# DATA WAREHOUSE JOBSHEET 4



Mochammad Audric Andhika H 2341760094

# PROGRAM STUDI D-IV SISTEM INFORMASI BISNIS JURUSAN TEKNOLOGI INFORMASI POLITEKNIK NEGERI MALANG

### Tujuan Praktikum

Setelah melakukan praktikum ini, mahasiswa diharapakan dapat lebih mengenal table dimensi, table fakta dan apa itu OLAP dengan banyak sumber.

### **Studi Kasus**

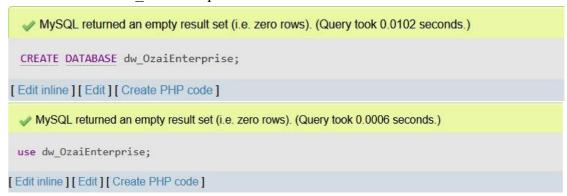
Pak Ozai merupakan staff di PT Indomarko Prismatama. PT Indomarko Prisma tama merupakan perusahaan retail yang memiliki usaha indomarti, superindi dan indigrosir. Pak ozai mempelajari proses bisnis dari perusahaan tersebut dan mencoba untuk membangun perusahaan sendiri dengan nama PT Ozai Enterprise 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. Berikut ketiga file tersebut:

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

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

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

1. Create Database dw OzaiEnterprise



2. Create tabel staging pada database dw OzaiEnterprise

```
✓ MySQL returned an empty result set (i.e. zero rows). (Query took 0.0110 seconds.)
CREATE TABLE stagingPenjualan (Tanggal DATE, NamaProduk VARCHAR(100), Kategori VARCHAR(50), Harga DECIMAL(10,2), Jumlah INT, Total DECIMAL(12,2), Cabang VARCHAR(50));
[Edit Inline] [Edit] [Create PHP code]
```

3. Menyiapkan 3 elemn Microsoft Excel Input untuk menginputkan ketiga data pada toko azura, zuko dan iroh





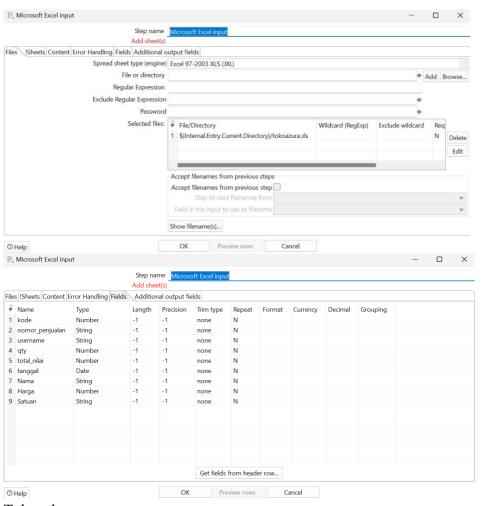


Microsoft Excel input

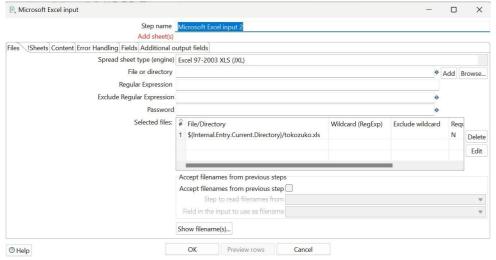
Microsoft Excel input 2

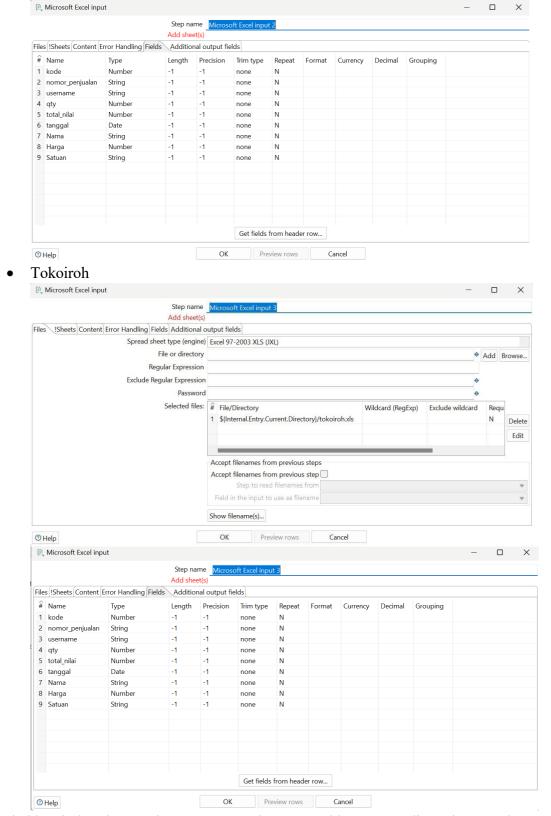
Microsoft Excel input 3

- 4. Drag & drop Excel Input sebanyak 3 kali untuk
  - Tokoazura

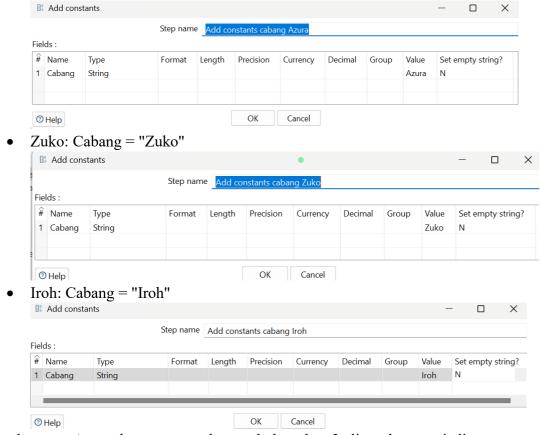


### Tokozuho

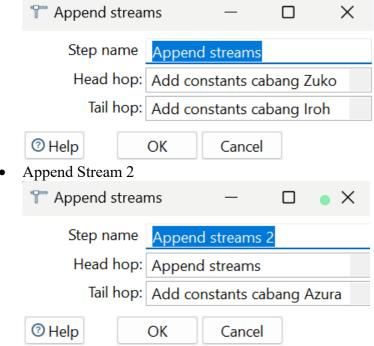




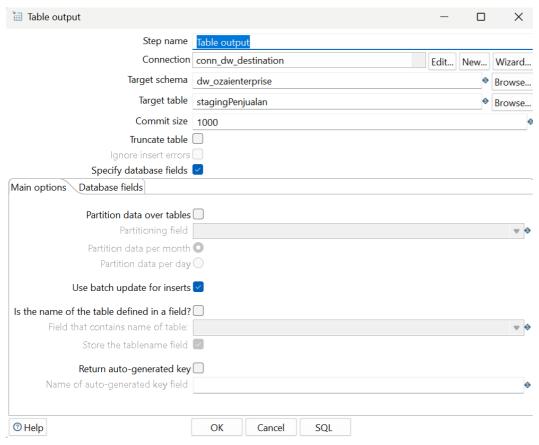
- 5. Tambahkan kolom baru Cabang menggunakan step Add constants digunakan untuk menambahkan kolom baru berisi nilai tetap (konstan) sesuai dengan nama toko cabangnya.
  - Azura: Cabang = "Azura"



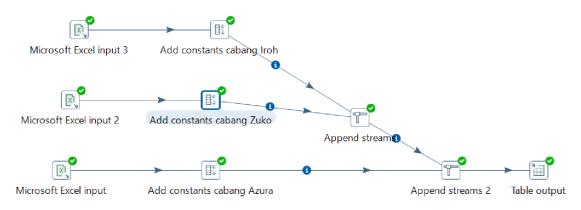
- 6. Gunakan step Append streams untuk menghubungkan 3 aliran data menjadi satu
  - Append Stream 1

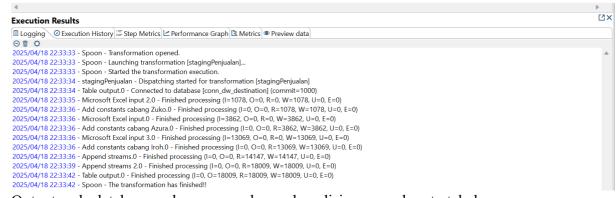


- 7. Drag & drop **Table Output**:
  - ➤ Koneksi: pilih conn dw destination (koneksi ke database).
  - ➤ Table: stagingPenjualan
  - Mapping: sesuai dengan field tabel.



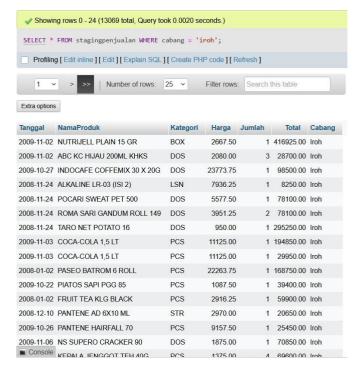
8. Hasil transformation:





9. Output pada database maka semua cabang akan disimpan pada satu tabel :





### A. Dimensi Waktu

 Buatlah sebuah database yang digunakan sebagai OLAP dengan nama dw OzaiEnterprise

```
✓ MySQL returned an empty result set (i.e. zero rows). (Query took 0.0006 seconds.)
use dw_OzaiEnterprise;
[Edit inline] [Edit] [ Create PHP code]
```

2. Buatlah table untuk menyimpan data master waktu atau yang disebut dengan **tabel dimensi.** Beri nama table tersebut dengan nama **dimDate**.

```
✓ MySQL returned an empty result set (i.e. zero rows). (Query took 0.0234 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]
```

Pada tahapan selanjutnya, untuk membuat tabel dimensi dimDate, maka diperlukan generate data tanggal. Data tanggal yang disiapkan pada tabel dimDate menyesuaikan dengan proses bisnis yang berjalan.

Proses bisnis pada LegendVehicle adalah 20 tahun. Sehingga data pada tabel dimdate yang harus tersedia adalah tanggal selama 15 tahun. Mulai dari 1 Januari 2005

- 1. Buka PDI Spoon. Buat Transformation baru -> File New Transformation.
- 2. Drag and Drop beberapa objek yaitu:
- Generate Rows: digunakan untuk membuat baris data baru.
- Add Sequence: digunakan untuk membuat sequence, dalam hal ini membuat data di setiap harinya.

- Calculator: digunakan untuk menjumlahkan hari dan mengambil data tahun, bulan dan hari.
- Select Values: digunakan untuk memilih field yang digunakan.
- **Database Lookup:** digunakan untuk melihat dan memastikan bahwa data yang akan dimasukkan kedalam tabel dimDate tidak kembar atau sama dengan data yang ada pada tabel dimDate itu sendiri.
- **Filter Rows:** digunakan untuk mengambil data yang belum ada pada table dimDate setelah dicek sebelumnya.
- Table Output: digunakan untuk menyimpan data pada tabel tujuan (dimDate).



- 1. Konfigurasi pada **Generate Rows** adalah merubah **limit** menjadi **1825** dimana memiliki arti bahwa data yang akan dibuat sebanyak 7300 data. 7300 merupakan jumlah hari dalam 20 tahun ( 365 hari x 20 tahun ).
- 2. Membuat fields baru bernama CurrentDate dengan type data Date dan format dd-MM-yyyy serta value awal 01-01-2005

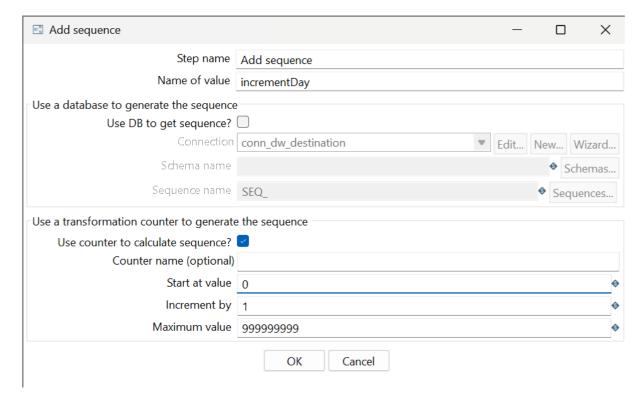


# Gambar konfigurasi generate rows

1. Hubungkan output dari Generate Rows menuju Add Sequence.



2. Konfigurasi pada Add Sequences adalah merubah Name of value menjadi incrementDay dengan start value bernilai 0 dan increment by bernilai 1

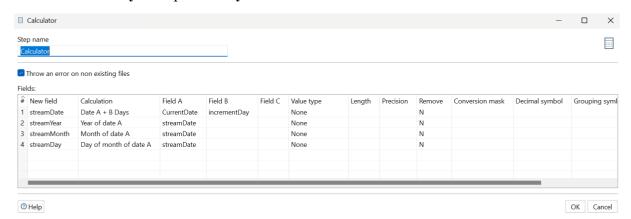


### Gambar konfigurasi add sequences

1. Hubungkan output dari add sequences menuju calculator.

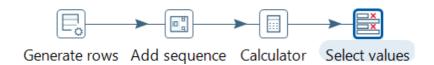


- 2. Konfigurasi pada calculator dengan membuat fields baru sebagai berikut:
- streamDate merupakan kalkulasi dari CurrentDate + incrementDay
- streamYear merupakan Year dari streamDate
- streamMonth merupakan Month dari streamDate
- streamDay merupakan Day of month dari streamDate

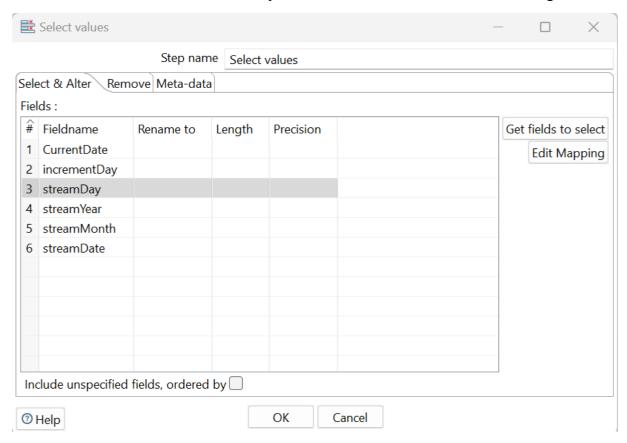


Gambar konfigurasi kalkulator

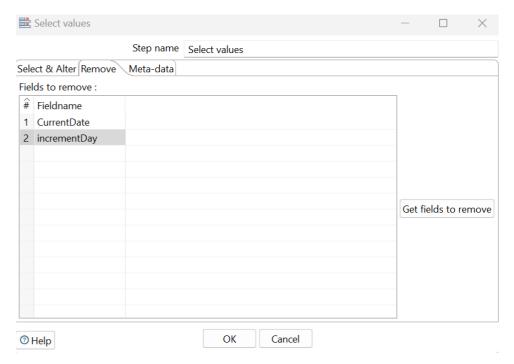
1. Hubungkan output dari calculator menuju Select values



- 2. Konfigurasi pada select values adalah dengan menekan tombol Get fields to select pada tab Select & Alter. Secara otomatis semua fields dari data input akan muncul.
- 3. Dikarenakan tidak semua fields digunakan, maka pada tab **Remove** diisikan fields **CurrentDate** dan **incrementDay** dikarenakan kedua fields tersebut tidak digunakan.



Gambar konfigurasi tab select & alter pada select values

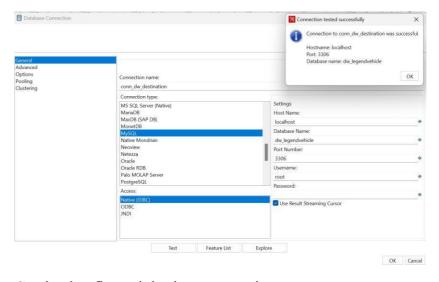


Gambar konfigurasi tab remove pada select values

1. Hubungkan output select values menuju database lookup.

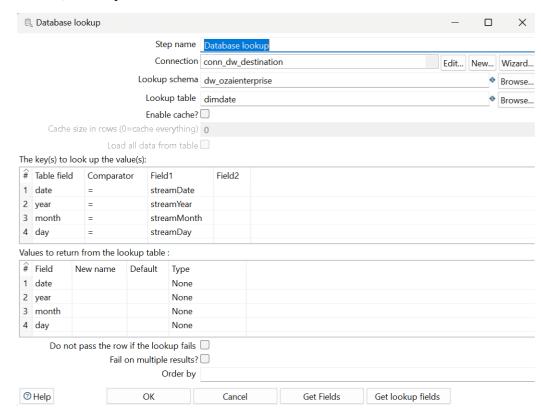


2. Sebelum melakukan konfigurasi pada database lookup, buatlah koneksi terlebih dahulu pada database melalui File - New - Database Connection. Gunakan Connection type MySQL dengan host name, database name, port number, username dan password sesuai konfigurasi MySQL pada device masing-masing. beri nama connection name tersebut dengan nama conn\_dw\_destination.



Gambar konfigurasi database connection

- 1. Konfigurasi pada **database lookup** adalah dengan memberikan **connection** dengan koneksi yang sudah dibuat pada step sebelumnya. dengan **schema** nama database yang digunakan dan **tabel dimdate** yang telah dibuat pada langkah pertama.
- 2. Field yang akan dicek untuk melihat kesamaan isi datanya agar tidak kembar adalah:
- field date pada table dimdate dengan field streamDate
- field year pada table dimdate dengan field streamYear
- field month pada table dimdate dengan field streamMonth
- field day pada table dimdate dengan field streamDay
- 1. Field yang akan di **retrive** adalah field yang ada pada table **dimDate** yaitu **date**, **year**, **month**, dan **day**.



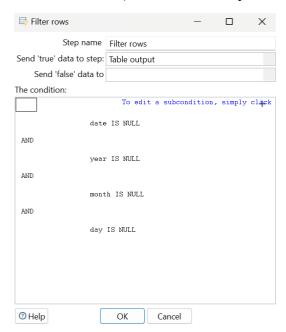
### Gambar konfigurasi database lookup

1. Hubungkan output dari database lookup dengan filter rows



2. Konfigurasi pada **filter rows** adalah dengan melakukan konfigurasi **output true data** pada **table output**. Pada bagian ini data yang tidak memiliki kesamaan pada tahapan sebelumnya akan dicek dimana jika **fields Stream** tidak memiliki kesamaan dengan

field dimDate, maka field dimDate tersebut akan bernilai null. Pada pernyataan kondisi tuliskan ( date is null and year is null and month is null and day is null)

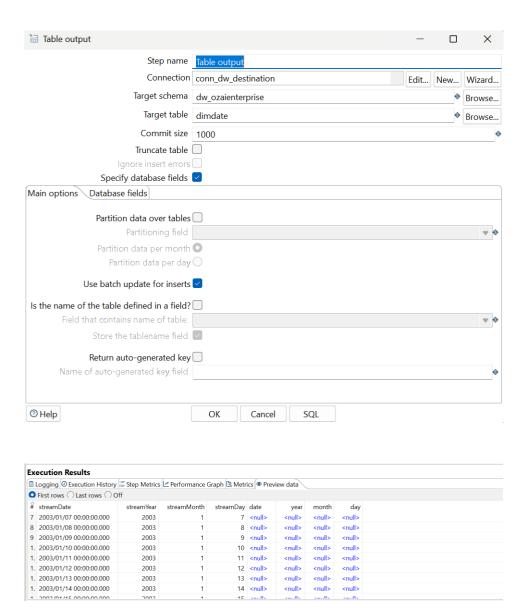


# Gambar konfigurasi filter rows

1. Hubungkan output dari filter rows menuju table output.



- 2. Konfigurasi pada table output adalah memberikan koneksi pada conn dw destination dengan schema dw legendvehicle dan table dimdate.
- 3. Aktifkan specify database fields.
- 4. Pada tab **Database fields**, mapping data input **streamDate**, **streamYear**, **streamMonth** dan **streamDay** dengan fields yang ada pada **dimDate**. Pada tahapan ini akan dilakukan insert data menuju tabel **dimDate**.



Gambar Konfiigurasi table output

1. cek isi table dimdate pada database. Jika sukses maka pada table dimdate akan terisi 1825 data.



Gambar isi tabel dimDate

### **TUGAS 1**

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.

Proses Objek	SS Data Input	SS Data Output	Keterangan
Generate rows	-	● First rows ○ Last rows ○ Off  # CurrentDate  1 01-01-2003  2 01-01-2003  3 01-01-2003  4 01-01-2003  5 01-01-2003  6 01-01-2003  7 01-01-2003  8 01-01-2003  0 01-01-2003	Menghasilkan kolom CurrentDate.
Add sequences	● First rows ○ Last rows ○ Off  # CurrentDate 1 01-01-2003 2 01-01-2003 3 01-01-2003 4 01-01-2003 5 01-01-2003 6 01-01-2003 7 01-01-2003 8 01-01-2003 8 01-01-2003 n_ost_ost_ost_ost_ost_ost_ost_ost_ost_ost	Execution Results	Kolom baru yang ditambahkan adalah incrementDay dengan nilai mulai dari 0 dan bertambah 1

			setiap barisnya.
Calculator	Execution Results	Execution Results   © Incomplete   Execution   Exec	incrementDay digunakan untuk menghitung streamDate dengan cara menambahkan hari ke tanggal awal (CurrentDate). Kemudian, streamDate dipecah menjadi beberapa kolom: streamYear, streamMonth, dan streamDay.
Select values	Execution Results		kolom yang dipilih adalah streamDate, streamYear, streamMonth, dan streamDay. Kolom lainnya diabaikan agar fokus pada data yang dibutuhkan saja.
Database lookup	Execution Results	Execution Results	terlihat kolom tambahan seperti date, year, month, dan day yang dicoba untuk diambil.
Filter rows	Securion Results		menyaring berdasarkan date, year, month dan day yang bernilai null

Table output	<b>Execution Results</b>							xecution Results
	B Logging © Execution History: - Step Metrics Id Performance Graph B: Metrics ◆ Preview data							Rogging © Decotion History     Step Metrics   & Performance Graph   B. Metrics   # Preview data   Kolom yang
	○ First rows ○ List rows ○ C	off						
	# streamDate	stream/Year	streamMonth		year	month	day	f streamDate streamNear streamMonth streamDay date year month day
	1 2003/01/01 00:00:00:00 2 2003/01/02 00:00:00:00	2003	1	1 crull>	<rul></rul>	<nul></nul>	coults	1 2003/01/01 00.00000000 2003 1 1 coulb coulb coulb coulb coulb coulb coulb
	3 2003/01/03 00:00:00:00	2003		2 englis	enulis	ends	enults	2 construction of the control of the
	4 2003/01/04 00:00:00:00	2003	1	4 ends	enulis	churs	crufts	
	5 2003/01/05 00:00:00:00	2003	1	5 coults	crolle	smile	enult-	4 2003/01/04 00:00:00:000 2003 1 4 cmilb cmilb cmilb antaranya:
	6 2003/01/06 00:00:00 000	2003	1	6 enults	<rul>&lt;</rul>	<nul></nul>	<rul></rul>	6 2003/01/06 00:00:00.000 2003 1 6 <null> <null> <null></null></null></null>
	7 2003/01/07 00:00:00:00:00 8 2003/01/08 00:00:00 000	2003		7 coults	enulis	<nul><li><nul><li></li></nul></li></nul>	enulis	7 2003/01/07 00:00:00:00 2003 1 7 crull> crull> crull> crull> crull>
	8 2003/01/08 00:00:00:000	2003	1	8 coults	coults	coults	enults	7 2003/01/07/000000000 2003 1 7 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
								streamYear,
								streamMontl
								streamDay,
								dan
								kemungkina
								Kemungkina
								kolom hasil
								lookup seper
								date, year,
								month, day