Laporan Praktikum Data Warehouse

Jobsheet 1: Data Warehouse Sederhana & Pengenalan PDI Spoon

Dosen Pengampu: Vipkas Al Hadid Firdaus, ST., MT



Disusun Oleh:

Queenadhynar Azarine Dwipa A.

2341760109

SIB 2B

JURUSAN TEKNOLOGI INFORMASI POLITEKNIK NEGERI MALANG 2023/2024

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:

- 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:

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

TUGAS 1

Analisa lah data tersebut!

- 1. Berapa jumlah kolom pada data tersebut?
 - ➤ Terdapat 7 kolom yang berisi (Customer, Product, Region, Date, Item unit price, No.Items, Total Sale
- 2. Apa arti atau isi data dari setiap kolom yang ada?
 - ➤ Customer > Nama Pelanggan

Product > Nama produk yang dibeli

Region > Wilayah tempat transaksi terjadi

Date > tanggal transaksi dalam format

Item unit price > Harga satuan dari produk yang dibeli

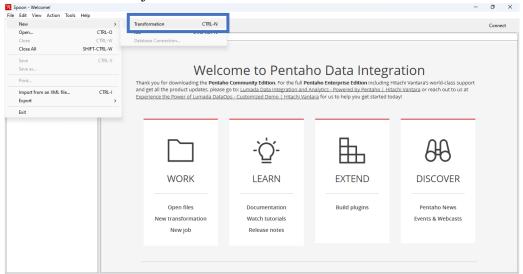
No. Items > Jumlah produk yang dibeli dalam satu transaksi

Total sale >Total penjualan dari transaksi tersebut

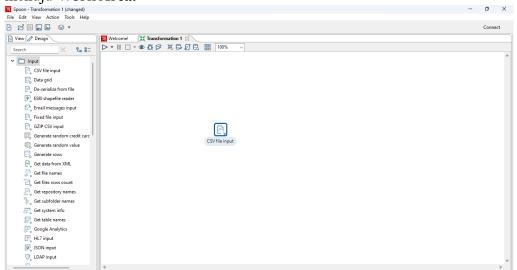
- 3. Adakah data yang memiliki nilai null / data yang tidak lengkap?
 - Pada baris 5 (A5) customer3 MLG, kolom tersebut tidak memiliki nama lengkap Pada baris 14 (A14) data tersebut tidak lengkap, hanya terdapat angka tanpa informasi lengkap mengenai produk, region atau tanggal Pada baris 17 (A17) nilai tidak lengkap pada kolom Region
- 4. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?
 - Ada beberapa yang tidak sesuai
 - 1. Kolom Customer: seharusnya berisi teks dengan format Customer X
 - 2. Kolom Region : Ada beberapa entri yang kosong
 - 3. Kolom Total Sale : Seharusnya berisi angka (hasil perhitungan harga × jumlah barang), tetapi bisa saja terdapat format yang tidak sesuai (misalnya, ada nilai yang kosong atau dipisahkan dengan koma yang bisa mengganggu pemrosesan data).

A. Pengambilan Data (Extract)

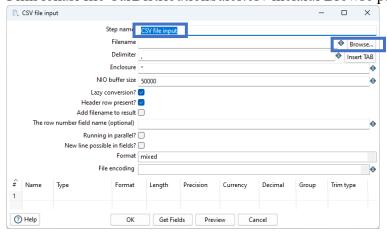
1. Membuka lembar kerja Transformation melalui toolbar File.

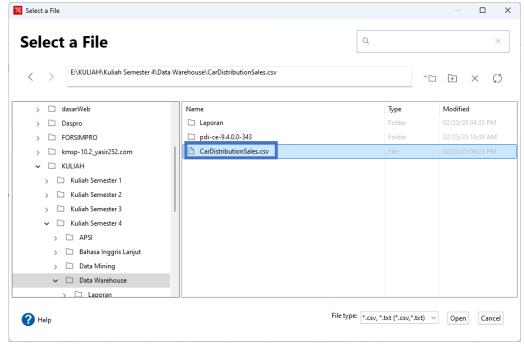


2. Cari lah objek **CSV file input** pada **Design Area**. Drag and drop objek tersebut menuju **Work Area**.

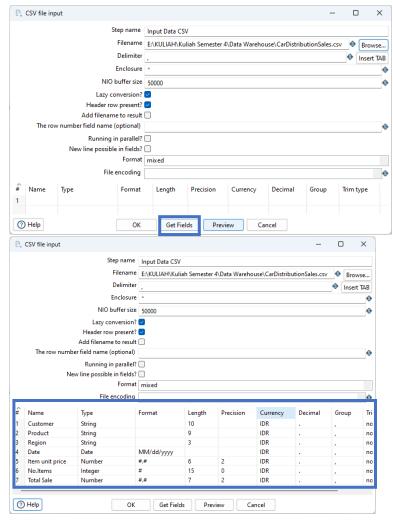


- 3. Double-click pada objek CSV file input hingga muncul jendela konfigurasinya.
- 4. Ubah nama Step name menjadi Input Data CSV, (penamaan digunakan untuk identifikasi saja, selanjutnya penamaan boleh menyesuaikan).
- 5. Pilih lokasi file CarDistributionSales.csv melalui Browse pada filename.

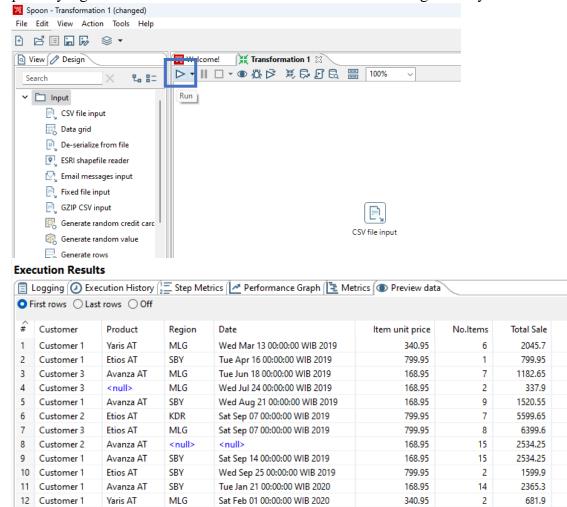




- 6. Jika data telah dipilih pastikan pada bagian Delimiter sesuai dengan kondisi pemisah data pada file yang dipilih. (Pada kesempatan ini adalah comma ", ").
- 7. Tekan Get Fields untuk mengambil judul pada setiap kolom, pastikan nama-nama kolom telah sesuai



- 8. Tutup jendela konfigurasi tersebut.
- 9. Tekan tombol "Run" pada pojok kanan atas Area Kerja, 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.



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

TUGAS 2

- 1. Apakah data hasil eksekusi sesuai dengan data aslinya?
 - > Iyaa, hasil sudah 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?

```
2025/02/25 18:48:55 - Spoon - Save file as...
2025/02/25 18:48:56 - Spoon - Transformation opened.
2025/02/25 18:48:56 - Spoon - Launching transformation [Transformation 1]...
2025/02/25 18:48:56 - Spoon - Started the transformation execution.
2025/02/25 18:48:57 - Transformation 1 - Dispatching started for transformation [Transformation 1]
2025/02/25 18:48:57 - Input Data CSV.0 - Header row skipped in file 'E:\KULIAH\Kuliah Semester 4\Data Warehouse\CarDistributionSales.csv'
2025/02/25 18:48:57 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)
2025/02/25 18:48:57 - Spoon - The transformation has finished!!
```

Berikut merupakan langkah – langkah

a. Membuka dan Menyimpan File

- Spoon save file untuk menyimpan file transformasi
- Spoon Transformation opened menandakan Transformasi telah dibuka

b. Menjalankan Transformasi

- Launching transformation [Transformation 1]... Proses transformasi dimulai
- Started the transformation execution. Menandakan Eksekusi transformasi berjalan.
- Dispatching started for transformation [Transformation 1] merupakan PDI mulai mendistribusikan tugas transformasi

c. Ekstraksi Data dari CSV

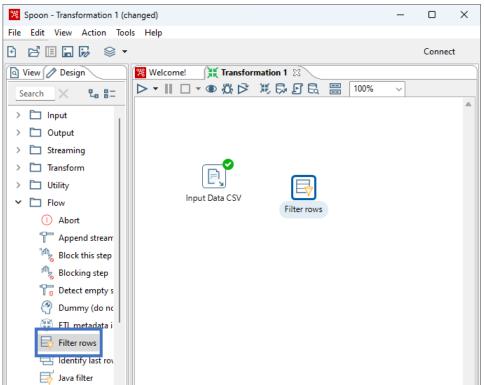
- Input Data CSV.0 Header row skipped in file ... CarDistributionSales.csv Baris header diabaikan (tidak diproses sebagai data).
- Input Data CSV.0 Finished processing (I=20, O=0, R=0, W=19, U=0, E=0) > Statistik pemrosesan
- 3. Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E ? 2024/02/17 06:51:48 Input Data CSV.0 Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)
 - ightharpoonup I=20
 ightharpoonup 20 baris dimasukkan (Input : jumlah total baris yang dibaca misal, CSV, database, API. Dalam log ini, terdapat 20 baris yang diambil dari file tsb).
 - ightharpoonup O=0
 ightharpoonup Tidak ada data yang dikeluarkan (Output : jumlah baris yang dikirim ke langkah selanjutnya dalam alur transformasi. Dalam log ini, tidak bada baris).
 - $ightharpoonup R=0 \rightarrow Tidak ada baris yang di-reject. ($
 - \triangleright W=19 \rightarrow 19 baris ditulis ke tujuan akhir.
 - \triangleright U=0, E=0 \rightarrow Tidak ada data yang diperbarui atau error.
- 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.

Nama	Processor	RAM	Storage	VGA	CPU	Waktu Eksekusi
Dhynar	AMD Ryzen 3 3250U with Radeon Graphics, 2600 Mhz, 2 Core(s), 4 Logical Processor(s)	4,00 GB	1 TB 5400 rpm 2.5" SATA Hard Drive	AMD Radeon TM Graphics	AMD Ryzen 3 3250U with Radeon Graphics	0,01
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	0,01

Danica	Intel ® Core ™ i5- 8365U CPU @ 1.60GHz 1.90 GHz	Core [™] i5- 3365U CPU @ 1.60GHz		Intel ® UHD Graphics 620	Intel ® Core TM i5-8365U CPU @ 1.60GHz 1.90 GHz	0,01
Babby	Intel ® Core TM i3- 6006U CPU @ 2.00 GHz	4,00 GB	99,4 GB	NVIDIA GeForce 920 MX Intel® HD Graphics 520	GeForce Intel ® 920 MX Core TM Intel® i3-6006U HD CPU @ Graphics 2.00 GHz	

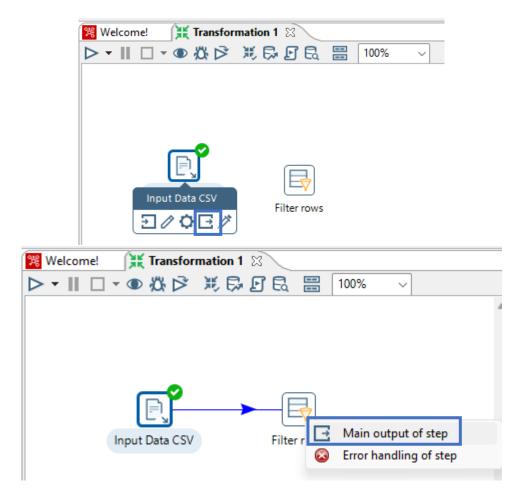
B. Filter Data (Transform) dan Pengemasan data (Load)

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

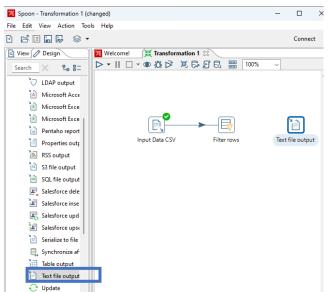


2. Hubungkan output pada Input Data CSV menuju Filter rows. Pilih koneksi Main output of step saat menghubungkan kedua objek tersebut.

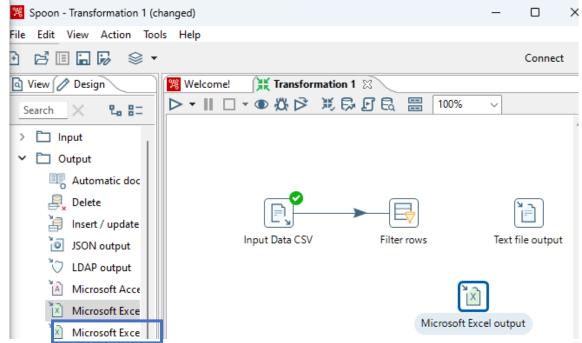
Pada tahap ini Input Data CSV dihubungkan oleh konektor pada Filter rows memiliki makna bahwa setelah proses Input Data CSV dilakukan maka proses selanjutnya adalah Filter rows



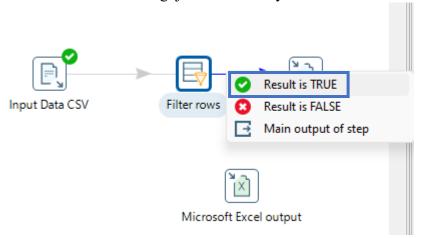
3. Carilah objek **Text file output** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**.



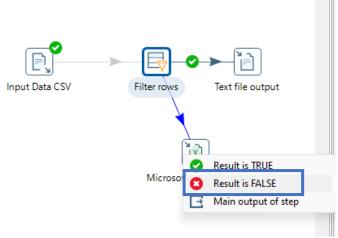
4. Carilah objek **Microsoft excel output** pada **Design Area.** Drag and drop objek tersebut pada **Work Area**



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



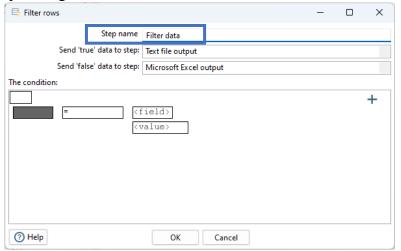
6. Hubungkan filter rows dengan microsoft excel output menggunakan konektor output dari filter rows. Pilih Result is false sebagi jenis konektornya.



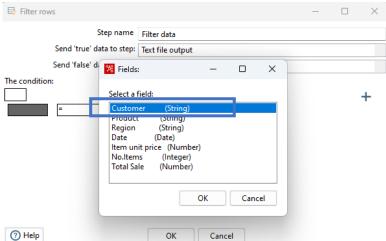
7. Double click pada filter rows hingga muncul jendela konfigurasi. ubah step name menjadi Filter data.

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

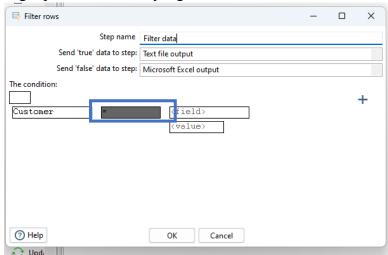
9. Klik <Field> pada bagian condition

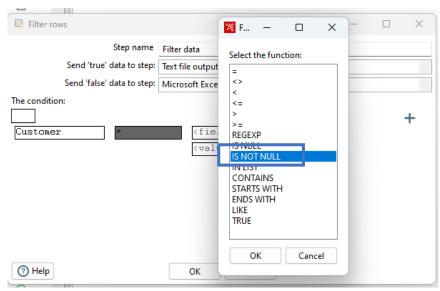


10. Pilih field yang diinginkan untuk digunakan sebagai kondisi. Pada kesempatan ini pilih **Customer**, kemudian klik **OK**.

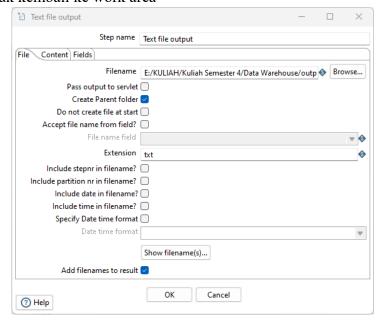


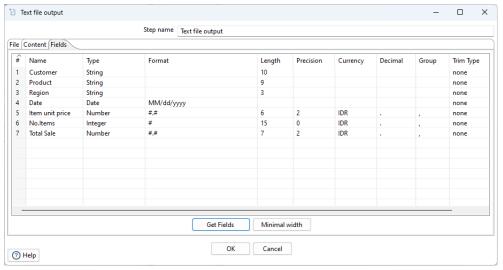
11. Pada bagian **value** pilih **IS NOT NULL** kemudian tekan **OK** , hal ini dilakukan karena data yang dipilih adalah data yang tidak Null



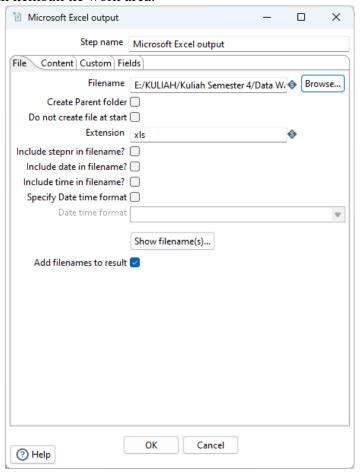


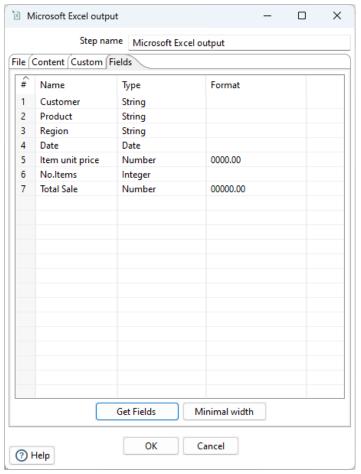
- 12. Tekan tombol + untuk menambah kondisi.
- 13. Pilih statement yang masih null untuk menambah kondisi.
- 14. Ulangi proses nomor 9,10,11 dengan mengganti semua field yang ada. hingga statement menjadi seperti pada gambar dibawah ini.
- 15. Gunakan operator AND untuk menghubung
- 16. Jika semua field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



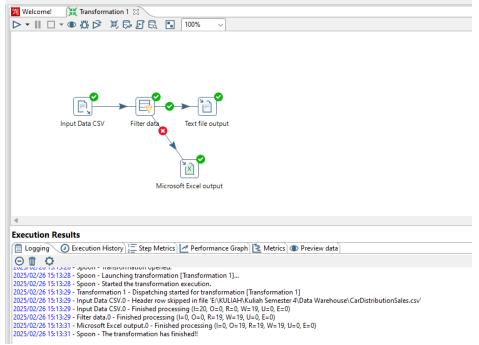


- 21. Hal yang sama dilakukan pada konfigurasi Microsoft Excel output, Double-click pada objek Microsoft Excel output hingga muncul jendela konfigurasi.
- 22. Pilih lokasi file untuk menyimpan file output dengan menekan tombol browse. Beri nama failedSalesCarDistribution. untuk extension tetap dengan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





25. 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?
 - File CSV adalah hasil extract awal sebelum dilakukan transformasi data, sedangkan file Excel sudah melewati proses pembersihan data
- 2. Jelaskan apa yang terjadi pada proses Filter rows!
 - Menginputkan data dari file yang dimasukkan ke dalam PDI
 - Filter rows digunakan untuk menentukan kondisi yang harus dipenuhi oleh data agar bisa lolos ke tahap berikutnya
 - > Data yang telah memenuhi kriteria akan terus diarahkan ke step selanjutnya
 - ➤ Hasil dari filter rows biasanya menghasilkan subset data lebih bersih dan relevan untuk analisis
- 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 Processor	RAM	Storage	VGA	CPU	Waktu Eksekusi
Dhynar	AMD Ryzen 3 3250U with Radeon Graphics, 2600 Mhz, 2 Core(s), 4 Logical Processor(s)	4,00 GB 1 TB 5400 rpm 2.5" SATA Hard Drive		AMD Radeon TM Graphics	AMD Ryzen 3 3250U with Radeon Graphics	0,05
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	0,5
Danica	Intel ® Core ™ i5- 8365U CPU @ 1.60GHz 1.90 GHz	16.0 GB	237 GB	Intel ® UHD Graphics 620	Intel ® Core TM i5-8365U CPU @ 1.60GHz 1.90 GHz	0,3
Babby	Intel ® Core TM i3- 6006U CPU @ 2.00 GHz	4,00 GB	99,4 GB	NVIDIA GeForce 920 MX Intel® HD Graphics 520	Intel ® Core TM i3-6006U CPU @ 2.00 GHz	0,10

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 pemerintahan

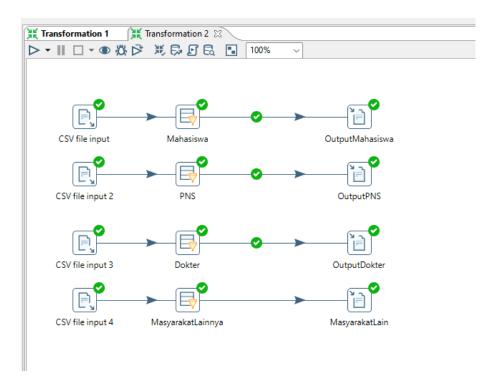
Data masyarakat berprofesi Dokter akan diberikan ke Pak Budi Pekerti selaku ketua RT wilayah kesehatan

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.

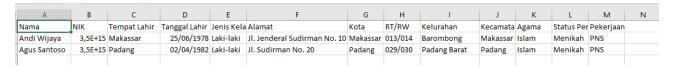
Dikarenakan proses tersebut akan repetisi setiap ada masyarakat Konoha yang berpindah ke IKN maka buatlah sistem tersebut menggunakan PDI Spoon.



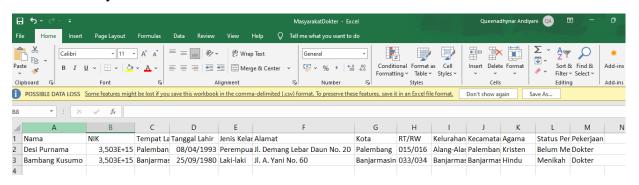
Hasil dari MasyarakatMahasiswa

Nama	NIK	Tempat La	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahar	Kecamatan	Agama	Status Perkawinan	Pekerjaan
Siti Rahayu	3,503E+15	Bandung	15/05/1990	Perempuan	Jl. Cendrawasih No. 45	Bandung	003/004	Cikutra	Bandung Wetan	Kristen	Belum Menikah	Mahasiswa
Adi Pratama	3,503E+15	Semarang	05/03/1995	Laki-laki	Jl. Pemuda No. 21	Semarang	009/010	Pandanara	Semarang Tengah	Buddha	Belum Menikah	Mahasiswa
Maya Wulandari	3,503E+15	Jayapura	23/10/1993	Perempuan	Jl. Proklamasi No. 80	Jayapura	039/040	Kota Jayar	Jayapura	Kristen	Belum Menikah	Mahasiswa
	· '					-						

Hasil dari MasyarakatPNS



Hasil dari MasyarakatDokter



Hasil dari MasyarakatLainnya

