

DATA WAREHOUSE  
PENTAHO DATA INTEGRATION  
JOBSHEET 1

Nama : Muhammad Reishi Fauzi  
Kelas : SIB-2A  
No : 23

<https://github.com/dik4rizky/datasources/blob/7e3f70f516a1b107635141d09862a8c30fbb6812/CarDistributionSales.csv>

Files

7e3f70f

Go to file

CarDistributionSales.csv

README.md

dhyk a

Preview Code Blame 20 lines (20 loc) · 982 Bytes

Search this file

	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
1							
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?
2. Apa arti atau isi data dari setiap kolom yang ada?
3. Adakah data yang memiliki nilai null / data yang tidak lengkap?
4. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

Jawab

1. Terdapat 7 kolom yang terdiri dari kolom Customer, Product, Region, Date, Items Unit Price, No. Items, Total Sale
2. Arti tiap kolom
  - a. Customer  
siapa pelanggan yang melakukan pembelian
  - b. Product  
nama mobil dan jenis transmisi yang dibeli
  - c. Region  
Nama Wilayah atau kota transaksi tersebut dilakukan
  - d. Date  
Tanggal transaksi dilakukan
  - e. Item Unit Price  
Harga satu unit mobil
  - f. No. Item  
jumlah unit mobil mobil yang dibeli dalam satu transaksi
  - g. Total Sale  
Total harga transaksi (Items Unit Price x No. Items)
3. Ya, ada yang memiliki nilai null  
yaitu pada baris ke-4 terdapat nilai null pada kolom product, lalu pada baris ke-8 terdapat nilai null pada kolom region dan date, kemudian pada baris ke-13 terdapat nilai null pada kolom product, region, date, setelah itu ada di baris ke-17 terdapat nilai null pada kolom product, dan yang terakhir pada baris ke-19 ada nilai null pada kolom region.
4. tidak ada data yang memiliki tipe data berbeda, adanya yang bernilai null.

## TUGAS 2

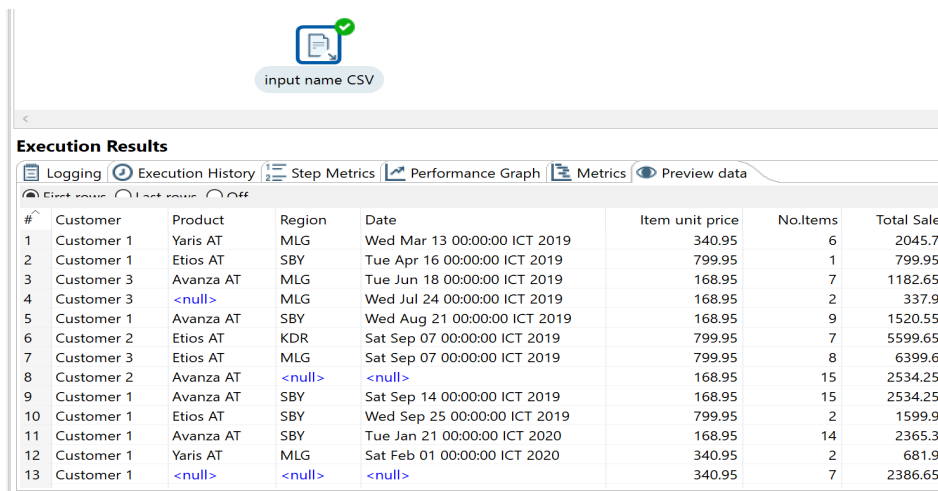
1. Apakah data hasil eksekusi sesuai dengan data aslinya?
2. PDI Spoon melakukan proses extract, perhatikan pada **Tab Logging di Execution Results Area**, langkah-langkah apa sajakah yang dilakukan PDI Spoon untuk melakukan extract data?
3. Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E ?

2024/02/17 06:51:48 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)

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.

Jawab

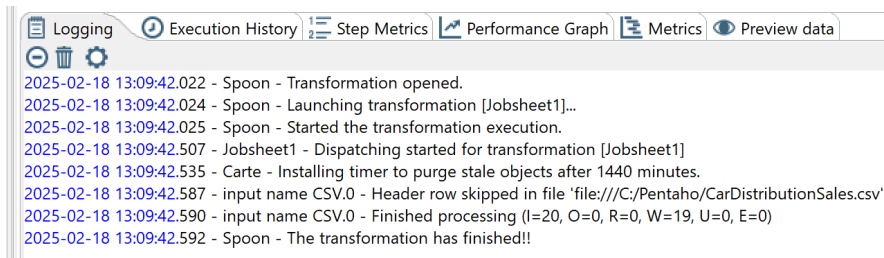
1. Ya, sudah sesuai



The screenshot shows the PDI Spoon interface. At the top, there is a button labeled 'input name CSV' with a green checkmark. Below it, the 'Execution Results' section is visible, showing a table with 13 rows of data. The table has columns for #, Customer, Product, Region, Date, Item unit price, No.Items, and Total Sale.

#	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 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	Etios AT	KDR	Sat Sep 07 00:00:00 ICT 2019	799.95	7	5599.65
7	Customer 3	Etios AT	MLG	Sat Sep 07 00:00:00 ICT 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 ICT 2019	168.95	15	2534.25
10	Customer 1	Etios AT	SBY	Wed Sep 25 00:00:00 ICT 2019	799.95	2	1599.9
11	Customer 1	Avanza AT	SBY	Tue Jan 21 00:00:00 ICT 2020	168.95	14	2365.3
12	Customer 1	Yaris AT	MLG	Sat Feb 01 00:00:00 ICT 2020	340.95	2	681.9
13	Customer 1	<null>	<null>	<null>	340.95	7	2386.65

2. tab logging



The screenshot shows the 'Logging' tab in the PDI Spoon interface. It displays a list of logs with timestamps and descriptions of the execution process.

Timestamp	Log Message
2025-02-18 13:09:42.022	- Spoon - Transformation opened.
2025-02-18 13:09:42.024	- Spoon - Launching transformation [Jobsheet1]...
2025-02-18 13:09:42.025	- Spoon - Started the transformation execution.
2025-02-18 13:09:42.507	- Jobsheet1 - Dispatching started for transformation [Jobsheet1]
2025-02-18 13:09:42.535	- Carte - Installing timer to purge stale objects after 1440 minutes.
2025-02-18 13:09:42.587	- input name CSV.0 - Header row skipped in file 'file:///C:/Pentaho/CarDistributionSales.csv'
2025-02-18 13:09:42.590	- input name CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)
2025-02-18 13:09:42.592	- Spoon - The transformation has finished!!

Proses dimulai dengan membuka dan menjalankan transformasi bernama Jobsheet1. Setelah transformasi dijalankan, sistem melakukan dispatching untuk mengelola eksekusi transformasi tersebut.

Sebagai bagian dari pengelolaan memori, PDI Spoon juga mengatur timer untuk membersihkan objek yang tidak terpakai setelah 1440 menit (24 jam). Hal ini berguna untuk memastikan sistem tetap optimal dalam menangani berbagai transformasi yang dijalankan. Selanjutnya, proses extract data dari file CSV dilakukan. Dalam tahap ini, Spoon membaca file CarDistributionSales.csv yang tersimpan di direktori C:/Pentaho/. Saat membaca file tersebut, sistem melewati baris header agar tidak diproses sebagai data utama.

3. Maksud dari adalah I, O, R, W, U, E adalah data yang di ambil dari file CSV dengan penjelasan sebagai berikut

- I (Input) = 20 → Jumlah baris yang dibaca.
- O (Output) = 0 → Tidak ada baris yang dikirim ke output.
- R (Rejected) = 0 → Tidak ada baris yang ditolak.
- W (Written) = 19 → 19 baris data berhasil ditulis ke output berikutnya.
- U (Updated) = 0 → Tidak ada baris yang diperbarui.
- E (Error) = 0 → Tidak ada kesalahan dalam proses

4. Perbandingan waktu ekstraksi

<https://docs.google.com/spreadsheets/d/1Qj3RbUFPrVZ8HGn5SqqJE36KSQmuldRyp8x7NaATEYs/edit?gid=0#gid=0>

NAMA	WAKTU EKSEKUSI	PROCCESOR	RAM	STORAGE	VGA	CPU
Muhammad Reishi Fauzi	1,1 detik	Intel Core i7-7820HQ	12GB	256 GB	NVIDIA GEFORCE 940MX	Intel core i7-7820HQ
Dahnir Davina	1 detik	Intel Core i5-6300U	8 GB	256 GB	Intel(R) HD Graphics 520	Intel Core i5-6300U
Yonanda Mayla	1 detik	Intel Core i5-12450H	16GB	512GB	NVIDIA GEFORCE RTX 3050	Intel Core i5-12450H
Hamdan Ubaidillah	1 detik	AMD Ryzen 5300U	12 GB	756 GB (500+256)	Radeon Ati Lucienne	AMD Ryzen 5300U

### TUGAS 3

1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?
2. Jelaskan apa yang terjadi pada proses **Filter rows**!
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

Jawab

1. Pada file csv, data yang masuk adalah data lengkap tidak terdapat nilai null.  
Sedangkan pada file excel data yang masuk adalah data yang tidak lengkap yang mengandung nilai null
2. pada filter rows, terdapat pilihan “send true data to step: Text File Output” dimana jika hasil dari filter yang kita buat is not null terpenuhi semua akan menjadi true dan masuk ke text file output, sedangkan jika ada kondisi yang tidak terpenuhi akan menjadi false dan masuk ke excel
3. Perbandingan waktu

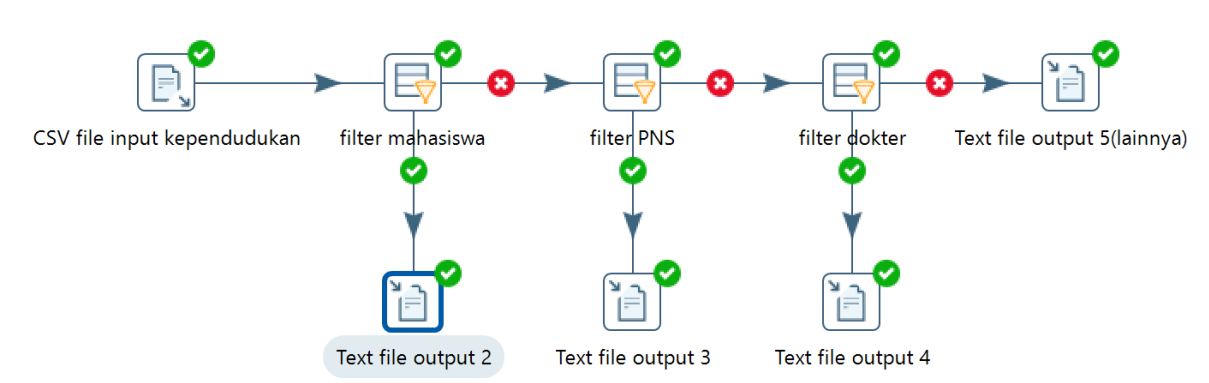
NAMA	WAKTU EKSEKUSI	PROCCESOR	RAM	STORAGE	VGA	CPU
Muhammad Reishi Fauzi	2 detik	Intel Core i7-7820HQ	12GB	256 GB	NVIDIA GEFORCE 940MX	Intel core i7-7820HQ
Dahnir Davina	2 detik	Intel Core i5-6300U	8 GB	256 GB	Intel(R) HD Graphics 520	Intel Core i5-6300U
Yonanda Mayla	2 detik	Intel Core i5-12450H	16GB	512GB	NVIDIA GEFORCE RTX 3050	Intel Core i5-12450H
Hamdan Ubaidillah	2 detik	AMD Ryzen 5300U	12 GB	756 GB (500+256)	Radeon Ati Lucienne	AMD Ryzen 5300U

## STUDI KASUS

Pak Pascanowo mendapatkan data penduduk yang akan pindah ke wilayahnya melalui link:  
<https://github.com/dik4rizky/datasources/blob/82821b8ec341176f979956c84b79720d4aeadd012/dataDummyKependudukan.csv>

Bantulah Pak Pascanowo untuk memisahkan 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



**Execution Results**

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2025-02-25 12:52:02.654 - Spoon - Transformation opened.  
2025-02-25 12:52:02.657 - Spoon - Launching transformation [Jobsheet1]...  
2025-02-25 12:52:02.657 - Spoon - Started the transformation execution.  
2025-02-25 12:52:03.149 - Jobsheet1 - Dispatching started for transformation [Jobsheet1]  
2025-02-25 12:52:03.192 - Carte - Installing timer to purge stale objects after 1440 minutes.  
2025-02-25 12:52:03.461 - input name CSV.0 - Header row skipped in file 'file:///C:/Pentaho/CarDistributionSales.csv'  
2025-02-25 12:52:03.461 - CSV file input kependudukan.0 - Header row skipped in file 'file:///C:/Pentaho/dataDummyKependudukan.csv'  
2025-02-25 12:52:03.464 - input name CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)  
2025-02-25 12:52:03.466 - CSV file input kependudukan.0 - Finished processing (I=21, O=0, R=0, W=20, U=0, E=0)  
2025-02-25 12:52:03.472 - Filter data.0 - Finished processing (I=0, O=0, R=19, W=19, U=0, E=0)  
2025-02-25 12:52:03.472 - filter mahasiswa.0 - Finished processing (I=0, O=0, R=20, W=20, U=0, E=0)  
2025-02-25 12:52:03.476 - Text file output 2.0 - Finished processing (I=0, O=4, R=3, W=3, U=0, E=0)  
2025-02-25 12:52:03.481 - filter PNS.0 - Finished processing (I=0, O=0, R=17, W=17, U=0, E=0)  
2025-02-25 12:52:03.481 - Text file output.0 - Finished processing (I=0, O=15, R=14, W=14, U=0, E=0)  
2025-02-25 12:52:03.485 - filter dokter.0 - Finished processing (I=0, O=0, R=15, W=15, U=0, E=0)  
2025-02-25 12:52:03.485 - Text file output 3.0 - Finished processing (I=0, O=3, R=2, W=2, U=0, E=0)  
2025-02-25 12:52:03.489 - Text file output 4.0 - Finished processing (I=0, O=3, R=2, W=2, U=0, E=0)  
2025-02-25 12:52:03.491 - Text file output 5(lainnya).0 - Finished processing (I=0, O=14, R=13, W=13, U=0, E=0)  
2025-02-25 12:52:03.587 - Microsoft Excel output (deprecated).0 - Finished processing (I=0, O=5, R=5, W=5, U=0, E=0)  
2025-02-25 12:52:03.596 - Spoon - The transformation has finished!!

- Masyarakat Mahasiswa

**Execution Results**

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First row Last row Off

	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
90	Perempuan	Jl. Cendrawasih No. 45	Bandung	003/004	Cikutra	Bandung Wetan	Kristen	Belum Menikah	Mahasiswa
95	Laki-laki	Jl. Pemuda No. 21	Semarang	009/010	Pandanaran	Semarang Tengah	Buddha	Belum Menikah	Mahasiswa
3	Perempuan	Jl. Proklamasi No. 80	Jayapura	039/040	Kota Jayapura	Jayapura	Kristen	Belum Menikah	Mahasiswa

- Masyarakat PNS

Execution Results									
Logging Execution History Step Metrics Performance Graph Metrics Preview data									
8	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
	Laki-laki	Jl. Jenderal Sudirman No. 10	Makassar	013/014	Barombong	Makassar Selatan	Islam	Menikah	PNS
	Laki-laki	Jl. Sudirman No. 20	Padang	029/030	Padang Barat	Padang	Islam	Menikah	PNS

- Masyarakat Dokter

Execution Results

Logging

Execution History

Step Metrics

Performance Graph

Metrics

Preview data

First row

Last row

Off

elamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
puan	Jl. Demang Lebar Daun No. 20	Palembang	015/016	Alang-Alang Lebar	Palembang Selatan	Kristen	Belum Menikah	Dokter
ki	Jl. A. Yani No. 60	Banjarmasin	033/034	Banjarmasin Tengah	Banjarmasin	Hindu	Menikah	Dokter

- Masyarakat Lainnya

Execution Results

Logging

Execution History

Step Metrics

Performance Graph

Metrics

Preview data

First row

Last row

Off

elamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
ki	Jl. Merdeka No. 123	Jakarta	001/002	Menteng	Jakarta Pusat	Islam	Menikah	Wiraswasta
ki	Jl. Diponegoro No. 67	Surabaya	005/006	Gubeng	Surabaya Pusat	Katolik	Menikah	Pegawai Swasta
uan	Jl. Malioboro No. 8	Yogyakarta	007/008	Sosromenduran	Yogyakarta Utara	Hindu	Belum Menikah	Guru
uan	Jl. Simpang Borobudur No. 15	Malang	011/012	Dinoyo	Malang Selatan	Kristen	Menikah	Wiraswasta
ki	Jl. Sudirman No. 50	Medan	017/018	Sei Sikambing	Medan Timur	Hindu	Menikah	Wiraswasta
uan	Jl. Gatot Subroto No. 30	Denpasar	019/020	Dauh Puri Kauh	Denpasar Utara	Buddha	Belum Menikah	Freelancer
ki	Jl. Pahlawan No. 55	Surakarta	021/022	Jebres	Surakarta Timur	Islam	Belum Menikah	Wiraswasta
uan	Jl. Pattimura No. 10	Ambon	023/024	Baguala	Ambon	Kristen	Menikah	Guru
ki	Jl. Ahmad Yani No. 30	Pontianak	025/026	Pontianak Barat	Pontianak Kota	Hindu	Menikah	Wiraswasta
uan	Jl. Wolter Monginsidi No. 17	Manado	027/028	Mapanget	Manado Selatan	Kristen	Belum Menikah	Akuntan
uan	Jl. Imam Bonjol No. 45	Bengkulu	031/032	Ratu Samban	Bengkulu Tengah	Kristen	Menikah	Wiraswasta
uan	Jl. Diponegoro No. 40	Manokwari	035/036	Manokwari Barat	Manokwari	Kristen	Belum Menikah	Penulis
ki	Jl. Soekarno-Hatta No. 70	Palu	037/038	Palu Barat	Palu	Islam	Menikah	Wiraswasta