

DATA WAREHOUSE

JOBSHEET 1

DATA WAREHOUSE SEDERAHAN DAN PENGENALAN PDI SPOON

TUGAS 1

Analisa data dari data penjualan Astro Mobil

1. Berapa jumlah kolom pada data tersebut?
⇒ Pada data tersebut terdapat **7** kolom
2. 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 ;
 - a. Customer : Kolom ini berisikan data dari tiap customer yang melakukan pembelian
 - b. Product : Kolom ini berisikan jenis product yang dibeli oleh customer yang bersangkutan
 - c. Region : Kolom ini menunjukkan dari daerah mana / cabang mana transaksi / penjualan tersebut dilakukan
 - d. Date : Kolom ini menunjukkan tanggal ketika transaksi / penjualan dilakukan
 - e. Item unit price : Kolom ini menunjukkan harga satuan dari product yang dibeli oleh customer
 - f. No.items : Kolom ini menunjukkan nomor items yang terjual / dibeli oleh customer
 - g. Total sale : Kolom ini berisikan total harga pembelian dari transaksi / penjualan tersebut
3. Adakah data yang memiliki nilai null / data tidak lengkap?
⇒ Ada, beberapa data memiliki nilai null atau data tidak lengkap diantaranya;
 - a. Baris no 5, tidak ada data pada kolom product
 - b. Baris no 9, tidak ada data pada kolom region dan juga date
 - c. Baris no 14, tidak ada data pada kolom
 - d. Baris no 18, tidak ada data pada kolom product
 - e. Baris no 10, tidak ada data pada kolom Region
4. 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

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

⇒ Ubah step name

CSV file input

Step name: Input data CSV

Filename: [Browse...]

Delimiter: , [Insert IAB]

Enclosure: " [Insert IAB]

NIO buffer size: 50000 [Insert IAB]

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional): [Insert IAB]

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

⇒ CarDistributionSales.csv

CSV file input

Step name: Input data CSV

Filename: D:\Kuliah\Semester 4\Data Warehouse\Dini Elminingtyas\CarDi... [Browse...]

Delimiter: , [Insert IAB]

Enclosure: " [Insert IAB]

NIO buffer size: 50000 [Insert IAB]

Lazy conversion? ☒

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

⇒ Sudah sesuai

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

⇒ Get Fields

CSV file input

Step name: Input data CSV

Filename: D:\Kuliah\Semester 4\Data Warehouse\Dini Elminingtyas\CarDi... [Browse...]

Delimiter: , [Insert IAB]

Enclosure: " [Insert IAB]

NIO buffer size: 50000 [Insert IAB]

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional): [Insert IAB]

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

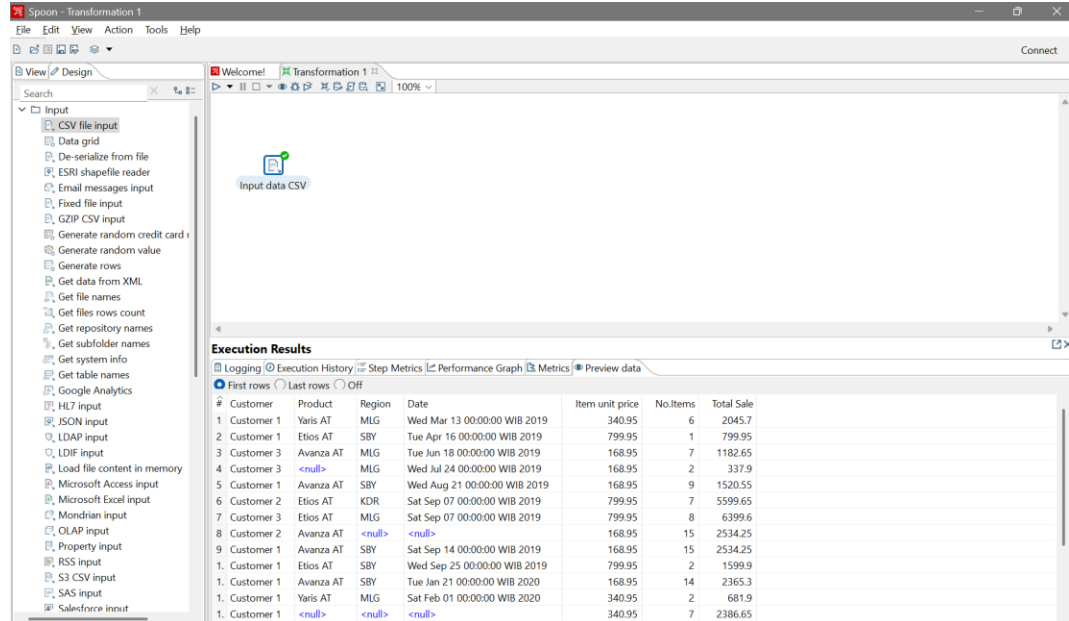
File encoding: [Insert IAB]

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	T
1	Customer	String		10		\$.	,	n
2	Product	String		9		\$.	,	n
3	Region	String		3		\$.	,	n
4	Date	Date	MM/dd/yyyy			\$.	,	n
5	Item unit price	Number	#,.	6	2	\$.	,	n
6	No.Items	Integer	#	15	0	\$.	,	n
7	Total Sale	Number	#,.	7	2	\$.	,	n

[Help] [OK] [Get Fields] [Preview] [Cancel]

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.

⇒ 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
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?
⇒ Langkah – langkah yang dilakukan untuk extract file data tersebut diantaranya adalah ;
 - a. Save file, atau menyimpan file terlebih dahulu
 - b. Kemudian membuka file Transformation
 - c. Launching file Transformation
 - d. Memulai eksekusi file Transformation
 - e. Mengirimkan data untuk file Transformation
 - f. Proses input data (header skipped)
 - g. Proses input data (finishing)

3. Perhatikan gambar dibawah ini, apakah 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)

⇒ 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

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.

⇒ Perbandingan dengan 3 teman saya :

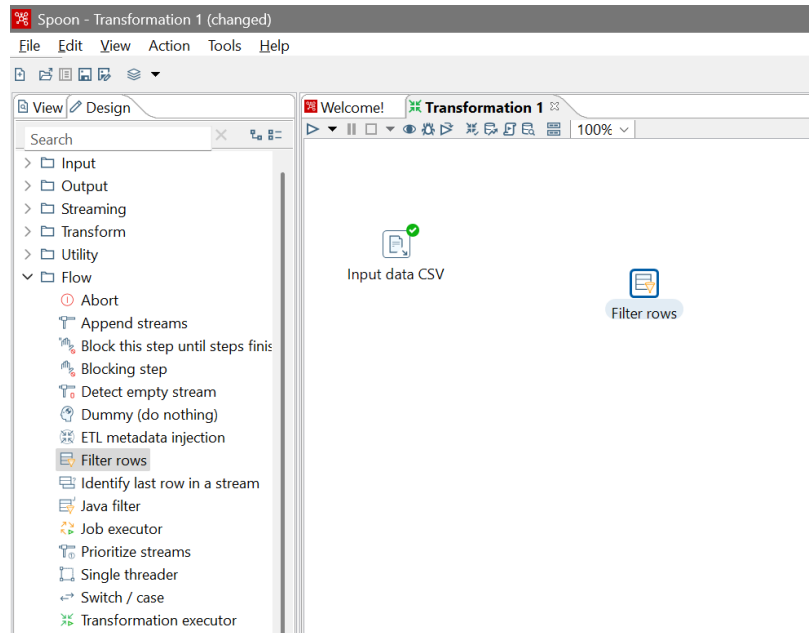
Name	Processor	RAM	Storage	VGA	CPU	Time
Dini(saya)	AMD Ryzen 7 5800HS with Radeon Graphics 3.20 GHz	16 GB	475GB	AMD Radeon Graphics	AMD Ryzen 7 5800HS with Radeon Graphics	1'ms

*teman saya belum buat

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

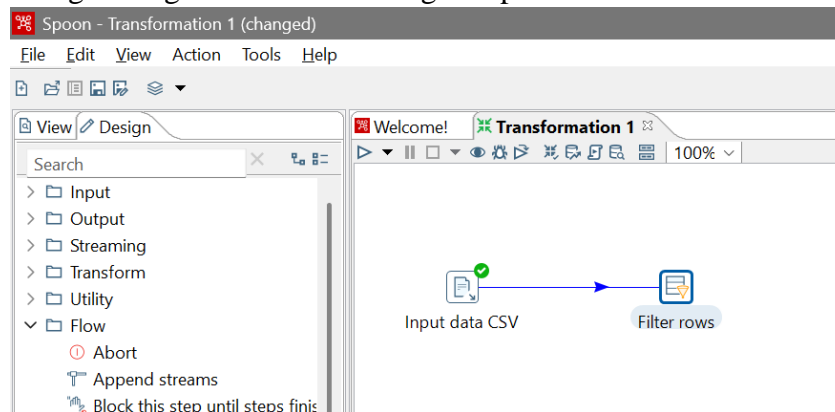
1. Carilah objek **Filter rows** pada **Design Area**. Drag and drop objek tersebut pada **Work Area**.

⇒ Filter Rows



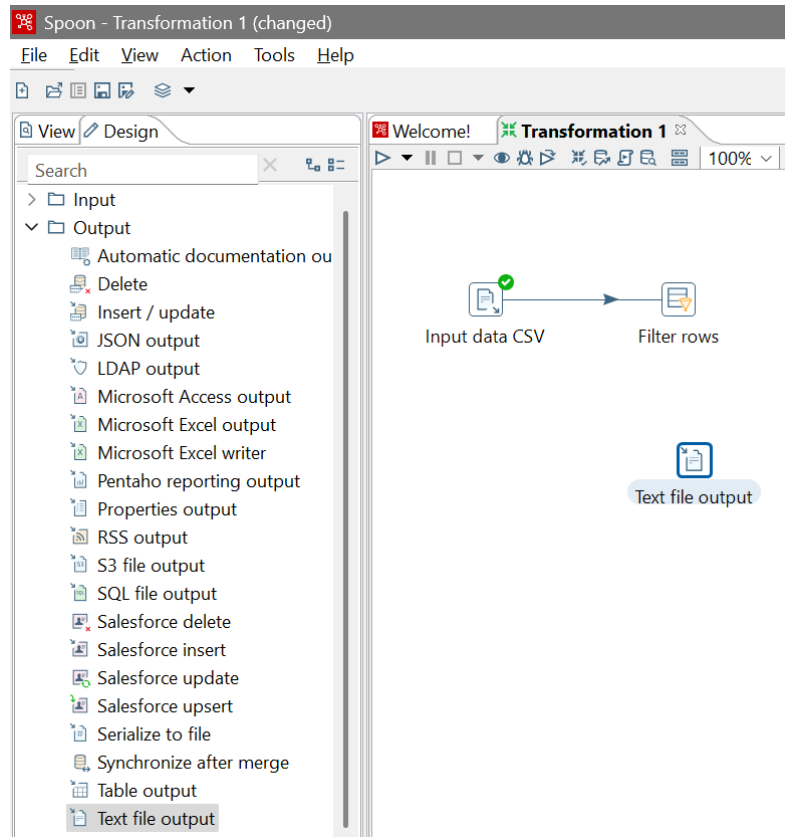
2. Hubungkan output pada Input Data CSV menuju Filter rows. Pilih koneksi Main output of step saat menghubungkan kedua objek tersebut. Pada tahap ini Input Data CSV dihubungkan oleh konektor pada Filter rows memiliki makna bahwa setelah proses Input Data CSV dilakukan maka proses selanjutnya adalah Filter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

⇒ Menghubungkan filter rows dengan input data csv



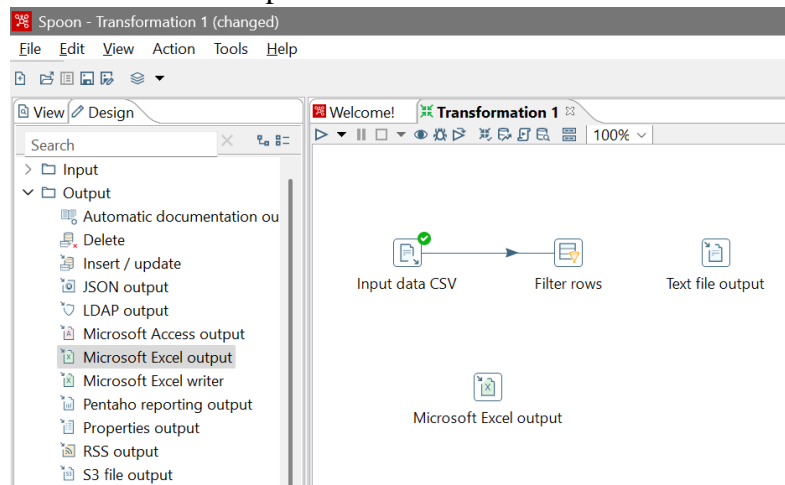
3. Carilah objek **Text file output** pada **Design Area**. Drag and drop objek tersebut pada **Work Area**.

⇒ Text file output



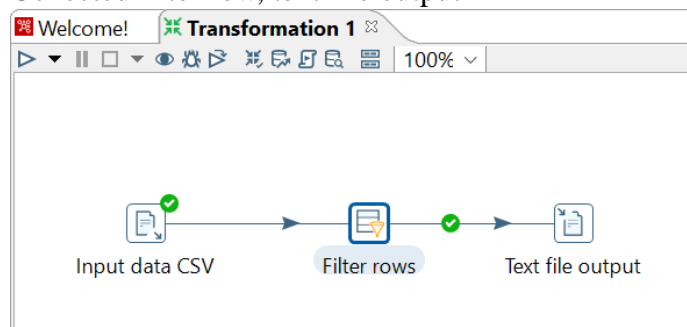
4. Carilah objek **Microsoft excel output** pada **Design Area**. Drag and drop objek tersebut pada **Work Area**.

⇒ Microsoft Excel output

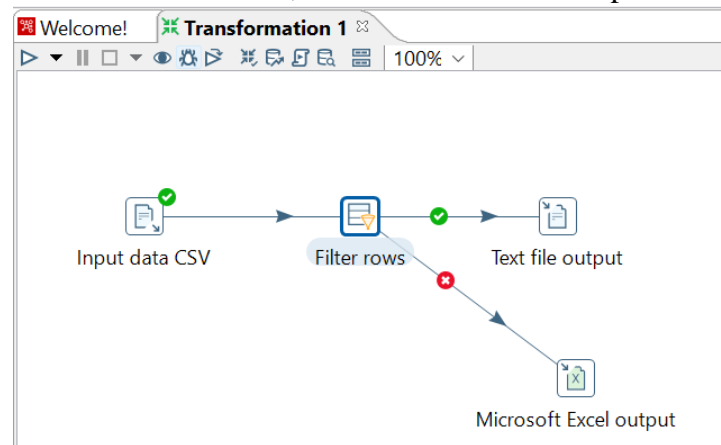


5. Hubungkan **filter rows** dengan **text file output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is true** sebagai jenis konektornya.

⇒ Conected filter row, text file output

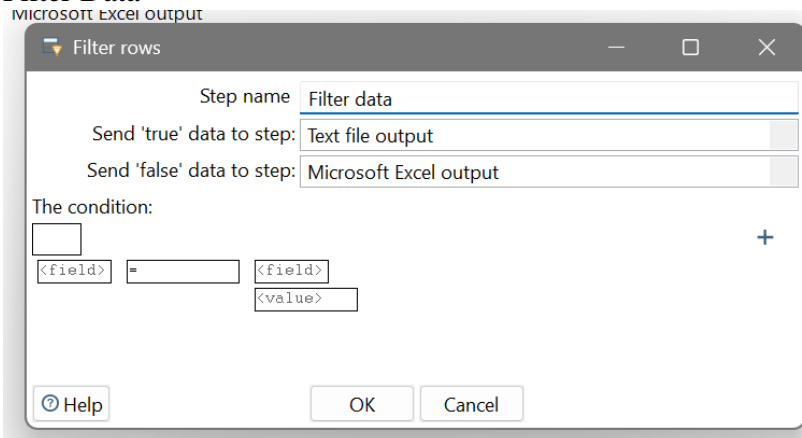


6. Hubungkan **filter rows** dengan **microsoft excel output** menggunakan **konektor output** dari **filter rows**. Pilih **Result is false** sebagai jenis konektornya.
⇒ Connected filter rows, dan Microsoft excel output



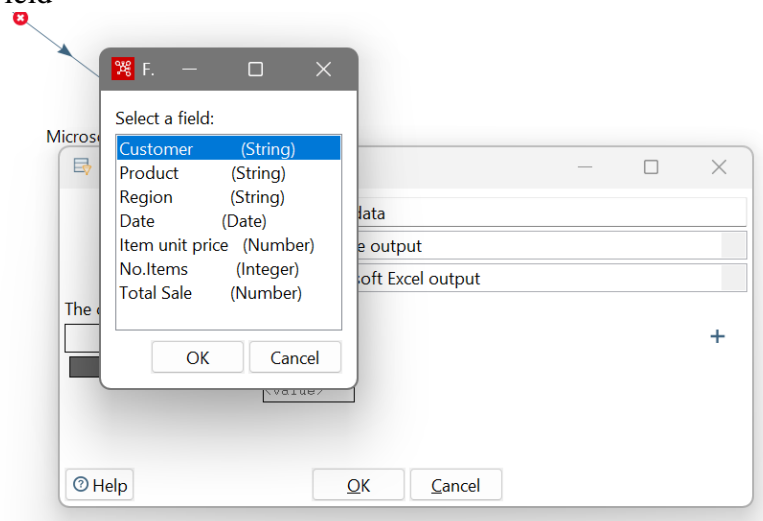
7. **Double click** pada **filter rows** hingga muncul jendela konfigurasi. ubah **step name** menjadi **Filter data**.

⇒ Filter Data



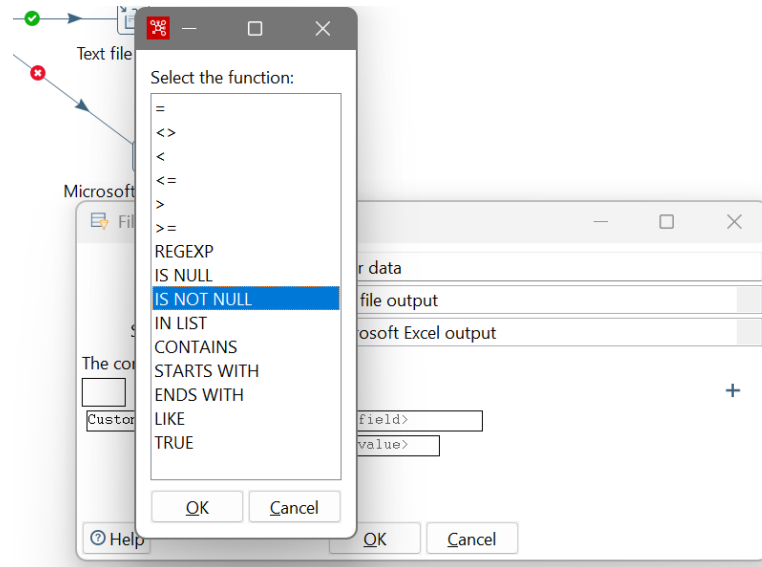
8. Cek bagian **send 'true' data to step** menjadi **Text file output** dan cek bagian **Send 'false' data to step** menjadi **Microsoft Excel output**.
9. Klik **<Field>** pada bagian **condition**
10. Pilih field yang diinginkan untuk digunakan sebagai kondisi. Pada kesempatan ini pilih **Customer**,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

⇒ Field



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

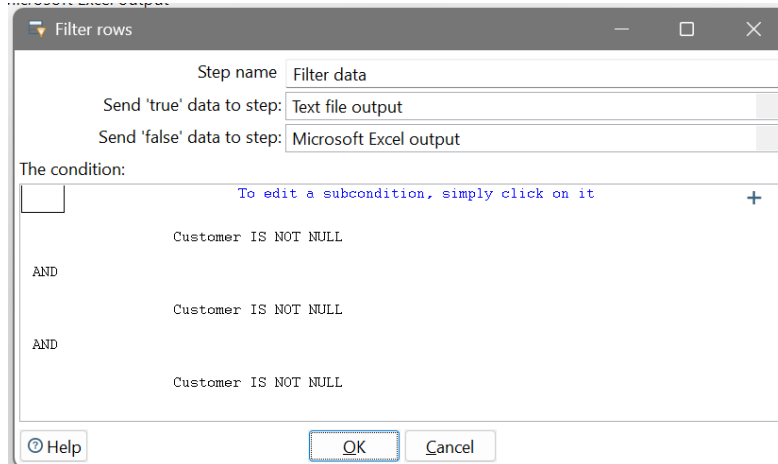
⇒ Is not null



12. Tekan tombol + untuk menambah kondisi.

13. Pilih statement yang masih null untuk menambah kondisi.

⇒ 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

⇒ Menambah condition

To edit a subcondition, si

Customer IS NOT NULL

AND

Product IS NOT NULL

AND

Region IS NOT NULL

AND

Date IS NOT NULL

AND

Item unit price IS NOT NULL

AND

No.Items IS NOT NULL

AND

Total Sale IS NOT NULL

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.

⇒ Ubah extension

Text file output

Step name Text file output

File Content Fields

Filename file

Pass output to servlet ☐

Create Parent folder ☒

Do not create file at start ☐

Accept file name from field? ☐

File name field

Extension csv

Include stepnr in filename? ☐

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

⇒ Mengambil field data

Text file output

Step name Text file output

File Content Fields

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim Type	N
1	Customer	String		10					none	
2	Product	String		9					none	
3	Region	String		3					none	
4	Date	Date	MM/dd/yyyy						none	
5	Item unit price	Number	#,.	6	2	\$.	,	none	
6	No.Items	Integer	#	15	0	\$.	,	none	
7	Total Sale	Number	#,.	7	2	\$.	,	none	

Get Fields Minimal width

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.

⇒ File location, dan extension

The screenshot shows the 'Microsoft Excel output' configuration window with the 'File' tab selected. The 'Step name' is 'Microsoft Excel output'. The 'Filename' field is set to 'D:/Kuliah/Semester 4/Data Warehouse' with a 'Browse...' button next to it. Below this, there are several checkboxes: 'Create Parent folder' (unchecked), 'Do not create file at start' (unchecked), 'Include stepnr in filename?' (unchecked), 'Include date in filename?' (unchecked), 'Include time in filename?' (unchecked), and 'Specify Date time format' (unchecked). The 'Extension' field is set to 'xls'. There is a 'Date time format' dropdown menu. A 'Show filename(s)...' button is also present. At the bottom, the 'Add filenames to result' checkbox is checked.

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

⇒ Fields

The screenshot shows the 'Microsoft Excel output' configuration window with the 'Fields' tab selected. The 'Step name' is 'Microsoft Excel output'. The 'Fields' tab displays a table with the following data:

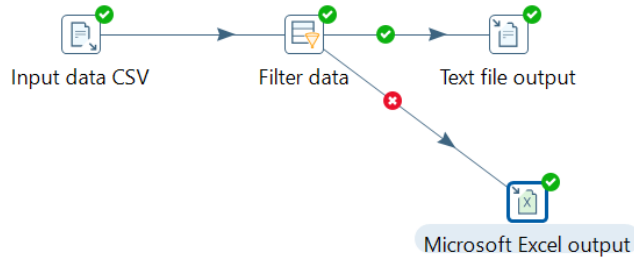
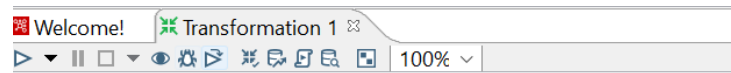
#	Name	Type	Format
1	Customer	String	
2	Product	String	
3	Region	String	
4	Date	Date	
5	Item unit price	Number	0000.00
6	No.Items	Integer	
7	Total Sale	Number	00000.00

At the bottom of the window, there are two buttons: 'Get Fields' and 'Minimal width'.

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.

⇒ Execution Results



- failedSalesCarDistributon

Execution Results

<div> Logging Execution History Step Metrics Performance Graph Metrics Preview data </div>							
<input checked="" type="radio"/> First rows <input type="radio"/> Last rows <input type="radio"/> Off							
#	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
1	Customer 3	<null>	MLG	Wed Jul 24 00:00:00 WIB 2019	168.95	2	337.9
2	Customer 2	Avanza AT	<null>	<null>	168.95	15	2534.25
3	Customer 1	<null>	<null>	<null>	340.95	7	2386.65
4	Customer 2	<null>	KDR	Fri Aug 07 00:00:00 WIB 2020	799.95	14	11199.3
5	Customer 1	Avanza AT	<null>	Wed Aug 12 00:00:00 WIB 2020	168.95	4	675.8

- outputSlesCarDistribution

Execution Results

<div> Logging Execution History Step Metrics Performance Graph Metrics Preview data </div>							
<input checked="" type="radio"/> First rows <input type="radio"/> Last rows <input type="radio"/> Off							
#	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
1	Customer 1	Yaris AT	MLG	Wed Mar 13 00:00:00 WIB 2019	340.95	6	2045.7
2	Customer 1	Etios AT	SBY	Tue Apr 16 00:00:00 WIB 2019	799.95	1	799.95
3	Customer 3	Avanza AT	MLG	Tue Jun 18 00:00:00 WIB 2019	168.95	7	1182.65
4	Customer 1	Avanza AT	SBY	Wed Aug 21 00:00:00 WIB 2019	168.95	9	1520.55
5	Customer 2	Etios AT	KDR	Sat Sep 07 00:00:00 WIB 2019	799.95	7	5599.65
6	Customer 3	Etios AT	MLG	Sat Sep 07 00:00:00 WIB 2019	799.95	8	6399.6
7	Customer 1	Avanza AT	SBY	Sat Sep 14 00:00:00 WIB 2019	168.95	15	2534.25
8	Customer 1	Etios AT	SBY	Wed Sep 25 00:00:00 WIB 2019	799.95	2	1599.9
9	Customer 1	Avanza AT	SBY	Tue Jan 21 00:00:00 WIB 2020	168.95	14	2365.3
1.	Customer 1	Yaris AT	MLG	Sat Feb 01 00:00:00 WIB 2020	340.95	2	681.9
1.	Customer 1	Etios AT	SBY	Sat May 30 00:00:00 WIB 2020	799.95	9	7199.55
1	Customer 3	Avanza AT	MLG	Sun Jun 07 00:00:00 WIB 2020	168.95	10	1689.5

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
2. 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.
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
⇒ 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. dari 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

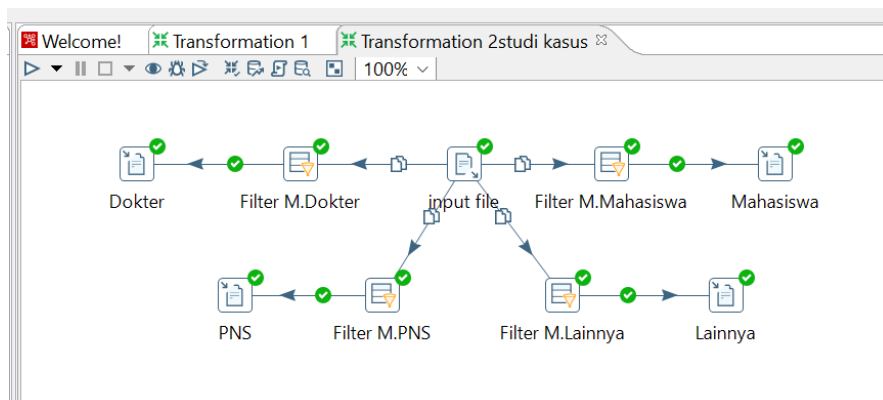
Data masyarakat berprofesi Dokter akan diberikan ke Pak Budi Pekerti selaku ketua RT wilayah kesehatan

data-data tersebut akan digunakan untuk analisa di masing-masing wilayah dan juga Dashboard untuk emmbantu mengambil keputusan.

Pak Pascanowo mendapatkan data penduduk yang akan pindah ke wilayahnya melalui link: <https://github.com/dik4rizky/datasources/blob/82821b8ec341176f979956c84b79720d4aead012/dummyDumyKependudukan.csv>

Bantulah Pak Pascanowo untuk memishkan data tersebut menjadi 4 file:
MasyarakatMahasiswa, MasyarakatPNS, MasyarakatDokter, MasyarakatLainnya.

PDI SISTEM ;



Execution Results

1. Masyarakat Mahasiswa

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecam
1	Siti Rahayu	3503010101010002	Bandung	Tue May 15 00:00:00 WIB 1990	Perempuan	Jl. Cendrawasih No. 45	Bandung	003/004	Cikutra	Bandu
2	Adi Pratama	3503010101010005	Semarang	Sun Mar 05 00:00:00 WIB 1995	Laki-laki	Jl. Pemuda No. 21	Semarang	009/010	Pandanaran	Semar
3	Maya Wulandari	3503010101010020	Jayapura	Sat Oct 23 00:00:00 WIB 1993	Perempuan	Jl. Proklamasi No. 80	Jayapura	039/040	Kota Jayapura	Jayapu

2. Masyarakat PNS

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecar
1	Andi Wijaya	3503010101010007	Makassar	Sun Jun 25 00:00:00 WIB 1978	Laki-laki	Jl. Jenderal Sudirman No. 10	Makassar	013/014	Barombong	Maka
2	Agus Santoso	3503010101010015	Padang	Fri Apr 02 00:00:00 WIB 1982	Laki-laki	Jl. Sudirman No. 20	Padang	029/030	Padang Barat	Padai

3. Masyarakat Dokter

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	
1	Desi Purnama	3503010101010008	Palembang	Thu Apr 08 00:00:00 WIB 1993	Perempuan	Jl. Demang Lebar Daun No. 20	Palembang	015/016	Alang-Alang	
2	Bambang Kusumo	3503010101010017	Banjarmasin	Thu Sep 25 00:00:00 WIB 1980	Laki-laki	Jl. A. Yani No. 60	Banjarmasin	033/034	Banjarmasin	

4. Masyarakat Lainnya

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	
1	Budi Santoso	3503010101010001	Jakarta	Tue Jan 01 00:00:00 WIB 1980	Laki-laki	Jl. Merdeka No. 123	Jakarta	001/002	Menteng	
2	Joko Susilo	3503010101010003	Surabaya	Fri Oct 10 00:00:00 WIB 1975	Laki-laki	Jl. Diponegoro No. 67	Surabaya	005/006	Gubeng	
3	Maya Dewi	3503010101010004	Yogyakarta	Sat Aug 20 00:00:00 WIB 1988	Perempuan	Jl. Malioboro No. 8	Yogyakarta	007/008	Sosromendura	
4	Rini Setiawan	3503010101010006	Malang	Sun Dec 12 00:00:00 WIB 1982	Perempuan	Jl. Simpang Borobudur No. 15	Malang	011/012	Dinoyo	
5	Fajar Prabowo	3503010101010009	Medan	Tue Sep 30 00:00:00 WIB 1986	Laki-laki	Jl. Sudirman No. 50	Medan	017/018	Sei Sikambing	
6	Rina Sari	3503010101010010	Denpasar	Fri Jul 17 00:00:00 WIB 1992	Perempuan	Jl. Gatot Subroto No. 30	Denpasar	019/020	Dauh Puri Kaut	
7	Ahmad Subhan	3503010101010011	Surakarta	Wed Jul 03 00:00:00 WIB 1991	Laki-laki	Jl. Pahlawan No. 55	Surakarta	021/022	Jebres	
8	Dian Pertiwi	3503010101010012	Maluku	Tue Dec 10 00:00:00 WIB 1985	Perempuan	Jl. Pattimura No. 10	Ambon	023/024	Baguala	
9	Rudi Hartono	3503010101010013	Pontianak	Mon May 28 00:00:00 WIB 1973	Laki-laki	Jl. Ahmad Yani No. 30	Pontianak	025/026	Pontianak Bara	
1.	Lina Fitriani	3503010101010014	Manado	Mon Aug 15 00:00:00 WIB 1994	Perempuan	Jl. Wolter Monginsidi No. 17	Manado	027/028	Mapanget	
1.	Rina Agustina	3503010101010016	Bengkulu	Mon Nov 19 00:00:00 WIB 1990	Perempuan	Jl. Imam Bonjol No. 45	Bengkulu	031/032	Ratu Samban	
1.	Nia Dewi Lestari	3503010101010018	Manokwari	Sun Jun 14 00:00:00 WIB 1987	Perempuan	Jl. Diponegoro No. 40	Manokwari	035/036	Manokwari Bar	
1.	Doni Setiawan	3503010101010019	Palu	Mon Mar 07 00:00:00 WIB 1977	Laki-laki	Jl. Soekarno-Hatta No. 70	Palu	037/038	Palu Barat	