

**LAPORAN PRAKTIKUM DATA WAREHOUSE**  
**JOBSHEET 1: DATA WAREHOUSE SEDERHANA DAN**  
**PENGENALAN PDI SPOON**



**Disusun Oleh:**  
**Maharani Wirawan (2341760111)**  
**SIB 2B**

**SISTEM INFORMASI BISNIS**  
**JURUSAN TEKNOLOGI INFORMASI**  
**POLITEKNIK NEGERI MALANG**

**2025**

## TUGAS 1

Analisa lah data tersebut!

1. Berapa jumlah kolom pada data tersebut?

Jawaban: Data tersebut memiliki 7 kolom yaitu Customer, Product, Region, Date, Item unit price, No.Items, Total Sale.

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

Jawaban:

- 1) Customer: Nama pelanggan
- 2) Product: Jenis produk yang dibeli
- 3) Region: Wilayah penjualan (transaksi terjadi)
- 4) Date: Tanggal transaksi
- 5) Item unit price: Harga satuan per item
- 6) No.Items: Jumlah item yang dibeli
- 7) Total Sale: Total penjualan per transaksi

3. Adakah data yang memiliki nilai null / data yang tidak lengkap?

Jawaban: Ada, yaitu:

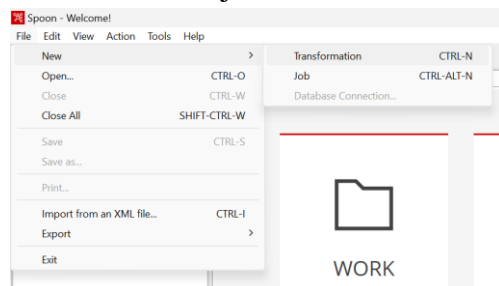
- Baris ke-4: Kolom Product kosong
- Baris ke-8: Kolom Region dan Date kosong
- Baris ke-13: Kolom Product, Region, dan Date kosong
- Baris ke-17: Kolom Product kosong
- Baris ke-19: Kolom Region kosong

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

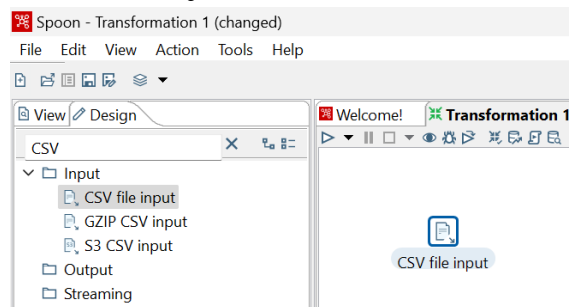
Jawaban:

### A. Pengambilan Data (Extract)

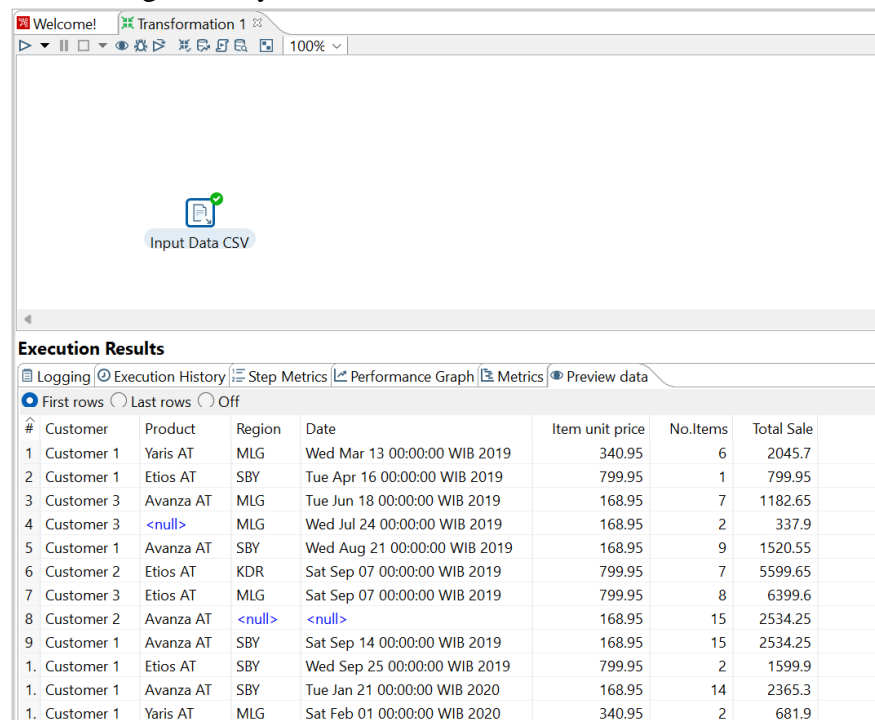
1. Buka lembar kerja Transformation melalui toolbar File.



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.



**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 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 3	<null>	MLG	Wed Jul 24 00:00:00 WIB 2019	168.95	2	337.9
5	Customer 1	Avanza AT	SBY	Wed Aug 21 00:00:00 WIB 2019	168.95	9	1520.55
6	Customer 2	Etios AT	KDR	Sat Sep 07 00:00:00 WIB 2019	799.95	7	5599.65
7	Customer 3	Etios AT	MLG	Sat Sep 07 00:00:00 WIB 2019	799.95	8	6399.6
8	Customer 2	Avanza AT	<null>	<null>	168.95	15	2534.25
9	Customer 1	Avanza AT	SBY	Sat Sep 14 00:00:00 WIB 2019	168.95	15	2534.25
1.	Customer 1	Etios AT	SBY	Wed Sep 25 00:00:00 WIB 2019	799.95	2	1599.9
1.	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

## TUGAS 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

Jawaban: 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?

Jawaban:

- Memulai proses eksekusi transformasi
- Menghubungkan ke sumber data
- Melakukan pembacaan data
- Menampilkan jumlah data yang berhasil diekstrak

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)

Jawaban:

- I (Input): Jumlah baris yang masuk ke dalam step tersebut = 20
- O (Output): Jumlah baris yang keluar dari step tersebut ke step berikutnya = 0
- R (Read): Jumlah baris yang dibaca dari sumber data = 0
- W (Written): Jumlah baris yang ditulis ke tujuan akhir (misalnya database atau file) =19
- U (Updated): Jumlah baris yang diperbarui dalam proses transformasi = 0
- E (Error): Jumlah baris yang mengalami kesalahan selama pemrosesan = 0

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
Maharani	AMD Athlon 300U with Radeon Vega Mobile Gfx	4,00 GB	202GB	AMD Radeon™ Vega 3 Graphics	AMD Athlon 300U with Radeon Vega Mobile Gfx	0.4s
Niriza	AMD Ryzen(TM) 3 7320U	8.00GB	1 TB	AMD Radeon Graphics	AMD Ryzen™ 3 7320U	0.1s
Dhevina	12th Gen	8,00 GB	152GB	Inter(R)	12 <sup>th</sup> Gen	0.0s

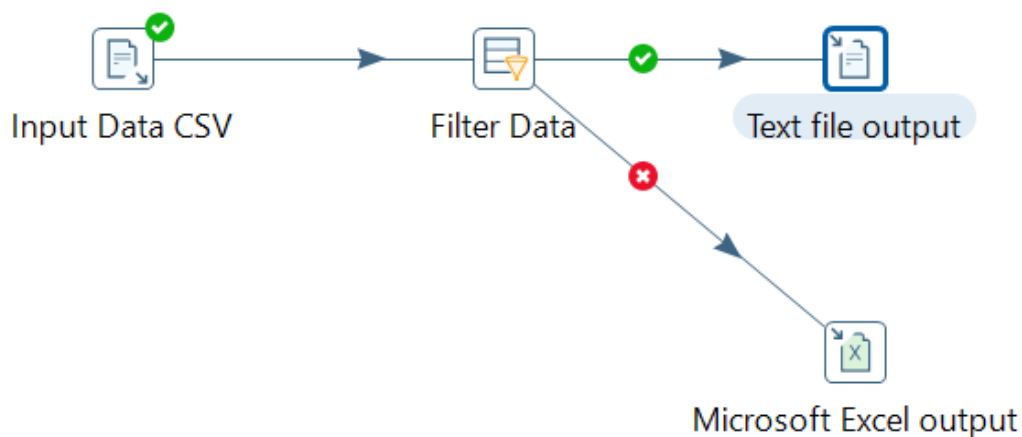
	Intel(R) Core(TM) i5 - 1235U 1.30 GHz			UHD Graphics	Intel(R) Core(TM) i5- 1235U	
Nimas	AMD Ryzen 3 5300U	16,0GB	475 GB	AMD Radeon Graphics	AMD Ryzen 3 5300U with Radeon Graphics	0.1s

## 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.
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.

Text file output									
Step name Text file output									
File Content Fields									
#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim Typ
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	IDR	.	,	none
6	No.Items	Integer	#	15	0	IDR	.	,	none
7	Total Sale	Number	#,.	7	2	IDR	.	,	none

20. Tekan ok untuk kembali ke work area
21. Hal yang sama dilakukan pada konfigurasi Microsoft Excel output, Doubleclick 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 screenshot shows the Spoon transformation interface. The workflow diagram at the top illustrates the process: 'Input Data CSV' (green checkmark) connects to 'Filter Data' (green checkmark), which then branches to 'Text file output' (green checkmark) and 'Microsoft Excel output' (green checkmark). A red 'X' mark is visible on the arrow connecting 'Filter Data' to 'Microsoft Excel output', indicating a failure in that path.

**Execution Results**

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2025/02/26 22:39:12 - Spoon - Transformation opened.  
 2025/02/26 22:39:12 - Spoon - Launching transformation [Transformation 1]...  
 2025/02/26 22:39:12 - Spoon - Started the transformation execution.  
 2025/02/26 22:39:13 - Transformation 1 - Dispatching started for transformation [Transformation 1]  
 2025/02/26 22:39:13 - Input Data CSV.0 - Header row skipped in file 'C:\Users\user\Downloads\CarDistributionSales.csv'  
 2025/02/26 22:39:13 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)  
 2025/02/26 22:39:13 - Filter Data.0 - Finished processing (I=0, O=0, R=19, W=19, U=0, E=0)  
 2025/02/26 22:39:13 - Text file output.0 - Finished processing (I=0, O=15, R=14, W=14, U=0, E=0)  
 2025/02/26 22:39:14 - Microsoft Excel output.0 - Finished processing (I=0, O=5, R=5, W=5, U=0, E=0)  
 2025/02/26 22:39:14 - Spoon - The transformation has finished!!

*Proses berhasil dijalankan*

**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 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

*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 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

*outputSalesCarDistribution*

## TUGAS 3

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

Jawaban: File CSV berisi data mentah yang diekstrak sebelum mengalami proses transformasi, sedangkan file Excel sudah melewati tahap pembersihan sehingga datanya lebih terstruktur dan siap digunakan.

2. Jelaskan apa yang terjadi pada proses Filter rows!

Jawaban:

- 1) Input data: Data dari sumber (seperti file) dimuat ke dalam PDI.
- 2) Menentukan kondisi penyaringan: Filter Rows berfungsi untuk menentukan aturan atau syarat yang harus dipenuhi oleh data agar dapat lanjut ke proses berikutnya.
- 3) Pemrosesan data: Data yang sesuai dengan kriteria yang ditentukan akan diproses lebih lanjut dalam alur transformasi ETL.
- 4) Output data: Output dari proses ini adalah sekumpulan data yang telah disaring, lebih bersih, dan relevan untuk keperluan analisis atau tahap pemrosesan selanjutnya.

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	Waktu Eksekusi
Maharani	AMD Athlon 300U with Radeon Vega Mobile Gfx	4,00 GB	202GB	AMD Radeon™ Vega 3 Graphics	AMD Athlon 300U with Radeon Vega Mobile Gfx	0.2s
Niriza	AMD Ryzen(TM) 3 7320U	8.00GB	1 TB	AMD Radeon Graphics	AMD Ryzen™ 3 7320U	0.3s



Dhevina	12 <sup>th</sup> Gen Intel(R) Core(TM) i5 - 1235U 1.30 GHz	8,00 GB	152GB	Inter(R) UHD Graphics	12 <sup>th</sup> Gen Intel(R) Core(TM) i5- 1235U	0.6s
Nimas	AMD Ryzen 3 5300U	16,0GB	475 GB	AMD Radeon Graphics	AMD Ryzen 3 5300U with Radeon Graphics	0.4s

### 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 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 membantu mengambil keputusan.

Pak Pascanowo mendapatkan data penduduk yang akan pindah ke wilayahnya melalui

link:

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

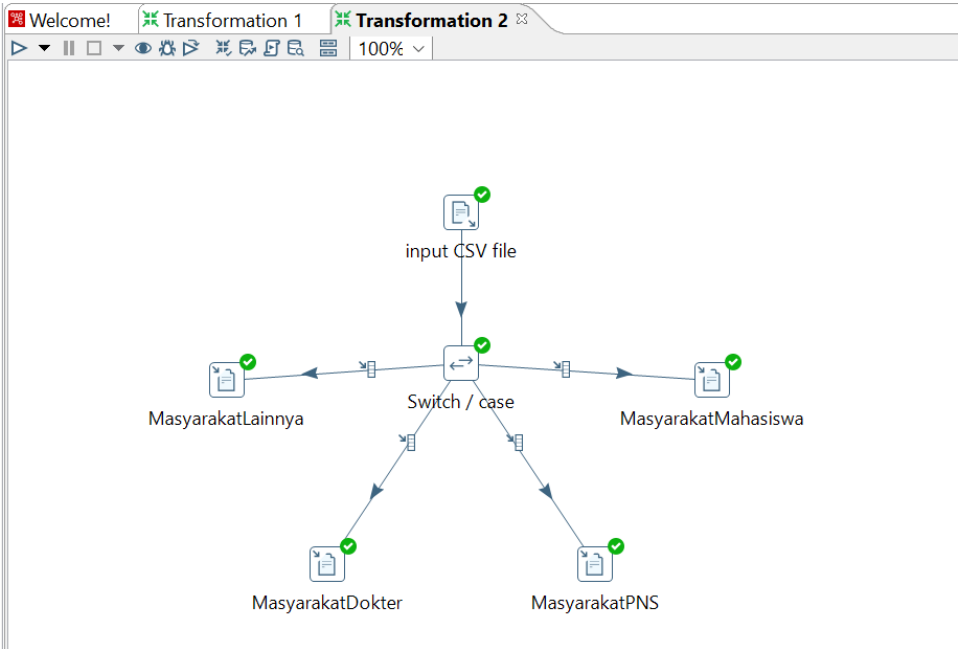
Bantulah Pak Pascanowo untuk memisahkan data tersebut menjadi 4 file:

**MasyarakatMahasiswa,MasyarakatPNS,MasyarakatDokter,MasyarakatL**

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

Jawaban:

Sistem:



Hasil run:

Execution Results	
Logging	Execution History
2025/02/27 06:10:05 - Spoon - Transformation opened.	
2025/02/27 06:10:05 - Spoon - Launching transformation [Transformation 2]...	
2025/02/27 06:10:05 - Spoon - Started the transformation execution.	
2025/02/27 06:10:05 - Transformation 2 - Dispatching started for transformation [Transformation 2]	
2025/02/27 06:10:05 - input CSV file.0 - Header row skipped in file 'C:\Users\user\Downloads\dataDummyKependudukan.csv'	
2025/02/27 06:10:05 - input CSV file.0 - Finished processing (I=21, O=0, R=0, W=20, U=0, E=0)	
2025/02/27 06:10:05 - Switch / case.0 - Finished processing (I=0, O=0, R=20, W=20, U=0, E=0)	
2025/02/27 06:10:05 - MasyarakatMahasiswa.0 - Finished processing (I=0, O=4, R=3, W=3, U=0, E=0)	
2025/02/27 06:10:05 - MasyarakatPNS.0 - Finished processing (I=0, O=3, R=2, W=2, U=0, E=0)	
2025/02/27 06:10:05 - MasyarakatDokter.0 - Finished processing (I=0, O=3, R=2, W=2, U=0, E=0)	
2025/02/27 06:10:05 - MasyarakatLainnya.0 - Finished processing (I=0, O=14, R=13, W=13, U=0, E=0)	
2025/02/27 06:10:05 - Spoon - The transformation has finished!!	

1. Masyarakat Profesi 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
1	Siti Rahayu	3503010101010002	Bandung	Tue May 15 00:00:00 WIB 1990	Perempuan	Jl. Cendrawasih No. 45	Bandung
2	Adi Pratama	3503010101010005	Semarang	Sun Mar 05 00:00:00 WIB 1995	Laki-laki	Jl. Pemuda No. 21	Semarang
3	Maya Wulandari	3503010101010020	Jayapura	Sat Oct 23 00:00:00 WIB 1993	Perempuan	Jl. Proklamasi No. 80	Jayapura
4							

2. Masyarakat Profesi 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
1	Andi Wijaya	3503010101010007	Makassar	Sun Jun 25 00:00:00 WIB 1978	Laki-laki	Jl. Jenderal Sudirman No. 10	Makassar
2	Agus Santoso	3503010101010015	Padang	Fri Apr 02 00:00:00 WIB 1982	Laki-laki	Jl. Sudirman No. 20	Padang

3. Masyarakat Profesi 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
1	Desi Purnama	3503010101010008	Palembang	Thu Apr 08 00:00:00 WIB 1993	Perempuan	Jl. Demang Lebar Daun No. 20	Palembang
2	Bambang Kusumo	3503010101010017	Banjarmasin	Thu Sep 25 00:00:00 WIB 1980	Laki-laki	Jl. A. Yani No. 60	Banjarmasi

4. Masyarakat Profesi 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
1	Budi Santoso	3503010101010001	Jakarta	Tue Jan 01 00:00:00 WIB 1980	Laki-laki	Jl. Merdeka No. 123	Jakarta
2	Joko Susilo	3503010101010003	Surabaya	Fri Oct 10 00:00:00 WIB 1975	Laki-laki	Jl. Diponegoro No. 67	Surabaya
3	Maya Dewi	3503010101010004	Yogyakarta	Sat Aug 20 00:00:00 WIB 1988	Perempuan	Jl. Malioboro No. 8	Yogyakarta
4	Rini Setiawan	3503010101010006	Malang	Sun Dec 12 00:00:00 WIB 1982	Perempuan	Jl. Simpang Borobudur No. 15	Malang
5	Fajar Prabowo	3503010101010009	Medan	Tue Sep 30 00:00:00 WIB 1986	Laki-laki	Jl. Sudirman No. 50	Medan
6	Rina Sari	3503010101010010	Denpasar	Fri Jul 17 00:00:00 WIB 1992	Perempuan	Jl. Gatot Subroto No. 30	Denpasar
7	Ahmad Subhan	3503010101010011	Surakarta	Wed Jul 03 00:00:00 WIB 1991	Laki-laki	Jl. Pahlawan No. 55	Surakarta
8	Dian Pertiwi	3503010101010012	Maluku	Tue Dec 10 00:00:00 WIB 1985	Perempuan	Jl. Pattimura No. 10	Ambon
9	Rudi Hartono	3503010101010013	Pontianak	Mon May 28 00:00:00 WIB 1973	Laki-laki	Jl. Ahmad Yani No. 30	Pontianak
1.	Lina Fitriani	3503010101010014	Manado	Mon Aug 15 00:00:00 WIB 1994	Perempuan	Jl. Wolter Monginsidi No. 17	Manado
1.	Rina Agustina	3503010101010016	Bengkulu	Mon Nov 19 00:00:00 WIB 1990	Perempuan	Jl. Imam Bonjol No. 45	Bengkulu
1.	Nia Dewi Lestari	3503010101010018	Manokwari	Sun Jun 14 00:00:00 WIB 1987	Perempuan	Jl. Diponegoro No. 40	Manokwari
1.	Doni Setiawan	3503010101010019	Palu	Mon Mar 07 00:00:00 WIB 1977	Laki-laki	Jl. Soekarno-Hatta No. 70	Palu