

JOBSHEET 4
MULTI RESOURCE
MATA KULIAH DATA WAREHOUSE

Oleh:
Reza Angelina Febriyanti
NIM 2341760015



PROGRAM STUDI D4 SISTEM INFORMASI BISNIS
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG
2025

STUDI KASUS

Pak Ozai merupakan staff di PT Indomarko Prisma. PT Indomarko Prisma merupakan perusahaan retail yang memiliki usaha indomart, superind, dan indigrosir. Pak Ozai mempelajari proses bisnis dari perusahaan tersebut dan mencoba untuk membangun perusahaan sendiri dengan nama PT Ozai Enterprise dan membangun 3 cabang. Untuk dapat bersaing, Pak Ozai perlu melakukan analisa dari penjualan di ketiga cabang tersebut. 3 cabang tersebut mencatat penjualannya pada satu file excel.

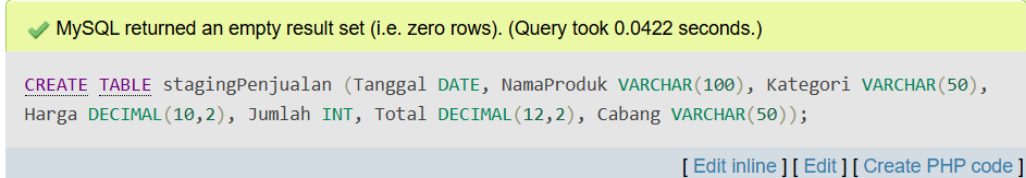
Toko Azura : <https://github.com/dik4rizky/datasources/blob/main/tokoazura.xls>

Toko Zuko : <https://github.com/dik4rizky/datasources/blob/main/tokozuko.xls>

Toko Iroh : <https://github.com/dik4rizky/datasources/blob/main/tokoiroh.xls>

Langkah Praktikum

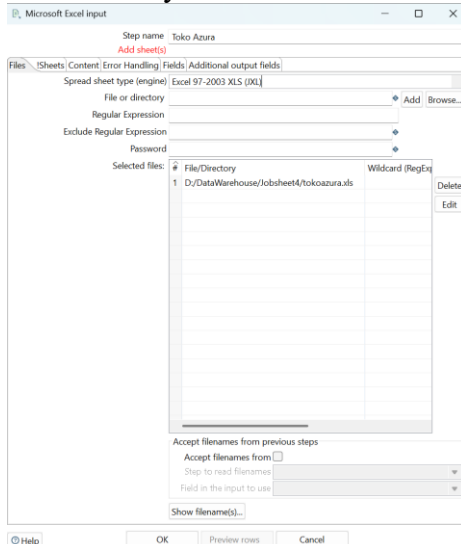
Buat tabel staging di database dw_OzaiEnterprise



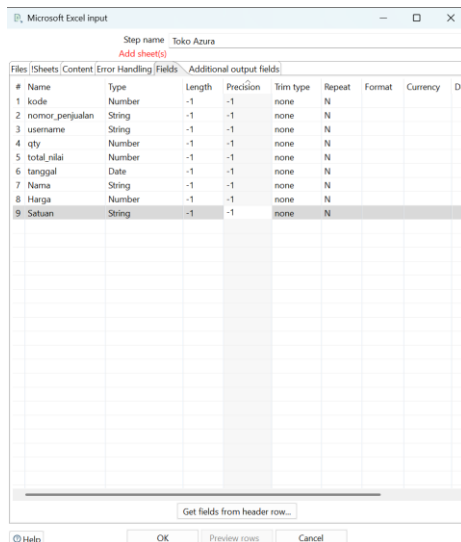
Drag and drop 3 Excel Input untuk data tokoazura, tokozuko, dan tokoiroh

Toko Azura

Upload file tokoazura tekan tombol browse dan klik ok. Jika berhasil maka muncul pada FileDirectory



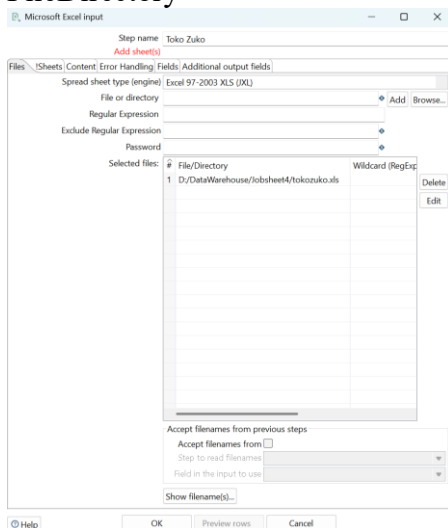
Kemudian klik tombol field lalu klik get field. Sesuaikan setiap field pada excel tokoazura seperti berikut



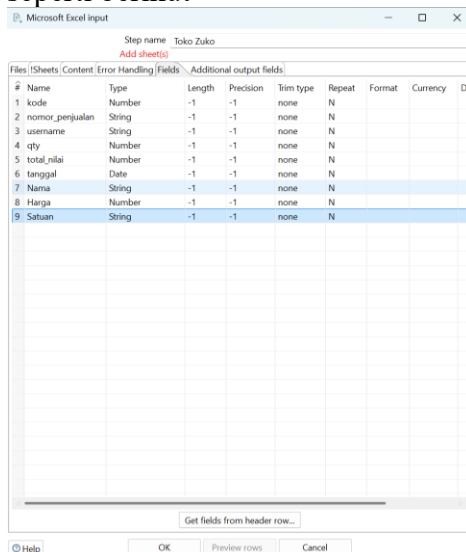
Kemudian klik ok

Toko Zuko

Upload file tokozuko tekan tombol browse dan klik ok. Jika berhasil maka muncul pada FileDirectory

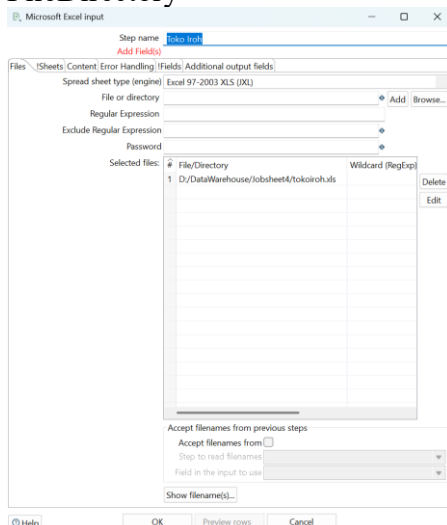


Kemudian klik tombol field lalu klik get field. Sesuaikan setiap field pada excel tokozuko seperti berikut

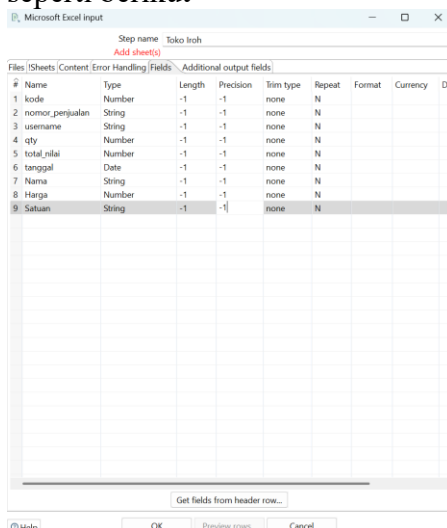


Toko Iroh

Upload file tokozuko tekan tombol browse dan klik ok. Jika berhasil maka muncul pada FileDirectory

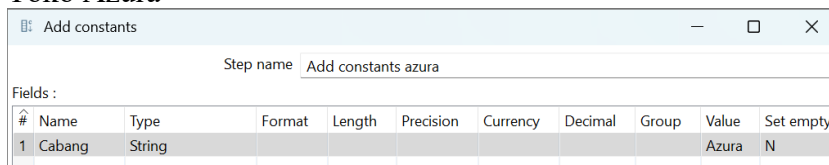


Kemudian klik tombol field lalu klik get field. Sesuaikan setiap field pada excel tokozuko seperti berikut

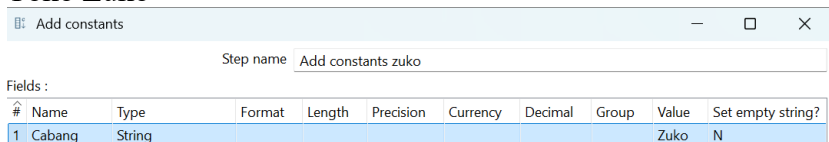


Tambahkan **add constants** (untuk mengisi field cabang dengan value constant)

Toko Azura



Toko Zuko



Toko Iroh

Add constants										
Step name Add constants iroh										
Fields :										
#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Value	Set empty string?
1	Cabang	String							Iroh	N

Lakukan union dengan Append Streams atau Merge Rows untuk menggabungkan ketiga data menjadi satu

Append streams 1 dan 2

Step name Append streams 1
Head hop: Add constants iroh
Tail hop: Add constants iroh

Step name Append streams 2
Head hop: Add constants azura
Tail hop: Add constants zuko

Konfigurasi untuk tabel output

Buat koneksi ke database dw_OzaiEnterprise

Database Connection

Connection name: conn_dw_destination_penjualan
Connection type: MSO
Settings: Host Name: localhost, Database Name: dw_OzaiEnterprise, Port Number: 3306, Username: root, Password:

Connection tested successfully
Connection to conn_dw_destination_penjualan was successful.
Hostname: localhost
Port: 3306
Database name: dw_OzaiEnterprise

Konfigurasi tabel output dan database fields nya sesuai dengan struktur tabel stagingPenjualan

Table output
Step name Table output
Connection conn_dw_destination_penjualan
Target schema dw_OzaiEnterprise
Target table stagingPenjualan
Commit size 1000
Truncate table
Ignore insert errors
Specify database fields

Main options Database fields
Fields to insert:

#	Table field	Stream field
1	Tanggal	Tanggal
2	NamaProduk	Nama
3	Kategori	Satuan
4	Harga	Harga
5	Jumlah	Qty
6	Total	total_nilai
7	Cabang	Cabang

Hasil akhir

The screenshot displays the Apache Spark Studio interface. At the top, there are tabs for 'Welcome!', 'Jobsheet4', and 'StudiKasus'. Below the tabs is a toolbar with various icons and a zoom level of 100%. The main workspace shows a data pipeline with the following components:

- Toko Iroh** (Source) → **Add constants iroh** (Transformation) → **Append streams 1** (Action)
- Toko Zuko** (Source) → **Add constants zuko** (Transformation) → **Append streams 2** (Action)
- Toko Azura** (Source) → **Add constants azura** (Transformation) → **Append streams 2** (Action)
- Append streams 1** and **Append streams 2** both connect to **Table output** (Sink).

Below the workspace, the **Execution Results** tab is active, showing a log of the transformation execution:

```
2025/04/22 11:14:25 - Spoon - Transformation opened.
2025/04/22 11:14:25 - Spoon - Launching transformation [StudiKasus]...
2025/04/22 11:14:26 - Spoon - Started the transformation execution.
2025/04/22 11:14:26 - StudiKasus - Dispatching started for transformation [StudiKasus]
2025/04/22 11:14:26 - Table output.0 - Connected to database [conn_dw_destination_penjualan] (commit=1000)
2025/04/22 11:14:27 - Toko Zuko.0 - Finished processing (I=1078, O=0, R=0, W=1078, U=0, E=0)
2025/04/22 11:14:27 - Add constants zuko.0 - Finished processing (I=0, O=0, R=1078, W=1078, U=0, E=0)
2025/04/22 11:14:27 - Toko Azura.0 - Finished processing (I=3862, O=0, R=0, W=3862, U=0, E=0)
2025/04/22 11:14:27 - Add constants azura.0 - Finished processing (I=0, O=0, R=3862, W=3862, U=0, E=0)
2025/04/22 11:14:28 - Toko Iroh.0 - Finished processing (I=13069, O=0, R=0, W=13069, U=0, E=0)
2025/04/22 11:14:28 - Add constants iroh.0 - Finished processing (I=0, O=0, R=13069, W=13069, U=0, E=0)
```

Isi tabel pada database :

Showing rows 0 - 49 (18009 total, Query took 0.0010 seconds.)

```
SELECT * FROM stagingpenjualan;
```

☐ Profiling [[Edit inline](#)] [[Edit](#)] [[Explain SQL](#)] [[Create PHP code](#)] [[Refresh](#)]

1 > >> | Number of rows: 50

Extra options

Tanggal	NamaProduk	Kategori	Harga	Jumlah	Total	Cabang
2008-02-18	INDOCAFE CAPUCINO KTK 5X25G	PCS	6916.25	2	353700.00	Azura
2008-02-18	ALKALINE LR-03 (ISI 2)	LSN	7936.25	1	8250.00	Azura
2008-01-02	NUTRIJELL STRAW 15 GR	BOX	2667.50	1	33000.00	Azura
2008-01-02	NUTRIJELL ANGGUR 15 GR	BOX	2667.50	1	33000.00	Azura
2008-01-02	NUTRIJELL PLAIN 15 GR	BOX	2667.50	1	33000.00	Azura
2008-01-02	REJOICE FAMILY C 175	PCS	10890.00	1	187500.00	Azura
2008-01-02	WINGS BIRU 500 K 209G	DOS	2250.00	2	46050.00	Azura
2008-01-02	SOKLIN MATIC FRONT LOAD 1KG	DOS	20000.00	1	26650.00	Azura
2008-01-07	SOKLIN HIGINIS 900	DOS	15625.00	1	20250.00	Azura
2008-01-17	AJINOMOTO 50GR	PCS	1418.75	2	56500.00	Azura

Dimensi Waktu

Membuat dabatase dw_OzaiEnterprise

✓ MySQL returned an empty result set (i.e. zero rows). (Query took 0.0242 seconds.)

```
CREATE DATABASE dw_OzaiEnterprise;
```

[\[Edit inline \]](#) [\[Edit \]](#) [\[Create PHP code \]](#)

✓ MySQL returned an empty result set (i.e. zero rows). (Query took 0.0004 seconds.)

```
use dw_OzaiEnterprise;
```

[\[Edit inline \]](#) [\[Edit \]](#) [\[Create PHP code \]](#)

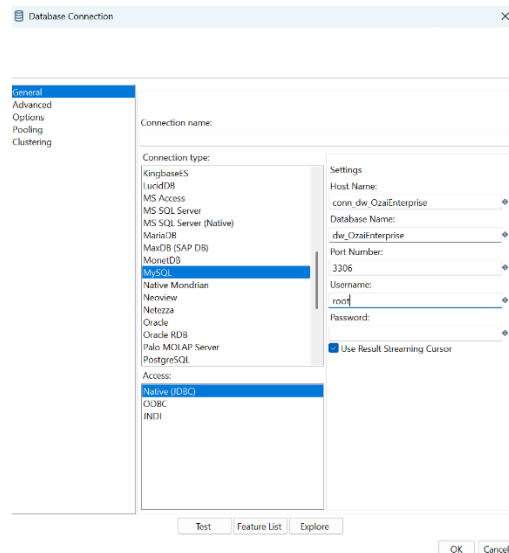
Membuat table dimDate

✓ MySQL returned an empty result set (i.e. zero rows). (Query took 0.0451 seconds.)

```
CREATE TABLE dimDate( id_dimDate int not null  
AUTO_INCREMENT PRIMARY KEY, date date, year int, month int,  
day int );
```

[\[Edit inline \]](#) [\[Edit \]](#) [\[Create PHP code \]](#)

Membuat koneksi ke database dw_OzaiEnterprise



Hasil Percobaan Praktikum dapat dilihat pada file Jobsheet4.krt

Percobaan berhasil karena table dimdate terisi 1825 rows

✓ Showing rows 0 - 499 (1825 total, Query took 0.0011 seconds.)

```
SELECT * FROM `dimdate`
```

Isi Table dimdate seperti berikut

id_dimDate	date	year	month	day
1	2003-01-01	2003	1	1
2	2003-01-02	2003	1	2
3	2003-01-03	2003	1	3
4	2003-01-04	2003	1	4
5	2003-01-05	2003	1	5

TUGAS 1

Buka preview tab pada execution result area di setiap proses object. amati input dan output data yang ada. bandingkan di setiap prosesnya. jelaskan perbedaan disetiap prosesnya.

Stepname	Copynr	Read	Written	Input	Output
Generate rows	0	0	1825	0	0
Add sequence	0	1825	1825	0	0
Calculator	0	1825	1825	0	0
Select values	0	1825	1825	0	0
Database lookup	0	1825	1825	0	0
Filter rows	0	1825	1825	0	0
Table output	0	1825	1825	0	1825

Generate rows

Input	Output	Keterangan
Tidak ada	# CurrentDate 1 01-01-2003 2 01-01-2003 3 01-01-2003 4 01-01-2003 5 01-01-2003	Membuat field currentdate yang berisi 1825 rows

Add Sequences

Input	Output	Keterangan
# CurrentDate 1 01-01-2003 2 01-01-2003 3 01-01-2003 4 01-01-2003 5 01-01-2003 Output dari generate rows	# CurrentDate incrementDay 1 01-01-2003 0 2 01-01-2003 1 3 01-01-2003 2 4 01-01-2003 3 5 01-01-2003 4	Terbentuk field baru incrementDay yang dimulai dari nilai 0 dan bertambah terus

Calculator

Input	Output	Keterangan
-------	--------	------------

#	CurrentDate	incrementDay	#	CurrentDate	incrementDay	streamDate	streamYear	streamMonth	streamDay	
1	01-01-2003	0	1	01-01-2003	0	2003/01/01 00:00:00.000	2003	1	1	Terbentuk field baru streamDate: tanggal awal 01-01-2003 bertambah sesuai baris incrementDay streamYear: mengambil year pada field streamDate streamMonth: mengambil month pada field streamDate streamDay: mengambil day pada field streamDate
2	01-01-2003	1	2	01-01-2003	1	2003/01/02 00:00:00.000	2003	1	2	
3	01-01-2003	2	3	01-01-2003	2	2003/01/03 00:00:00.000	2003	1	3	
4	01-01-2003	3	4	01-01-2003	3	2003/01/04 00:00:00.000	2003	1	4	
5	01-01-2003	4	5	01-01-2003	4	2003/01/05 00:00:00.000	2003	1	5	

Output dari Add
Sequences

Select values

Input						Output				Keterangan		
#	CurrentDate	incrementDay	streamDate	streamYear	streamMonth	streamDay	#	streamDate	streamYear	streamMonth	streamDay	Menghapus field currentdate & incerementday
1	01-01-2003	0	2003/01/01 00:00:00.000	2003	1	1	1	2003/01/01 00:00:00.000	2003	1	1	
2	01-01-2003	1	2003/01/02 00:00:00.000	2003	1	2	2	2003/01/02 00:00:00.000	2003	1	2	
3	01-01-2003	2	2003/01/03 00:00:00.000	2003	1	3	3	2003/01/03 00:00:00.000	2003	1	3	
4	01-01-2003	3	2003/01/04 00:00:00.000	2003	1	4	4	2003/01/04 00:00:00.000	2003	1	4	
5	01-01-2003	4	2003/01/05 00:00:00.000	2003	1	5	5	2003/01/05 00:00:00.000	2003	1	5	
Output dari Calculator												

Output dari Calculator

Database lookup

Input					Output													Keterangan
#	streamDate	streamYear	streamMonth	streamDay	#	streamDate	streamYear	streamMonth	streamDay	date	year	month	day	Muncul field date year, month, day				
1	2003/01/01 00:00:00.000	2003	1	1	1	2003/01/01 00:00:00.000	2003	1	1	<null>	<null>	<null>	<null>					
2	2003/01/02 00:00:00.000	2003	1	2	2	2003/01/02 00:00:00.000	2003	1	2	<null>	<null>	<null>	<null>					
3	2003/01/03 00:00:00.000	2003	1	3	3	2003/01/03 00:00:00.000	2003	1	3	<null>	<null>	<null>	<null>					
4	2003/01/04 00:00:00.000	2003	1	4	4	2003/01/04 00:00:00.000	2003	1	4	<null>	<null>	<null>	<null>					
5	2003/01/05 00:00:00.000	2003	1	5	5	2003/01/05 00:00:00.000	2003	1	5	<null>	<null>	<null>	<null>					
Output select values																		

Output select values

Filter rows

Input									Output									Keterangan
#	streamDate	streamYear	streamMonth	streamDay	date	year	month	day	#	streamDate	streamYear	streamMonth	streamDay	date	year	month	day	Muncul field date year, month, day
1	2003/01/01 00:00:00.000	2003	1	1	<null>	<null>	<null>	<null>	1	2003/01/01 00:00:00.000	2003	1	1	<null>	<null>	<null>	<null>	
2	2003/01/02 00:00:00.000	2003	1	2	<null>	<null>	<null>	<null>	2	2003/01/02 00:00:00.000	2003	1	2	<null>	<null>	<null>	<null>	
3	2003/01/03 00:00:00.000	2003	1	3	<null>	<null>	<null>	<null>	3	2003/01/03 00:00:00.000	2003	1	3	<null>	<null>	<null>	<null>	
4	2003/01/04 00:00:00.000	2003	1	4	<null>	<null>	<null>	<null>	4	2003/01/04 00:00:00.000	2003	1	4	<null>	<null>	<null>	<null>	
5	2003/01/05 00:00:00.000	2003	1	5	<null>	<null>	<null>	<null>	5	2003/01/05 00:00:00.000	2003	1	5	<null>	<null>	<null>	<null>	
Output database lookup																		

Output database lookup

Table Output

Input									Output									Keterangan
#	streamDate	streamYear	streamMonth	streamDay	date	year	month	day	#	streamDate	streamYear	streamMonth	streamDay	date	year	month	day	Muncul field date year, month, day
1	2003/01/01 00:00:00.000	2003	1	1	<null>	<null>	<null>	<null>	1	2003/01/01 00:00:00.000	2003	1	1	<null>	<null>	<null>	<null>	
2	2003/01/02 00:00:00.000	2003	1	2	<null>	<null>	<null>	<null>	2	2003/01/02 00:00:00.000	2003	1	2	<null>	<null>	<null>	<null>	
3	2003/01/03 00:00:00.000	2003	1	3	<null>	<null>	<null>	<null>	3	2003/01/03 00:00:00.000	2003	1	3	<null>	<null>	<null>	<null>	
4	2003/01/04 00:00:00.000	2003	1	4	<null>	<null>	<null>	<null>	4	2003/01/04 00:00:00.000	2003	1	4	<null>	<null>	<null>	<null>	
5	2003/01/05 00:00:00.000	2003	1	5	<null>	<null>	<null>	<null>	5	2003/01/05 00:00:00.000	2003	1	5	<null>	<null>	<null>	<null>	
Output filter rows																		

Output filter rows