

LAPORAN PRAKTIKUM

JOBSHEET 1: DATA WAREHOUSE SEDERHANA DAN PENGENALAN PDI SPOON

Dosen Pengampu:
Vipkas Al Hadid Firdaus., ST., MT.



Disusun Oleh:
Yusra Yusuf
NIM: 2341760044

POLITEKNIK NEGERI MALANG
JURUSAN TEKNOLOGI INFORMASI
PROGRAM STUDY SISTEM INFORMASI BISNIS

2025

Jobsheet 1: Data Warehouse Sederhana dan Pengenalan PDI Spoon

TUGAS 1

	A	B	C	D	E	F	G	H
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	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	
21								

1. Berapa jumlah kolom pada data tersebut?

Jumlah Kolom: Dataset memiliki **7 kolom**.

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

- **Customer:** Nama pelanggan.
- **Product:** Model mobil yang dibeli.
- **Region:** Wilayah pembelian.
- **Date:** Tanggal transaksi (format teks).
- **Item unit price:** Harga satuan mobil (angka desimal).
- **No.Items:** Jumlah unit mobil yang dibeli (bilangan bulat).
- **Total Sale:** Total penjualan (angka desimal)

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

- **Product:** 3 nilai kosong.
- **Region:** 3 nilai kosong.
- **Date:** 2 nilai kosong.

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

TIDAK ADA

TUGAS 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

SUDAH SAMA

Examine preview data

Rows of step: Input Data CSV (19 rows)

#	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
1	Customer 1	Yaris AT	MLG	03/13/2019	340.9	6	2045.7
2	Customer 1	Etios AT	SBY	04/16/2019	800	1	800
3	Customer 3	Avanza AT	MLG	06/18/2019	168.9	7	1182.7
4	Customer 3		MLG	07/24/2019	168.9	2	337.9
5	Customer 1	Avanza AT	SBY	08/21/2019	168.9	9	1520.5
6	Customer 2	Etios AT	KDR	09/07/2019	800	7	5599.6
7	Customer 3	Etios AT	MLG	09/07/2019	800	8	6399.6
8	Customer 2	Avanza AT			168.9	15	2534.2
9	Customer 1	Avanza AT	SBY	09/14/2019	168.9	15	2534.2
10	Customer 1	Etios AT	SBY	09/25/2019	800	2	1599.9
11	Customer 1	Avanza AT	SBY	01/21/2020	168.9	14	2365.3
12	Customer 1	Yaris AT	MLG	02/01/2020	340.9	2	681.9
13	Customer 1				340.9	7	2386.7
14	Customer 1	Etios AT	SBY	05/30/2020	800	9	7199.6
15	Customer 3	Avanza AT	MLG	06/07/2020	168.9	10	1689.5
16	Customer 3	Yaris AT	MLG	06/09/2020	340.9	4	1363.8
17	Customer 2		KDR	08/07/2020	800	14	11199.3
18	Customer 2	Avanza AT	KDR	08/08/2020	168.9	12	2027.4
19	Customer 1	Avanza AT		08/12/2020	168.9	4	675.8

2. Saat PDI Spoon melakukan proses extract, langkah-langkahnya meliputi inisialisasi job atau transformasi, membuka koneksi ke sumber data, menjalankan query atau membaca file, serta mengambil dan memproses data. Setelah itu, Spoon melakukan parsing, validasi awal, dan transformasi jika diperlukan sebelum mengirim data ke langkah berikutnya. Sepanjang proses, Spoon mencatat informasi eksekusi di **Tab Logging**, termasuk jumlah data yang diambil, waktu eksekusi, serta potensi error seperti kegagalan koneksi atau format data yang tidak sesuai.
3. Huruf **I, O, R, W, U, E** biasanya muncul di **Tab Logging** atau indikator pada langkah-langkah transformasi dan memiliki arti sebagai berikut:
- **I (Input)** → Jumlah record yang masuk ke dalam step.
 - **(Output)** → Jumlah record yang keluar dari step.
 - **R (Read)** → Jumlah record yang dibaca dari sumber data.

[illegible]

CSV hanya berisi data mentah, sedangkan Excel dapat menyajikan data dengan format lebih rapi dan kaya fitur.

2. Jelaskan apa yang terjadi pada proses Filter rows!

Filter Rows adalah langkah dalam transformasi ETL yang digunakan untuk menyaring data berdasarkan kriteria tertentu. Dalam **Pentaho Spoon**, Filter Rows bekerja dengan cara:

- Mengevaluasi setiap baris data berdasarkan kondisi yang telah ditentukan (misalnya, hanya menampilkan data dengan total penjualan > 1000).
- Memisahkan data yang memenuhi kriteria (data yang "lolos" filter) untuk diproses lebih lanjut.
- Mengabaikan atau membuang data yang tidak sesuai dengan filter.

3. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract Transfer Load data ini. Sangat cepat