Nama : Dini Elminingtyas

Kelas/No.absen : SIB 2B / 07

NIM : 2341760180

**DATA WAREHOUSE**

**JOBSHEET 1**

**DATA WAREHOOUSE SEDERAHAN DANPENGENALAN PDI SPOON**

TUGAS 1

Analisa data dari data penjualan Astro Mobil

1. Berapa jumlah kolom pada data tersebut?

* Pada data tersebut terdapat **7** kolom

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

* Data tersebut merupakan data hasil penjualan Sales Astro Mobil dari beberapa cabang. Pada data tersebut terdapat beberapa kolom yang berbeda diantaranya ;

1. Customer : Kolom ini berisikan data dari tiap customer yang melakukan pembelian
2. Product : Kolom ini berisikan jenis product yang dibeli oleh customer yang bersangkutan
3. Region : Kolom ini menunjukkan dari daerah mana / cabang mana transaksi / penjualan tersebut dilakukan
4. Date : Kolom ini menunjukkan tanggal ketika transaksi / penjualan dilakukan
5. Item unit price : Kolom ini menujukkan harga satuan dari product yang dibeli oleh customer
6. No.items : Kolom ini menunjukkan nomor items yang terjual / dibeli oleh customer
7. Total sale : Kolom ini berisikan total harga pembelian dari transaksi / penjualan tersebut
8. Adakah data yang memiliki nilai null / data tidak lengkap?

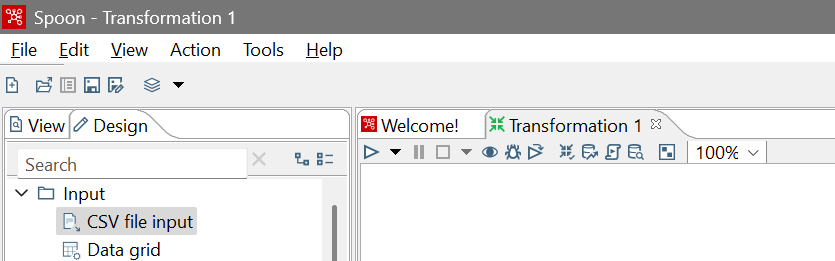
* Ada, beberapa data memiliki nilai null atau data tidak lengkap diantaranya;

1. Baris no 5, tidak ada data pada kolom product
2. Baris no 9, tidak ada data pada kolom region dan juga date
3. Baris no 14, tidak ada data pada kolom
4. Baris no 18, tidak ada data pada kolom product
5. Baris no 10, tidak ada data pada kolom Region
6. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

* Pada data tersebut, tidak ada tipe data yang berbeda dengan data lainnya, tetapi ada beberapa data yang kosong

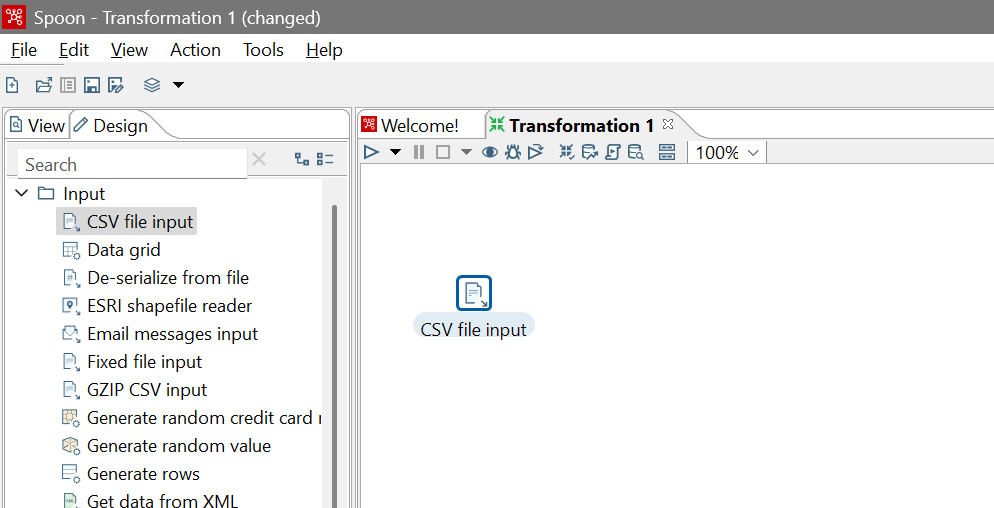
1. Pengambilan Data (Extract)
2. Buka lembar kerja Transformation melalui toolbar file

* Transformation



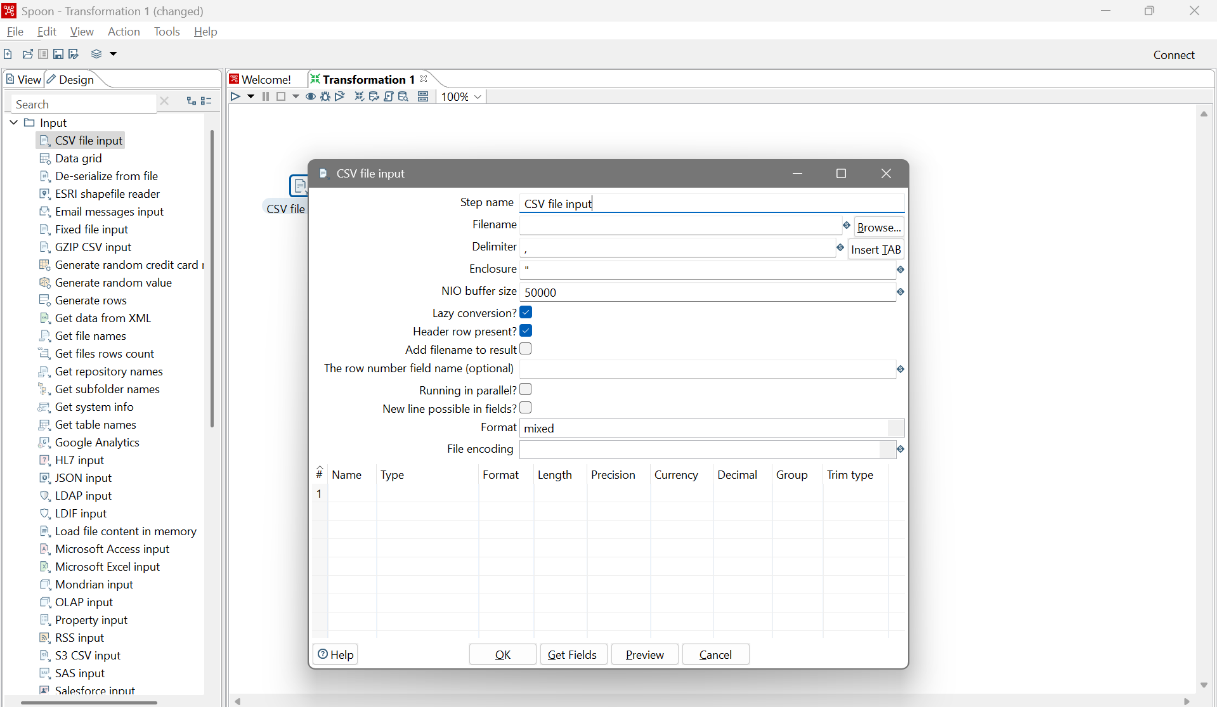
1. Cari lah objek **CSV file input** pada **Design Area**. Drag and drop objek tersebut menuju **Work Area.**

* Csv file input



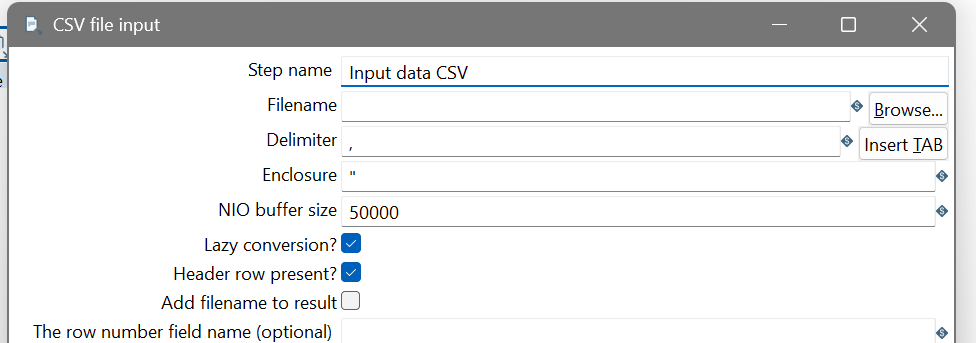
1. Double-click pada objek **CSV file input** hingga muncul jendela konfigurasinya.

* Jendela konfigurasi



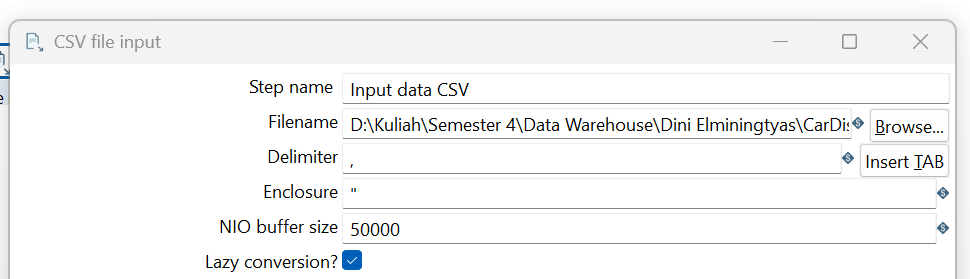
1. Ubah nama **Step name** menjadi **Input Data CSV**, (penamaan digunakan untuk identifikasi saja, selanjutnya penamaan boleh menyesuaikan).

* Ubah step name



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

* CarDistributionSales.csv

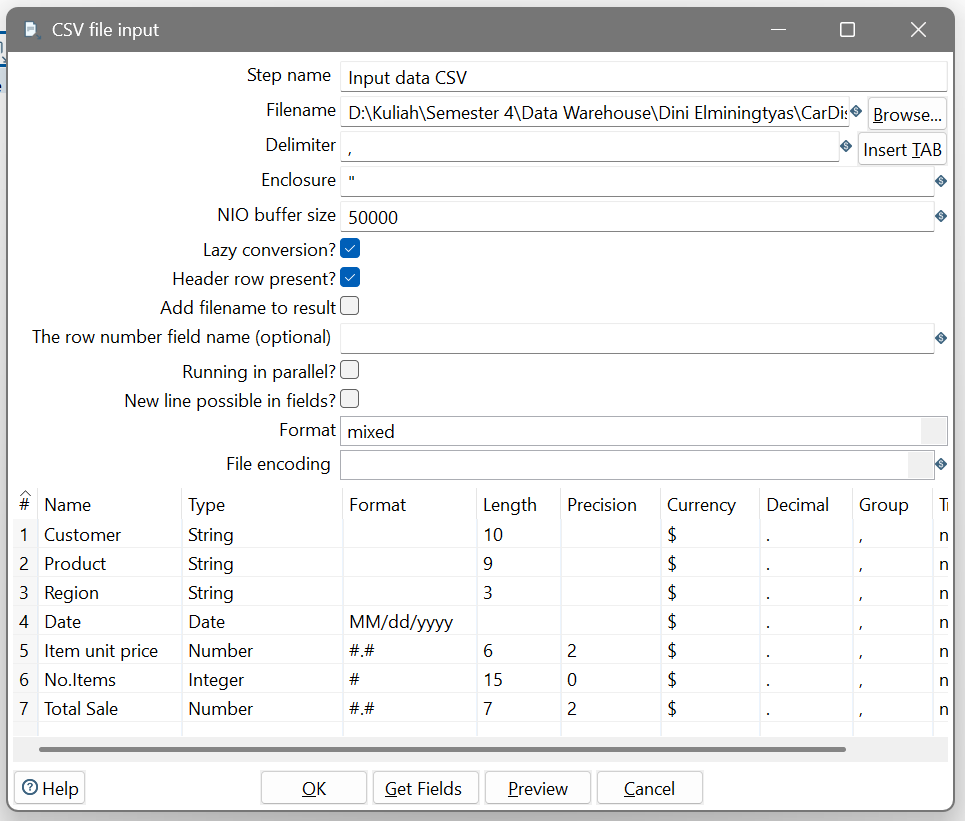


1. Jika data telah dipilih pastikan pada bagian **Delimiter** sesuai dengan kondisi pemisah data pada file yang dipilih. (Pada kesempatan ini adalah comma " , " ).

* Sudah sesuai

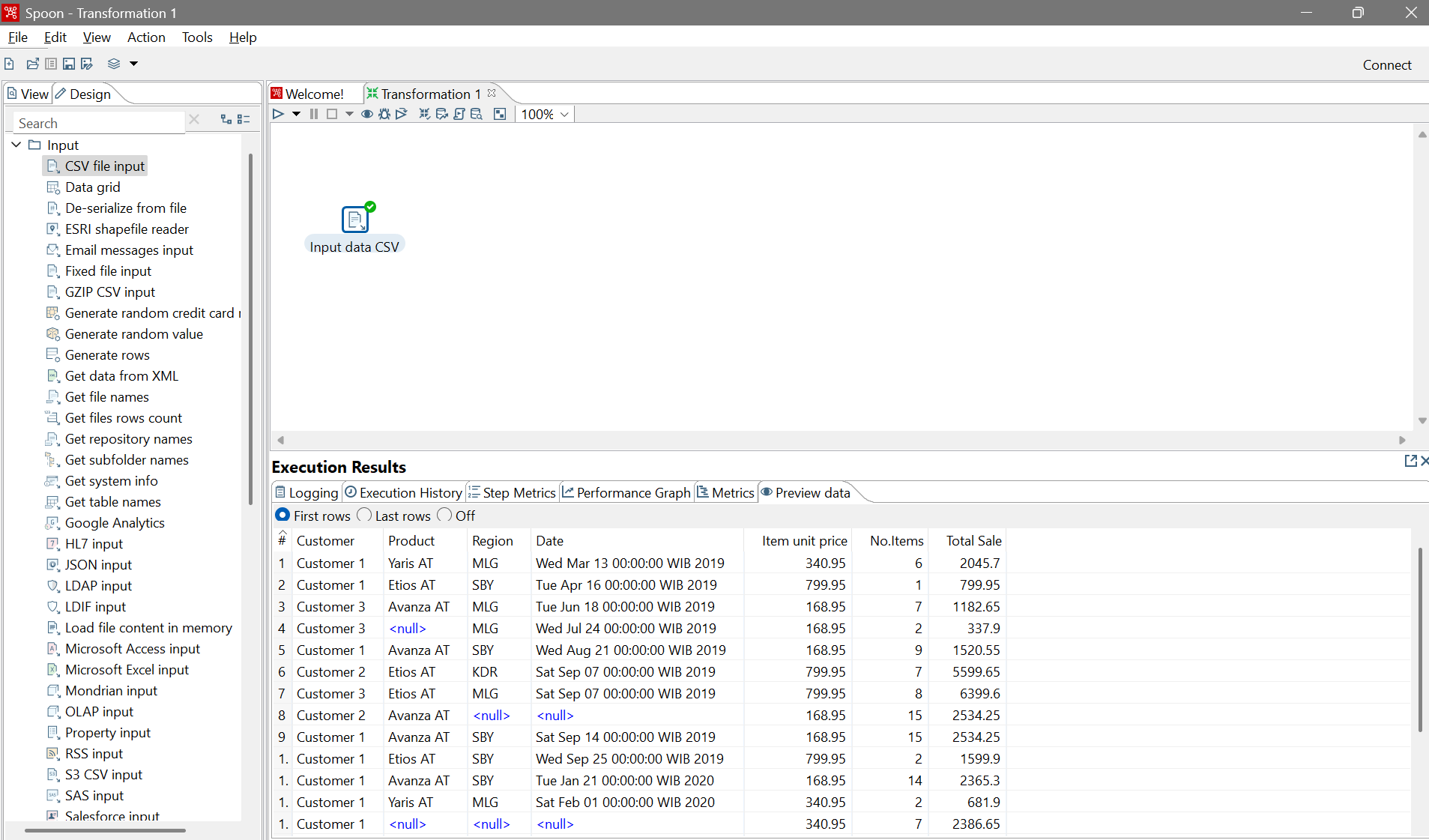
1. Tekan **Get Fields** untuk mengambil judul pada setiap kolom, pastikan nama-nama kolom telah sesuai

* Get Fields



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.

* Hasil run



Sudah sesuai dengan data aslinya

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

TUGAS 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

* Ya, hasil eksekusi sudah 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?

* Langkah – langkah yang dilakukan untuk extract file data tersebut diantaranya adalah ;
  1. Save file, atau menyimpan file terlebih dahulu
  2. Kemudian membuka file Transformation
  3. Launching file Transformation
  4. Memulai eksekusi file Transformation
  5. Mengirimkan data untuk file Transformation
  6. Proses input data (header skipped)
  7. Proses input data (finishing)

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



* Huruf huruf tersebut merepresentasikan ;
* I (Input) => Jumlah data yang diinput
* O (Output) => Jumlah data yang berhasil dikeluarkan
* R (Rejected) => Jumlah data yang tertolak atau gagal diproses
* W (Written) => Jumlah data yang berhasil tertulis
* U (Updated) => Jumlah data yang diperbarui
* E (Errors) => Jumlah kesalahan yang terjadi pada saat proses input data

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.

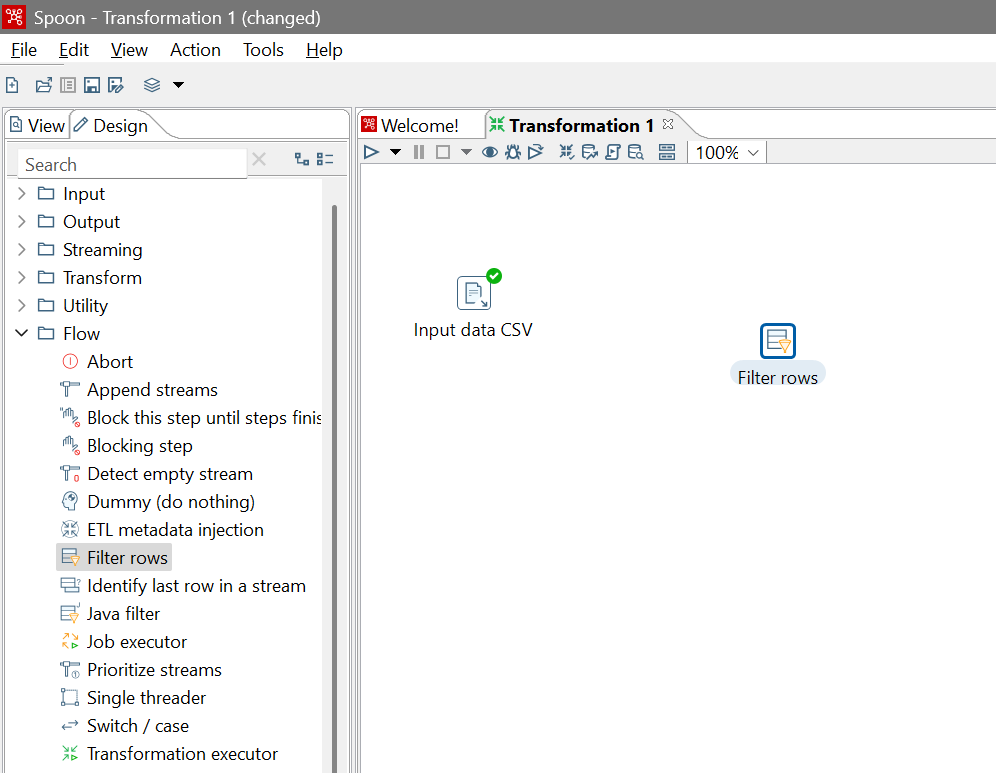
* Perbandingan dengan 3 teman saya :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | Processor | RAM | Storage | VGA | CPU | Time |
| Dini(saya) | AMD Ryzen 7 5800HS with Radeon Graphics 3.20 GHz | 16 GB | 475GB | AMD Radeon Graphics | AMD Ryzen 7 5800HS with Radeon Graphics | 1’ms |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

\*teman saya belum buat

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

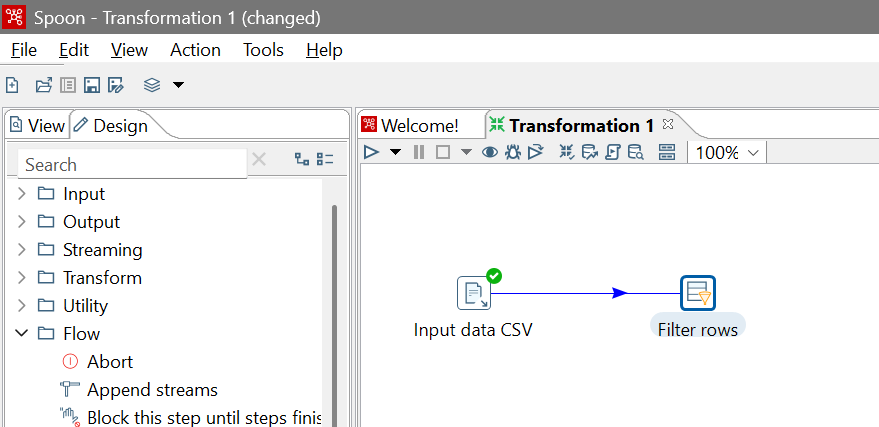
* Filter Rows



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

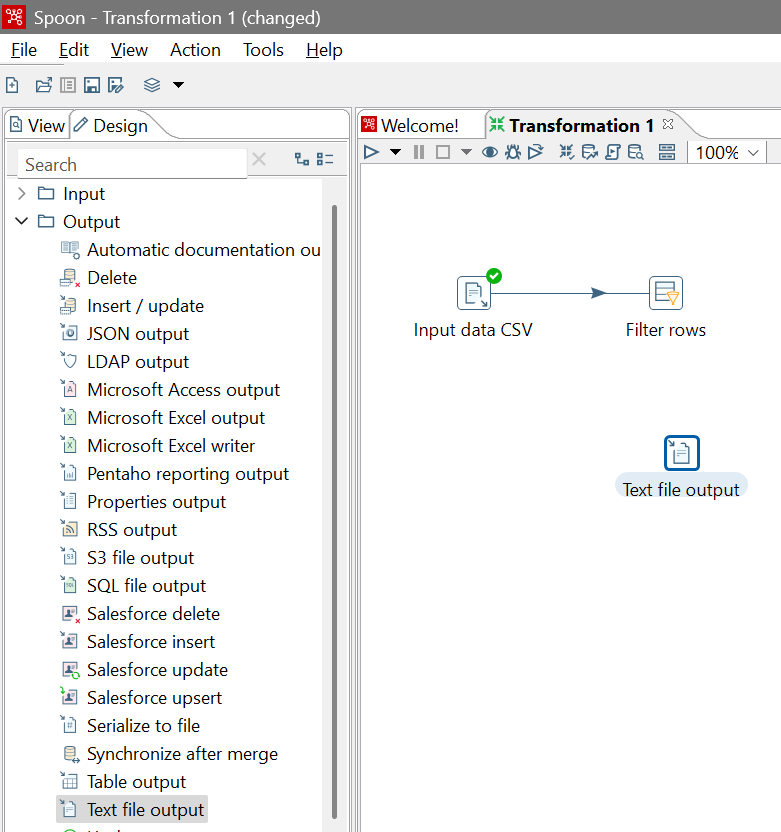
Pada tahap ini Input Data CSV dihubungkan oleh konektor pada Filter rows memiliki makna bahwa setelah proses Input Data CSV dilakukan maka proses selanjutnya adalah Filter rows.

* Menghubungkan filter rows dengan input data csv



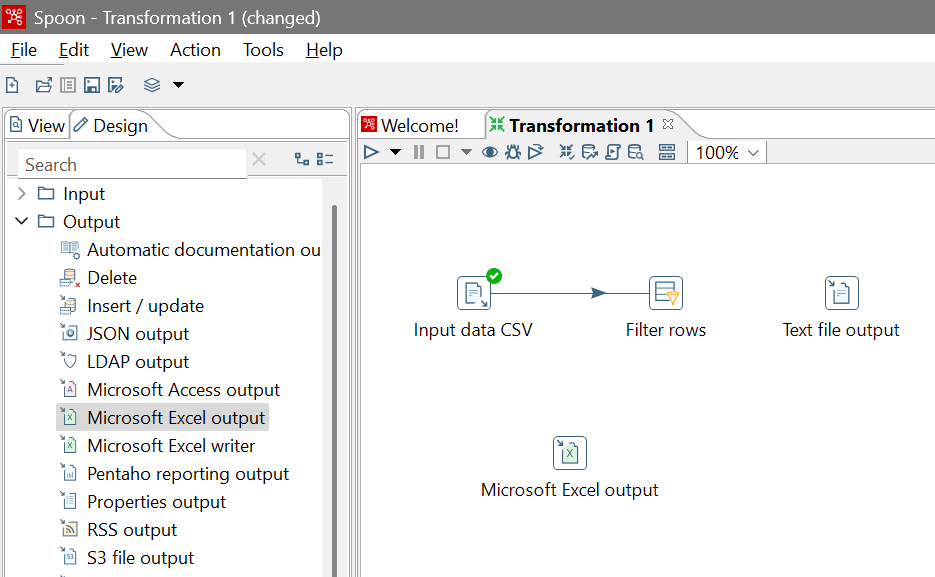
1. Carilah objek **Text file output** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.

* Text file output

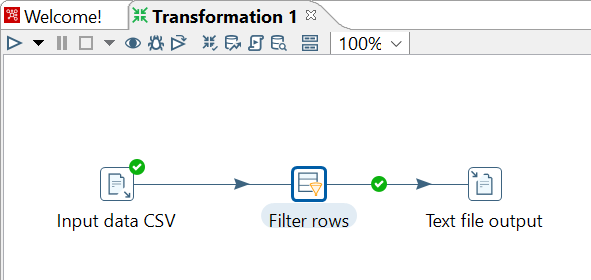


1. Carilah objek **Microsoft excel output** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.

* Microsoft Excel output

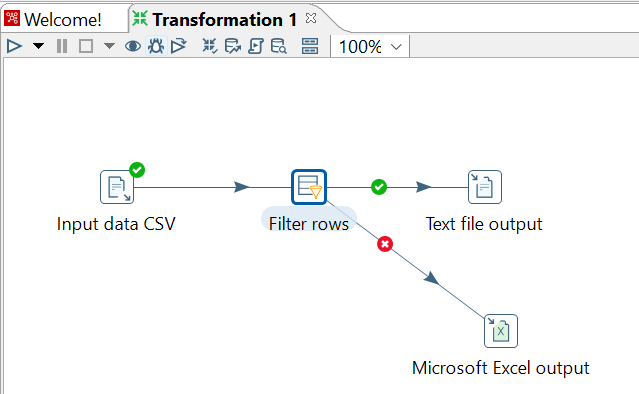


1. Hubungkan **filter rows** dengan **text file output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is true** sebagi jenis konektornya.

* Conected filter row, text file output

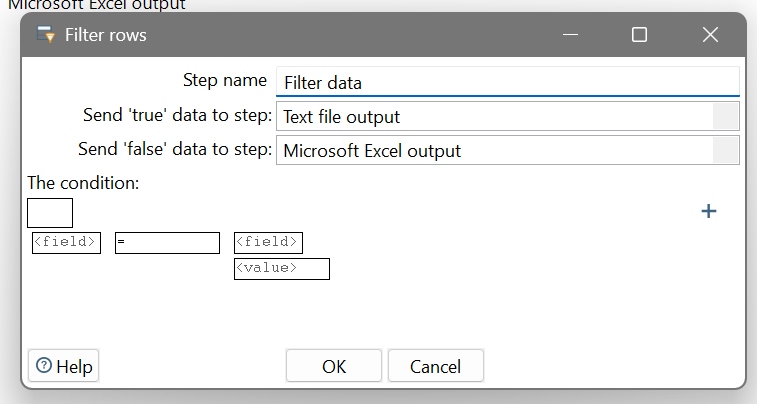
1. Hubungkan **filter rows** dengan **microsoft excel output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is false** sebagi jenis konektornya.

* Connected filter rows, dan Microsoft excel output

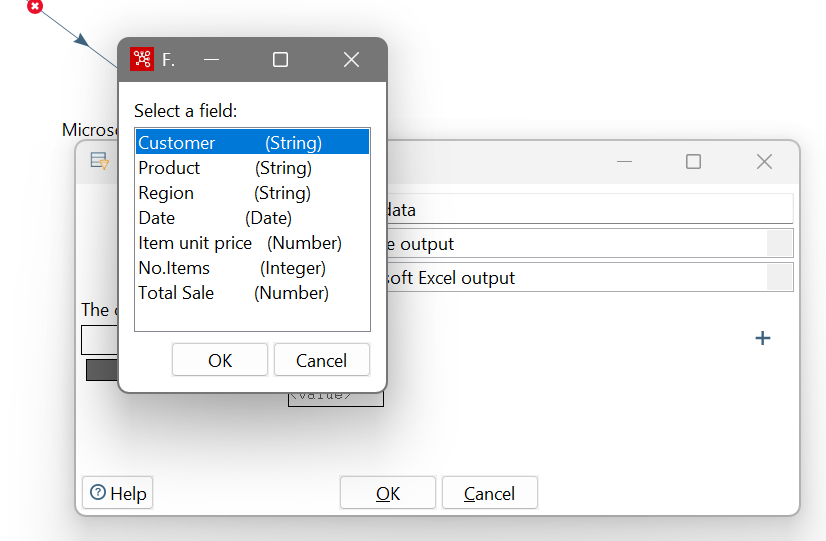


1. **Double click** pada **filter rows** hingga muncul jendela konfigurasi**.** ubah **step name** menjadi **Filter data**.

* Filter Data

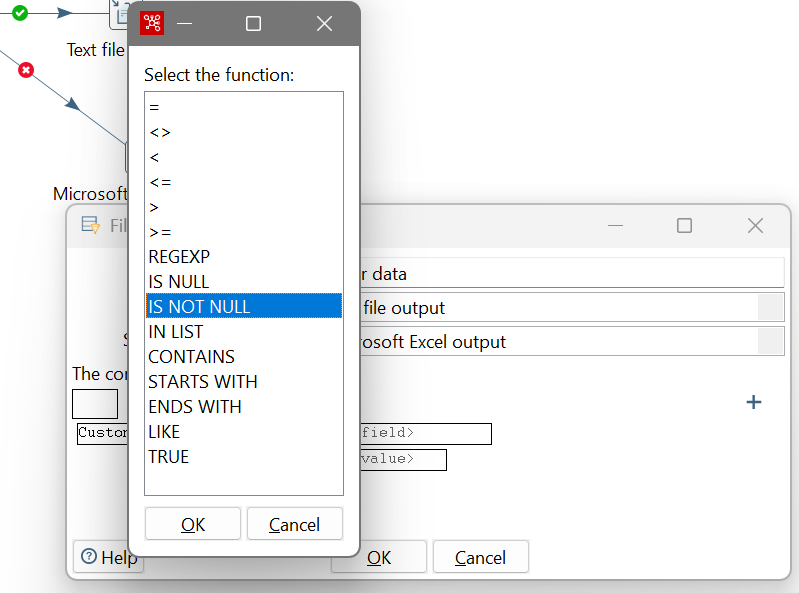


1. Cek bagian **send 'true' data to step** menjadi **Text file output** dan cek bagian **Send 'false' data to step** menjadi **Microsoft Excel output**.
2. Klik **<Field>** pada bagian **condition**
3. Pilih field yang diinginkan untuk digunakan sebagai kondisi. Pada kesempatan ini pilih **Customer,** kemudian klik **OK.**

* Field

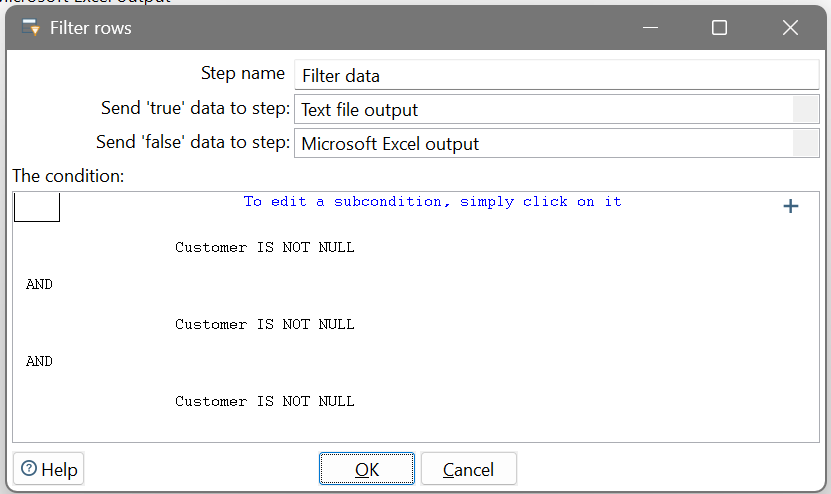
1. Pada bagian **value** pilih **IS NOT NULL** kemudian tekan **OK** , hal ini dilakukan karena data yang dipilih adalah data yang tidak Null

* Is not null



1. Tekan tombol **+** untuk menambah kondisi.
2. Pilih statement yang masih null untuk menambah kondisi.

* Menambah kondisi



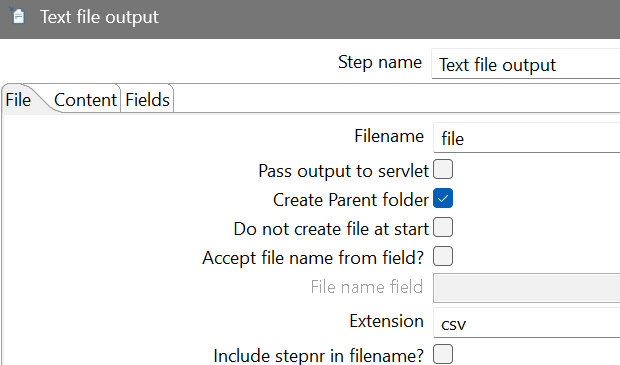
1. Ulangi proses nomor 9,10,11 dengan mengganti semua field yang ada. hingga statement menjadi seperti pada gambar dibawah ini.
2. Gunakan operator AND untuk menghubung

* Menambah condition



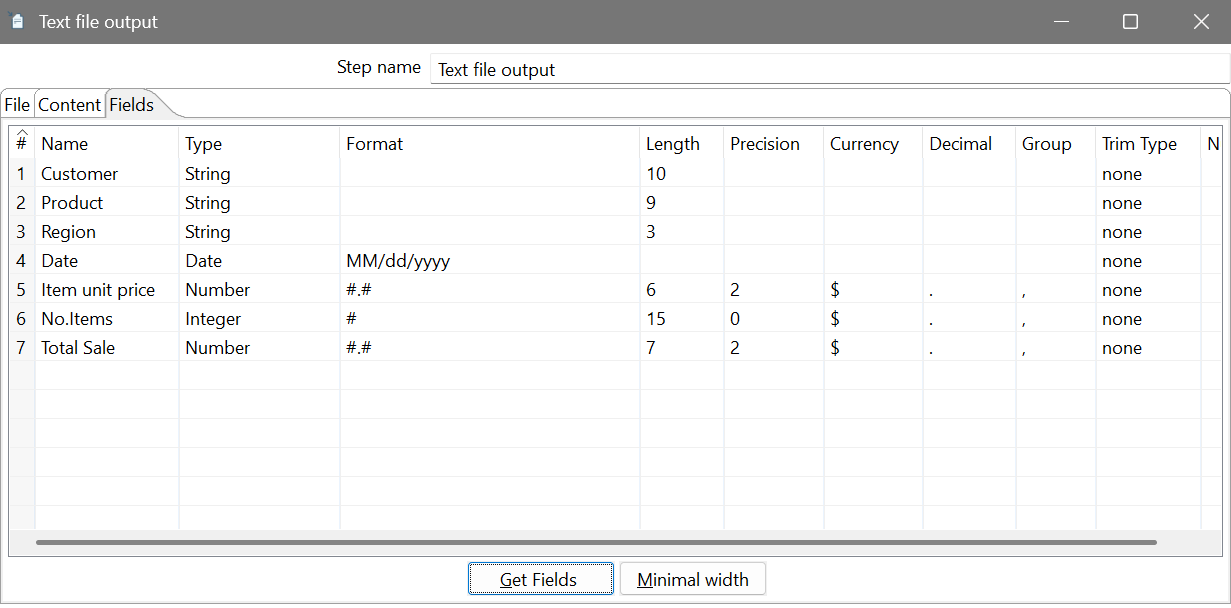
1. Jika semua field sudah masuk dalam kondisi maka tekan OK.
2. Double-click pada text file output, pilih lokasi file untuk menyimpan hasil output file dengan menekan tombol browse. beri nama outputSalesCarDistribution.
3. Ubah extension menjadi csv.

* Ubah extension



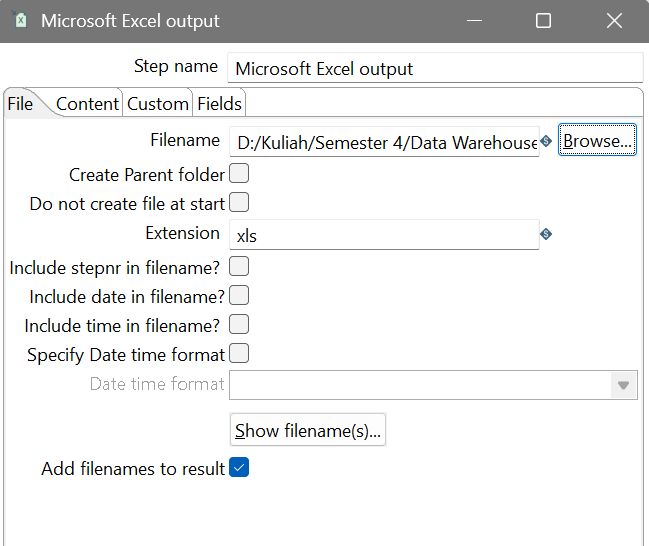
1. Pada tab Fields, tekan tombol Get fields untuk mengambil field data.

* Mengambil field data



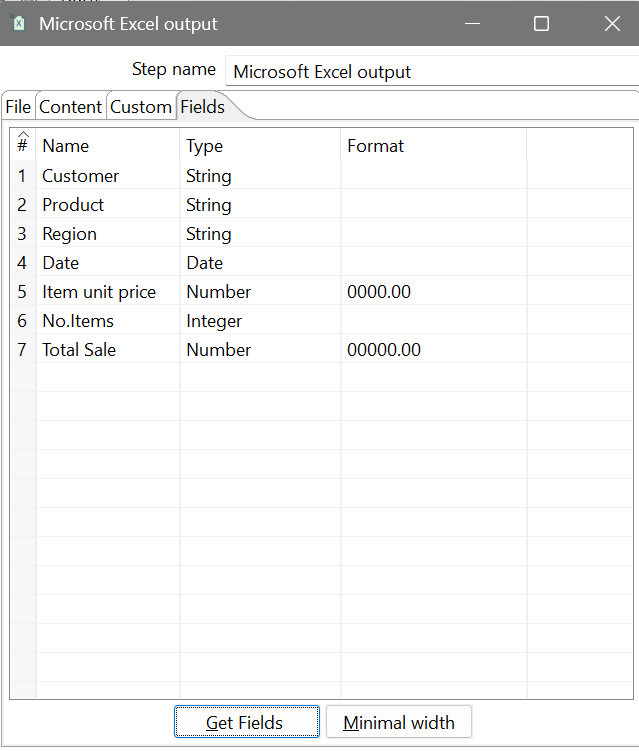
1. Tekan ok untuk kembali ke work area
2. Hal yang sama dilakukan pada konfigurasi Microsoft Excel output, Double-click pada objek Microsoft Excel output hingga muncul jendela konfigurasi.
3. Pilih lokasi file untuk menyimpan file output dengan menekan tombol browse. Beri nama failedSalesCarDistribution.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

* File location, dan extension



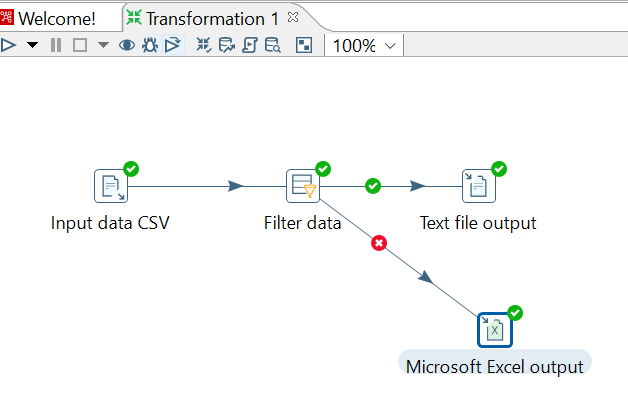
1. Pada tab fields tekan tombol Get fields untuk mengambil field data.

* Fields

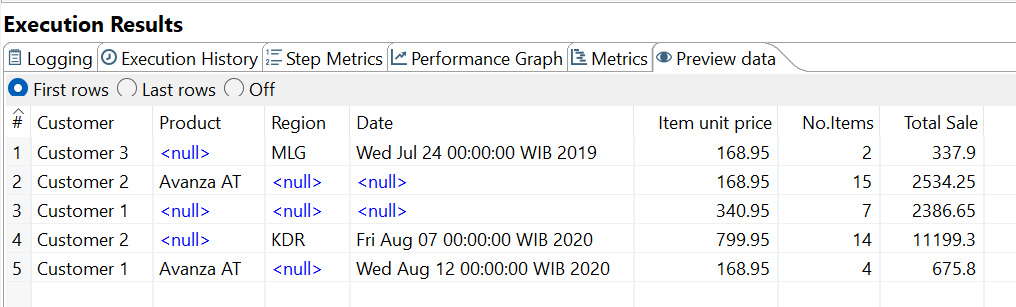


1. Tekan ok untuk kembali ke work area.
2. Tekan tombul **Run** pada pojok kiri atas **Work Area**. jika berhasil maka akan muncul tanda centang hijau di setiap proses.

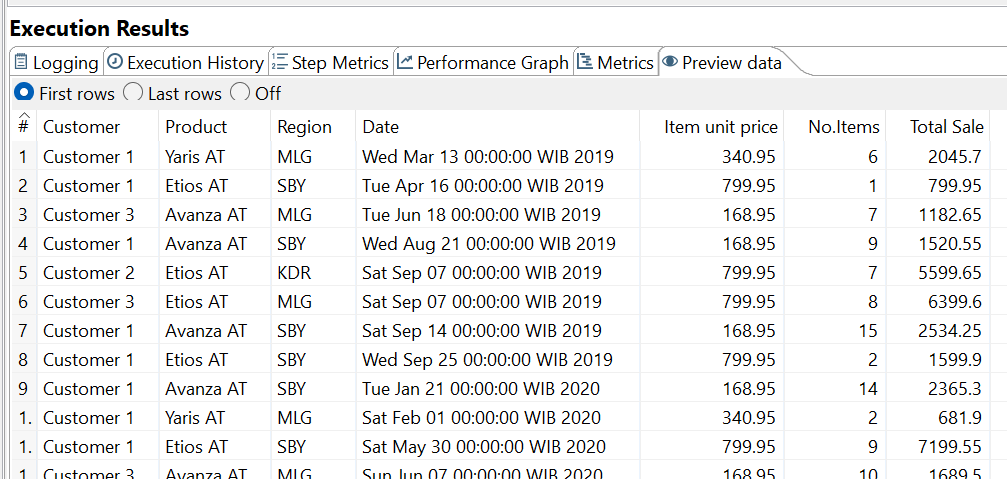
* Execution Results



* failedSalesCarDistributon



* outputSlesCarDistribution



TUGAS 3

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

* Pada file csv berisikan data lengkap tanpa ada data yang kosong, sedangkan pada file excel hanya menampilkan beberapa data, yang masih memiliki data null atau data yang kosong

1. Jelaskan apa yang terjadi pada proses Filter rows!

* Filter rows memisahkan data sebelumnya menjadi dua bagian, yaitu data yang lengkap disimpan pada file csv, sedangkan data yang masih kosong atau memiliki nilai null disimpan pada file excel.

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 tabel

* Perbandingan

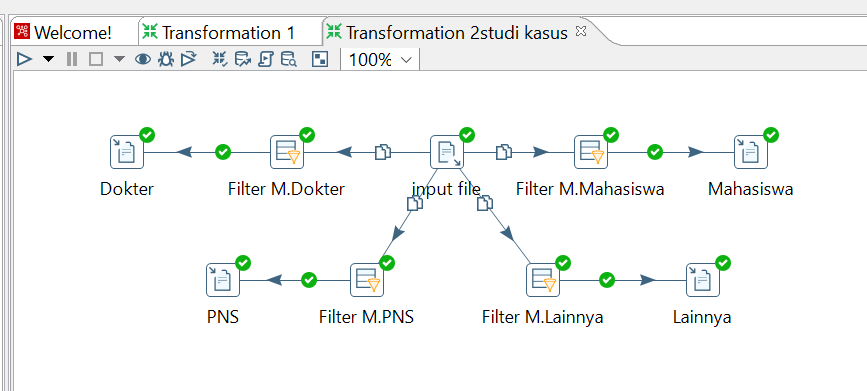
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | Processor | RAM | Storage | VGA | CPU | Time |
| Dini(saya) | AMD Ryzen 7 5800HS with Radeon Graphics 3.20 GHz | 16 GB | 475GB | AMD Radeon Graphics | AMD Ryzen 7 5800HS with Radeon Graphics | 1’ms |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

\*Teman saya belum :>

STUDI KASUS

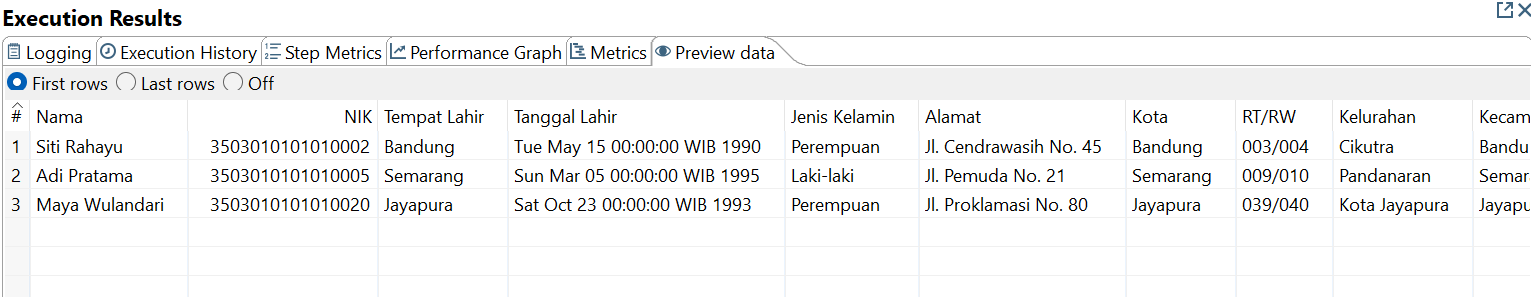
Pak Pascanowo seorang ketua RW di Ibu Kota Nusabangsa (IKN). Terdapat 20 orang dari negara Konoha yang akan pindah dan berdomisili di IKN. Tata letak pemukiman pada IKN telah dikondisikan agar berkelompok disesuaikan dengan lokasi pekerjaannya agar akses terjangkau.Sebagai contoh, mahasiswa akan dikumpulkan 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 pemerintahanData masyarakat berprofesi Dokter akan diberikan ke Pak Budi Pekerti selaku ketua RT wilayah kesehatandata-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.csvBantulah Pak Pascanowo untuk memishkan data tersebut menjadi 4 file:MasyarakatMahasiswa, MasyarakatPNS, MasyarakatDokter, MasyarakatLainnya.

PDI SISTEM ;

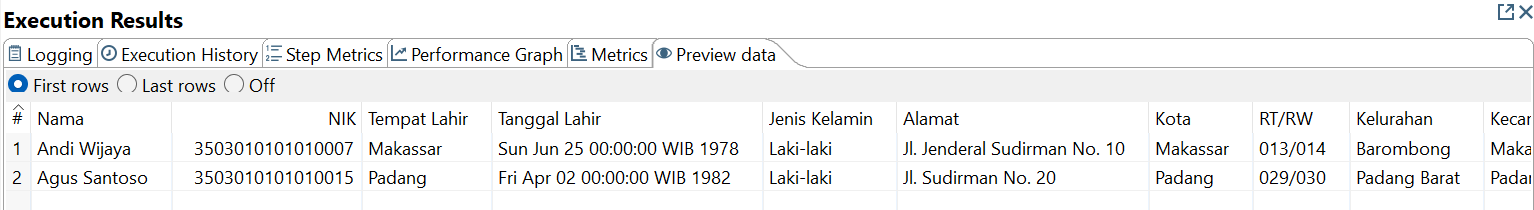


Execution Results

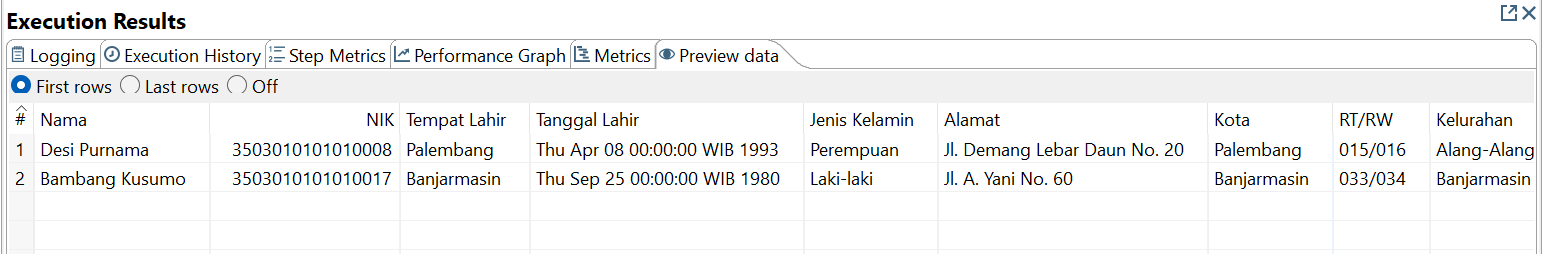
1. Masyarakat Mahasiswa



1. Masyarakat PNS



1. Masyarakat Dokter



1. Masyarakat Lainnya

