

LAPORAN HASIL PRAKTIKUM

DATA WAREHOUSE

JOB SHEET 1



Oleh:

DZULFIKAR MUHAMMAD AL GHIFARI

NIM. 2341760071

SIB-2B / 08

D-IV SISTEM INFORMASI BISNIS

JURUSAN TEKNOLOGI INFORMASI

POLITEKNIK NEGERI MALANG

Tugas 1

1. Berapa jumlah kolom pada data tersebut?

7 Kolom

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

Data report penjualan Sales Astro Mobil dari berbagai cabang

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

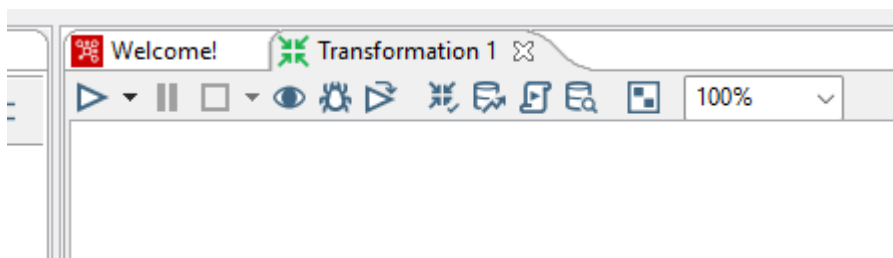
Ada, terdapat data yang memiliki nilai null

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

Tidak ada

A. Pengambilan Data (Extract)

a. membuka workspace transformation



b. import data csv

CSV file input

Step name: CSV file input

Filename: D:\1POLINEMA\4Genap 2024-2025\DW\pertemuan1\CarDistributionSales. 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 | Tri |
|---|-----------------|---------|------------|--------|-----------|----------|---------|-------|-----|
| 1 | Customer | String | | 10 | | IDR | . | , | no |
| 2 | Product | String | | 9 | | IDR | . | , | no |
| 3 | Region | String | | 3 | | IDR | . | , | no |
| 4 | Date | Date | MM/dd/yyyy | | | IDR | . | , | no |
| 5 | Item unit price | Number | #,. | 6 | 2 | IDR | . | , | no |
| 6 | No.Items | Integer | # | 15 | 0 | IDR | . | , | no |
| 7 | Total Sale | Number | #,. | 7 | 2 | IDR | . | , | no |

Help OK Get Fields Preview Cancel

c. run transformation

Run Options

Run configurations:

Pentaho local

Options

☒ Clear log before running

Log level: Basic

☐ Enable safe mode

☒ Gather performance metrics

Parameters Variables

| Parameter | Default value | Value | Description |
|-----------|---------------|-------|-------------|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

Arguments (legacy)

☒ Always show dialog on run

Help

Run Cancel

d. preview data execution

Execution Results

Logging

Execution History

Step Metrics

Performance Graph

Metrics

Preview data

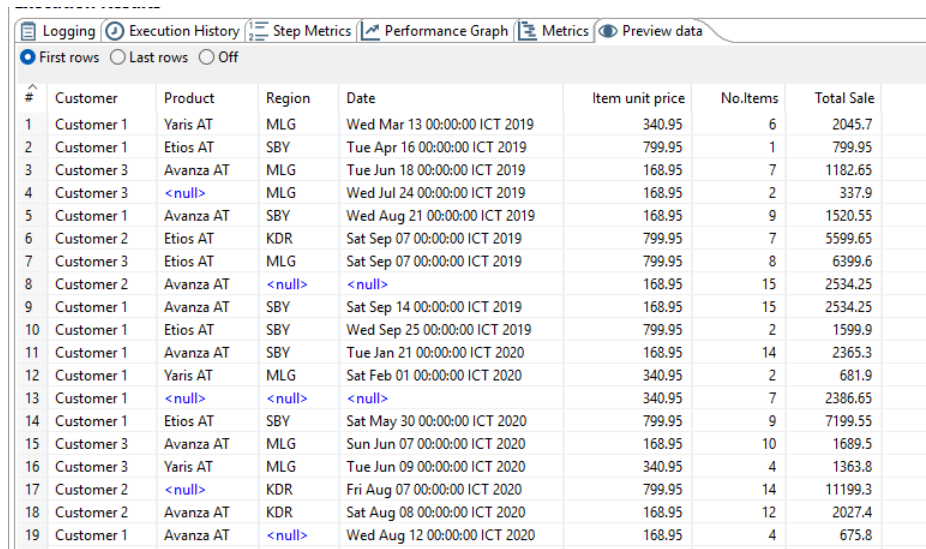
☒ First rows
 ☐ Last rows
 ☐ Off

| # | 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 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 | SRV | Wed Sep 25 00:00:00 ICT 2019 | 799.95 | 2 | 1599.9 |

TUGAS 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

Ya, sama



| # | 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 |
| 14 | Customer 1 | Etios AT | SBY | Sat May 30 00:00:00 ICT 2020 | 799.95 | 9 | 7199.55 |
| 15 | Customer 3 | Avanza AT | MLG | Sun Jun 07 00:00:00 ICT 2020 | 168.95 | 10 | 1689.5 |
| 16 | Customer 3 | Yaris AT | MLG | Tue Jun 09 00:00:00 ICT 2020 | 340.95 | 4 | 1363.8 |
| 17 | Customer 2 | <null> | KDR | Fri Aug 07 00:00:00 ICT 2020 | 799.95 | 14 | 11199.3 |
| 18 | Customer 2 | Avanza AT | KDR | Sat Aug 08 00:00:00 ICT 2020 | 168.95 | 12 | 2027.4 |
| 19 | Customer 1 | Avanza AT | <null> | Wed Aug 12 00:00:00 ICT 2020 | 168.95 | 4 | 675.8 |

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 yang dilakukan PDI Spoon pada saat extract data:

1. Transformation opened.
2. Launching transformation
3. Started the transformation execution.
4. Dispatching started for transformation
5. Input Data CSV.0 - Header row skipped
6. Input Data CSV.0 - Finished processing

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) = Jumlah data yang diinput

O (Output) = Jumlah data yang berhasil dikeluarkan

R (Rejected) = Jumlah data yang ditolak atau gagal diproses

W (Written) = Jumlah data yang berhasil ditulis ke output

U (Updated) = Jumlah data yang diperbarui dalam sistem

E (Errors) = Jumlah kesalahan yang terjadi selama pemrosesan

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.

Time : 14ms

RAM : 16 GB (1x16)

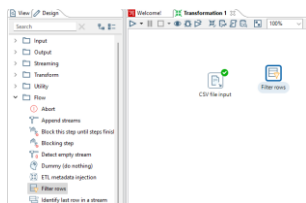
Processor/CPU : i5 Gen 11

VGA : Intel Iris X

Storage : 512GB Nvme

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

- a. add filter rows to workspace



- b. connect csv to filter



- c. add text file output



- d. add Microsoft excel output



- e. connect filter as true for text file output, and as false for microsoft excel output



f. set filter

Filter rows

Step name: Filter Data

Send 'true' data to step: Text file output

Send 'false' data to step: Microsoft Excel output

The condition:

Customer IS NOT NULL

AND

Product IS NOT NULL

AND

Region IS NOT NULL

OK Cancel

g. set path file text output and get field

Text file output

Step name: Text file output

File Content Fields

| # | Name | Type | Format | Length | Precision | Currency | Decimal | Group | Trim | Type | Null |
|---|-----------------|---------|------------|--------|-----------|----------|---------|-------|------|------|------|
| 1 | Customer | String | | 10 | | | | | | none | |
| 2 | Product | String | | 9 | | | | | | none | |
| 3 | Region | String | | 3 | | | | | | none | |
| 4 | Date | String | 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 | |

Get Fields Minimal width

h. set path excel output and get field

Microsoft Excel output

Step name: Microsoft Excel output

File Content Custom Fields

Filename: D:/POLINEMA/4Genap 2024-2025/DWI

Create Parent folder: ☐

Do not create file at start: ☐

Accept file names from field: ☐

File name field:

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: ☒

OK Cancel

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 |

i. run work area

Run Options

Run configuration: Pentaho local

Options

☒ Clear log before running

☐ Enable safe mode

☒ Gather performance metrics

Log level: Basic

Parameters Variables

| Parameter | Default value | Value | Description |
|-----------|---------------|-------|-------------|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

Arguments (legacy)

☒ Always show dialog on run

Help Run Cancel

Result text file

| 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,Elios 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,Elios AT,KDR,9/7/2019,799.95,7,5599.65 |
| 8 | Customer 3,Elios 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,Elios AT,SBY,9/25/2019,799.95,2,1599.9 |
| 12 | Customer 1,Avanza AT,SBY,11/21/2020,168.95,14,2365.3 |
| 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,Elios 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 |
| 21 | |
| 22 | |

Result xls file

| | A | B | C | D | E | F | G | H |
|---|----------|-----------|--------|-------|-------------|----------|------------|---|
| 1 | Customer | Product | Region | Date | Item unit p | No.Items | Total Sale | |
| 2 | Customer | | MLG | ##### | 0168.95 | 2.00 | 00337.90 | |
| 3 | Customer | Avanza AT | | | 0168.95 | 15.00 | 02534.25 | |
| 4 | Customer | | | | 0340.95 | 7.00 | 02386.65 | |
| 5 | Customer | | KDR | ##### | 0799.95 | 14.00 | 11199.30 | |
| 6 | Customer | Avanza AT | | ##### | 0168.95 | 4.00 | 00675.80 | |
| 7 | | | | | | | | |

TUGAS 3

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

File csv berisikan data lengkap, sedangkan file excel terdapat data yang kosong

2. Jelaskan apa yang terjadi pada proses Filter rows!

Melakukan filterisasi terhadap data : memisahkan data yang lengkap ke file csv sedangkan isi data yang terdapat null masuk ke 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 table

Time : 131ms

RAM : 16 GB (1x16)

Processor/CPU : i5 Gen 11

VGA : Intel Iris X

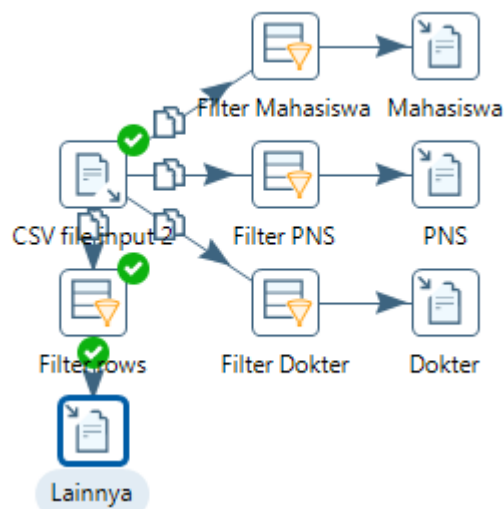
Storage : 512GB Nvme

Tugas Studi Kasus

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.



Mahasiswa

[illegible]

PNS

[illegible]

[illegible][illegible]