



Name : Azaria Cindy Sahasika
Number Id : 2341760169 / 06
Class : 2G – Business Information System
Lesson : Advanced Web Programing
Github Link : <https://github.com/azariacindy>

Jobsheet 1

Data Warehouse Sederhana dan Pengenalan PDI Spoon

Study Case:

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.

Tugas 1

Analisa lah data tersebut!

1. Berapa jumlah kolom pada data tersebut?

	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,,,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,,,,,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,,8/12/2020,168.95,4,675.8							

➔ Jumlah kolom ada 7 yaitu kolom 'Customer', 'Product', 'Region', 'Date', 'Item unit price', 'No.Items' dan 'Total Sale'.

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

- a. Customer: nama pelanggan
- b. Product: nama product
- c. Region: wilayah tempat transaksi dilakukan
- d. Date: tanggal transaksi
- e. Item unit price: harga per unit product
- f. No. Item: jumlah unit produk yang dibeli dalam transaksi
- g. Total Sale: total transaksi dari perkalian 'item unit price' dan 'no. item'

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

➔ Ada beberapa baris yang isi kolomnya tidak lengkap, seperti pada kolom 'product', 'region' dan 'date'.

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

➔ Pada beberapa kolom yang kosong atau 'null' otomatis akan memiliki tipe data yang berbeda. Tetapi pada dasarnya tidak ada tipe data yang berbeda dpada kolom yang sama.

Tugas 2

Pengambilan Data (Extract)

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

#	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
10	Customer 1	Etios AT	SBY	Wed Sep 25 00:00:00 WIB 2019	799.95	2	1599.9
11	Customer 1	Avanza AT	SBY	Tue Jan 21 00:00:00 WIB 2020	168.95	14	2365.3
12	Customer 1	Yaris AT	MLG	Sat Feb 01 00:00:00 WIB 2020	340.95	2	681.9
13	Customer 1	<null>	<null>	<null>	340.95	7	2386.65
15	Customer 3	Avanza AT	MLG	Sun Jun 07 00:00:00 WIB 2020	168.95	10	1689.5
16	Customer 3	Yaris AT	MLG	Tue Jun 09 00:00:00 WIB 2020	340.95	4	1363.8
17	Customer 2	<null>	KDR	Fri Aug 07 00:00:00 WIB 2020	799.95	14	11199.3
18	Customer 2	Avanza AT	KDR	Sat Aug 08 00:00:00 WIB 2020	168.95	12	2027.4
19	Customer 1	Avanza AT	<null>	Wed Aug 12 00:00:00 WIB 2020	168.95	4	675.8

➔ Hasil eksekusi sama 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?

- Membaca sumber data
- Membersihkan data
- Transformasi data
- Filtering dan sorting
- Memvalidasi data sebelum diproses
- Data dicatat pada file csv baru

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/jumlah data yang diambil
- O = output/jumlah data yang dihasilkan

- R = read/jumlah data yang dibaca
 - W = write/jumlah data yang berhasil dicatat
 - U = update/jumlah data yang diperbarui
 - E = error/jumlah error selama data diproses
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	Waktu	Processor	RAM	Storage	VGA	CPU
Cindy	6s	Intel i5-10300H	12 GB	HDD 1 TB	NVIDIA GTX 1050	60%
Evan	5s	Intel i7-10750H	16 GB	SSD NVMe 512 GB	NVIDIA GTX 1650 Ti	70%
Kiya	7s	Intel Celeron N4020	6 GB	HDD	Integrated	50%
Afif	6s	AMD Ryzen 5 4600H	8 GB	SSD SATA 256 GB	Integrated Vega 6	70%

Tugas 3

Filter Data (Transform) dan Pengemasan Data (Load)

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

Customer	Product	Region	Date	Item unit	No.Items	Total Sale
Customer 1	Yaris AT	MLG	3/13/2019	340.9	6	2045.7
Customer 1	Etios AT	SBY	4/16/2019	800	1	800
Customer 3	Avanza AT	MLG	6/18/2019	168.9	7	1182.7
Customer 1	Avanza AT	SBY	8/21/2019	168.9	9	1520.5
Customer 2	Etios AT	KDR	09/07/2019	800	7	5599.6
Customer 3	Etios AT	MLG	09/07/2019	800	8	6399.6
Customer 1	Avanza AT	SBY	9/14/2019	168.9	15	2534.2
Customer 1	Etios AT	SBY	9/25/2019	800	2	1599.9
Customer 1	Avanza AT	SBY	1/21/2020	168.9	14	2365.3
Customer 1	Yaris AT	MLG	02/01/2020	340.9	2	681.9
Customer 1	Etios AT	SBY	5/30/2020	800	9	7199.6
Customer 3	Avanza AT	MLG	06/07/2020	168.9	10	1689.5
Customer 3	Yaris AT	MLG	06/09/2020	340.9	4	1363.8
Customer 2	Avanza AT	KDR	08/08/2020	168.9	12	2027.4

➔ File csv berisi data yang tidak ada nilai null

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

➔ File xls berisi data yang ada nilai null atau tidak lengkap

2. Jelaskan apa yang terjadi pada proses Filter rows!

➔ Langkah transformasi yang bertujuan untuk menyaring data berdasarkan kondisi tertentu. Jika true akan diteruskan ke langkah selanjutnya, jika false maka dapat dihapus atau diteruskan ke langkah yang lainnya.

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 table

Nama	Waktu	Processor	RAM	Storage	VGA	CPU
Cindy	6s	Intel i5-10300H	12 GB	HDD 1 TB	NVIDIA GTX 1050	60%
Evan	5s	Intel i7-10750H	16 GB	SSD NVMe 512 GB	NVIDIA GTX 1650 Ti	70%
Kiya	7s	Intel Celeron N4020	6 GB	HDD	Integrated	50%
Afif	6s	AMD Ryzen 5 4600H	8 GB	SSD SATA 256 GB	Integrated Vega 6	70%

Studi 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 kondisikan 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. Dari 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 tersebut akan digunakan untuk analisa di masing-masing wilayah dan juga Dashboard untuk membantu mengambil keputusan. Pak Pascanowo mendapatkan data penduduk yang akan pindah ke wilayahnya melalui link: <https://github.com/dik4rizky/datasources/blob/82821b8ec341176f979956c84b79720d4aead012/dataDummyKependudukan.csv>

Bantulah Pak Pascanowo untuk memisahkan 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.

Jawaban Studi Kasus

Spoon - data penduduk (changed)

File Edit View Action Tools Help

Connect

View Design

Search

Transformations

- data penduduk
 - Run configurati
 - Database connx
 - Steps
 - Hops
 - Partition schem
 - Slave server
 - Kettle cluster sc
 - VFS Connector

CSV file input

Filter data mhs

mhs

Filter data pns

pns

Filter data dokter

dokter

Filter data lainnya

lainnya

Filter rows

Step name: Filter data mhs

Send 'true' data to step: mhs

Send 'false' data to step:

The condition:

Pekerjaan * Mahasiswa (String)

Text file output

Step name: mhs

File Content Fields

Filename: D:/cooleyah/smstr4/data warehouse/assignment/MasyarakatMa

Pass output to servlet

Create Parent folder

Do not create file at start

Accept file name from field?

File name field

Extension: csv

Include stepnr in filename?

Include partition nr in filename?

Include dste in filename?

Include time in filename?

Specify Date time format

Date time format

Show filename(s)...

Add filenames to result

Filter rows

Step name: filter data pns

Send 'true' data to step: pns

Send 'false' data to step:

The condition:

=

PNS (String)

OK Cancel

Text file output

Step name: pns

File Content Fields

Filename: D:/cooleyah/smstr4/data warehouse/assignment/MasyarakatPNS Browse...

Pass output to servlet: ☐

Create Parent folder: ☒

Do not create file at start: ☐

Accept file name from field: ☐

File name field:

Extension: csv

Include stepnr in filename: ☐

Include partition nr in filename: ☐

Include date in filename: ☐

Include time in filename: ☐

Specify Date time format:

Date time format:

Show filename(s)...

Add filenames to result: ☒

OK Cancel

Filter rows

Step name: filter data dokter

Send 'true' data to step: dokter

Send 'false' data to step:

The condition:

=

Dokter (String)

OK Cancel

Text file output

Step name: dokter

File Content Fields

Filename: D:/cooleyah/smstr4/data warehouse/assignment/MasyarakatDok Browse...

Pass output to servlet: ☐

Create Parent folder: ☒

Do not create file at start: ☐

Accept file name from field: ☐

File name field:

Extension: csv

Include stepnr in filename: ☐

Include partition nr in filename: ☐

Include date in filename: ☐

Include time in filename: ☐

Specify Date time format:

Date time format:

Show filename(s)...

Add filenames to result: ☒

OK Cancel

Filter rows

Step name

Filter data lainnya

Send 'true' data to step:

lainnya

Send 'false' data to step:

The condition:

+

Pekerjaan <> [Mahasiswa]

AND

Pekerjaan <> [PNS]

AND

Pekerjaan <> [Dokter]

Help

OK

Cancel

Text file output

Step name

lainnya

File

Content

Fields

Filename

D:/cooleyah/smstr4/data warehouse/assignment/MasyarakatLain

Browse...

Pass output to servlet

Create Parent folder

Do not create file at start

Accept file name from field?

File name field

Extension

csv

Include stepnr in filename?

Include partition nr in filename?

Include date in filename?

Include time in filename?

Specify Date time format

Date time format

Show filename(s)...

Add filenames to result

Help

OK

Cancel

Output hasil data:

- MasyarakatMahasiswa

Nama	NIK	Tempat La	Tanggal La	Jenis Kela	Alamat	Kota	RT/RW	Kelurahan	Kecamata	Agama	Status Per	Pekerjaan
Siti Rahay	3,5E+15	Bandung	#####	Perempu	Jl. Cendra	Bandung	003/004	Cikutra	Bandung	Kristen	Belum Me	Mahasisw
Adi Pratar	3,5E+15	Semarang	#####	Laki-laki	Jl. Pemuda	Semarang	009/010	Pandanar	Semarang	Buddha	Belum Me	Mahasisw
Maya Wul	3,5E+15	Jayapura	#####	Perempu	Jl. Proklan	Jayapura	039/040	Kota Jaya	Jayapura	Kristen	Belum Me	Mahasisw

- MasyarakatPNS

Nama	NIK	Tempat La	Tanggal La	Jenis Kela	Alamat	Kota	RT/RW	Kelurahan	Kecamata	Agama	Status Per	Pekerjaan
Andi Wija	3,5E+15	Makassar	#####	Laki-laki	Jl. Jender	Makassar	013/014	Barombor	Makassar	Islam	Menikah	PNS
Agus Sant	3,5E+15	Padang	#####	Laki-laki	Jl. Sudirm	Padang	029/030	Padang Ba	Padang	Islam	Menikah	PNS

- MasyarakatDokter

Nama	NIK	Tempat La	Tanggal La	Jenis Kela	Alamat	Kota	RT/RW	Kelurahan	Kecamata	Agama	Status Per	Pekerjaan
Desi Purni	3,5E+15	Palemban	#####	Perempu	Jl. Deman	Palemban	015/016	Alang-Ala	Palemban	Kristen	Belum Me	Dokter
Bambang	3,5E+15	Banjarmas	#####	Laki-laki	Jl. A. Yani	Banjarmas	033/034	Banjarmas	Banjarmas	Hindu	Menikah	Dokter

- MasyarakatLainnya

Nama	NIK	Tempat La	Tanggal La	Jenis Kela	Alamat	Kota	RT/RW	Kelurahan	Kecamata	Agama	Status Per	Pekerjaan
Budi Santo	3,5E+15	Jakarta	#####	Laki-laki	Jl. Merdeka	Jakarta	001/002	Menteng	Jakarta Pu	Islam	Menikah	Wiraswast
Joko Susil	3,5E+15	Surabaya	#####	Laki-laki	Jl. Dipone	Surabaya	005/006	Gubeng	Surabaya	Katolik	Menikah	Pegawai S
Maya Dew	3,5E+15	Yogyakarta	#####	Perempu	Jl. Maliob	Yogyakarta	007/008	Sosromen	Yogyakarta	Hindu	Belum Me	Guru
Rini Setia	3,5E+15	Malang	#####	Perempu	Jl. Simpan	Malang	011/012	Dinoyo	Malang Se	Kristen	Menikah	Wiraswast
Fajar Prab	3,5E+15	Medan	#####	Laki-laki	Jl. Sudirm	Medan	017/018	Sei Sikam	Medan Tir	Hindu	Menikah	Wiraswast
Rina Sari	3,5E+15	Denpasar	#####	Perempu	Jl. Gatot S	Denpasar	019/020	Dauh Puri	Denpasar	Buddha	Belum Me	Freelance
Ahmad Su	3,5E+15	Surakarta	#####	Laki-laki	Jl. Pahlaw	Surakarta	021/022	Jebres	Surakarta	Islam	Belum Me	Wiraswast
Dian Perti	3,5E+15	Maluku	#####	Perempu	Jl. Pattim	Ambon	023/024	Baguala	Ambon	Kristen	Menikah	Guru
Rudi Harto	3,5E+15	Pontianak	#####	Laki-laki	Jl. Ahmad	Pontianak	025/026	Pontianak	Pontianak	Hindu	Menikah	Wiraswast
Lina Fitria	3,5E+15	Manado	#####	Perempu	Jl. Wolter	Manado	027/028	Mapanget	Manado S	Kristen	Belum Me	Akuntan
Rina Agus	3,5E+15	Bengkulu	#####	Perempu	Jl. Imam B	Bengkulu	031/032	Ratu Sami	Bengkulu	Kristen	Menikah	Wiraswast
Nia Dewi	3,5E+15	Manokwa	#####	Perempu	Jl. Dipone	Manokwa	035/036	Manokwa	Manokwa	Kristen	Belum Me	Penulis
Doni Setia	3,5E+15	Palu	#####	Laki-laki	Jl. Soekarr	Palu	037/038	Palu Barat	Palu	Islam	Menikah	Wiraswast