

JOBSHEET 1 DATA WAREHOUSE



NAMA: BAYU TRIWIBOWO

KELAS: SIB 2A

NIM: 2341760073

TUGAS 1

Analisa lah data tersebut!

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?

JAWABAN

1. Terdapat 7 kolom
2.
 - Customer: Nama pelanggan (String)
 - Product: Jenis mobil yang dibeli (String)
 - Region: Wilayah pembelian (String)
 - Date: Tanggal transaksi (Date)
 - Item unit price: Harga satuan mobil (Float)
 - No.Items: Jumlah unit yang dibeli (Integer)
 - Total Sale: Total penjualan untuk transaksi tersebut (Float)
3.
 - Product: 3 nilai kosong
 - Region: 3 nilai kosong
 - Date: 2 nilai kosong
4. Tidak ada data yang memiliki tipe yang berbeda dalam kolom yang sama.

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 ?
4. 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.

JAWABAN

1. Data hasil eksekusi **harus** sesuai dengan data aslinya jika proses Extract, Transform, dan Load (ETL) berjalan dengan benar. Untuk memastikan hal ini, periksa jumlah record yang diekstrak, bandingkan nilai dari sumber data dengan hasilnya, dan pastikan tidak ada perubahan data yang tidak diinginkan selama proses ETL.
2. PDI Spoon mencatat semua langkah yang dilakukan selama eksekusi di **Tab Logging**. Beberapa langkah umum dalam proses ekstraksi data adalah:
 - **Menghubungkan ke sumber data** (misalnya database, file, API).
 - **Mengeksekusi query atau membaca file** untuk mengambil data.
 - **Membaca dan memproses record** satu per satu atau dalam batch.
 - **Menampilkan jumlah record yang berhasil diambil.**
 - **Menangani error atau exception jika ada kesalahan** dalam membaca data.
3.
 - **I (Input)** → Jumlah record yang masuk ke dalam step.

- **O (Output)** → Jumlah record yang keluar dari step ke step berikutnya.
- **R (Read)** → Jumlah record yang dibaca dari sumber data.
- **W (Written)** → Jumlah record yang ditulis ke target (misalnya database atau file).
- **U (Updated)** → Jumlah record yang diperbarui dalam target.
- **E (Error)** → Jumlah error yang terjadi selama eksekusi.

4.

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	CSV file input	0	0	19	20	0	0	0	0	Finished	0.0s	2,222	-

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?
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

JAWABAN

1. Pada file csv data yang ditampilkan yaitu valuenya not null, sedangkan file excel yang ditampilkan adalah data yang mempunyai value null
2. **-Membaca data dari sumber** (misalnya database, file CSV, atau Excel).
-Mengevaluasi setiap baris data berdasarkan kondisi yang telah ditentukan.
-Memisahkan data menjadi dua aliran:
 - **Data yang memenuhi kondisi** diteruskan ke langkah (step) berikutnya.
 - **Data yang tidak memenuhi kondisi** akan dibuang atau dikirim ke jalur alternatif jika ditentukan.

3.

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	CSV file input	0	0	19	20	0	0	0	0	Finished	0.0s	2,222	-
2	Filter data	0	19	19	0	0	0	0	0	Finished	0.0s	1,267	-
3	Text file output	0	19	19	0	20	0	0	0	Finished	0.3s	78	-
4	Microsoft Excel output	0	0	0	0	0	0	0	0	Finished	0.3s	0	-

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%