# LAPORAN TUGAS DATA WAREHOUS JOBSHEET 1- Data Warehouse Sederhana dan pengenalan PDI Spoon



Disusun Oleh:

Muhammad Afif Khosyidzaki 2341760159

**2G** 

# D4 SISTEM INFORMASI BISNIS TEKNOLOGI INFORMAS POLITEKNIK NEGERI MALANG 2025

	Α	В	С	D	E	F	G	Н
1	Customer	,Product,R	egion,Date	, Item uni	t price , No	.Items , To	tal Sale	
2	Customer	1,Yaris AT	MLG,3/13/	2019,340.9	5,6,2045.7			
3	Customer	1,Etios AT	SBY,4/16/	2019,799.9	5,1,799.95			
4	Customer	3,Avanza	AT,MLG,6/1	18/2019,16	8.95,7,1182	2.65		
5	Customer	3,,MLG,7/	24/2019,16	8.95,2,337.	9			
6	Customer	1,Avanza	AT,SBY,8/2	1/2019,168	.95,9,1520.	.55		
7	Customer	2,Etios AT	KDR,9/7/2	019,799.95	,7,5599.65			
8	Customer	3,Etios AT	,MLG,9/7/2	019,799.95	,8,6399.6			
9	Customer	2,Avanza	AT,,,168.95	,15,2534.25	5			
10	Customer	1,Avanza	AT,SBY,9/1	4/2019,168	.95,15,253	4.25		
11	Customer	1,Etios AT	,SBY,9/25/	2019,799.9	5,2,1599.9			
12	Customer	1,Avanza	AT,SBY,1/2	1/2020,168	.95,14,236	5.3		
13	Customer	1,Yaris AT	MLG,2/1/2	020,340.95	,2,681.9			
14	Customer	1,,,,340.95	,7,2386.65					
15	Customer	1,Etios AT	,SBY,5/30/	2020,799.9	5,9,7199.55	i		
16	Customer	3,Avanza	AT,MLG,6/7	7/2020,168	.95,10,1689	9.5		
17	Customer	3,Yaris AT	MLG,6/9/2	020,340.95	,4,1363.8			
18	Customer	2,,KDR,8/7	7/2020,799.	95,14,1119	9.3			
19	Customer	2,Avanza	AT,KDR,8/8	/2020,168.	95,12,2027	.4		
20	Customer	1,Avanza	AT,,8/12/20	20,168.95,	4,675.8			
			Gamhar	Data Peni	ıalan Mohi	il		

#### **TUGAS 1**

- 1. Analisa lah data tersebut!
- 2. Berapa jumlah kolom pada data tersebut?
- 3. Apa arti atau isi data dari setiap kolom yang ada?
- 4. Adakah data yang memiliki nilai null / data yang tidak lengkap?
- 5. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama? **Jawab:**

#### **Analisis Data**

- 1. Data memiliki 7 kolom, yaitu: Customer,Product,Region,Date,Item unit price,No.Items,Total Sale
- 2. Arti atau Isi Data dari Setiap Kolom
  - Customer: Nama atau kode unik pelanggan, misalnya Customer 1, Customer 2, dst.
  - **Product**: Nama tipe produk yang dibeli pelanggan seperti Yaris AT, Etios AT, atau Avanza AT.
  - **Region**: Kota atau wilayah tempat produk dijual seperti MLG (Malang), SBY (Surabaya), atau KDR (Kediri).
  - Date: Tanggal penjualan produk dalam format MM/DD/YYYY.
  - Item unit price: Harga satuan untuk produk yang dibeli.
  - **No.Items**: Jumlah unit produk yang dibeli.
  - Total Sale: Total penjualan, yang biasanya dihitung dari Item unit price

### 3. Cek Data dengan Nilai Null

Region: Pada baris ke-5, ke-9, ke-14, ke 15, dan ke-20, kolom "Region" tidak memiliki nilai.

Date: Pada baris ke-9 dan ke-20, kolom "Date" kosong.

**Product**: Pada baris ke-5 dan ke-15, kolom "Product" kosong.

### 4. Perbedaan Tipe Data pada Kolom yang Sama

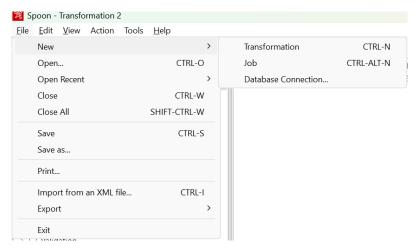
**Customer**: Konsisten berupa teks, meskipun beberapa datanya tampak lebih singkat (contoh: baris 5 hanya "Customer 3,").

**Date**: Beberapa nilai kosong (null) terlihat di kolom ini, sehingga tipe data tidak konsisten antara tanggal dan kosong.

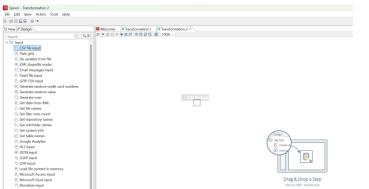
**Region**: Sama seperti kolom "Date", nilai kosong menyebabkan ada inkonsistensi antara tipe data teks dan kosong.

Total Sale: Konsisten berupa angka numerik, namun kemungkinan ada masalah dengan null jika ada konversi lebih lanjut.

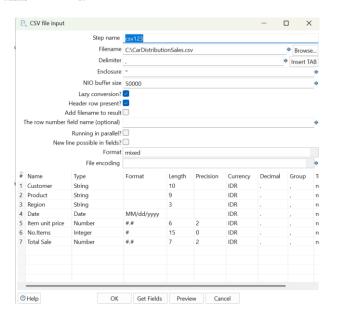
#### STEP 1



## STEP 2



## STEP 3



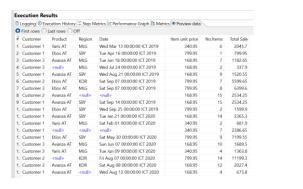
## STEP 4



#### STEP 5



#### STEP 6



#### **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?
- 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. Untuk menentukan apakah data hasil eksekusi sesuai dengan data aslinya, perlu dilakukan validasi data. Pada Pentaho Data Integration (PDI), Anda dapat memvalidasi hasil transformasi atau proses extract data dengan membandingkan jumlah record input dan output di Execution Results Area (Tab Logging), terlihat hasil proses sebagai berikut:
- Total Input (I) = 20
- Total Output (W) = 19
- Ada 1 record yang mungkin hilang atau ada masalah dengan transformasi. Oleh karena itu, hasil eksekusi belum sepenuhnya sesuai dengan data aslinya.

#### 2. A. Load File/Source

Gunakan langkah seperti "Input Data CSV", "Table Input", atau "Data Input" lainnya untuk membaca data dari sumbernya. Pastikan koneksi ke file/source tersambung dengan benar.

#### B. Validasi Data

Periksa bahwa semua baris data memiliki kolom lengkap (tidak ada kolom kosong yang tidak diharapkan).Gunakan langkah "Select Values" untuk memvalidasi dan memfilter data yang dibutuhkan.

C. Transformasi Awal (Opsional)

Lakukan pengolahan awal seperti pengecekan null values atau formatisasi tanggal jika diperlukan.

### D. Log Aktivitas

Buka Execution Results Area untuk memantau jumlah input, output, atau log error yang terjadi untuk setiap langkah. Save/Output Results

E. Ekstrak data yang sudah valid ke tujuan (file/DB) menggunakan step seperti "Output CSV" atau koneksi database.

3. I (Input): Jumlah baris data yang dibaca dari sumber data (Input).

Pada log ini, jumlah Input adalah 20.

O (Output): Jumlah baris data yang berhasil dikeluarkan dari proses tertentu untuk langkah berikutnya.

Jumlah Output dalam log ini adalah 0 karena mungkin proses ini hanya untuk membaca data.

R (Read): Jumlah record yang dibaca oleh step khusus di transformasi.

Pada log, 0 menunjukkan tidak ada pembacaan lain.

W (Write): Jumlah baris data yang berhasil ditulis ke output.

Pada log, 19, mengindikasikan bahwa 19 baris berhasil diproses dan ditulis keluar.

U (Updated): Jumlah data yang diperbarui (biasanya terkait database update).

Di  $\log$  ini, U = 0 menunjukkan tidak ada pembaruan data.

E (Error): Jumlah error yang terjadi selama proses.

Error adalah 0, menunjukkan tidak ada kesalahan eksplisit.

4. A. Saya sendiri (Muhammad Afif Khosyidzaki)

- PROCESSOR : AMD RYZEN 7 7000series

- RAM : 16 GB

- STORGAE : SSD 1.526 GB

- GPU : NVIDIA GeForce RTX 4050

- Waktu Eksekusi : 2 Detik

B. Azaria Cindy Sahasika

- PROCESSOR : Intel Core i5 (3.0 GHz)

- RAM : 8 GB

STORGAE : SSD 256 GBGPU : Integrated GPU

- Waktu Eksekusi : 5 Detik

C. Naufal Ardian

- PROCESSOR : AMD RYZEN 5 5000series

- RAM : 16 GB

- STORGAE : SSD 512 GB

- GPU : NVIDIA GeForce RTX 4050

- Waktu Eksekusi : 3 Detik

D. Evan Dhianta Favian

- PROCESSOR : AMD RYZEN 9 9000series

- RAM : 16 GB - STORGAE : 1.526 GB

- GPU : NVIDIA GeForce RTX 4070

- Waktu Eksekusi : 2 Detik

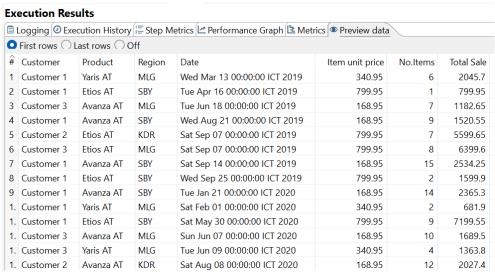
### B. Filter Data (Transform) dan Pengemasan data (Load)

Ini adalah hasil Step by Step bagian B

#### 1. Microsoft Excel Output

Ex	ecution Res	ults					
	Logging @ Exe	ecution History	√ 🔚 Step M	etrics 🗠 Performance Graph 🕒 Met	rics Preview data		
0	First rows	Last rows 🔾 (	Off				
<b>#</b>	Customer	Product	Region	Date	Item unit price	No.Items	Total Sal
1	Customer 3	<null></null>	MLG	Wed Jul 24 00:00:00 ICT 2019	168.95	2	337.9
2	Customer 2	Avanza AT	<null></null>	<null></null>	168.95	15	2534.25
3	Customer 1	<null></null>	<null></null>	<null></null>	340.95	7	2386.65
4	Customer 2	<null></null>	KDR	Fri Aug 07 00:00:00 ICT 2020	799.95	14	11199.3
5	Customer 1	Avanza AT	<null></null>	Wed Aug 12 00:00:00 ICT 2020	168.95	4	675.8

#### 2. Text File Output



### **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 table

### Jawab:

- 1.  **File CSV**: Data disusun dalam format teks terpisah oleh koma (comma-separated values). Dalam file CSV pertama, data hanya terdiri dari nilai angka dan teks tanpa format atau styling.
  - **File Excel**: Data dalam file Excel lebih terstruktur, menggunakan kolom dan baris yang dapat diformat dengan lebih fleksibel. Terdapat data yang lebih lengkap dengan kolom tambahan dan baris yang lebih banyak jika dibandingkan dengan file CSV.

Di file Excel kedua, terlihat bahwa data memiliki lebih banyak entri dan terdapat variasi produk, tanggal, serta penjualan yang lebih rinci.

2. Biasanya, dalam proses Filter Rows, data yang tidak relevan atau tidak diperlukan akan difilter, sehingga hanya data yang diperlukan muncul. Misalnya, jika kita hanya ingin

melihat data berdasarkan produk, wilayah, atau rentang waktu tertentu, maka proses filter akan menghapus baris-baris yang tidak memenuhi kriteria tersebut.

# 3. A. Saya sendiri (Muhammad Afif Khosyidzaki)

- PROCESSOR : AMD RYZEN 7 7000series

- RAM : 16 GB

- STORGAE : SSD 1.526 GB

- GPU : NVIDIA GeForce RTX 4050

- Waktu ETL : 2 Detik

## B. Azaria Cindy Sahasika

- PROCESSOR : Intel Core i5 (3.0 GHz)

- RAM : 8 GB

- STORGAE : SSD 256 GB - GPU : Integrated GPU

- Waktu ETL : 3 Detik

#### C. Naufal Ardian

- PROCESSOR : AMD RYZEN 5 5000series

- RAM : 16 GB

- STORGAE : SSD 512 GB

- GPU : NVIDIA GeForce RTX 4050

- Waktu ETL : 0 Detik

### D. Evan Dhianta Favian

- PROCESSOR : AMD RYZEN 9 9000series

- RAM : 16 GB - STORGAE : 1.526 GB

- GPU : NVIDIA GeForce RTX 4070

- Waktu ETL : 0 Detik

#### 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. darai 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

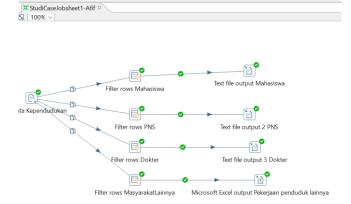
data-data tersbut akan digunakan untuk analisa di masing-masing wilayah dan juga Dashboard untuk emmbantu mengambil keputusan.

Pak Pascanowo mendapatakan data penduduk yang akan pindah ke wilayahnya melalui link: <a href="https://github.com/dik4rizky/datasources/blob/82821b8ec341176f979956c84b79720d4aead0">https://github.com/dik4rizky/datasources/blob/82821b8ec341176f979956c84b79720d4aead0</a> 12/dataDummyKependudukan.csv

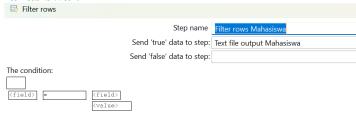
Bantulah Pak Pascanowo untuk memishkan data tersebut menjadi 4 file: MasyarakatMahasiswa, MasyarakatPNS, MasyarakatDokter, MasyarakatLainnya.

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

### Jawab:



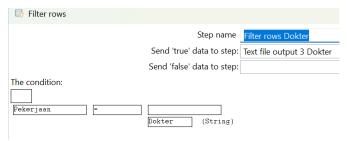
## **Condition Filter Mahasiswa:**



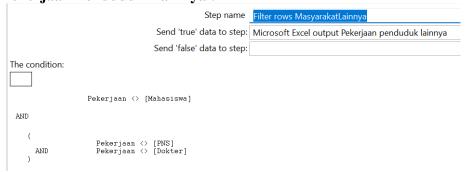
## **Condition Filter PNS:**



## **Condition Dokter:**



# Condtion Pekerjaan Penduduk Lainnya:



# OUTPUT FILE MASING MASING YANG DI INGKAN PAK PASCANOWO

Pak emir menerima Hasil Data Masyarakat Berprofrsi Mahasiswa Selaku Ketua RT Wilayah Pendidikan

# MahasiswaOutput:

		,										
# Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
1 Siti Rahayu	3503010101010002	Bandung	15-05-1990	Perempuan	Jl. Cendrawasih No. 45	Bandung	003/004	Cikutra	Bandung Wetan	Kristen	Belum Menikah	Mahasiswa
2 Adi Pratama	3503010101010005	Semarang	05-03-1995	Laki-laki	Jl. Pemuda No. 21	Semarang	009/010	Pandanaran	Semarang Tengah	Buddha	Belum Menikah	Mahasiswa
3 Maya Wulandari	3503010101010020	Jayapura	23-10-1993	Perempuan	Jl. Proklamasi No. 80	Jayapura	039/040	Kota Jayapura	Jayapura	Kristen	Belum Menikah	Mahasiswa

# Pak Yasana Laili menerima Hasil Data Masyarakat Berprofesi PNS selaku ketua Rt wilayah Pemerintahan

# **PNSOutput:**

₩ Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
1 Andi Wijaya	3503010101010007	Makassar	25-06-1978	Laki-laki	Jl. Jenderal Sudirman No. 10	Makassar	013/014	Barombong	Makassar Selatan	Islam	Menikah	PNS
2 Agus Santoso	3503010101010015	Padang	02-04-1982	Laki-laki	Jl. Sudirman No. 20	Padang	029/030	Padang Barat	Padang	Islam	Menikah	PNS

# Pak Budi pekerti menerima hasil data Masyarakt berprofesi Dokter Selaku ketua RT wilayah Kesehatan

# **DokterOutput:**

Ĥ	Nama	NIK Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
1	Desi Purnama	3503010101010008 Palembang	08-04-1993	Perempuan	Jl. Demang Lebar Daun No. 20	Palembang	015/016	Alang-Alang Lebar	Palembang Selatan	Kristen	Belum Menikah	Dokter
2	Bambang Kusumo	3503010101010017 Banjarmasin	25-09-1980	Laki-laki	Jl. A. Yani No. 60	Banjarmasin	033/034	Banjarmasin Tengah	Banjarmasin	Hindu	Menikah	Dokter

# MasyrakatLainnyaOutput:

₩ Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
1 Budi Santoso	3503010101010001	Jakarta	01-01-1980	Laki-laki	Jl. Merdeka No. 123	Jakarta	001/002	Menteng	Jakarta Pusat	Islam	Menikah	Wiraswasta
2 Joko Susilo	3503010101010003	Surabaya	10-10-1975	Laki-laki	Jl. Diponegoro No. 67	Surabaya	005/006	Gubeng	Surabaya Pusat	Katolik	Menikah	Pegawai Swasta
3 Maya Dewi	3503010101010004	Yogyakarta	20-08-1988	Perempuan	Jl. Malioboro No. 8	Yogyakarta	007/008	Sosromenduran	Yogyakarta Utara	Hindu	Belum Menikah	Guru
Rini Setiawan	3503010101010006	Malang	12-12-1982	Perempuan	Jl. Simpang Borobudur No. 15	Malang	011/012	Dinoyo	Malang Selatan	Kristen	Menikah	Wiraswasta
5 Fajar Prabowo	3503010101010009	Medan	30-09-1986	Laki-laki	Jl. Sudirman No. 50	Medan	017/018	Sei Sikambing	Medan Timur	Hindu	Menikah	Wiraswasta
5 Rina Sari	3503010101010010	Denpasar	17-07-1992	Perempuan	Jl. Gatot Subroto No. 30	Denpasar	019/020	Dauh Puri Kauh	Denpasar Utara	Buddha	Belum Menikah	Freelancer
7 Ahmad Subhan	3503010101010011	Surakarta	03-07-1991	Laki-laki	Jl. Pahlawan No. 55	Surakarta	021/022	Jebres	Surakarta Timur	Islam	Belum Menikah	Wiraswasta
B Dian Pertiwi	3503010101010012	Maluku	10-12-1985	Perempuan	Jl. Pattimura No. 10	Ambon	023/024	Baguala	Ambon	Kristen	Menikah	Guru
Rudi Hartono	3503010101010013	Pontianak	28-05-1973	Laki-laki	Jl. Ahmad Yani No. 30	Pontianak	025/026	Pontianak Barat	Pontianak Kota	Hindu	Menikah	Wiraswasta
1. Lina Fitriani	3503010101010014	Manado	15-08-1994	Perempuan	Jl. Wolter Monginsidi No. 17	Manado	027/028	Mapanget	Manado Selatan	Kristen	Belum Menikah	Akuntan
Rina Agustina	3503010101010016	Bengkulu	19-11-1990	Perempuan	Jl. Imam Bonjol No. 45	Bengkulu	031/032	Ratu Samban	Bengkulu Tengah	Kristen	Menikah	Wiraswasta
Nia Dewi Lestari	3503010101010018	Manokwari	14-06-1987	Perempuan	Jl. Diponegoro No. 40	Manokwari	035/036	Manokwari Barat	Manokwari	Kristen	Belum Menikah	Penulis
. Doni Setiawan	3503010101010019	Palu	07-03-1977	Laki-laki	Jl. Soekarno-Hatta No. 70	Palu	037/038	Palu Barat	Palu	Islam	Menikah	Wiraswasta