

LAPORAN PRAKTIKUM

DATA WAREHOUSE

JOB SHEET 1

DATA WAREHOUSE SEDERHANA DAN PENGENALAN PDI SPOON



Oleh: Rizky Roza Rahim

2341760150

SIB 2B

PROGRAM STUDI SISTEM INFORMASI BISNIS

JURUSAN TEKNOLOGI INFORMASI

POLITEKNIK NEGERI MALANG

TUGAS 1

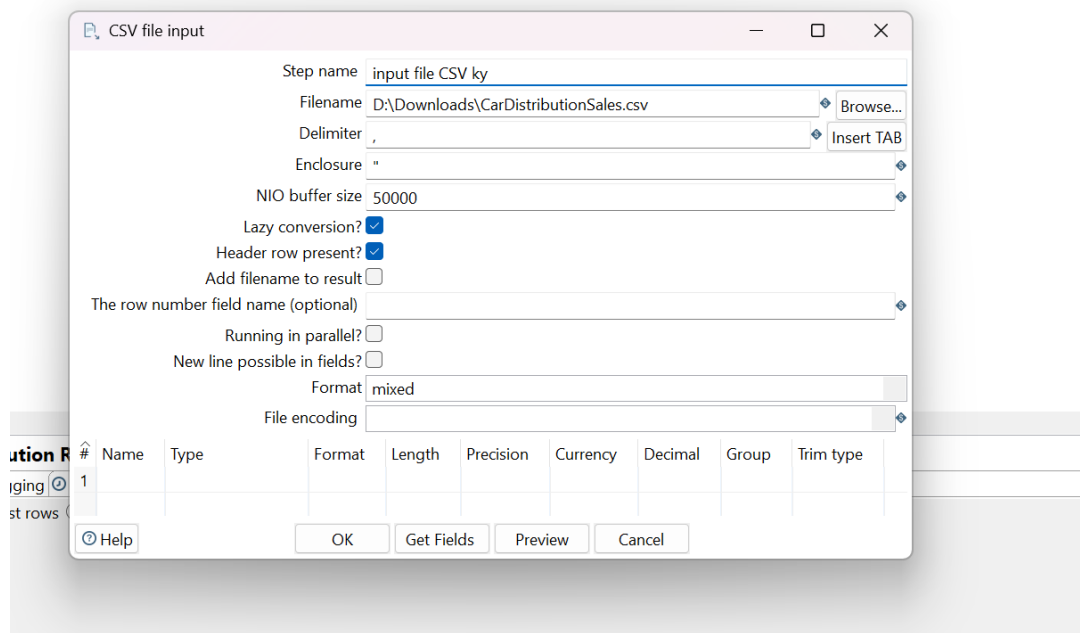
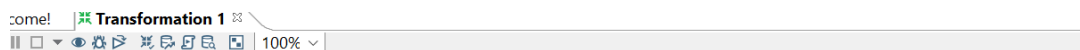
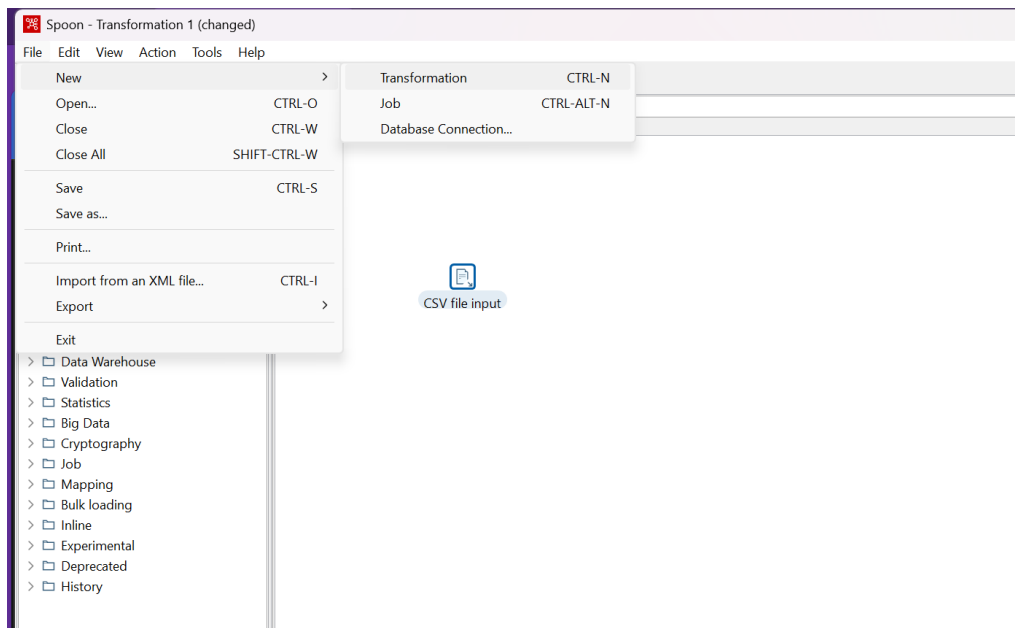
Analisa lah data tersebut!

1. Berapa jumlah kolom pada data tersebut?
 - Ada 7 kolom
2. Apa arti atau isi data dari setiap kolom yang ada?
 - Customer = siapa yang menjadi customer
 - Product = produk yang di beli customer
 - Region = Alamat customer
 - Date = tanggal transaksi
 - Item unit price = harga barang
 - No.Item = No item yang di beli
 - Total item = total pembelian
3. Adakah data yang memiliki nilai null / data yang tidak lengkap?
 - Ada, baris 4,8,13,17,19
4. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

Tidak ada, semua tipe sesuai, tetapi banyak beberapa kolom yang tidak diisi

A. Pengambilan data

1. Buka 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, 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.



CSV file input

Step name: input data CSV ky

Filename: D:\Downloads\CarDistributionSales.csv [Browse...](#)

Delimiter: , [Insert TAB](#)

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

File encoding:

#	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](#)

Search

Input

- CSV file input
- Data grid
- De-serialize from file
- ESRI shapefile reader
- Email messages input
- Fixed file input
- GZIP CSV input
- Generate random credit card
- Generate random value
- Generate rows
- Get data from XML
- Get file names
- Get files rows count
- Get repository names
- Get subfolder names
- Get system info
- Get table names
- Google Analytics
- HL7 input
- JSON input
- LDAP input
- LDIF input
- Load file content in memory
- Microsoft Access input
- Microsoft Excel input
- Mondrian input
- OLAP input
- Property input
- RSS input

input file csv ky

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
1	Customer 1	Yaris AT	MLG	Wed Mar 13 00:00:00 ICT 2019	340.95	6	2045.7
2	Customer 1	Elios AT	SBY	Tue Apr 16 00:00:00 ICT 2019	799.95	1	799.95
3	Customer 3	Avanza AT	MLG	Tue Jun 18 00:00:00 ICT 2019	168.95	7	1182.65
4	Customer 3	<null>	MLG	Wed Jul 24 00:00:00 ICT 2019	168.95	2	337.9
5	Customer 1	Avanza AT	SBY	Wed Aug 21 00:00:00 ICT 2019	168.95	9	1520.55
6	Customer 2	Elios AT	KDR	Sat Sep 07 00:00:00 ICT 2019	799.95	7	5599.65
7	Customer 3	Elios AT	MLG	Sat Sep 07 00:00:00 ICT 2019	799.95	8	6399.6
8	Customer 2	Avanza AT	MLG	Sat Sep 07 00:00:00 ICT 2019	168.95	15	2534.25

TUGAS 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?
 - Sesuai, nilai null nya pun sesuai
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?

2025/02/21 19:09:18 - Spoon - Save file as...

2025/02/21 19:09:18 - Spoon - Transformation opened.

2025/02/21 19:09:18 - Spoon - Launching transformation [Transformation1]

2025/02/21 19:09:18 - Spoon - Started the transformation execution.

2025/02/21 19:09:18 - Transformation1 - Dispatching started for transformation [Transformation1]

2025/02/21 19:09:18 - input file csv ky.0 - Header row skipped in file 'D:\Downloads\CarDistributionSales.csv'

2025/02/21 19:09:18 - input file csv ky.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)

2025/02/21 19:09:18 - Spoon - The transformation has finished!!

Transformasi dibuka

Mengeksekusi transformasi

Melakukan pengiriman transformasi

selesai

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)

- I = input
- O = output
- R = rejected
- W = written
- U = update
- E = errors

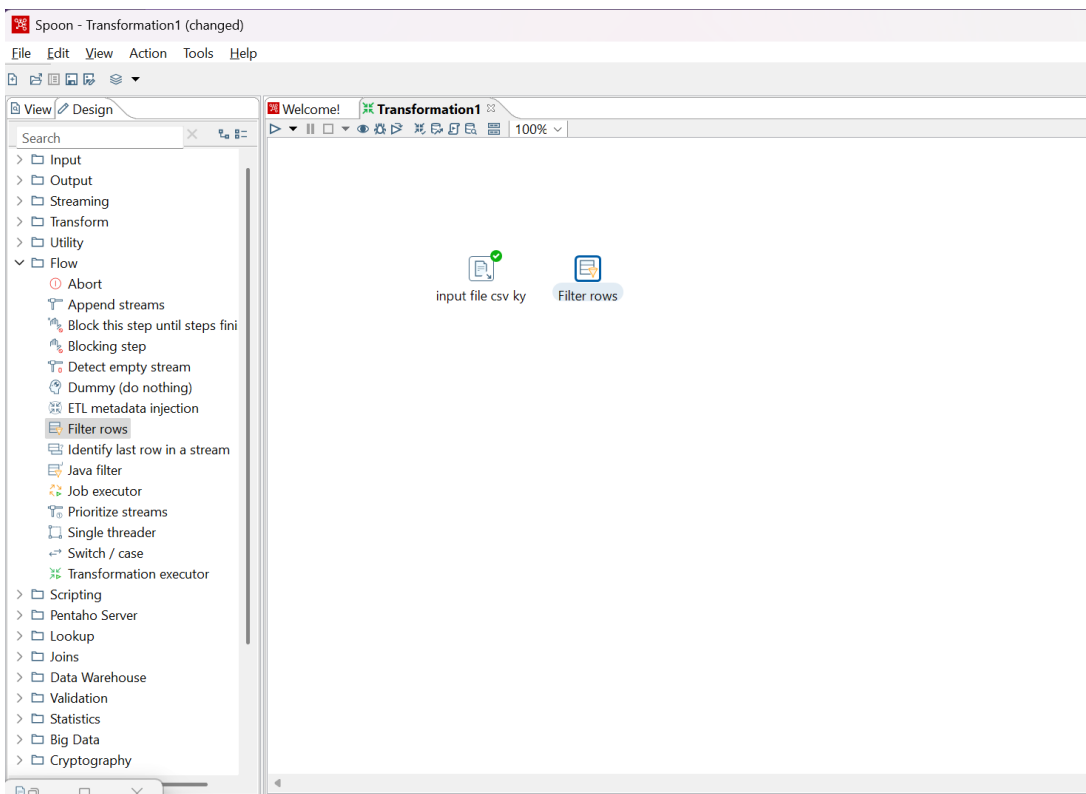
4. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract data ini? Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU). Bandingkan dalam bentuk tabel.

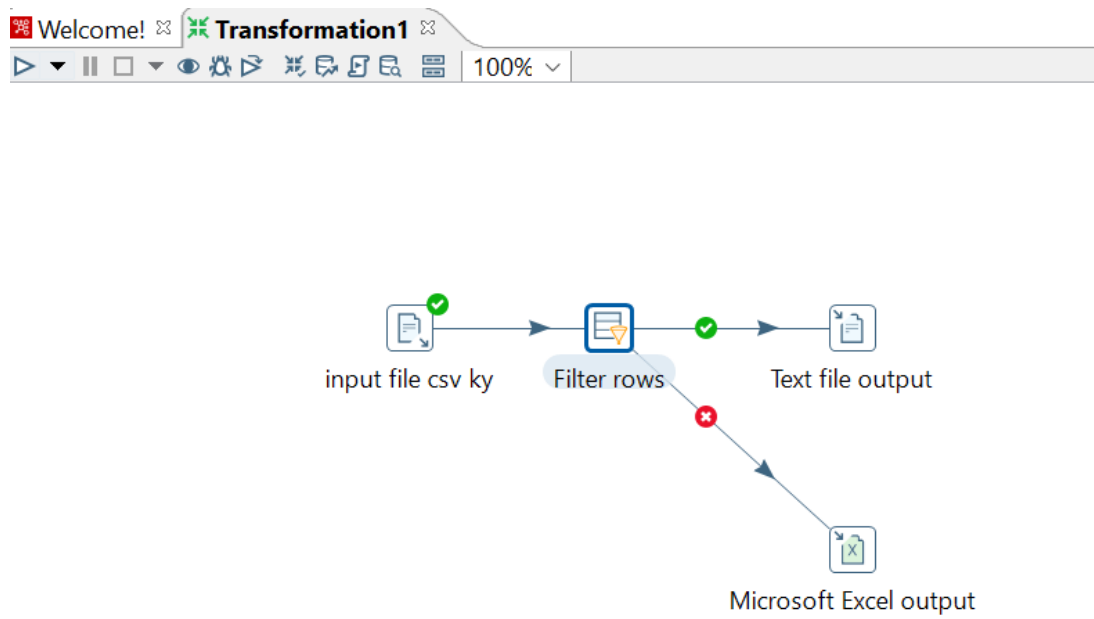
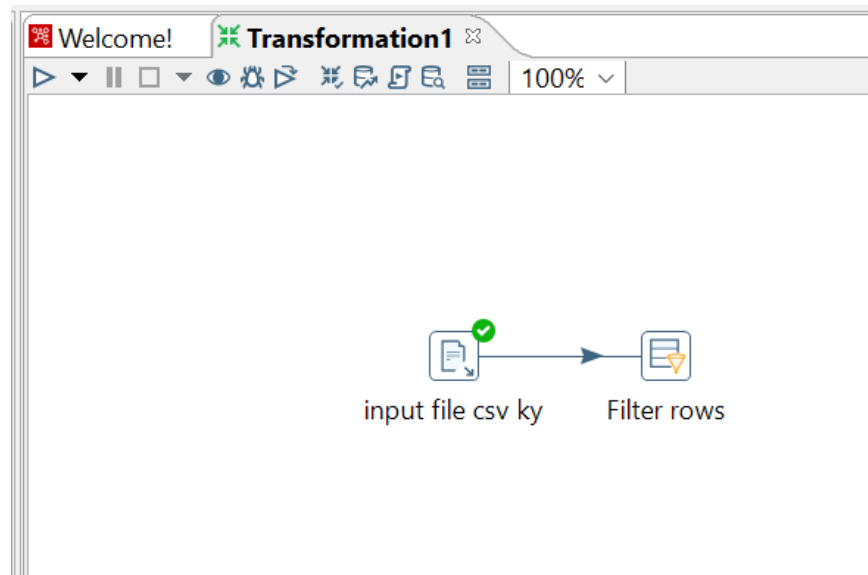
Nama	Processor	Ram	Storage	Vga	Cpu	Time
Rizky	AMD RYZEN 3 7320U with Radeon Graphics 2.40 GHz	8,00 gb	512 gb	AMD RADEON GRAPHICS	AMD RYZEN 3 7320U with Radeon Graphics	0.1s
Bagas	Intel(R) Core(TM) i5-53000 CPU 2.30GHz	16,0 GB	256 GB	Intel (R) HD Graphics 5500	Intel(R) Core(TM) i5-53000 CPU 2.30GHz	0,1s
Daffa	AMD A9 9425 RADEON R5	8,0 GB	1,2 TB	AMD RADEON (TM) R5 Graphics	AMD A9 9425 RADEON R5	0.3s
Dhevina	12th Gen Intel(R) Core(TM) i5 - 1235U 1.30 GHz	8,00 GB	152 GB	Inter(R) UHD Graphics	12th Gen Intel(R) Core(TM) i5 - 1235U	0.0s

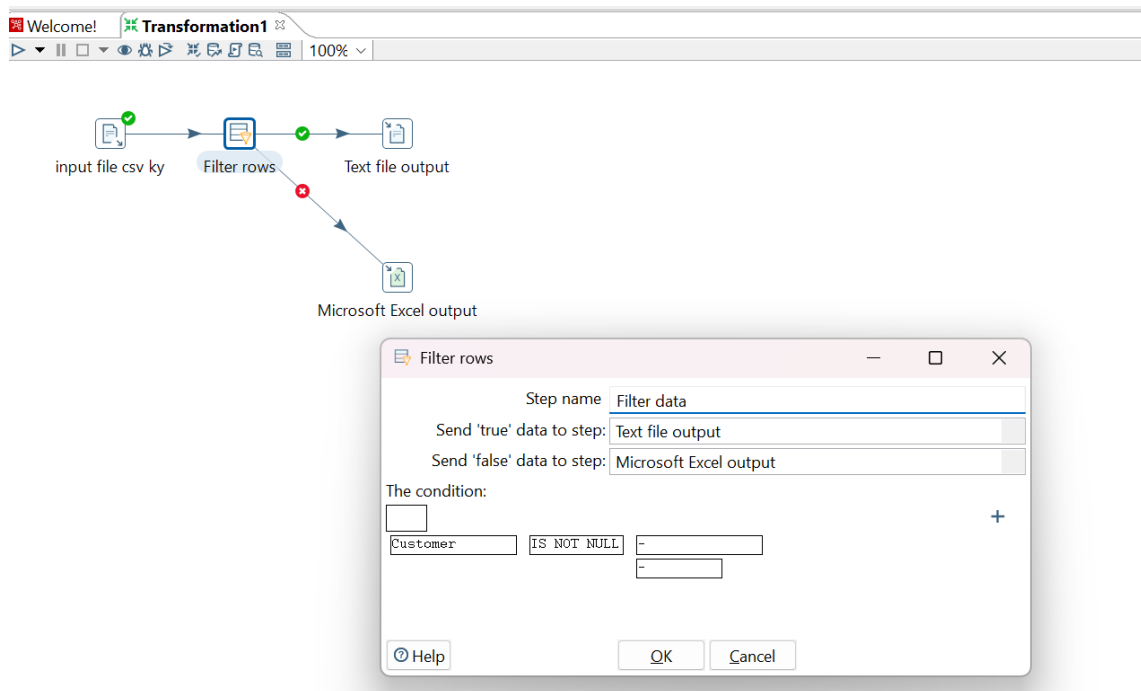
B. Filter Data (Transform) dan Pengemasan data (Load)

1. Carilah objek Filter rows pada Design Area. Drag and drop objek tersebut pada Work Area. Gambar objek filter rows pada area kerja
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 sebagai jenis konektornya.
6. Hubungkan filter rows dengan microsoft excel output menggunakan konektor output dari filter rows. Pilih Result is false sebagai 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 menghubungkan
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. 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.







The condition:

AND

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

Microsoft Excel output

Step name: Microsoft Excel output

File Content Custom Fields

#	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

Get Fields Minimal width

Help OK Cancel

Microsoft Excel output

Step name: Microsoft Excel output

File Content Custom Fields

Filename: D:/KULIAH/s4/dw/failedSalesCarDistri Browse...

Create Parent folder ☐

Do not create file at start ☐

Extension: xls

Include stepnr in filename? ☐

Include date in filename? ☐

Include time in filename? ☐

Specify Date time format ☐

Date time format:

Show filename(s)...

Add filenames to result ☒

Help OK Cancel

Execution Results

Logging

Execution History

Step Metrics

Performance Graph

Metrics

Preview data

☒ First rows
 ☐ Last rows
 ☐ Off

#	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
1	Customer 3	<null>	MLG	Wed Jul 24 00:00:00 ICT 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 ICT 2020	799.95	14	11199.3
5	Customer 1	Avanza AT	<null>	Wed Aug 12 00:00:00 ICT 2020	168.95	4	675.8

Gambar failedSalesCarDistribution

Execution Results

Logging

Execution History

Step Metrics

Performance Graph

Metrics

Preview data

First rows

Last rows

Off

#	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
1	Customer 1	Yaris AT	MLG	Wed Mar 13 00:00:00 ICT 2019	340.95	6	2045.7
2	Customer 1	Etios AT	SBY	Tue Apr 16 00:00:00 ICT 2019	799.95	1	799.95
3	Customer 3	Avanza AT	MLG	Tue Jun 18 00:00:00 ICT 2019	168.95	7	1182.65
4	Customer 1	Avanza AT	SBY	Wed Aug 21 00:00:00 ICT 2019	168.95	9	1520.55
5	Customer 2	Etios AT	KDR	Sat Sep 07 00:00:00 ICT 2019	799.95	7	5599.65
6	Customer 3	Etios AT	MLG	Sat Sep 07 00:00:00 ICT 2019	799.95	8	6399.6
7	Customer 1	Avanza AT	SBY	Sat Sep 14 00:00:00 ICT 2019	168.95	15	2534.25
8	Customer 1	Etios AT	SBY	Wed Sep 25 00:00:00 ICT 2019	799.95	2	1599.9
9	Customer 1	Avanza AT	SBY	Tue Jan 21 00:00:00 ICT 2020	168.95	14	2365.3
1.	Customer 1	Yaris AT	MLG	Sat Feb 01 00:00:00 ICT 2020	340.95	2	681.9
1.	Customer 1	Etios AT	SBY	Sat May 30 00:00:00 ICT 2020	799.95	9	7199.55
1.	Customer 3	Avanza AT	MLG	Sun Jun 07 00:00:00 ICT 2020	168.95	10	1689.5
1.	Customer 3	Yaris AT	MLG	Tue Jun 09 00:00:00 ICT 2020	340.95	4	1363.8

Gambar outputSalesCarDistribution

TUGAS 3

1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?
 - Csv hasil extract diawal sebelum melakukan transformasi data, jika file excel sudah melewati proses pembersihan data

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

- Input
- Proses
- output

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	Ram	Storage	Vga	Cpu	Time
Rizky	AMD RYZEN 3 7320U with Radeon Graphics 2.40 GHz	8,00 gb	512 gb	AMD RADEON GRAPHICS	AMD RYZEN 3 7320U with Radeon Graphics	0.2s
Bagas	Intel(R) Core(TM) i5-53000 CPU 2.30GHz	16,0 GB	256 GB	Intel (R) HD Graphics 5500	Intel(R) Core(TM) i5-53000 CPU 2.30GHz	0,7s
Daffa	AMD A9 9425 RADEON R5	8,0 GB	1,2 TB	AMD RADEON (TM) R5 Graphics	AMD A9 9425 RADEON R5	0.12s
Dhevina	12th Gen Intel(R) Core(TM) i5 - 1235U 1.30 GHz	8,00 GB	152 GB	Inter(R) UHD Graphics	12th Gen Intel(R) Core(TM) i5 - 1235U	0.6s

C. Studi kasus

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 kondisikan 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

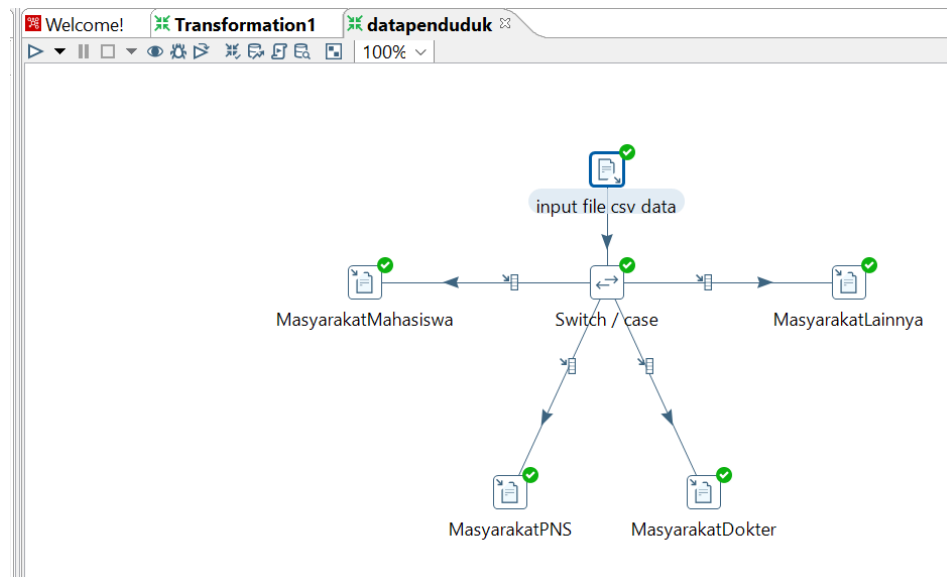
data-data tersebut akan digunakan untuk analisa di masing-masing wilayah dan juga Dashboard untuk emmbantu mengambil keputusan.

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

<https://github.com/dik4rizky/datasources/blob/82821b8ec341176f979956c84b79720d4aead012/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.



Switch / case

Step name
Switch / case

Field name to switch
Pekerjaan

Use string contains
☐

Case value data type
None

Case value conversion mask

Case value decimal symbol

Case value grouping symbol

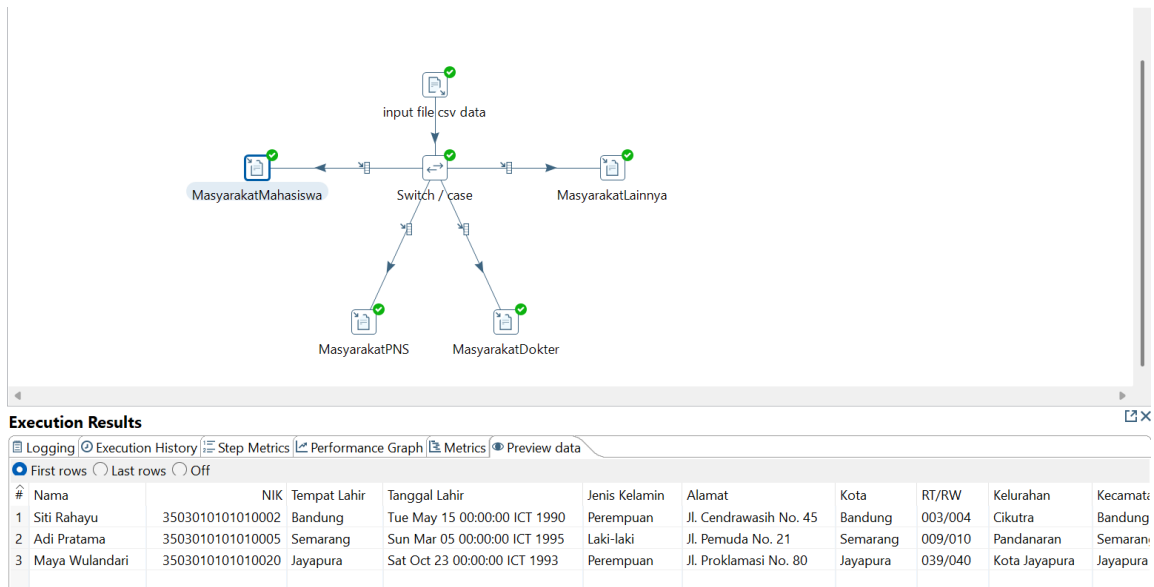
Case values

#	Value	Target step
1	Mahasiswa	MasyarakatMahasiswa
2	PNS	MasyarakatPNS
3	Dokter	MasyarakatDokter

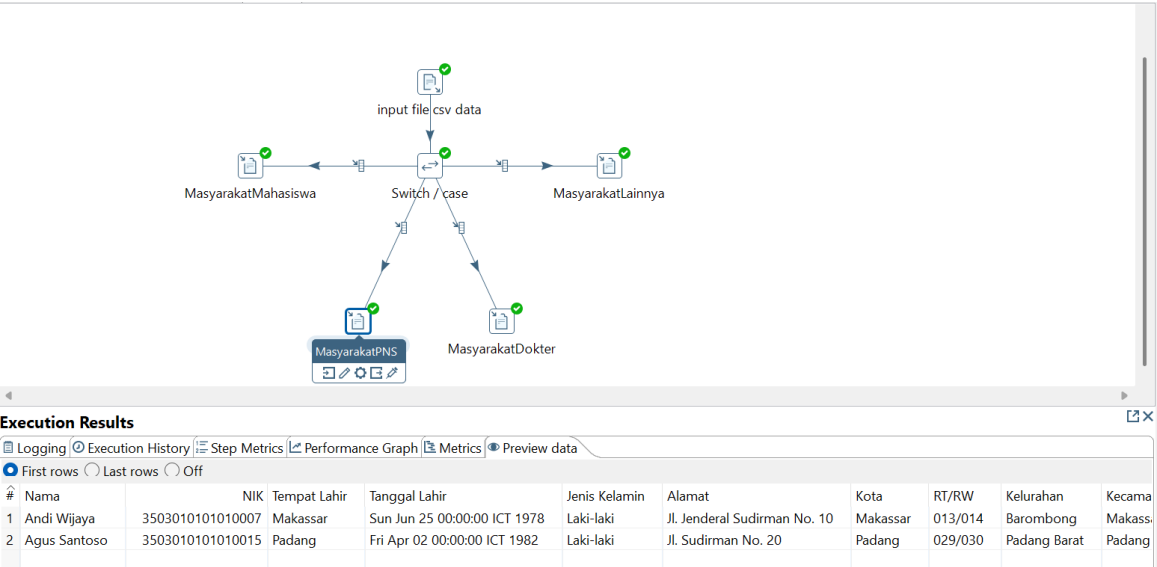
Default target step
MasyarakatLainnya

Help
OK
Cancel

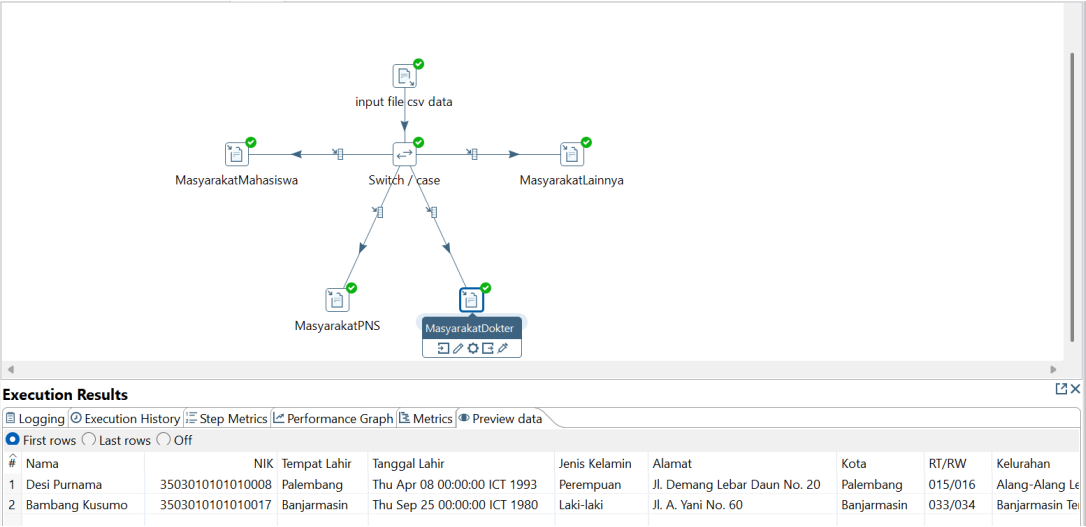
MAHASISWA



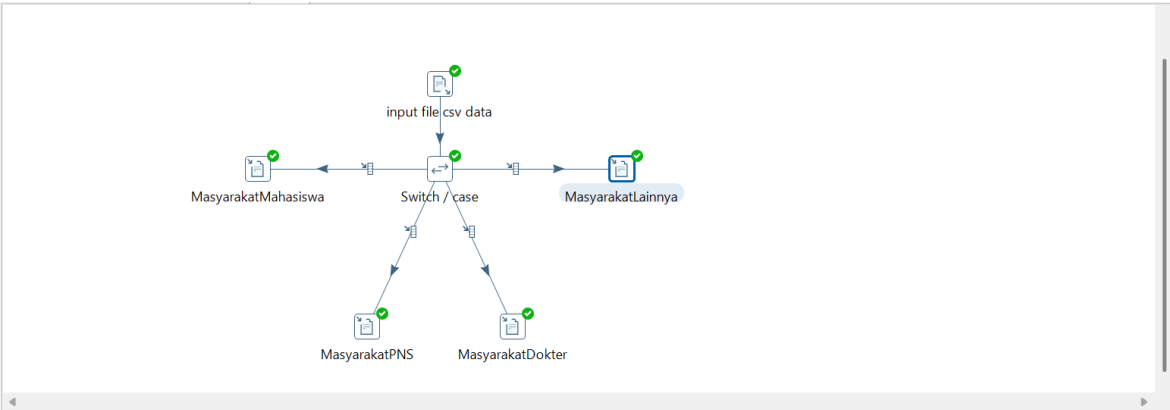
PNS



DOKTER



LAINNYA



Execution Results

LoggingExecution HistoryStep MetricsPerformance GraphMetricsPreview data

First rowsLast rowsOff

#	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan
1	Budi Santoso	3503010101010001	Jakarta	Tue Jan 01 00:00:00 ICT 1980	Laki-laki	Jl. Merdeka No. 123	Jakarta	001/002	Menteng
2	Joko Susilo	3503010101010003	Surabaya	Fri Oct 10 00:00:00 ICT 1975	Laki-laki	Jl. Diponegoro No. 67	Surabaya	005/006	Gubeng
3	Maya Dewi	3503010101010004	Yogyakarta	Sat Aug 20 00:00:00 ICT 1988	Perempuan	Jl. Malioboro No. 8	Yogyakarta	007/008	Sosromendur.
4	Rini Setiawan	3503010101010006	Malang	Sun Dec 12 00:00:00 ICT 1982	Perempuan	Jl. Simpang Borobudur No. 15	Malang	011/012	Dinoyo
5	Fajar Prabowo	3503010101010009	Medan	Tue Sep 30 00:00:00 ICT 1986	Laki-laki	Jl. Sudirman No. 50	Medan	017/018	Sei Sikambing
6	Rina Sari	3503010101010010	Denpasar	Fri Jul 17 00:00:00 ICT 1992	Perempuan	Jl. Gatot Subroto No. 30	Denpasar	019/020	Dauh Puri Kau
7	Ahmad Subhan	3503010101010011	Surakarta	Wed Jul 03 00:00:00 ICT 1991	Laki-laki	Jl. Pahlawan No. 55	Surakarta	021/022	Jebres