

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 diharapkan 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/CatDistributionSales.csv>

	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		

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

2. 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.

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

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

4. 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

A. Pengambilan Data (Extract)

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
1..	Customer 1	Etios AT	SBY	Wed Sep 25 00:00:00 ICT 2019	799.95	2	1599.9
1..	Customer 1	Avanza AT	SBY	Tue Jan 21 00:00:00 ICT 2020	168.95	14	2365.3
1..	Customer 1	Yaris AT	MLG	Sat Feb 01 00:00:00 ICT 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 ICT 2020	799.95	9	7199.55
1..	Customer 3	Avanza AT	MLG	Sun Jun 07 00:00:00 ICT 2020	168.95	10	1689.5
1..	Customer 3	Yaris AT	MLG	Tue Jun 09 00:00:00 ICT 2020	340.95	4	1363.8
1..	Customer 2	<null>	KDR	Fri Aug 07 00:00:00 ICT 2020	799.95	14	11199.3
1..	Customer 2	Avanza AT	KDR	Sat Aug 08 00:00:00 ICT 2020	168.95	12	2027.4
1..	Customer 1	Avanza AT	<null>	Wed Aug 12 00:00:00 ICT 2020	168.95	4	675.8




TUGAS 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?

Yes, the executed data matches the original data.

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?

Execution Results

Logging	Execution History	Step Metrics	Performance Graph	Metrics	Preview data
  					
2025/03/07 14:18:33 - Spoon - Save file as...					
2025/03/07 14:18:33 - Spoon - Transformation opened.					
2025/03/07 14:18:33 - Spoon - Launching transformation [Transformation 1]...					
2025/03/07 14:18:33 - Spoon - Started the transformation execution.					
2025/03/07 14:18:33 - Transformation 1 - Dispatching started for transformation [Transformation 1]					
2025/03/07 14:18:33 - CSV file input.0 - Header row skipped in file 'C:\SEMESTER 4\DATA WAREHOUSE\TUGAS\CarDistributionSales.csv'					
2025/03/07 14:18:33 - CSV file input.0 - Finished processing (I=20, O=0, R=0, W=19, U=0, E=0)					
2025/03/07 14:18:33 - Spoon - The transformation has finished!!					

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.

3. 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.

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.

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

B. Filter Data (Transform) dan Pengemasan data (Load)

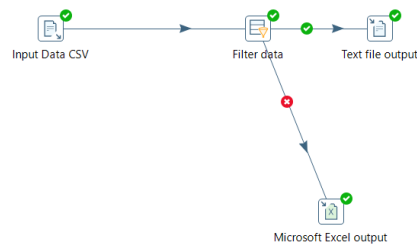
The screenshot shows the "Filter rows" dialog box with the following configuration:

- Step name:** filter data
- Send 'true' data to step:** Text file output
- Send 'false' data to step:** Microsoft Excel output
- The condition:** A field icon (rectangle) followed by a plus sign (+).

The condition expression is as follows:

```
Customer IS NOT NULL
AND
(
    Product IS NOT NULL
    AND
    (
        Region IS NOT NULL
        AND
        (
            Date IS NOT NULL
            AND
            (
                Item unit price IS NOT NULL
                AND
                (
                    No. Items IS NOT NULL
                    AND
                    (
                        Total Sale IS NOT NULL
                        null = {}
                    )
                )
            )
        )
    )
)
```

At the bottom of the dialog are buttons for Help, OK, and Cancel.



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
1..	Customer 1	Etios AT	SBY	Wed Sep 25 00:00:00 ICT 2019	799.95	2	1599.9
1..	Customer 1	Avanza AT	SBY	Tue Jan 21 00:00:00 ICT 2020	168.95	14	2365.3
1..	Customer 1	Yaris AT	MLG	Sat Feb 01 00:00:00 ICT 2020	340.95	2	681.9

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.

2. 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.

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

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.

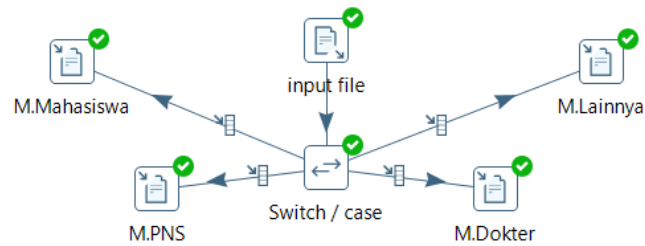
Pak Pascanowo akan mendapatkan data dari pemerintah pusat tentang Masyarakat Konoha yang akan pindah ke IKN. dari data tersebut Pak Pascanowo akan membagi menjadi 4 kelompok,

- **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/dummyKependudukan.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.



Switch / case

Step name: Switch / case

Field name to switch: Pekerjaan

Use string contains comparison: ☐

Case value data type: String

Case value conversion mask:

Case value decimal symbol:

Case value grouping symbol:

#	Value	Target step
1	Mahasiswa	M.Mahasiswa
2	PNS	M.PNS
3	Dokter	M.Dokter

Default target step: M.Lainnya

Help OK Cancel

Rows of step: M.Mahasiswa (3 rows)

#	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
1	Siti Rahayu	3503010101010002	Bandung	15-05-1990	Perempuan	Jl. Cendrawasih No. 45	Bandung	003/004	Cikutra	Bandung Wetan	Kristen	Belum Menikah	Mahasiswa
2	Adi Pratama	3503010101010005	Semarang	05-03-1995	Laki-laki	Jl. Pemuda No. 21	Semarang	009/010	Pandanaran	Semarang Tengah	Buddha	Belum Menikah	Mahasiswa
3	Maya Wulandari	3503010101010020	Jayapura	23-10-1993	Perempuan	Jl. Proklamasi No. 80	Jayapura	039/040	Kota Jayapura	Jayapura	Kristen	Belum Menikah	Mahasiswa

Rows of step: M.PNS (2 rows)

#	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
1	Andi Wijaya	3503010101010007	Makassar	25-06-1978	Laki-laki	Jl. Jenderal Sudirman No. 10	Makassar	013/014	Barombong	Makassar Selatan	Islam	Menikah	PNS
2	Agus Santoso	3503010101010015	Padang	02-04-1982	Laki-laki	Jl. Sudirman No. 20	Padang	029/030	Padang Barat	Padang	Islam	Menikah	PNS

Rows of step: M.Dokter (2 rows)

#	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
1	Desi Purnama	3503010101010008	Palembang	08-04-1993	Perempuan	Jl. Demang Lebar Daun No. 20	Palembang	015/016	Alang-Alang Lebar	Palembang Selatan	Kristen	Belum Menikah	Dokter
2	Bambang Kusumo	3503010101010017	Bandjarmasin	25-09-1980	Laki-laki	Jl. A. Yani No. 60	Bandjarmasin	033/034	Bandjarmasin Tengah	Bandjarmasin	Hindu	Menikah	Dokter

Rows of step: M.Lainnya (13 rows)

#	Nama	NIK	Tempat Lahir	Tanggal Lahir	Jenis Kelamin	Alamat	Kota	RT/RW	Kelurahan	Kecamatan	Agama	Status Perkawinan	Pekerjaan
1	Budi Santoso	3503010101010001	Jakarta	01-01-1980	Laki-laki	Jl. Merdeka No. 123	Jakarta	001/002	Menteng	Jakarta Pusat	Islam	Menikah	Wiraswasta
2	Joko Susilo	3503010101010003	Surabaya	10-10-1975	Laki-laki	Jl. Diponegoro No. 67	Surabaya	005/006	Gubeng	Surabaya Pusat	Katolik	Menikah	Pegawai Swasta
3	Maya Dewi	3503010101010004	Yogyakarta	20-08-1988	Perempuan	Jl. Malioboro No. 8	Yogyakarta	007/008	Sosromenduran	Yogyakarta Utara	Hindu	Belum Menikah	Guru
4	Rini Setiawan	3503010101010006	Malang	12-12-1982	Perempuan	Jl. Sempang Borobudur No. 15	Malang	011/012	Dinoyo	Malang Selatan	Kristen	Menikah	Wiraswasta
5	Fajar Prabowo	3503010101010009	Medan	30-09-1986	Laki-laki	Jl. Sudirman No. 50	Medan	017/018	Sei Sikambing	Medan Timur	Hindu	Menikah	Wiraswasta
6	Rina Sari	3503010101010010	Denpasar	17-07-1992	Perempuan	Jl. Gatot Subroto No. 30	Denpasar	019/020	Dauh Puri Kauh	Denpasar Utara	Buddha	Belum Menikah	Freelancer
7	Ahmad Subhan	3503010101010011	Surakarta	03-07-1991	Laki-laki	Jl. Pahlawan No. 55	Surakarta	021/022	Jebres	Surakarta Timur	Islam	Belum Menikah	Wiraswasta
8	Dian Pertiwi	3503010101010012	Maluku	10-12-1985	Perempuan	Jl. Pattimura No. 10	Ambon	023/024	Baguala	Ambon	Kristen	Menikah	Guru
9	Rudi Hartono	3503010101010013	Pontianak	28-05-1973	Laki-laki	Jl. Ahmad Yani No. 30	Pontianak	025/026	Pontianak Barat	Pontianak Kota	Hindu	Menikah	Wiraswasta
1..	Lina Fitriani	3503010101010014	Manado	15-08-1994	Perempuan	Jl. Wolter Monginsidi No. 17	Manado	027/028	Mapanget	Manado Selatan	Kristen	Belum Menikah	Akuntan
1..	Rina Agustina	3503010101010016	Bengkulu	19-11-1990	Perempuan	Jl. Imam Bonjol No. 45	Bengkulu	031/032	Ratu Samban	Bengkulu Tengah	Kristen	Menikah	Wiraswasta
1..	Nia Dewi Lestari	3503010101010018	Manokwari	14-06-1987	Perempuan	Jl. Diponegoro No. 40	Manokwari	035/036	Manokwari Barat	Manokwari	Kristen	Belum Menikah	Penulis
1..	Doni Setiawan	3503010101010019	Palu	07-03-1977	Laki-laki	Jl. Soekarno-Hatta No. 70	Palu	037/038	Palu Barat	Palu	Islam	Menikah	Wiraswasta

