**Jobsheet 1**

**Data Warehouse**

****

|  |  |
| --- | --- |
| **Hanin Mariam Abiyyah Hendrik** | **2341760154** |

**STUDY PROGRAM D-IV BUSINESS INFORMATION SYSTEMS**

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**STATE POLYTECHNIC OF MALANG**

**YEAR 2024**

**Jobsheet 1: Data Warehouse Sederhana dan Pengenalan PDI Spoon**

**Tujuan Praktikum**

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

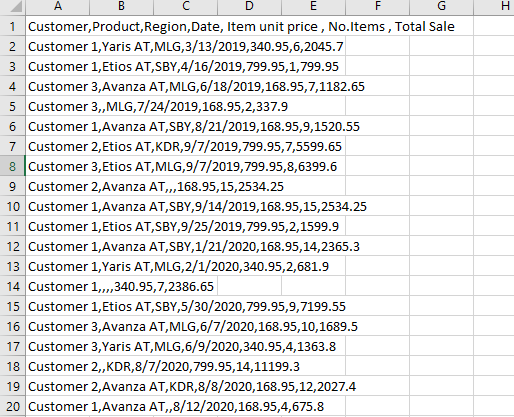
**Studi Kasus**

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.

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?**

Has 7 columns which are:

* Customer
* Product
* Region
* Date
* Item unit price
* No.Items
* Total Sale

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

* Customer: The identity of the customer (e.g. “Customer 1”, “Customer 2”).
* Product: The name of the purchased product (e.g. “Yaris AT”, “Etios AT”, “Avanza AT”).
* Region: Sales region/location (e.g. “MLG”, “SBY”, “KDR”).
* Date: Transaction date in full format with time and time zone.
* Item unit price: The price per unit of the purchased product.
* No.Items: The number of units of the product purchased.
* Total Sale: Total sales value calculated from Item unit price \* No.Items.

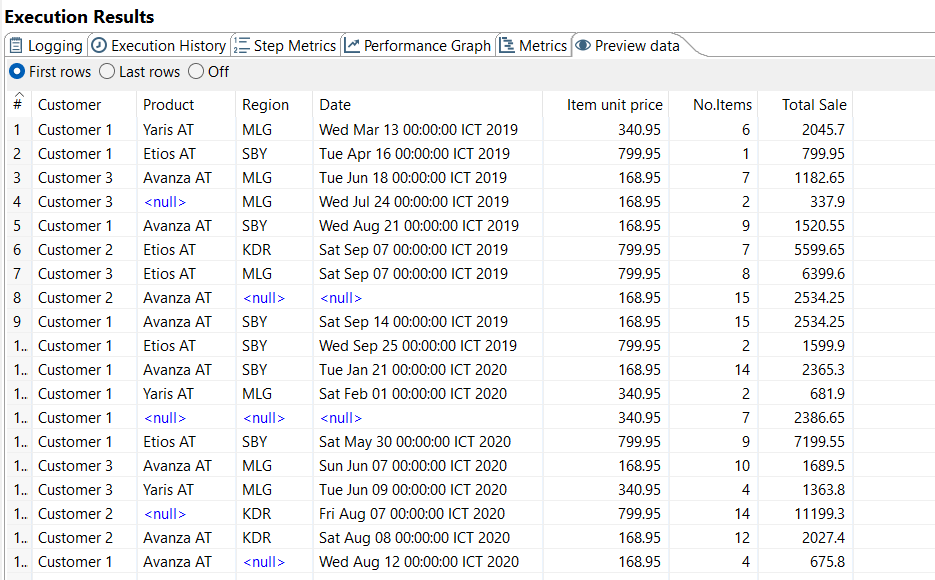
1. **Adakah data yang memiliki nilai null / data yang tidak lengkap?**

Yes, there are data with null values in the product, region and date columns.

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

No, all data seems to have a consistent data type in each column

1. **Pengambilan Data (Extract)**

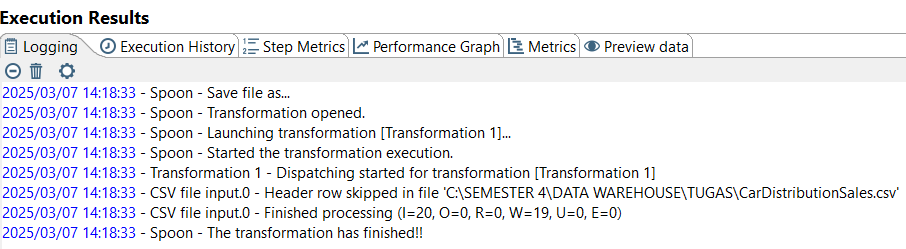


**TUGAS 2**

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

Yes, the executed data matches the original data.

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?**



The log shows the steps of the extract process in PDI Spoon:

* The transformation file is saved and opened.
* The transformation was run and execution started.
* Data from the CSV file was read (20 lines), headers were ignored, and 19 lines were successfully written to the next step without error.
* The transformation completes successfully.

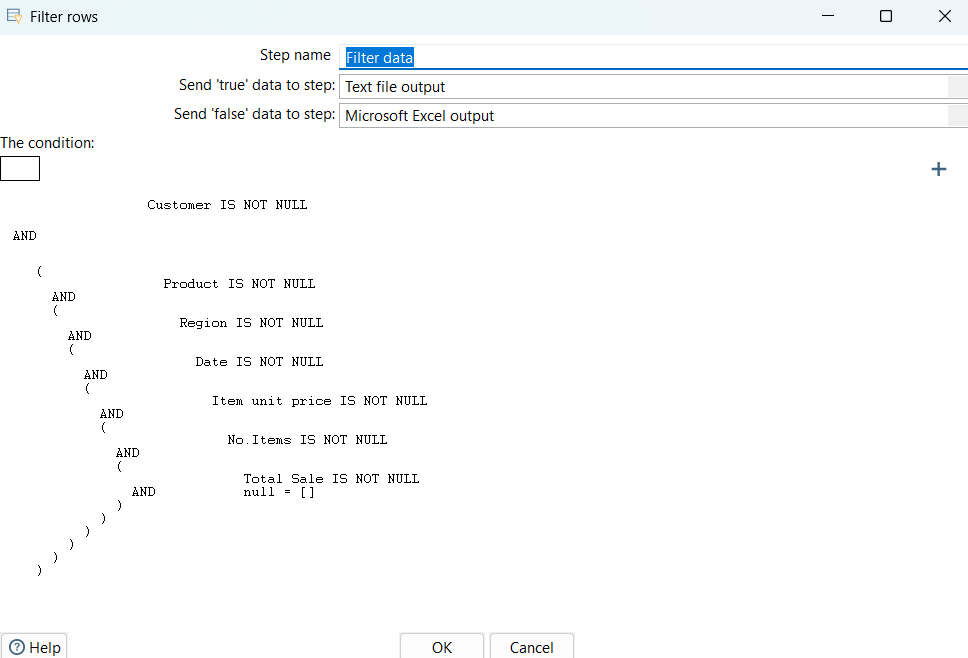
1. **Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E ?**

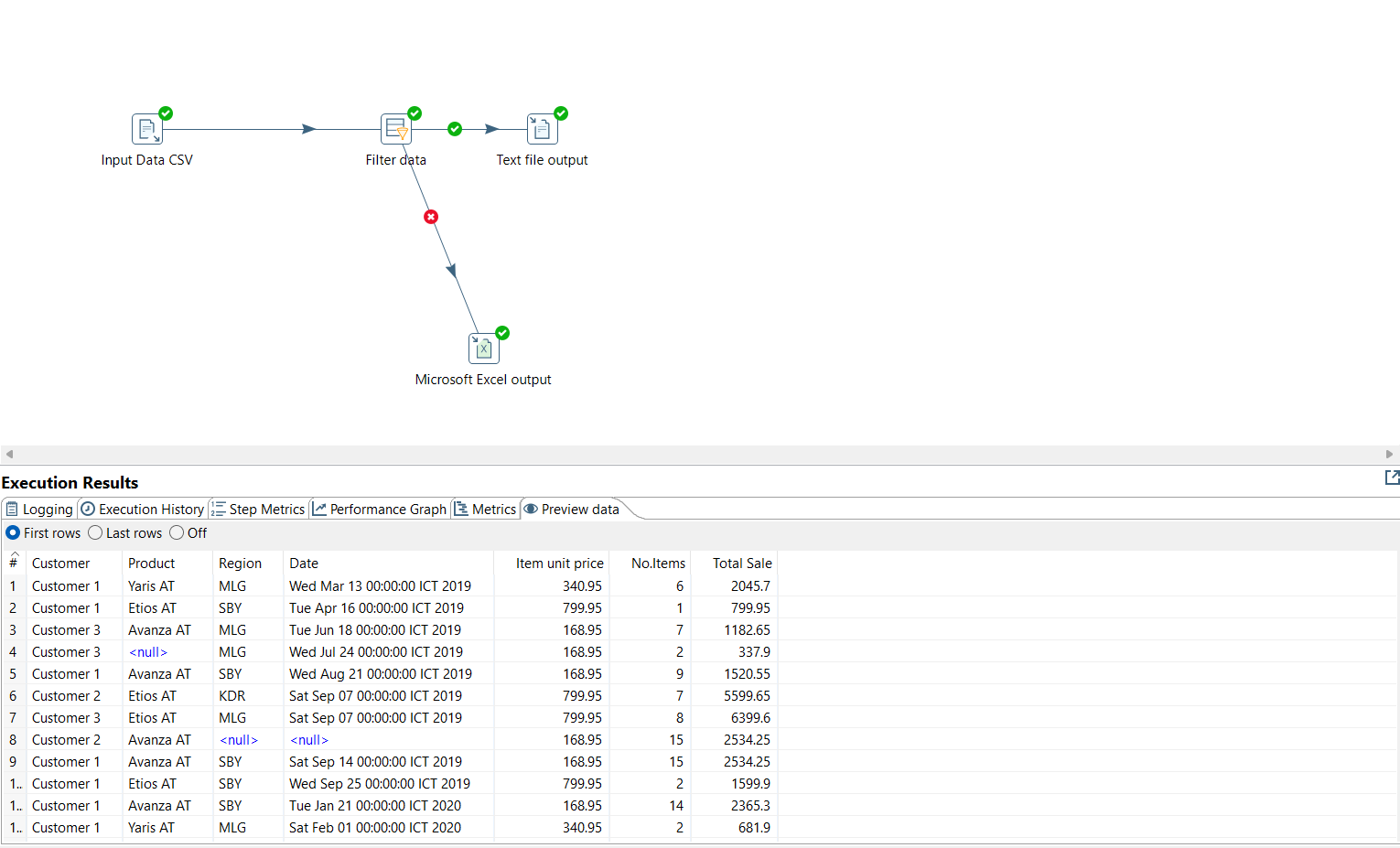
* I (Input) = 20 lines of data read from file.
* (Output) = 0 lines of data were written directly to the output (in this context, the data may be being further processed).
* R (Read) = 0 lines of data were re-read.
* W (Written) = 19 lines of data were successfully written to the next step.
* U (Updated) = 0 rows of data were updated.
* E (Error) = 0 errors occurred during the process.

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** | **Name** | **Processor** | **RAM** | **Storage** | **VGA** | **CPU** | **Time of Execution** |
| 1. | Hanin | Intel® Core™ i7-12650H | 16 GB | 512 GB SSD | GeForce RTX 2050 | 13% | 0 seconds (or less than 1 second) |
| 2. | Fakhira | Intel® Core™ i7-13650H | 16 GB | 477 GB SSD | GeForce RTX 4050 | 7% | 1 second |
| 3. | Sabrina | Intel i5-7300U | 8 GB | Samsung 239 GB | Intel® Graphics 620 | 50% | 1 second |
| 4. | Nada | 11th Gen Intel(R) Core(TM) i5-1155G7 | 16384MB | SSD NVMe 477GB | Intel(R) Iris(R) Xe Graphics | 23% | 1 second |

1. **Filter Data (Transform) dan Pengemasan data (Load)**

****

****

**TUGAS 3**

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

The CSV file contains all the raw data, while the Excel file may be empty or incomplete due to a process failure.

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

Splitting the data according to certain conditions, some data is passed to the text file, and the other part fails to be written to the Excel file.

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** | **Name** | **Processor** | **RAM** | **Storage** | **VGA** | **CPU** | **Time of Execution** |
| 1. | Hanin | Intel® Core™ i7-12650H | 16 GB | 512 GB SSD | GeForce RTX 2050 | 13% | 1 second |
| 2. | Fakhira | Intel® Core™ i7-13650H | 16 GB | 477 GB SSD | GeForce RTX 4050 | 7% | 0 second |
| 3. | Sabrina | Intel i5-7300U | 8 GB | Samsung 239 GB | Intel® Graphics 620 | 50% | 1 second |
| 4. | Nada | 11th Gen Intel(R) Core(TM) i5-1155G7 | 16384MB | SSD NVMe 477GB | Intel(R) Iris(R) Xe Graphics | 23% | 1 second |

**Studi Kasus**

Disclaimer: Data yang tercantum di studi kasus ini adalah fiktif dan hanya digunakan untuk tujuan pembelajaran saja. Nama tempat, tanggal lahir, alamat, dan informasi lainnya yang tertera hanya tipuan belaka dan tidak memiliki kaitan dengan individu yang sebenarnya.

**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: <https://github.com/dik4rizky/datasources/blob/82821b8ec341176f979956c84b79720d4aead012/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.

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |