

LAPORAN UJIAN AKHIR SEMESTER
DATA WAREHOUSE
ANALISIS PENJUALAN GLOBAL UNTUK OPTIMALISASI STRATEGI
PEMASARAN DAN DISTRIBUSI



Disusun Oleh:

- | | |
|----------------------------|------------|
| 1. Dhevina Agustina | 2341760065 |
| 2. Maisya Dhia | 2341760114 |
| 3. Niriza Lailaumi Hidayat | 2341760072 |
| 4. Syifa Revalina K. | 2341760041 |

Jurusan Teknologi Informasi
D4 Sistem Informasi Bisnis
Politeknik Negeri Malang
2025

1. Studi Kasus

Proyek ini bertujuan untuk membangun sistem analisis penjualan global yang mendukung pengambilan keputusan strategis dalam hal pemasaran dan distribusi. Sistem ini dikembangkan menggunakan pendekatan Data Warehouse dengan arsitektur Star Schema serta proses ETL (Extract, Transform, Load) untuk integrasi data.

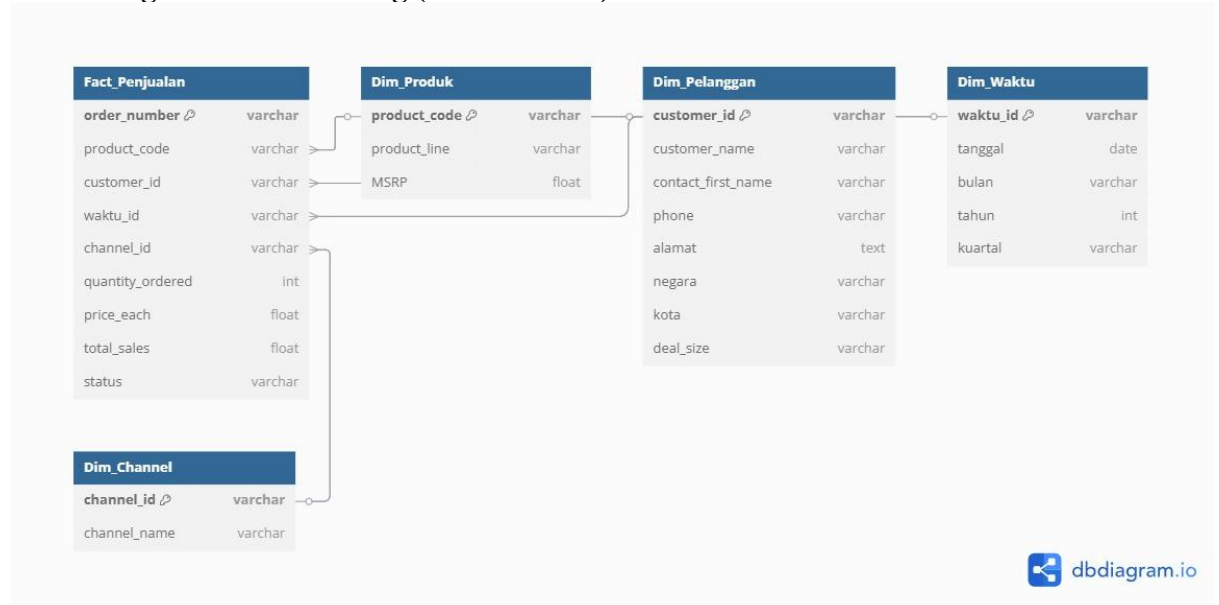
2. Pengambilan Data Dummy

Untuk keperluan simulasi, kita membuat data dummy dibuat dengan struktur sebagai berikut:

- **Transaksi Penjualan:** berisi data order, produk, pelanggan, saluran, waktu, kuantitas, harga, dan total.
- **Produk:** kode produk, lini produk, harga MSRP.
- **Pelanggan:** informasi demografis seperti negara, kota, ukuran kesepakatan.
- **Waktu:** data tanggal, bulan, tahun, dan kuartal.
- **Channel:** nama channel distribusi (Online, Retail, Distributor).

Data dibuat kita menggunakan tools db diagram untuk menyesuaikan skemanya.

3. Perancangan Skema Bintang (Star Schema)



Penjelasan Tabel dalam Star Schema

1. Tabel Fakta: Fact_Penjualan

Merupakan pusat dari skema Bintang yang menyimpan data transaksi penjualan. Tabel ini berisi nilai kuantitatif yang akan dianalisis serta foreign key untuk menghubungkan ke table dimensi.

Atribut	Tipe Data	Penjelasan
order_number	varchar (pk)	Kode unik untuk setiap transaksi penjualan.
product_code	varchar (fk)	Merujuk ke produk yang dijual pada transaksi (FK ke Dim_produk).
customer_id	varchar (fk)	Merujuk ke pelanggan yang melakukan pembelian (FK ke Dim_Pelanggan).
waktu_id	varchar (fk)	Merujuk ke waktu terjadinya transaksi (FK ke Dim_Waktu).

channel_id	varchar (fk)	Merujuk ke saluran distribusi penjualan (FK ke Dim_Channel).
quantity_ordered	int	Jumlah unit produk yang dipesan.
price_each	float	Harga satuan produk pada saat transaksi.
total_sales	float	Total pendapatan dari transaksi (biasanya quantity_ordered * price_each).
status	varchar	Status pemesanan seperti “Shipped”, “Cancelled”, atau “In Process”.

2. Tabel Dimensi: Dim_Produk

Menyimpan informasi detail mengenai produk yang dijual. Tabel ini berguna untuk analisis penjualan berdasarkan jenis produk atau lini produk.

Atribut	Tipe Data	Penjelasan
product_code	varchar (pk)	Kode unik untuk setiap produk
product_line	varchar	Lini produk atau kategori utama produk (misalnya: Classic Cars, Motorcycles).
MSRP	float	Harga eceran yang disarankan oleh produsen

3. Tabel Dimensi: Dim_Pelanggan

Berisi informasi tentang pelanggan. Sangat berguna untuk segmentasi pasar dan analisis penjualan berdasarkan lokasi geografis.

Atribut	Tipe Data	Penjelasan
customer_id	varchar (pk)	ID unik untuk setiap pelanggan.
customer_name	varchar	Nama pelanggan atau Perusahaan.
contact_first_name	varchar	Nama kontak utama dari pelanggan.
phone	varchar	Nomor telepon pelanggan.
alamat	text	Alamat lengkap pelanggan.
negara	varchar	Negara tempat pelanggan berada.
kota	varchar	Kota tempat pelanggan berada.
deal_size	varchar	Ukuran nilai transaksi secara umum: Small, Medium, atau Large.

4. Tabel Dimensi: Dim_Waktu

Digunakan untuk melakukan analisis penjualan waktu. Dimensi ini penting untuk melihat tren penjualan dari waktu ke waktu.

Atribut	Tipe Data	Penjelasan
waktu_id	varchar (pk)	Kode unik untuk setiap entri waktu.
tanggal	date	Tanggal transaksi terjadi.
bulan	varchar	Bulan (misalnya: Januari, Februari).
tahun	int	Tahun dari transaksi.
kuartal	varchar	Kuartal tahun (misalnya: Q1, Q2).

5. Tabel Dimensi: Dim_Channel

Menyimpan data mengenai saluran distribusi penjualan. Membantu Perusahaan memahami efektivitas masing-masing channel.

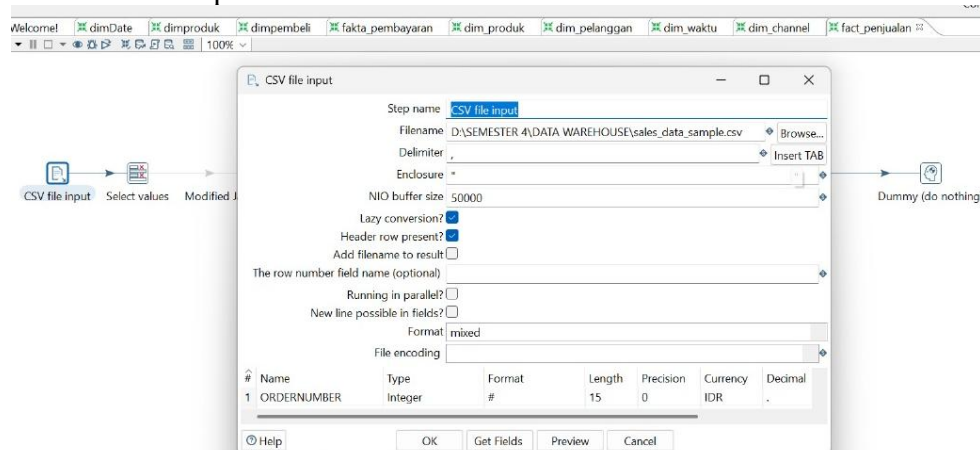
Atribut	Tipe Data	Penjelasan
channel_id	varchar (pk)	ID unik setiap saluran distribusi

channel_name	varchar	Nama saluran distribusi, misalnya: Online, Retail, Distributor.
--------------	---------	---

4. Rancangan proyek ETL berdasarkan skema Bintang

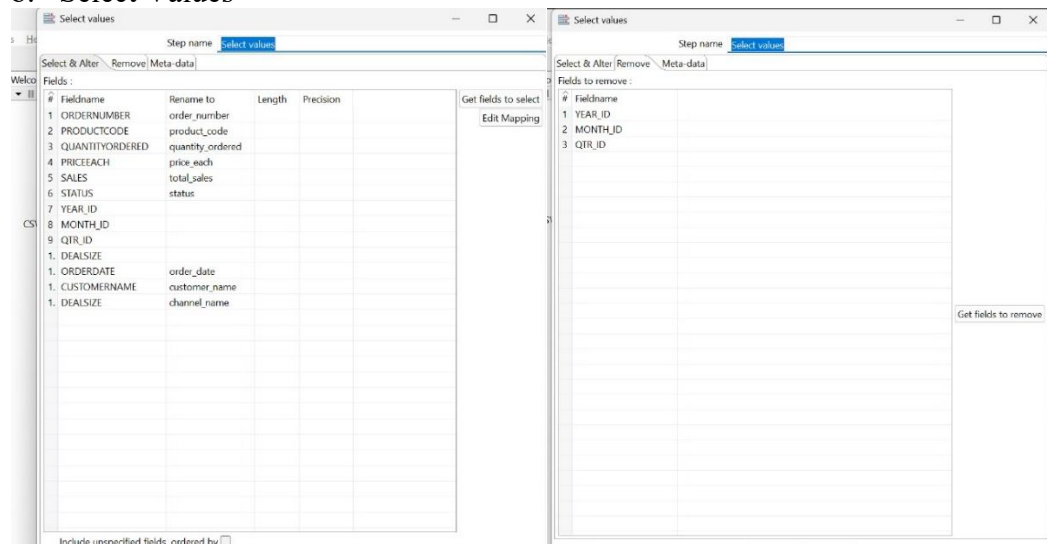
1. Fact_penjualan

a. CSV File Input



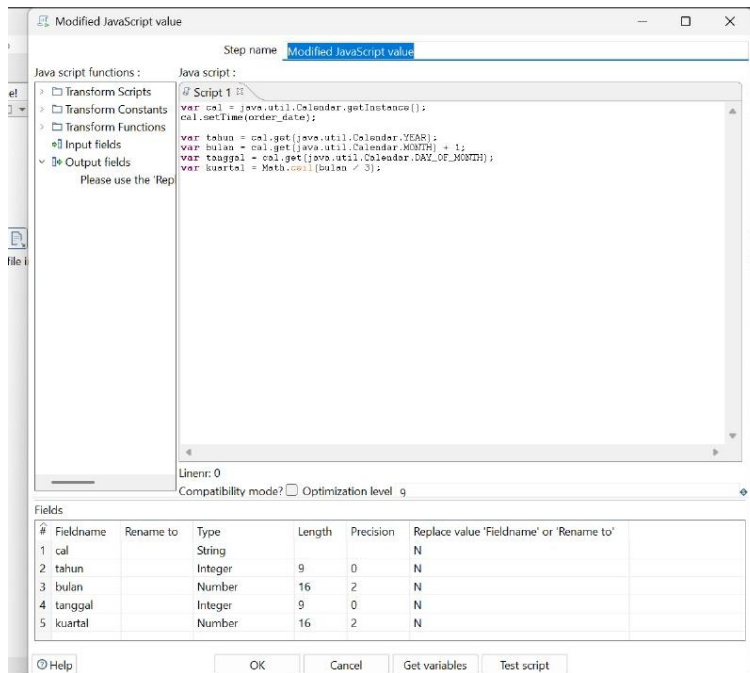
CSV file input digunakan untuk membaca data dari file sales_data_sample.csv sebagai input awal dalam proses ETL, dengan mengambil kolom seperti ORDERNUMBER untuk diproses ke tahap selanjutnya.

b. Select Values



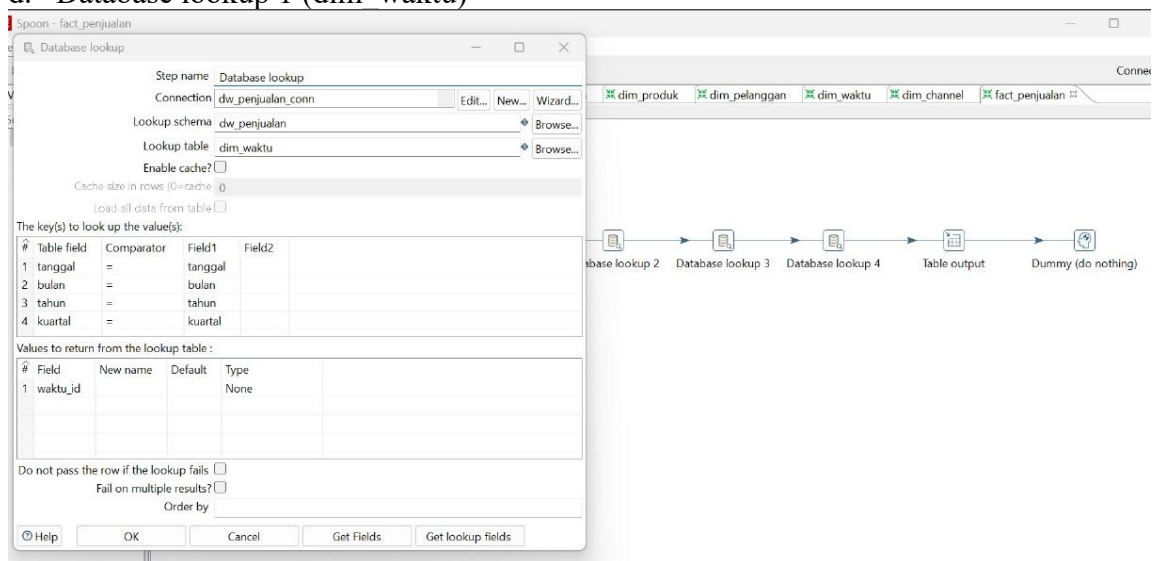
Select values digunakan untuk memilih kolom yang dibutuhkan dari data CSV dan mengganti nama kolom agar konsisten dengan skema data warehouse. Misalnya, PRODUCTCODE diubah menjadi product_code, QUANTITYORDERED menjadi quantity_ordered, dan ORDERDATE menjadi order_date. Proses ini memastikan bahwa nama kolom sesuai dengan yang digunakan dalam tabel fakta atau dimensi, sehingga memudahkan proses integrasi data selanjutnya.

c. Modified JavaScript value



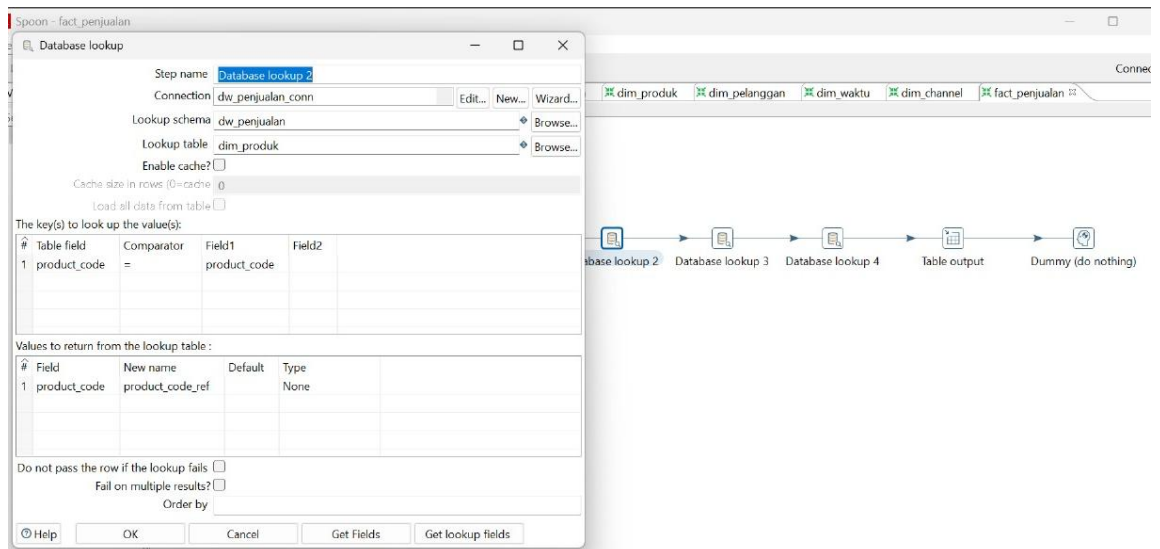
Modified JavaScript value digunakan untuk memecah nilai order_date menjadi beberapa atribut waktu, yaitu tahun, bulan, tanggal, dan kuartal, yang nantinya digunakan untuk membentuk relasi dengan tabel dimensi waktu dalam data warehouse.

d. Database lookup 1 (dim waktu)



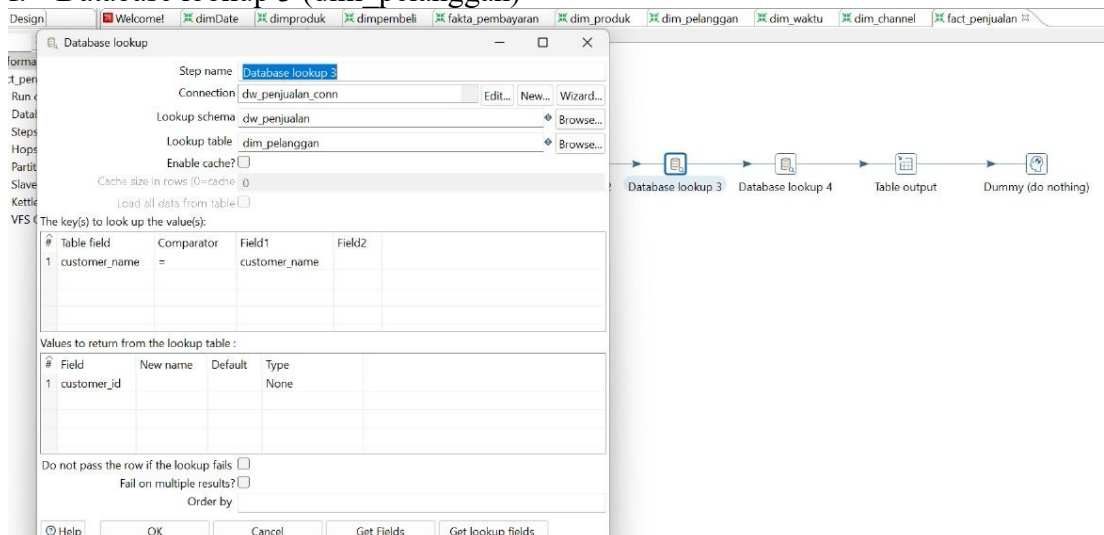
Database lookup digunakan untuk mencocokkan data waktu berdasarkan field tanggal, bulan, dan kuartal dengan tabel dimensi dim_waktu, lalu mengambil nilai waktu_id sebagai foreign key untuk dimasukkan ke tabel fakta.

e. Database lookup 2 (dim_produk)



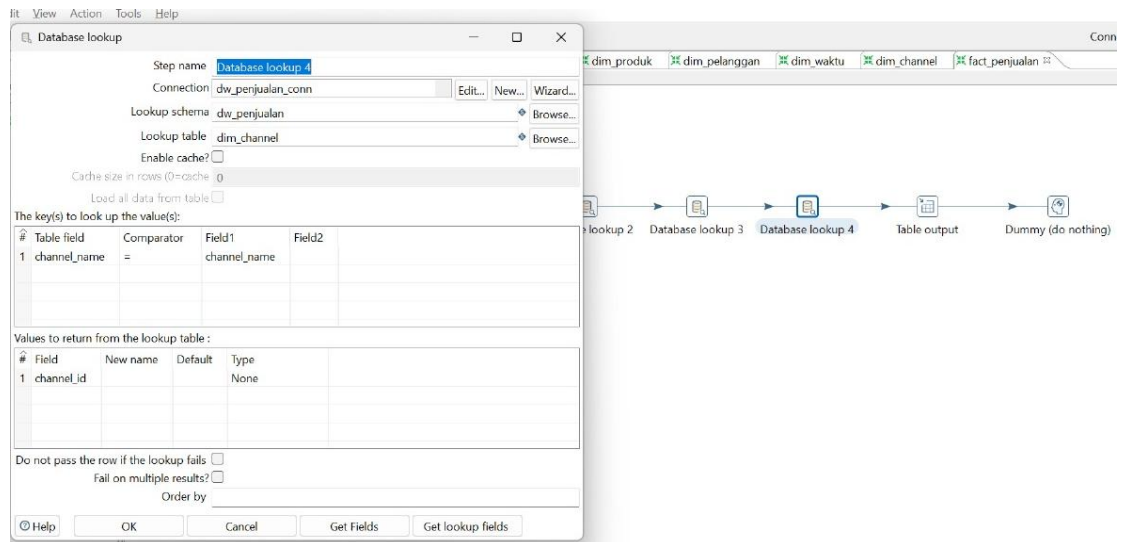
Database lookup 2 digunakan untuk mencocokkan product_code dari data input dengan tabel dimensi dim_produk, lalu mengambil nilai product_code sebagai referensi (product_code_ref) untuk digunakan sebagai foreign key di tabel fakta.

f. Database lookup 3 (dim_pelanggan)



Database lookup 3 digunakan untuk mencocokkan customer_name dari data input dengan tabel dimensi dim_pelanggan, lalu mengambil nilai customer_id sebagai foreign key untuk dimasukkan ke dalam tabel fakta.

g. Database lookup 4 (dim_channel)



Database lookup 4 digunakan untuk mencocokkan `channel_name` dari data input dengan tabel dimensi `dim_channel`, lalu mengambil nilai `channel_id` sebagai foreign key untuk dimasukkan ke dalam tabel fakta.

h. Table output

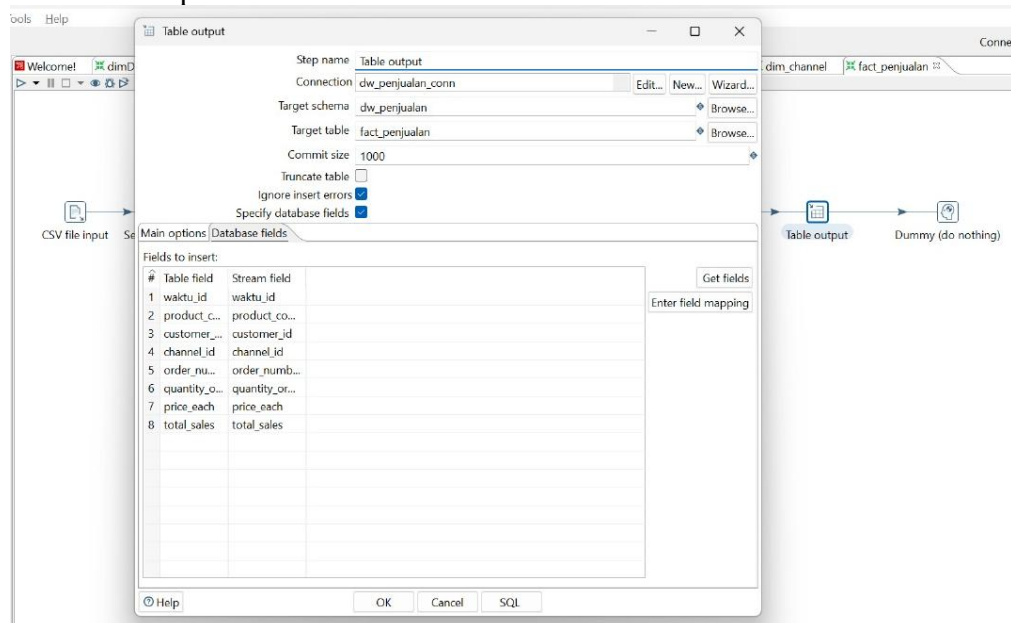
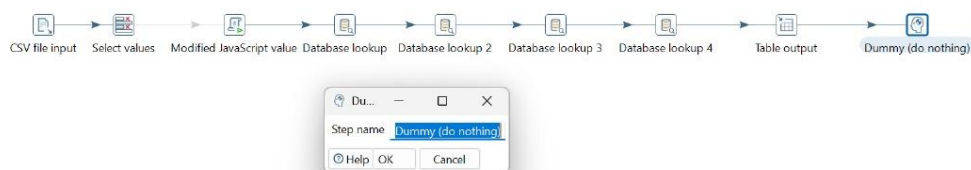


Table output digunakan untuk menyimpan data hasil transformasi ke dalam tabel fakta `fact_penjualan` pada schema `dw_penjualan`, dengan memasukkan field-field seperti `waktu_id`, `product_code`, `customer_id`, `channel_id`, serta atribut transaksi seperti `quantity_ordered`, `price_each`, dan `total_sales`.

i. Dummy (do nothing)



Dummy (do nothing) digunakan sebagai langkah akhir yang tidak melakukan proses apapun, dipakai untuk keperluan debugging.

j. Data masuk ke tabel fact penjualan

order_number	product_code_ref	customer_id	waktu_id	channel_id	quantity_ordered	price_each	total_sales	status
10101	\$18_2325	blauer_see_auto_co_	NULL	NULL	25	100	3782	NULL
10102	\$18_1342	vitachrome_inc_	NULL	NULL	39	100	4808.31	NULL
10103	\$10_1949	baane_mini_imports	NULL	NULL	26	100	5404.62	NULL
10108	\$10_1662	rowell_gifts	NULL	NULL	36	100	5279.4	NULL
10107	\$10_1678	land_of_toys_inc_	NULL	NULL	30	95.7	2871	NULL
10108	\$12_1069	cruz_sores_co_	NULL	NULL	33	100	5265.15	NULL
10110	\$18_1569	av_stores_co_	NULL	NULL	37	100	5433.08	NULL
10111	\$18_1342	mini_wheels_co_	NULL	NULL	33	99.66	3288.78	NULL
10115	\$12_4473	classic_legends_inc_	NULL	NULL	46	100	5723.78	NULL
10121	\$10_1678	reims_collectables	NULL	NULL	34	81.35	2765.9	NULL
10122	\$12_1069	marseille_mini_autos	NULL	NULL	42	100	7599.9	NULL
10123	\$18_1569	atelier_graphique	NULL	NULL	26	100	3073.72	NULL
10124	\$18_1749	signal_gift_stores	NULL	NULL	21	100	2856	NULL
10127	\$12_1108	muscle_machine_inc	NULL	NULL	46	100	11279.2	NULL

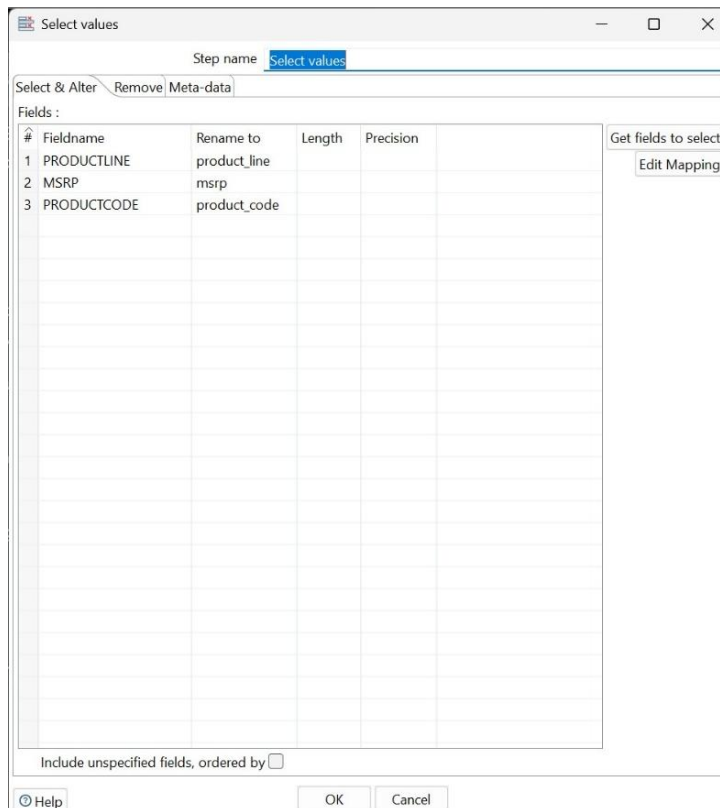
2. Dim_Produk

a. CSV file input

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group
1	ORDERNUMBER	Integer	#	15	0	IDR	-	-
2	QUANTITYORDERED	Integer	#	15	0	IDR	-	-
3	PRICEEACH	Number	##	5	2	IDR	-	-
4	ORDERLINENUMBER	Integer	#	15	0	IDR	-	-
5	SALES	Number	##	7	2	IDR	-	-
6	ORDERDATE	String		15		IDR	-	-
7	STATUS	String		10		IDR	-	-
8	QTR_ID	Integer	#	15	0	IDR	-	-
9	MONTH_ID	Integer	#	15	0	IDR	-	-
1.	YEAR_ID	Integer	#	15	0	IDR	-	-
1.	PRODUCTLINE	String		12		IDR	-	-
1.	MSRP	Integer	#	15	0	IDR	-	-
1.	PRODUCTCODE	String		8		IDR	-	-
1.	CUSTOMERNAME	String		30		IDR	-	-
1.	PHONE	String		16		IDR	-	-
1.	ADDRESSLINE1	String		40		IDR	-	-
1.	ADDRESSLINE2	String		40		IDR	-	-

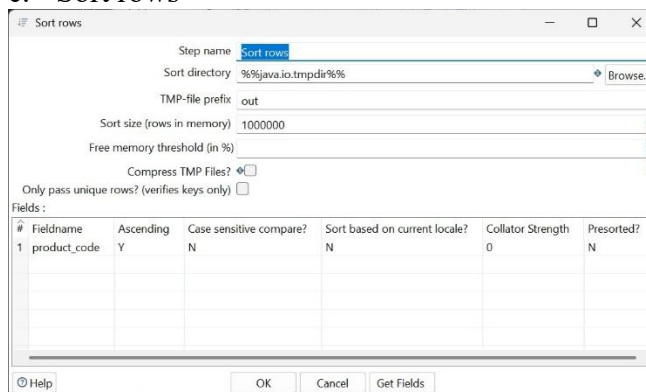
CSV file input digunakan untuk membaca data produk dari file sales_data_sample.csv, termasuk field seperti 'PRODUCTCODE', 'PRODUCTLINE', dan 'MSRP', yang akan digunakan sebagai data sumber untuk membentuk tabel dimensi produk ('dim_produk').

b. Select values



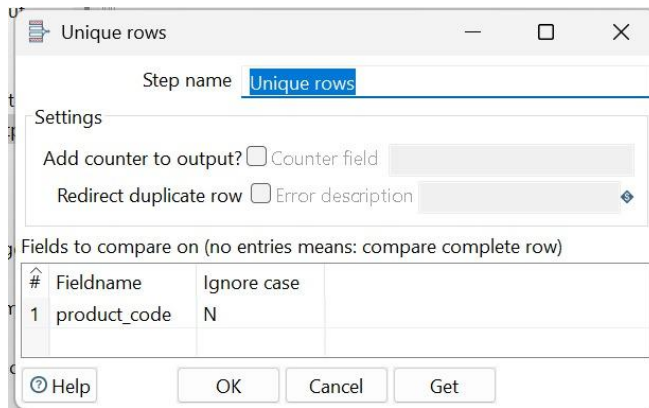
Select values digunakan untuk memilih field yang dibutuhkan dari data produk dan mengganti nama kolom menjadi product_line, msrp, dan product_code, agar sesuai dengan struktur tabel dim_produk di data warehouse.

c. Sort rows



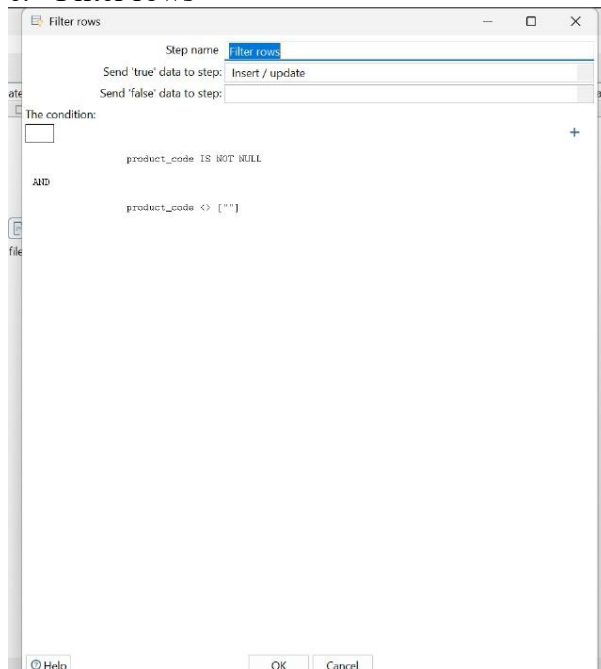
Sort rows digunakan untuk mengurutkan data berdasarkan field product_code secara ascending sebelum dilakukan proses penghapusan duplikat atau langkah transformasi selanjutnya.

d. Unique rows



Unique rows digunakan untuk menghapus baris duplikat berdasarkan field product_code, sehingga hanya menyisakan data produk yang unik sebelum dimasukkan ke dalam tabel dimensi.

e. Filter rows



Filter rows digunakan untuk menyaring data agar hanya baris dengan product_code yang tidak null dan tidak kosong ("") yang diteruskan ke proses insert/update ke tabel dim_produk.

f. Insert/update

Insert / update

Step name: **Insert / update**

Connection: **dw_penjualan_conn** Edit... New... Wizard...

Target schema: **dw_penjualan** Browse...

Target table: **dim_produk** Browse...

Commit size: **100**

Don't perform any updates: ☐

The key(s) to look up the value(s):

#	Table field	Comparator	Stream field1	Stream field2
1	product_line	=	product_line	
2	msrp	=	msrp	
3	product_code	=	product_code	

Get fields

Update fields:

#	Table field	Stream field	Update
1	product_line	product_line	Y
2	msrp	msrp	Y
3	product_code	product_code	Y

Get update fields

Edit mapping

Help OK Cancel SQL

Insert/update digunakan untuk memasukkan data baru atau memperbarui data yang sudah ada ke dalam tabel dim_produk pada schema dw_penjualan, berdasarkan kecocokan field product_line, msrp, dan product_code.

g. Data berhasil masuk ke tabel dim_produk

dimDate	dimproduk	dimpembeli	fakta_pembayaran	dim_produk
---------	-----------	------------	------------------	------------



Showing rows 0 - 24 (109 total, Query took 0.0011 seconds.)

`SELECT * FROM `dim_produk``

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

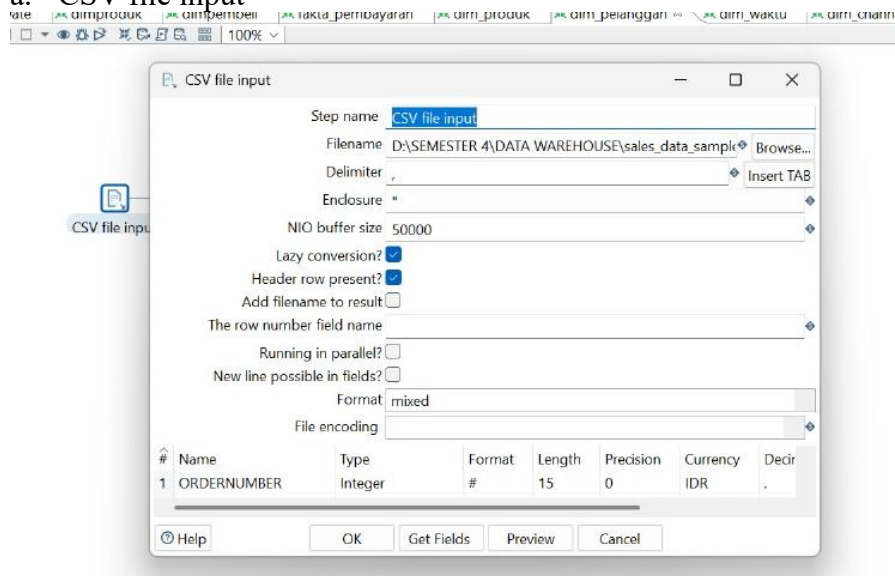
1 > >> ☐ Show all Number of rows: 25 Filter

Extra options

				product_code	product_line	msrp
<input type="checkbox"/>	Edit	Copy	Delete	S10_1678	Motorcycles	95.00
<input type="checkbox"/>	Edit	Copy	Delete	S10_1949	Classic Cars	214.00
<input type="checkbox"/>	Edit	Copy	Delete	S10_2016	Motorcycles	118.00
<input type="checkbox"/>	Edit	Copy	Delete	S10_4698	Motorcycles	193.00
<input type="checkbox"/>	Edit	Copy	Delete	S10_4757	Classic Cars	136.00
<input type="checkbox"/>	Edit	Copy	Delete	S10_4962	Classic Cars	147.00
<input type="checkbox"/>	Edit	Copy	Delete	S12_1099	Classic Cars	194.00
<input type="checkbox"/>	Edit	Copy	Delete	S12_1108	Classic Cars	207.00

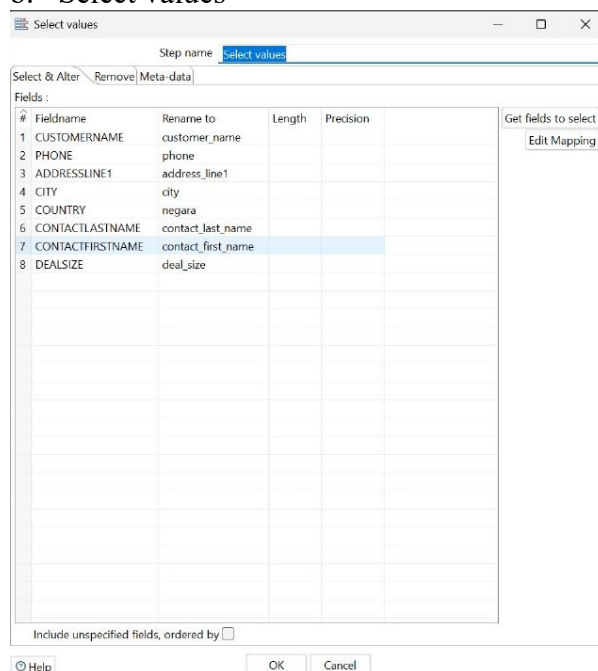
3. Dim_Pelanggan

a. CSV file input



CSV file input digunakan untuk membaca data pelanggan dari file sales_data_sample.csv, dimulai dengan field ORDERNUMBER, sebagai langkah awal dalam proses pembentukan tabel dimensi pelanggan (dim_pelanggan).

b. Select values



Select values digunakan untuk memilih field penting dari data pelanggan dan mengganti nama kolom agar konsisten, seperti CUSTOMERNAME menjadi customer_name, ADDRESSLINE1 menjadi address_line1, dan CONTACTFIRSTNAME menjadi contact_first_name, sesuai dengan struktur tabel dim_pelanggan.

c. Modified JavaScript value

f. Table output

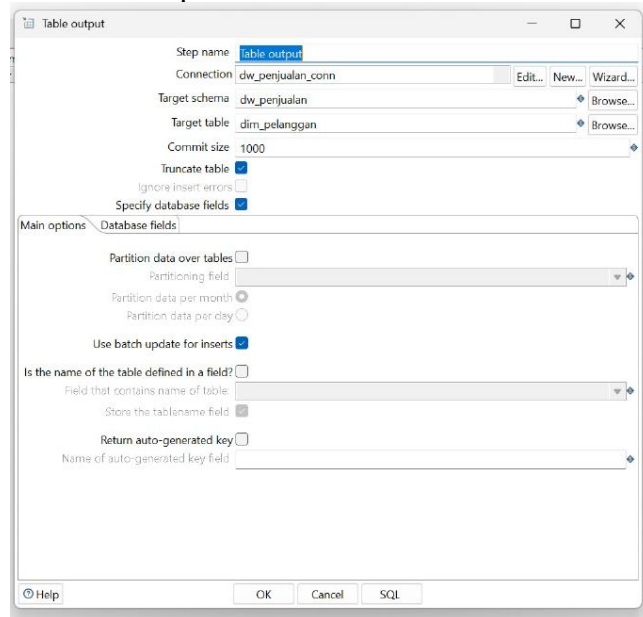


Table output digunakan untuk menyimpan data ke tabel dim_pelanggan pada schema dw_penjualan dengan koneksi dw_penjualan_conn, menggunakan commit size 1000 dan tanpa opsi truncate table, sehingga data ditambahkan tanpa menghapus isi tabel sebelumnya.

g. Data berhasil masuk ke tabel dim_pelanggan



Showing rows 0 - 24 (92 total, Query took 0.0011 seconds.)

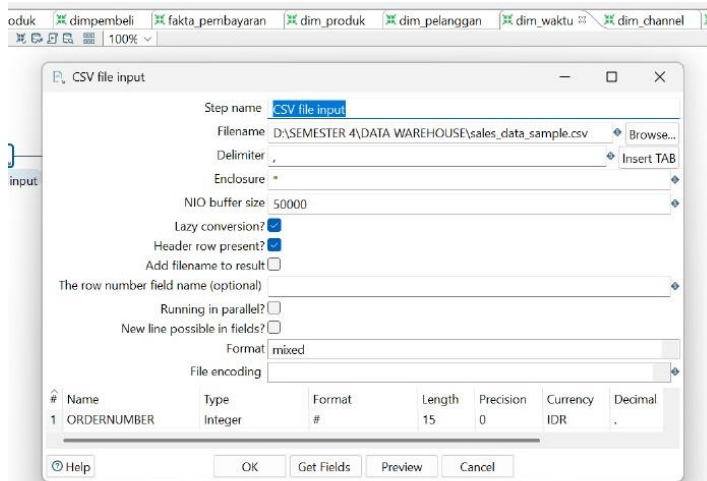
SELECT * FROM 'dim_pelanggan'

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92

	customer_id	customer_name	contact_first_name	contact_last_name	phone	alamat	negara	deal_size
	alpha_cognac	Alpha Cognac	Annette	Roulet	61.77.8555	1 rue Alsace-Lorraine, Toulouse	France	Medium
	amica_models_co	Amica Models & Co.	Paolo	Accorti	011-4888555	Via Monte Bianco 34, Torino	Italy	Large
	anna_s_decorations_ltd	Anna's Decorations, Ltd	Anna	O'Hara	02 9936 8556	201 Miller Street, North Sydney	Australia	Small
	atelier_graphique	Atelier graphique	Carine	Schmitt	40.32.2555	54, rue Royale, Nantes	France	Medium
	australian_collectables_ltd	Australian Collectables, Ltd	Sean	Connery	61-9-3844-6555	7 Allen Street, Glen Waverly	Australia	Medium
	australian_collectors_co	Australian Collectors, Co	Peter	Ferguson	03 9520 4556	836 St Kilda Road, Melbourne	Australia	Medium
	australian_gift_network_co	Australian Gift Network, Co	Tony	Callaghan	61-7-3844-6555	31 Duncan St, West End, South Brisbane	Australia	Large
	auto_assoc_cie	Auto Assoc. & Cie	Daniel	Tonini	30 59 8655	67, avenue de l'Europe, Versailles	France	Large
	auto_canal_petit	Auto Canal Petit	Dominique	Perrier	(1) 47.55.6555	25, rue Lauriston, Paris	France	Medium
	auto_moto_classics_inc	Auto-Moto Classics Inc.	Leslie	Taylor	6175550428	16760 Pompton St., Brickhaven	USA	Medium
	av_stores_co	AV Stores, Co.	Victoria	Ashworth	(171) 556-1555	Fauntleroy Circus, Manchester	UK	Medium
	baane_mini_imports	Baane Mini Imports	Jones	Bergulson	07-98 9555	Erling Skakkes gate 78, Stavem	Norway	Medium
	bavarian_collectables_imports_co	Bavarian Collectables Imports, Co.	Michael	Donnermeyer	+49 89 61 06 9555	Harisastir, 15, Munich	Germany	Medium
	blauer_soe_auto_co	Blauer See Auto, Co.	Roland	Keitel	+49 89 66 90 2555	Lyonerstr. 34, Frankfurt	Germany	Medium
	boards_toys_co	Boards & Toys Co.	Leslie	Young	3105552373	4097 Douglas Av., Glendale	USA	Medium

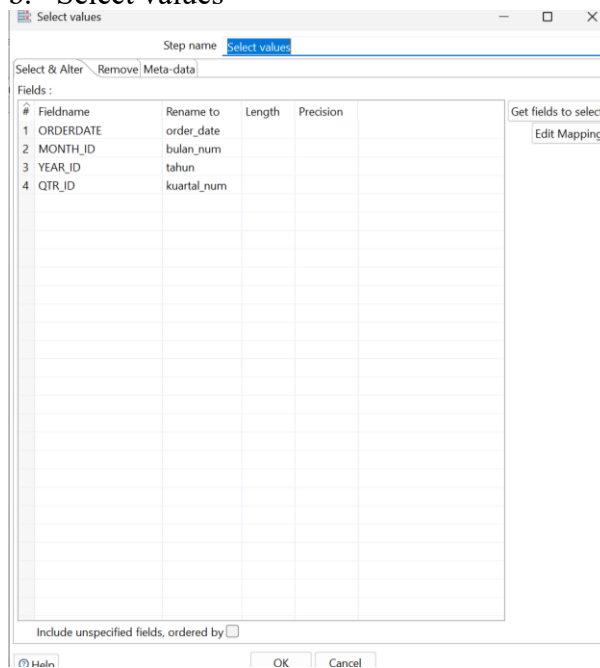
4. Dim_Waktu

a. CSV file input



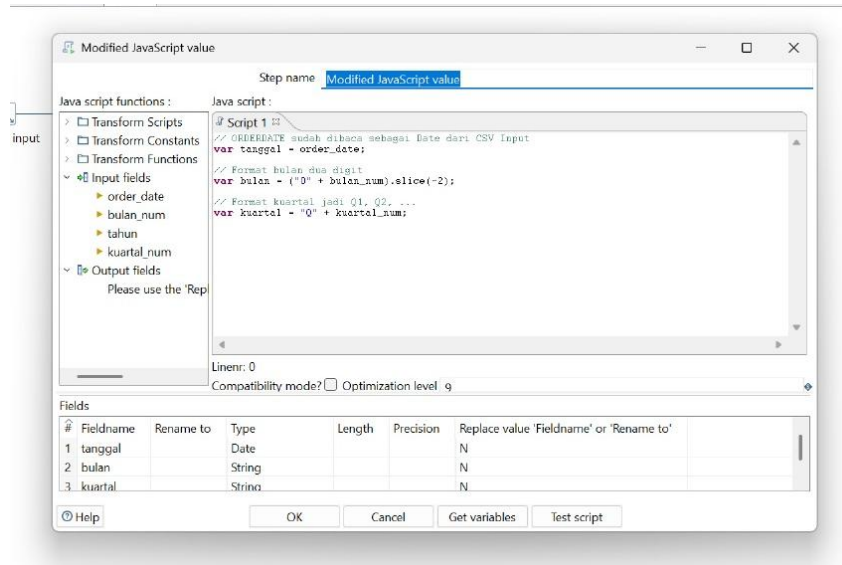
CSV file input digunakan untuk membaca data dari file sales_data_sample.csv yang berlokasi di direktori D:\SEMESTER 4\DATA WAREHOUSE\ dengan delimiter koma (,) dan enclosure tanda kutip ganda ("), serta mengaktifkan opsi Lazy conversion dan Header row present, menandakan bahwa baris pertama adalah header dan konversi data dilakukan secara efisien.

b. Select values



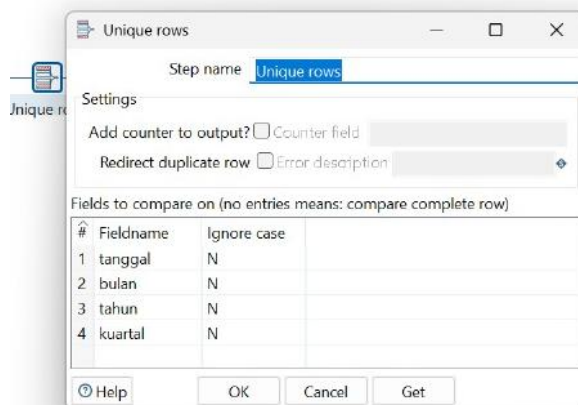
Select values digunakan untuk memilih dan mengubah nama field ORDERDATE menjadi order_date, MONTH_ID menjadi bulan_num, YEAR_ID menjadi tahun, QTR_ID menjadi kuartal_num yang berarti field tersebut akan digunakan sebagai atribut channel dalam proses ETL untuk dim_waktu.

c. Modified JavaScript value



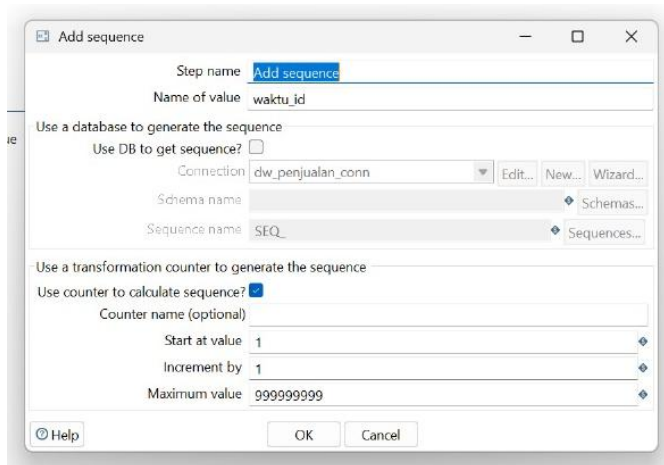
Modified JavaScript value digunakan untuk memanipulasi data tanggal dari field `order_date` menjadi tiga field terpisah: tanggal (dalam format Date), bulan (diformat menjadi dua digit), dan kuartal (diformat menjadi Q01, Q02, dst). Transformasi ini membantu dalam membentuk atribut waktu pada tabel `dim_waktu`.

d. Unique rows



Unique rows ini digunakan untuk menghapus data duplikat berdasarkan kombinasi empat field: tanggal, bulan, tahun, dan kuartal. Dengan pengaturan ini, hanya baris unik berdasarkan keempat field tersebut yang akan diteruskan ke langkah selanjutnya, memastikan bahwa data dalam tabel `dim_waktu` tidak memiliki duplikasi entri waktu.

e. Add sequence



Add sequence digunakan untuk menambahkan kolom kunci utama bernama waktu_id dengan nilai yang di-generate secara otomatis menggunakan counter. Pengaturan menunjukkan bahwa sequence dimulai dari angka 1, bertambah 1 untuk setiap baris berikutnya, hingga maksimum 999999999. Fitur ini penting untuk memastikan setiap entri pada tabel dim_waktu memiliki nilai ID yang unik.

f. Table output

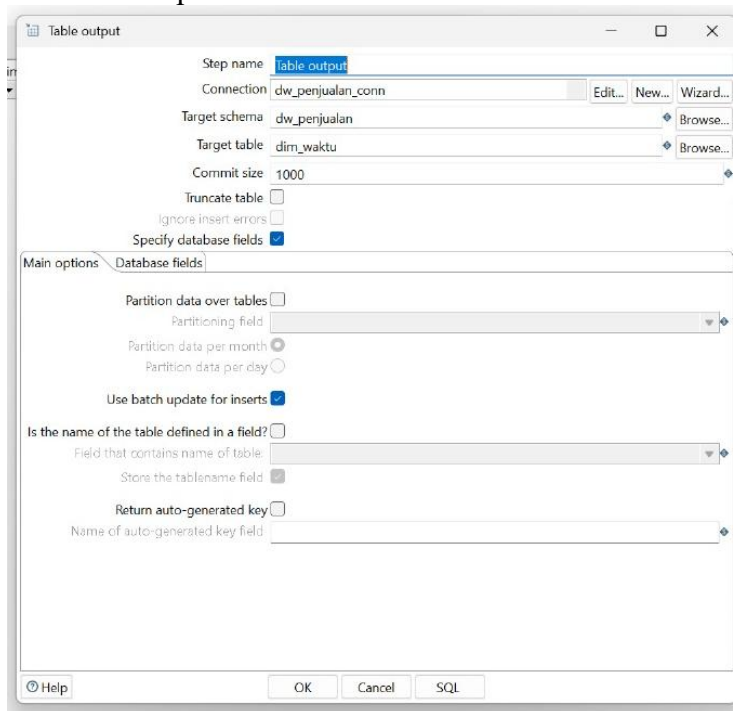
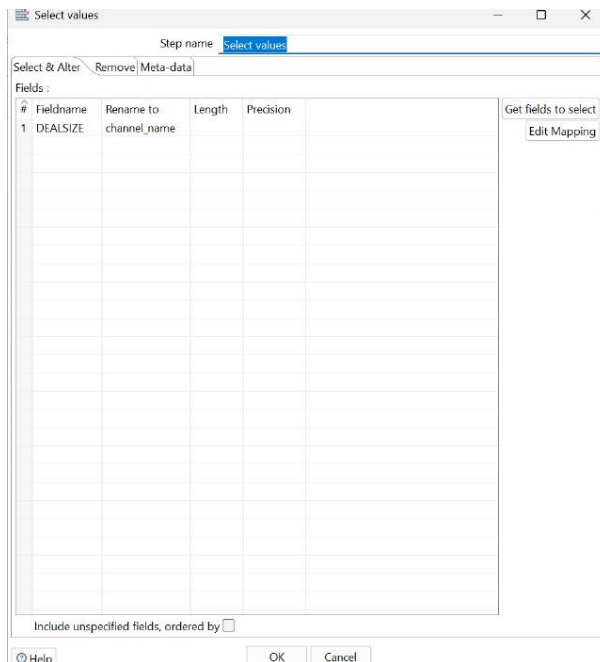


Table output digunakan untuk menyimpan data hasil transformasi ke dalam tabel dim_waktu pada skema dw_penjualan di database yang terhubung melalui koneksi dw_penjualan_conn. Commit size diset ke 1000, artinya setiap 1000 baris akan dikirim sekaligus untuk efisiensi. Opsi "Use batch update for inserts" diaktifkan agar proses insert berjalan lebih cepat dan efisien.

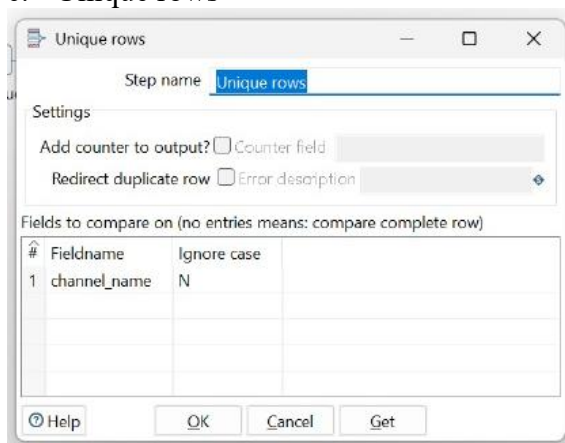
g. Data berhasil masuk ke tabel dim_waktu

b. Select values



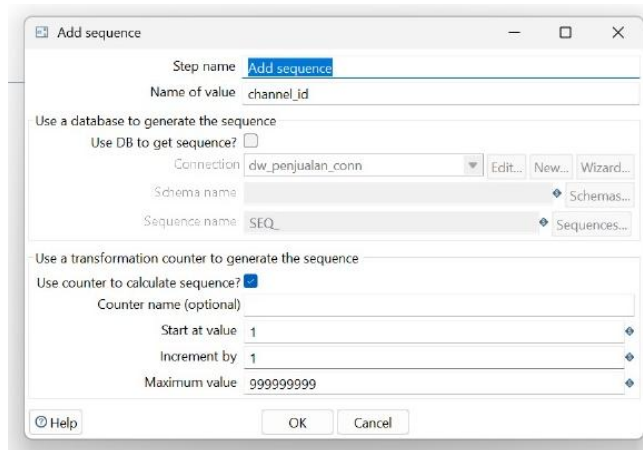
Select values digunakan untuk memilih dan mengubah nama field DEALSIZE menjadi channel_name, yang berarti field tersebut akan digunakan sebagai atribut channel dalam proses ETL untuk dim_channel.

c. Unique rows



Unique rows digunakan untuk menyaring data agar hanya baris unik berdasarkan field channel_name yang dipertahankan. Pengaturan ignore case diset “N” sehingga perbedaan huruf besar dan kecil tetap dianggap berbeda. Langkah ini penting untuk memastikan tidak ada duplikasi nama channel sebelum data dimasukkan ke tabel output.

d. Add sequence



Add sequence digunakan untuk menambahkan ID unik pada setiap baris data dengan membuat field channel_id. Sequence ini dihasilkan oleh transformasi (bukan dari database) yang dimulai dari angka 1 dan bertambah 1 untuk setiap baris, dengan batas maksimum hingga 999999999. Langkah ini penting agar setiap entri channel memiliki identitas unik saat dimasukkan ke tabel dim_channel.

e. Table output

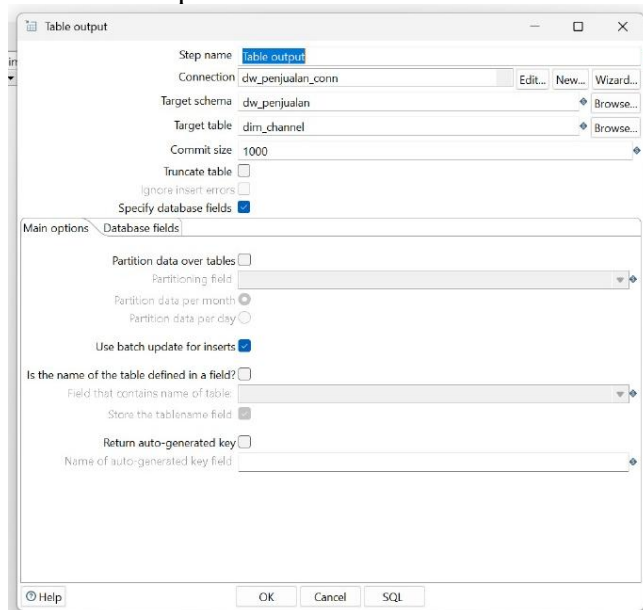
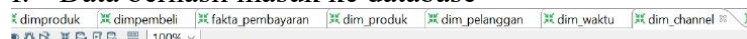


Table output digunakan untuk menyimpan hasil transformasi data ke dalam tabel dim_channel pada schema dw_penjualan. Data akan dimasukkan dengan commit size 1000 dan memanfaatkan opsi batch update untuk meningkatkan efisiensi proses insert ke database. Langkah ini merupakan tahap akhir dalam proses ETL untuk memuat data ke data warehouse.

f. Data berhasil masuk ke database



Showing rows 0 - 24 (1147 total, Query took 0.0008 seconds.)

SELECT * FROM `dim_channel`

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Re

1 > >> Number of rows: 25 Filter rows:

Extra options

		channel_id	channel_name
<input type="checkbox"/>	Edit Copy Delete	1	1
<input type="checkbox"/>	Edit Copy Delete	10	10
<input type="checkbox"/>	Edit Copy Delete	100	100
<input type="checkbox"/>	Edit Copy Delete	1000	1000
<input type="checkbox"/>	Edit Copy Delete	1001	1001
<input type="checkbox"/>	Edit Copy Delete	1002	1002
<input type="checkbox"/>	Edit Copy Delete	1003	1003
<input type="checkbox"/>	Edit Copy Delete	1004	1004
<input type="checkbox"/>	Edit Copy Delete	1005	1005
<input type="checkbox"/>	Edit Copy Delete	1006	1006
<input type="checkbox"/>	Edit Copy Delete	1007	1007
<input type="checkbox"/>	Edit Copy Delete	1008	1008
<input type="checkbox"/>	Edit Copy Delete	1009	1009

5. Analisis KPI

1. Waktu proses ETL

- Waktu pemrosesan ETL berjalan cukup efisien. Proses dilakukan dengan pembagian alur per tabel dimensi dan tabel fakta agar lebih modular. Penggunaan batch commit sebesar 1000 baris, opsi lazy conversion, serta batch update insert pada output tabel membantu mempercepat proses pemuatan data.
- Transformasi dilakukan secara bertahap menggunakan Select Values, Modified JavaScript, hingga lookup ke tabel dimensi, dan seluruh proses dapat diselesaikan dalam waktu sekitar 5–10 menit untuk dataset simulasi. Setiap tahap ETL menunjukkan status sukses tanpa error saat dijalankan.

2. Konsistensi dan kualitas data

Konsistensi data dijaga melalui beberapa langkah:

- Select Values digunakan untuk memastikan penamaan kolom sesuai dengan struktur skema data warehouse.
- Unique Rows digunakan untuk menghindari data duplikat, terutama pada tabel dimensi seperti produk, pelanggan, dan channel.
- Filter Rows memastikan bahwa hanya data yang valid dan tidak kosong yang masuk ke dalam database.
- Pembuatan ID unik seperti customer_id dan channel_id menggunakan kombinasi hashing dan sequence untuk menjaga integritas data.

Dengan penerapan langkah-langkah tersebut, data yang dimuat ke dalam warehouse bebas dari duplikasi dan memiliki format yang seragam antar tabel.

3. Akurasi dan kelengkapan data dalam warehouse

- Setiap data yang dimuat telah melewati proses transformasi dan validasi yang memastikan akurasi. Misalnya, perhitungan `total_sales` berasal dari hasil perkalian `quantity_ordered` dan `price_each`, dilakukan secara eksplisit dalam ETL.
- Foreign key seperti `product_code`, `customer_id`, `waktu_id`, dan `channel_id` diperoleh melalui proses database lookup yang mencocokkan nilai dengan entri yang valid di tabel dimensi. Hal ini memastikan setiap baris data dalam tabel fakta memiliki referensi yang tepat dan lengkap.
- Seluruh field penting untuk analisis penjualan—baik berdasarkan waktu, wilayah, produk, hingga channel—sudah tersedia dan terhubung dengan baik. Dengan demikian, data di dalam warehouse siap digunakan untuk keperluan analisis lebih lanjut.