


Nama: Adam Zannuba	 Praktikum Data Warehouse	MODUL 9
NIM: 065001900022		Nama Dosen: Ir. Teddy Siswanto, MMSi
Hari/Tanggal: 6 Juni 2022		Nama Asisten Labratorium: 1. Azhar Rizki Zulma 065001900001 2. Nadiya Amanda Rizkania 064001900003

Proyek Akhir Praktikum I

1. Teori Singkat

Data warehouse adalah jenis sistem manajemen data yang dirancang untuk memungkinkan dan mendukung kegiatan business intelligence (BI), terutama analitik. Gudang data semata-mata dimaksudkan untuk melakukan kueri dan analisis dan sering berisi sejumlah besar data historis. Data dalam gudang data biasanya berasal dari berbagai sumber seperti file log aplikasi dan aplikasi transaksi. Gudang data memusatkan dan mengkonsolidasikan sejumlah besar data dari berbagai sumber. Kemampuan analitisnya memungkinkan organisasi untuk memperoleh wawasan bisnis yang berharga dari data mereka untuk meningkatkan pengambilan keputusan. Seiring waktu, ia membangun catatan sejarah yang dapat sangat berharga bagi para ilmuwan data dan analis bisnis. Karena kemampuan ini, gudang data dapat dianggap sebagai "sumber kebenaran tunggal" organisasi.

Cleansing Data

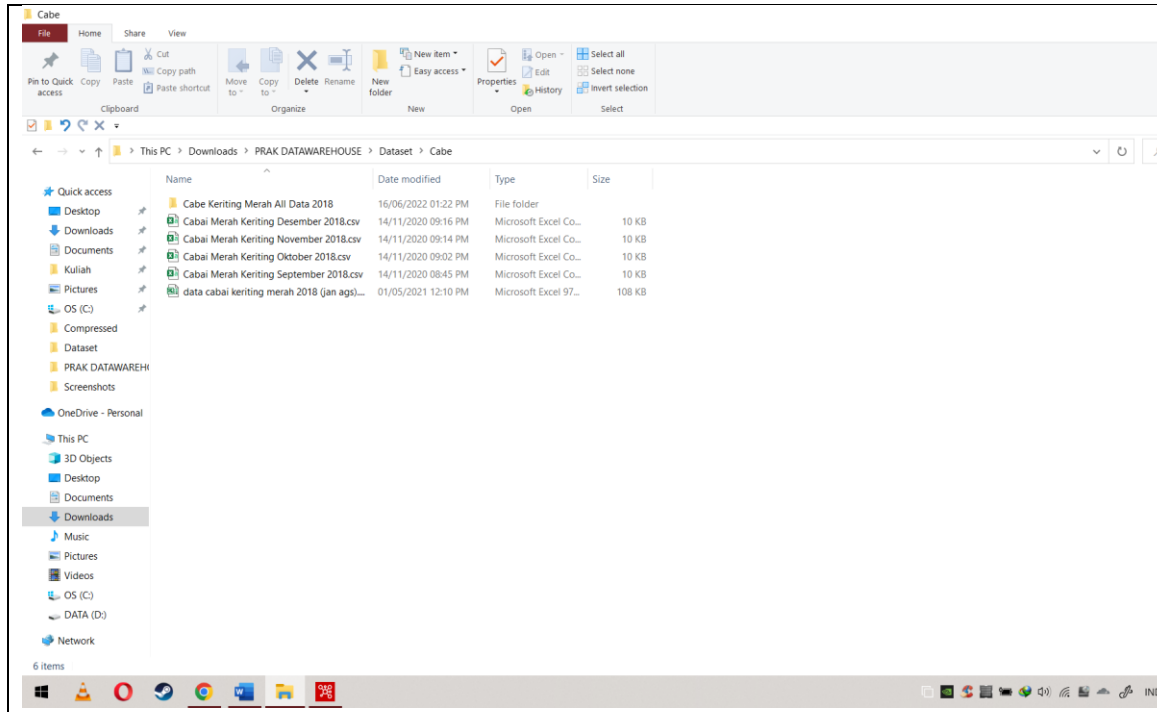
Cleansing Data adalah suatu proses mendeteksi dan memperbaiki (atau menghapus) data set, tabel, dan database yang korup atau tidak akurat. Data tidak akurat tersebut berupa data yang kosong dan tidak memiliki nilai atau berisi nilai NULL/NaN. Tujuan dari cleansing data ialah agar ketika data diproses nanti, tidak terjadi adanya error dari pengolahan data yang dilakukan akibat data yang tidak lengkap ataupun data hilang (missing value).

Berikut ini adalah langkah langkah dalam melakukan cleansing data menggunakan Spoon Pentaho:



1. Data Preparation

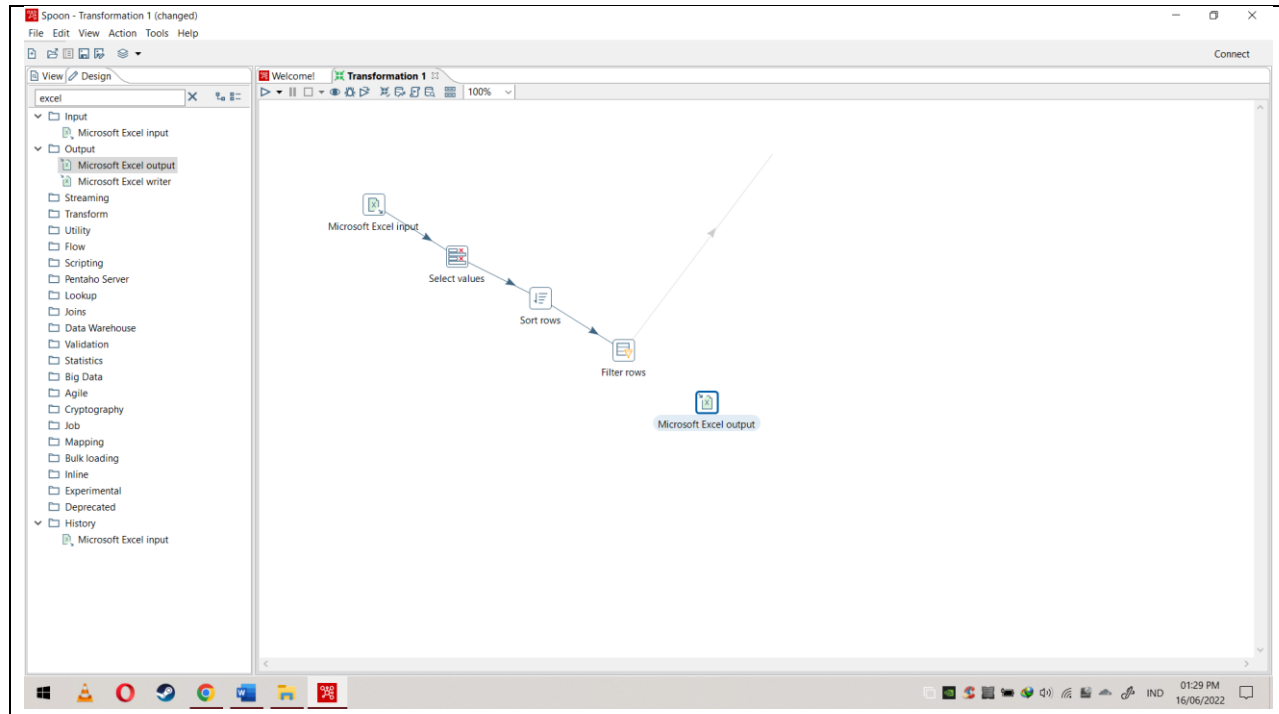
Unduh serta ekstrak datanya dan satukan kedalam folder seperti pada gambar dibawah.



2. Excel Input

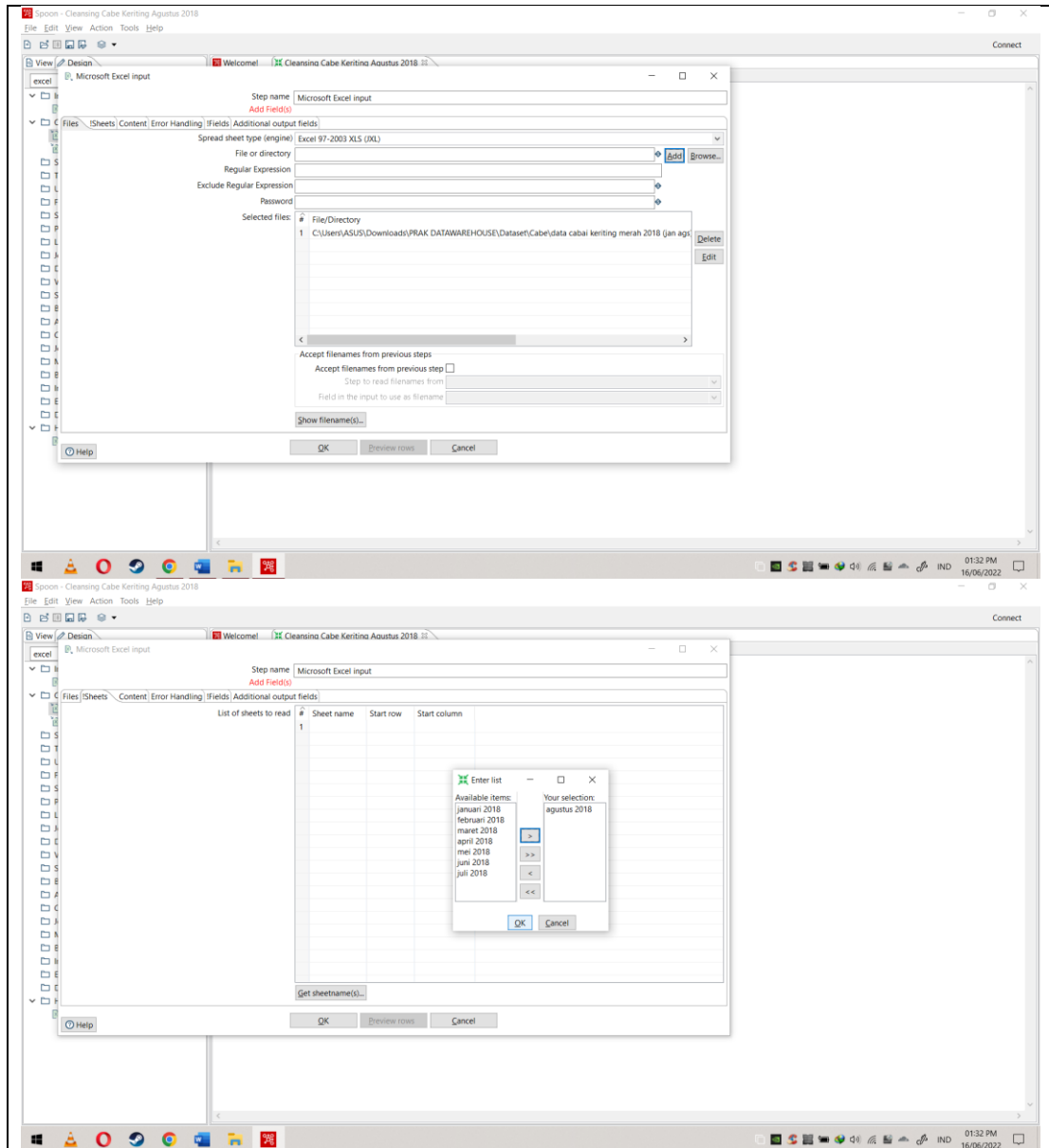
- Buka Pentaho Spoon Data Integration lalu buat struktur seperti pada gambar dibawah ini lalu simpan dengan format ***Cleansing Cabai Merah Keriting [Bulan] [Tahun].ktr*** seperti ***Cleansing Cabai Merah Keriting Agustus 2018.ktr***

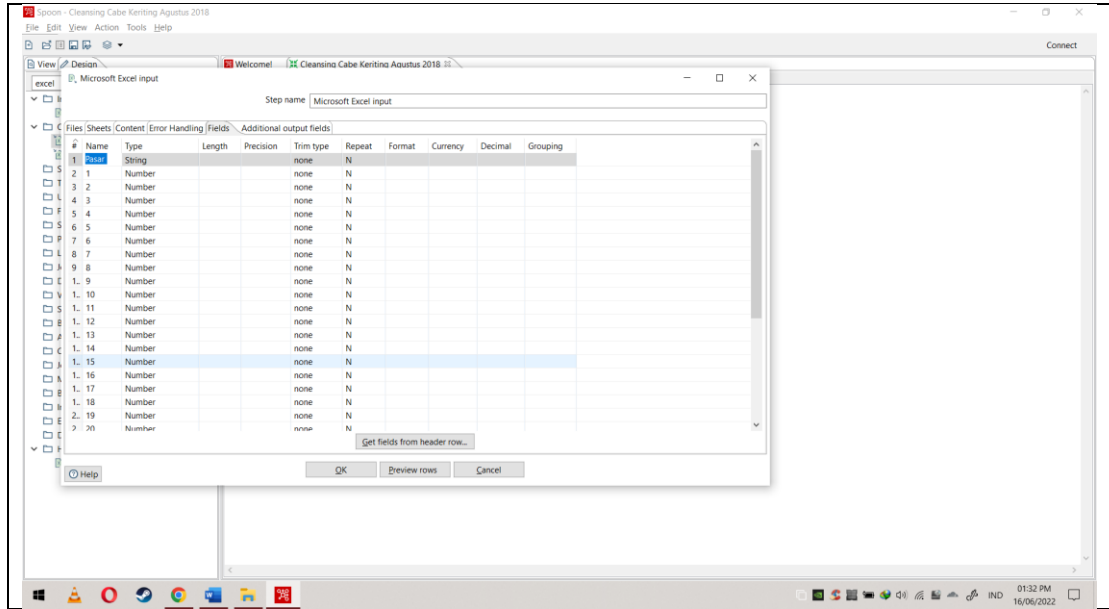




- Berikut ini adalah konfigurasi pada step *Microsoft Excel Input* (Files, Sheets, Fields)

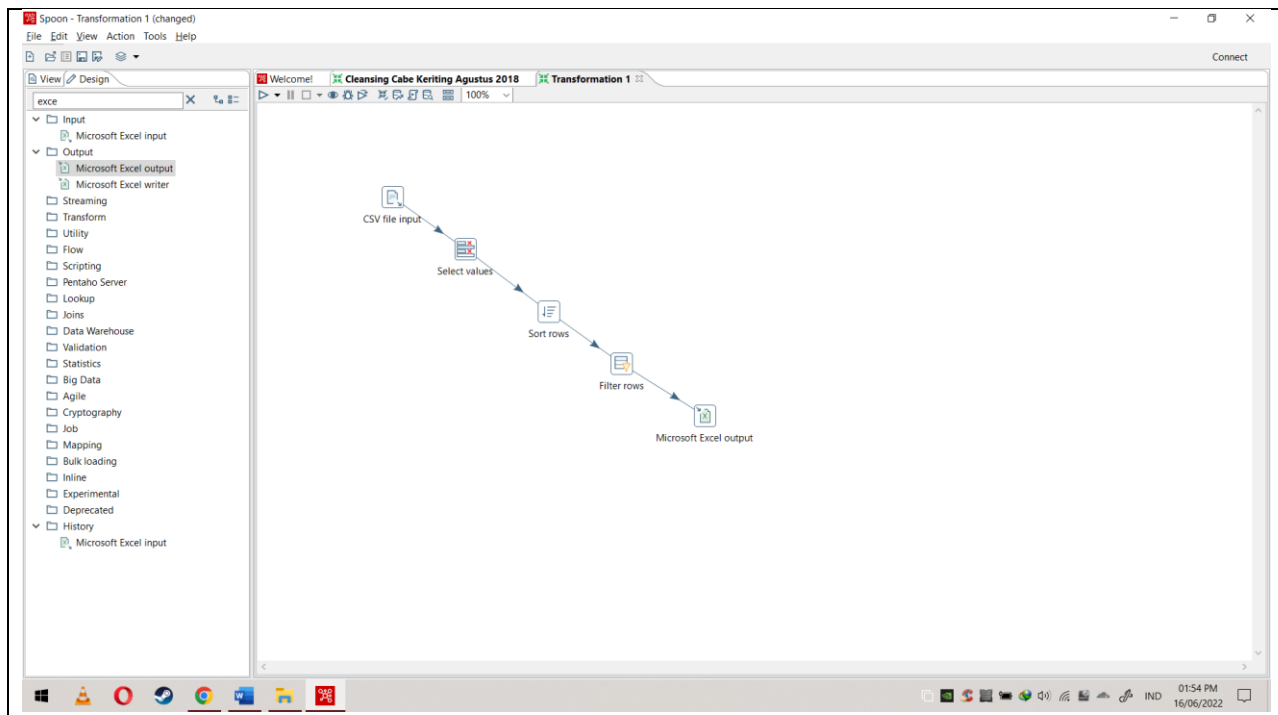




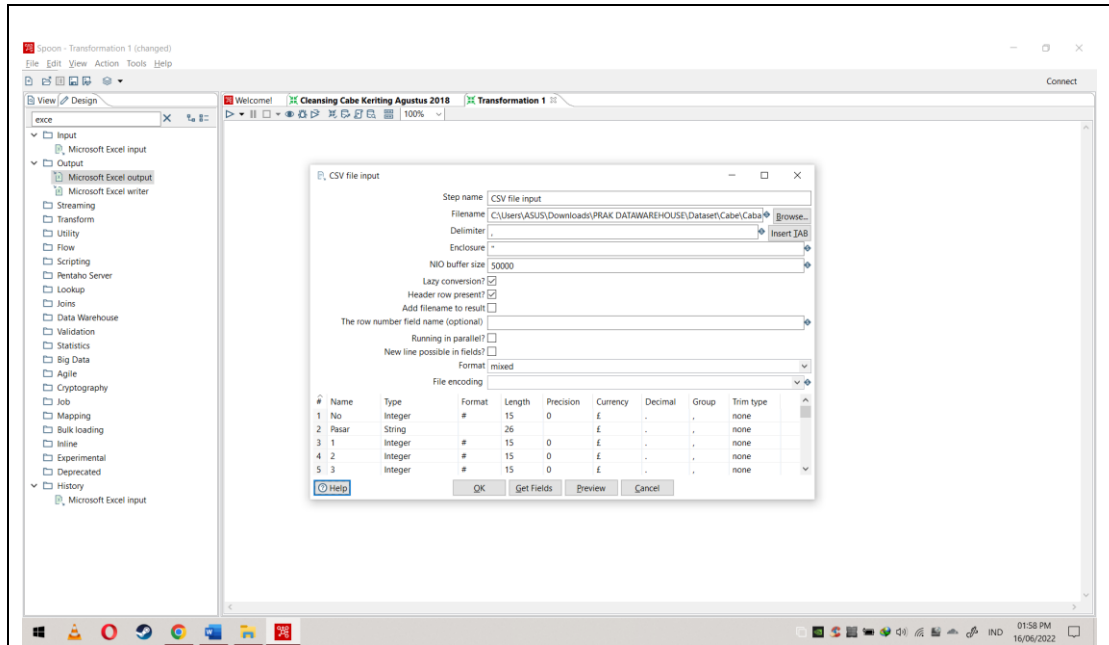


3. CSV Input

- Buka Pentaho Spoon Data Integration lalu buat struktur seperti pada gambar dibawah ini lalu simpan dengan format ***Cleansing Cabe Merah Keriting [Bulan] [Tahun].ktr*** seperti ***Cleansing Cabe Merah Keriting Oktober 2018.ktr***

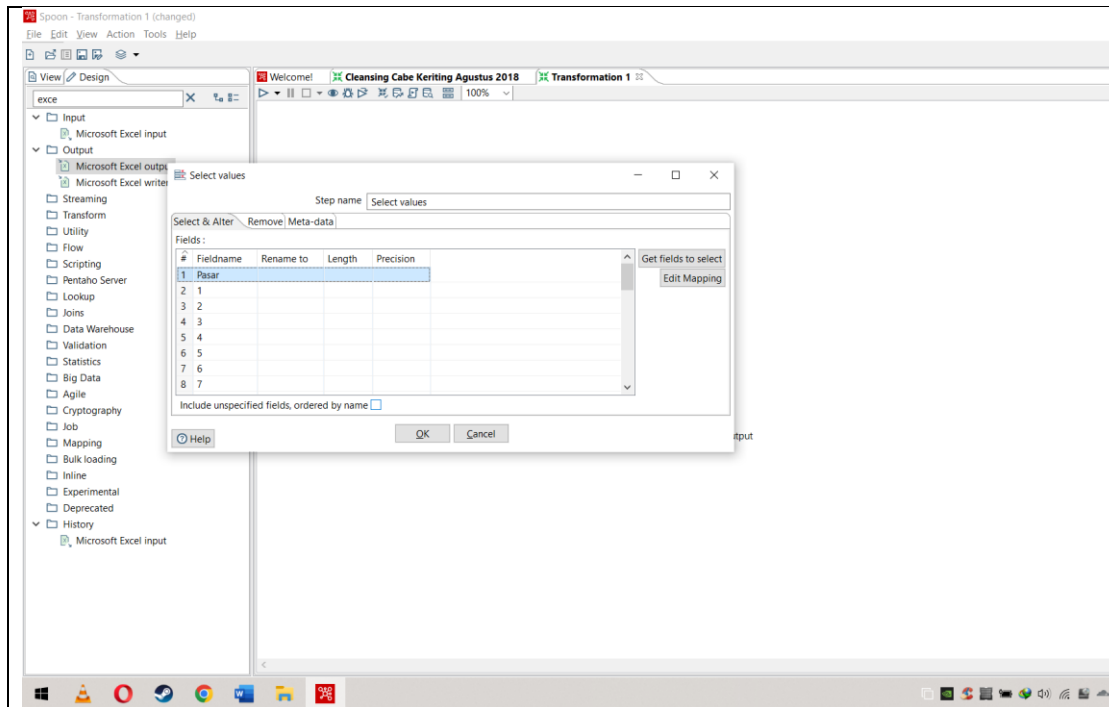


- Berikut ini adalah konfigurasi pada step CSV File Input.



4. Data Filtering

- Step selanjutnya adalah melakukan select value yang berguna untuk memilah kolom pada excel seperti pada gambar berikut. Kemudian Delete kolom No saja, karena nomor tidak diperlukan lalu Klik OK.



- Kemudian lakukan ascending data pada proses step Sort Rows

Sort rows

Step name:

Sort directory:

TMP-file prefix:

Sort size (rows in memory):

Free memory threshold (in %):

Compress TMP Files? ☒

Only pass unique rows? (verifies keys only) ☐

Fields:

#	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1	Pasar	Y	N	N	2	N
2	1	Y	N	N	2	N
3	2	Y	N	N	2	N
4	3	Y	N	N	2	N
5	4	Y	N	N	2	N

- Kemudian pada step Filter Rows, buang beberapa nama pasar seperti contoh pada dibawah ini. Untuk menambahkan condition klik icon + di pojok kanan atas.

Filter rows

Step name:

Send 'true' data to step:

Send 'false' data to step:

The condition:

NOT (Pasar = [Pasar Anyer Bahari])

AND

NOT (Pasar = [Pasar Gondangdia])

AND

NOT (Pasar = [Pasar Induk Kramat Jati])

AND

NOT (Pasar = [Pasar Kalibaru])

AND

NOT (Pasar = [Pasar Kalideres])

AND

NOT (Pasar = [Pasar Klender SS])

AND

- Berikut List Data Nama Pasar yang mau dibuang (2018)

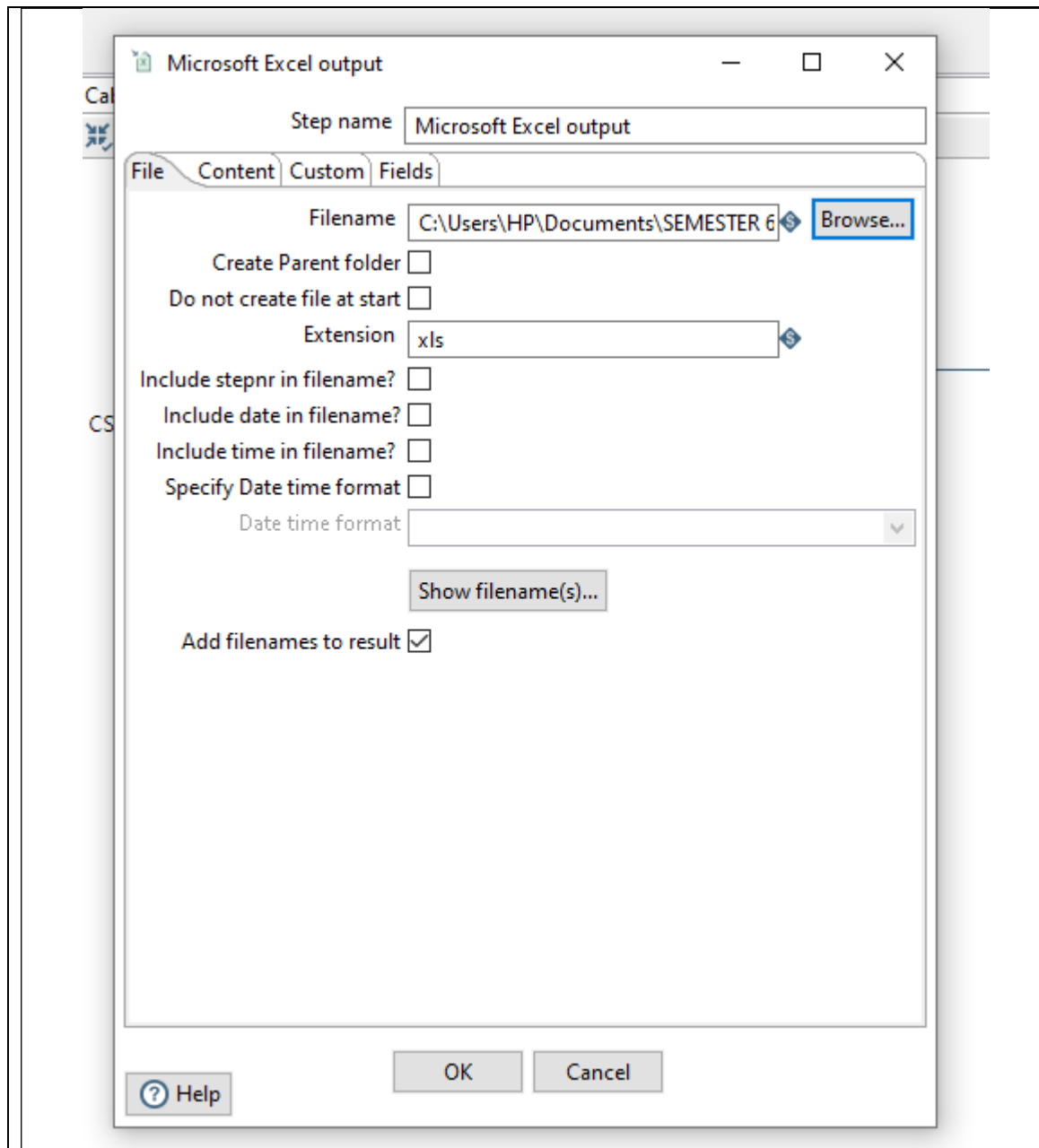


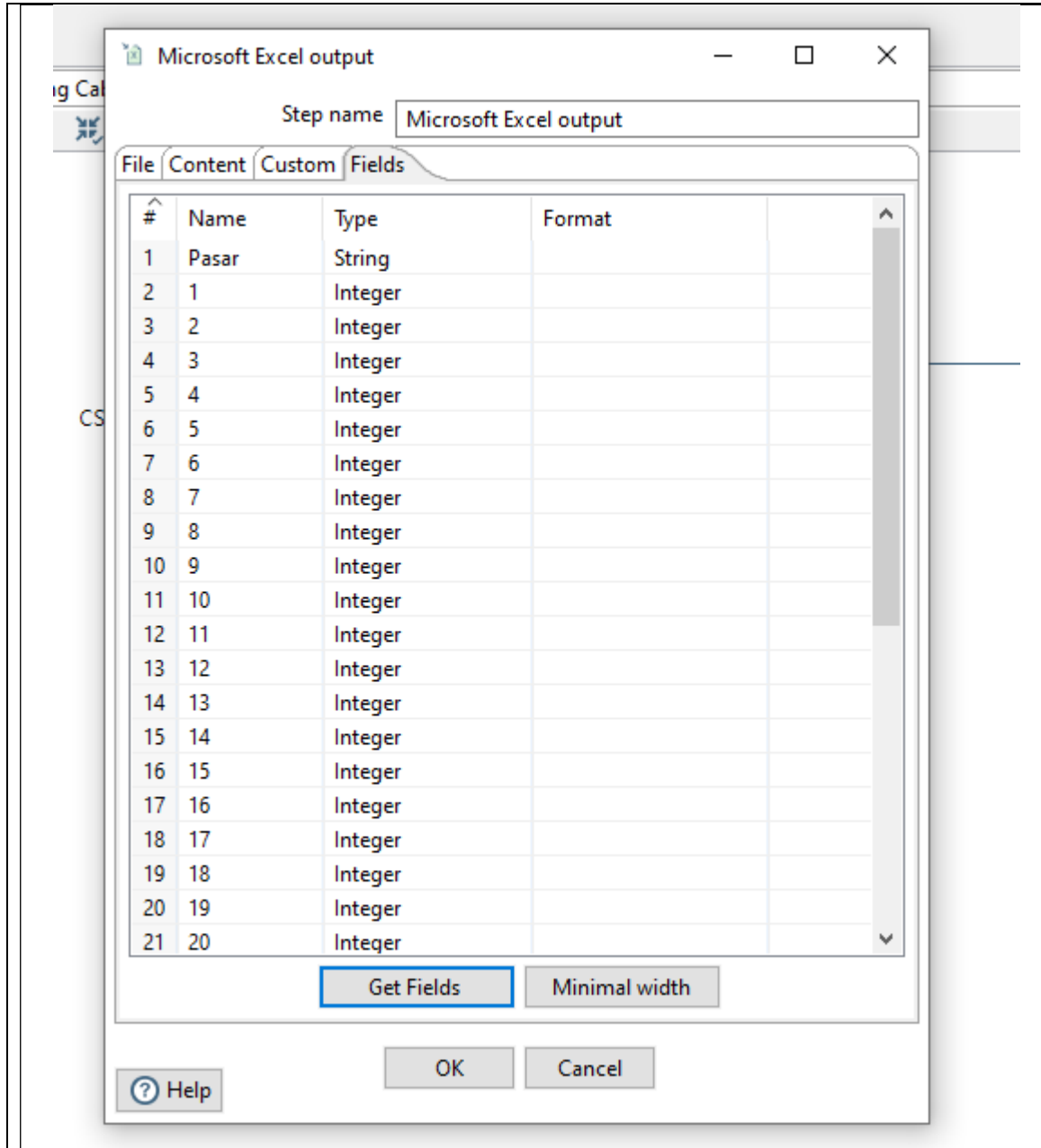
	NOT (Pasar = [Pasar Anyer Bahari])
AND	NOT (Pasar = [Pasar Gondangdia])
AND	NOT (Pasar = [Pasar Induk Kramat Jati])
AND	NOT (Pasar = [Pasar Kalibaru])
AND	NOT (Pasar = [Pasar Kalideres])
AND	NOT (Pasar = [Pasar Klender SS])
AND	NOT (Pasar = [Pasar Koja Baru])
AND	NOT (Pasar = [Pasar Rawamangun])
AND	NOT (Pasar = [Pasar Pramuka])
AND	NOT (Pasar = [Pasar Pluit])
AND	NOT (Pasar = [Pasar Pulo Gadung])
AND	NOT (Pasar = [Pasar Pesanggrahan])
AND	NOT (Pasar = [Pasar Glodok])
AND	NOT (Pasar = [Pasar Matraman KK])
AND	NOT (Pasar = [Pasar Jatinegara])

5. Excel Output

Langkah terakhir simpan data menggunakan step *Microsoft Excel Output* dengan konfigurasi sebagai berikut lalu simpan dengan format nama **Cabai Merah Keriting [Bulan] [Tahun].xls** seperti **Cabai Merah Keriting Agustus 2018.xls** (File & Sheet)







2. Alat dan Bahan

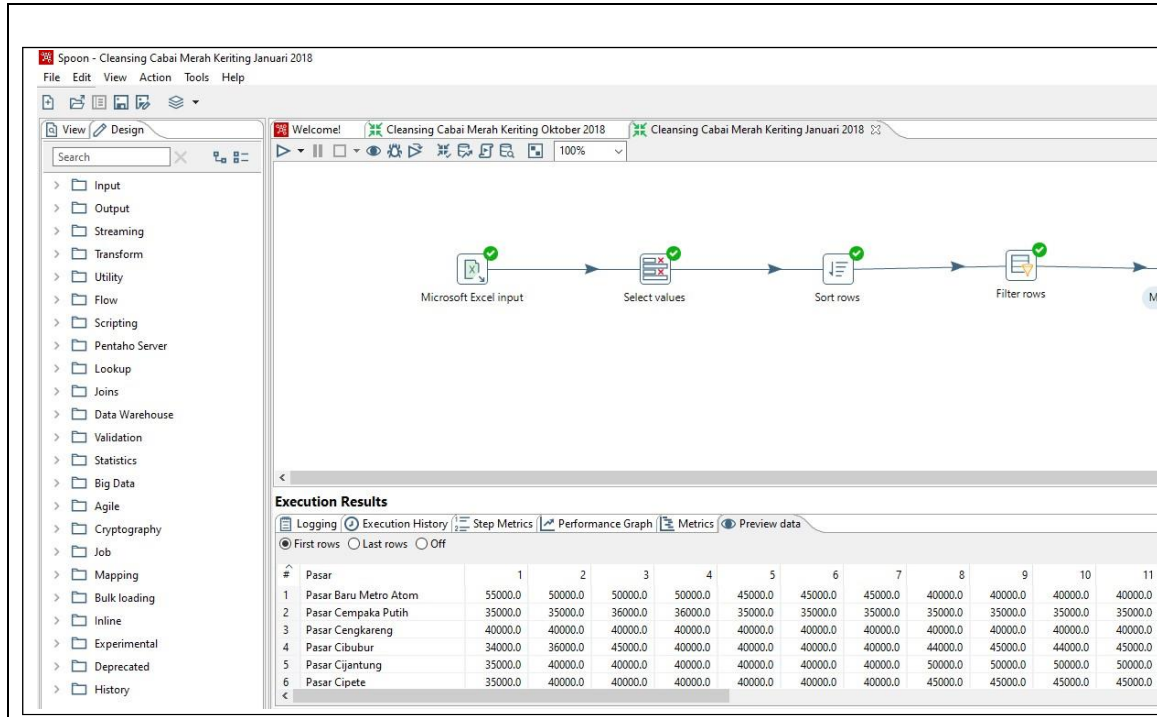
Hardware : Laptop/PC

Software : Spoon Pentaho from Hitachi Vantara

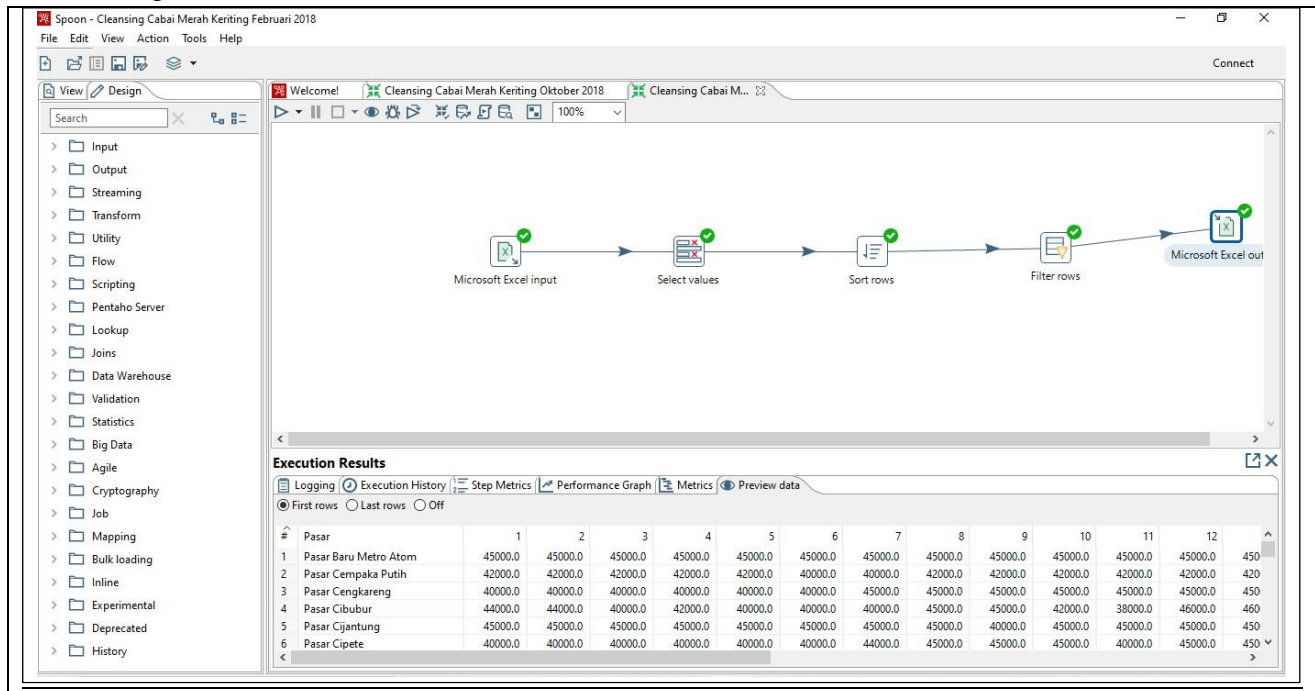
3. Elemen Kompetensi

- a. Tugas Akhir Praktikum I – Screenshot Struktur yang sudah berhasil berjalan
 1. Cleansing Januari 2018





2. Cleansing Februari 2018



3. Cleansing Maret 2018



Spoon - Cleansing Cabai Merah Maret 2018

File Edit View Action Tools Help

View Design

Search

- Input
- Output
- Streaming
- Transform
- Utility
- Flow
- Scripting
- Pentaho Server
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics
- Big Data
- Agile
- Cryptography
- Job
- Mapping
- Bulk loading
- Inline
- Experimental
- Deprecated
- History

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	
1	Pasar Baru Metro Atom	50000.0	50000.0	50000.0	50000.0	50000.0	50000.0	50000.0	50000.0	50000.0	50000.0	50000.0	50000.0	550
2	Pasar Cempaka Putih	13000.0	40000.0	42000.0	42000.0	42000.0	42000.0	42000.0	42000.0	40000.0	42000.0	42000.0	42000.0	420
3	Pasar Cengkareng	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	450
4	Pasar Cibubur	40000.0	46000.0	46000.0	40000.0	40000.0	40000.0	42000.0	45000.0	45000.0	42000.0	42000.0	40000.0	400
5	Pasar Cijantung	45000.0	45000.0	50000.0	50000.0	50000.0	50000.0	45000.0	45000.0	45000.0	45000.0	45000.0	50000.0	500
6	Pasar Cipete	45000.0	40000.0	40000.0	40000.0	45000.0	45000.0	45000.0	45000.0	45000.0	48000.0	48000.0	50000.0	500

4. Cleansing April 2018

Spoon - Cleansing Cabai Merah April 2018

File Edit View Action Tools Help

View Design

Search

- Input
- Output
- Streaming
- Transform
- Utility
- Flow
- Scripting
- Pentaho Server
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics
- Big Data
- Agile
- Cryptography
- Job
- Mapping
- Bulk loading
- Inline
- Experimental
- Deprecated
- History

Execution Results

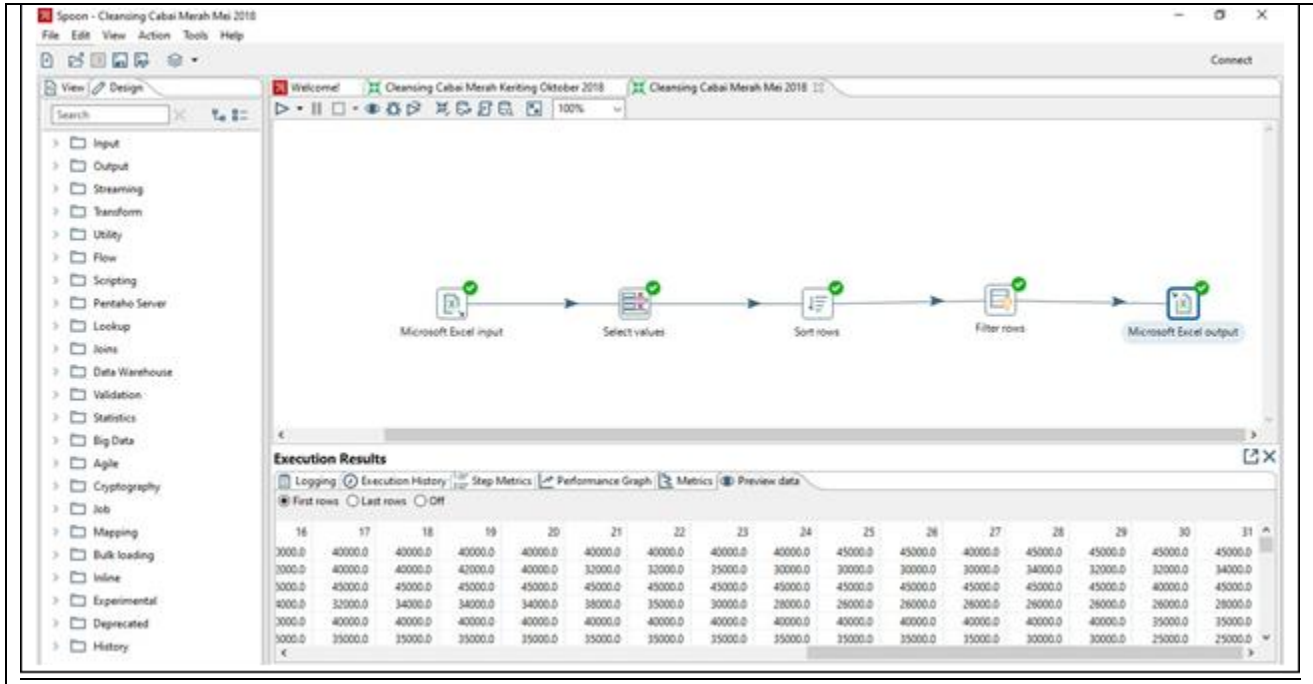
Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

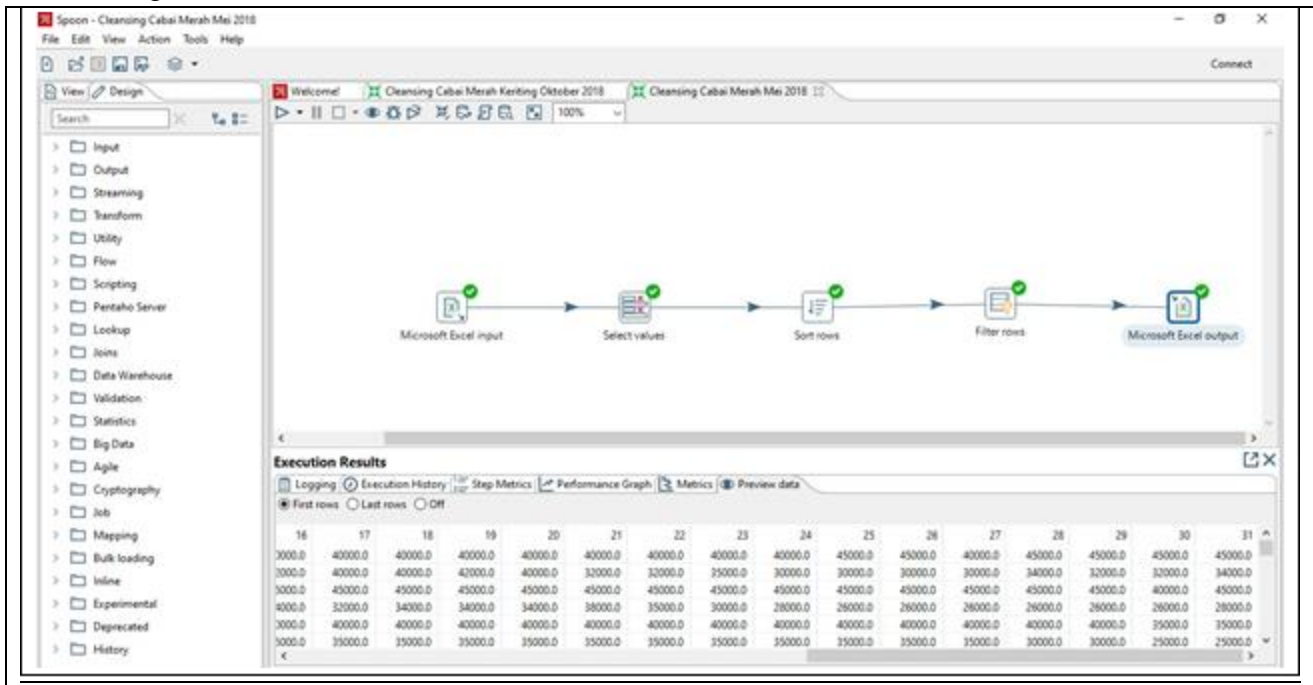
#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	
30	Pasar Sunter Podomoro	48000.0	48000.0	46000.0	47000.0	47000.0	47000.0	47000.0	47000.0	48000.0	48000.0	47000.0	47000.0	470
31	Pasar Tanah Abang Blok A-G	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	40000.0	400
32	Pasar Tebet Barat	60000.0	50000.0	50000.0	50000.0	50000.0	50000.0	50000.0	50000.0	50000.0	50000.0	50000.0	50000.0	500
33	Pasar Tomang Barat	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	45000.0	450
34	Pasar Ujung Menteng	44000.0	44000.0	44000.0	44000.0	44000.0	44000.0	44000.0	44000.0	44000.0	44000.0	44000.0	44000.0	440

5. Cleansing Mei 2018



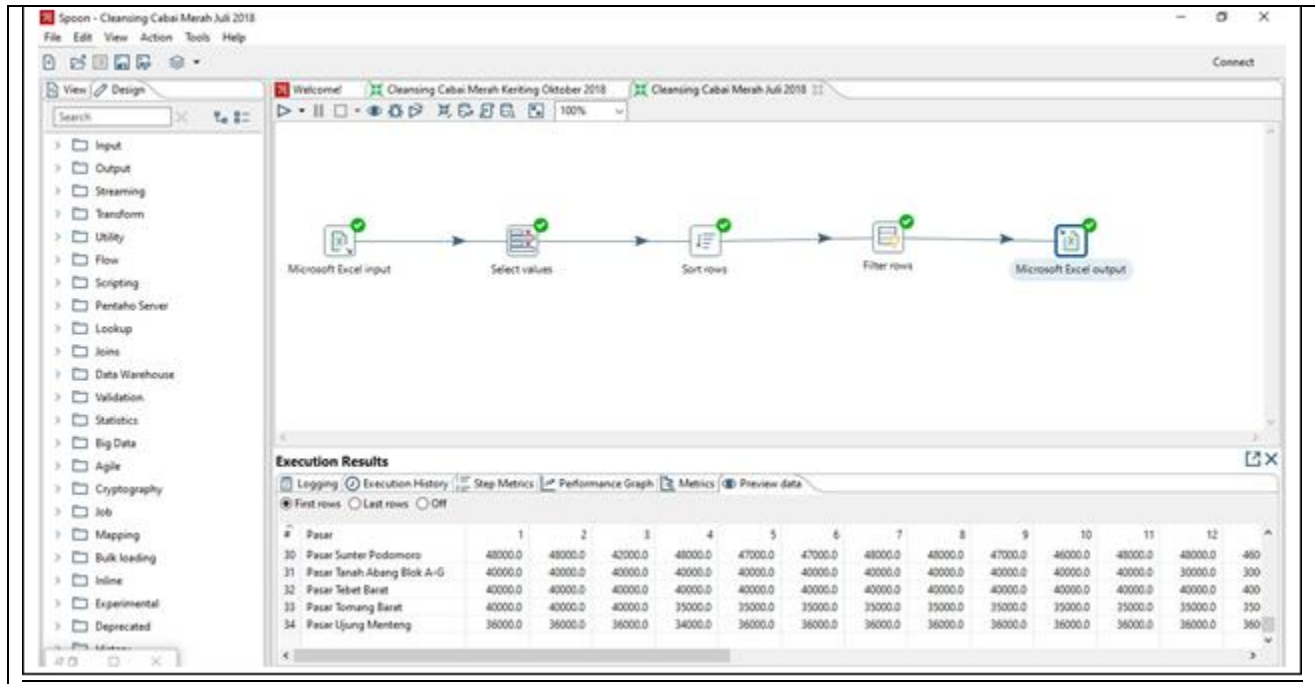


6. Cleansing Juni 2018

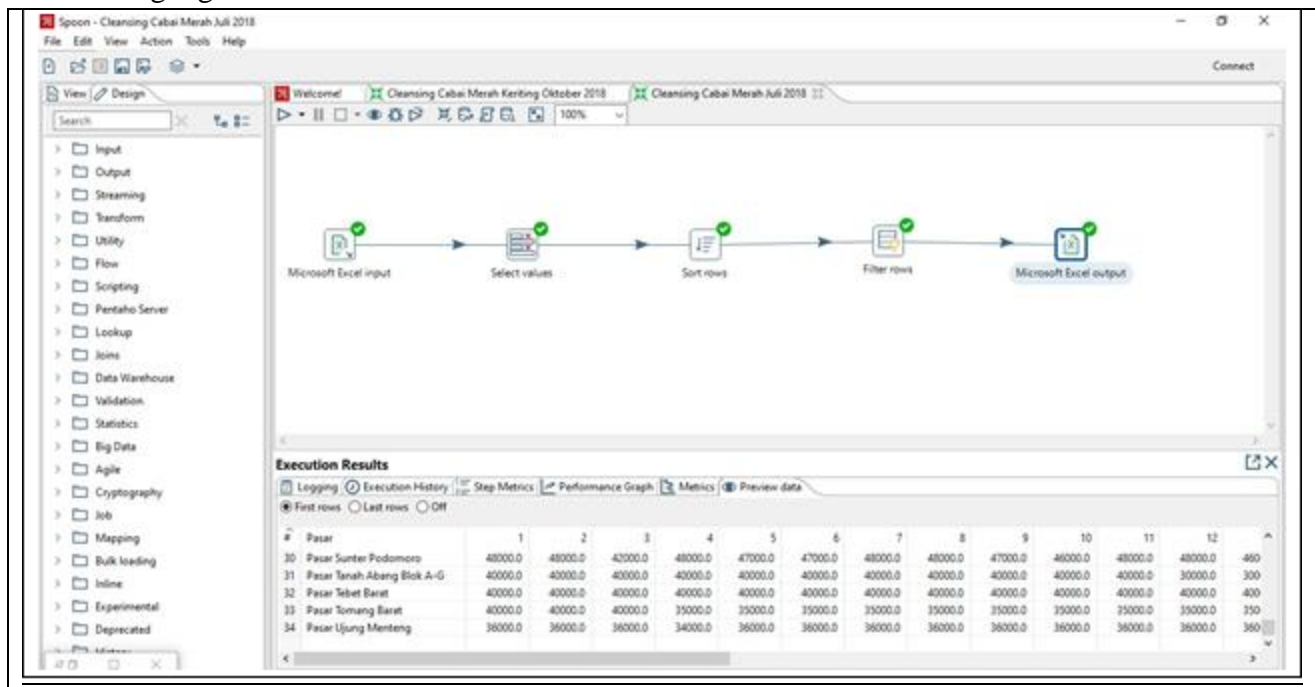


7. Cleansing Juli 2018





8. Cleansing Agustus 2018



9. Cleansing September 2018



Spoon - Cleansing Cabai Merah September 2018

File Edit View Action Tools Help

View Design

Search

Input
Output
Streaming
Transform
Utility
Flow
Scripting
Pentaho Server
Lookup
Joins
Data Warehouse
Validation
Statistics
Big Data
Agile
Cryptography
Job
Mapping
Bulk loading
Inline
Experimental
Deprecated
History

Connect

CSV file input → Select values → Sort rows → Filter rows → Microsoft Excel output

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14
30	Pasar Sunter Podomoro	46000	46000	48000	47000	46000	46000	48000	48000	48000	47000	46000	46000	46000	4
31	Pasar Tanah Abang Blok A-G	30000	30000	30000	30000	30000	30000	30000	30000	30000	25000	25000	25000	25000	2
32	Pasar Tebet Barat	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	4
33	Pasar Tomang Barat	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	3
34	Pasar Ujung Menteng	30000	30000	26000	26000	26000	26000	26000	26000	26000	26000	25000	25000	30000	3

10. Cleansing Oktober 2018

Spoon - Cleansing Cabai Merah Kerting Oktober 2018

File Edit View Action Tools Help

View Design

Search

Transformations
Cleansing Cabai Merah Kerting
Run configurations
Database connections
Steps
Hops
Partition schemas
Slave server
Kettle cluster schemas
Data Services
Hadoop clusters
VFS Connections

Connect

CSV file input → Select values → Sort rows → Filter rows → Microsoft Excel output

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Pasar Ekaru Metro Atom	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	35000	35000	35000	3
2	Pasar Cempaka Putih	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	2
3	Pasar Cengkareng	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	3
4	Pasar Cibubur	24000	24000	23000	23000	24000	24000	23000	25000	25000	28000	25000	28000	28000	2
5	Pasar Cijantung	35000	35000	35000	35000	35000	35000	40000	40000	40000	40000	40000	40000	40000	4
6	Pasar Cipete	27000	26000	35000	35000	35000	35000	35000	30000	30000	32000	32000	32000	32000	3

11. Cleansing November 2018



Execution Results

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
44000	44000	44000	44000	46000	46000	48000	43000	43000	43000	42000	44000	43000	43000	43000	42000	41000	42000
40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	35000	35000	35000	40000	40000	35000	35000
40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	45000	45000
45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000
40000	40000	40000	40000	35000	35000	35000	35000	35000	35000	35000	35000	35000	36000	35000	35000	35000	35000

12. Cleansing Desember 2018

Execution Results

18	19	20	21	22	23	24	25	26	27	28	29	30	31	Tertinggi	Rata-Rata	Terendah
34000	27000	31000	31000	30000	30000	31000	32000	31000	34000	34000	31000	31000	30000	34000	31548	27000
30000	30000	30000	30000	30000	30000	30000	35000	35000	30000	35000	35000	35000	35000	35000	32742	30000
35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	40000	36935	35000
25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	32000	32000	32000	32000	35000	28806	25000
30000	30000	30000	30000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	32552	30000

4. File Praktikum

Github Repository:

<https://github.com/adamzannuba11/PraktikumDWH>

5. Soal Latihan



Jurusan Teknik Informatika & Sistem Informasi
Fakultas Teknologi Industri – Universitas Trisakti

Soal:

1. Apa itu Cleansing Data?
2. Mengapa Data perlu dilakukan Cleansing?

Jawaban:

1. Cleansing Data adalah suatu proses mendeteksi dan memperbaiki (atau menghapus) data set, tabel, dan database yang korup atau tidak akurat. Data tidak akurat tersebut berupa data yang kosong dan tidak memiliki nilai atau berisi nilai NULL/NaN. Tujuan dari cleansing data ialah agar ketika data diproses nanti, tidak terjadi adanya error dari pengolahan data yang dilakukan akibat data yang tidak lengkap ataupun data hilang (missing value).
2. Tujuan dari cleansing data ialah agar ketika data diproses nanti, tidak terjadi adanya error dari pengolahan data yang dilakukan akibat data yang tidak lengkap ataupun data hilang (missing

6. Kesimpulan

- a. Dalam pengerjaan praktikum Data Warehouse, kita harus benar-benar teliti dalam menginputkan suatu fungsi untuk menampilkan suatu keluaran pada layar dengan sesuai.
- b. Kita dapat mengetahui...

7. Cek List (✓)

No	Elemen Kompetensi	Penyelesaian	
		Selesai	Tidak Selesai
1.	Tugas Akhir Praktikum I	...	

8. Formulir Umpan Balik

No	Elemen Kompetensi	Waktu Pengerjaan	Kriteria
1.	Tugas Akhir Praktikum I	... Menit	...

Keterangan:

1. Menarik
2. Baik
3. Cukup
4. Kurang

