Jobsheet 1

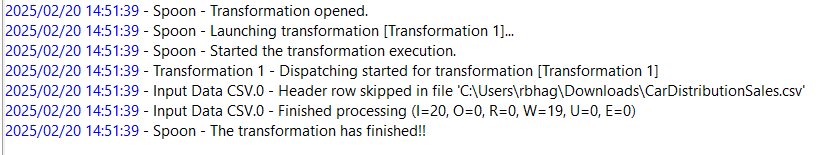
Data Warehouse

**Tugas 1**

1. **Berapa jumlah kolom pada data tersebut?**
   * Customer
   * Product
   * Region
   * Date
   * Item unit price
   * No. Items
   * Total Sale
2. **Apa arti atau isi data dari setiap kolom yang ada?**
   * **Customer**: Nama atau ID pelanggan yang melakukan pembelian.
   * **Product**: Jenis mobil yang dibeli.
   * **Region**: Lokasi tempat pembelian dilakukan.
   * **Date**: Tanggal transaksi pembelian.
   * **Item unit price**: Harga satuan dari produk yang dibeli.
   * **No. Items**: Jumlah unit produk yang dibeli.
   * **Total Sale**: Total nilai penjualan, biasanya dihitung sebagai (Item unit price \* No. Items).
3. **Adakah data yang memiliki nilai null / data yang tidak lengkap?**  
   Ya, terdapat beberapa baris yang tampak tidak lengkap, misalnya:
   * Baris ke-5 (Customer 3,,MLG,7/24/2019,168.95,2,337.9), di mana terdapat kolom kosong setelah "Customer 3," yang mungkin seharusnya berisi nama produk.
   * Baris ke-14 (Customer 1,,,340.95,7,2386.65), yang kehilangan informasi di kolom "Product" dan "Region".
4. **Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?**
   * Beberapa kolom seharusnya memiliki tipe data numerik (misalnya "Item unit price", "No. Items", dan "Total Sale"), tetapi ada kemungkinan beberapa nilai dituliskan dalam format string jika terjadi kesalahan input.
   * Kolom "Date" memiliki format tanggal, tetapi ada kemungkinan beberapa nilai ditulis dalam format yang tidak seragam, seperti 4/16/2019 vs 8/21/2019.
   * Kolom "Customer" terkadang memiliki angka atau karakter yang tidak seragam, misalnya ada baris dengan hanya "Customer 1," yang tampak tidak lengkap.

**Tugas 2**

1. **Apakah data hasil eksekusi sesuai dengan data aslinya?**
   * Hasil data eksekusi sama 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?**



1. **Perhatikan gambar di bawah ini! Apa yang dimaksud dengan I,O,R,W,U,E?**
   * Pada log proses extract, terdapat kode berikut:
   * 2024/02/17 06:51:48 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)

Keterangan dari masing-masing huruf:

* + 1. **I (Input)** = 20 → Jumlah baris data yang diambil dari sumber.
    2. **O (Output)** = 0 → Jumlah baris data yang dikirim ke output (mungkin belum diproses ke tujuan akhir).
    3. **R (Rejected)** = 0 → Jumlah baris data yang ditolak karena kesalahan atau aturan validasi.
    4. **W (Written)** = 19 → Jumlah baris data yang berhasil ditulis atau diproses ke output.
    5. **U (Updated)** = 0 → Jumlah baris data yang diperbarui (jika prosesnya melibatkan pembaruan data).
    6. **E (Error)** = 0 → Jumlah error yang terjadi selama proses extract.

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.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nama | Waktu Eksekusi | Processor | RAM | Storage | VGA |
| Nimas | 0,1s | AMD Ryzen 3 5300U | 16,0 GB | 475 GB | AMD Radeon Graphics |
| Daffa | 0,1s | AMD A9-9425 RADEON R5 | 8,0 GB | 1,2 TB | AMD RADEON (TM) R5 Graphics |
| Dhevina | 0,0s | 12th Gen Intel(R) Core(TM) i5 - 1235U 1.30 GHz | 8,00 GB | 152 GB | Inter(R) UHD Graphics |

**Tugas 3**

1. **Apa perbedaan isi data output dilihat dari isi file CSV dan file Excel?**

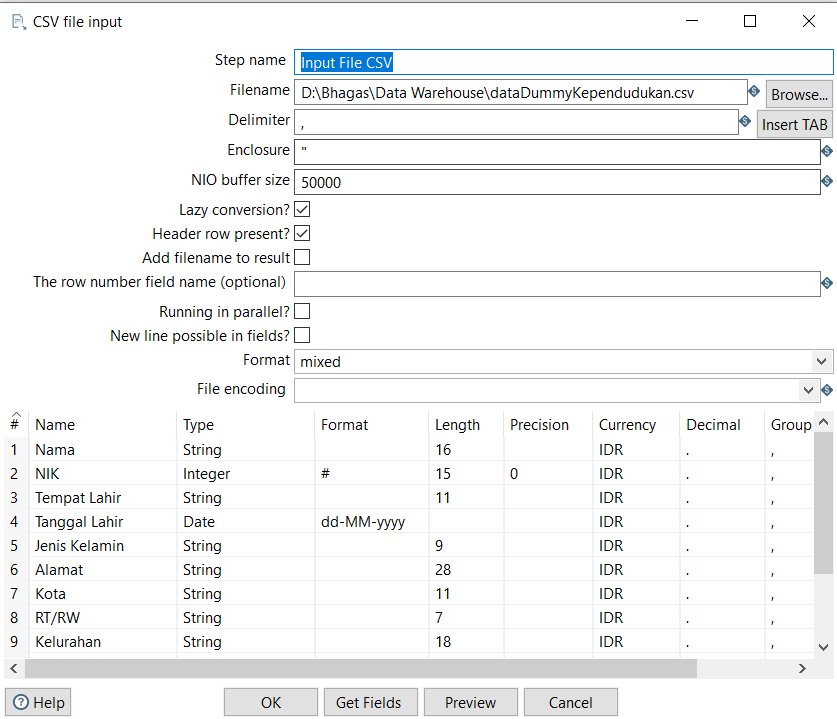
File CSV merupakan hasil awal sebelum dilakukan transformasi data, sedangkan file excel sudah melewati proses pembersihan data

1. **Jelaskan apa yang terjadi pada proses *Filter rows*!**
   * Proses *Filter rows* dalam ETL (Extract, Transform, Load) digunakan untuk menyaring baris data berdasarkan kondisi tertentu.
   * Data yang tidak memenuhi kriteria akan dibuang atau tidak diteruskan ke langkah berikutnya.
2. **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.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nama | Waktu Eksekusi | Processor | RAM | Storage | VGA |
| Nimas | 0,1s | AMD Ryzen 3 5300U | 16,0 GB | 475 GB | AMD Radeon Graphics |
| Daffa | 0,12s | AMD A9-9425 RADEON R5 | 8,0 GB | 1,2 TB | AMD RADEON (TM) R5 Graphics |
| Dhevina | 0,6s | 12th Gen Intel(R) Core(TM) i5 - 1235U 1.30 GHz | 8,00 GB | 152 GB | Inter(R) UHD Graphics |

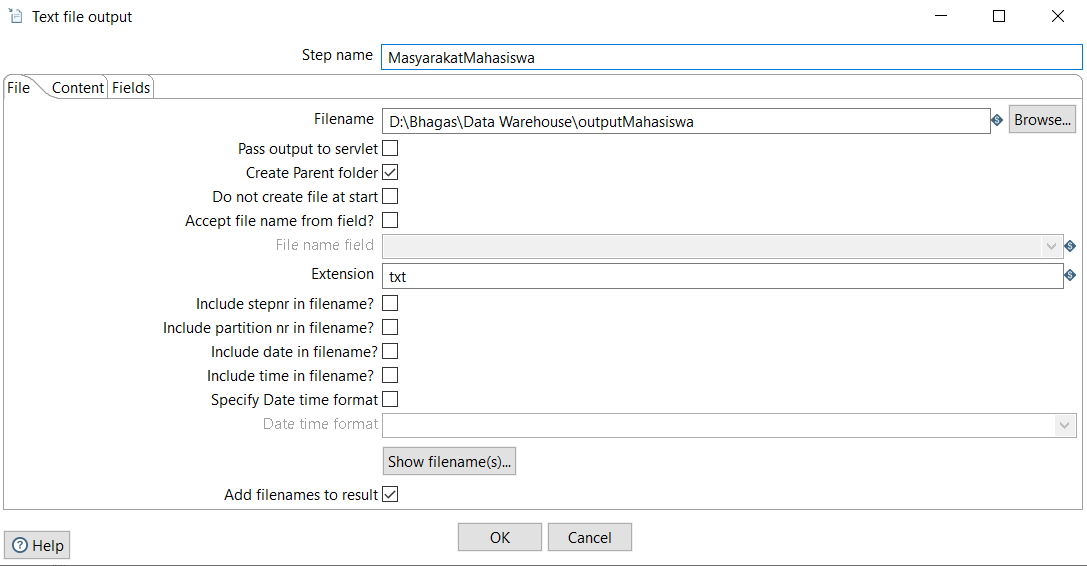
**STUDI KASUS**

**Data Input Kependudukan**





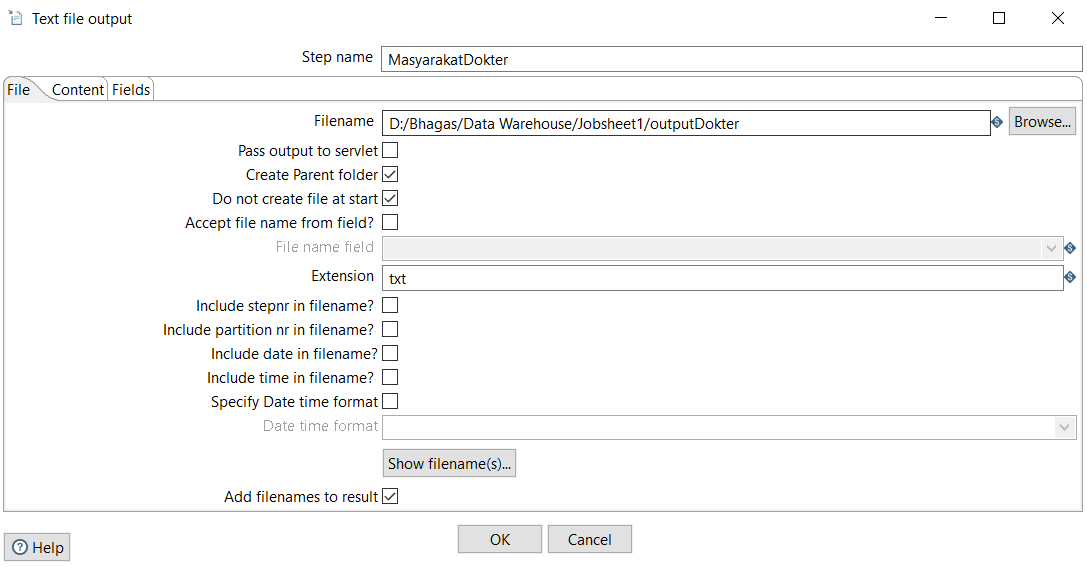
**Data Masyarakat Mahasiswa**

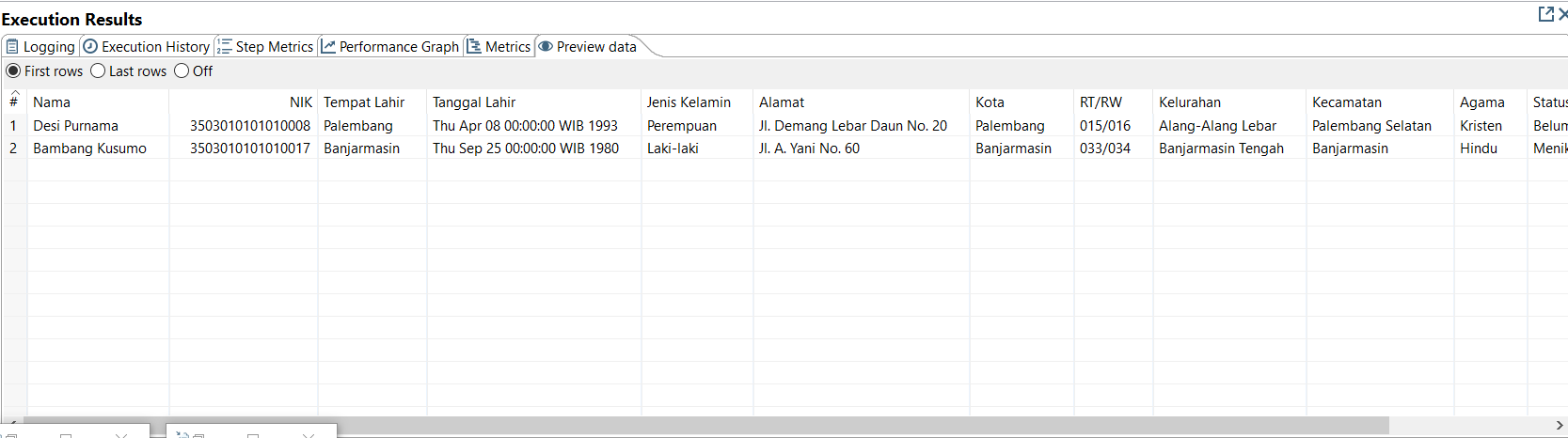




Terdapat 3 Value dalam Data Masyarakat Mahasiswa

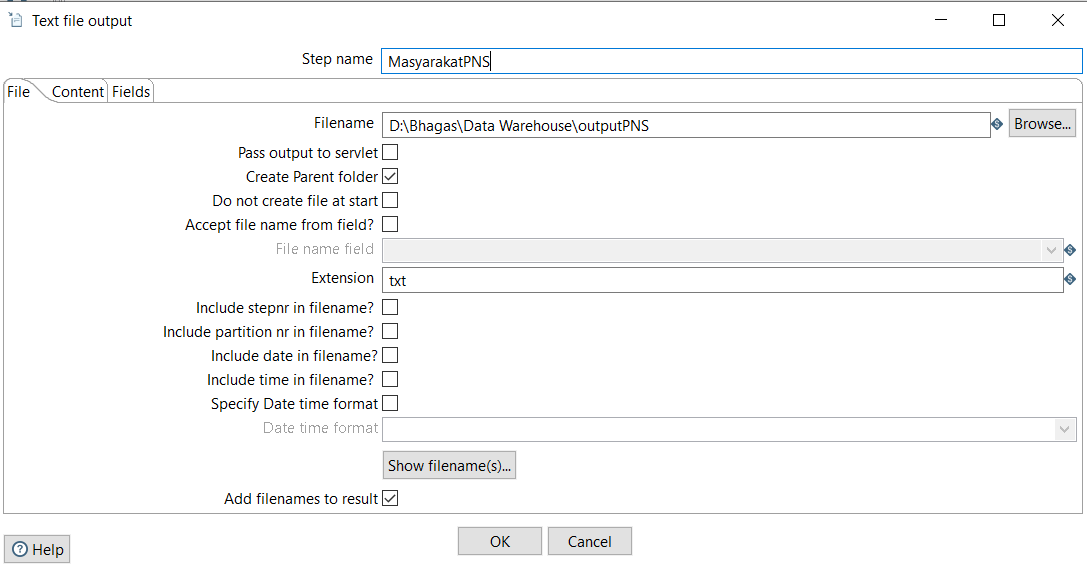
**Data Masyarakat Dokter**





Terdapat 2 Value dalam Data Masyarakat Dokter

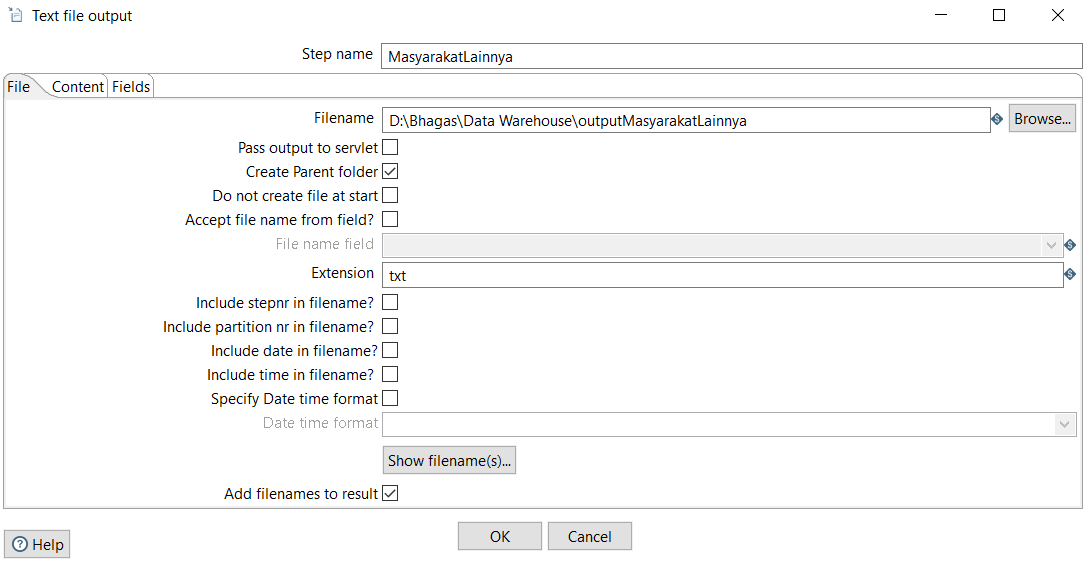
**Data Masyarakat PNS**

****

****

Terdapat 2 Value dalam Data Masyrakat PNS

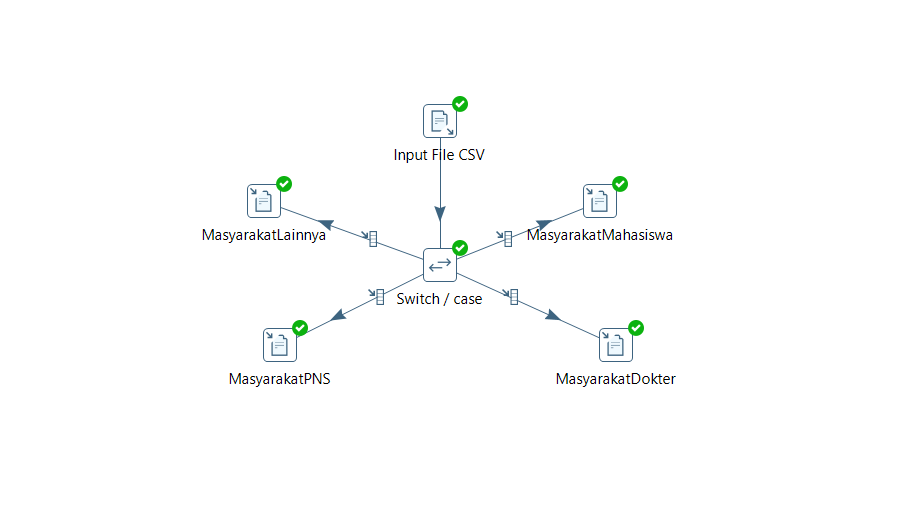
**Data Masyarakat Lainnya**

****

****

Terdapat 13 Value dalam Data Masyarakat (Penulis,Wiraswasta,Freelancer,Guru,Pegawai Swasta,Akuntan)

**Output :**

****