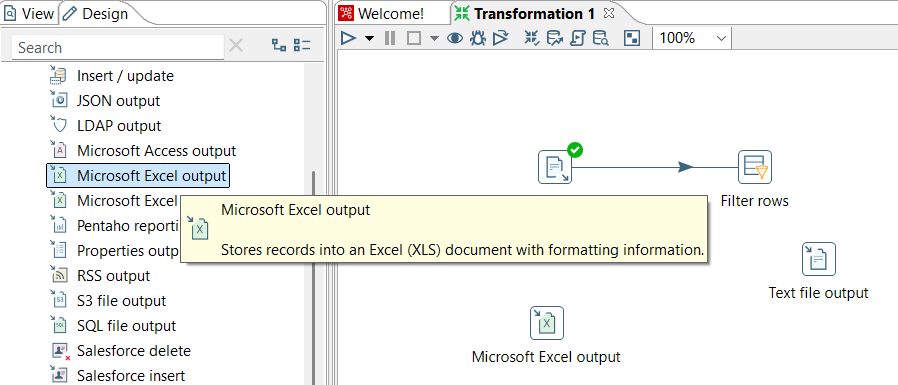




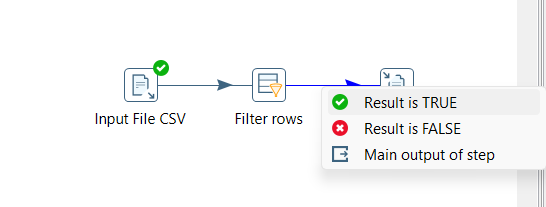
1. Carilah objek **Text file output** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**. Design - Output – Text File Output



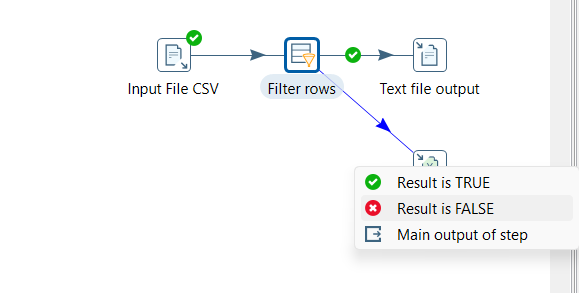
1. Carilah objek **Microsoft excel output** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.



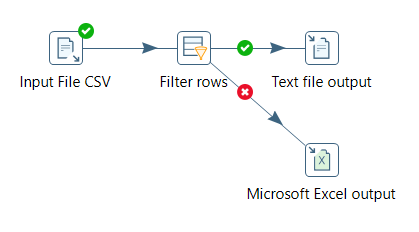
1. Hubungkan **filter rows** dengan **text file output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is true** sebagi jenis konektornya.



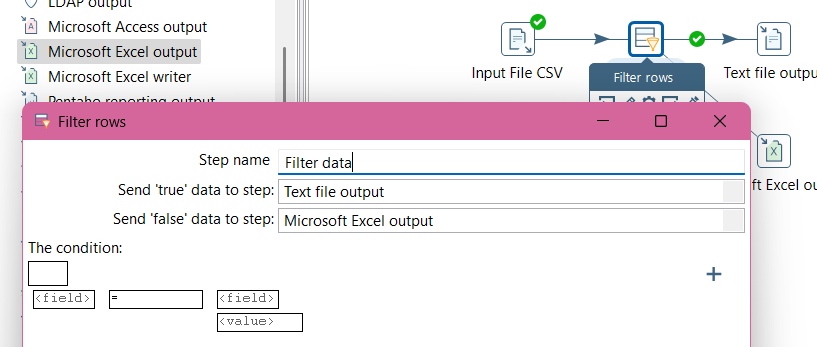
1. Hubungkan **filter rows** dengan **microsoft excel output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is false** sebagi jenis konektornya.



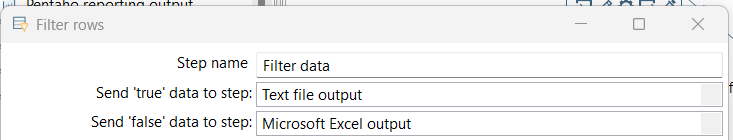
Result



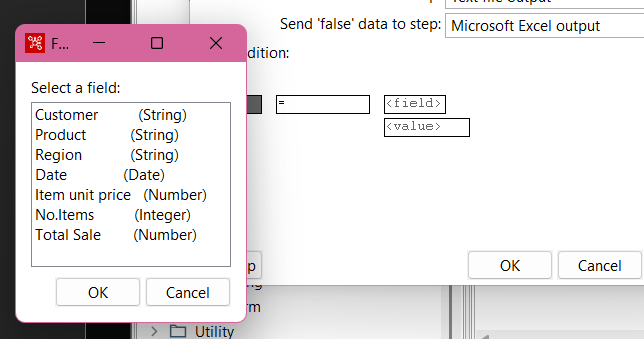
1. **Double click** pada **filter rows** hingga muncul jendela konfigurasi**.** ubah **step name** menjadi **Filter data**.



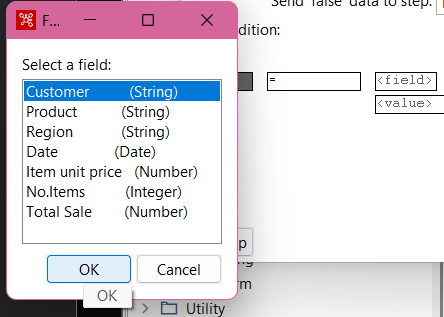
1. Cek bagian **send 'true' data to step** menjadi **Text file output** dan cek bagian **Send 'false' data to step** menjadi **Microsoft Excel output**.



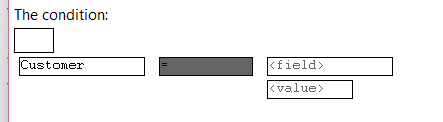
1. Klik **<Field>** pada bagian **condition**

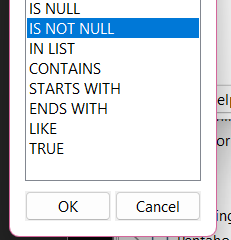


1. Pilih field yang diinginkan untuk digunakan sebagai kondisi. Pada kesempatan ini pilih **Customer,** kemudian klik **OK.**

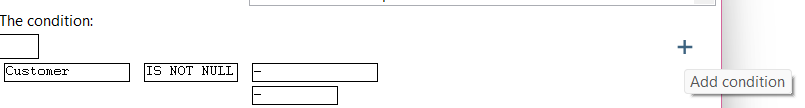


1. Pada bagian **value** pilih **IS NOT NULL** kemudian tekan **OK** , hal ini dilakukan karena data yang dipilih adalah data yang tidak Null

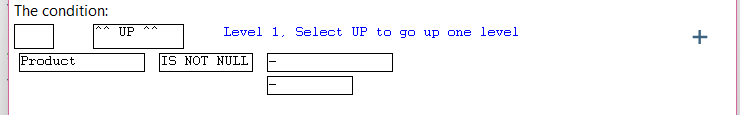




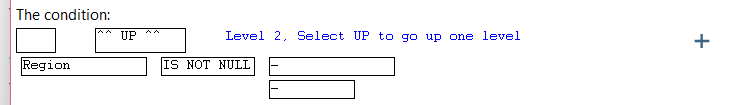
1. Tekan tombol **+** untuk menambah kondisi.



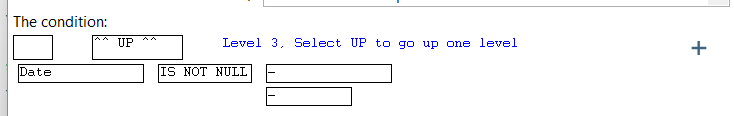
1. Pilih **statement** yang masih null untuk menambah kondisi.
2. Ulangi proses nomor 9,10,11 dengan mengganti semua **field** yang ada. hingga statement menjadi seperti pada gambar dibawah ini.
3. Product



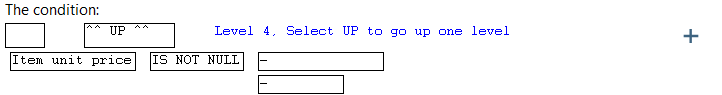
1. Region



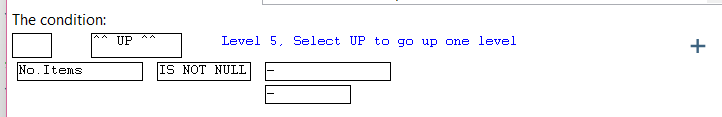
1. Date



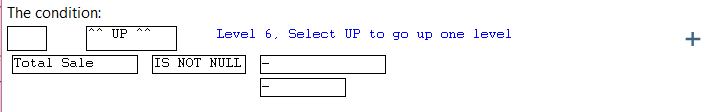
1. Item unit price



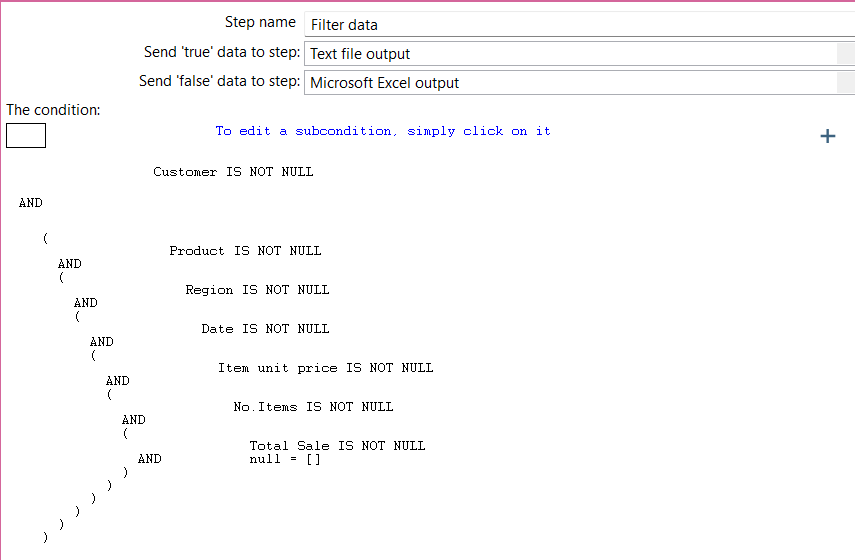
1. No.Items



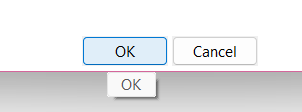
1. Total Sale



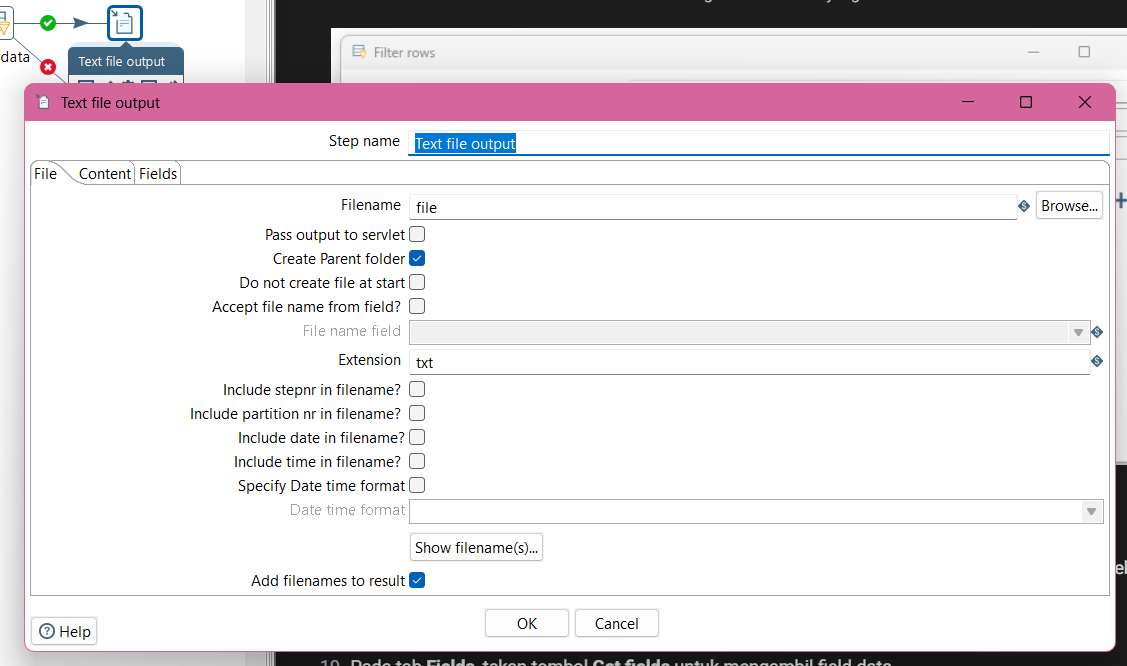
1. Gunakan operator **AND** untuk menghubung



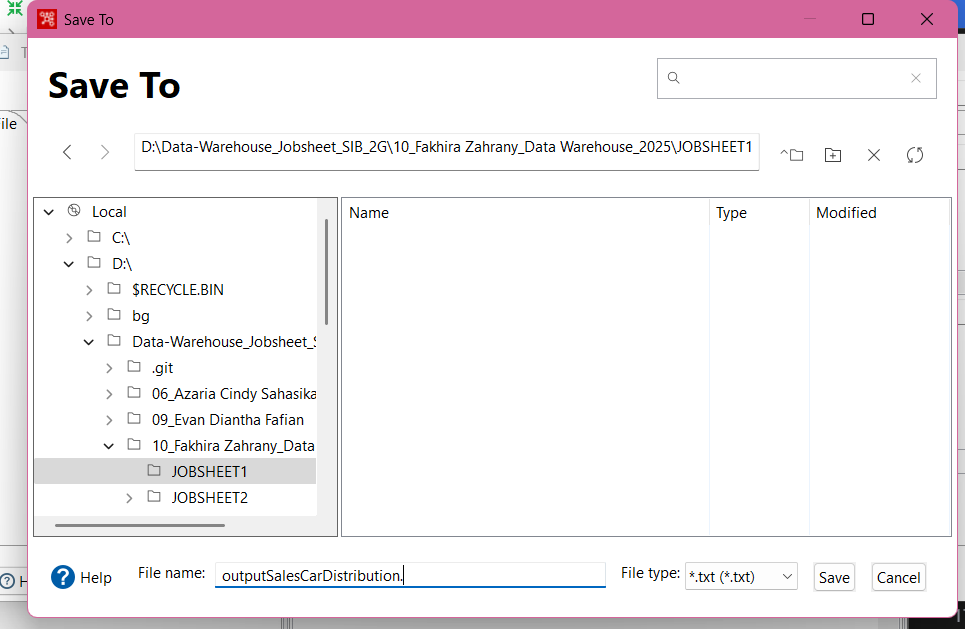
1. Jika semua field sudah masuk dalam kondisi maka tekan **OK**.



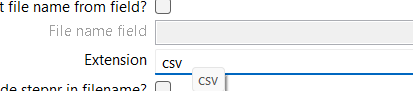
1. Double-click pada **text file output**, pilih lokasi file untuk menyimpan hasil output file dengan menekan tombol **browse**. beri nama **outputSalesCarDistribution**



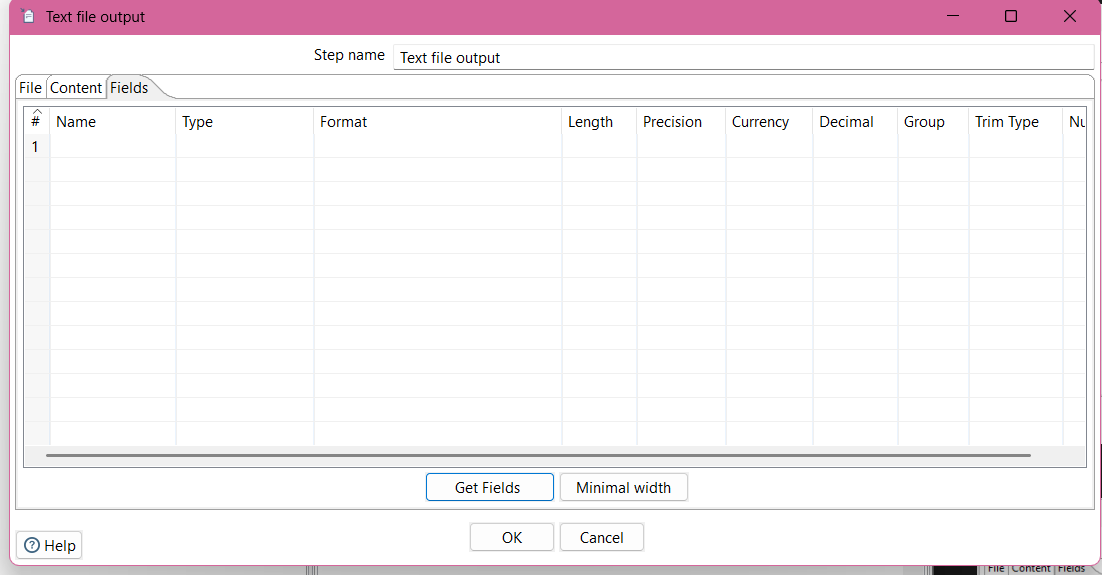
Browse - D:\Data-Warehouse\_Jobsheet\_SIB\_2G\10\_Fakhira Zahrany\_Data Warehouse\_2025\JOBSHEET1\outputSalesCarDistribution. And save there

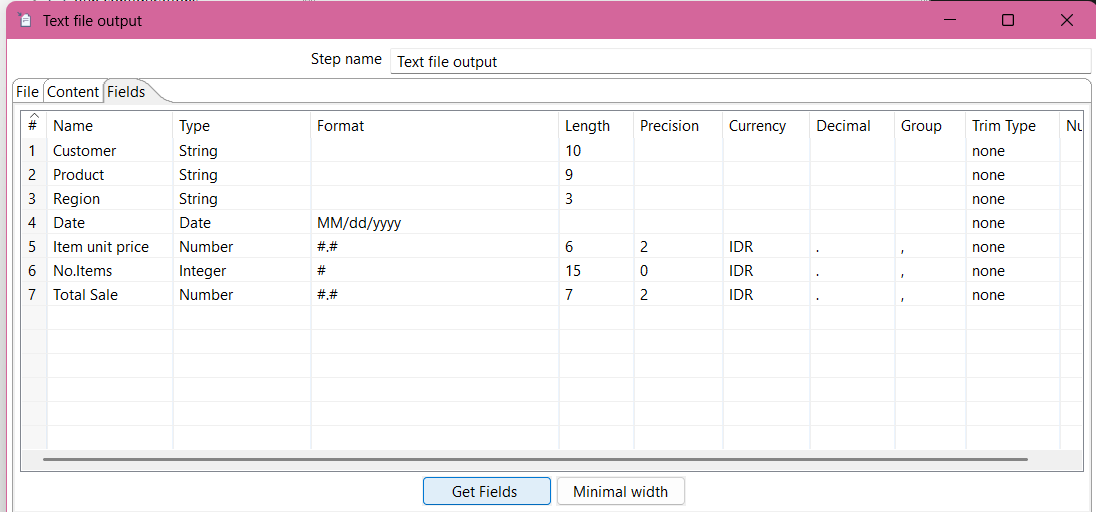


1. Ubah extension menjadi **csv.**

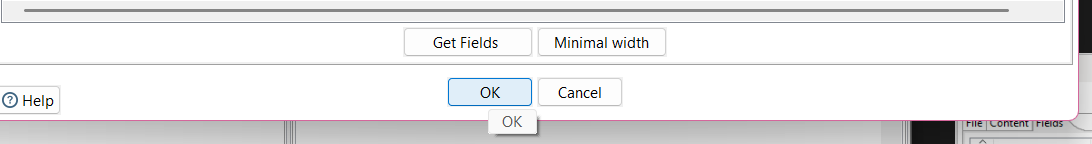


1. Pada tab **Fields**, tekan tombol **Get fields** untuk mengambil field data.

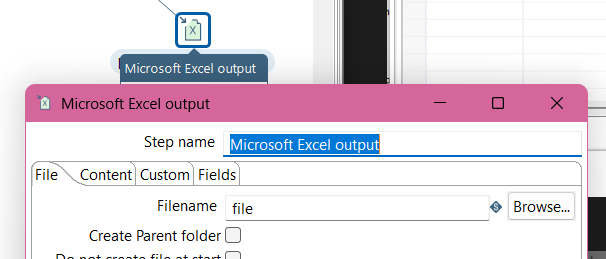




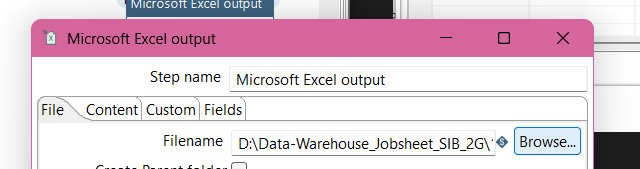
1. Tekan **ok** untuk kembali ke **work area**



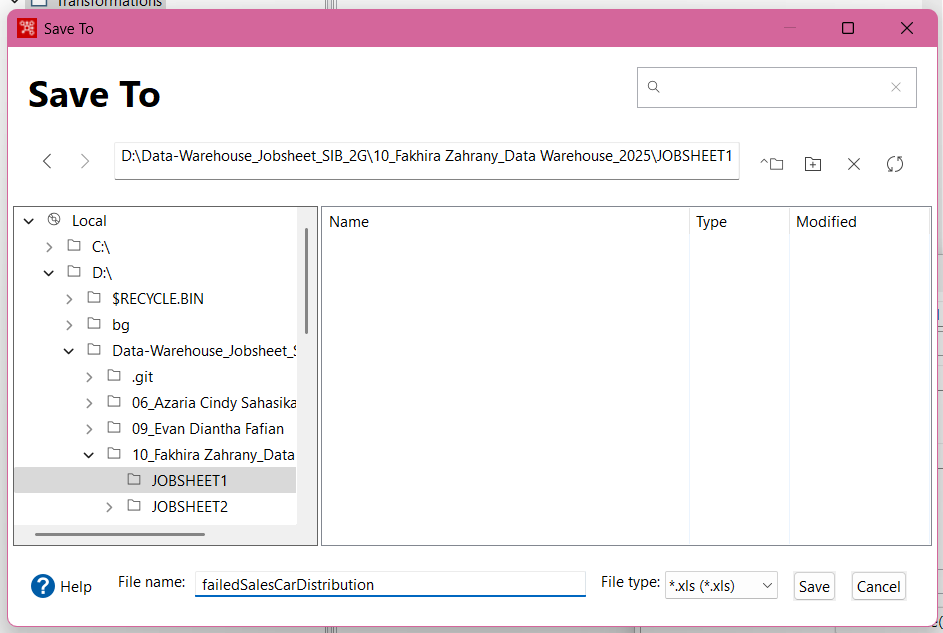
1. Hal yang sama dilakukan pada konfigurasi **Microsoft Excel output,** **Double-click** pada objek **Microsoft Excel output** hingga muncul jendela konfigurasi.



1. Pilih lokasi file untuk menyimpan file output dengan menekan tombol **browse**. Beri nama **failedSalesCarDistribution.** untuk extension tetap dengan extension **xls**.



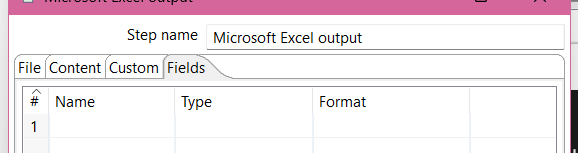
Browse - D:\Data-Warehouse\_Jobsheet\_SIB\_2G\10\_Fakhira Zahrany\_Data Warehouse\_2025\JOBSHEET1\failedSalesCarDistribution



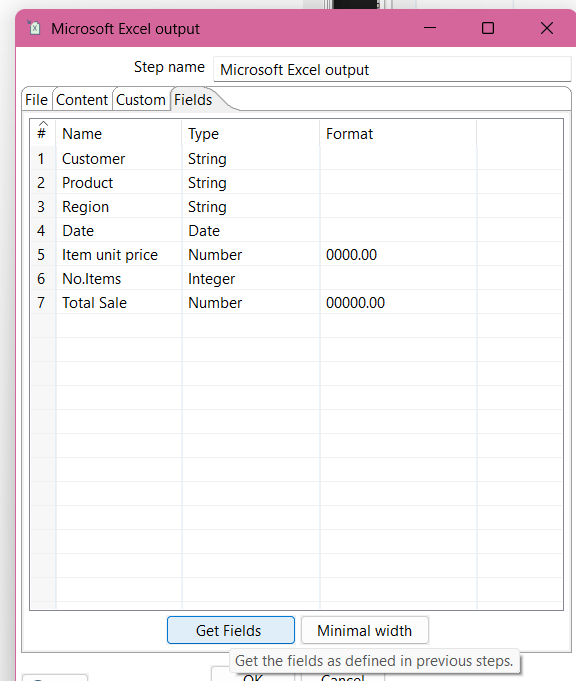
The extention still xls.



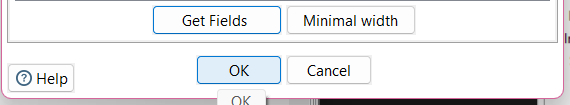
1. Pada tab **fields** tekan tombol **Get fields** untuk mengambil field data.



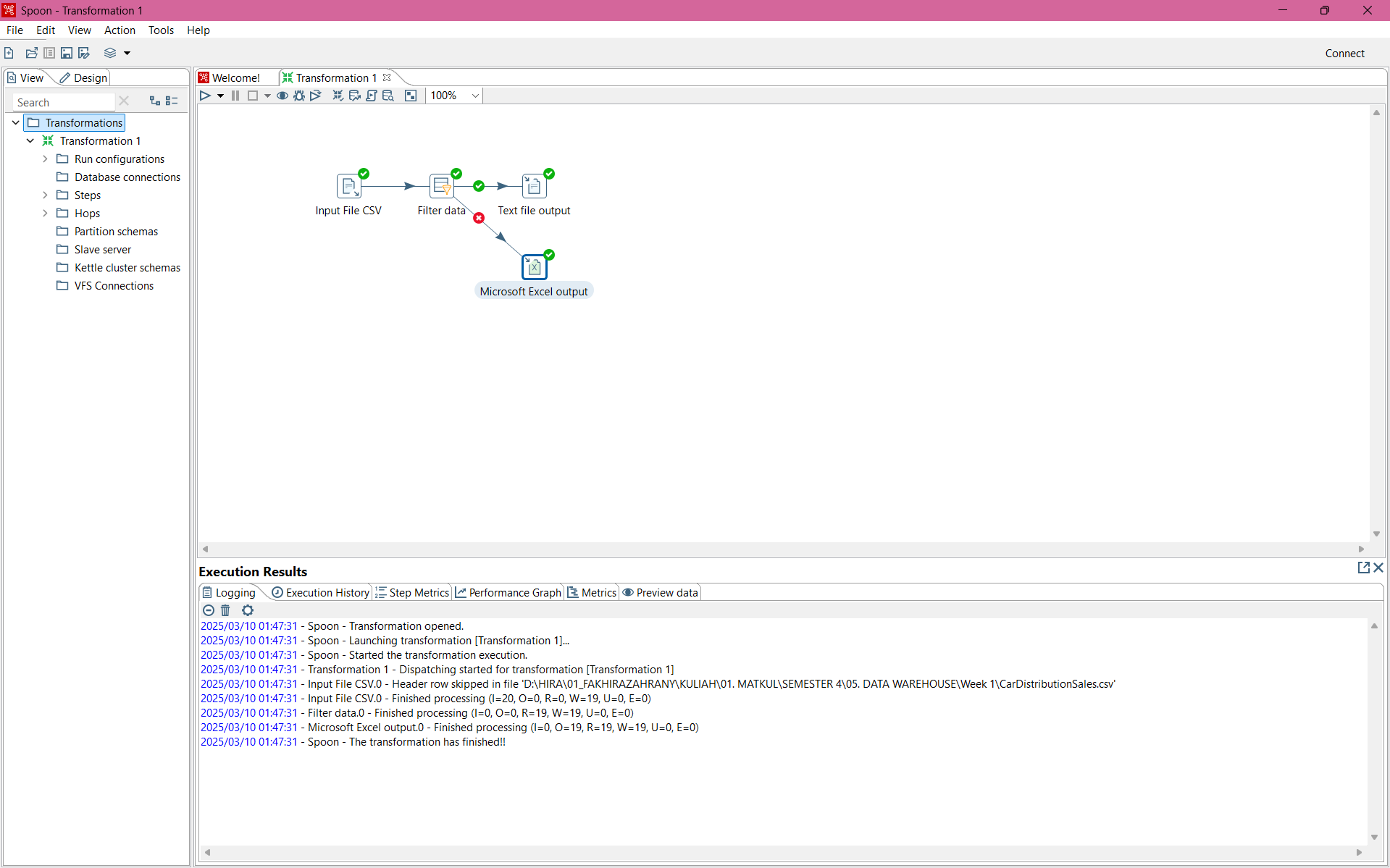
Push Get filed



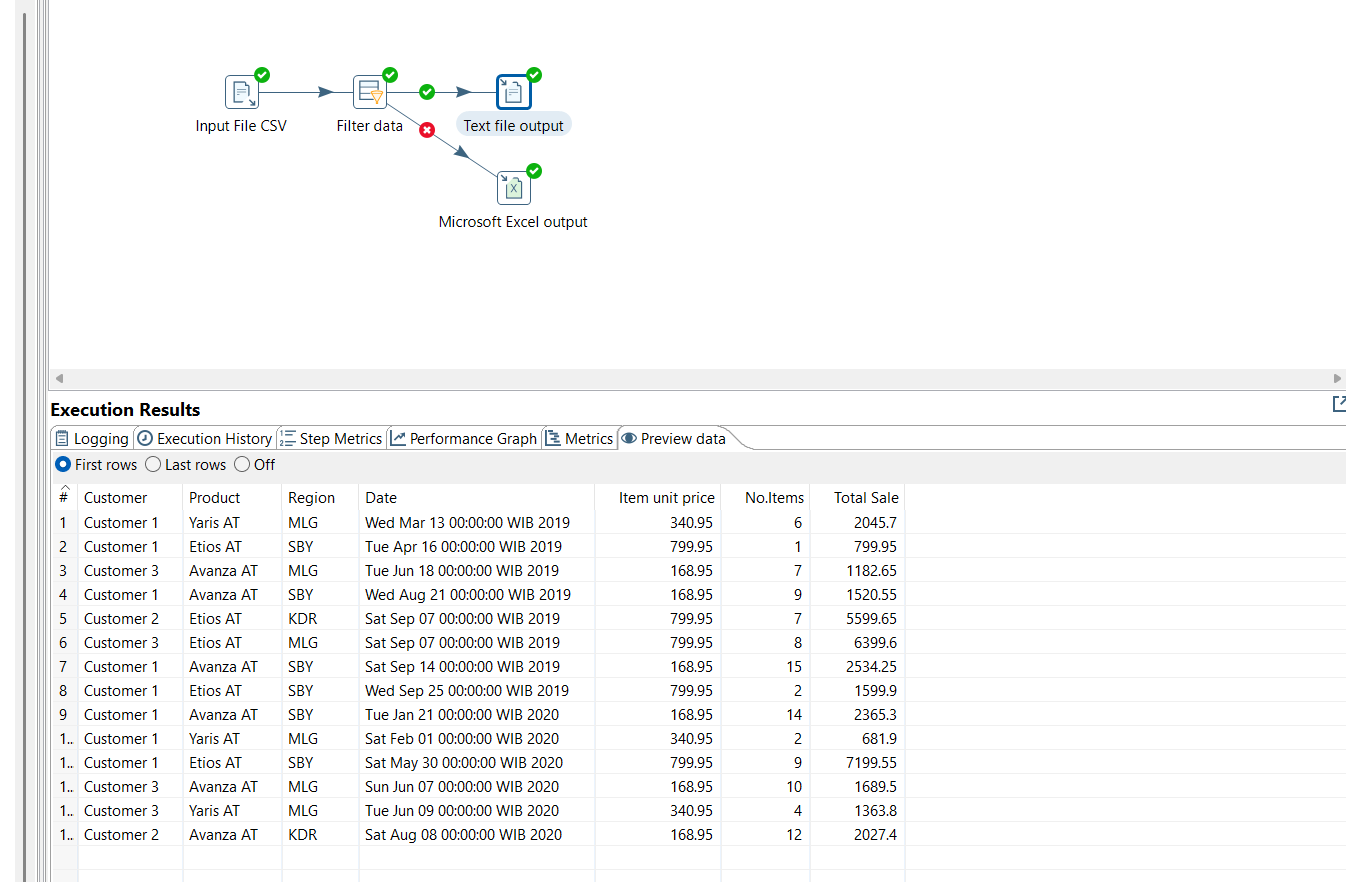
1. Tekoan **ok** untuk kembali ke **work area**.



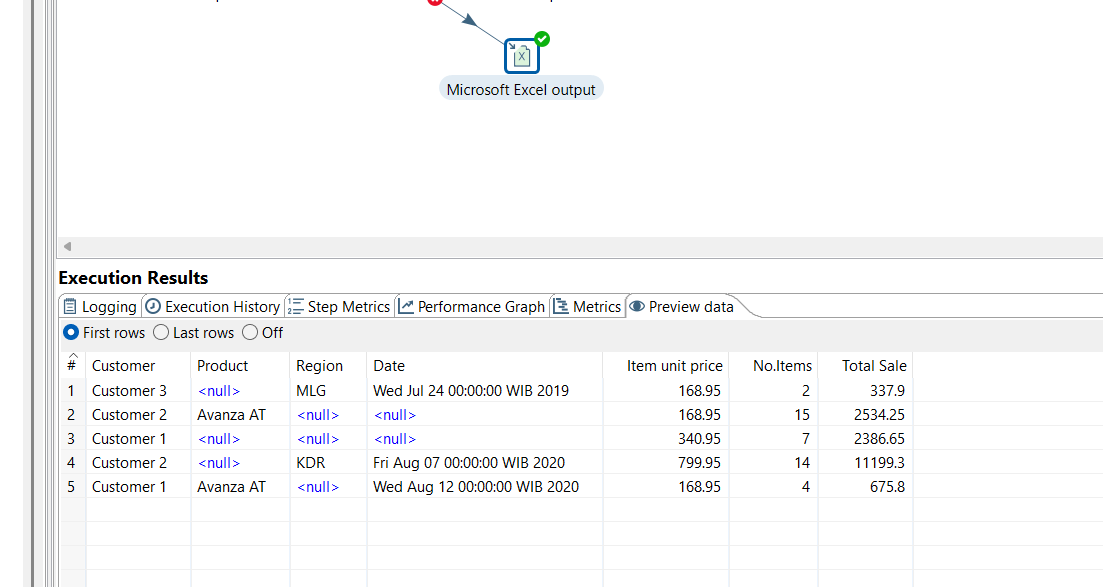
1. Tekan tombul **Run** pada pojok kiri atas **Work Area**. jika berhasil maka akan muncul tanda centang hijau di setiap proses.



1. Text File Output (outputSalesCarDistribution)



1. MCS Exel Output (failedSalesCarDistribution)



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.

**TUGAS 3**

1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?

For the CSV File appear the complate data ( not null ), beside on exel the data appear within Null condition wich is uncomplate

1. Jelaskan apa yang terjadi pada proses **Filter rows**!

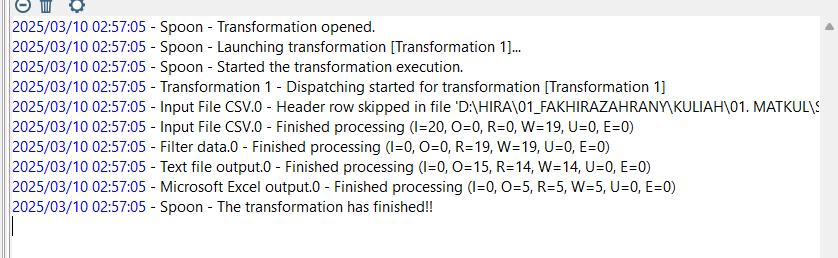
|  |
| --- |
| 2025/03/10 02:57:05 - Input File CSV.0 - Header row skipped in file 'D:\HIRA\01\_FAKHIRAZAHRANY\KULIAH\01. MATKUL\SEMESTER 4\05. DATA WAREHOUSE\Week 1\CarDistributionSales.csv' |
| 2025/03/10 02:57:05 - Input File CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0) |
| 2025/03/10 02:57:05 - Filter data.0 - Finished processing (I=0, O=0, R=19, W=19, U=0, E=0) |
| 2025/03/10 02:57:05 - Text file output.0 - Finished processing (I=0, O=15, R=14, W=14, U=0, E=0) |
| 2025/03/10 02:57:05 - Microsoft Excel output.0 - Finished processing (I=0, O=5, R=5, W=5, U=0, E=0) |
| 2025/03/10 02:57:05 - Spoon - The transformation has finished!! |

Based on this Log, on Filter Processed the system first reads the CarDistributionSales.csv file. Based on the log, the header row was skipped, meaning the first row was not treated as data. A total of 20 rows were read, but only 19 rows were successfully processed, possibly because one row was empty or had an issue.

Next, the data went through the Filter Rows step. The log shows that 19 rows were processed and written, which means the filter executed, but no rows were actually removed.

After filtering, the data that passed the filter was sent to the Text File Output step. The log indicates that 15 rows entered this process, but only 14 were successfully written to the file. Meanwhile, the data that did not meet the filter criteria was sent to Microsoft Excel Output, where 5 rows were written into an Excel file.

1. 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



Based on my log, the process of Extract, Transform, and Load (ETL) was completed in less than one second. The transformation started at 02:57:05, and by the same timestamp, it had already finished. This indicates that the execution was very fast, likely due to the small dataset or the efficiency of the processing pipeline.