

LAPORAN DATA WAREHOUSE

JOBSHEET 1



Oleh :
GEGAS ANUGRAH DERAJAT (2341760140)

JURUSAN TEKNOLOGI INFORMASI
PRODI D-IV SISTEM INFORMASI BISNIS
POLITEKNIK NEGERI MALANG
2024

	A	B	C	D	E	F	G	H
1	Customer,Product,Region,Date, Item unit price , No.Items , Total Sale							
2	Customer 1	Yaris AT	MLG	3/13/2019	340.95	6	2045.7	
3	Customer 1	Etios AT	SBY	4/16/2019	799.95	1	799.95	
4	Customer 3	Avanza AT	MLG	6/18/2019	168.95	7	1182.65	
5	Customer 3	MLG	7/24/2019	168.95	2	337.9		
6	Customer 1	Avanza AT	SBY	8/21/2019	168.95	9	1520.55	
7	Customer 2	Etios AT	KDR	9/7/2019	799.95	7	5599.65	
8	Customer 3	Etios AT	MLG	9/7/2019	799.95	8	6399.6	
9	Customer 2	Avanza AT	MLG	10/15/2019	168.95	15	2534.25	
10	Customer 1	Avanza AT	SBY	9/14/2019	168.95	15	2534.25	
11	Customer 1	Etios AT	SBY	9/25/2019	799.95	2	1599.9	
12	Customer 1	Avanza AT	SBY	1/21/2020	168.95	14	2365.3	
13	Customer 1	Yaris AT	MLG	2/1/2020	340.95	2	681.9	
14	Customer 1	MLG	5/30/2020	340.95	7	2386.65		
15	Customer 1	Etios AT	SBY	5/30/2020	799.95	9	7199.55	
16	Customer 3	Avanza AT	MLG	6/7/2020	168.95	10	1689.5	
17	Customer 3	Yaris AT	MLG	6/9/2020	340.95	4	1363.8	
18	Customer 2	KDR	8/7/2020	799.95	14	11199.3		
19	Customer 2	Avanza AT	KDR	8/8/2020	168.95	12	2027.4	
20	Customer 1	Avanza AT	MLG	8/12/2020	168.95	4	675.8	

Gambar Data Penjualan Mobil

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?

1. Terdapat 7 kolom dalam data.

2. Kolom Customer untuk nama customer, kolom Product untuk nama produk, kolom Region untuk lokasi pembelian, kolom Date untuk tanggal pembelian, kolom Item unit price untuk harga satuan produk, kolom No.Items untuk jumlah produk terjual, dan kolom Total Sale untuk Total nilai penjualan.

3. Ada, terdapat pada baris 5, 9, 14, 18, dan 20

4. Tidak ada

Execution Results

Logging

Execution History

Step Metrics

Performance Graph

Metrics

Preview data

First rows

Last rows

Off

#	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
1	Customer 1	Yaris AT	MLG	Wed Mar 13 00:00:00 ICT 2019	340.95	6	2045.7
2	Customer 1	Etios AT	SBY	Tue Apr 16 00:00:00 ICT 2019	799.95	1	799.95
3	Customer 3	Avanza AT	MLG	Tue Jun 18 00:00:00 ICT 2019	168.95	7	1182.65
4	Customer 3	<null>	MLG	Wed Jul 24 00:00:00 ICT 2019	168.95	2	337.9
5	Customer 1	Avanza AT	SBY	Wed Aug 21 00:00:00 ICT 2019	168.95	9	1520.55
6	Customer 2	Etios AT	KDR	Sat Sep 07 00:00:00 ICT 2019	799.95	7	5599.65
7	Customer 3	Etios AT	MLG	Sat Sep 07 00:00:00 ICT 2019	799.95	8	6399.6
8	Customer 2	Avanza AT	<null>	<null>	168.95	15	2534.25
9	Customer 1	Avanza AT	SBY	Sat Sep 14 00:00:00 ICT 2019	168.95	15	2534.25

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 ?

2024/02/17 06:51:48 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)

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.

1. Tidak sama, terdapat beberapa perbedaan pada beberapa baris.

2. Menyimpan file transformasi, membuka transformasi, meluncurkan transformasi, memulai eksekusi, menjalankan proses transformasi, membaca file csv dan melewati header, memproses data, dan transformasi selesai.

3.

I = Baris yang diambil dari data

O = Baris yang dikeluarkan ke output berikutnya

R = Baris yang dihapus

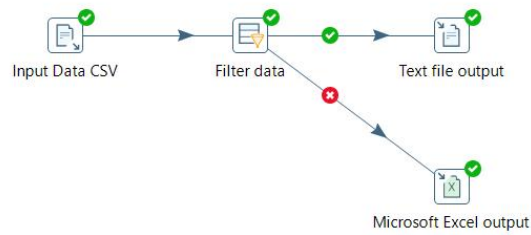
W = Baris data yang ditulis ke tahap berikutnya

U = Pembaruan data

E = Error

4.

NO	Nama	Processor	RAM	Storage	VGA	CPU	Waktu
1	Devin	Intel	4		Intel HD	Intel i5	1 detik
2	Bayu	AMD	8	512	RTX	Ryzen 7	0.5 detik
3	Audric	Intel	8	256	RTX	Intel i5	0.5 detik



Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

025/03/04 06:41:39 - Transformation 1 - Dispatching started for transformation [Transformation 1]
 025/03/04 06:41:39 - Input Data CSV.0 - Header row skipped in file 'C:\Users\Pongo\Downloads\CarDistributionSales.csv'
 025/03/04 06:41:39 - Input Data CSV.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)
 025/03/04 06:41:39 - Filter data.0 - Finished processing (I=0, O=0, R=19, W=19, U=0, E=0)
 025/03/04 06:41:39 - Text file output.0 - Finished processing (I=0, O=15, R=14, W=14, U=0, E=0)
 025/03/04 06:41:39 - Microsoft Excel output.0 - Finished processing (I=0, O=5, R=5, W=5, U=0, E=0)
 025/03/04 06:41:39 - Spoon - The transformation has finished!!

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

1. Pada file csv outputnya adalah data yang lengkap, sedangkan pada file excel outputnya adalah data yang tidak lengkap.

2. Filter row digunakan untuk memisahkan baris, dimana baris yang benar akan dikirim ke Text file output dan baris yang belum benar akan dikirim ke Excel output.

3.

NO	Nama	Processor	RAM	Storage	VGA	CPU	Waktu
1	Devin	Intel	4		Intel HD	Intel i5	1 detik
2	Bayu	AMD	8	512	RTX	Ryzen 7	0.5 detik
3	Audric	Intel	8	256	RTX	Intel i5	0.5 detik