

## JOBSHEET 1

### Data Warehouse Sederhana dan Pengenalan PDI Spoon

Nama : Reza Angelina Febriyanti

Kelas / No Absen : SIB 2A / 27

NIM : 2341760015

---

## TUGAS 1

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	Customer	String		10		IDR	.	,	none
2	Product	String		9		IDR	.	,	none
3	Region	String		3		IDR	.	,	none
4	Date	Date	MM/dd/yyyy			IDR	.	,	none
5	Item unit price	Number	#,.	6	2	IDR	.	,	none
6	No.Items	Integer	#	15	0	IDR	.	,	none
7	Total Sale	Number	#,.	7	2	IDR	.	,	none

*Struktur Data File CarDistributionSales.csv*

1. Berapa jumlah kolom pada data tersebut?

Ada 7 column diantaranya: *Column Customer*, *Column Product*, *Column Region*, *Column Date*, *Column Item unit price*, *Column No.Items*, *Column Total Sale*

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

*Column Customer* berisi *id customer* yang berbeda

*Column Product* berisi *merk mobil*

*Column Region* berisi kota tempat tinggal *customer*

*Column Date* berisi hari, tanggal, bulan, tahun, dan waktu

*Column Item unit price* berisi harga satuan produk

*Column No.Items* berisi jumlah produk yang dibeli

*Column Total Sale* berisi total harga pembelian

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

Ya, terletak pada baris 4;8;12;13;17;20. *Column* yang kosong diantaranya *column product*, *column region*, *column date*

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

## TUGAS 2

#	Customer	Product	Region	Date	Item unit price	No.Items	Total Sale
1	Customer 1	Yaris AT	MLG	Wed Mar 13 00:00:00 WIB 2019	340.95	6	2045.7
2	Customer 1	Etios AT	SBY	Tue Apr 16 00:00:00 WIB 2019	799.95	1	799.95
3	Customer 3	Avanza AT	MLG	Tue Jun 18 00:00:00 WIB 2019	168.95	7	1182.65
4	Customer 3	<null>	MLG	Wed Jul 24 00:00:00 WIB 2019	168.95	2	337.9
5	Customer 1	Avanza AT	SBY	Wed Aug 21 00:00:00 WIB 2019	168.95	9	1520.55
6	Customer 2	Etios AT	KDR	Sat Sep 07 00:00:00 WIB 2019	799.95	7	5599.65
7	Customer 3	Etios AT	MLG	Sat Sep 07 00:00:00 WIB 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 WIB 2019	168.95	15	2534.25
1.	Customer 1	Etios AT	SBY	Wed Sep 25 00:00:00 WIB 2019	799.95	2	1599.9
1.	Customer 1	Avanza AT	SBY	Tue Jan 21 00:00:00 WIB 2020	168.95	14	2365.3
1.	Customer 1	Yaris AT	MLG	Sat Feb 01 00:00:00 WIB 2020	340.95	2	681.9
1.	Customer 1	<null>	<null>	<null>	340.95	7	2386.65
1.	Customer 1	Etios AT	SBY	Sat May 30 00:00:00 WIB 2020	799.95	9	7199.55
1.	Customer 3	Avanza AT	MLG	Sun Jun 07 00:00:00 WIB 2020	168.95	10	1689.5
1.	Customer 3	Yaris AT	MLG	Tue Jun 09 00:00:00 WIB 2020	340.95	4	1363.8
1.	Customer 2	<null>	KDR	Fri Aug 07 00:00:00 WIB 2020	799.95	14	11199.3
1.	Customer 2	Avanza AT	KDR	Sat Aug 08 00:00:00 WIB 2020	168.95	12	2027.4
1.	Customer 1	Avanza AT	<null>	Wed Aug 12 00:00:00 WIB 2020	168.95	4	675.8

*Hasil Execute CarDistributionSales.csv*

1. Apakah data hasil eksekusi sesuai dengan data aslinya? Ya, terdapat 20 baris seperti data awal
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. Menentukan sumber data
  2. Konfigurasi koneksi ke sumber data bila menggunakan database
  3. Menentukan file
  4. Menjalankan transformasi untuk melihat sample data dengan mengklik Preview Rows
  5. Menangani error atau pembersihan data awal apabila diinginkan
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)

I (Input) : Data yang masuk ke step

O (Output) : Data yang dikirim ke step lain

R (Read) : Data yang dibaca ulang

W (Written) : Data yang ditulis ke output

U (Updated) : Data yang diperbarui

E (Error) : Jumlah error dalam proses

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.

Nama	Processor	RAM (GB)	Storage (GB)	VGA	CPU	Time (s)
Reza	AMD Ryzen 7 7730U	16	512	-	AMD Ryzen 7 7730U with Radeon Graphics (16 CPUs), 2.0 GHz	0,0
Louise	Inter(R) Core™ i3	8	256	-	-	0,0
Aqila	AMD Ryzen 5 5600H	8	512	AMD Radeon(TM) Graphics	AMD Ryzen 5 5600H with Radeon Graphics	1
Dahniar	Inter(R) Core™ i5-12450H	8	256	Intel(R) HD Graphics 520	CPU @ 2.40GHz 2.50 GHz	1

Kesimpulan: kecepatan *execute* Luise sama dengan Reza padahal jenis *processor* nya dibawah Dahniar dan Aqila. *Processor* Louise merupakan generasi yang mendukung untuk proses *exexute* data dengan cepat dan *space* RAM yang tersedia masih banyak. Jadi kecepatan *execute* data dipengaruhi jenis *processor* beserta generasinya dan RAM yang masih tersedia.

## TUGAS 3

	A	B	C	D	E	F	G
1	Customer	Product	Region	Date	Item unit price	No. Items	Total Sale
2	Customer 1	Yaris AT	MLG	3/13/2019	340.9	6	2045.7
3	Customer 1	Elios AT	SBY	4/16/2019	800	1	800
4	Customer 3	Avanza AT	MLG	6/18/2019	168.9	7	1182.7
5	Customer 1	Avanza AT	SBY	8/21/2019	168.9	9	1520.5
6	Customer 2	Elios AT	KDR	9/7/2019	800	7	5599.6
7	Customer 3	Elios AT	MLG	9/7/2019	800	8	6399.6
8	Customer 1	Avanza AT	SBY	9/14/2019	168.9	15	2534.2
9	Customer 1	Elios AT	SBY	9/25/2019	800	2	1599.9
10	Customer 1	Avanza AT	SBY	1/21/2020	168.9	14	2365.3
11	Customer 1	Yaris AT	MLG	2/1/2020	340.9	2	681.9
12	Customer 1	Elios AT	SBY	5/30/2020	800	9	7199.6
13	Customer 3	Avanza AT	MLG	6/7/2020	168.9	10	1689.5
14	Customer 3	Yaris AT	MLG	6/9/2020	340.9	4	1363.8
15	Customer 2	Avanza AT	KDR	8/8/2020	168.9	12	2027.4

Data pada baris yang not null pada kolom

[https://drive.google.com/file/d/1\\_JceeNEvipn1eQ0C-EM5oB2XQu\\_haOWT/view?usp=drive\\_link](https://drive.google.com/file/d/1_JceeNEvipn1eQ0C-EM5oB2XQu_haOWT/view?usp=drive_link)

	A	B	C	D	E	F	G
1	Customer	Product	Region	Date	Item unit p	No.Items	Total Sale
2	Customer		MLG	24/07/2019 00:00	0168,95	2,00	00337,90
3	Customer	Avanza AT			0168,95	15,00	02534,25
4	Customer				0340,95	7,00	02386,65
5	Customer		KDR	07/08/2020 00:00	0799,95	14,00	11199,30
6	Customer	Avanza AT		12/08/2020 00:00	0168,95	4,00	00675,80

Data pada baris yang null pada kolom

[https://docs.google.com/spreadsheets/d/12lmkd0pmGX\\_lIEHrHwObx57dZi2POvxG/edit?usp=drive\\_link&ouid=103798971040024129365&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/12lmkd0pmGX_lIEHrHwObx57dZi2POvxG/edit?usp=drive_link&ouid=103798971040024129365&rtpof=true&sd=true)

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

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

*File csv*

	A	B	C	D	E	F	G
1	Customer	Product	Region	Date	Item unit p	No.Items	Total Sale
2	Customer		MLG	24/07/2019 00:00	0168,95	2,00	00337,90
3	Customer	Avanza AT			0168,95	15,00	02534,25
4	Customer				0340,95	7,00	02386,65
5	Customer		KDR	07/08/2020 00:00	0799,95	14,00	11199,30
6	Customer	Avanza AT		12/08/2020 00:00	0168,95	4,00	00675,80

*File excel*

**CSV (Comma-Separated Values)** : format teks biasa dimana data dipisahkan oleh koma; tidak mendukung format sel, warna, atau gaya teks; tidak mendukung multiple sheet untuk menyimpan output yang berbeda

**Excel (XLS/XLSX)** : bisa berupa format teks dan biner dimana data dipisahkan oleh garis; mendukung berbagai fitur seperti warna, font, rumus, format angka, dan grafik; mendukung multiple sheet untuk menyimpan output yang berbeda

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

Filter rows adalah step dalam PDI yang digunakan untuk menyaring (*filtering*) data berdasarkan kondisi tertentu.

Cara Kerja: pertama filter rows menerima input data; kedua mengevaluasi inputan sesuai kondisi yang diinginkan; terakhir memisahkan data sesuai hasil evaluasi dan meneruskan ke step output

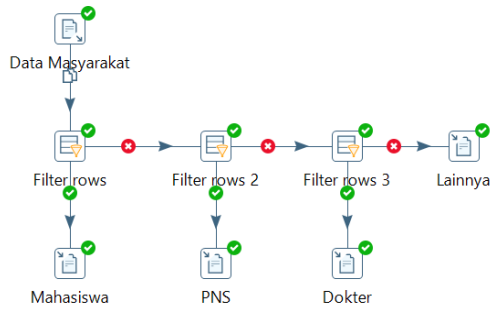
Pada tugas 3 filter rows menyaring data dengan kondisi *Column Customer*, *Column Product*, *Column Region*, *Column Date*, *Column Item unit price*, *Column No.Items*, *Column Total Sale* bernilai *not null*.

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

Nama	Processor	RAM (GB)	Storage (GB)	VGA	CPU	Time (s)
Reza	AMD Ryzen 7 7730U	16	512	-	AMD Ryzen 7 7730U with Radeon Graphics (16 CPUs), 2.0 GHz	0,0 s
Vita	Intel (R) Core (TM) i5-8250U	8	1000 HDD	NVIDIA® GeForce® MX150 2 GB Intel® UHD Graphics 620	AMD Ryzen 5 5600H with Radeon Graphics	0,6 s
Aqila	AMD Ryzen 5 5600H	8	512	AMD Radeon(TM) Graphics	AMD Ryzen 5 5600H with Radeon Graphics	2 s
Dahnier	Intel(R) Core(TM) i5-6300U	8	256	Intel(R) HD Graphics 520	CPU @ 2.40GHz 2.50 GHz	1 s

Kesimpulan: Kecepatan *execute* data Dahnier lebih lambat daripada Vita padahal jenis *processor* nya sama yaitu i5 namun berbeda generasi dan storage Dahnier lebih maju. Jadi kecepatan *execute* data dipengaruhi jenis *processor* dengan generasinya, walau bertipe sama namun tipe generasi ada yang lebih maju maka yang lebih cepat adalah *processor* dengan tipe generasi yang lebih maju

## STUDI KASUS



1. Data masyarakat berprofesi Mahasiswa: diperoleh 3 data penduduk dengan nama Siti Rahayu, Adi Pratama, dan Maya Wulandari

[https://drive.google.com/file/d/17RR6VV-fcmNBIRLWdWvNwQku9PQfEDz/view?usp=drive\\_link](https://drive.google.com/file/d/17RR6VV-fcmNBIRLWdWvNwQku9PQfEDz/view?usp=drive_link)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
2	Siti Rahayu	3503010101010002	Bandung	15-05-1990	Perempuan	Jl. Cendrawasih No. 1	Bandung	003/004	Cikutra	Bandung Wetan	Kristen	Belum Menikah	Mahasiswa
3	Adi Pratama	3503010101010005	Semarang	05-03-1995	Laki-laki	Jl. Pemuda No. 21	Semarang	009/010	Pandananan	Semarang Tengah	Buddha	Belum Menikah	Mahasiswa
4	Maya Wulandari	3.50301E+15	Jayapura	23-10-1993	Perempuan	Jl. Proklamasi No. 8C	Jayapura	039/040	Kota Jayapura	Jayapura	Kristen	Belum Menikah	Mahasiswa

2. Data masyarakat berprofesi PNS: diperoleh 2 data penduduk dengan nama Andi Wijaya dan Agus Santoso

[https://drive.google.com/file/d/1w5O\\_qPqKvvWUcWhUffXu6-L2iLzg1PzD/view?usp=drive\\_link](https://drive.google.com/file/d/1w5O_qPqKvvWUcWhUffXu6-L2iLzg1PzD/view?usp=drive_link)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
2	Andi Wijaya	3503010101010007	Malassar	25-06-1978	Laki-laki	Jl. Jenderal Sudirman	Malassar	015/014	Barombong	Malassar Selatan	Islam	Menikah	PNS
3	Agus Santoso	3503010101010015	Padang	02-04-1982	Laki-laki	Jl. A. Sudirman No. 20	Padang	029/030	Padang Barat	Padang	Islam	Menikah	PNS

3. Data masyarakat berprofesi Dokter: diperoleh 2 data penduduk dengan nama Desi Purnama dan Bambang Kusumo

[https://drive.google.com/file/d/1pHSmcoDxbSUc6P-wqP6HO6ow9fTp\\_Li8/view?usp=drive\\_link](https://drive.google.com/file/d/1pHSmcoDxbSUc6P-wqP6HO6ow9fTp_Li8/view?usp=drive_link)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
2	Desi Purnama	3503010101010008	Palembang	08-04-1993	Perempuan	Jl. Demang Lebar Di	Palembang	015/016	Alang-Alang Lebar	Palembang Selatan	Kristen	Belum Menikah	Dokter
3	Bambang Kusumo	3503010101010017	Banjarasin	25-09-1980	Laki-laki	Jl. A. Yani No. 60	Banjarasin	033/034	Banjarasin Tengah	Banjarasin	Hindu	Menikah	Dokter

4. Data masyarakat berprofesi Lainnya: diperoleh 13 data penduduk dengan nama seperti dibawah ini

[https://drive.google.com/file/d/1ttA\\_SN3vMBtEuiXtv-x8Ykd4XPduCcWS/view?usp=drive\\_link](https://drive.google.com/file/d/1ttA_SN3vMBtEuiXtv-x8Ykd4XPduCcWS/view?usp=drive_link)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
2	Budi Santoso	3503010101010001	Jakarta	01-01-1980	Laki-laki	Jl. Merdeka No. 123	Jakarta	001/002	Merentang	Jakarta Pusat	Islam	Menikah	Wirawasta
3	Joko Susilo	3503010101010003	Surabaya	10-10-1975	Laki-laki	Jl. Diponegoro No. 6	Surabaya	005/006	Gubeng	Surabaya Pusat	Katolik	Menikah	Pegawai Swasta
4	Maya Dewi	3503010101010004	Yogyakarta	20-08-1988	Perempuan	Jl. Malioboro No. 8	Yogyakarta	007/008	Sosomenduran	Yogyakarta Utara	Hindu	Belum Menikah	Guru
5	Rini Setiawan	3503010101010006	Malang	12-12-1982	Perempuan	Jl. Simpang Borembu	Malang	011/012	Dinoyo	Malang Selatan	Kristen	Menikah	Wirawasta
6	Fajar Prabowo	3503010101010009	Medan	30-09-1986	Laki-laki	Jl. Sudirman No. 50	Medan	017/018	Sei Sikambing	Medan Timur	Hindu	Menikah	Wirawasta
7	Rina Sari	3.50301E+15	Denpasar	17-07-1992	Perempuan	Jl. Gatot Subroto No. Denpasar	Denpasar	019/020	Dauh Puri Kauh	Denpasar Utara	Buddha	Belum Menikah	Freelancer
8	Ahmad Subhan	3503010101010011	Surakarta	03-07-1991	Laki-laki	Jl. Pahlawan No. 55	Surakarta	021/022	Jebres	Surakarta Timur	Islam	Belum Menikah	Wirawasta
9	Dian Pertiwi	3503010101010012	Maluku	10-12-1985	Perempuan	Jl. Pattimura No. 10	Ambon	023/024	Baguala	Ambon	Kristen	Menikah	Guru
10	Rudi Hartono	3503010101010013	Pontianak	28-05-1973	Laki-laki	Jl. Ahmad Yani No. 1	Pontianak	025/026	Pontianak Barat	Pontianak Kota	Hindu	Menikah	Wirawasta
11	Lina Fitriani	3503010101010014	Manado	15-08-1994	Perempuan	Jl. Wolter Monginsidi	Manado	027/028	Mapangot	Manado Selatan	Kristen	Belum Menikah	Akuntan
12	Rina Agudina	3503010101010016	Bengkulu	19-11-1990	Perempuan	Jl. Imam Bonjol No.	Bengkulu	031/032	Ratu Samban	Bengkulu Tengah	Kristen	Menikah	Wirawasta
13	Nia Dewi Lestari	3503010101010018	Manokwari	14-06-1987	Perempuan	Jl. Diponegoro No. 4	Manokwari	035/036	Manokwari Barat	Manokwari	Kristen	Belum Menikah	Penulis
14	Doni Setiawan	3503010101010019	Palu	07-03-1977	Laki-laki	Jl. Soekarno-Hatta N	Palu	037/038	Palu Barat	Palu	Islam	Menikah	Wirawasta