**Penugasan Data Warehouse**

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

Jobsheet 1 : Data Warehouse Sederhana dan Pengenalan PDI Spoon



**Satria Rakhmadani**

**D-IV Sistem Informasi Bisnis | SIB 2A**

**2341760106**

**JURUSAN TEKNOLOGI INFORMASI**

**POLITEKNIK NEGERI MALANG**

**Tujuan Praktikum**

Setelah melakukan praktikum ini, mahasiswa diharapakan dapat lebih mengenal apa itu data warehouse berdasarkan uji coba membuat data warehouse sederhana

**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.

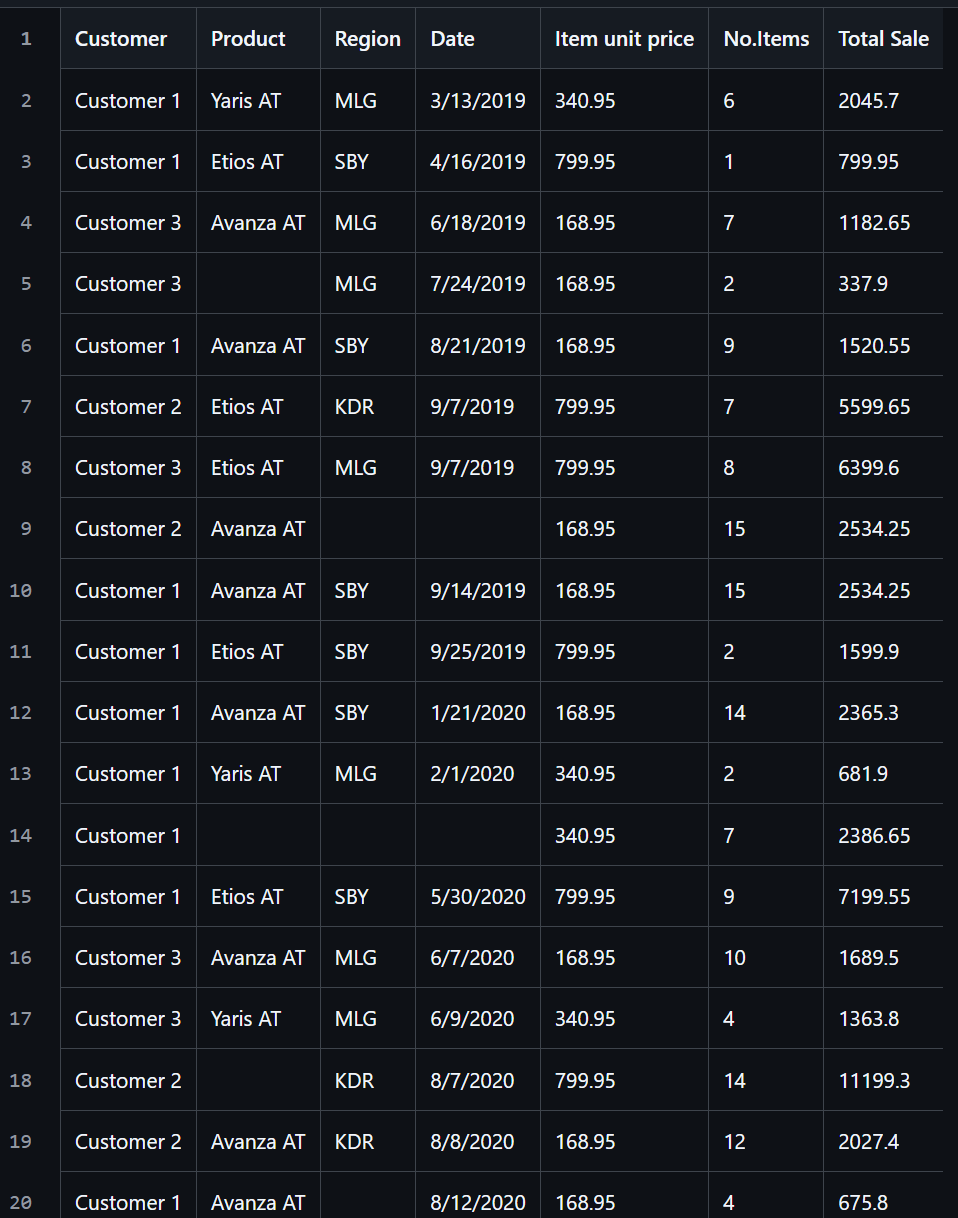
Dari studi kasus tersebut maka, akan dilakukan:

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.

Untuk Data Penjualan berada pada link berikut:

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

Hasil table sebagai berikut:



**TUGAS 1**

Analisa lah data tersebut!

1. Berapa jumlah kolom pada data tersebut?

*Jumlah kolom pada table distribusi mobil ada 7 kolom.*

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

*Kolom* ***Customer*** *sebagai nomor pelanggan, Kolom* ***Produk*** *sebagai produk yang dibeli pelanggan, Kolom* ***Region*** *sebagai wilayah, Kolom* ***Date*** *sebagai tanggal pembelian, Kolom* ***Item Unit Price*** *sebagai harga dari produk, Kolom* ***no.items*** *sebagai jumlah produk yang dibeli, dan Kolom* ***Total Sale*** *sebagai total harga dari harga satuan dikali jumlah produk.*

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

*Ada, dari tidak adanya produk, region, hingga date pada suatu data.*

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

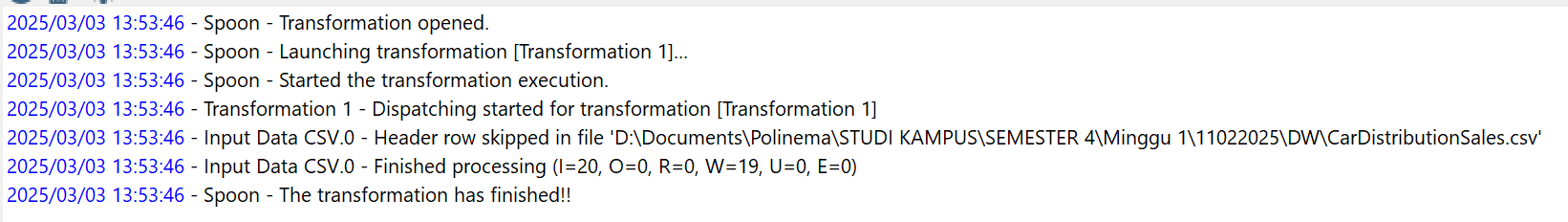
*Tidak ada, semua data yang dimasukkan sesuai sama kolom.*

**TUGAS 2**

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

*Sesuai dengan data asli pada table sebelumnya.*

1. 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?



**Spoon - Transformation opened.**  
*- Pentaho Spoon membuka transformasi yang telah dibuat.*

**Launching transformation [Transformation 1]...**  
*- Transformasi bernama Transformation 1 sedang diluncurkan.*

**Started the transformation execution.**  
*- Eksekusi transformasi telah dimulai.*

**Dispatching started for transformation [Transformation 1]**  
*- Proses dispatching data untuk transformasi telah dimulai.*

**Input Data CSV.0 - Header row skipped in file 'D:\Documents\Polinema\STUDI KAMPUS\SEMESTER 4\Minggu 1\11022025\DW\CarDistributionSales.csv'**  
*- Baris header di file CSV CarDistributionSales.csv dilewati dalam proses ekstraksi.*

**Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)**  
*- Data CSV selesai diproses dengan statistik berikut:*

* ***I=20*** *→ 20 baris data dibaca dari file input.*
* ***W=19*** *→ 19 baris berhasil ditulis ke output.*
* ***O, R, U, E = 0*** *→ Tidak ada data yang dibuang, update, atau mengalami error.*

**Spoon - The transformation has finished!!**  
- *Proses transformasi selesai tanpa error.*

1. Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E ?



**I (Input) = 20**  
- Jumlah baris data yang pada Input Data CSV

**O (Output) = 0**  
- Jumlah baris yang dikirimkan ke step berikutnya sebagai output langsung.

**R (Rejected) = 0**  
- Jumlah baris yang ditolak karena tidak memenuhi syarat tertentu.

**W (Written) = 19**  
- Jumlah baris data yang berhasil ditulis ke output.

**U (Updated) = 0**  
- Jumlah baris yang diperbarui.

**E (Error) = 0**  
- Jumlah baris yang mengalami error selama proses transformasi.

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.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | Nama | Waktu Eksekusi (detik) | Processor | RAM | Storage (SSD/HDD) | VGA | CPU Usage (%) |
| 1 | Satria | 0.1 detik | AMD Ryzen 7 5800H | 8GB | SSD 512GB | RTX 3050 | 40% |
| 2 | Reishi | 1 detik | Inter(R) CoreTM i7-7820HQ | 16GB | SSD 256GB | NVIDIA FeForce 940MX | 25% |

**TUGAS 3**

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

**CSV (Comma-Separated Values)**

* Format penyimpanan berupa teks biasa dengan pemisah koma/tab.
* Ukuran file lebih kecil karena hanya berisi data tanpa format tambahan.
* Tidak mendukung formula atau pemformatan teks.
* Tidak dapat menyimpan lebih dari satu sheet.
* Kompatibel dengan hampir semua sistem.
* Lebih cepat dalam akses dan pemrosesan.

**Excel (XLSX/XLS)**

* Format penyimpanan dalam bentuk biner atau XML yang lebih kompleks.
* Ukuran file lebih besar karena menyimpan metadata, format, dan formula.
* Mendukung formula, grafik, dan pemformatan teks.
* Bisa menyimpan data dalam beberapa sheet.
* Membutuhkan software seperti Microsoft Excel untuk membuka.
* Lebih lambat dibandingkan CSV karena ada pemrosesan tambahan.

1. Jelaskan apa yang terjadi pada proses **Filter rows**!

**Input Data:** Data mentah masuk dari file CSV, database, atau sumber lainnya.

**Kondisi Filter:** Hanya data yang memenuhi syarat yang akan diteruskan.

**Output Data:**

* Data yang sesuai dengan kondisi akan diproses lebih lanjut.
* Data yang tidak memenuhi syarat akan dibuang atau dialihkan ke jalur lain.

1. 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

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | Nama | Waktu Eksekusi (detik) | Processor | RAM | Storage (SSD/HDD) | VGA | CPU Usage (%) |
| 1 | Satria | 0.3 detik | AMD Ryzen 7 5800H | 8GB | SSD 512GB | RTX 3050 | 40% |
| 2 | Reishi | 1 detik | Inter(R) CoreTM i7-7820HQ | 16GB | SSD 256GB | NVIDIA FeForce 940MX | 25% |