**Jobsheet 1: Data Warehouse**

**Sederhana dan Pengenalan PDI Spoon**

Tujuan Praktikum

Setelah melakukan praktikum ini, mahasiswa diharapakan 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.

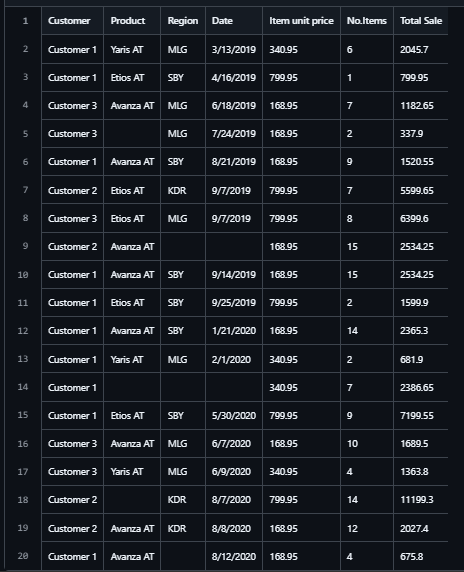
Dari studi kasus tersebut maka, akan dilakukan:

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

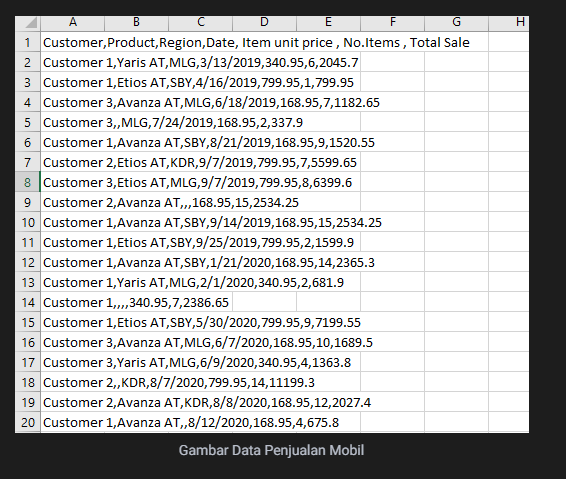
\*Data penjualan tersebut dapat diakses dan di download melalui link berikut:

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

* Data pada github



* Data pada exel :



TUGAS 1 :

1. Analisa lah data tersebut!
2. Berapa jumlah kolom pada data tersebut?

Data memiliki 7 kolom, yaitu:

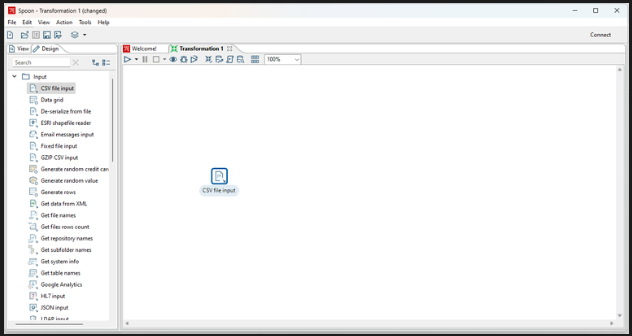
1. Customer = Identitas pelanggan.
2. Product = Produk yang dibeli.
3. Region = Wilayah tempat transaksi terjadi.
4. Date = Tanggal transaksi.
5. Item unit price = Harga satuan dari produk.
6. No.Items = Jumlah unit produk yang dibeli.
7. Total Sale = Total harga berdasarkan jumlah produk yang dibeli.
8. Apa arti atau isi data dari setiap kolom yang ada?
9. Customer: Identitas pelanggan (Customer X).
10. Product: Jenis kendaraan yang dibeli (Yaris AT, Etios AT, Avanza AT).
11. Region: Wilayah transaksi (MLG, SBY, KDR).
12. Date: Tanggal transaksi (MM/DD/YYYY).
13. Item unit price: Harga satuan produk (desimal).
14. No.Items: Jumlah unit yang dibeli.
15. Total Sale: Hasil perkalian harga satuan dan jumlah unit.
16. Adakah data yang memiliki nilai null / data yang tidak lengkap?
    * Ya, terdapat data yang tidak lengkap. Contohnya:

* Pada baris yang hanya memiliki nilai angka tanpa informasi pelanggan, produk, atau wilayah (baris ke-14 di file Excel).
* Ada beberapa baris di mana Product atau Region kosong.

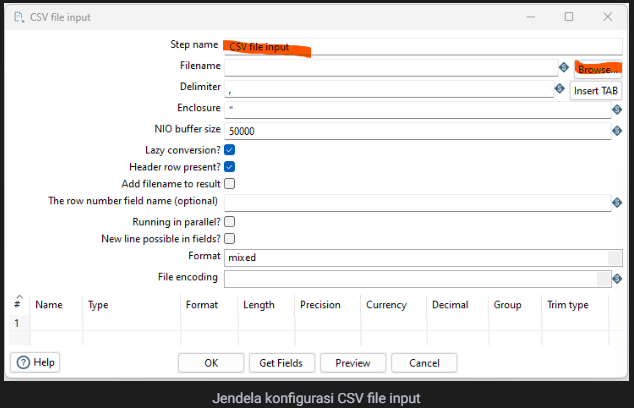
1. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

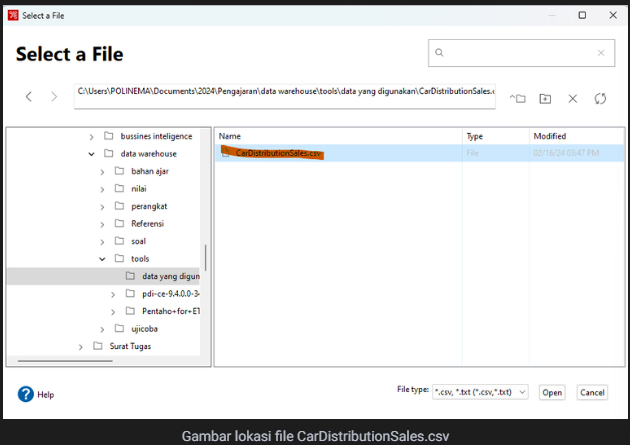
* Ya, terdapat perbedaan tipe data:
* Pada kolom Region, ada sel kosong yang seharusnya berisi teks.
* Pada kolom Product, beberapa entri kosong yang seharusnya berisi nama kendaraan.
* Pada kolom Customer, ada entri yang tidak sesuai format standar "Customer X".

1. **Pengambilan Data (Extract)**
2. Buka lembar kerja Transformation melalui toolbar File.
3. Cari lah objek CSV file input pada Design Area. Drag and drop objek tersebut menuju Work Area.

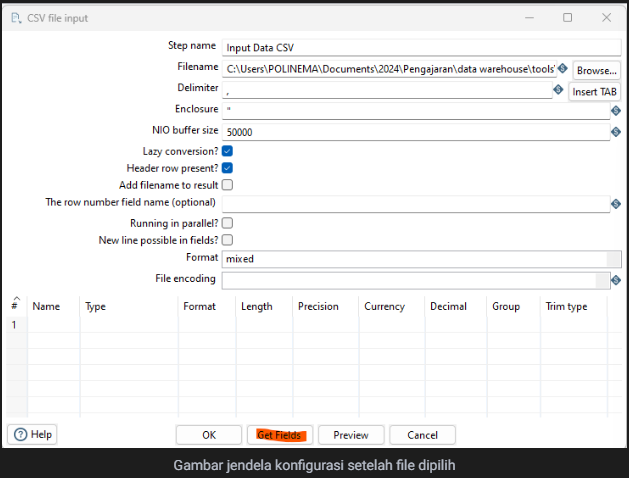


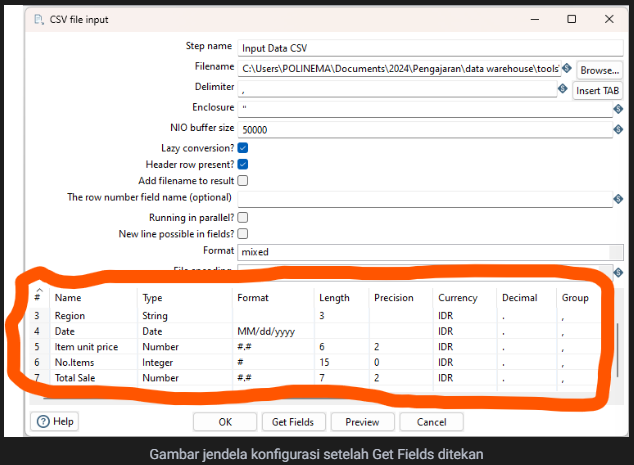
1. Double-click pada objek CSV file input hingga muncul jendela konfigurasinya.
2. Ubah nama Step name menjadi Input Data CSV, (penamaan digunakan untuk identifikasi saja, selanjutnya penamaan boleh menyesuaikan).
3. Pilih lokasi file CarDistributionSales.csv melalui Browse pada filename.

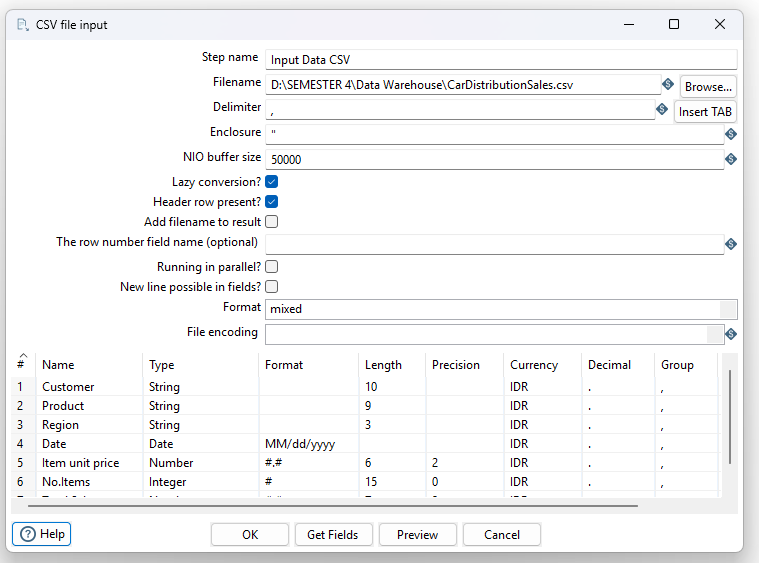




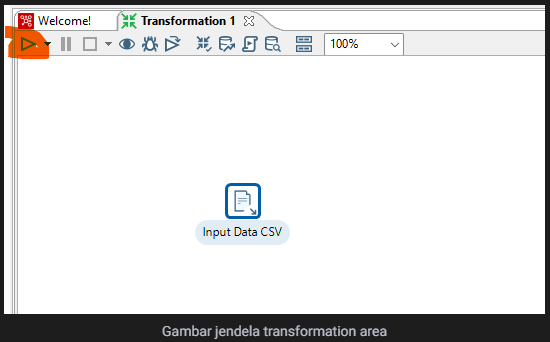
1. Jika data telah dipilih pastikan pada bagian Delimiter sesuai dengan kondisi pemisah data pada file yang dipilih. (Pada kesempatan ini adalah comma " , " ).
2. Tekan Get Fields untuk mengambil judul pada setiap kolom, pastikan nama-nama kolom telah sesuai

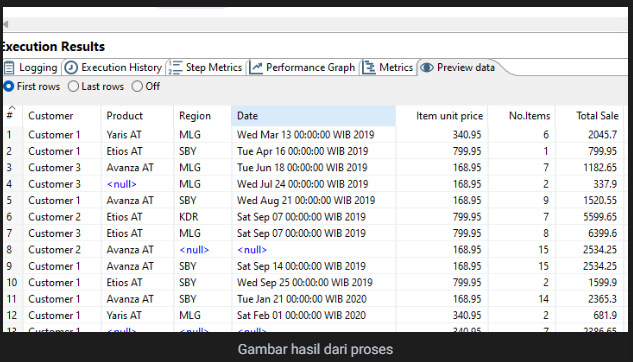




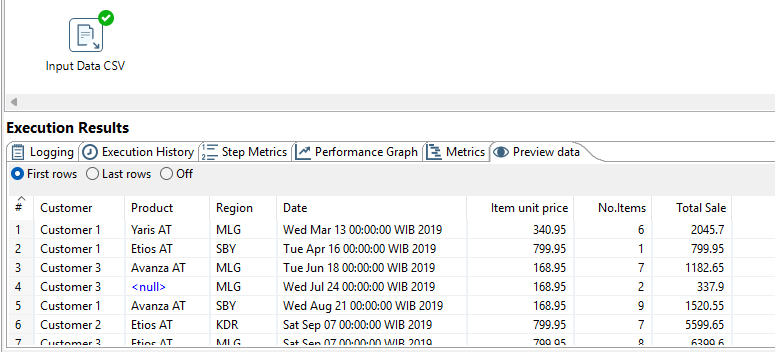


1. Tutup jendela konfigurasi tersebut.
2. Tekan tombol "Run" pada pojok kanan atas Area Kerja, makan akan muncul Execution Result Area. Pilih tab Prieview Data pada Execution Result Area. Jika proses yang dilakukan benar maka data akan muncul sesuai dengan aslinya.





My answer :

Output

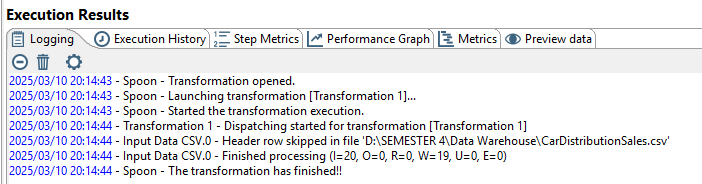
Jika langkah-langkah diatas telah berhasil, maka proses extract telah berhasil dilakukan.

TUGAS 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

Ya, sesuai dengan data aslinya

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. **Membuka Transformasi yang sudah dikonfigurasi.**

* **2025/03/10 20:14:43 - Spoon - Transformation opened.**

1. **Menjalankan Transformasi**

* **2025/03/10 20:14:43 - Spoon - Launching transformation [Transformation 1]...**
* **2025/03/10 20:14:43 - Spoon - Started the transformation execution.**

1. **Memulai Proses Extract Data**

* **2025/03/10 20:14:44 - Transformation 1 - Dispatching started for transformation [Transformation 1]**

1. **Membaca Data dari File CSV**

* **2025/03/10 20:14:44 - Input Data CSV.0 - Header row skipped in file 'D:\SEMESTER 4\Data Warehouse\CarDistributionSales.csv'**
* **Spoon membaca file CarDistributionSales.csv**, **tetapi melewatkan** **baris header**.

1. **Menyelesaikan Proses Extract Data**

* **2025/03/10 20:14:44 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)**
* **Spoon selesai memproses file CSV dengan:**
* **I = 20 (baris yang dibaca)**
* **W = 19 (baris yang berhasil ditulis)**
* **E = 0 (tidak ada error)**

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

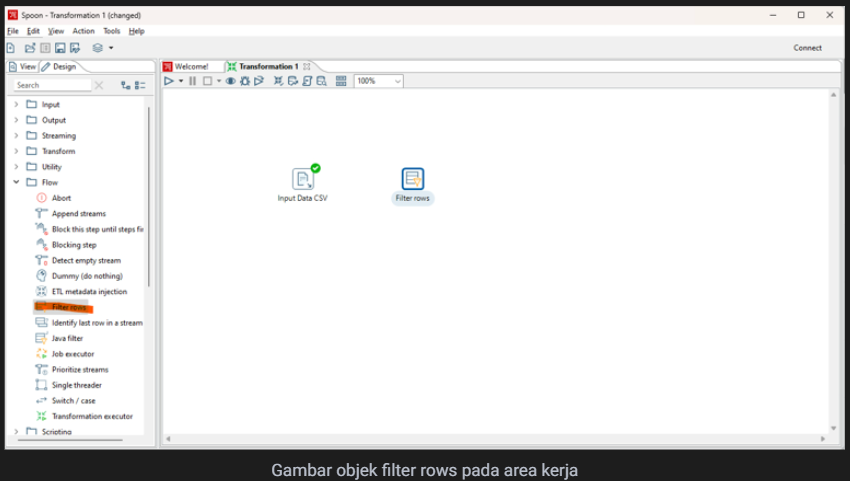


* **I (Input)** = Jumlah baris data yang dibaca. (**I=20**, berarti 20 baris dibaca).
* **O (Output)** = Jumlah baris yang dikirim ke langkah berikutnya. (**O=0**, tidak ada data dikirim).
* R (Rejected) = Jumlah baris yang ditolak. (R=0, tidak ada data ditolak).
* W (Written) = Jumlah baris yang berhasil ditulis ke output. (W=19, berarti 19 baris ditulis).
* U (Updated) = Jumlah baris yang diperbarui. (U=0, tidak ada data diperbarui).
* E (Error) = Jumlah baris yang gagal diproses. (E=0, tidak ada error).
* Dari 20 baris yang dibaca, 19 berhasil ditulis tanpa error.

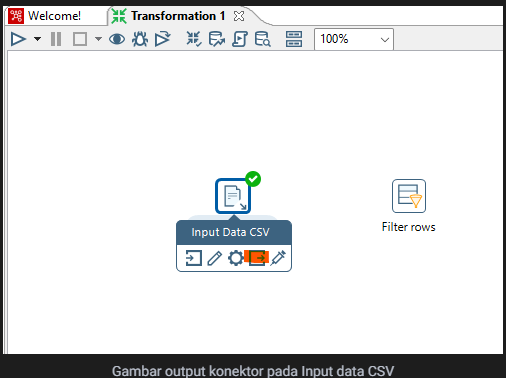
1. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract data ini? Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU ). Bandingkan dalam bentuk tabel.

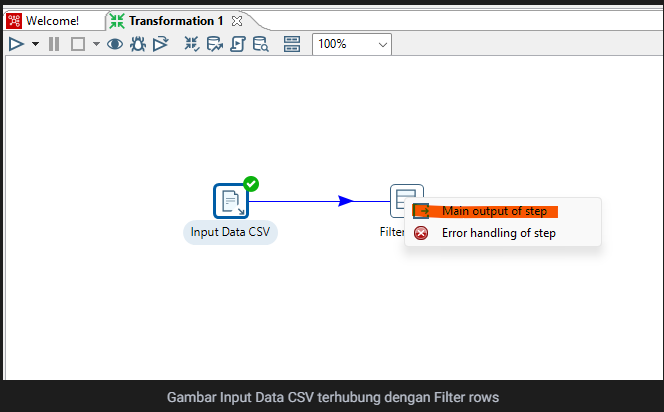
* Based on my log, the execution took 1 second to execute
* Based on my log, Start time: 2025/03/10 20:14:43 → "Started the transformation execution"
* End time: 2025/03/10 20:14:44 → "The transformation has finished!!"
* Execution time difference: 1 second (20:14:44 - 20:14:43) to execute

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | Execution Time | Processor | RAM | Storage | VGA | CPU Usage |
| (HP Laptop 14-dq2078wm) | 1s | Intel Core i5-1115G4 (2 Cores, 4 Threads) | 8GB DDR4 | 256GB SSD | Intel UHD Graphics | 12% |
| MSI GF63 Thin 11SC | 1s | Intel® Core™ i7-13650HX (13th Gen, 20 Cores) | 16GB | SSD 477GB | Intel® UHD Graphics + NVIDIA RTX 4050 | 7% |
| Acer Aspire 5 | 2s | AMD Ryzen 5 5500U | 8GB | SSD 256GB | AMD Radeon Graphics | 9% |
| ASUS VivoBook 14 | 1s | Intel® Core™ i5-1135G7 | 12GB | SSD 512GB | Intel® Iris Xe Graphics | 8% |

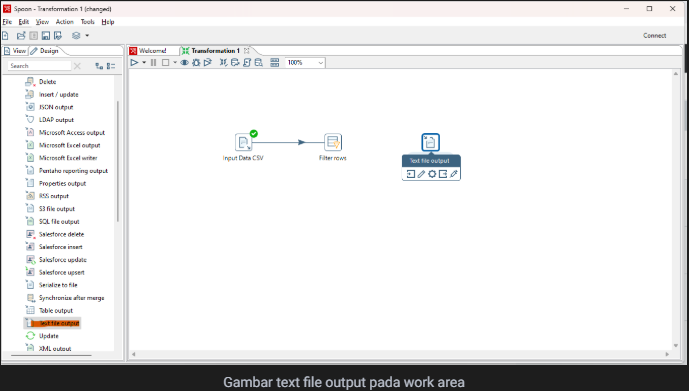
1. **Filter Data (Transform) dan Pengemasan data (Load)**
2. Carilah objek Filter rows pada Design Area. Drag and drop objek tersebut pada Work Area.
3. Hubungkan output pada Input Data CSV menuju Filter rows. Pilih koneksi Main output of step saat menghubungkan kedua objek tersebut.

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.

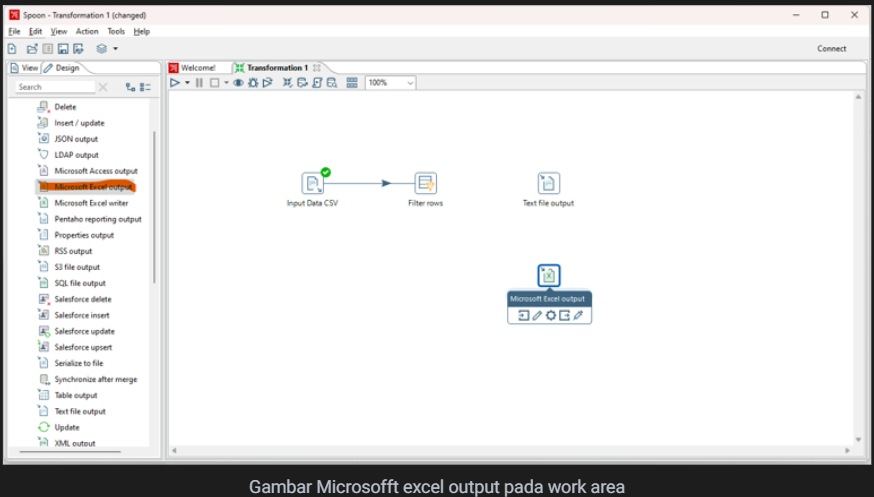




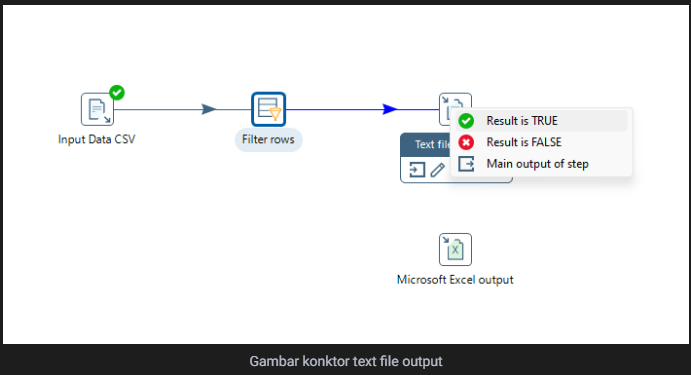
1. Carilah objek Text file output pada Design Area. Drag and drop objek tersebut pada Work Area.



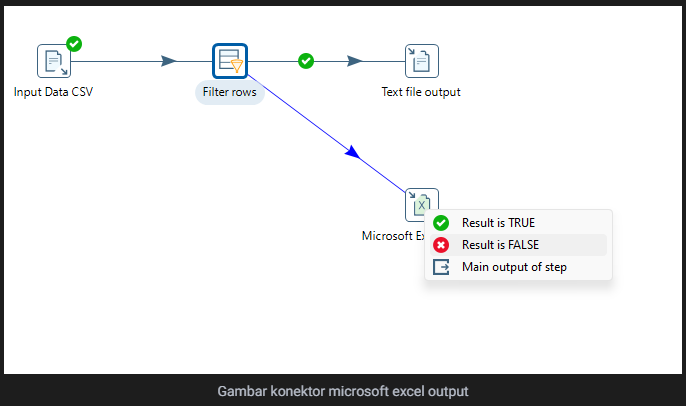
1. Carilah objek Microsoft excel output pada Design Area. Drag and drop objek tersebut pada Work Area.



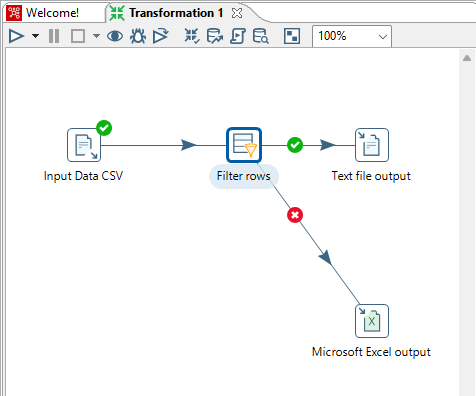
1. Hubungkan filter rows dengan text file output menggunakan konektor output dari filter rows. Pilih Result is true sebagi jenis konektornya.



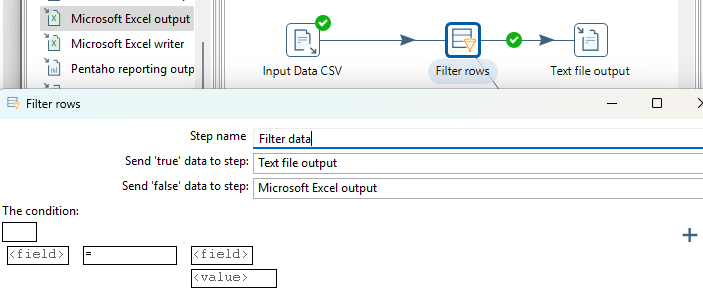
1. Hubungkan filter rows dengan microsoft excel output menggunakan konektor output dari filter rows. Pilih Result is false sebagi jenis konektornya.



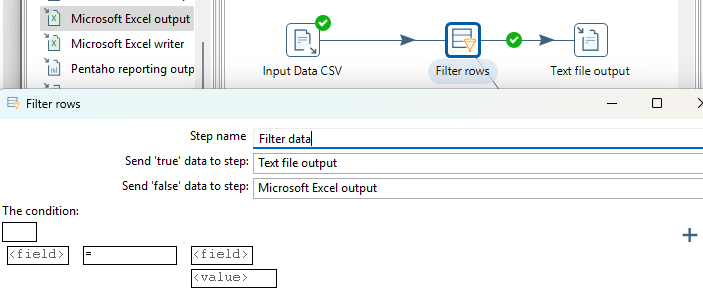
Result :



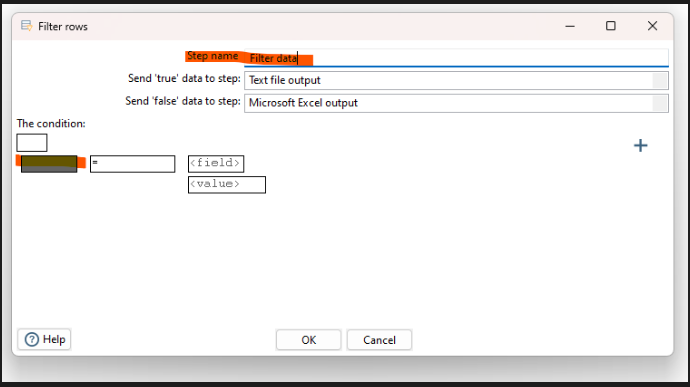
1. Double click pada filter rows hingga muncul jendela konfigurasi. ubah step name menjadi Filter data.



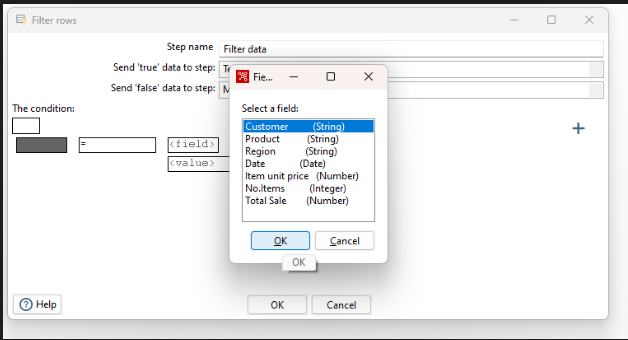
1. Cek bagian send 'true' data to step menjadi Text file output dan cek bagian Send 'false' data to step menjadi Microsoft Excel output.

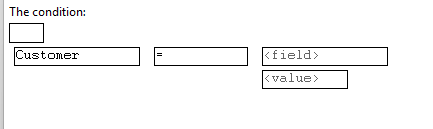


1. Klik <Field> pada bagian condition

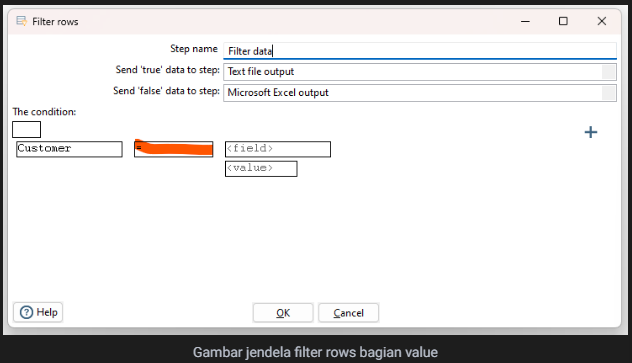


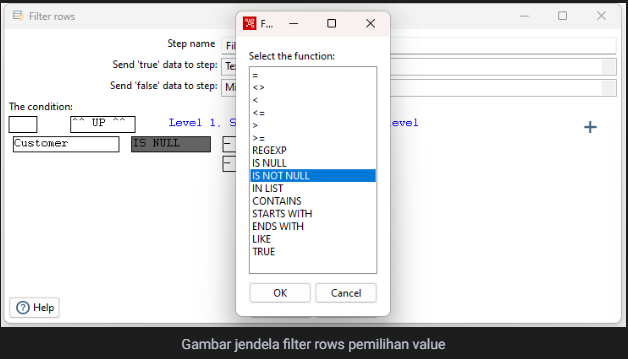
1. Pilih field yang diinginkan untuk digunakan sebagai kondisi. Pada kesempatan ini pilih Customer, kemudian klik OK.

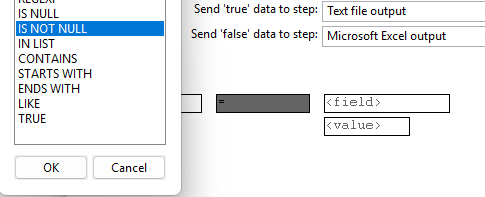




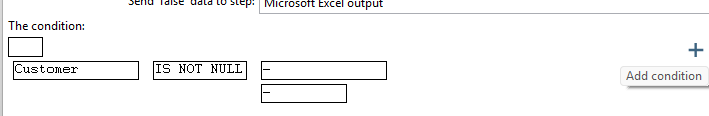
1. Pada bagian value pilih IS NOT NULL kemudian tekan OK , hal ini dilakukan karena data yang dipilih adalah data yang tidak Null



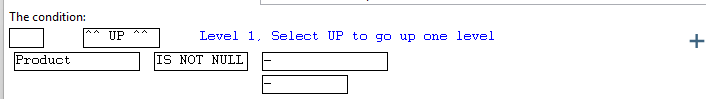




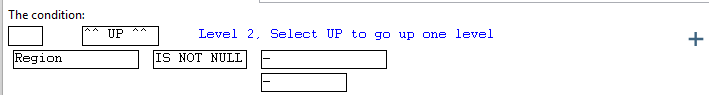
1. Tekan tombol + untuk menambah kondisi.



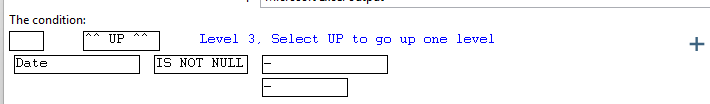
1. Pilih statement yang masih null untuk menambah kondisi.
2. Ulangi proses nomor 9,10,11 dengan mengganti semua field yang ada. hingga statement menjadi seperti pada gambar dibawah ini.
3. Product



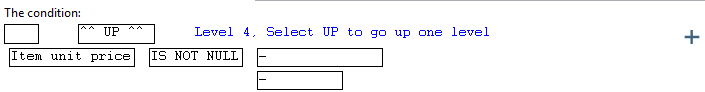
1. Region



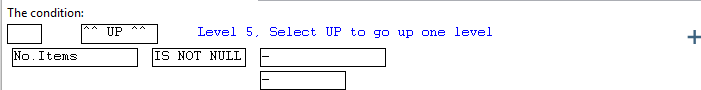
1. Date



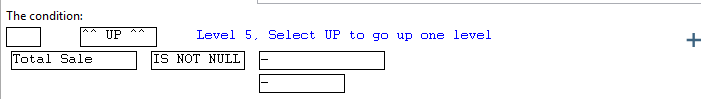
1. Item unit price



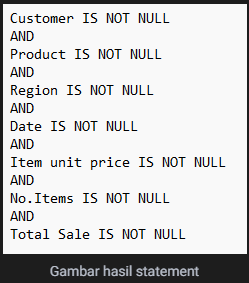
1. No.items

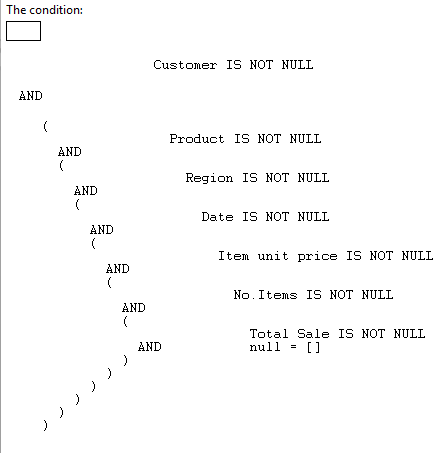


1. Total sale

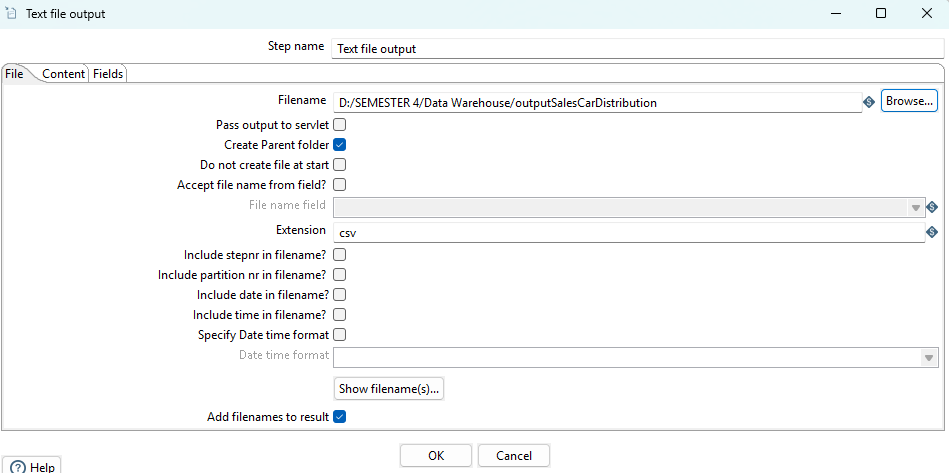


1. Gunakan operator AND untuk menghubung

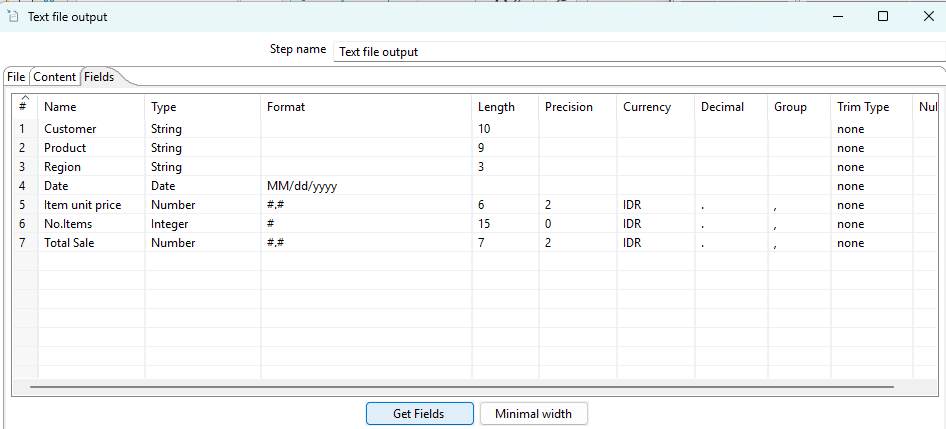




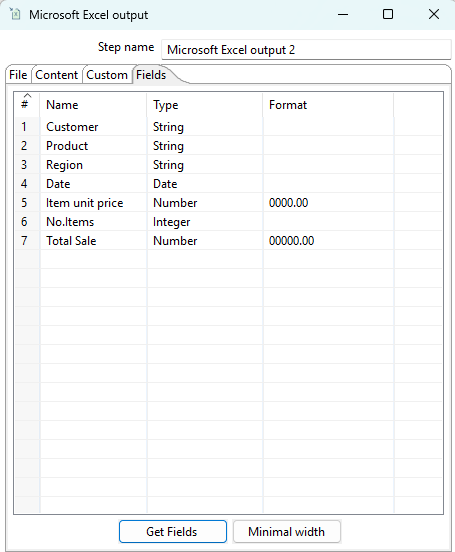
1. Jika semua field sudah masuk dalam kondisi maka tekan OK.
2. Double-click pada text file output, pilih lokasi file untuk menyimpan hasil output file dengan menekan tombol browse. beri nama outputSalesCarDistribution.
3. Ubah extension menjadi csv.



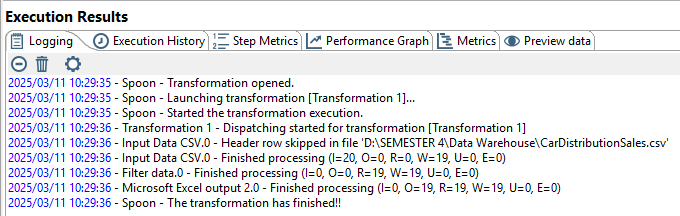
1. Pada tab Fields, tekan tombol Get fields untuk mengambil field data.

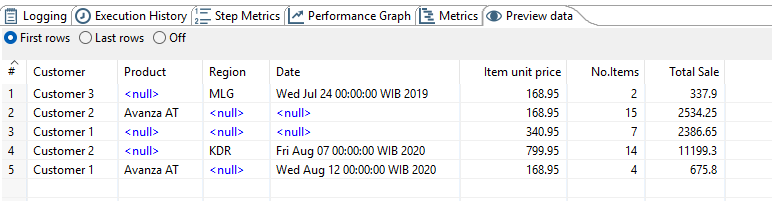


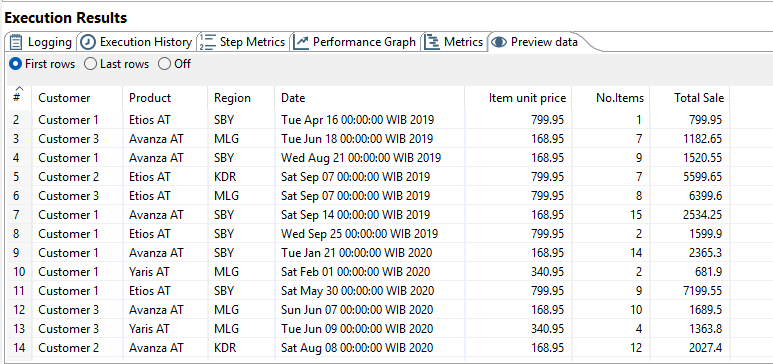
1. Tekan ok untuk kembali ke work area
2. Hal yang sama dilakukan pada konfigurasi Microsoft Excel output, Double-click pada objek Microsoft Excel output hingga muncul jendela konfigurasi.
3. Pilih lokasi file untuk menyimpan file output dengan menekan tombol browse. Beri nama failedSalesCarDistribution. untuk extension tetap dengan extension xls.
4. Pada tab fields tekan tombol Get fields untuk mengambil field data.



1. Tekoan ok untuk kembali ke work area.
2. Tekan tombul Run pada pojok kiri atas Work Area. jika berhasil maka akan muncul tanda centang hijau di setiap proses.







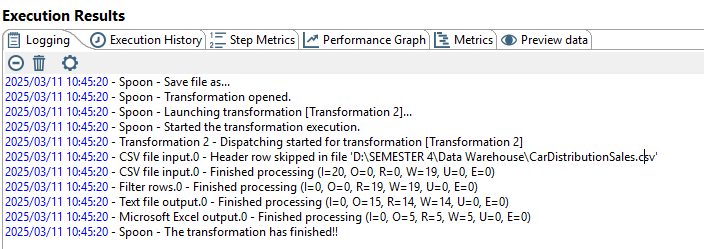
Langkah-langkah diatas merupakan proses melakukan filter data, dimana data yang lengkap akan diletakkan pada text file csv dan data yang belum lengkap akan diletakkan pada ms excel.

TUGAS 3

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

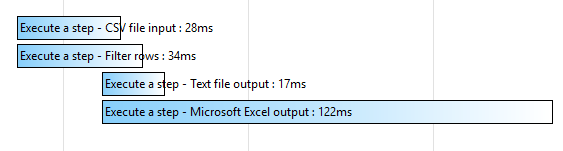
Perbedaanya terjadi pada pemisahan antara null dan tidak null

1. Jelaskan apa yang terjadi pada proses Filter rows!



* Pada proses Filter rows! dalam transformasi ETL menggunakan Pentaho Data Integration (PDI) (Spoon), terjadi penyaringan data berdasarkan kondisi yang ditentukan, di mana Filter rows.0 selesai diproses dengan menerima 19 baris data (R=19) dari langkah sebelumnya, menuliskan seluruhnya ke langkah berikutnya (W=19), tanpa input langsung (I=0), tanpa output langsung setelah langkah ini (O=0), tanpa pembaruan data (U=0), dan tanpa error (E=0), sehingga tidak ada data yang dibuang atau ditolak oleh filter ini.

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



* Total Waktu Eksekusi
* Total waktu yang dibutuhkan untuk proses ETL:
* 28ms + 34ms + 17ms + 122ms = 201ms (0.201 detik)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Nama | Waktu Eksekusi | Processor | RAM | Storage | VGA | CPU Usage |
| (HP Laptop 14-dq2078wm) | 201ms | Intel® Core™ i5-13650HX (13th Gen, 20 Cores) | 16GB | SSD 256GB | Intel® UHD Graphics + NVIDIA RTX 4050 | 12% |
| Laptop 1 (ASUS ROG Strix G15) | 250ms | Intel® Core™ i7-11800H (8 Cores, 16 Threads) | 16GB | SSD 512GB | NVIDIA GeForce RTX 3070 | 10% |
| Laptop 2 (Acer Predator Helios 300) | 250ms | Intel® Core™ i7-11800H (8 Cores, 16 Threads) | 8 GB | SSD 512GB | NVIDIA GeForce RTX 3070 | 10% |
| Laptop 3 (Lenovo Legion 5 Pro) | 230ms | AMD Ryzen 9 5900HX (8 Cores, 16 Threads) | 32GB | SSD 1TB | NVIDIA GeForce RTX 3080 | 8% |

**Studi Kasus**

**Disclaimer: Data yang tercantum di studi kasus ini adalah fiktif dan hanya digunakan untuk tujuan pembelajaran saja. Nama tempat, tanggal lahir, alamat, dan informasi lainnya yang tertera hanya tipuan belaka dan tidak memiliki kaitan dengan individu yang sebenarnya.**

**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. darai 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 tersbut akan digunakan untuk analisa di masing-masing wilayah dan juga Dashboard untuk emmbantu mengambil keputusan.**

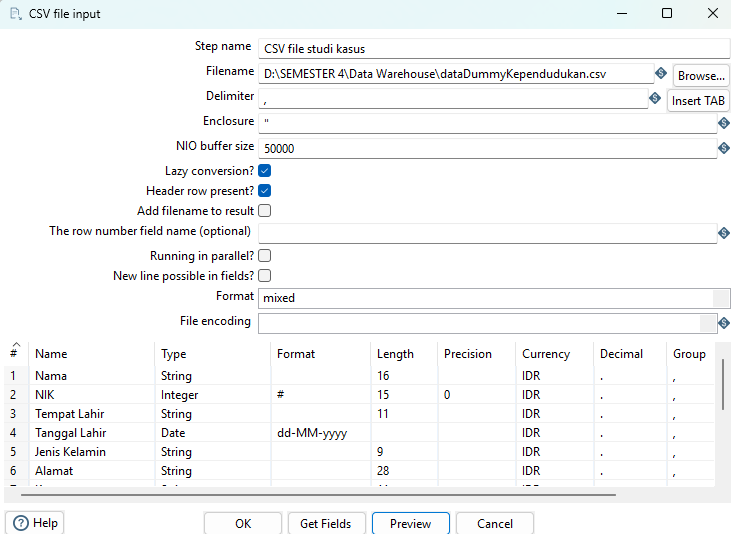
**Pak Pascanowo mendapatakan data penduduk yang akan pindah ke wilayahnya melalui link:**

**https://github.com/dik4rizky/datasources/blob/82821b8ec341176f979956c84b79720d4aead012/dataDummyKependudukan.csv**

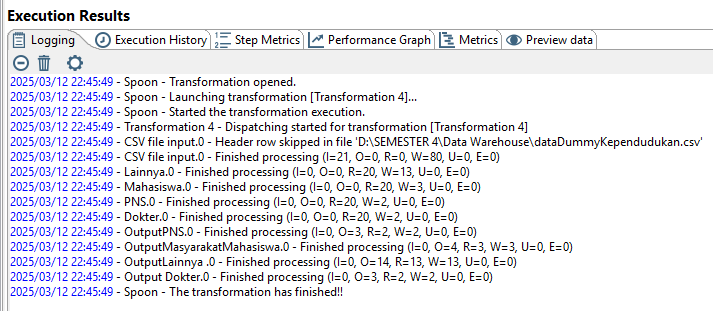
**Bantulah Pak Pascanowo untuk memishkan data tersebut menjadi 4 file:**

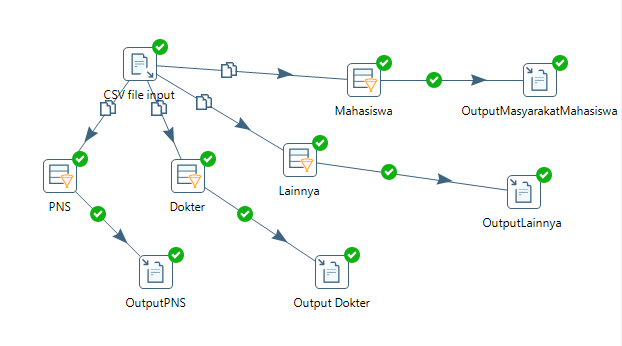
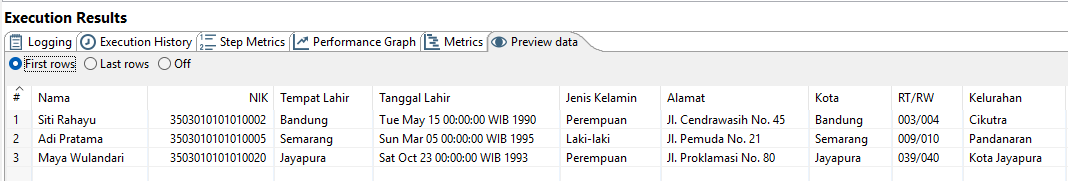
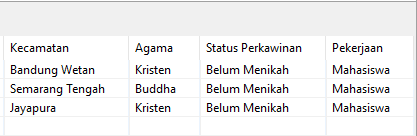
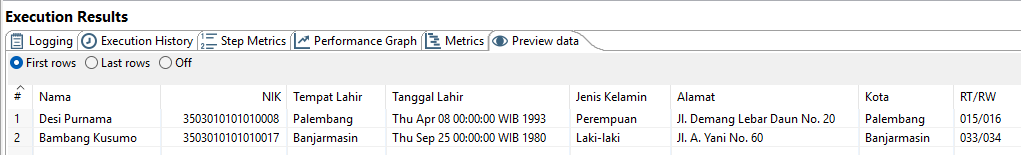
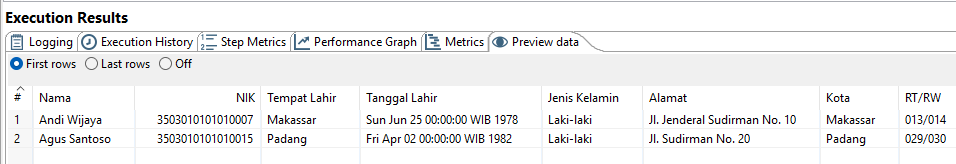
**MasyarakatMahasiswa, MasyarakatPNS, MasyarakatDokter, MasyarakatLainnya.**

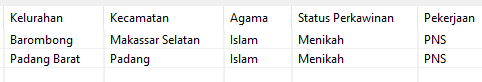
**Dikarenakan proses tersebut akan repetisi setiap ada masyarakat Konoha yang berpindah ke IKN maka buatlah sistem tersebut menggunakan PDI Spoon.**

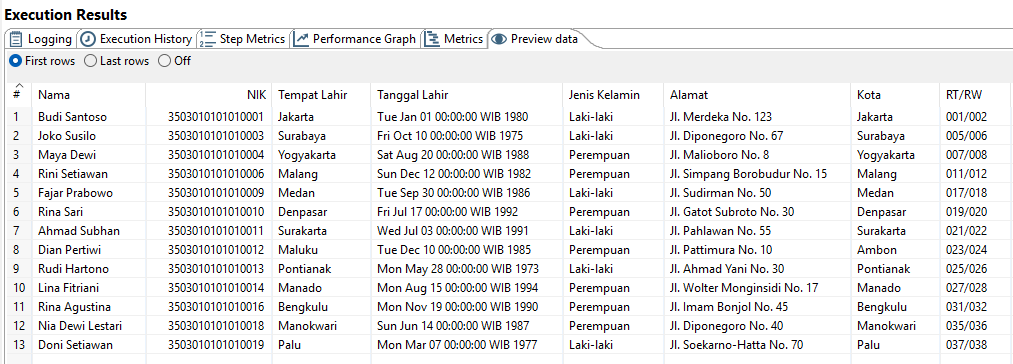
****

1. proses Filter rows

****

1. Hubungkan filter rows dengan microsoft ****
2. **Output MasyarakatMahasiswa**
3. **Output Dokter**
4. **Output PNS**

****

1. **Output masyarakat lainnya**

****

**Selamat explore 😄**