**LAPORAN PRAKTIKUM**

**MATA KULIAH DATA WAREHOUSE**

Dosen Pengampu : Vipkas Al Hadid Firdaus, S.T., M.T.

**JOBSHEET 1: DATA WAREHOUSE DAN PENGENALAN PDI SPOON**

****

Nama : Yonanda Mayla Rusdiaty

NIM : 2341760184

Prodi : D-IV Sistem Informasi Bisnis

**JURUSAN TEKNOLOGI INFORMASI**

**POLITEKNIK NEGERI MALANG**

**2024**

1. Studi Kasus

Pak Amir adalah Manager Sales Astro Mobil, yang bergerak di bidang distributor mobil wilayah Jawa Timur. Pak Amir meminta data penjualan dari beberapa cabang kepada supervisor. Data tersebut akan digunakan untuk membuat suatu Dashboard pengambilan keputusan. Dikarenakan proses permintaan ini dilakukan secara berulang setiap hari setelah jam kantor maka, Pak Amir akan membuat proses untuk menarik data yang ada pada file milik supervisor tersebut secara otomatis. Namun, kadang data tersebut tidak lengkap. Sehingga Pak Amir akan mengambil data yang lengkap saja dan mengembalikan data yang tidak lengkap.

Dari studi kasus tersebut maka, akan dilakukan:

1. Cek dan Analisa data penjualan.
2. Import data dari file (Extract)
3. Identifikasi data yang tidak lengkap (missing data) dan meletakkan pada file yang berbeda (Transform)
4. Memindahkan data yang sudah lengkap ke file dashboard. (Load)
5. Mengumpulkan data yang belum lengkap untuk dikembalikan.

\*Data penjualan tersebut dapat diakses dan di download melalui link berikut:

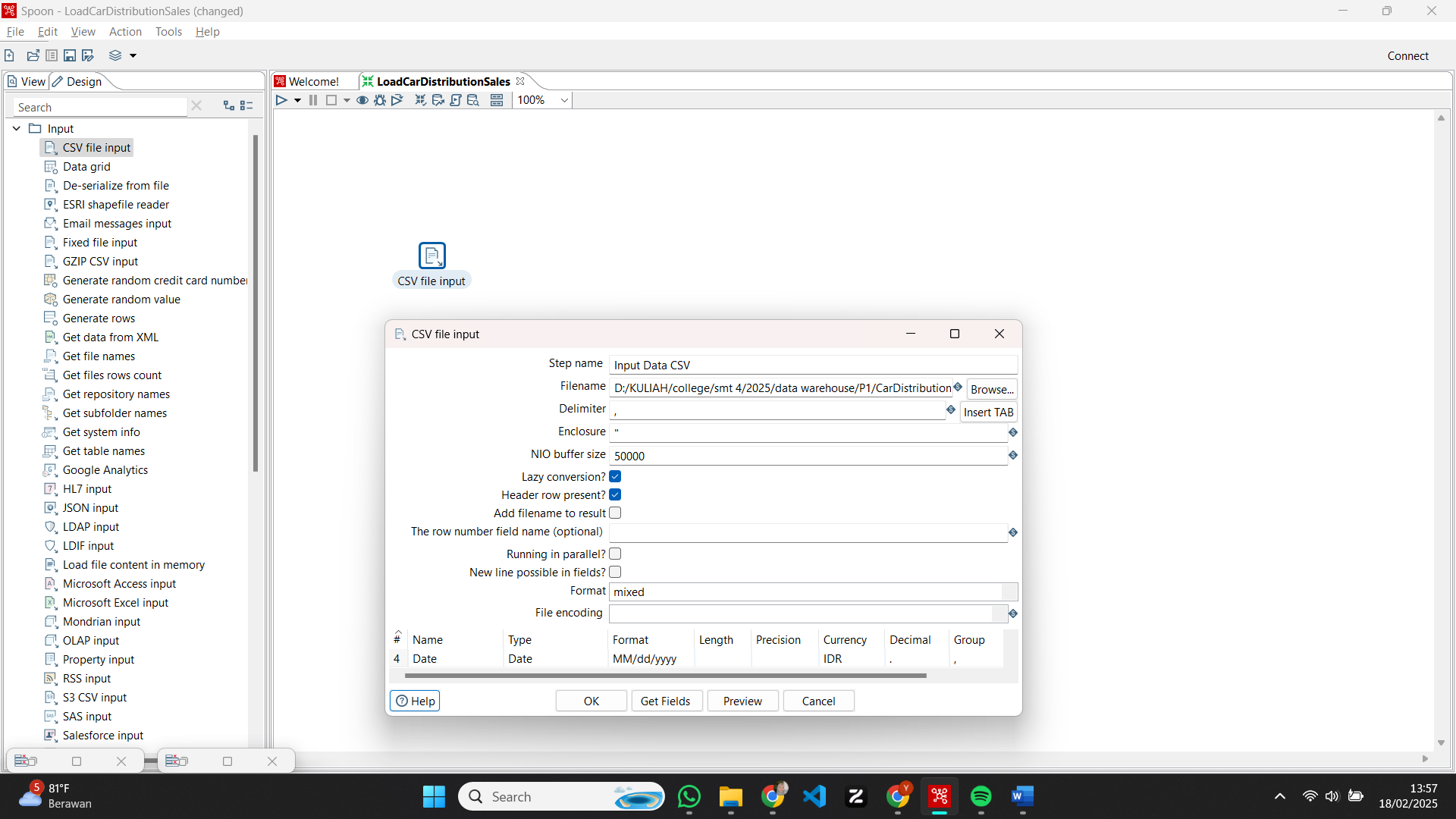
<https://github.com/dik4rizky/datasources/blob/7e3f70f516a1b107635141d09862a8c30fbb6812/CarDistributionSales.csv>

**TUGAS 1:**

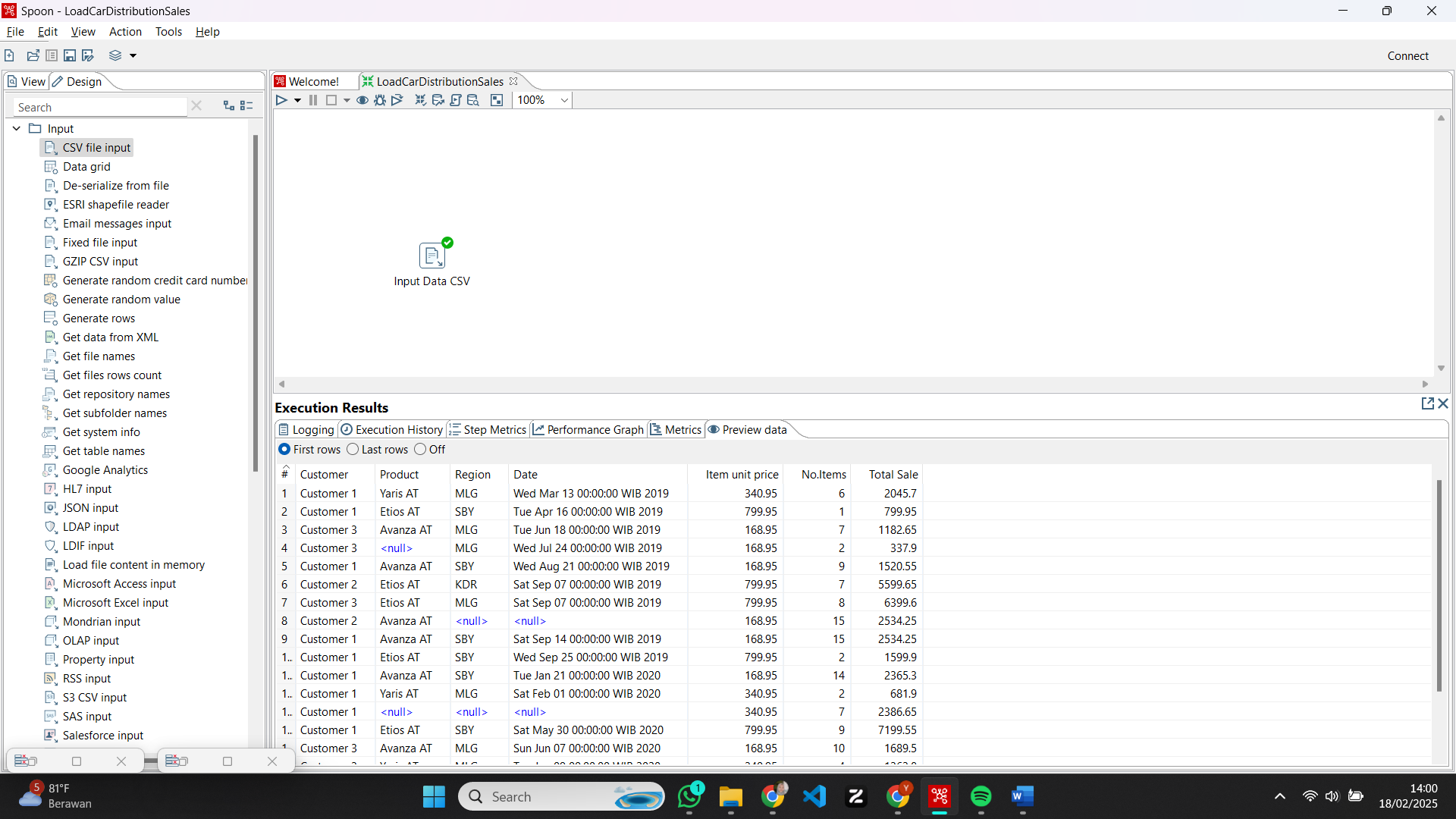
1. Berapa jumlah kolom pada data tersebut?
2. Apa arti atau isi data dari setiap kolom yang ada?
3. Adakah data yang memiliki nilai null / data yang tidak lengkap?
4. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?

**JAWAB:**

1. Ada 7 kolom
2. Berikut adalah arti atau isi data dari setiap kolom yang ada:
3. Customer: Nama pelanggan (type data: teks)
4. Product: Nama produk yang dibeli (type data: teks)
5. Region: Wilayah tempat produk terjual (type data: teks)
6. Date: Tanggal transaksi (type data: teks)
7. Item unit price: Harga per unit dari produk (tipe data: numerik - float)
8. No.Items: Jumlah item yang dibeli (tipe data: numerik - int)
9. Total Sale: Total penjualan dari transaksi (tipe data: numerik - float)
10. Ada, berikut datanya:
11. Product: 3 data kosong
12. Region: 3 data kosong
13. Date: 2 data kosong
14. Ada, yaitu Product, Region, dan Date memiliki tipe data yang tidak konsisten (terdapat nilai dengan tipe berbeda dalam kolom tersebut)
15. **Extract**
16. Step satu kita akan mengekstraksi datanya. Buka **File** lalu **Transformation** dan input file CSV CarDistributionSales.csv



1. Berikut adalah hasil setelah di Run



**TUGAS 2**

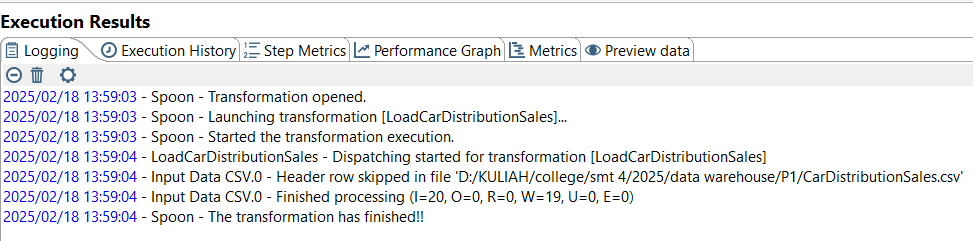
1. Apakah data hasil eksekusi sesuai dengan data aslinya?
2. PDI Spoon melakukan proses extract, perhatikan pada Tab Logging di Execution Results Area, langkah-langkah apa sajakah yang dilakukan PDI Spoon untuk melakukan extract data?
3. Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E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



1. 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, sama dengan data aslinya. Karena pada tahap extract, pentaho hanya membaca dan mengambil data dari sumber tanpa melakukan perubahan
2. Berikut tab Logging pada Execution Results Area



1. Berikut adalah artinya:
2. I (Input) = 20 → Total 20 baris data yang dibaca dari sumber CSV.
3. O (Output) = 0 → Tidak ada data yang ditulis ke output (bisa jadi karena langkah transformasi belum dihubungkan ke output).
4. R (Rejected) = 0 → Tidak ada baris yang ditolak atau gagal diproses.
5. W (Written) = 19 → 19 baris berhasil ditulis ke langkah berikutnya (mungkin ada 1 baris yang terlewati karena header).
6. U (Updated) = 0 → Tidak ada baris yang diperbarui (biasanya untuk database update).
7. E (Error) = 0 → Tidak ada error yang terjadi selama proses.
8. Berikut adalah hasil perbandingannya:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | Nama | Waktu Eksekusi | Processor | RAM | Storage | VGA | CPU |
| 1 | Yonanda | 1 detik | Intel Core i5-12450H (12 CPU), ~2.0GHz | 16GB | 512GB | NVIDIA GeForce RTX 3050 | Intel Core i5-12450H (12 CPU), ~2.0GHz |
| 2 | Dahniar | 1 detik | Intel(R) Core(TM) i5-6300U | 8GB | 256GB | Intel(R) HD Graphics 520 | Intel(R) Core(TM) i5-6300U |
| 3 | Hamdan | 1 detik | AMD Ryzen 5300U | 12GB | 500GB + 256 GB | AMD Ati | AMD Ryzen 5300U |
| 4 | Aqila | 1 detik | AMD Ryzen 5600H | 8GB | 512GB | AMD Radeon(TM) Graphics | AMD Ryzen 5600H |

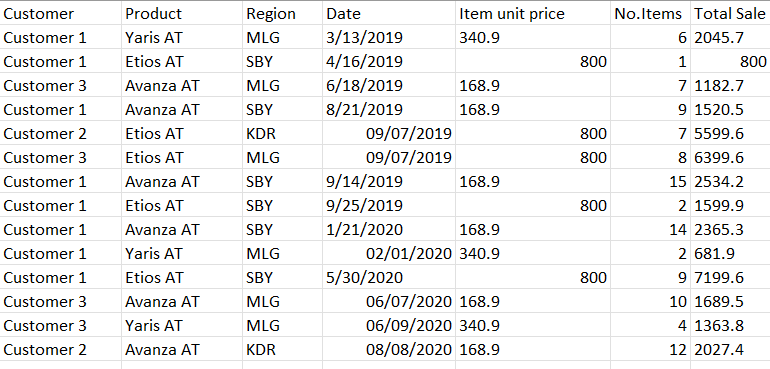
1. **Filter Data (Transformation2) dan Pengemasan Data (Load)**

**TUGAS 3:**

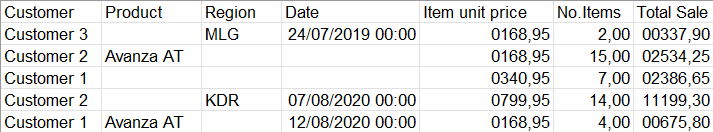
1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?
2. Jelaskan apa yang terjadi pada proses **Filter rows**!
3. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract Transfer Load data ini? Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU ). bandingkan dalam bentuk tabel

**JAWAB:**

1. Berikut adalah isi data output dari file csv dan file excel:
2. File csv



1. File excel



1. Filter rows merupakan proses filtrasi data pada kondisi tertentu. Filter rows mendapatkan data dari input file csv. Hasilnya akan dibagi dalam 2 kondisi yaitu:
2. Result is true: Baris yang memenuhi kondisi akan diarahkan ke langkah Text File Output.
3. Result is false: Baris yang tidak memenuhi kondisi akan diarahkan ke langkah Microsoft Excel Output.

Selanjutnya kita akan mengonfigurasi kondisi filter dengan menambahkan kondisi untuk setiap kolom (misalnya, Customer, Product, Region, Date) dengan operator **IS NOT NULL.** Hal ini memastikan hanya baris dengan semua kolom terisi yang akan diteruskan ke output "true"

Langkah terakhir kita akan mengonfigurasi output:

1. Text File Output: Menyimpan baris yang lengkap (tidak ada nilai null) ke dalam file CSV.
2. Microsoft Excel Output: Menyimpan baris yang memiliki nilai null ke dalam file Excel.
3. Berikut adalah hasil perbandingannya

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | Nama | Waktu Eksekusi | Processor | RAM | Storage | VGA | CPU |
| 1 | Yonanda | 2 detik | Intel Core i5-12450H (12 CPU), ~2.0GHz | 16GB | 512GB | NVIDIA GeForce RTX 3050 | Intel Core i5-12450H (12 CPU), ~2.0GHz |
| 2 | Dahniar | 2 detik | Intel(R) Core(TM) i5-6300U | 8GB | 256GB | Intel(R) HD Graphics 520 | Intel(R) Core(TM) i5-6300U |
| 3 | Hamdan | 2 detik | AMD Ryzen 5300U | 12GB | 500GB + 256 GB | AMD Ati | AMD Ryzen 5300U |
| 4 | Aqila | 2 detik | AMD Ryzen 5600H | 8GB | 512GB | AMD Radeon(TM) Graphics | AMD Ryzen 5600H |

**STUDI KASUS**

