

PROJEK BUSINESS INTELLIGENT

**RANCANG BANGUN ETL DAN DATA WAREHOUSE MENGGUNAKAN DATASET
RETAIL KAGGLE (Northwind Traders)**

Dosen Pembimbing:

Endah Septa Sintiya, S.Pd., M.Kom.



KELAS : TI-3B

Disusun oleh:

Kelompok 6

Azzahra Attaqina	(2341720224)
Luthfiyyah Adzka N.S	(2341720148)
Najwa Alya Nurizzah	(2341720230)
Tri Sukma Sarah	(2341720051)
Wahyu Rizky Cahyana	(2341720006)

JURUSAN TEKNOLOGI INFORMASI

PROGRAM STUDI D-IV TEKNIK INFORMATIKA

POLITEKNIK NEGERI MALANG

2025

DAFTAR ISI

DAFTAR ISI.....	1
BAB 1	
PENDAHULUAN.....	3
1.1 Latar Belakang.....	3
1.2 Rumusan Masalah.....	4
1.3 Tujuan Proyek.....	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Proyek.....	5
BAB 2	
STUDI KASUS DAN DESKRIPSI DATA.....	6
2.1 Pemilihan Studi Kasus.....	6
2.2 Deskripsi Umum Studi Kasus.....	6
2.3 Sumber Dataset.....	6
2.4 Struktur Dataset.....	7
2.5 Alasan Pemilihan Studi Kasus.....	7
BAB 3	
PERANCANGAN DATA WAREHOUSE (STAR SCHEMA).....	9
3.1 Identifikasi Tabel Fakta.....	9
3.2 Identifikasi Tabel Dimensi.....	10
Tabel Dimensi yang digunakan.....	11
1. dim_customer.....	11
Kesimpulan dimensi : tabel dimensi yang digunakan dalam proyek ini ada 5, yaitu: dim_customer, dim_employee, dim_product, dim_shipper, dan dim_time.	12
3.3 Deskripsi Atribut Fakta dan Dimensi.....	12
3.4 Diagram Star Schema.....	15
3.5 Hubungan Antar Tabel (Relational Mapping).....	16
BAB 4	
PERANCANGAN PROSES ETL.....	17
4.1 Gambaran Umum ETL.....	17
4.2 Tahap Extract.....	17
4.2.1 Sumber Data.....	17
4.2.2 Teknik dan Tools Extract.....	17
4.3 Tahap Transform.....	24
4.3.1 Data Cleaning.....	24
4.3.2 Data Transformation.....	29
4.3.3 Data Aggregation.....	38
4.3.4 Data Integration.....	38
4.4 Tahap Load.....	42
4.4.1 Target Data Warehouse.....	42
4.4.2 Mapping Fakta dan Dimensi.....	46

BAB 5	
IMPLEMENTASI ETL PIPELINE.....	50
5.1 Lingkungan dan Tools yang Digunakan.....	50
5.2 Struktur Folder Proyek.....	51
5.3 Desain Pipeline ETL.....	51
5.4 Penjelasan Setiap Komponen Step dalam Pipeline.....	52
5.5 Hasil Eksekusi ETL.....	76
5.6 Penyimpanan File .etl.....	82
Pada proyek ini, hasil implementasi pipeline ETL disimpan dalam format file bawaan Pentaho Data Integration. File ETL terdiri dari :.....	82
- Transformation (.ktr), yaitu file yang berisi rangkaian step ETL untuk masing-masing tabel dimensi dan tabel fakta.....	82
BAB 6	
HASIL ANALISIS DAN VISUALISASI.....	83
6.1 Data Warehouse Final.....	83
6.1.1 Tabel Dimensi.....	83
6.1.2 Tabel Fakta.....	85
6.1.3 ANALISIS KPI.....	86
6.1.4 Visualisasi.....	87
BAB 7	
KESIMPULAN DAN SARAN.....	89
7.1 Kesimpulan.....	89
7.2 Saran.....	89

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan dunia bisnis saat ini menuntut perusahaan untuk mampu mengambil keputusan dengan cepat dan tepat berdasarkan data yang dimiliki. Pada umumnya, data perusahaan tersimpan dalam sistem operasional yang disebut OLTP (Online Transaction Processing). OLTP berfungsi untuk mendukung kegiatan transaksi harian seperti pencatatan pemesanan barang, pengelolaan pelanggan, pengaturan stok produk, hingga proses pengiriman barang. Walaupun OLTP sangat efektif untuk transaksi, struktur tabelnya bersifat terpisah-pisah dan saling terhubung secara relasional, sehingga kurang efisien jika digunakan untuk kebutuhan analisis.

Dalam konteks Business Intelligence (BI), perusahaan membutuhkan sistem yang mampu menganalisis data secara cepat untuk melihat pola penjualan, tren waktu, produk terlaris, hingga performa karyawan. Untuk itu, dibutuhkan data warehouse (OLAP) yang disusun menggunakan model analitis seperti star schema. Model ini mengorganisasi data ke dalam tabel fakta dan tabel dimensi sehingga query agregasi menjadi lebih sederhana dan performa analisis meningkat.

Pada proyek ini, dataset yang digunakan adalah dataset Northwind, yaitu dataset publik yang menggambarkan bisnis penjualan produk retail. Dataset ini memiliki tabel transaksi penjualan (orders dan order_details) serta tabel pendukung lainnya seperti products, categories, customers, employees, dan shippers. Dataset tersebut kemudian dipisahkan menjadi dua area, yaitu OLTP sebagai data sumber dan OLAP sebagai data warehouse untuk analisis. Proses perpindahan data dari OLTP ke OLAP dilakukan melalui tahapan ETL (Extract, Transform, Load).

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah pada proyek ini adalah sebagai berikut:

1. Bagaimana memahami dataset Northwind sebagai data operasional (OLTP) yang akan dijadikan sumber data warehouse?
2. Bagaimana merancang skema data warehouse menggunakan star schema sehingga mampu mendukung proses analisis penjualan?
3. Bagaimana menyusun proses ETL yang meliputi extract, transform, dan load untuk memindahkan data dari OLTP ke OLAP?
4. Bagaimana membangun tabel dimensi dan tabel fakta penjualan yang konsisten serta siap digunakan untuk analisis KPI?
5. Bagaimana menghasilkan informasi analitis seperti total penjualan per periode dan top produk terlaris berdasarkan data warehouse yang dibangun?

1.3 Tujuan Proyek

Adapun tujuan dari proyek Business Intelligence ini adalah:

1. Menggunakan dataset Northwind sebagai studi kasus untuk memahami alur implementasi data warehouse pada bisnis retail/penjualan.
2. Merancang data warehouse berbasis star schema dengan menentukan tabel fakta utama serta tabel dimensi pendukung yang sesuai.
3. Membuat rancangan proses ETL yang jelas dan terstruktur untuk memastikan data dari OLTP dapat diubah menjadi data analitis pada OLAP.
4. Mengimplementasikan ETL untuk membangun seluruh dimensi (dim_customer, dim_employee, dim_product, dim_shipper, dim_time) dan tabel fakta (fact_sales).
5. Menyediakan data warehouse yang dapat digunakan untuk perhitungan KPI dan pembuatan laporan/dashboard penjualan.

1.4 Batasan Masalah

Agar pembahasan lebih fokus dan sesuai ruang lingkup proyek, maka batasan masalah pada laporan ini yaitu:

1. Dataset yang digunakan hanya dataset Northwind meliputi tabel categories, products, customers, employees, shippers, orders, dan order_details.
2. Model data warehouse yang dibangun berbentuk star schema sederhana dengan satu tabel fakta penjualan dan lima tabel dimensi pendukung.
3. Proses ETL berfokus pada integrasi data historis untuk kebutuhan analisis penjualan, bukan untuk transaksi real-time.
4. Implementasi dan evaluasi BI dibatasi pada pembuatan KPI utama penjualan sesuai instruksi proyek.

1.5 Manfaat Proyek

Manfaat dari proyek ini antara lain:

1. Memberikan pemahaman praktis tentang cara mengonversi data OLTP ke OLAP melalui perancangan data warehouse.
2. Mempermudah proses analisis bisnis, khususnya pada kasus penjualan, karena struktur OLAP lebih efisien untuk kebutuhan agregasi.
3. Menjadi dasar untuk pembuatan laporan dan dashboard BI yang berisi informasi penting seperti tren penjualan, produk unggulan, dan performa karyawan.
4. Melatih kemampuan teknis dalam menyusun dan menerapkan ETL sebagai bagian utama dalam implementasi Business Intelligence.

BAB 2

STUDI KASUS DAN DESKRIPSI DATA

2.1 Pemilihan Studi Kasus

Studi kasus pada proyek ini diambil dari dataset publik Northwind. Pemilihan dataset ini didasarkan pada kesesuaiannya dengan karakteristik tugas BI, yaitu dataset harus memiliki data transaksi dan data master yang saling berkaitan. Northwind menggambarkan perusahaan perdagangan produk yang melayani pelanggan dari berbagai wilayah. Dataset tersebut banyak digunakan pada pembelajaran BI karena strukturnya cukup lengkap untuk membangun sebuah data warehouse penjualan.

Selain itu, Northwind memiliki entitas transaksi yang jelas (orders dan order_details) serta entitas pendukung seperti pelanggan, karyawan, kategori, produk, dan pengirim. Dengan demikian, dataset ini cocok untuk dimodelkan menjadi star schema dan dianalisis menggunakan KPI penjualan.

2.2 Deskripsi Umum Studi Kasus

Dataset Northwind berisi data aktivitas bisnis penjualan, mulai dari pemesanan produk oleh pelanggan, proses pencatatan order, pencantuman detail produk yang dipesan, hingga proses pengiriman barang melalui jasa shipper. Setiap order memiliki informasi pelanggan yang membeli, karyawan yang menangani, tanggal order, serta shipper yang mengirim. Order kemudian dirinci pada order_details yang menjelaskan produk apa saja yang dibeli, berapa jumlahnya, berapa harga satuannya, serta diskon yang diberikan.

2.3 Sumber Dataset

Dataset yang digunakan pada proyek ini adalah Northwind Sample Database yang diambil dari Kaggle, yaitu dataset publik yang sering dipakai sebagai contoh studi kasus bisnis retail/penjualan. Northwind menyediakan data transaksi pemesanan produk beserta data master pendukung seperti pelanggan, karyawan, produk, kategori, dan pengiriman.

Pada proyek ini, dataset Northwind diperoleh dalam bentuk file SQL (categories.sql, products.sql, customers.sql, employees.sql, shippers.sql, orders.sql, dan order_details.sql). Seluruh file tersebut kemudian di-*import* ke dalam database OLTP cy6er_bi menggunakan DBMS MySQL, sehingga menjadi sumber utama dalam proses ETL menuju data warehouse OLAP cy6er_bi.olap

2.4 Struktur Dataset

Ringkasan umum struktur dataset (Northwind Traders)

Nama Tabel	Peran / Fungsi tabel di OLTP	Relasi Utama
categories	Master data kategori produk (mengelompokkan produk berdasarkan jenisnya).	1 kategori dapat memiliki banyak produk (1-M ke products).
products	Menyimpan data produk seperti nama produk, harga, dan status discontinued.	Banyak produk berada dalam satu kategori (M-1 ke categories) dan bisa muncul di banyak order_details (1-M ke order_details).
customers	Menyimpan profil pelanggan seperti nama perusahaan, kontak, kota, dan negara.	1 customer dapat melakukan banyak order (1-M ke orders).
employees	Menyimpan data karyawan/pegawai yang menangani transaksi penjualan.	1 employee dapat menangani banyak order (1-M ke orders)
shippers	Menyimpan data jasa/perusahaan pengiriman barang.	1 shipper dapat mengirim banyak order (1-M ke orders).
orders	Tabel transaksi header berisi info umum order (customer, employee, tanggal, shipper, freight).	1 order dimiliki 1 customer, 1 employee, 1 shipper (M-1), dan punya banyak detail item (1-M ke order_details)
order_details	Tabel transaksi detail berisi rincian produk dalam order (unit Price, quantity, discount).	Banyak detail terhubung ke 1 order (M-1 ke orders) dan 1 produk bisa muncul di banyak detail (M-1 ke products).

2.5 Alasan Pemilihan Studi Kasus

Alasan pemilihan dataset Northwind adalah:

1. Dataset bersifat publik, legal digunakan, dan mudah diakses.
2. Memiliki struktur OLTP yang jelas sehingga memudahkan perancangan OLAP.
3. Data transaksi penjualan lengkap, mendukung pembuatan tabel fakta penjualan.
4. Banyak referensi penggunaan Northwind pada BI, sehingga valid untuk studi kasus.

BAB 3

PERANCANGAN DATA WAREHOUSE (STAR SCHEMA)

3.1 Identifikasi Tabel Fakta

Tabel fakta merupakan inti dari star schema karena berisi data transaksi utama yang menjadi bahan perhitungan analisis dan KPI. Semua nilai numerik yang ingin dihitung (misalnya total penjualan, jumlah barang terjual, rata-rata diskon, dan sebagainya) akan disimpan di tabel fakta. Karena itu, menentukan tabel fakta harus benar-benar mewakili aktivitas bisnis inti.

Pada kasus dataset Northwind, aktivitas bisnis inti yang dianalisis adalah penjualan produk. Transaksi penjualan paling detail tercatat pada tabel order_details (OLTP), karena tabel ini memuat item per produk yang dibeli pada setiap order, lengkap dengan harga, jumlah, dan diskon. Maka dari itu, tabel fakta yang dirancang dalam data warehouse adalah:

Tabel Fakta yang digunakan

1. fact_sales (*tabel fakta utama*)

A. Sumber OLTP:

- a. orders (informasi order header seperti customer, employee, shipper, tanggal order)
- b. order_details (detail transaksi produk: unit Price, quantity, discount)
- c. products (untuk memastikan item valid dan referensi produk)

B. Grain (tingkat detail):

1 baris pada fact_sales merepresentasikan 1 produk dalam 1 order.

Artinya, jika satu order membeli 5 produk, maka akan ada 5 baris di fact_sales.

2. Isi fact_sales :

A. Measure (nilai numerik yang dianalisis):

- a. unit_price → harga satuan produk saat transaksi
- b. quantity → jumlah produk yang dibeli
- c. discount → potongan harga dalam transaksi

- d. total_sales → nilai penjualan bersih setelah diskon dengan rumus:

$$\text{total_sales} = \text{unit_price} \times \text{quantity} \times (1 - \text{discount})$$
$$\text{total_sales} = \text{unit_price} \times \text{quantity} \times (1 - \text{discount})$$

B. **Foreign Key (penghubung ke dimensi):**

- a. customer_id → menghubungkan ke dim_customer
- b. employee_id → menghubungkan ke dim_employee
- c. product_id → menghubungkan ke dim_product
- d. shipper_id → menghubungkan ke dim_shipper
- e. time_id → menghubungkan ke dim_time

Dengan struktur ini, fact_sales mampu merepresentasikan seluruh konteks penjualan, yaitu:

- a. siapa yang membeli (customer)
- b. produk apa yang dibeli (product & category)
- c. kapan transaksi terjadi (time)
- d. siapa yang melayani (employee)
- e. dan dikirim oleh siapa (shipper)

Kesimpulan : tabel fakta yang digunakan dalam proyek ini hanya satu, yaitu fact_sales, karena fokus analisis kalian adalah penjualan. Fact_sales menjadi pusat star schema yang nanti dipakai untuk semua KPI utama.

3.2 Identifikasi Tabel Dimensi

Tabel dimensi digunakan untuk menyimpan atribut deskriptif yang menjadi sudut pandang analisis. Dimensi berfungsi menjawab pertanyaan “siapa, apa, kapan, oleh siapa, dikirim oleh siapa” pada transaksi penjualan. Dengan adanya dimensi, data pada tabel fakta dapat dianalisis dari berbagai sisi seperti pelanggan, produk, waktu, karyawan, maupun pengiriman.

Pada proyek Northwind ini, tabel dimensi yang digunakan dalam OLAP adalah sebagai berikut:

Tabel Dimensi yang digunakan

1. **dim_customer**

- a. **Sumber OLTP:** customers
- b. **Fungsi:** menyimpan informasi pelanggan agar analisis dapat dilakukan berdasarkan customer dan lokasi (kota/negara).
- c. **Atribut utama:**
customer_id, company_name, contact_name, contact_title, city, country

2. **dim_employee**

- a. **Sumber OLTP:** employees
- b. **Fungsi:** menyimpan data karyawan/penjual sehingga performa penjualan bisa dianalisis per employee.
- c. **Atribut utama:**
employee_id, employee_name, title, city, country

3. **dim_product**

- a. **Sumber OLTP:** products join categories
- b. **Fungsi:** menyimpan detail produk beserta kategori agar analisis produk dan kategori terlaris lebih mudah dilakukan.
- c. **Atribut utama:**
product_id, product_name, category, unit_price

4. **dim_shipper**

- a. **Sumber OLTP:** shippers
- b. **Fungsi:** menyimpan data perusahaan pengiriman sehingga dapat dianalisis jasa pengiriman mana yang paling sering digunakan.
- c. **Atribut utama:**
shipper_id, shipper_name

5. **dim_time**

- a. **Sumber OLTP:** orders.orderDate
- b. **Fungsi:** menyimpan kalender waktu agar analisis penjualan bisa dilakukan per hari, bulan, kuartal, atau tahun.
- c. **Atribut utama:**
time_id, full_date, day, month, month_name, quarter, year, weekday

Kesimpulan dimensi : tabel dimensi yang digunakan dalam proyek ini ada 5, yaitu: **dim_customer, dim_employee, dim_product, dim_shipper, dan dim_time.**

Kelima dimensi ini sudah cukup untuk mendukung analisis penjualan Northwind secara fleksibel dan sesuai kebutuhan KPI.

3.3 Deskripsi Atribut Fakta dan Dimensi

1. Tabel Fakta

Tabel fakta mencatat peristiwa transaksi penjualan dan metrik kuantitatif yang akan ianalisis.

TABEL	KOLOM	TIPE KUNCI	DESKRIPSI
fact_sales	fact_id	Primary Key	Kunci unik untuk setiap baris data pada tabel fakta.
	order_id	Foreign Key	Kunci order dari transaksi sumber (mewakili nomor pesanan).
	customer_id	Foreign Key	Terhubung ke dim_customer , menunjukkan pelanggan yang melakukan transaksi.
	employee_id	Foreign Key	Terhubung ke dim_employee , menunjukkan karyawan yang menangani order
	product_id		Terhubung ke dim_product , menunjukkan produk yang dijual.
	shipper_id	Foreign Key	Terhubung ke dim_shipper , menunjukkan jasa pengiriman yang digunakan
	time_id	Foreign Key	Terhubung ke dim_time , menunjukkan tanggal terjadinya transaksi.
	unit_price	Measure (Numerik)	Harga satuan produk pada saat transaksi berlangsung.
	quantity	Measure (Numerik)	Jumlah unit produk yang dibeli
	discount	Measure (Numerik)	Besaran diskon yang diberikan pada item transaksi.
	total_sales	Measure (Numerik)	Nilai penjualan bersih setelah diskon. Rumus: unit_price × quantity × (1-discount)

2. Tabel Dimensi

Tabel dimensi memberikan konteks deskriptif sebagai sudut pandang analisis (produk, pelanggan, karyawan, pengiriman, waktu).

A. dimensi Customer (dim_customer)

TABEL	KOLOM	TIPE KUNCI	DESKRIPSI
dim_customer	customer_id	Primary key	Kunci unik setiap pelanggan, diPrimary gunakan di fact_sales.
	company_name	Deskriptif	Nama perusahaan/customer yang membeli produk.
	contact_name	Deskriptif	Nama orang yang bisa dihubungi dari customer.
	contact_title	Deskriptif	Jabatan/kedudukan contact customer
	city	Deskriptif	Kota asal customer untuk analisis wilayah
	country	Deskriptif	Negara asal customer untuk analisis penjualan per negara

B. Dimensi Employee (dim_employee)

Tabel	Kolom	Tipe Kunci	Deskripsi
dim_employee	employee_id	Primary Key	Kunci unik karyawan, digunakan di fact_sales
	employee_name	Deskriptif	Nama lengkap karyawan yang melayani transaksi.
	title	Deskriptif	Jabatan karyawan (mis. Sales Representative)
	city	Deskriptif	Kota domisili karyawan.
	country	Deskriptif	Negara domisili karyawan

C. Dimensi Product (dim_product)

Tabel	Kolom	Tipe Kunci	Deskripsi
dim_product	product_id	Primary Key	Kunci unik produk, digunakan di fact_sales

D. Dimensi Shipper (dim_shipper)

Tabel	Kolom	Tipe Kunci	Deskripsi
dim_shipper	shipper_id	Primary Key	Kunci unik shipper, digunakan di fact_sales.

E. Dimensi Time (dim_time)

Tabel	Kolom	Tipe Kunci	Deskripsi
dim_date	time_id	Primary Key	Kunci unik tanggal (surrogate key), digunakan di fact_sales

3.4 Diagram Star Schema

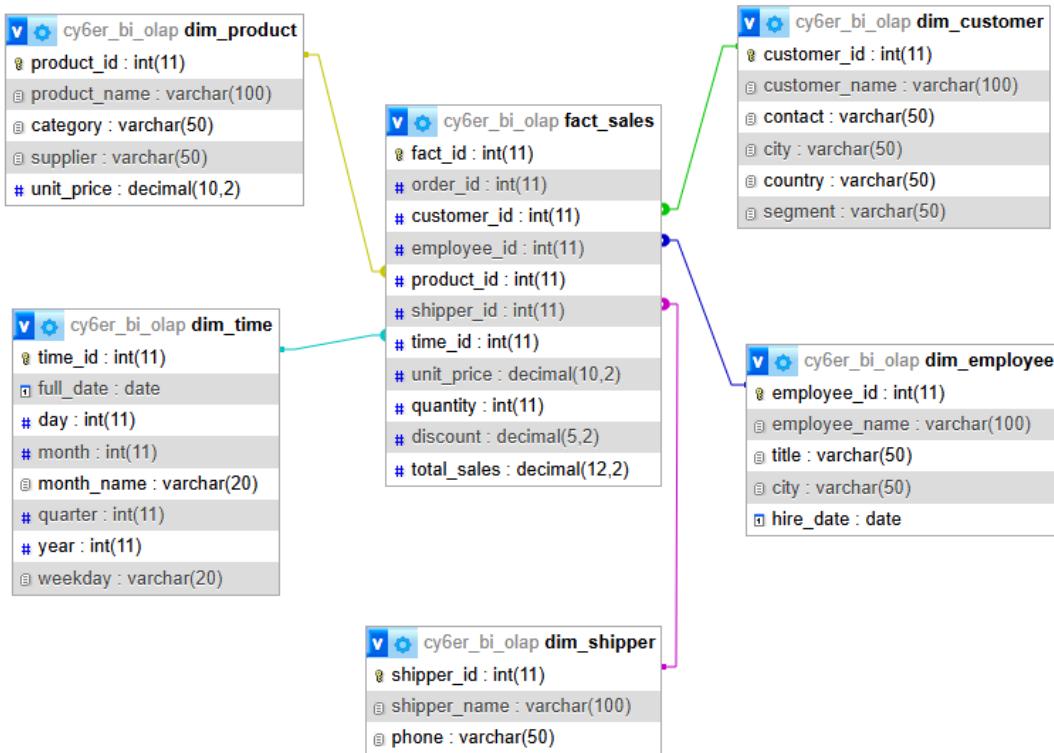


Diagram di atas menunjukkan rancangan **star schema** pada database OLAP cy6er_bi_olap. Struktur ini menempatkan tabel **fact_sales** sebagai pusat karena berisi data transaksi inti penjualan. Tabel fakta ini menyimpan metrik utama seperti `unit_price`, `quantity`, `discount`, dan `total_sales`.

Di sekelilingnya terdapat lima tabel dimensi yang terhubung langsung melalui foreign key, yaitu:

1. **dim_product** → memberi informasi detail produk dan kategorinya.
2. **dim_customer** → memberi konteks pelanggan dan lokasi pembeli.
3. **dim_employee** → memberi informasi karyawan yang menangani transaksi.
4. **dim_shipper** → memberi informasi jasa pengiriman yang digunakan.
5. **dim_time** → memberi konteks waktu transaksi (hari, bulan, kuartal, tahun).

Dengan model ini, analisis penjualan menjadi lebih mudah dan cepat karena setiap transaksi di `fact_sales` dapat dilihat dari berbagai sudut pandang dimensi tanpa perlu melakukan join kompleks seperti pada OLTP.

3.5 Hubungan Antar Tabel (Relational Mapping)

BAB 4

PERANCANGAN PROSES ETL

4.1 Gambaran Umum ETL

ETL (Extract, Transform, Load) adalah proses utama yang digunakan untuk memindahkan data dari OLTP ke OLAP. Proses ini memastikan data operasional yang bersifat transaksi dapat diubah menjadi data analitis yang terstruktur. ETL pada proyek ini dilakukan dari schema cy6er_bi (OLTP) menuju cy6er_bi.olap (OLAP).

4.2 Tahap Extract

4.2.1 Sumber Data

Sumber data berasal dari OLTP dengan database bernama cy6er_bi :

- orders
- order_details
- products
- categories
- customers
- employees
- shippers

4.2.2 Teknik dan Tools Extract

1. Dim_Product

- a. Tools yang digunakan :
 - Table Input
 - Select Values
 - Database Lookup
 - Filter Rows
 - Table Output

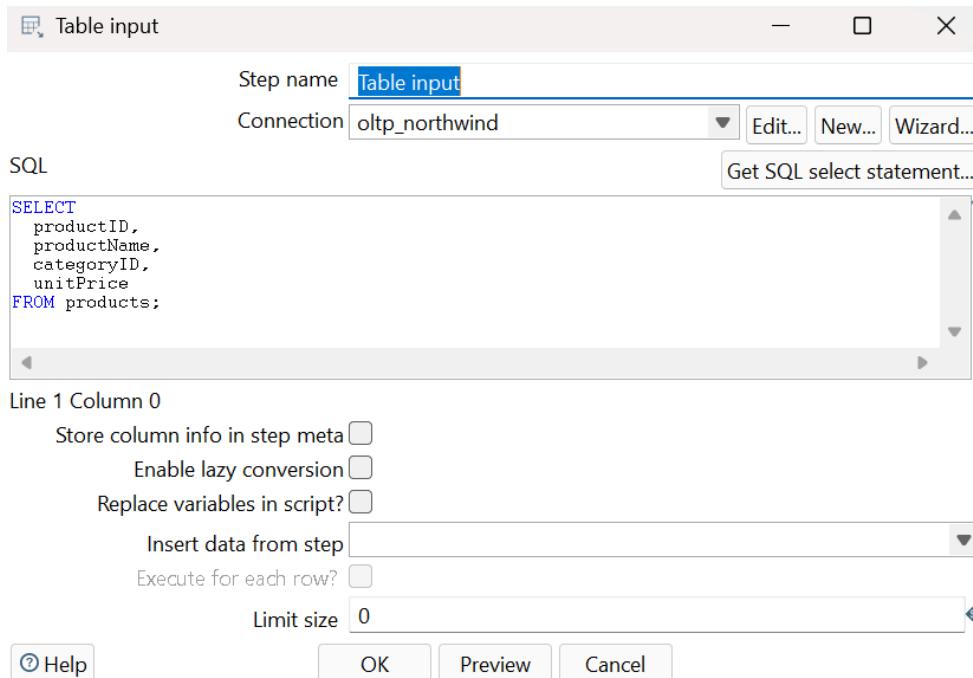


Penjelasan :

- Step Table Input mengambil data produk dari tabel OLTP cy6er_bi_products.
- Extract mencakup field productID, productName, categoryID, unitPrice.
- Data kategori dilengkapi dengan Database Lookup ke tabel cy6er_bi_categories.

- Setelah data valid, hasil extract dikirim untuk dimuat ke tabel OLAP dim_product.

b. Extract untuk Dim_Product :



Pada tahap extract digunakan step **Table Input** di Pentaho. Step ini terhubung ke database **oltp_northwind** lalu menjalankan query SQL:

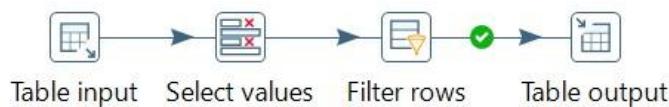
SELECT productID, productName, categoryID, unitPrice FROM products;

Artinya, sistem mengambil data dari tabel **products** dengan kolom yang diperlukan (ID produk, nama produk, kategori, dan harga) untuk diproses ke tahap berikutnya.

2. Dim_Customer

a. Tools yang digunakan :

- Table Input
- Select Values
- Filter Rows
- Table Output

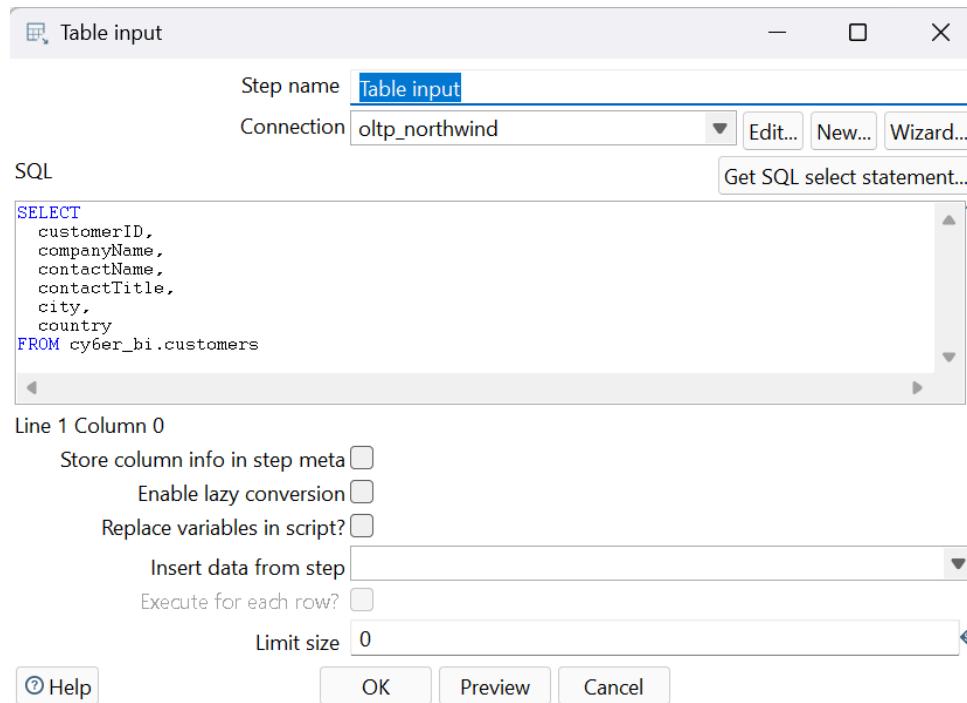


Penjelasan :

- Extract dilakukan melalui Table Input yang mengambil data dari OLTP cy6er_bi_customers.

- Field yang diambil antara lain customerID, companyName, contactName, contactTitle, city, country.
- Select Values menyesuaikan nama kolom agar sesuai skema OLAP.
- Filter Rows memastikan data customer_id tidak NULL sebelum dimuat.

b. Extract untuk Dim_Customer :



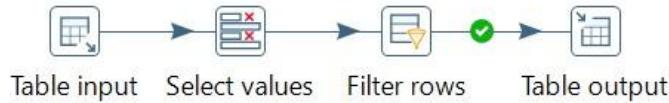
Pada tahap extract ini digunakan step **Table Input** di Pentaho dengan koneksi **oltp_northwind**. Step ini menjalankan query:

```
SELECT customerID, companyName, contactName, contactTitle, city, country
FROM cyber_bi.customers;
```

Artinya, sistem mengambil data pelanggan dari tabel **customers** (ID, nama perusahaan, nama & jabatan kontak, kota, negara) secara penuh karena **limit size = 0**.

3. Dim_Employee

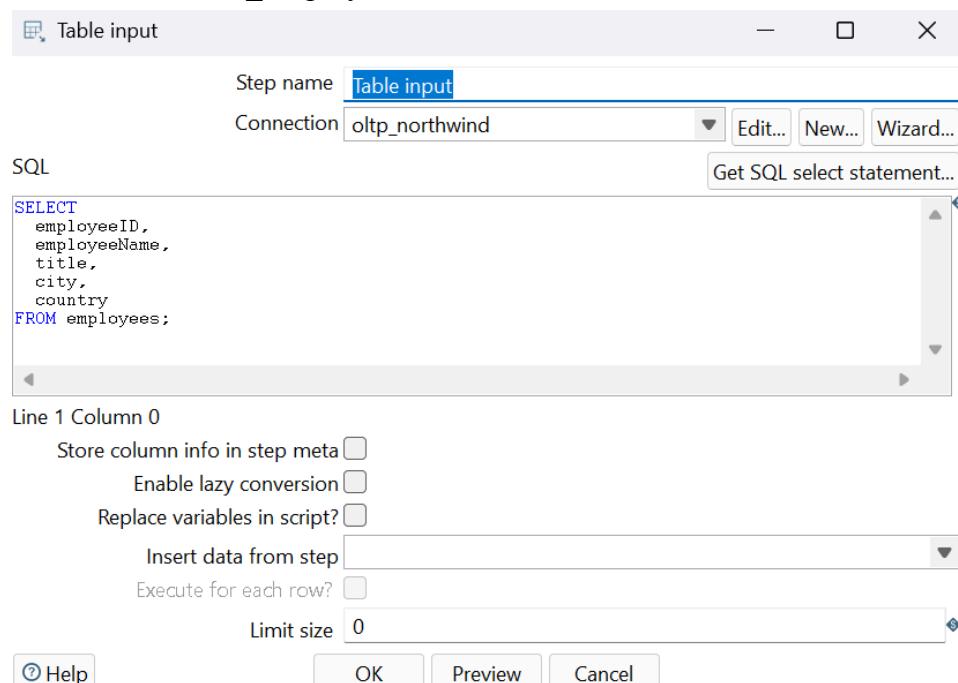
- Tools yang digunakan :
 - Table Input
 - Select Values
 - Filter Rows
 - Table Output



Penjelasan :

- Table Input mengekstrak data karyawan dari OLTP cy6er_bi_employees.
- Data yang diambil meliputi employeeID, employeeName, title, city, country.
- Select Values melakukan mapping nama field ke struktur dim_employee.
- Filter Rows membersihkan data yang tidak valid (ID/nama kosong).

b. Extract untuk Dim_Employee :



Pada tahap extract ini dipakai step **Table Input** di Pentaho dengan koneksi **oltp_northwind**. Query yang dijalankan:

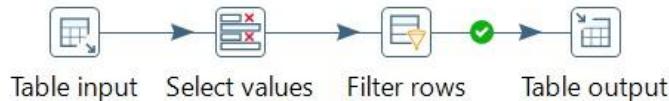
SELECT employeeID, employeeName, title, city, country FROM employees;

Artinya, sistem mengambil data karyawan dari tabel **employees** (ID, nama, jabatan, kota, negara) untuk digunakan di proses ETL selanjutnya, dengan **limit size = 0** sehingga semua data diambil.

4. Dim_Shipper

- Tools yang digunakan :
 - Table Input

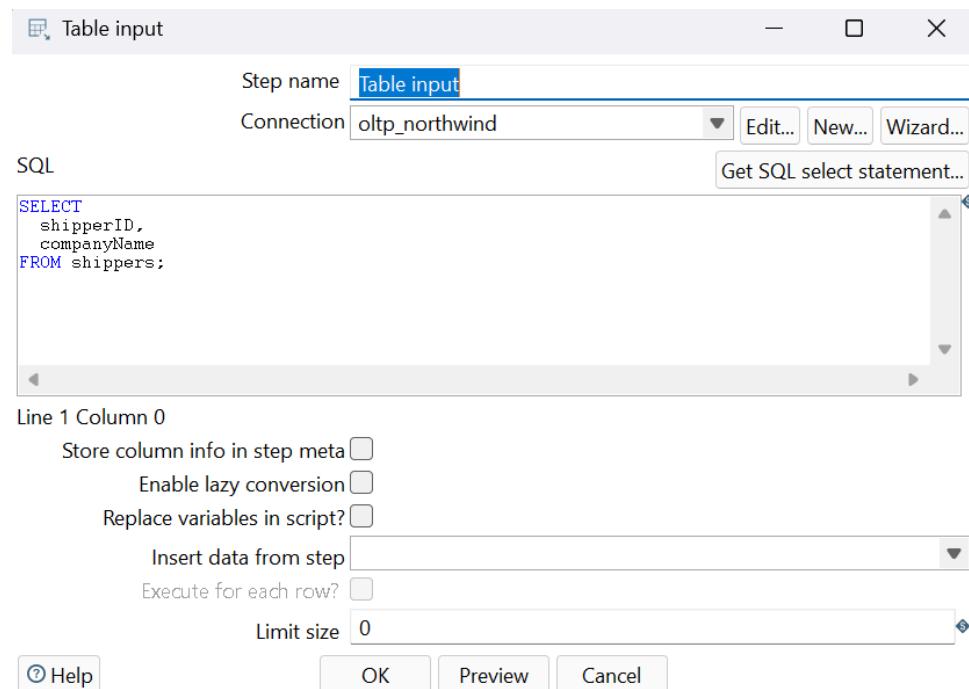
- Select Values
- Filter Rows
- Table Output



Penjelasan :

- Extract dilakukan dari tabel OLTP cy6er_bi_shippers menggunakan Table Input.
- Field yang diambil shipperID dan companyName.
- Select Values menyesuaikan nama field menjadi shipper_id dan shipper_name.
- Baris invalid dibuang lewat Filter Rows sebelum load.

b. Extract untuk Dim_Shipper :



Pada tahap extract ini digunakan step **Table Input** di Pentaho dengan koneksi **oltp_northwind**. Query:

SELECT shipperID, companyName FROM shippers;

Artinya, sistem mengambil data **pengirim/ekspedisi** dari tabel **shippers** (ID shipper dan nama perusahaan) untuk dipakai di proses ETL berikutnya, dengan **limit size = 0** sehingga semua data diambil.

5. Dim_Time :

a. Tools yang digunakan :

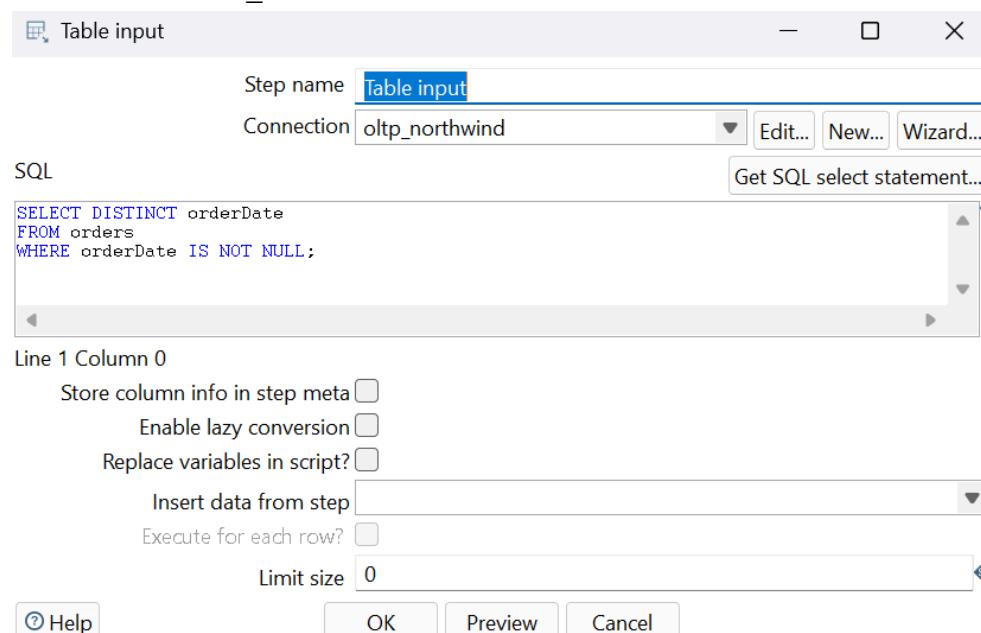
- Table Input
- Select Values
- Calculator
- Modified JavaScript Value
- Filter Rows
- Table Output



Penjelasan :

- Sumber tanggal diekstrak dari tabel cy6er_bi_orders (kolom orderDate).
- Table Input mengambil DISTINCT orderDate sebagai dasar dimensi waktu.
- Calculator menghasilkan atribut waktu (day, month, year, quarter, weekday_num).
- Modified JavaScript Value membentuk time_id format YYYYMMDD, month_name, dan weekday.
- Data hasil extract dimuat ke tabel OLAP dim_time.

b. Extract untuk Dim_Time :



Pada tahap extract ini digunakan step **Table Input** di Pentaho dengan koneksi **oltp_northwind**. Query:

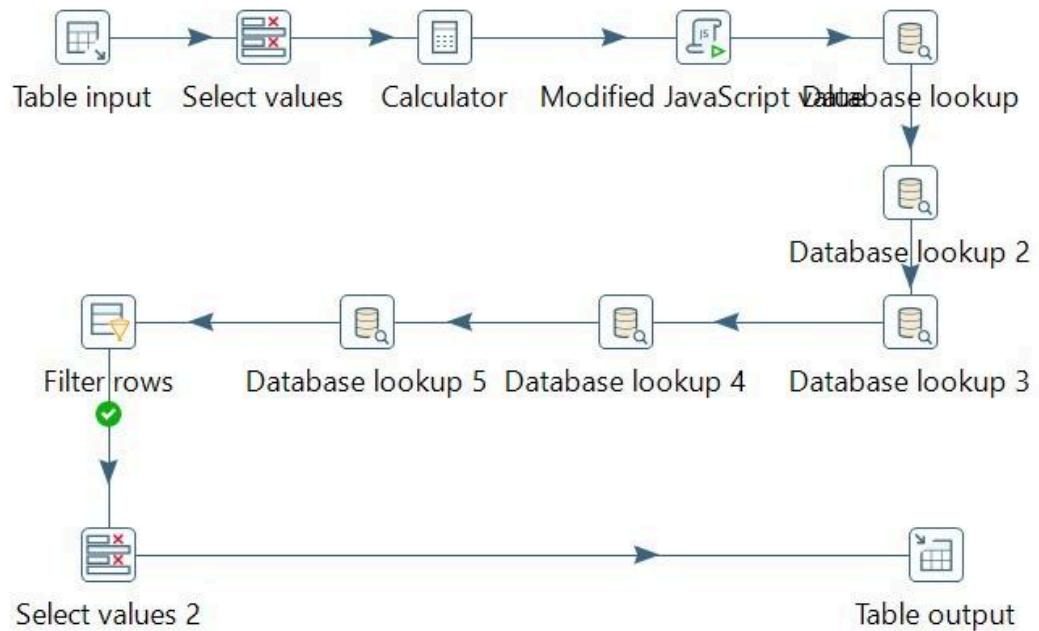
```
SELECT DISTINCT orderDate FROM orders WHERE orderDate IS NOT NULL;
```

Artinya, sistem mengambil **daftar tanggal order yang unik** dari tabel **orders**, hanya yang tidak kosong (NOT NULL), untuk kebutuhan proses ETL selanjutnya.

6. Fact_Sales

a. Tools yang digunakan :

- Table Input (JOIN OLTP)
- Select Values
- Calculator
- Modified JavaScript Value
- Database Lookup (validasi FK ke semua dimensi)
- Filter Rows
- Select Values 2
- Table Output

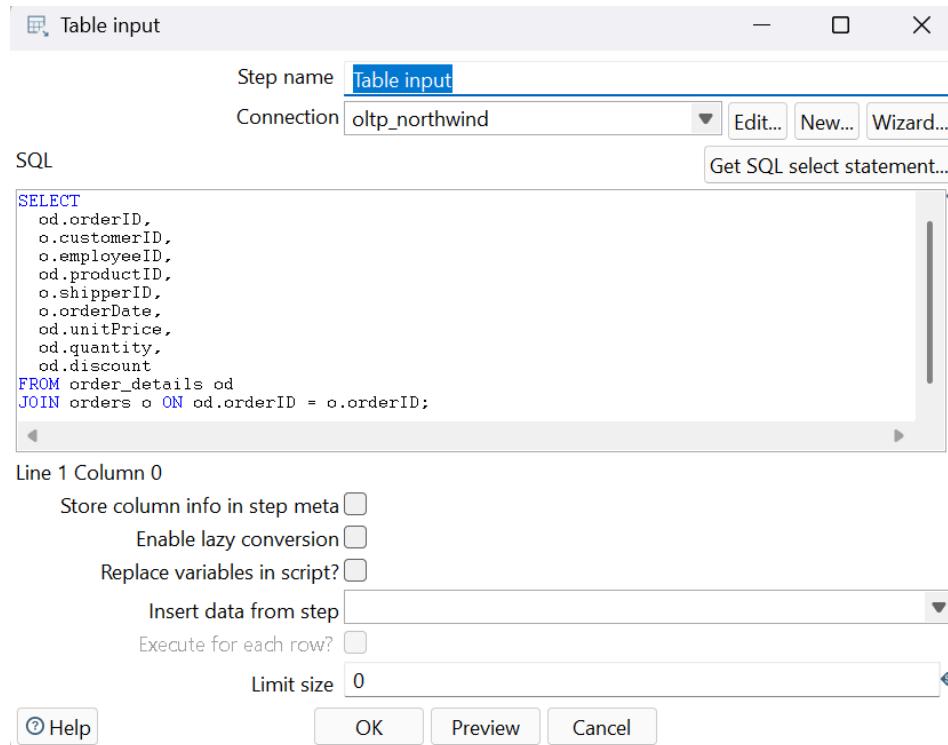


Penjelasan :

- Fact sales mengekstrak data transaksi dari hasil join OLTP orders dan order_details.
- Table Input mengambil field orderID, customerID, employeeID, productID, shipperID, orderDate, unitPrice, quantity, discount.
- Calculator menghitung measure penjualan (subtotal, discount_amount, total_sales).
- JavaScript membentuk time_id dari orderDate.

- Database Lookup memastikan semua foreign key ada di tabel dimensi OLAP.
- Data valid dimuat ke fact_sales.

b. Extract untuk Fact_Sales :



Pada tahap extract ini digunakan step **Table Input** di Pentaho dengan koneksi **oltp_northwind**. Query mengambil data transaksi dengan **join** antara tabel **order_details (od)** dan **orders (o)** berdasarkan **orderID**.

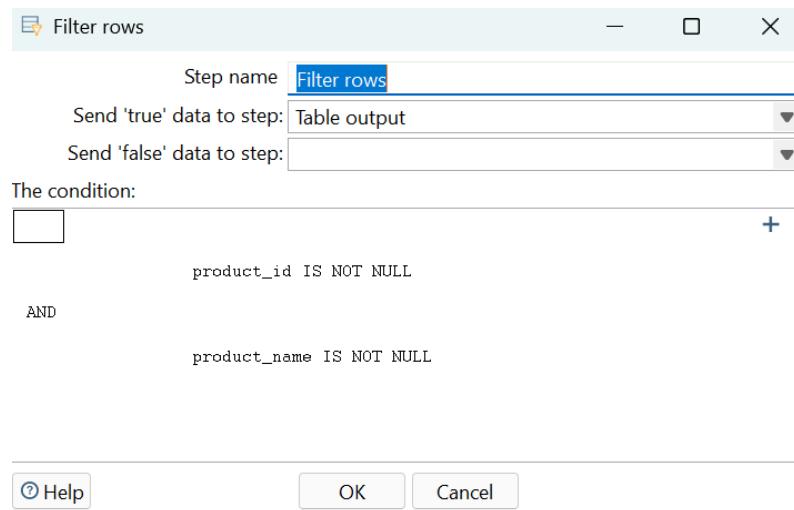
Data yang diekstrak mencakup informasi order dan detailnya seperti **customerID**, **employeeID**, **productID**, **shipperID**, **orderDate**, **unitPrice**, **quantity**, dan **discount**, untuk dipakai di proses ETL berikutnya.

4.3 Tahap Transform

4.3.1 Data Cleaning

Menjelaskan proses pembersihan data hasil extract sebelum dimuat ke data warehouse. Data cleaning dilakukan untuk menghilangkan data yang tidak valid, seperti nilai NULL pada atribut penting atau data yang tidak sesuai aturan bisnis, sehingga kualitas data tetap terjaga.

1. Data Cleaning untuk Dim_Product



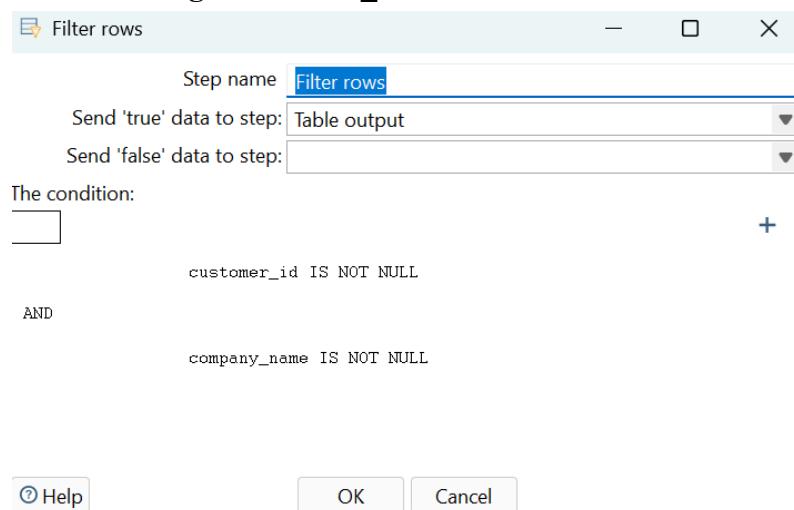
Pada tahap transformation (data cleaning) digunakan step **Filter Rows**. Step ini menyaring data agar hanya record yang valid yang diteruskan ke proses berikutnya (**Table Output**).

Kondisi filter yang dipakai adalah:

- `product_id IS NOT NULL`
- `product_name IS NOT NULL`

Artinya, data yang `product_id` dan `product_name`-nya tidak kosong akan disimpan/di-load, sedangkan data yang salah satu kolomnya kosong akan dibuang.

2. Data Cleaning untuk Dim_Customer



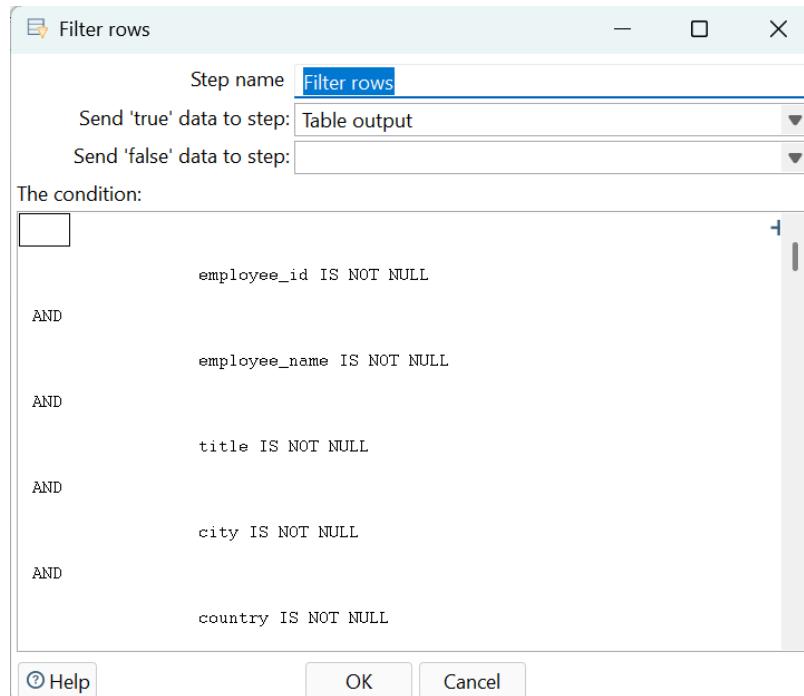
Pada tahap data cleaning ini digunakan step **Filter Rows** untuk memastikan data customer yang masuk sudah lengkap.

Filter yang dipakai:

- customer_id IS NOT NULL
- company_name IS NOT NULL

Artinya, hanya data pelanggan yang punya **customer_id** dan **company_name** (tidak kosong) yang diteruskan ke **Table Output**, sedangkan yang kosong dibuang.

3. Data Celaning untuk Dim_Employee



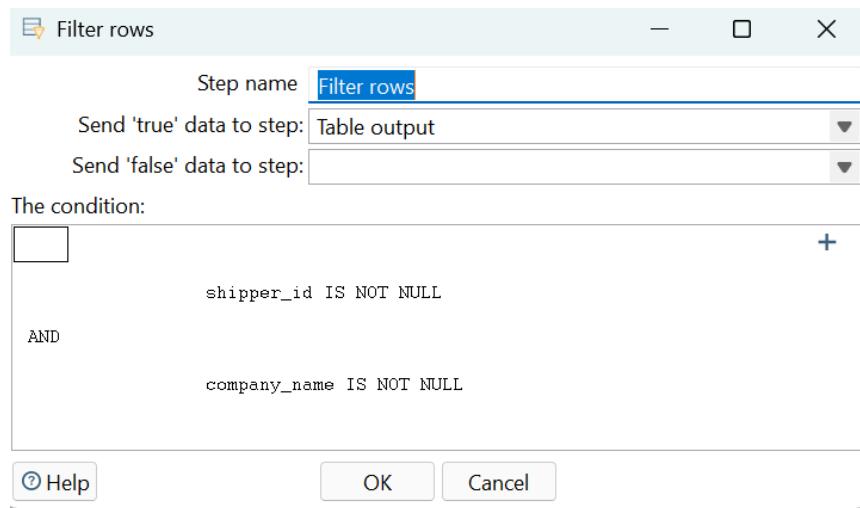
Pada tahap data cleaning ini digunakan step Filter Rows untuk menyaring data karyawan yang lengkap.

Kondisi filter:

- employee_id IS NOT NULL
- employee_name IS NOT NULL
- title IS NOT NULL
- city IS NOT NULL
- country IS NOT NULL

Artinya, hanya record karyawan yang semua kolom tersebut terisi (tidak null) yang diteruskan ke Table Output, sedangkan data yang ada nilai kosong akan dibuang.

4. Data Cleaning untuk Dim_Shipper



Pada tahap data cleaning ini digunakan step **Filter Rows** untuk memastikan data shipper valid.

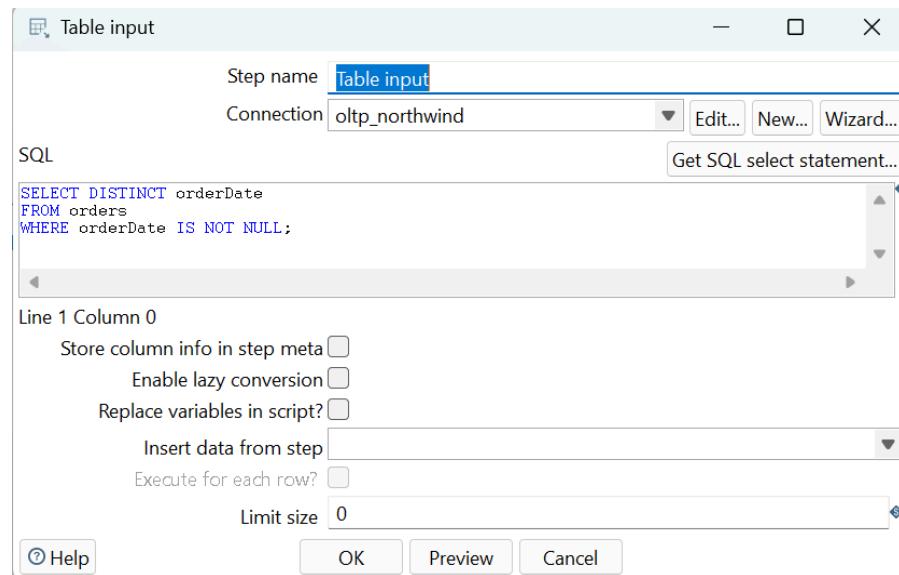
Kondisi filter:

- shipper_id IS NOT NULL
- company_name IS NOT NULL

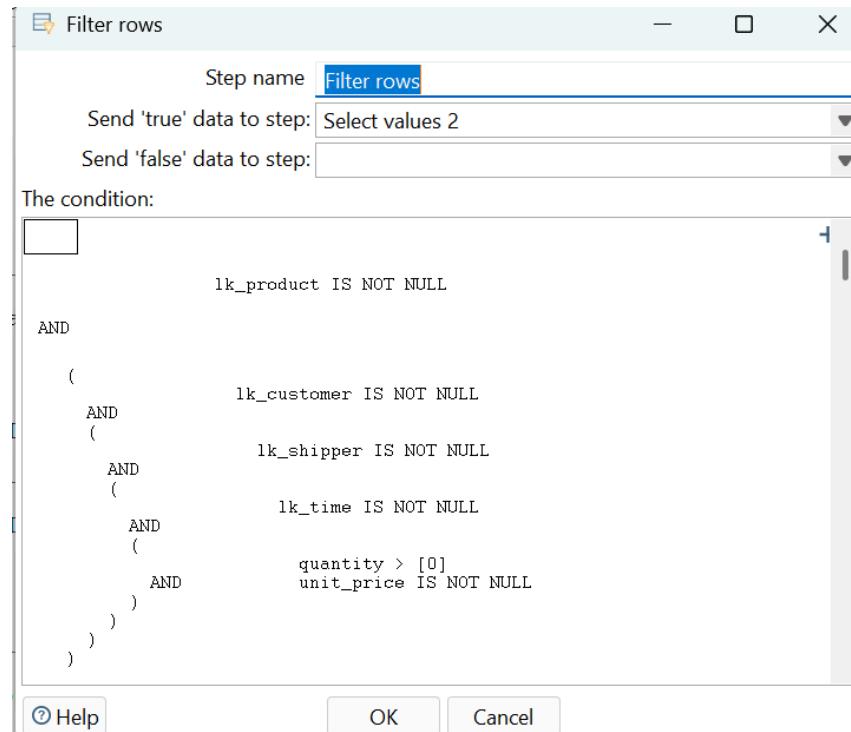
Artinya, hanya data shipper yang punya **ID shipper** dan **nama perusahaan** (tidak kosong) yang diteruskan ke **Table Output**, sedangkan yang kosong dibuang.

5. Data Cleaning untuk Dim_Time

Pada dim_time, proses data cleaning tidak dilakukan menggunakan Filter Rows, karena pembersihan sudah diterapkan sejak tahap extract. Hal ini terlihat pada query Table Input yang menggunakan WHERE orderDate IS NOT NULL untuk membuang tanggal kosong/NULL, serta SELECT DISTINCT orderDate untuk menghilangkan duplikasi tanggal. Dengan demikian, data tanggal yang masuk ke proses transformasi sudah bersih dan unik sehingga tidak memerlukan cleaning tambahan. Berikut adalah tampilan query table input dari dim_time :



6. Data Cleaning untuk Fact_Sales



Pada tahap data cleaning ini digunakan Filter Rows untuk menyaring data transaksi yang benar-benar valid sebelum lanjut ke step Select values 2.

Data hanya diteruskan jika memenuhi kondisi:

- lk_product IS NOT NULL
- lk_customer IS NOT NULL
- lk_shipper IS NOT NULL
- lk_time IS NOT NULL
- quantity > 0

- unit_price IS NOT NULL

Artinya, hanya transaksi yang punya semua key relasi lengkap, jumlah beli lebih dari 0, dan harga tidak kosong yang diproses дальше; sisanya dibuang.

4.3.2 Data Transformation

Proses transformasi data yang bertujuan menyesuaikan struktur data OLTP agar sesuai dengan skema bintang OLAP. Transformasi mencakup perubahan nama kolom, konversi tipe data, serta pembuatan atribut turunan yang diperlukan pada tabel dimensi maupun tabel fakta.

1. Data Transformation untuk Dim_Product

#	Fieldname	Rename to	Length	Precision	
1	productID	product_id			<input type="button" value="Get fields to select"/>
2	productName	product_name			<input type="button" value="Edit Mapping"/>
3	categoryID	category_id			
4	unitPrice	unit_price			

Include unspecified fields, ordered

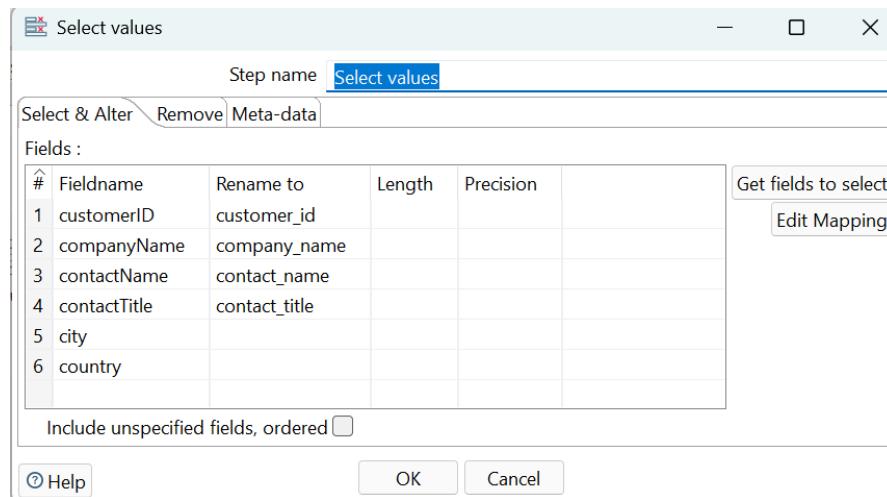
Pada tahap **data transformation** ini digunakan step **Select Values** untuk menyesuaikan struktur data dari OLTP ke OLAP. Step ini melakukan **perubahan nama kolom (rename)** agar sesuai dengan format skema bintang.

Perubahan yang dilakukan:

- productID → product_id
- productName → product_name
- categoryID → category_id
- unitPrice → unit_price

Hasilnya, atribut produk menjadi lebih konsisten dan siap dimasukkan ke tabel dimensi/fakta pada OLAP.

2. Data Transformation untuk Dim_Customer



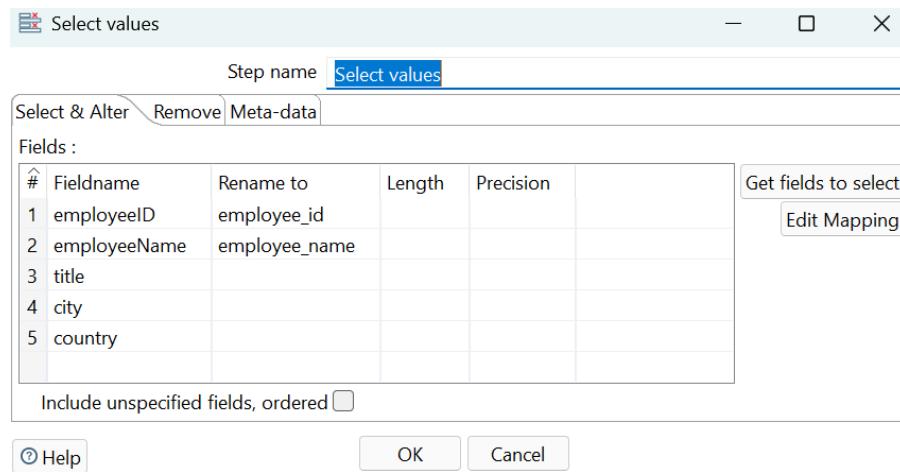
Pada tahap **data transformation** ini digunakan step **Select Values** untuk menyesuaikan nama kolom data customer dari OLTP agar sesuai skema OLAP.

Rename yang dilakukan:

- customerID → customer_id
- companyName → company_name
- contactName → contact_name
- contactTitle → contact_title

Kolom city dan country tetap digunakan tanpa perubahan. Hasil transformasi ini membuat atribut customer konsisten untuk tabel dimensi.

3. Data Transformation untuk Dim_Employee



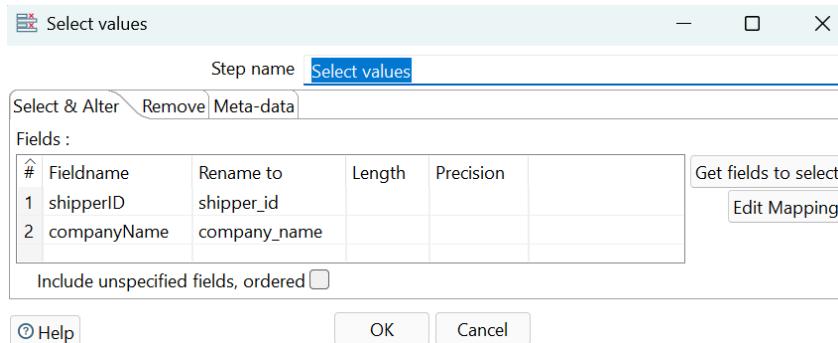
Pada tahap **data transformation** ini digunakan step **Select Values** untuk menyesuaikan nama kolom data karyawan dari OLTP ke format OLAP.

Rename yang dilakukan:

- employeeID → employee_id
- employeeName → employee_name

Kolom title, city, dan country tetap dipakai tanpa perubahan. Transformasi ini membuat struktur data employee konsisten untuk tabel dimensi.

4. Data Transformation untuk Dim_Shipper



Pada tahap **data transformation** ini digunakan step **Select Values** untuk menyesuaikan nama kolom data shipper agar sesuai skema OLAP.

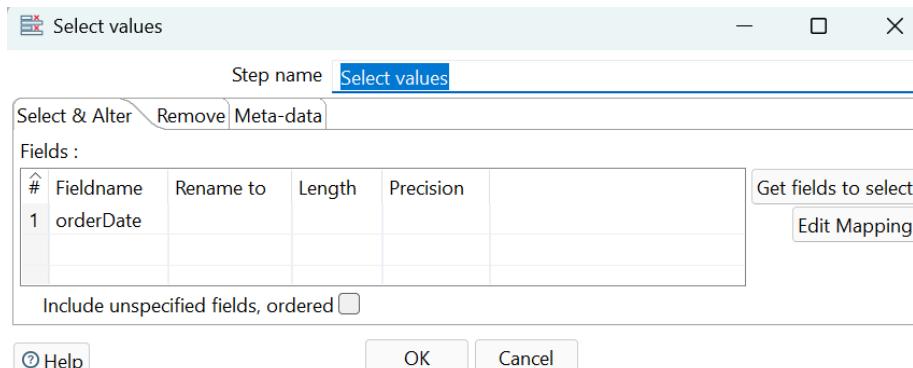
Rename yang dilakukan :

- shipperID → shipper_id
- companyName → company_name

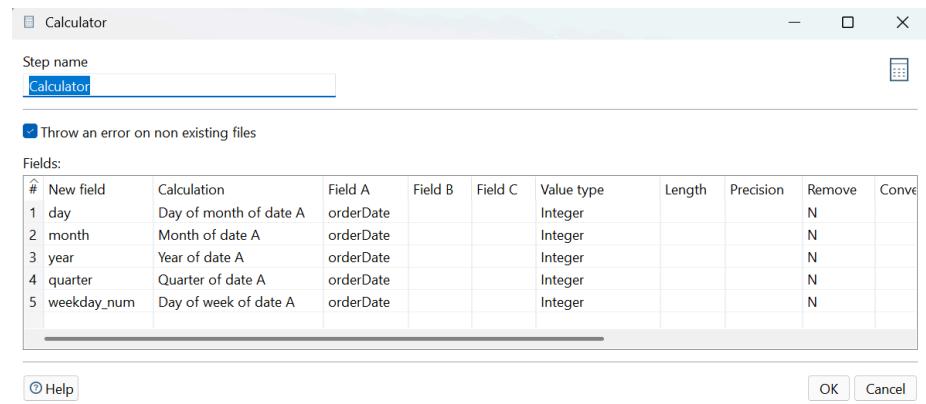
Hasilnya, atribut shipper menjadi konsisten dan siap dimasukkan ke tabel dimensi.

5. Data Transformation untuk Dim_Time

Pada dim_time, proses transformasi berbeda dari dimensi lain karena data yang diambil masih berupa tanggal mentah, sehingga perlu dibentuk menjadi atribut waktu turunan sebelum dimuat ke tabel dimensi waktu. Berikut langkah-langkahnya :



Mengubah nama kolom orderDate menjadi full_date sebagai tanggal utama di dimensi waktu.

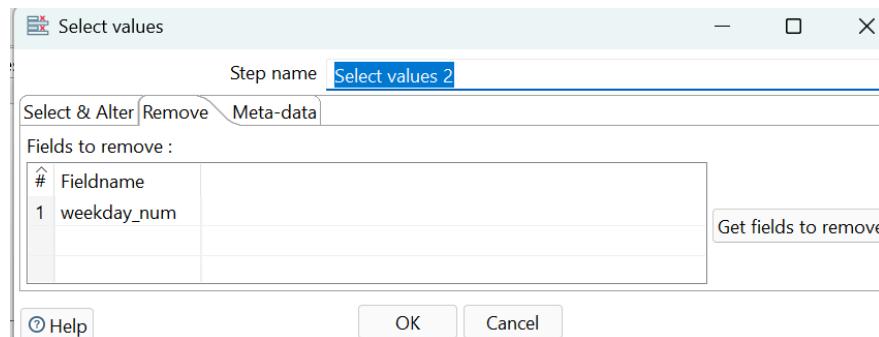


Membuat atribut turunan dari `full_date` seperti `day`, `month`, `year`, `quarter`, dan `weekday_num`.

The 'Fields' table below lists three output fields: 'time_id' (Integer), 'month_name' (String), and 'weekday' (String). Buttons for 'Help', 'OK', 'Cancel', 'Get variables', and 'Test script' are at the bottom."/>

#	Fieldname	Rename to	Type	Length	Precision	Replace value 'Fieldname' or 'Rename to'
1	time_id		Integer			N
2	month_name		String			N
3	weekday		String			N

Membentuk `time_id` dengan format `YYYYMMDD`, serta menambahkan `month_name` dan `weekday` (nama bulan & hari).

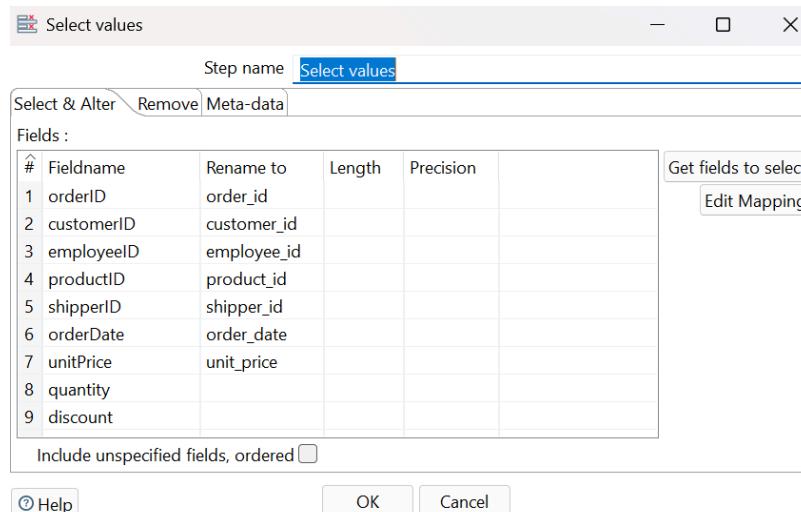


Menghapus field yang tidak dipakai (misalnya weekday_num) agar struktur dim_time rapi dan sesuai skema bintang.

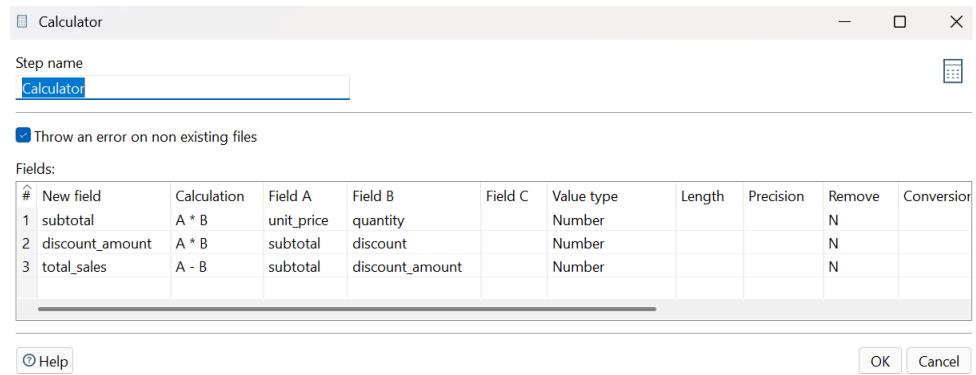
Maka, Setelah tanggal unik diperoleh, proses transformasi dilakukan untuk membentuk atribut dimensi waktu. Calculator menghasilkan atribut numerik (day, month, year, quarter, weekday_num), sedangkan Modified JavaScript Value membentuk time_id format YYYYMMDD serta nama bulan dan nama hari. Tahap akhir menggunakan Select Values 2 untuk menyisakan field yang sesuai dengan struktur dim_time sebelum diload ke OLAP.

6. Data Transformation untuk Fact_Sales

Data transformation pada tabel fact_sales dilakukan secara bertahap untuk menyesuaikan data transaksi OLTP agar sesuai dengan struktur tabel fakta pada skema bintang OLAP. Proses transformasi ini melibatkan dua tahap Select Values, perhitungan measure, pembentukan time_id, serta finalisasi field sebelum load. Berikut langkah-langkahnya :

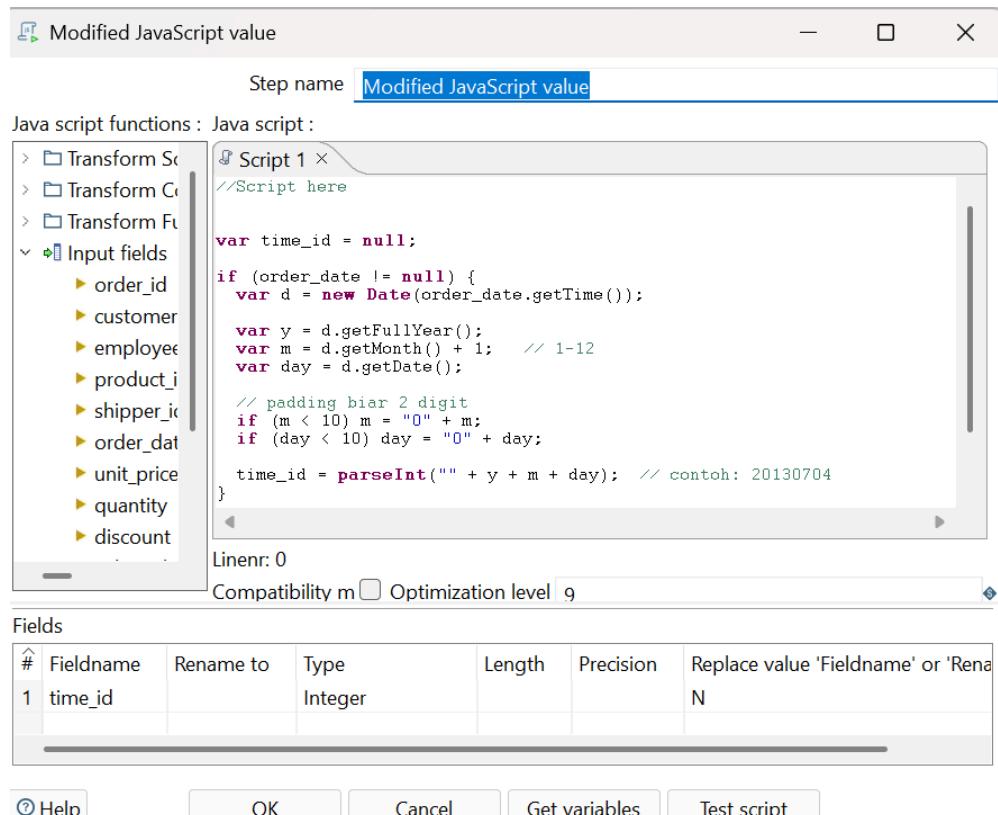


Select Values pertama digunakan untuk memilih dan mengganti nama kolom hasil extract dari OLTP agar sesuai dengan atribut fact_sales.



Setelah field utama terbentuk, Calculator digunakan untuk menghitung measure penjualan:

- subtotal = unit_price * quantity
- discount_amount = subtotal * discount
- total_sales = subtotal - discount_amount



Step JavaScript digunakan untuk mengubah order_date menjadi time_id dengan format YYYYMMDD, sehingga dapat direlasikan dengan dim_time.

Database lookup

Step name **Database lookup**

Connection **dw_northwind.olap**

Lookup schema **cy6er.bi.olap**

Lookup table **dim_product**

Enable cache?

Cache size in rows (0=cache) **0**

Load all data from table

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

#	Table field	Comparator	Field1	Field2
1	product_id	=	product_id	

Values to return from the lookup table :

#	Field	New name	Default	Type
1	product_id	lk_product		None

Do not pass the row if the lookup fails

Fail on multiple results?

Order by

Database lookup

Step name **Database lookup 2**

Connection **dw_northwind.olap**

Lookup schema **cy6er.bi.olap**

Lookup table **dim_customer**

Enable cache?

Cache size in rows (0=cache) **0**

Load all data from table

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

#	Table field	Comparator	Field1	Field2
1	customer_id	=	customer_id	

Values to return from the lookup table :

#	Field	New name	Default	Type
1	customer_id	lk_customer		None

Do not pass the row if the lookup fails

Fail on multiple results?

Order by

Database lookup

Step name: Database lookup 3

Connection: dw_northwind_olap

Lookup schema: cy6er_bi_olap

Lookup table: dim_employee

Enable cache?

Cache size in rows (0=cache) 0

Load all data from table

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

#	Table field	Comparator	Field1	Field2
1	employee_id	=	employee_id	

Values to return from the lookup table :

#	Field	New name	Default	Type
1	employee_id	lk_employee		None

Do not pass the row if the lookup fails

Fail on multiple results?

Order by _____

Database lookup

Step name: Database lookup 4

Connection: dw_northwind_olap

Lookup schema: cy6er_bi_olap

Lookup table: dim_shipper

Enable cache?

Cache size in rows (0=cache) 0

Load all data from table

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

#	Table field	Comparator	Field1	Field2
1	shipper_id	=	shipper_id	

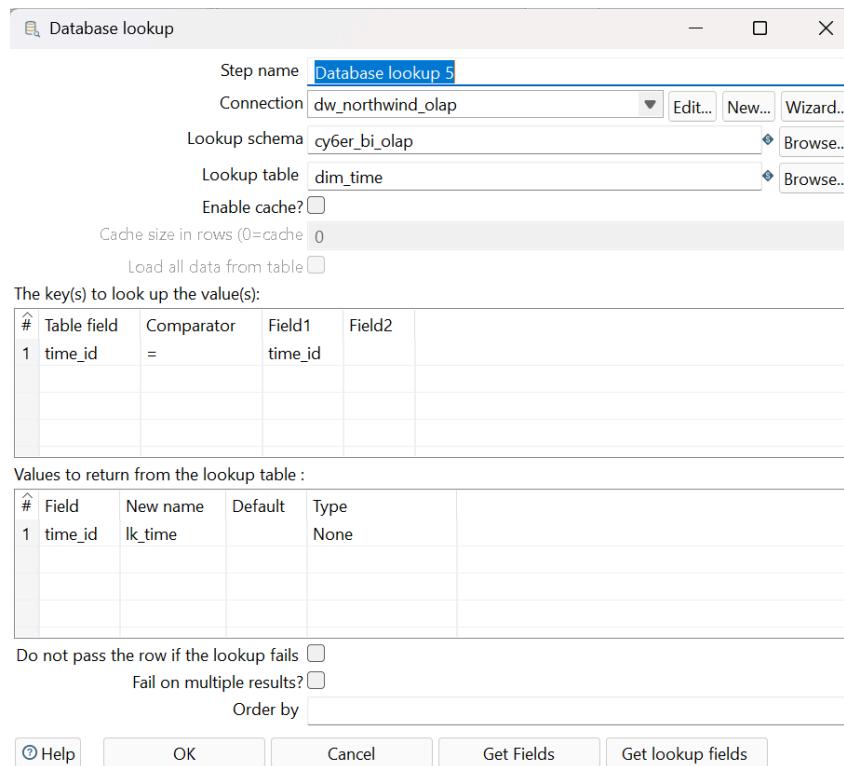
Values to return from the lookup table :

#	Field	New name	Default	Type
1	shipper_id	lk_shipper		None

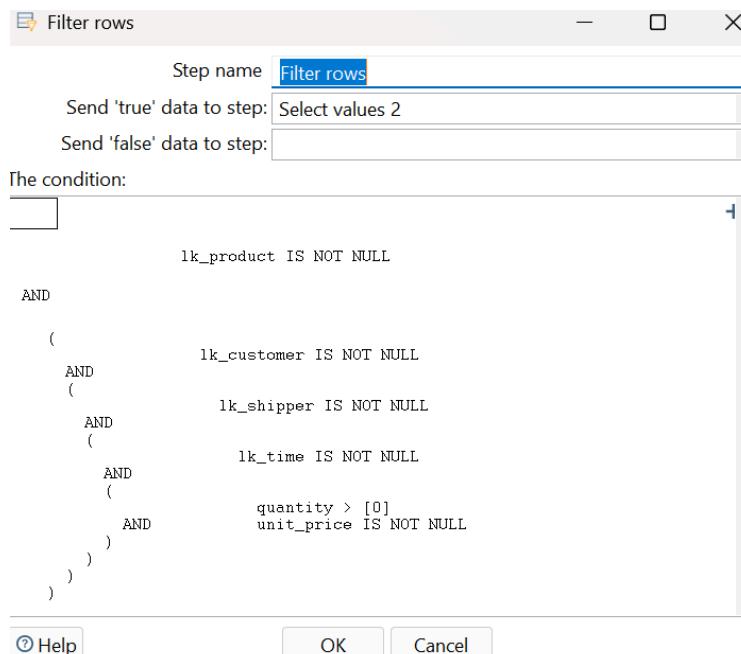
Do not pass the row if the lookup fails

Fail on multiple results?

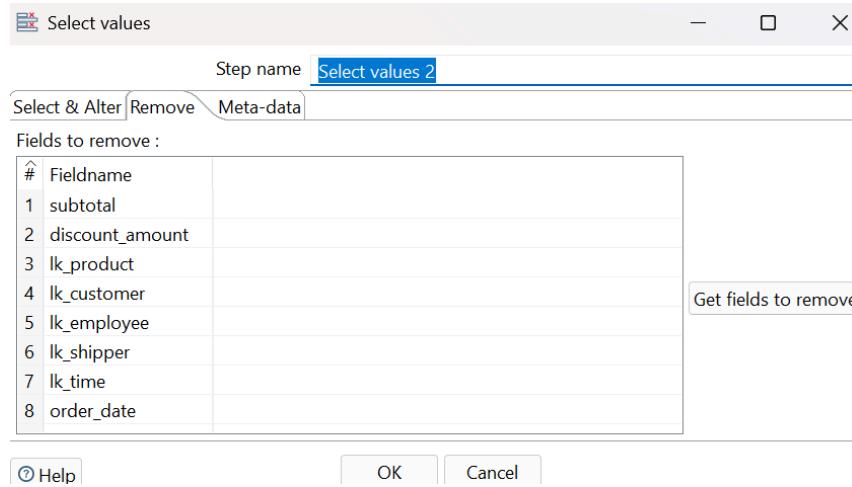
Order by _____



Database Lookup dilakukan berurutan untuk memastikan setiap foreign key pada fact_sales memiliki pasangan di tabel dimensi (dim_product, dim_customer, dim_employee, dim_shipper, dim_time). Jika lookup gagal, baris dianggap tidak valid.



Filter Rows digunakan untuk membuang transaksi yang tidak valid.



Select Values kedua digunakan untuk menghapus field sementara yang tidak perlu disimpan di OLAP.

4.3.3 Data Aggregation

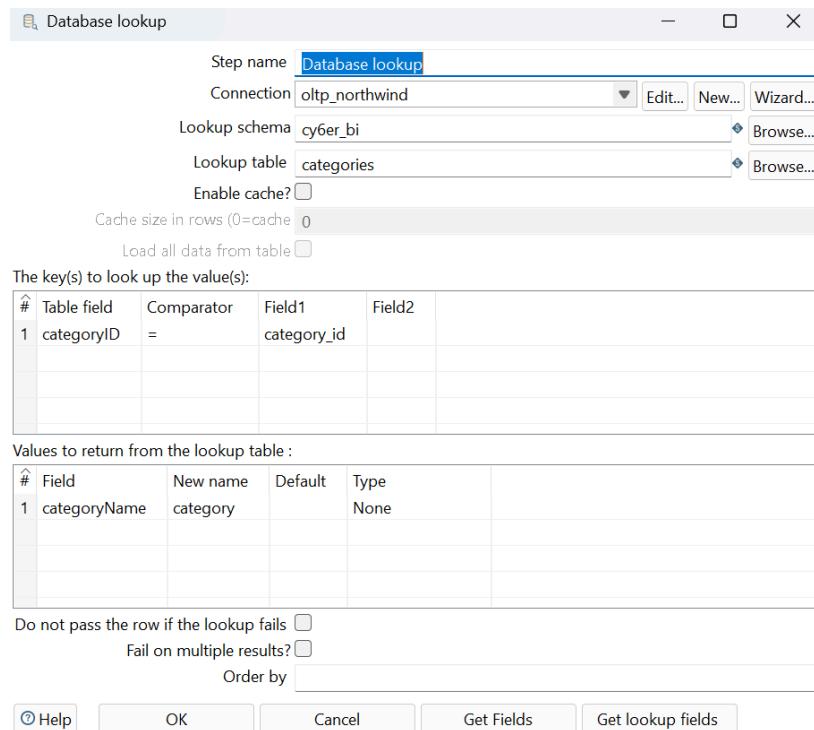
Pada implementasi ETL ini, **agregasi tidak dilakukan pada pipeline Pentaho**. Tabel fact_sales disimpan dalam bentuk **data transaksi detail (granular)**, dimana setiap baris merepresentasikan satu item produk pada satu order. Penyimpanan data secara detail ini bertujuan agar analisis OLAP lebih fleksibel, karena agregasi dapat dilakukan sesuai kebutuhan pengguna pada tahap query atau dashboard. Agregasi seperti **total sales per bulan, top 5 produk terlaris, atau total sales per customer** dilakukan pada tahap KPI menggunakan query SQL atau visualisasi dashboard.

4.3.4 Data Integration

Menjelaskan proses integrasi data, yaitu penggabungan data dari beberapa tabel sumber agar menghasilkan informasi yang lengkap dan konsisten sesuai kebutuhan data warehouse. Integrasi dilakukan melalui join dan lookup untuk menambahkan atribut pendukung serta memastikan keterkaitan antar dimensi dan fakta valid.

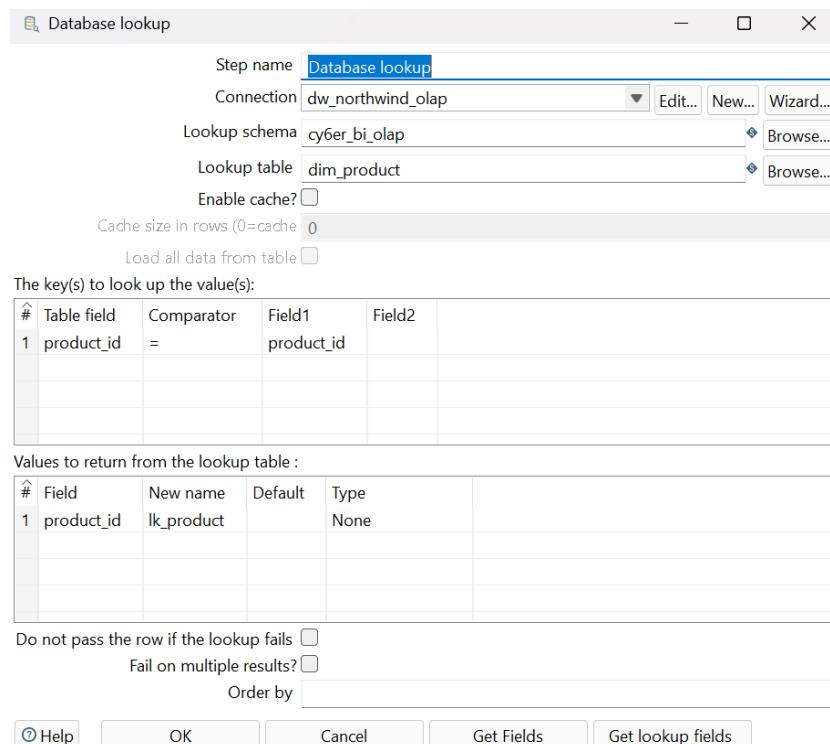
Pada proyek ini, integrasi data terjadi pada dua bagian utama, yaitu :

1. **Integrasi pada Dim_Product**



Dimensi produk membutuhkan informasi kategori agar setiap produk memiliki atribut kategori yang jelas. Oleh karena itu, setelah data produk diekstrak dari tabel products, dilakukan integrasi dengan tabel categories menggunakan step **Database Lookup**. Lookup ini mengambil categoryName berdasarkan categoryID sehingga output dim_product berisi informasi produk dan kategorinya secara lengkap.

2. Integrasi pada Fact_Sales



Database lookup

Step name: Database lookup 2

Connection: dw_northwind.olap

Lookup schema: cy6er.bi.olap

Lookup table: dim_customer

Enable cache?

Cache size in rows (0=cache) 0

Load all data from table

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

#	Table field	Comparator	Field1	Field2
1	customer_id	=	customer_id	

Values to return from the lookup table :

#	Field	New name	Default	Type
1	customer_id	lk_customer		None

Do not pass the row if the lookup fails

Fail on multiple results?

Order by

Database lookup

Step name: Database lookup 3

Connection: dw_northwind.olap

Lookup schema: cy6er.bi.olap

Lookup table: dim_employee

Enable cache?

Cache size in rows (0=cache) 0

Load all data from table

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

#	Table field	Comparator	Field1	Field2
1	employee_id	=	employee_id	

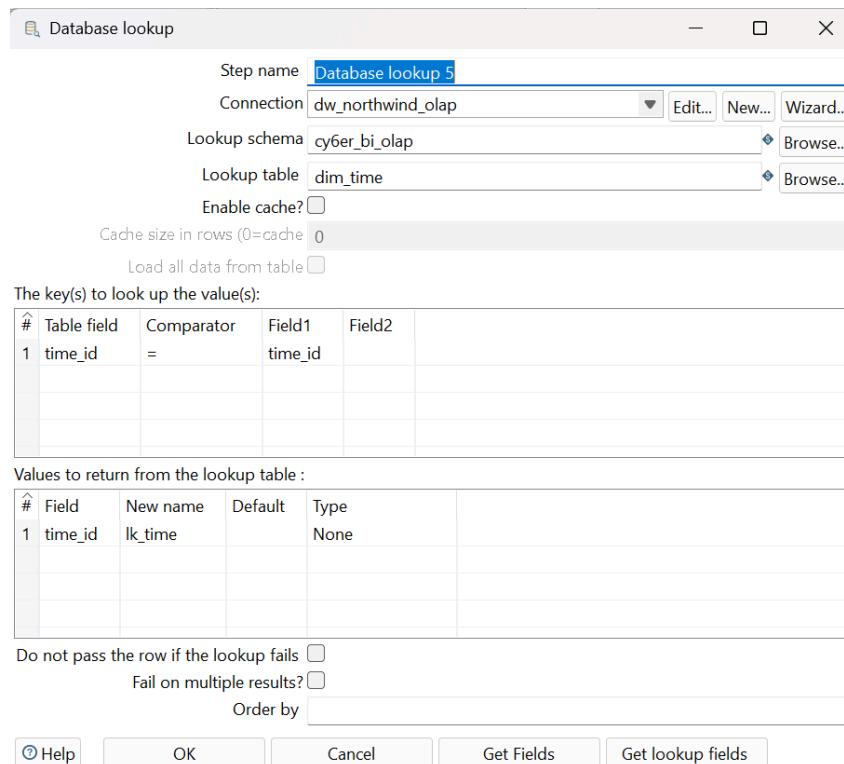
Values to return from the lookup table :

#	Field	New name	Default	Type
1	employee_id	lk_employee		None

Do not pass the row if the lookup fails

Fail on multiple results?

Order by



Data transaksi penjualan dibentuk dari hasil penggabungan tabel OLTP orders dan order_details. Penggabungan ini dilakukan di step **Table Input (JOIN)**, sehingga menghasilkan data transaksi lengkap per produk per order.

Setelah transaksi terbentuk, dilakukan integrasi dengan tabel dimensi OLAP melalui **Database Lookup berurutan** untuk memvalidasi foreign key:

- lookup ke **dim_product** untuk product_id
- lookup ke **dim_customer** untuk customer_id
- lookup ke **dim_employee** untuk employee_id
- lookup ke **dim_shipper** untuk shipper_id
- lookup ke **dim_time** untuk time_id

Setiap lookup hanya berfungsi memastikan nilai key tersebut tersedia di dimensi terkait. Jika salah satu lookup tidak menemukan pasangan dimensi, maka nilai lk_* menjadi NULL dan baris transaksi tersebut dibuang pada Filter Rows. Dengan cara ini, hanya transaksi yang memiliki keterkaitan dimensi lengkap yang dimuat ke tabel fact_sales.

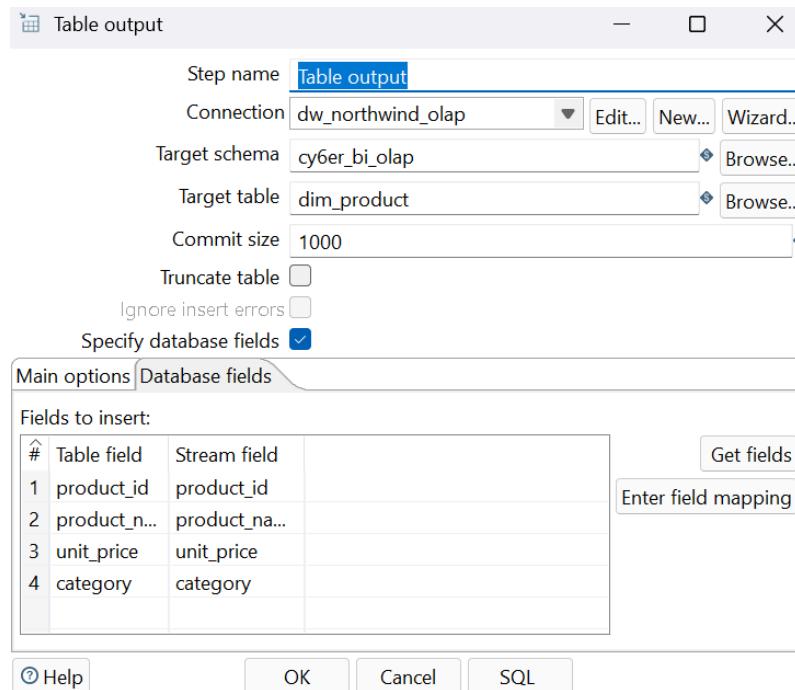
4.4 Tahap Load

4.4.1 Target Data Warehouse

- Target = database OLAP (cyber_bi.olap)/ data warehouse.
- Tabel target :
 - dim_product
 - dim_customer
 - dim_employee
 - dim_shipper
 - dim_time
 - fact_sales
- Load dilakukan dengan **Table Output** di Pentaho.
- Connection yang dipakai: **dw_northwind.olap**

Tahap ini dilakukan dengan memasukkan data hasil transformasi ke database OLAP sebagai data warehouse target. Proses load menggunakan step **Table Output** pada setiap transformation, dengan koneksi menuju database OLAP. Tabel dimensi dimuat terlebih dahulu (dim_product, dim_customer, dim_employee, dim_shipper, dim_time), kemudian dilanjutkan dengan pemuatan tabel fakta fact_sales.

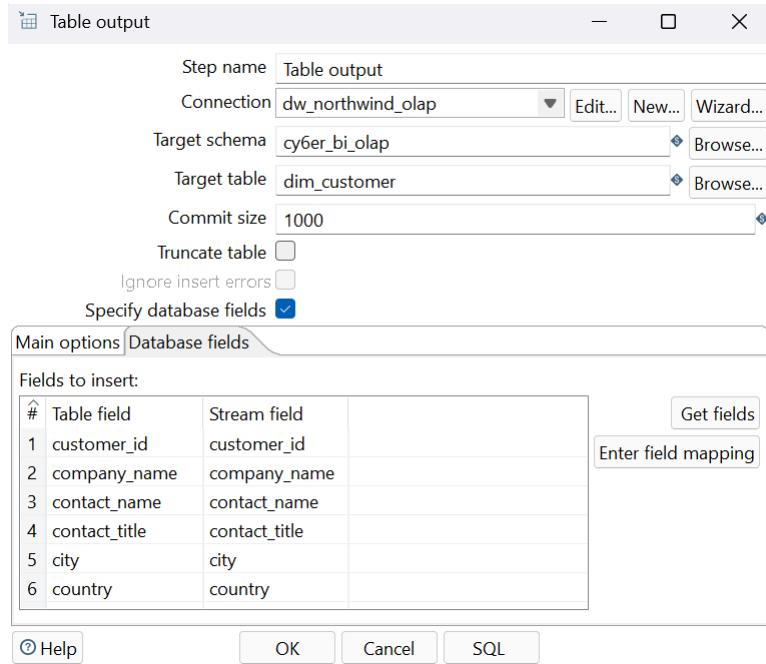
1. Target Data Warehouse untuk Dim_Product



Pada tahap **target data warehouse (load)** ini digunakan step **Table Output**. Step ini mengirim data hasil transformasi ke database OLAP melalui koneksi **dw_northwind.olap**, dengan target schema **cyber.bi.olap** dan tabel tujuan **dim_product**.

Field yang dimasukkan sudah dipetakan sesuai kolom di tabel dimensi (product_id, product_name, unit_price, category). Commit size 1000 berarti data disimpan per 1000 baris agar proses load lebih stabil.

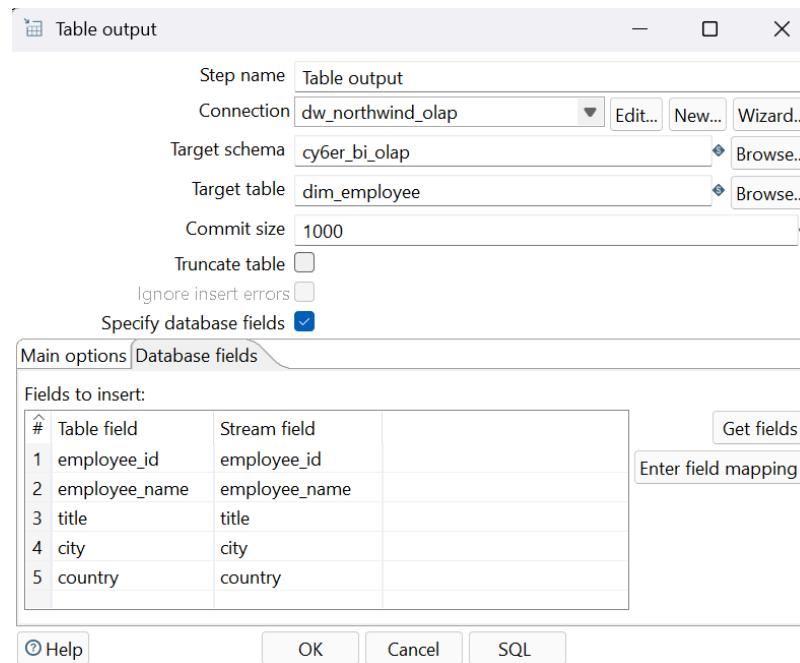
2. Target Data Warehouse untuk Dim_Customer



Pada tahap **load ke data warehouse** ini digunakan step **Table Output** untuk memasukkan data hasil transformasi ke tabel **dim_customer** pada database OLAP.

Koneksi yang dipakai **dw_northwind.olap** dengan schema **cyber_bi.olap**. Kolom yang dimuat sudah dipetakan sesuai tabel dimensi, yaitu **customer_id**, **company_name**, **contact_name**, **contact_title**, **city**, dan **country**. Commit size 1000 berarti penyimpanan dilakukan per 1000 baris agar proses load lebih aman.

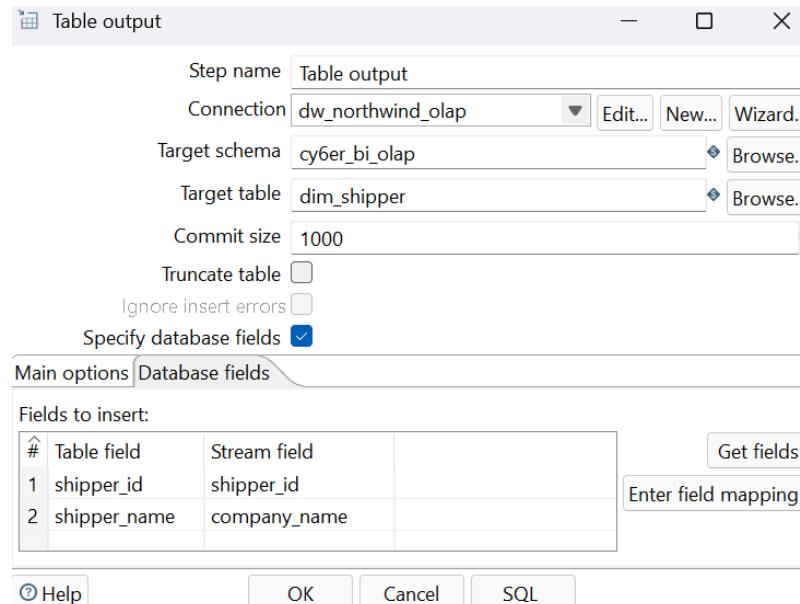
3. Target Data Warehouse untuk Dim_Employee



Pada tahap **load ke data warehouse** ini digunakan step **Table Output** untuk menyimpan data hasil transformasi ke tabel **dim_employee** di database OLAP.

Step terhubung ke **dw_northwind.olap** dengan schema **cyber_bi.olap**. Field yang dimuat sesuai mapping tabel dimensi yaitu **employee_id**, **employee_name**, **title**, **city**, dan **country**. Commit size 1000 berarti data di-insert per 1000 baris agar proses load stabil.

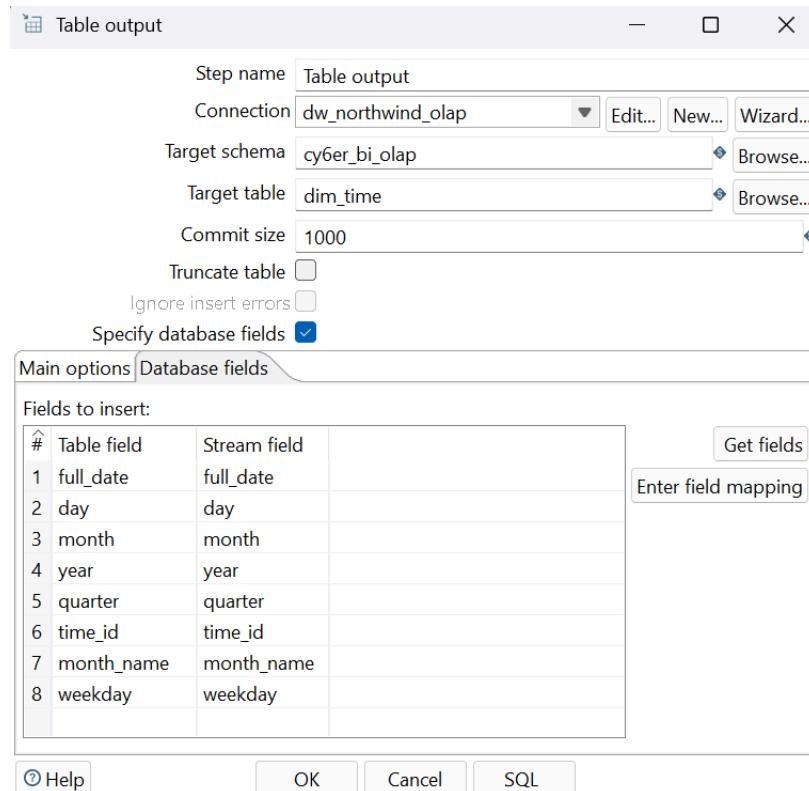
4. Target Data Warehouse untuk Dim_Shipper



Pada tahap **load ke data warehouse** ini digunakan step **Table Output** untuk memasukkan data hasil transformasi ke tabel **dim_shipper** pada database OLAP.

Step memakai koneksi **dw_northwind.olap** dengan schema **cyber.bi.olap**. Field yang dimuat adalah **shipper_id** dan **shipper_name**, dimana **company_name** dari stream dipetakan menjadi **shipper_name**. Commit size 1000 berarti insert dilakukan bertahap per 1000 baris.

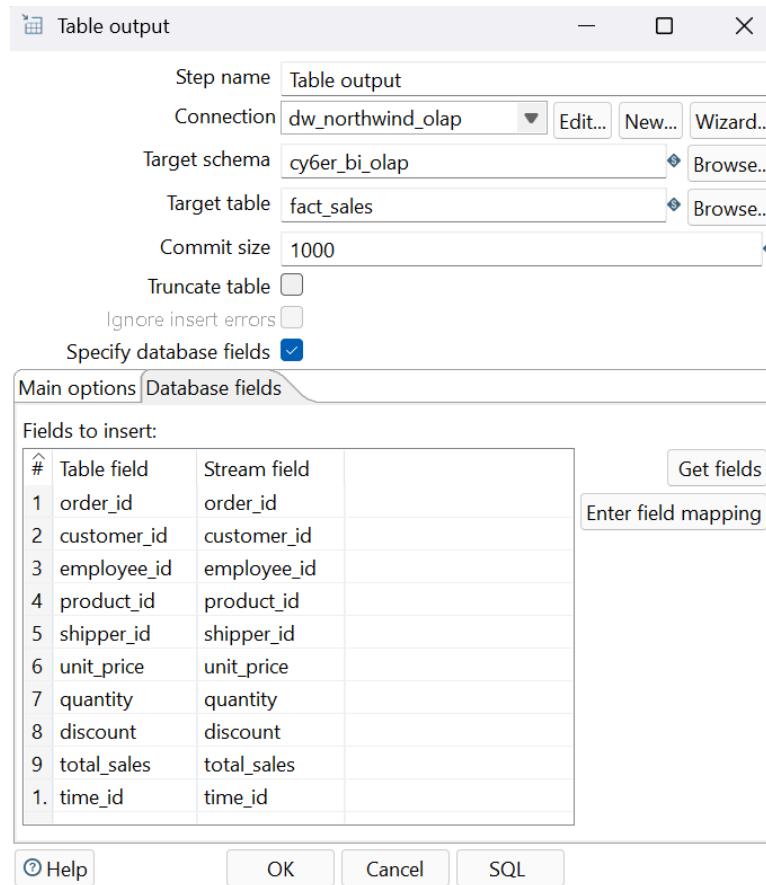
5. Target Data Warehouse untuk Dim_Time



Pada tahap **load ke data warehouse** ini digunakan step **Table Output** untuk menyimpan hasil transformasi dimensi waktu ke tabel **dim_time** pada database OLAP.

Koneksi yang dipakai **dw_northwind.olap** dengan schema **cyber.bi.olap**. Kolom yang dimuat meliputi **full_date**, **day**, **month**, **year**, **quarter**, **time_id**, **month_name**, dan **weekday** sesuai mapping dari stream. Commit size 1000 berarti data dimasukkan per 1000 baris agar proses load stabil.

6. Target Data Warehouse untuk Fact_Sales



Pada tahap **load ke data warehouse** ini digunakan step **Table Output** untuk memasukkan data hasil transformasi ke tabel fakta **fact_sales** pada database OLAP.

Step terhubung ke **dw_northwind.olap** dengan schema **cyber.bi.olap**. Kolom yang dimuat mencakup key dimensi (**order_id**, **customer_id**, **employee_id**, **product_id**, **shipper_id**, **time_id**) serta nilai transaksi (**unit_price**, **quantity**, **discount**, **total_sales**). Commit size 1000 berarti insert dilakukan per 1000 baris agar proses load tetap stabil.

4.4.2 Mapping Fakta dan Dimensi

Pemetaan atribut dari tabel sumber OLTP (cy6er.bi) menuju tabel dimensi dan fakta pada OLAP (cy6er.bi.olap). Mapping dilakukan berdasarkan hasil transformasi Pentaho sehingga struktur data sesuai dengan star schema.

A. Mapping Tabel Dimensi

1. Dim_Product

Sumber OLTP: products + categories

Fields to insert:

#	Table field	Stream field
1	product_id	product_id
2	product_name	product_name
3	unit_price	unit_price
4	category	category

Mapping dim_product dilakukan pada step Table Output. Berdasarkan gambar tersebut field hasil transformasi dari OLTP dipetakan langsung ke tabel OLAP dim_product. Atribut yang dimuat meliputi product_id, product_name, unit_price, dan category. Karena nama stream field sudah sesuai dengan nama table field, proses mapping dilakukan secara one-to-one tanpa perubahan tambahan.

2. Dim_Customer

Sumber OLTP: customers

Fields to insert:

#	Table field	Stream field
1	customer_id	customer_id
2	company_name	company_name
3	contact_name	contact_name
4	contact_title	contact_title
5	city	city
6	country	country

Mapping dim_customer dilakukan pada step **Table Output**.

Berdasarkan gambar tersebut, field hasil transformasi dari OLTP dipetakan langsung ke tabel OLAP dim_customer. Atribut yang dimuat meliputi customer_id, company_name, contact_name, contact_title, city, dan country. Karena nama pada stream field sudah sama dengan nama table field, mapping dilakukan secara one-to-one tanpa perubahan tambahan sehingga data siap masuk ke dimensi customer.

3. Dim_Employee

Sumber OLTP: employees

Fields to insert:

#	Table field	Stream field
1	employee_id	employee_id
2	employee_name	employee_name
3	title	title
4	city	city
5	country	country

Mapping dim_employee dilakukan pada step **Table Output**.

Berdasarkan gambar tersebut, field hasil transformasi dari OLTP dipetakan langsung ke tabel OLAP dim_employee. Atribut yang dimuat meliputi employee_id, employee_name, title, city, dan country. Karena nama stream field sudah sesuai dengan nama table field, proses mapping dilakukan secara one-to-one tanpa perubahan tambahan, sehingga data siap dimuat ke dimensi employee.

4. Dim_Shipper

Sumber OLTP: shippers

Fields to insert:

#	Table field	Stream field
1	shipper_id	shipper_id
2	shipper_name	company_name

Mapping dim_shipper dilakukan pada step **Table Output**. Berdasarkan gambar tersebut, field hasil transformasi dari OLTP dipetakan ke tabel OLAP dim_shipper. Atribut shipper_id dipetakan langsung secara one-to-one. Sedangkan atribut shipper_name diisi dari stream field company_name, yaitu hasil rename dari kolom nama perusahaan shipper pada OLTP. Dengan mapping ini, dim_shipper memiliki identitas shipper dan nama shipper yang sesuai untuk kebutuhan analisis OLAP.

5. Dim_Time

Sumber OLTP: orders.orderDate

Fields to insert:

#	Table field	Stream field
1	full_date	full_date
2	day	day
3	month	month
4	year	year
5	quarter	quarter
6	time_id	time_id
7	month_name	month_name
8	weekday	weekday

Mapping dim_time dilakukan pada step **Table Output**. Berdasarkan gambar tersebut, seluruh field hasil transformasi dipetakan langsung ke tabel OLAP dim_time secara one-to-one. Atribut yang dimuat meliputi full_date, day, month, year, quarter, time_id, month_name, dan weekday. Field-field tersebut merupakan hasil pembentukan atribut waktu melalui Calculator dan Modified JavaScript Value, sehingga

data yang dimuat sudah lengkap dan sesuai kebutuhan analisis berbasis waktu pada OLAP.

B. Mapping Tabel Fakta

1. Fact_Sales

Sumber OLTP: JOIN orders + order_details

Fields to insert:

#	Table field	Stream field
1	order_id	order_id
2	customer_id	customer_id
3	employee_id	employee_id
4	product_id	product_id
5	shipper_id	shipper_id
6	unit_price	unit_price
7	quantity	quantity
8	discount	discount
9	total_sales	total_sales
1.	time_id	time_id

Mapping fact_sales dilakukan pada step **Table Output**. Berdasarkan gambar tersebut , field hasil transformasi dipetakan langsung ke tabel OLAP fact_sales. Kolom order_id, customer_id, employee_id, product_id, shipper_id, dan time_id berperan sebagai foreign key yang menghubungkan tabel fakta dengan tabel dimensi terkait. Sementara itu, unit_price, quantity, discount, dan total_sales merupakan measure utama penjualan yang akan digunakan dalam analisis OLAP.

Karena nama stream field sudah sama dengan nama table field, proses mapping dilakukan secara one-to-one tanpa perubahan tambahan sehingga data siap dimuat ke tabel fakta.

BAB 5

IMPLEMENTASI ETL PIPELINE

5.1 Lingkungan dan Tools yang Digunakan

Pada tahap implementasi ETL, proses extract–transform–load dilakukan menggunakan **Pentaho Data Integration (PDI/Spoon)** sebagai tools utama untuk membangun pipeline ETL. Database yang digunakan adalah **MySQL** dengan dua lingkungan data, yaitu **OLTP** sebagai sumber data dan **OLAP/Data Warehouse** sebagai target penyimpanan hasil ETL.

A. Tools / Software

Tools yang digunakan dalam implementasi ETL adalah :

1. **Pentaho Data Integration (Spoon)**
Digunakan untuk membuat transformation (.ktr) dan (jika ada) job (.kjb) untuk mengeksekusi ETL secara berurutan.
2. **MySQL Database**
Menyimpan data sumber (OLTP) dan data target (OLAP).
3. **DB Client (opsional)** seperti phpMyAdmin / MySQL Workbench
Digunakan untuk mengecek hasil load tabel dimensi dan fakta.

B. Lingkungan Database

Implementasi ETL menggunakan dua koneksi database :

1. **Koneksi OLTP (Sumber):** oltp_northwind
Berisi tabel transaksi seperti orders, order_details, products, categories, customers, employees, dan shippers.
2. **Koneksi OLAP (Target Data Warehouse):** (*sesuaikan nama koneksi kamu, mis. dw_northwind.olap / cyber.bi.olap*)
Berisi tabel star schema yaitu dim_product, dim_customer, dim_employee, dim_shipper, dim_time, dan fact_sales.

Kedua koneksi tersebut digunakan pada step utama Pentaho :

- **Table Input** untuk mengambil data dari OLTP
- **Table Output** untuk memasukkan data hasil transformasi ke OLAP

C. Perangkat / Sistem Operasi

Proses ETL dijalankan pada perangkat dengan sistem operasi **Windows** dan seluruh file ETL disimpan dalam satu folder proyek agar mudah dikelola dan dieksekusi ulang.

5.2 Struktur Folder Proyek

Pada implementasi ETL, seluruh file dan komponen proyek disimpan dalam struktur folder yang terorganisir agar proses pengembangan, eksekusi ulang (re-run), dan pemeliharaan pipeline lebih mudah. File ETL Pentaho terdiri dari dua jenis utama :

- **Transformation (.ktr)** untuk proses ETL per tabel (dimensi dan fakta).
- **Job (.kjb)** untuk menjalankan beberapa transformation secara berurutan (jika digunakan).

Struktur folder proyek ETL disusun sebagai berikut :

1. /ETL_Northwind/

- **/ktr/**

Berisi file transformation (.ktr) untuk setiap tabel OLAP, yaitu:

- dim_product.ktr
- dim_customer.ktr
- dim_employee.ktr
- dim_shipper.ktr
- dim_time.ktr
- fact_sales.ktr

- **/output/**

Berisi hasil export (Excel) karena pipeline menghasilkan file output berupa excel.

Pemisahan folder berdasarkan jenis file tersebut bertujuan agar setiap komponen ETL mudah ditemukan, jelas fungsinya, dan memudahkan proses dokumentasi maupun demonstrasi saat pengujian.

5.3 Desain Pipeline ETL

Pada tahap implementasi, pipeline ETL dirancang untuk memindahkan data dari sistem OLTP ke data warehouse OLAP berbentuk **star schema**. Desain pipeline mengikuti prinsip utama bahwa **tabel dimensi harus dimuat terlebih dahulu** sebelum tabel fakta, karena tabel fakta membutuhkan foreign key yang mengacu pada dimensi (product, customer, employee, shipper, dan time).

A. Urutan Eksekusi Pipeline

Pipeline dijalankan dengan urutan berikut :

1. ETL_DimProduct
2. ETL_DimCustomer
3. ETL_DimEmployee
4. ETL_DimShipper
5. ETL_DimTime
6. ETL_FactSales

Urutan ini memastikan seluruh data referensi (dimensi) sudah tersedia di OLAP sehingga proses validasi foreign key pada fact_sales melalui lookup dapat berjalan dengan baik.

B. Arsitektur Pipeline ETL

Secara umum, setiap transformation dimensi menggunakan pola pipeline yang sederhana, yaitu :

1. Table Input (extract dari OLTP)
2. Select Values (penyesuaian field)
3. (*opsional*) Database Lookup (melengkapi atribut, seperti kategori pada dim_product)
4. Filter Rows (cleaning data invalid)
5. Table Output (load ke OLAP)

Sedangkan transformation fact_sales memiliki proses lebih kompleks karena :

- Mengambil data transaksi dari join orders dan order_details
- Melakukan perhitungan measure (subtotal/total_sales)
- Membuat time_id
- Melakukan validasi foreign key ke semua dimensi menggunakan Database Lookup berurutan.

C. Mekanisme Eksekusi

Pipeline dapat dieksekusi dengan dua cara :

1. **Menjalankan transformation (.ktr) secara manual** sesuai urutan dimensi → fakta.
2. **Menggunakan Job (.kjb)** untuk menjalankan seluruh transformation secara otomatis berurutan, sehingga eksekusi ETL lebih konsisten dan mudah diulang.

5.4 Penjelasan Setiap Komponen Step dalam Pipeline

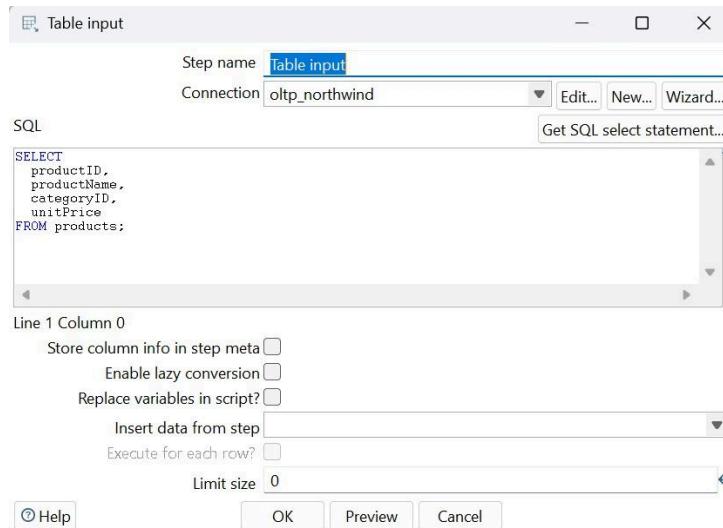
Komponen (step) yang digunakan dalam pipeline ETL pada Pentaho Data Integration. Setiap step memiliki peran tertentu dalam proses extract, transform, dan load, sehingga keseluruhan pipeline dapat menghasilkan data OLAP yang valid dan siap dianalisis, berikut adalah step pada setiap dimensi dan fakta :

A. Implementasi Pipeline Untuk Dim_Product

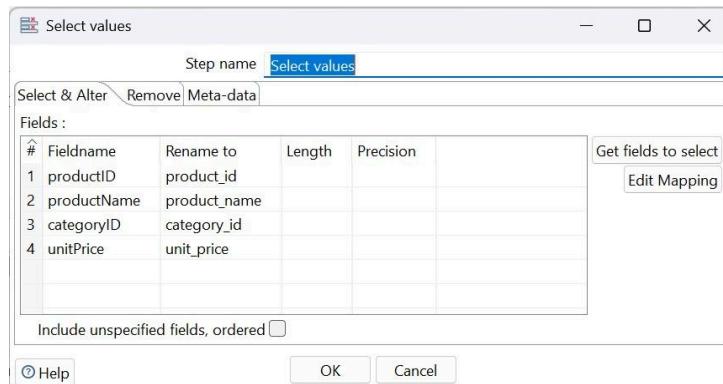
1. Tambahkan transformasi baru dan beri nama **ETL_DimProduct.ktr**
2. Drag & drop table input, select values, database lookup, filter rows, dan table output



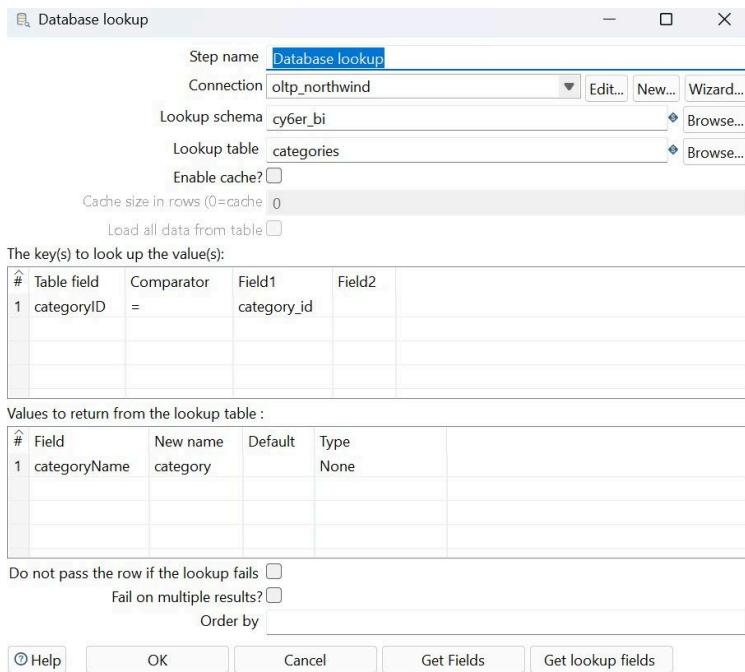
- Konfigurasi tabel input dan koneksi ke **oltp_northwind** lalu ketikkan query lalu klik preview untuk memastikan data keluar lalu klik ok



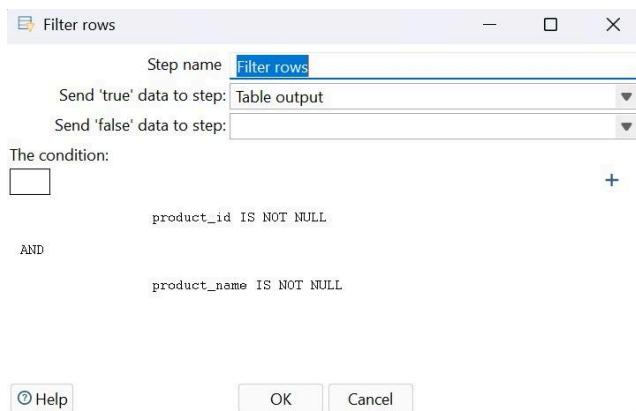
- Konfigurasi select values lalu klik get fields dan rename field tersebut sesuai kebutuhan lalu klik ok



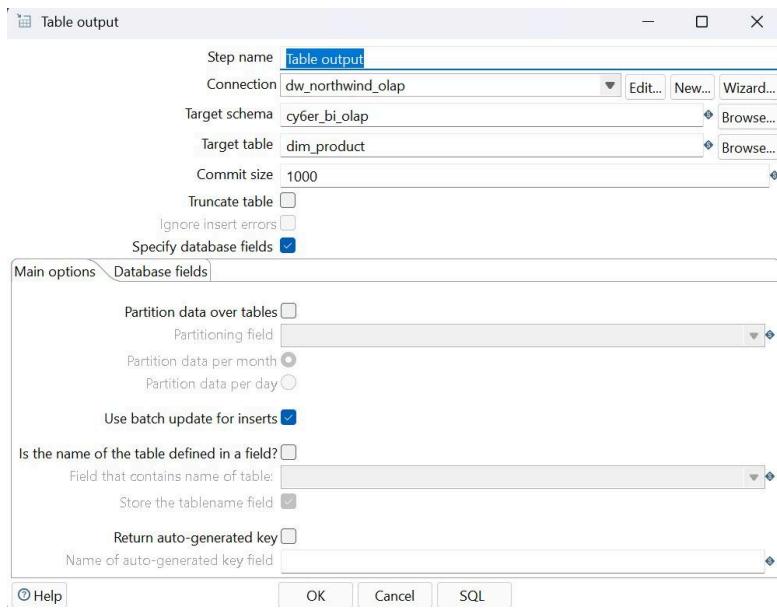
- Konfigurasi database lookup untuk mengambil kategori, koneksi terlebih dahulu ke **oltp_northwind** lalu pilih lookup table dan lookup schema sesuai dengan yang dituju, serta isikan bagian key(s) dan values nya lalu klik ok



6. Konfigurasi filter rows untuk cleaning data dengan memfilter kondisinya



7. Konfigurasi table output untuk load ke OLAP dengan mengkoneksikan ke **dw_northwind_olap** dan lalu pilih target table dan target schema sesuai dengan yang dituju, serta pada database fields dan klik get fields dan pastikan mapping sesuai dengan yang dituju lalu klik ok



- Jalankan transformasi dengan klik tombol run dan pastikan semua step hijau dan log menunjukkan proses selesai tanpa eror



Execution Results

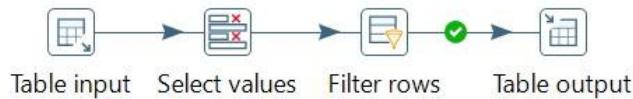
Logging	Execution History	Step Metrics	Performance Graph	Metrics	Preview data
2025-12-03 18:11:30.849 - Spoon - Transformation opened.					
2025-12-03 18:11:30.849 - Spoon - Launching transformation [ETL_DimProduct]...					
2025-12-03 18:11:30.849 - Spoon - Started the transformation execution.					
2025-12-03 18:11:30.868 - ETL_DimProduct - Dispatching started for transformation [ETL_DimProduct]					
2025-12-03 18:11:30.879 - Table output.0 - Connected to database [dw_northwind.olap] (commit=1000)					
2025-12-03 18:11:30.910 - Table input.0 - Finished reading query, closing connection					
2025-12-03 18:11:30.914 - Table input.0 - Finished processing (I=77, O=0, R=0, W=77, U=0, E=0)					
2025-12-03 18:11:30.923 - Select values.0 - Finished processing (I=0, O=0, R=77, W=77, U=0, E=0)					
2025-12-03 18:11:30.962 - Database lookup.0 - Finished processing (I=77, O=0, R=77, W=77, U=0, E=0)					
2025-12-03 18:11:30.968 - Filter rows.0 - Finished processing (I=0, O=0, R=77, W=77, U=0, E=0)					
2025-12-03 18:11:31.022 - Table output.0 - Finished processing (I=0, O=77, R=77, W=77, U=0, E=0)					
2025-12-03 18:11:31.024 - Spoon - The transformation has finished!!					

- Buka database OLAP dan cek tabel **dim_product** sudah terisi

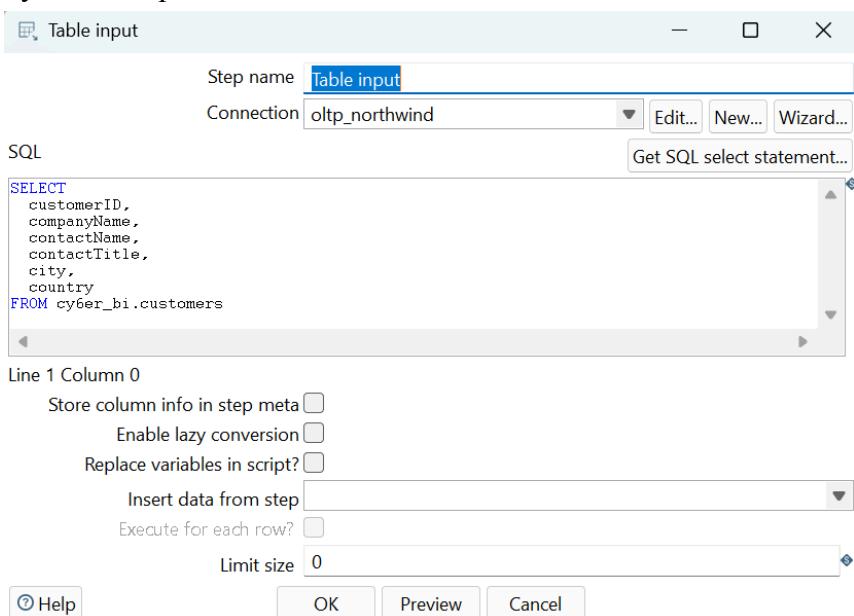
	product_id	product_name	category	unit_price
<input type="checkbox"/>	1	Chai	Beverages	18.00
<input type="checkbox"/>	2	Chang	Beverages	19.00
<input type="checkbox"/>	3	Aniseed Syrup	Condiments	10.00
<input type="checkbox"/>	4	Chef Anton's Cajun Seasoning	Condiments	22.00
<input type="checkbox"/>	5	Chef Anton's Gumbo Mix	Condiments	21.35
<input type="checkbox"/>	6	Grandma's Boysenberry Spread	Condiments	25.00
<input type="checkbox"/>	7	Uncle Bob's Organic Dried Pears	Produce	30.00
<input type="checkbox"/>	8	Northwoods Cranberry Sauce	Condiments	40.00
<input type="checkbox"/>	9	Mishi Kobe Niku	Meat & Poultry	97.00
<input type="checkbox"/>	10	Ikura	Seafood	31.00
<input type="checkbox"/>	11	Queso Cabrales	Dairy Products	21.00
<input type="checkbox"/>	12	Queso Manchego La Pastora	Dairy Products	38.00
<input type="checkbox"/>	13	Konbu	Seafood	6.00
<input type="checkbox"/>	14	Tofu	Produce	23.25
<input type="checkbox"/>	15	Genen Shouyu	Condiments	15.50
<input type="checkbox"/>	16	Pavlova	Confections	17.45
<input type="checkbox"/>	17	Alice Mutton	Meat & Poultry	39.00
<input type="checkbox"/>	18	Camarvon Tigers	Seafood	62.50
<input type="checkbox"/>	19	Teatime Chocolate Biscuits	Confections	9.20
<input type="checkbox"/>	20	Sir Rodney's Marmalade	Confections	81.00
<input type="checkbox"/>	21	Sir Rodney's Scones	Confections	10.00
<input type="checkbox"/>	22	Gustaf's Knäckebrot	Grains & Cereals	21.00
<input type="checkbox"/>	23	Tunnbröd	Grains & Cereals	9.00
<input type="checkbox"/>	24	Guarana Fantastica	Beverages	4.50
<input type="checkbox"/>	25	NuNuCa Nuß-Nougat-Creme	Confections	14.00

B. Implementasi Pipeline Untuk Dim_Customer

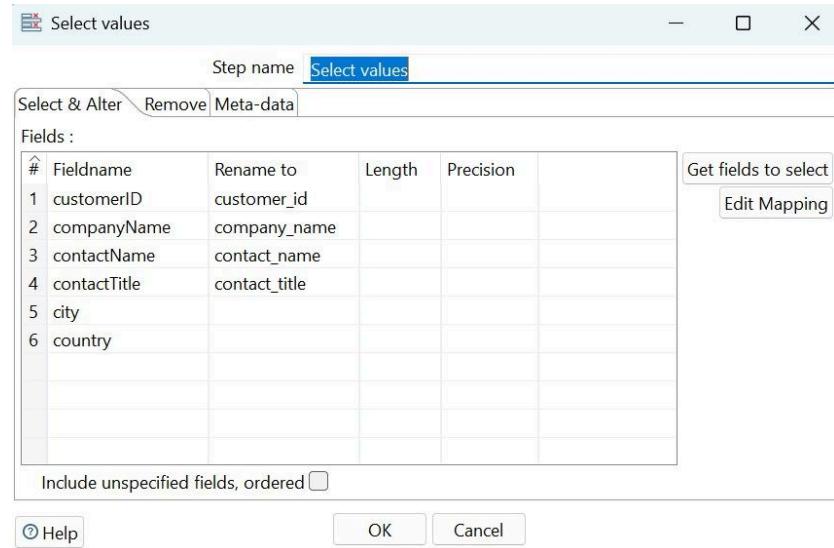
1. Tambahkan transformasi baru dan beri nama **ETL_DimCustomer.ktr**
2. Drag & drop table input, select values, filter rows, dan table output



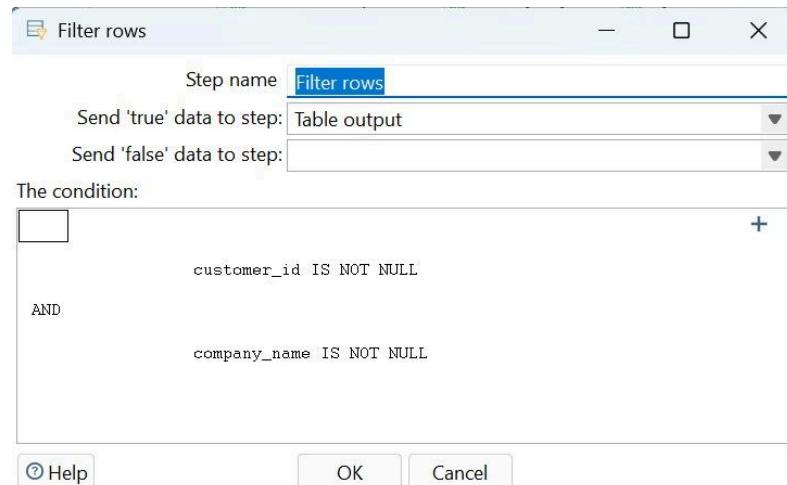
3. Konfigurasi tabel input dan koneksi ke **oltp_northwind** lalu ketikkan query lalu klik preview untuk memastikan data keluar lalu klik ok



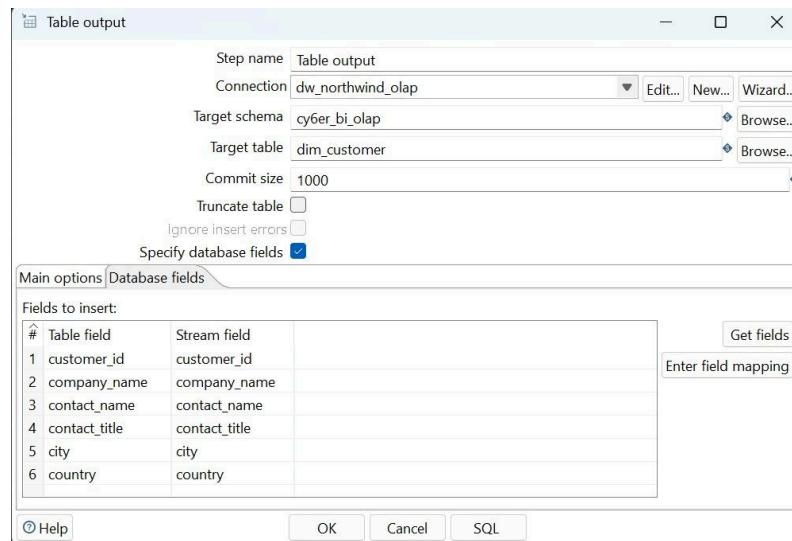
4. Konfigurasi select values lalu klik get fields dan rename field tersebut sesuai kebutuhan lalu klik ok



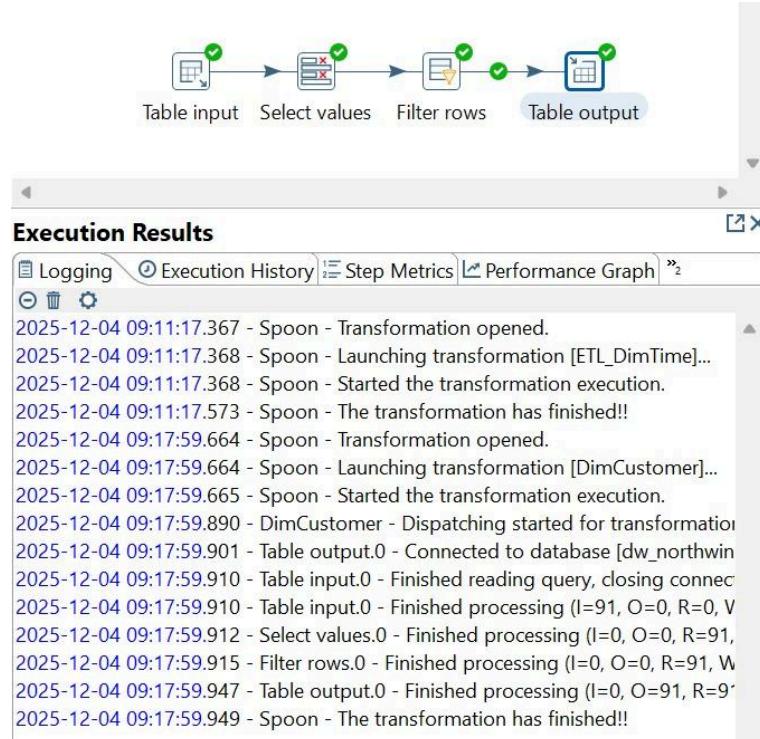
5. Konfigurasi filter rows untuk cleaning data dengan memfilter kondisinya



6. Konfigurasi table output untuk load ke OLAP dengan mengkoneksikan ke **dw_northwind.olap** dan lalu pilih target table dan target schema sesuai dengan yang dituju, serta pada database fields and klik get fields dan pastikan mapping sesuai dengan yang dituju lalu klik ok



7. Jalankan transformasi dengan klik tombol run dan pastikan semua step hijau dan log menunjukkan proses selesai tanpa eror

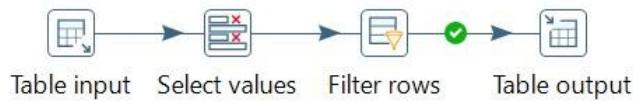


8. Buka database OLAP dan cek tabel **dim_customer** sudah terisi

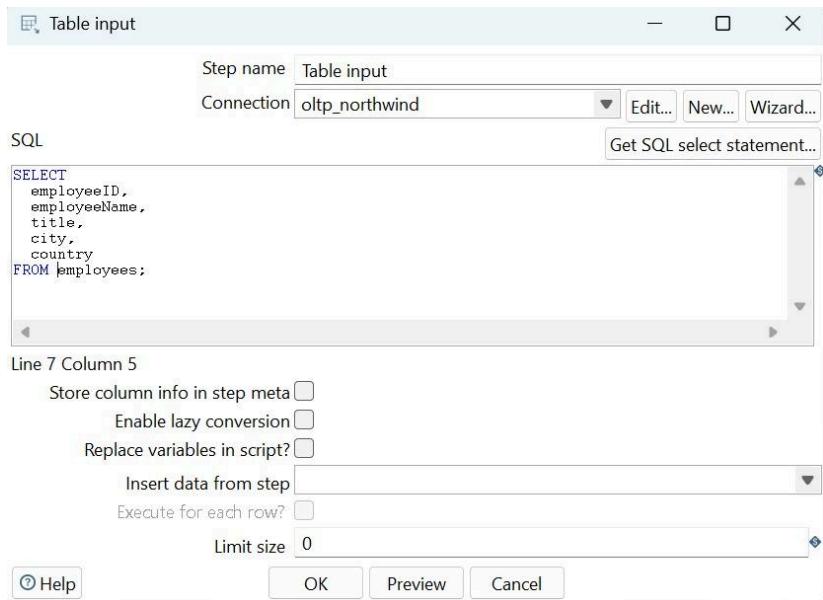
	customer_id	company_name	contact_name	contact_title	city	country
<input type="checkbox"/>	ALFKI	Alfreds Futterkiste	Maria Anders	Sales Representative	Berlin	Germany
<input type="checkbox"/>	ANATR	Aña Trujillo Emparedados y helados	Aña Trujillo	Owner	Mexico City	Mexico
<input type="checkbox"/>	ANTON	Antonio Moreno Taquería	Antonio Moreno	Owner	Mexico City	Mexico
<input type="checkbox"/>	AROUT	Around the Horn	Thomas Hardy	Sales Representative	London	UK
<input type="checkbox"/>	BERGS	Berglunds snabbköp	Christina Berglund	Order Administrator	Luleå	Sweden
<input type="checkbox"/>	BLAUS	Blauer See Delikatessen	Hanna Moos	Sales Representative	Mannheim	Germany
<input type="checkbox"/>	BLONP	Blondesdsl père et fils	Frédérique Citeaux	Marketing Manager	Strasbourg	France
<input type="checkbox"/>	BOLID	Bólido Comidas preparadas	Martin Sommer	Owner	Madrid	Spain
<input type="checkbox"/>	BONAP	Bon app'	Laurence Lebihan	Owner	Marseille	France
<input type="checkbox"/>	BOTTIM	Bottom-Dollar Markets	Elizabeth Lincoln	Accounting Manager	Tsawassen	Canada
<input type="checkbox"/>	BSBEV	B's Beverages	Victoria Ashworth	Sales Representative	London	UK
<input type="checkbox"/>	CACTU	Cactus Comidas para llevar	Patricia Simpson	Sales Agent	Buenos Aires	Argentina
<input type="checkbox"/>	CENTC	Centro comercial Moctezuma	Francisco Chang	Marketing Manager	Mexico City	Mexico
<input type="checkbox"/>	CHOPS	Chop-suey Chinese	Yang Wang	Owner	Bern	Switzerland
<input type="checkbox"/>	COMM1	Comércio Mineiro	Pedro Afonso	Sales Associate	Sao Paulo	Brazil
<input type="checkbox"/>	CONSH	Consolidated Holdings	Elizabeth Brown	Sales Representative	London	UK
<input type="checkbox"/>	DRACD	Drachenblut Delikatessen	Sven Ottlieb	Order Administrator	Aachen	Germany
<input type="checkbox"/>	DUMON	Du monde entier	Janine Labrune	Owner	Nantes	France
<input type="checkbox"/>	EASTC	Eastern Connection	Ann Devon	Sales Agent	London	UK
<input type="checkbox"/>	ERNSH	Ernst Handel	Roland Mendel	Sales Manager	Graz	Austria
<input type="checkbox"/>	FAMIA	Familia Arquibaldo	Aria Cruz	Marketing Assistant	Sao Paulo	Brazil
<input type="checkbox"/>	FISSA	FISSA Fábrica Inter. Salchichas S.A.	Diego Roel	Accounting Manager	Madrid	Spain
<input type="checkbox"/>	FOLIG	Folies gourmandes	Martine Rancé	Assistant Sales Agent	Lille	France
<input type="checkbox"/>	FOLKO	Folk och fä HB	Maria Larsson	Owner	Bracke	Sweden
<input type="checkbox"/>	FRANK	Frankenversand	Peter Franken	Marketing Manager	München	Germany

C. Implementasi Pipeline Untuk Dim_Employee

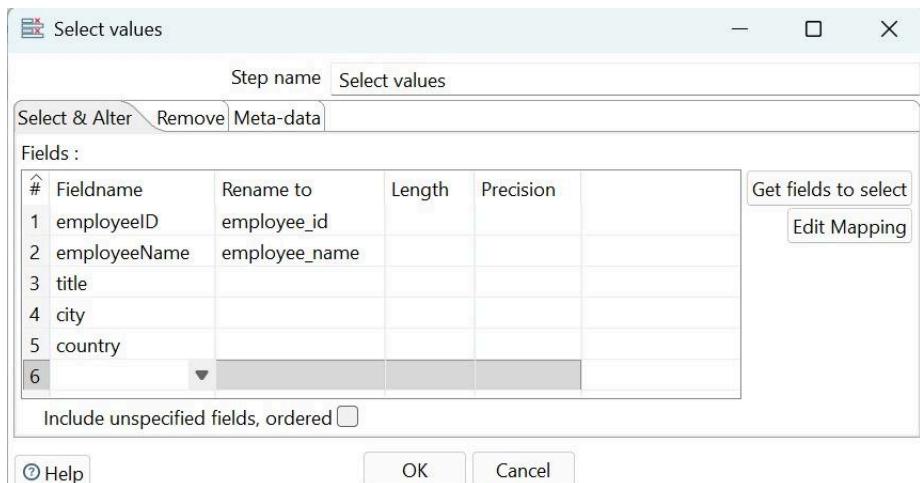
1. Tambahkan transformasi baru dan beri nama **ETL_DimEmployee.ktr**
2. Drag & drop table input, select values, filter rows, dan table output



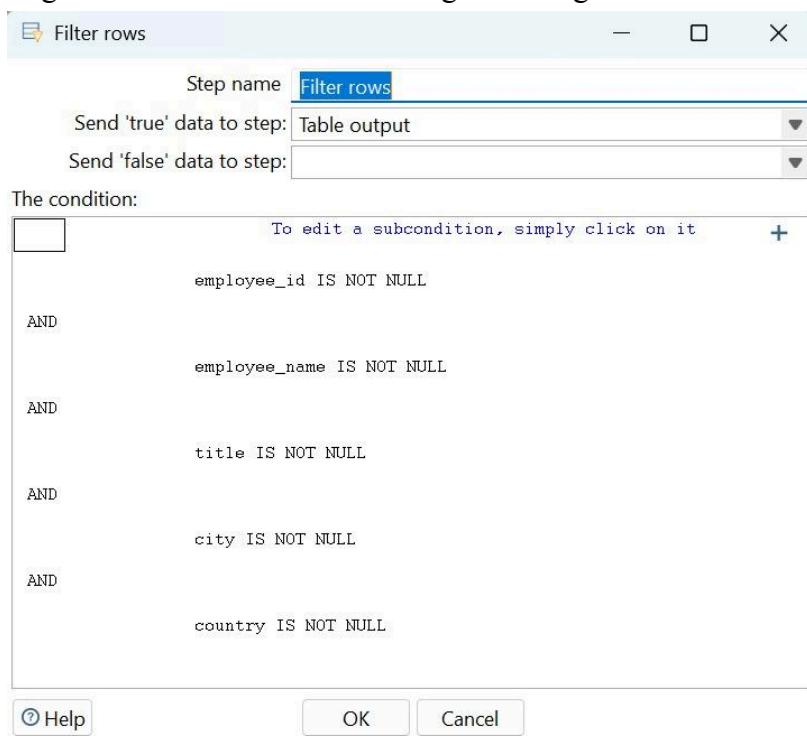
3. Konfigurasi tabel input dan koneksi ke **oltp_northwind** lalu ketikkan query lalu klik preview untuk memastikan data keluar lalu klik ok



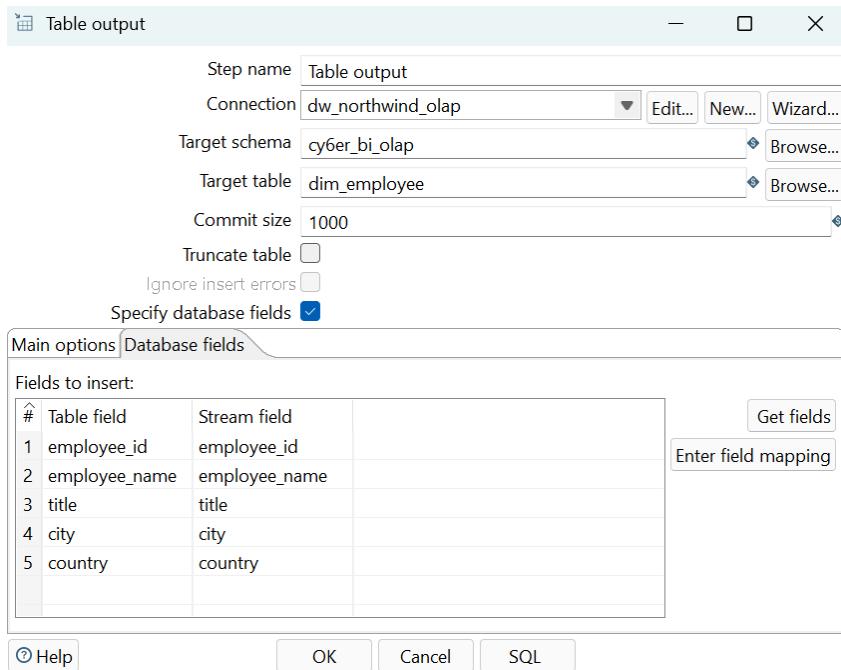
4. Konfigurasi select values lalu klik get fields dan rename field tersebut sesuai kebutuhan lalu klik ok



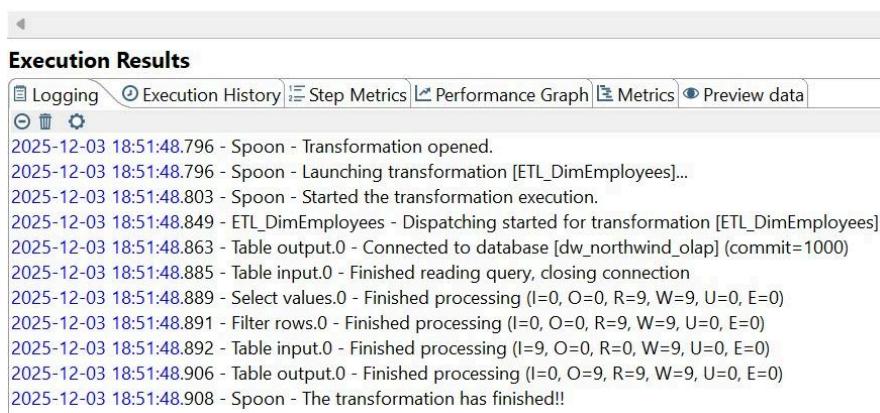
5. Konfigurasi filter rows untuk cleaning data dengan memfilter kondisinya



6. Konfigurasi table output untuk load ke OLAP dengan mengkoneksikan ke **dw_northwind_olap** dan lalu pilih target table dan target schema sesuai dengan yang dituju, serta pada database fields and klik get fields dan pastikan mapping sesuai dengan yang dituju lalu klik ok



7. Jalankan transformasi dengan klik tombol run dan pastikan semua step hijau dan log menunjukkan proses selesai tanpa eror



8. Buka database OLAP dan cek tabel dim_employee sudah terisi

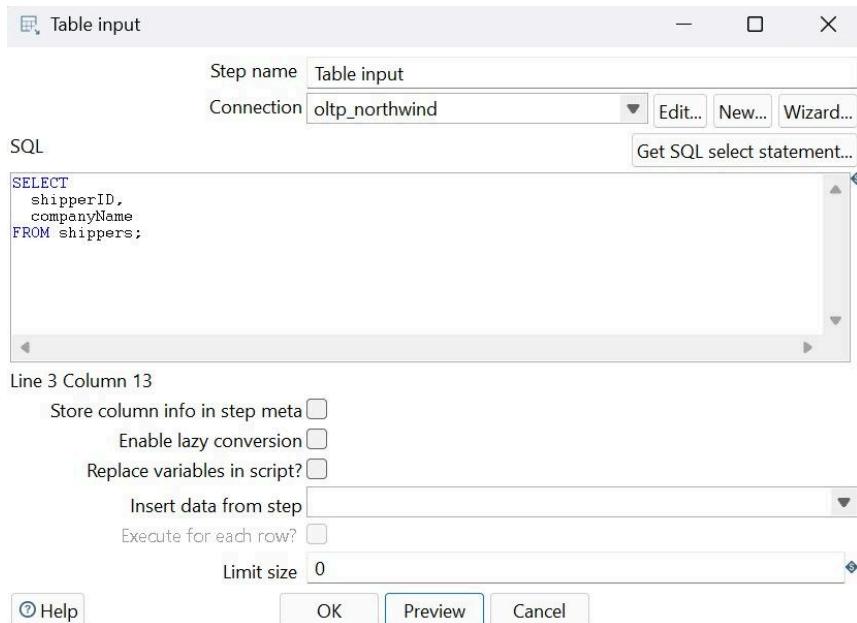
	employee_id	employee_name	title	city	country
<input type="checkbox"/>	1	Nancy Davolio	Sales Representative	New York	USA
<input type="checkbox"/>	2	Andrew Fuller	Vice President Sales	New York	USA
<input type="checkbox"/>	3	Janet Leverling	Sales Representative	New York	USA
<input type="checkbox"/>	4	Margaret Peacock	Sales Representative	New York	USA
<input type="checkbox"/>	5	Steven Buchanan	Sales Manager	London	UK
<input type="checkbox"/>	6	Michael Suyama	Sales Representative	London	UK
<input type="checkbox"/>	7	Robert King	Sales Representative	London	UK
<input type="checkbox"/>	8	Laura Callahan	Sales Manager	New York	USA
<input type="checkbox"/>	9	Anne Dodsworth	Sales Representative	London	UK

D. Implementasi Pipeline Untuk Dim_Shipper

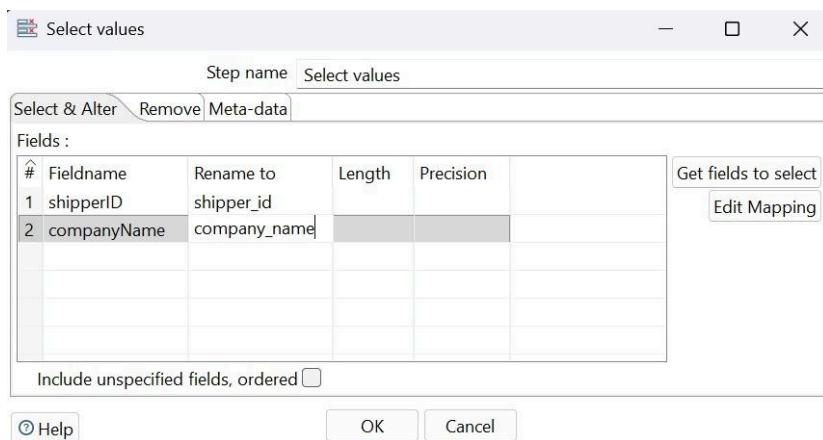
1. Tambahkan transformasi baru dan beri nama **ETL_DimShipper.ktr**
2. Drag & drop table input, select values, filter rows, dan table output



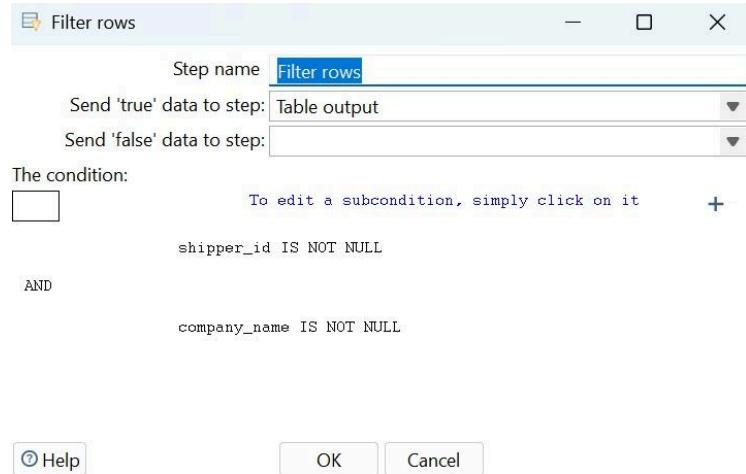
3. Konfigurasi tabel input dan koneksi ke **oltp_northwind** lalu ketikkan query lalu klik preview untuk memastikan data keluar lalu klik ok



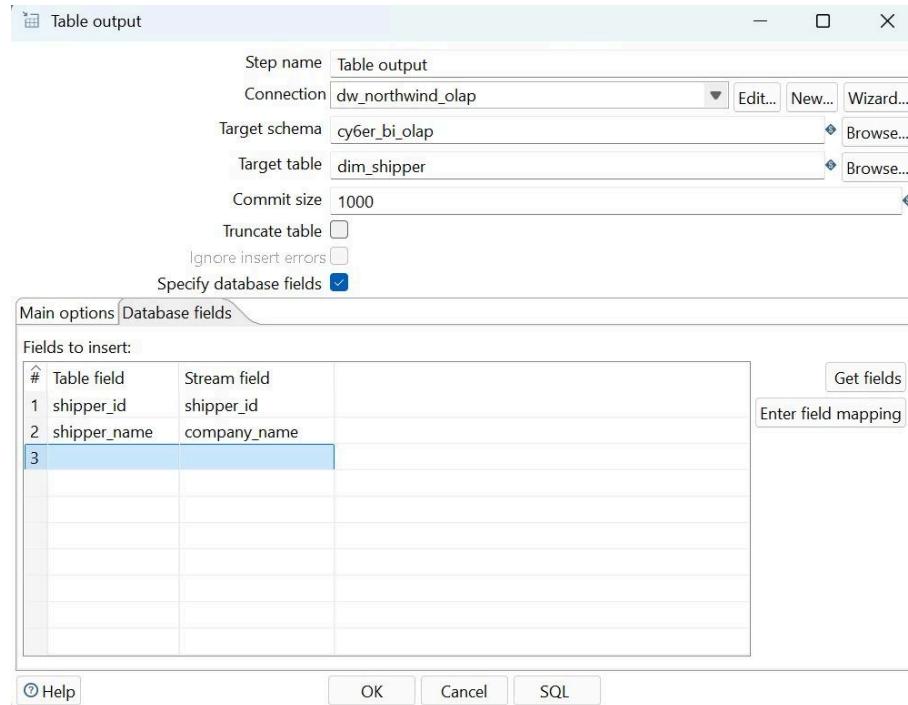
4. Konfigurasi select values lalu klik get fields dan rename field tersebut sesuai kebutuhan lalu klik ok



5. Konfigurasi filter rows untuk cleaning data dengan memfilter kondisinya



- Konfigurasi table output untuk load ke OLAP dengan mengkoneksikan ke **dw_northwind_olap** dan lalu pilih target table dan target schema sesuai dengan yang dituju, serta pada database fields dan klik get fields dan pastikan mapping sesuai dengan yang dituju lalu klik ok



- Jalankan transformasi dengan klik tombol run dan pastikan semua step hijau dan log menunjukkan proses selesai tanpa eror



Execution Results

- Logging
- Execution History
- Step Metrics
- Performance Graph
- Metrics
- Preview data

2025-12-03 18:34:18.992 - Spoon - Transformation opened.
 2025-12-03 18:34:18.992 - Spoon - Launching transformation [ETL_DimShipper]...
 2025-12-03 18:34:18.992 - Spoon - Started the transformation execution.
 2025-12-03 18:34:19.029 - ETL_DimShipper - Dispatching started for transformation [ETL_DimShipper]
 2025-12-03 18:34:19.036 - Table output.0 - Connected to database [dw_northwind_olap] (commit=1000)
 2025-12-03 18:34:19.051 - Table input.0 - Finished reading query, closing connection
 2025-12-03 18:34:19.054 - Select values.0 - Finished processing (I=0, O=0, R=3, W=3, U=0, E=0)
 2025-12-03 18:34:19.054 - Table input.0 - Finished processing (I=3, O=0, R=0, W=3, U=0, E=0)
 2025-12-03 18:34:19.056 - Filter rows.0 - Finished processing (I=0, O=0, R=3, W=3, U=0, E=0)
 2025-12-03 18:34:19.068 - Table output.0 - Finished processing (I=0, O=3, R=3, W=3, U=0, E=0)
 2025-12-03 18:34:19.074 - Spoon - The transformation has finished!!

8. Buka database OLAP dan cek tabel **dim_shipper** sudah terisi

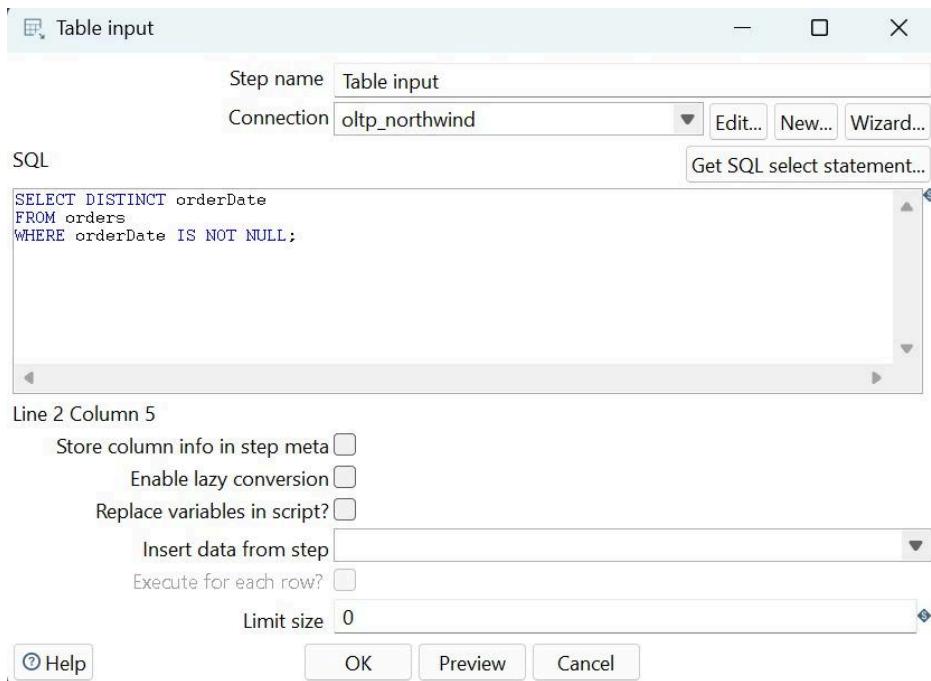
	shipper_id	shipper_name
<input type="checkbox"/>	1	Speedy Express
<input type="checkbox"/>	2	United Package
<input type="checkbox"/>	3	Federal Shipping

E. Implementasi Pipeline Untuk Dim_Time

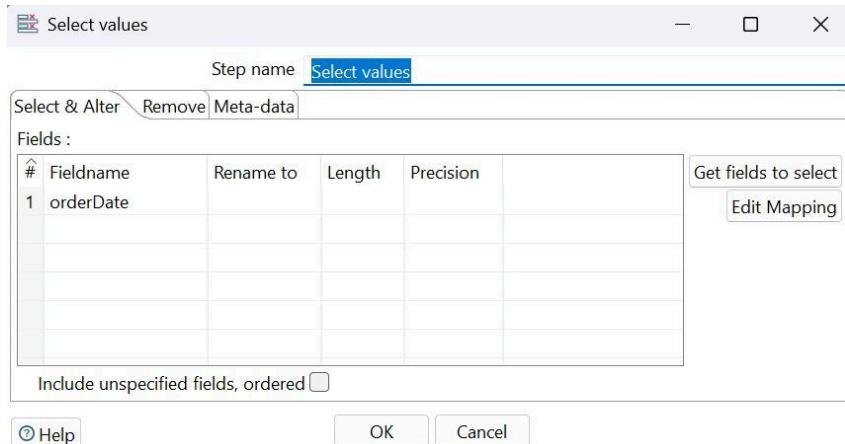
1. Tambahkan transformasi baru dan beri nama **ETL_DimTime.ktr**
2. Drag & drop table input, select values, calculator, modified JavaScript values, select values 2, dan table output



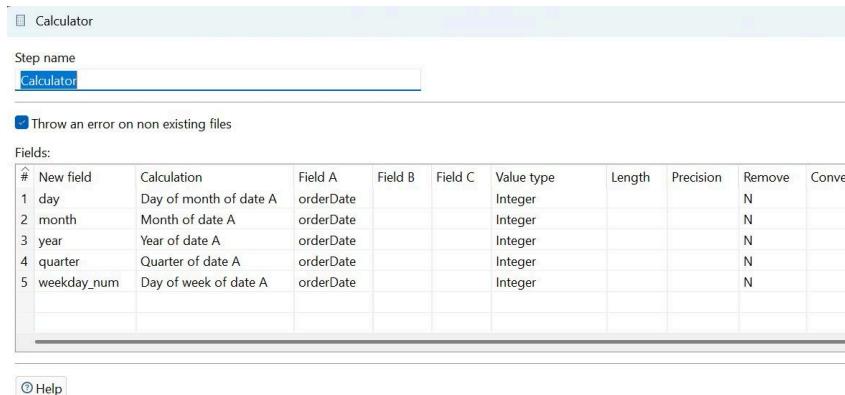
3. Konfigurasi tabel input dan koneksi ke **oltp_northwind** lalu ketikkan query lalu klik preview untuk memastikan data keluar lalu klik ok



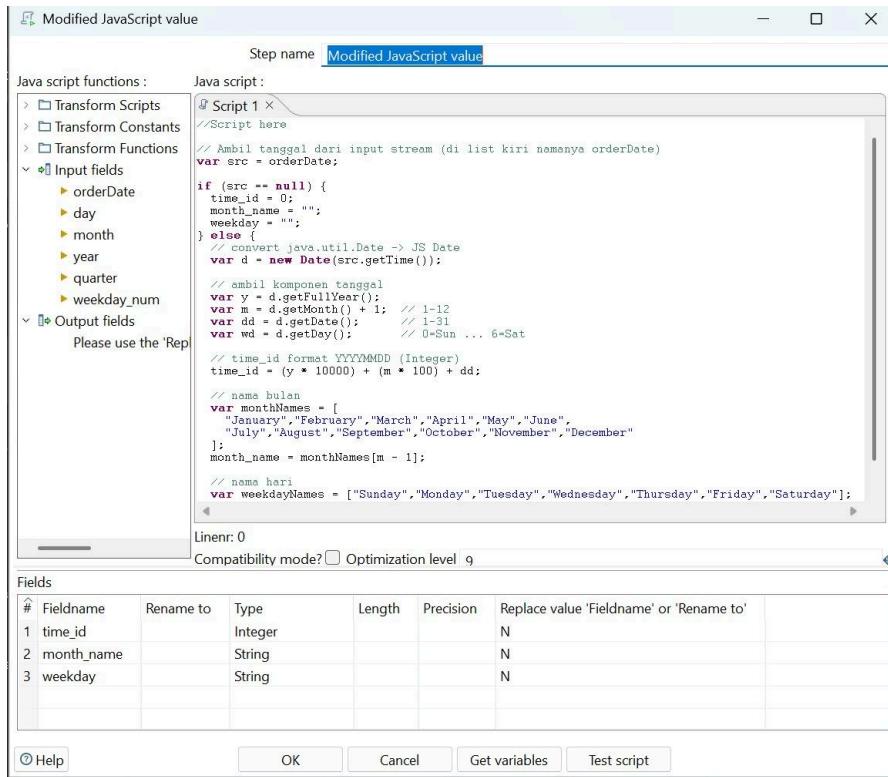
- Konfigurasi select values 1 lalu klik get fields dan rename field tersebut sesuai kebutuhan lalu klik ok



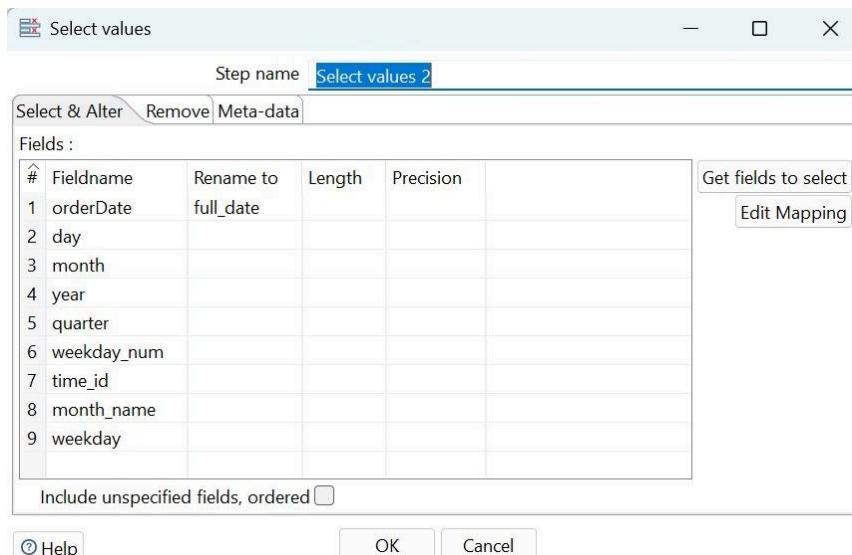
- Konfigurasi calculator untuk membuat atribut waktu angka dengan menambahkan field baru berupa day, month, year, quarter, weekday_num yang dimana kita memasukkan field A nya sesuai yang dibutuhkan, lalu klik ok



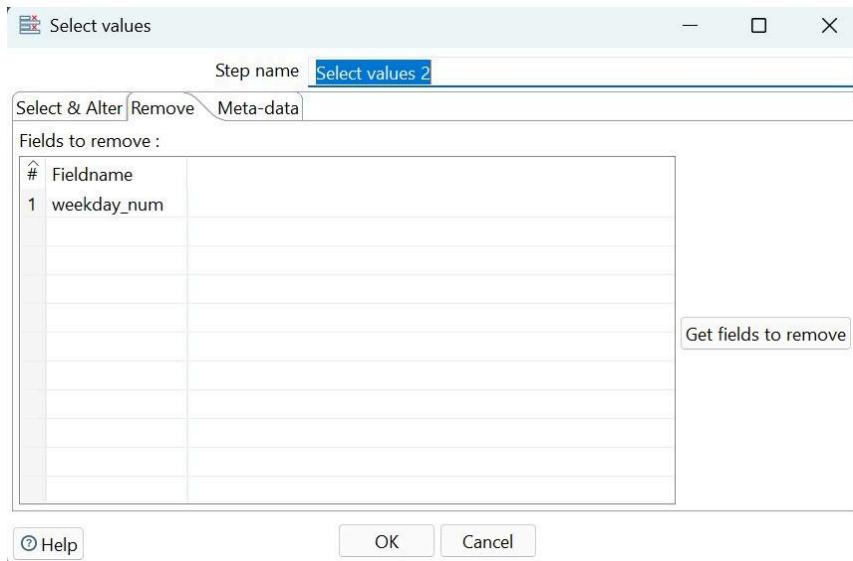
6. Konfigurasi modified JavaScript value untuk membuat time_id, month_name dan weekday dengan cara menambahkan script yang sesuai lalu mengisi pada output field berupa time_id, month_name, weekday lalu klik ok



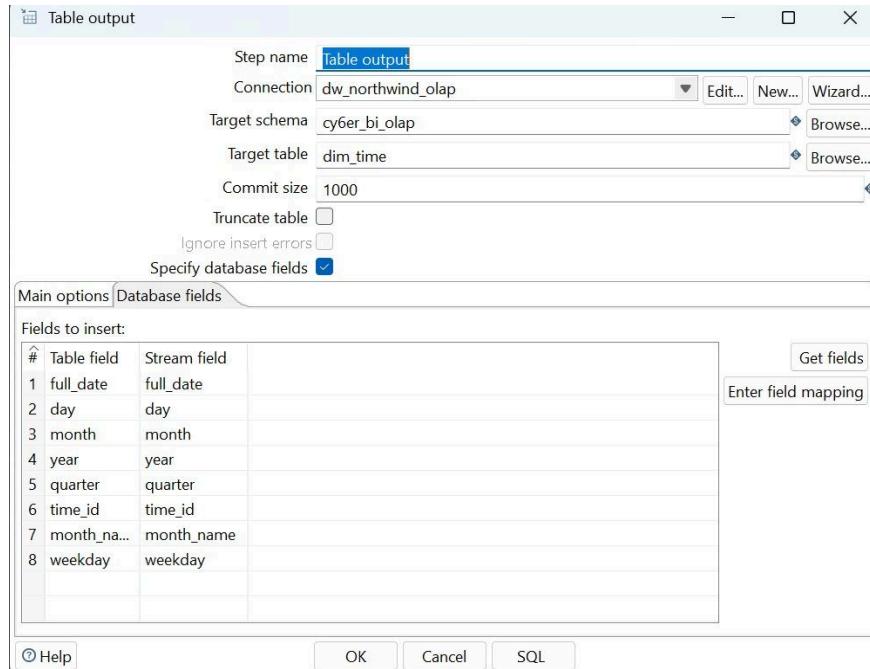
7. Konfigurasi select values 2 untuk finalisasi dan juga membuang field yang tidak dipakai di tabel OLAP dan memastikan yang tersisa hanya field yang akan dimuat, lalu klik ok



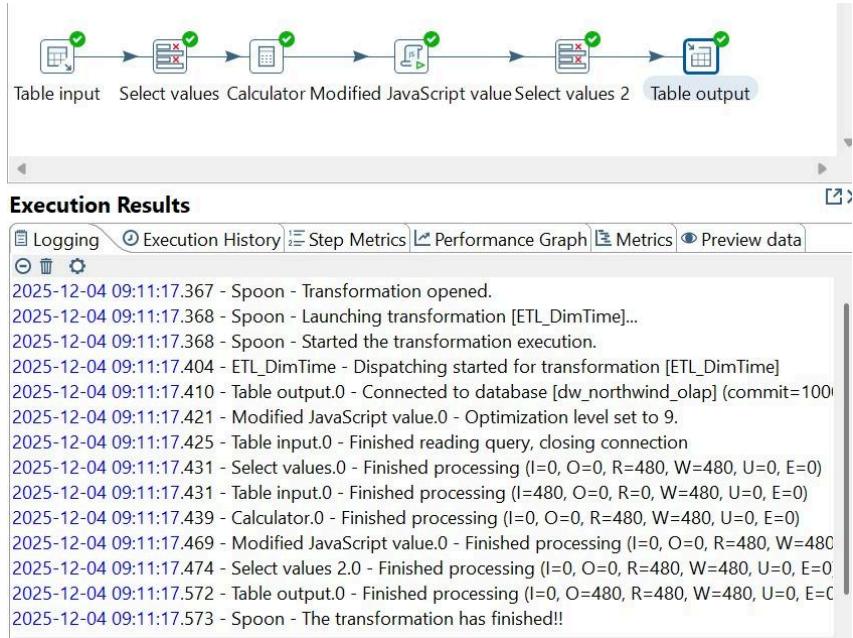
Field yang dihapus :



8. Konfigurasi table output untuk load ke OLAP dengan mengkoneksikan ke **dw_northwind_olap** dan lalu pilih target table dan target schema sesuai dengan yang dituju, serta pada database fields and klik get fields dan pastikan mapping sesuai dengan yang dituju lalu klik ok



9. Jalankan transformasi dengan klik tombol run dan pastikan semua step hijau dan log menunjukkan proses selesai tanpa eror



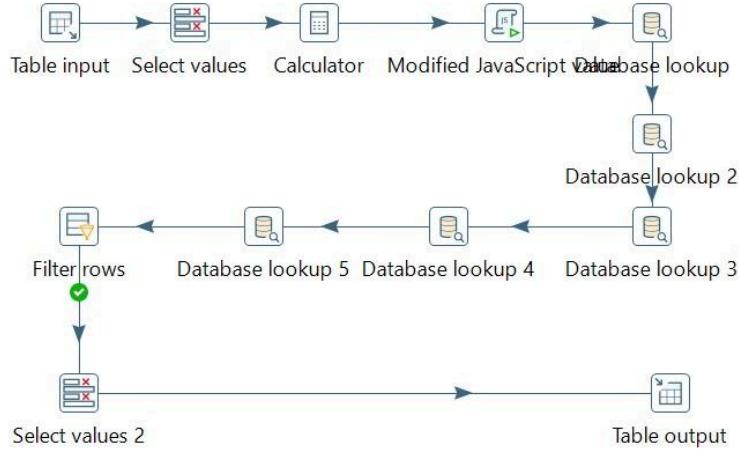
10. Buka database OLAP dan cek tabel **dim_time** sudah terisi

	time_id	full_date	day	month	month_name	quarter	year	weekday
Delete	20130704	2013-07-04	4	7	July	3	2013	Thursday
Delete	20130705	2013-07-05	5	7	July	3	2013	Friday
Delete	20130708	2013-07-08	8	7	July	3	2013	Monday
Delete	20130709	2013-07-09	9	7	July	3	2013	Tuesday
Delete	20130710	2013-07-10	10	7	July	3	2013	Wednesday
Delete	20130711	2013-07-11	11	7	July	3	2013	Thursday
Delete	20130712	2013-07-12	12	7	July	3	2013	Friday
Delete	20130715	2013-07-15	15	7	July	3	2013	Monday
Delete	20130716	2013-07-16	16	7	July	3	2013	Tuesday
Delete	20130717	2013-07-17	17	7	July	3	2013	Wednesday
Delete	20130718	2013-07-18	18	7	July	3	2013	Thursday
Delete	20130719	2013-07-19	19	7	July	3	2013	Friday
Delete	20130722	2013-07-22	22	7	July	3	2013	Monday
Delete	20130723	2013-07-23	23	7	July	3	2013	Tuesday
Delete	20130724	2013-07-24	24	7	July	3	2013	Wednesday
Delete	20130725	2013-07-25	25	7	July	3	2013	Thursday
Delete	20130726	2013-07-26	26	7	July	3	2013	Friday
Delete	20130729	2013-07-29	29	7	July	3	2013	Monday
Delete	20130730	2013-07-30	30	7	July	3	2013	Tuesday
Delete	20130731	2013-07-31	31	7	July	3	2013	Wednesday
Delete	20130801	2013-08-01	1	8	August	3	2013	Thursday
Delete	20130802	2013-08-02	2	8	August	3	2013	Friday
Delete	20130805	2013-08-05	5	8	August	3	2013	Monday
Delete	20130806	2013-08-06	6	8	August	3	2013	Tuesday
Delete	20130807	2013-08-07	7	8	August	3	2013	Wednesday

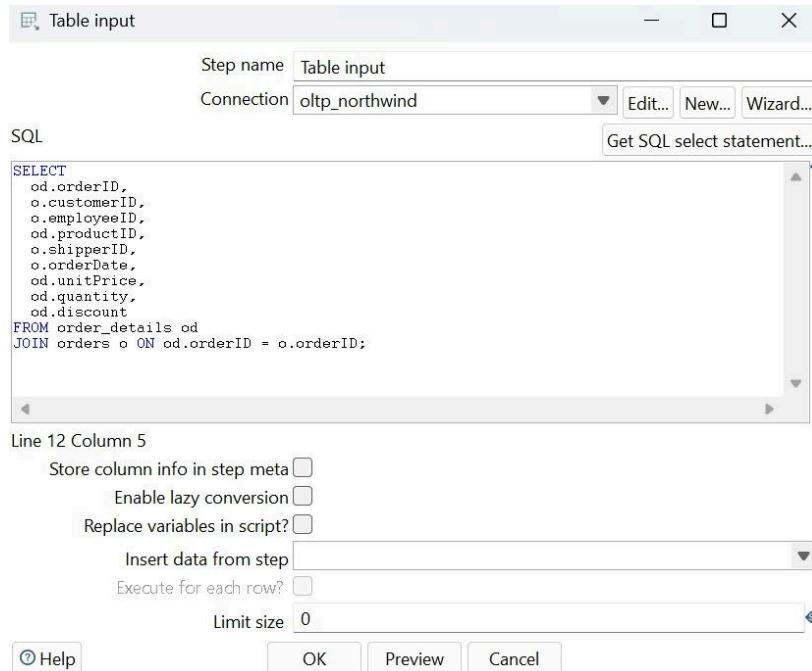
F. Implementasi Pipeline Untuk Fact_Sales

1. Tambahkan transformasi baru dan beri nama **ETL_Fact_Sales.ktr**

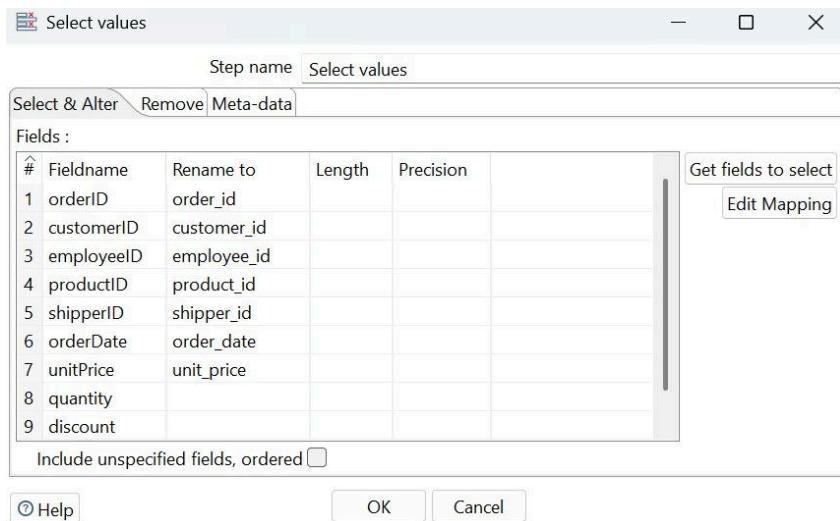
- Drag & drop table input, select values, calculator, modified JavaScript value, database lookup 1, database lookup 2, database lookup 3, database lookup 4, database lookup 5, filter rows, select values 2 dan table output



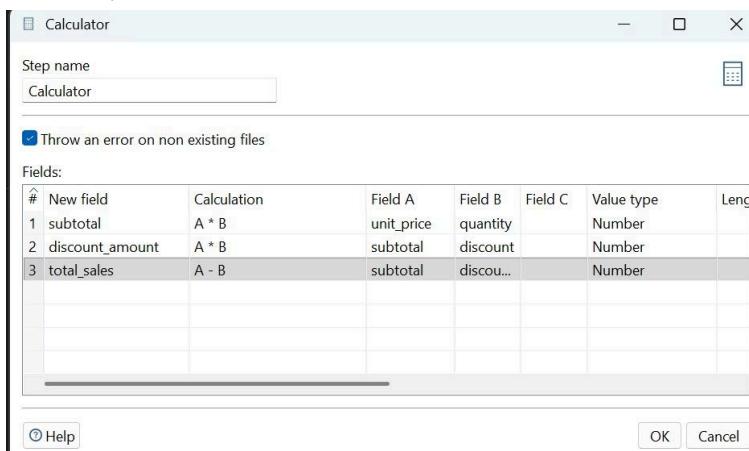
- Konfigurasi tabel input dan koneksi ke **oltp_northwind** lalu ketikkan query lalu klik preview untuk memastikan data keluar lalu klik ok



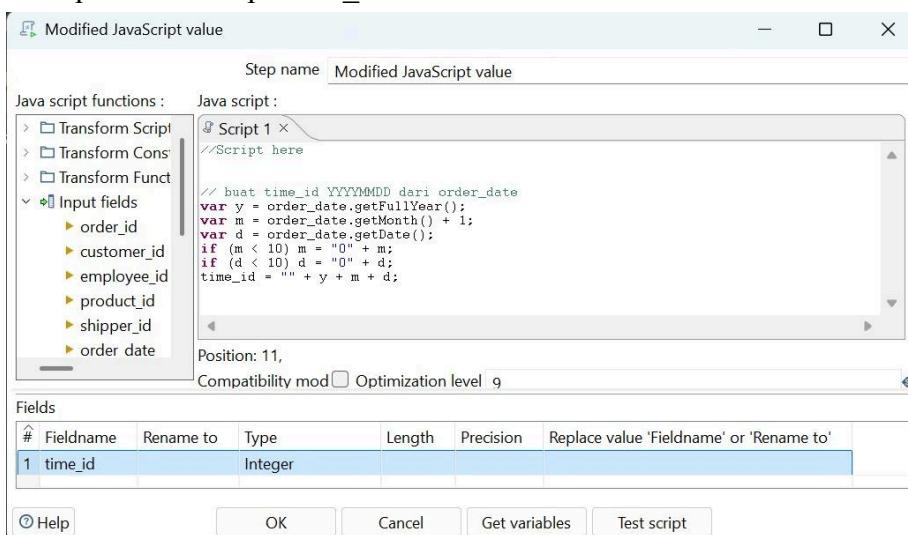
- Konfigurasi select values 1 lalu klik get fields dan rename field tersebut sesuai kebutuhan lalu klik ok



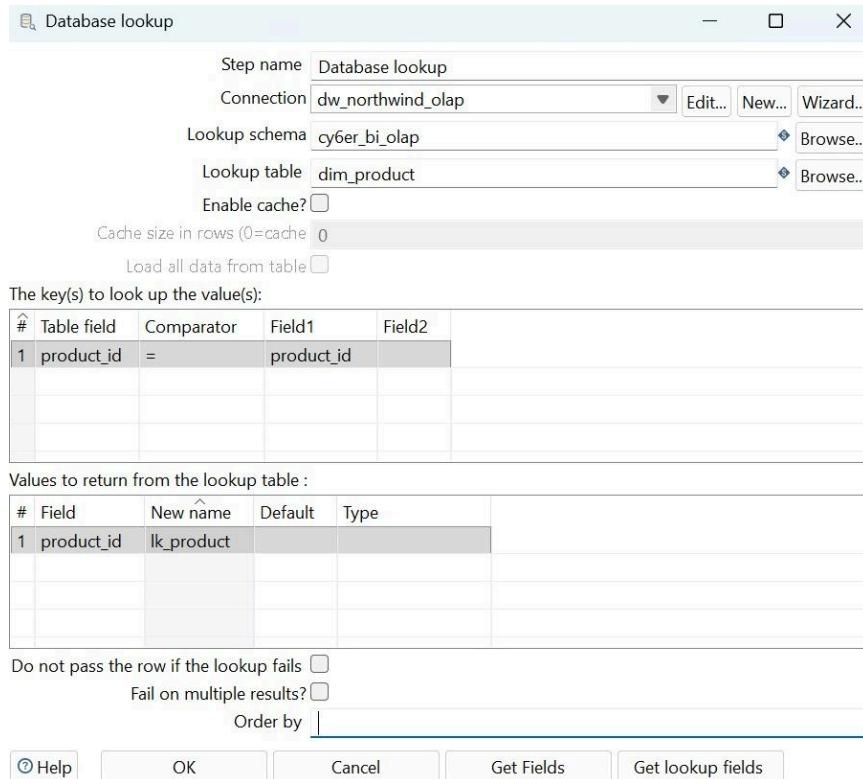
5. Konfigurasi calculator untuk membuat atribut waktu angka dengan menambahkan field baru berupa subtotal, discount_amount, dan total_sales yang dimana kita memasukkan field A nya sesuai yang dibutuhkan, lalu klik ok

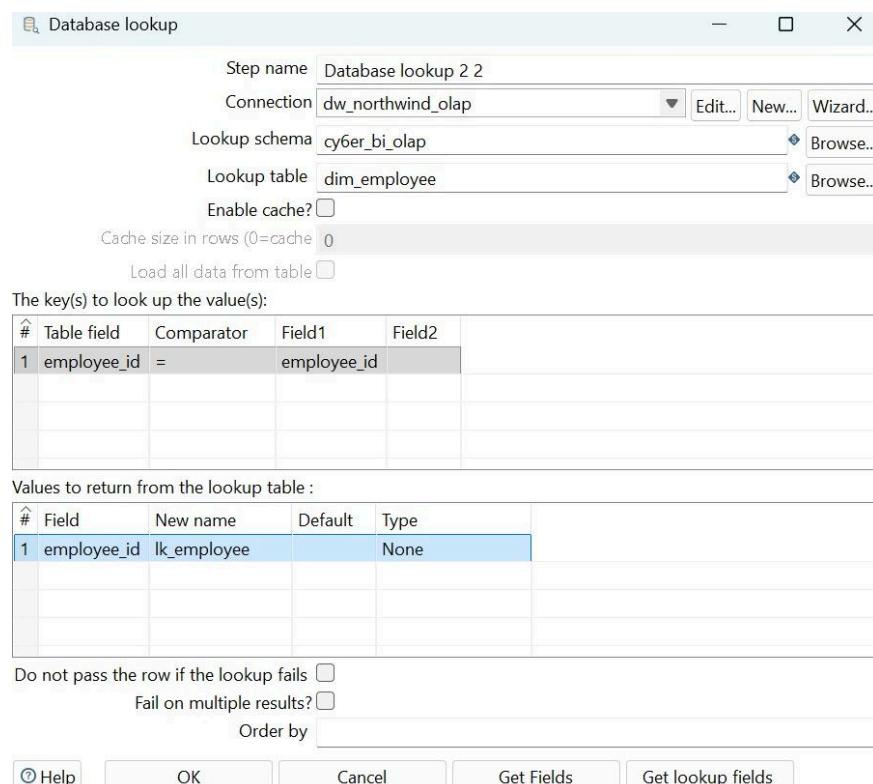
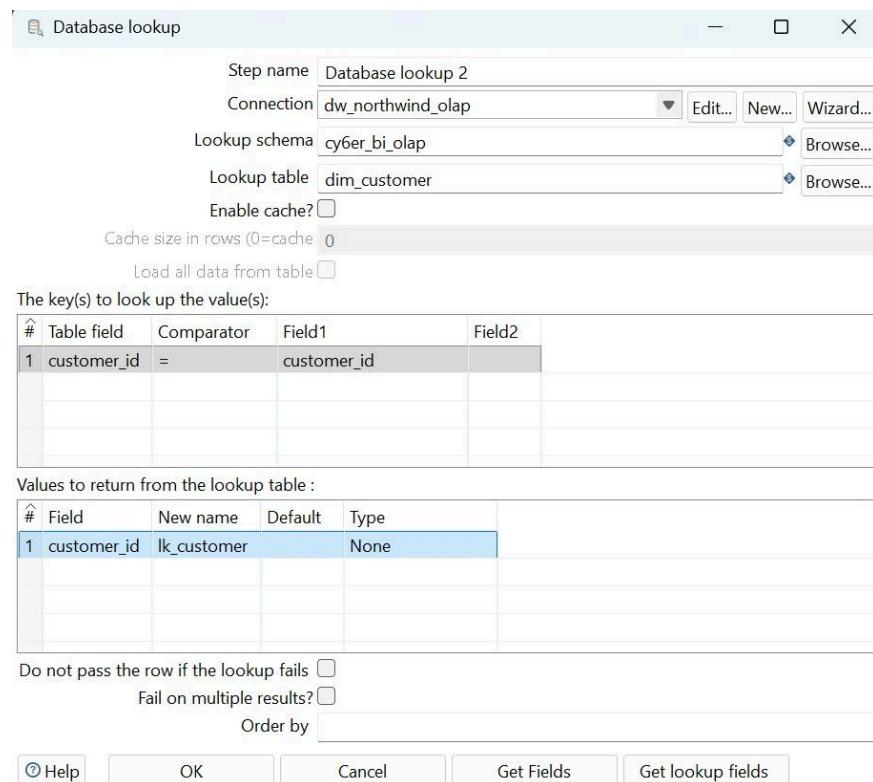


6. Konfigurasi modified JavaScript value untuk membuat time_id dari order_date, dengan cara menambahkan script yang sesuai lalu mengisi pada output field berupa time_id lalu klik ok



7. Konfigurasi database lookup 1, database lookup 2, database lookup 3, database lookup 4, database lookup 5, untuk mengambil kategori, koneksi terlebih dahulu ke **oltp_northwind** lalu pilih lookup table dan lookup schema sesuai dengan yang dituju, serta isikan bagian key(s) dan values nya lalu klik ok





Database lookup

Step name: Database lookup 2 2 2

Connection: dw_northwind_olap

Lookup schema: cy6er_bi_olap

Lookup table: dim_shipper

Enable cache?

Cache size in rows (0=cache): 0

Load all data from table

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

#	Table field	Comparator	Field1	Field2
1	shipper_id	=	shipper_id	

Values to return from the lookup table:

#	Field	New name	Default	Type
1	shipper_id	lk_shipper		None

Do not pass the row if the lookup fails

Fail on multiple results?

Order by

Help

Database lookup

Step name: Database lookup 2 2 2 2

Connection: dw_northwind_olap

Lookup schema: cy6er_bi_olap

Lookup table: dim_time

Enable cache?

Cache size in rows (0=cache): 0

Load all data from table

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

#	Table field	Comparator	Field1	Field2
1	time_id	=	time_id	

Values to return from the lookup table:

#	Field	New name	Default	Type
1	time_id	lk_time		None

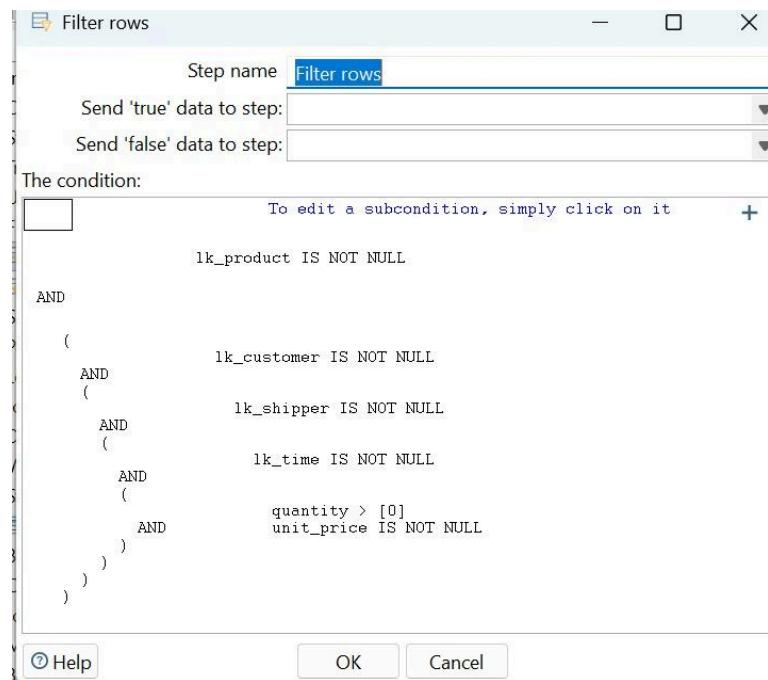
Do not pass the row if the lookup fails

Fail on multiple results?

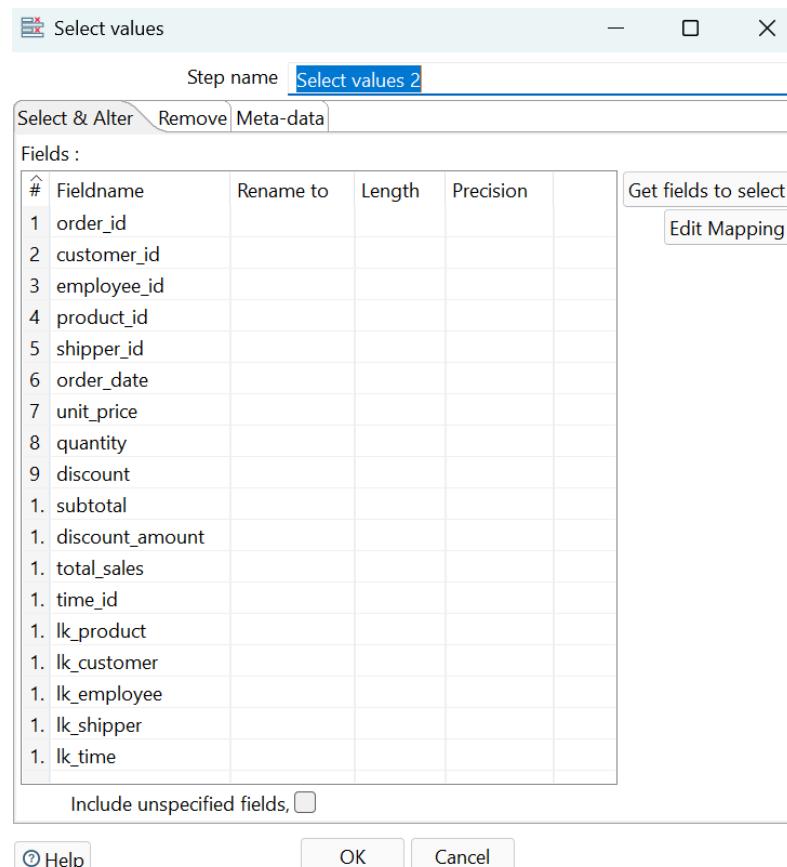
Order by

Help

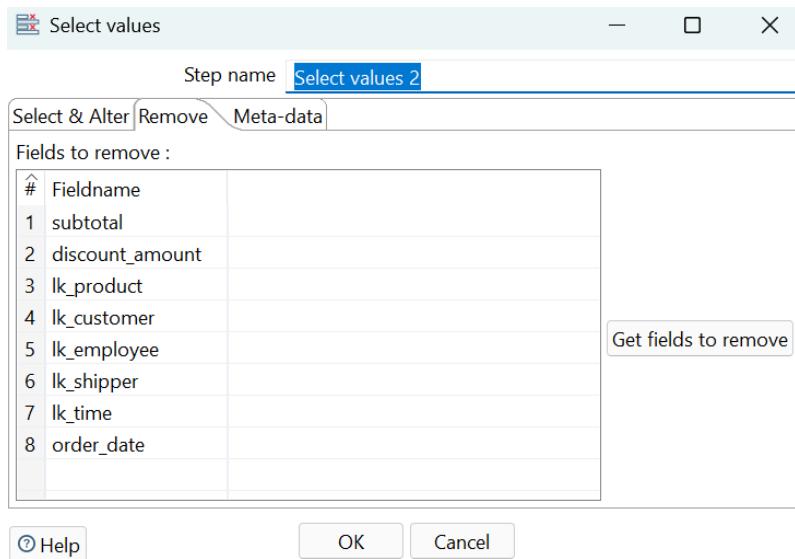
8. Konfigurasi filter rows untuk cleaning data dengan memfilter kondisinya



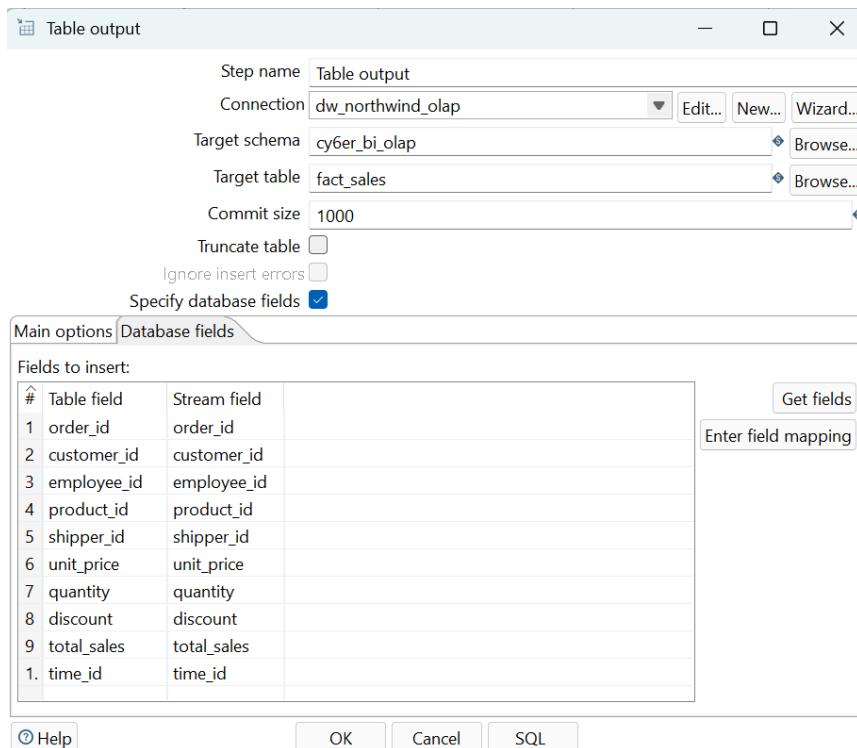
9. Konfigurasi select values 2 untuk finalisasi dan juga membuang field yang tidak dipakai di tabel OLAP dan memastikan yang tersisa hanya field yang akan dimuat, lalu klik ok



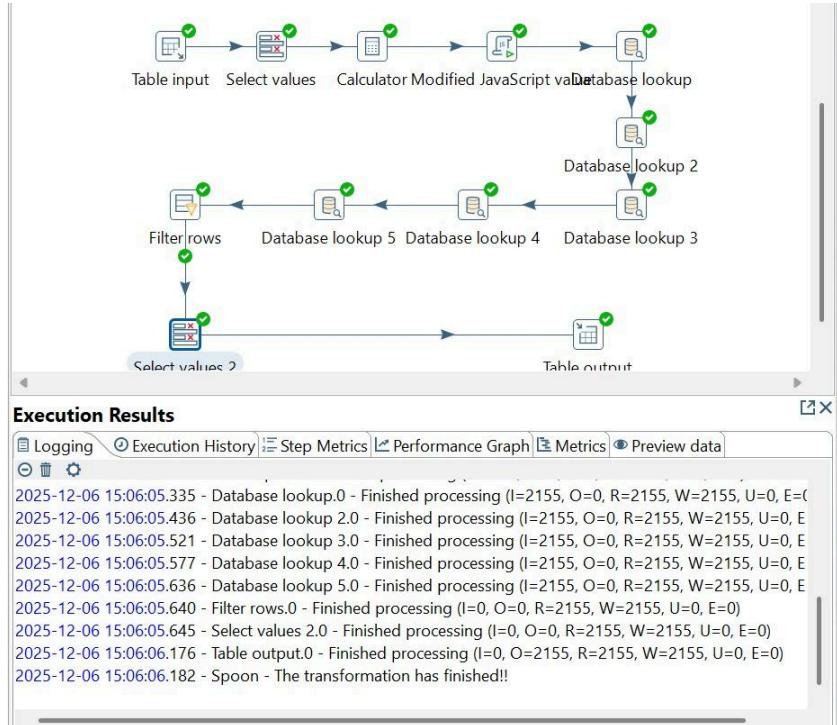
Field yang dihapus :



10. Konfigurasi table output untuk load ke OLAP dengan mengkoneksikan ke **dw_northwind.olap** dan lalu pilih target table dan target schema sesuai dengan yang dituju, serta pada database fields and klik get fields dan pastikan mapping sesuai dengan yang dituju lalu klik ok



11. Jalankan transformasi dengan klik tombol run dan pastikan semua step hijau dan log menunjukkan proses selesai tanpa eror



12. Buka database OLAP dan cek tabel fact_sales sudah terisi

	fact_id	order_id	customer_id	employee_id	product_id	shipper_id	time_id	unit_price	quantity	discount	total_sales
<input type="checkbox"/>	1	10248	VINET	5	11	3	20130704	14.00	12	0.00	168.00
<input type="checkbox"/>	2	10248	VINET	5	42	3	20130704	9.80	10	0.00	98.00
<input type="checkbox"/>	3	10248	VINET	5	72	3	20130704	34.80	5	0.00	174.00
<input type="checkbox"/>	4	10249	TOMSP	6	14	1	20130705	18.60	9	0.00	167.40
<input type="checkbox"/>	5	10249	TOMSP	6	51	1	20130705	42.40	40	0.00	1696.00
<input type="checkbox"/>	6	10250	HANAR	4	41	2	20130708	7.70	10	0.00	77.00
<input type="checkbox"/>	7	10250	HANAR	4	51	2	20130708	42.40	35	0.15	1261.40
<input type="checkbox"/>	8	10250	HANAR	4	65	2	20130708	16.80	15	0.15	214.20
<input type="checkbox"/>	9	10251	VICTE	3	22	1	20130708	16.80	6	0.05	95.76
<input type="checkbox"/>	10	10251	VICTE	3	57	1	20130708	15.60	15	0.05	222.30
<input type="checkbox"/>	11	10251	VICTE	3	65	1	20130708	16.80	20	0.00	336.00
<input type="checkbox"/>	12	10252	SUPRD	4	20	2	20130709	64.80	40	0.05	2462.40
<input type="checkbox"/>	13	10252	SUPRD	4	33	2	20130709	2.00	25	0.05	47.50
<input type="checkbox"/>	14	10252	SUPRD	4	60	2	20130709	27.20	40	0.00	1088.00
<input type="checkbox"/>	15	10253	HANAR	3	31	2	20130710	10.00	20	0.00	200.00
<input type="checkbox"/>	16	10253	HANAR	3	39	2	20130710	14.40	42	0.00	604.80
<input type="checkbox"/>	17	10253	HANAR	3	49	2	20130710	16.00	40	0.00	640.00
<input type="checkbox"/>	18	10254	CHOPS	5	24	2	20130711	3.60	15	0.15	45.90
<input type="checkbox"/>	19	10254	CHOPS	5	55	2	20130711	19.20	21	0.15	342.72
<input type="checkbox"/>	20	10254	CHOPS	5	74	2	20130711	8.00	21	0.00	168.00
<input type="checkbox"/>	21	10255	RICSU	9	2	3	20130712	15.20	20	0.00	304.00
<input type="checkbox"/>	22	10255	RICSU	9	16	3	20130712	13.90	35	0.00	486.50
<input type="checkbox"/>	23	10255	RICSU	9	36	3	20130712	15.20	25	0.00	380.00
<input type="checkbox"/>	24	10255	RICSU	9	59	3	20130712	44.00	30	0.00	1320.00
<input type="checkbox"/>	25	10256	WELLI	3	53	2	20130715	26.20	15	0.00	393.00

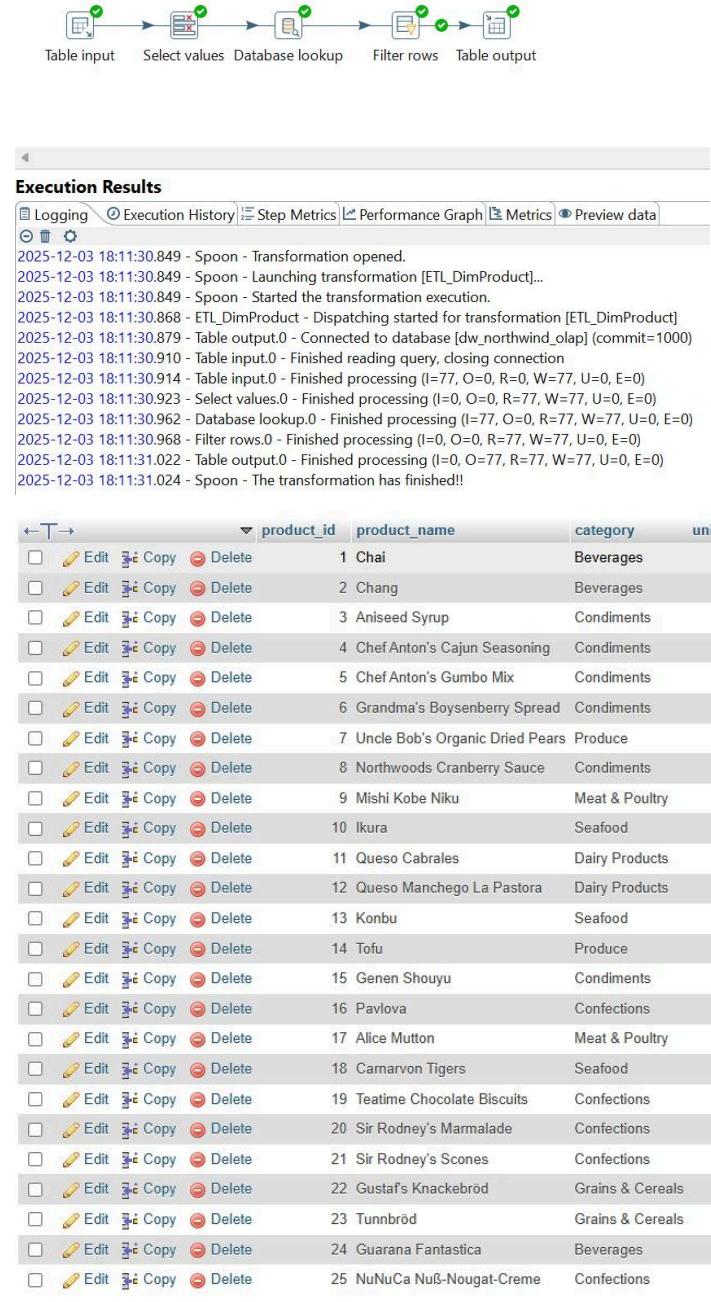
5.5 Hasil Eksekusi ETL

Pada tahap ini dilakukan eksekusi seluruh transformation ETL untuk memuat tabel dimensi dan tabel fakta pada data warehouse. Eksekusi dilakukan sesuai urutan pipeline, yaitu dimensi terlebih dahulu kemudian tabel fakta, sehingga foreign key pada fact_sales dapat tervalidasi terhadap seluruh dimensi, yaitu sebagai berikut :

1. Hasil Eksekusi Dim_Product

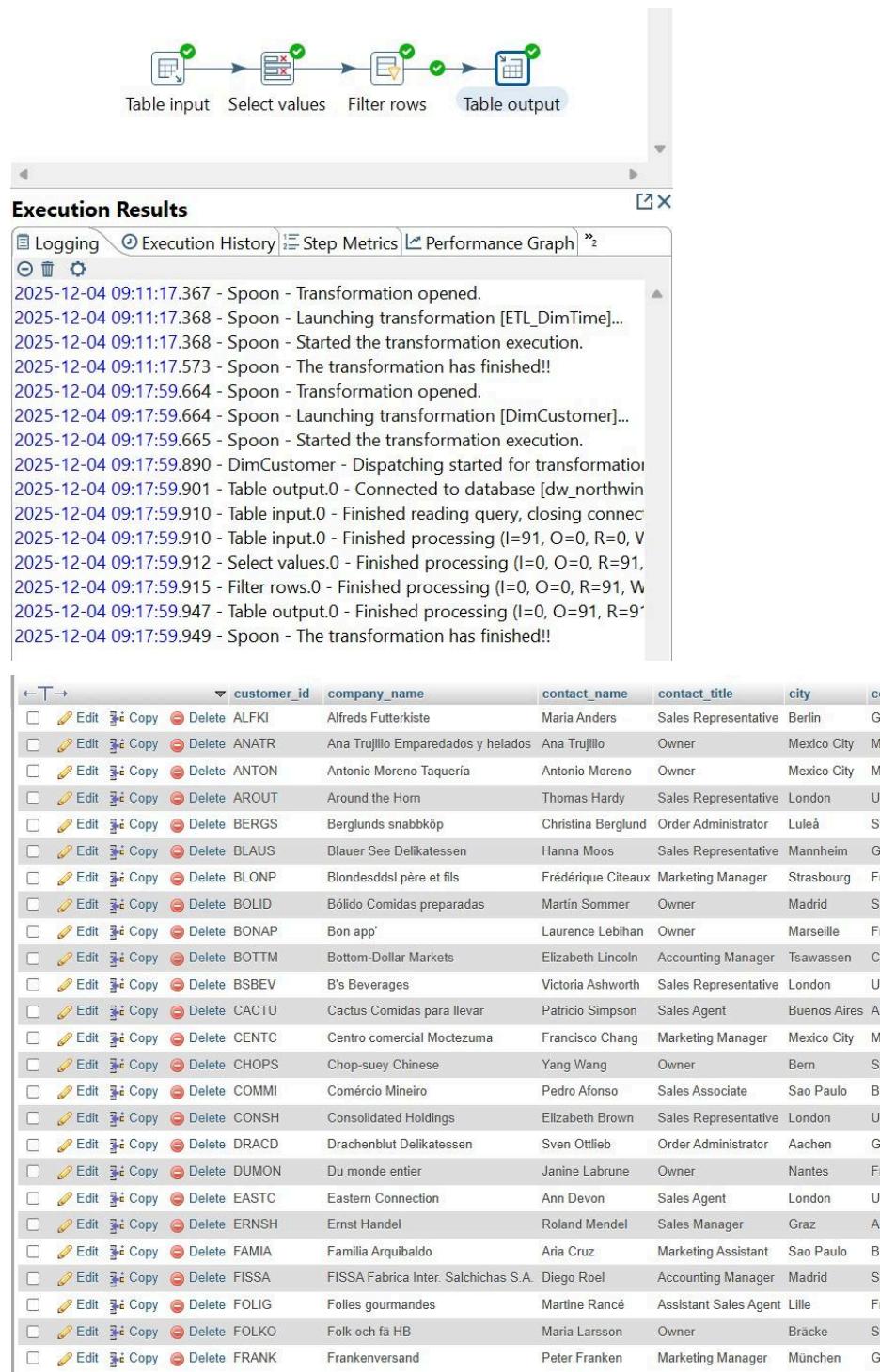
Eksekusi transformation dim_product berhasil dijalankan tanpa error. Seluruh step pada pipeline, mulai dari proses extract, transform, hingga load, berstatus sukses

yang ditandai dengan indikator hijau pada setiap step. Data produk berhasil dimuat ke tabel dim_product pada data warehouse dan dapat ditampilkan dengan lengkap, mencakup informasi identitas produk, nama produk, kategori, dan harga satuan.



2. Hasil Eksekusi Dim_Customer

Transformation dim_customer berhasil dijalankan dengan baik. Proses extract data pelanggan dari OLTP, transformasi struktur data, serta load ke tabel dim_customer pada OLAP berjalan tanpa kendala. Seluruh data pelanggan yang valid berhasil tersimpan dan siap digunakan sebagai dimensi dalam analisis OLAP.



3. Hasil Eksekusi Dim_Employee

Hasil eksekusi ETL pada dim_employee menunjukkan bahwa data karyawan berhasil diproses dan dimuat ke data warehouse. Tidak ditemukan error selama proses eksekusi, dan seluruh field yang diperlukan berhasil tersimpan sesuai dengan struktur tabel dimensi.



Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2025-12-03 18:51:48.796 - Spoon - Transformation opened.
2025-12-03 18:51:48.796 - Spoon - Launching transformation [ETL_DimEmployees]...
2025-12-03 18:51:48.803 - Spoon - Started the transformation execution.
2025-12-03 18:51:48.849 - ETL_DimEmployees - Dispatching started for transformation [ETL_DimEmployees]
2025-12-03 18:51:48.863 - Table output.0 - Connected to database [dw_northwind.olap] (commit=1000)
2025-12-03 18:51:48.885 - Table input.0 - Finished reading query, closing connection
2025-12-03 18:51:48.889 - Select values.0 - Finished processing (I=0, O=0, R=9, W=9, U=0, E=0)
2025-12-03 18:51:48.891 - Filter rows.0 - Finished processing (I=0, O=0, R=9, W=9, U=0, E=0)
2025-12-03 18:51:48.892 - Table input.0 - Finished processing (I=9, O=0, R=0, W=9, U=0, E=0)
2025-12-03 18:51:48.906 - Table output.0 - Finished processing (I=0, O=9, R=9, W=9, U=0, E=0)
2025-12-03 18:51:48.908 - Spoon - The transformation has finished!!

		employee_id	employee_name	title	city	country
<input type="checkbox"/>		1	Nancy Davolio	Sales Representative	New York	USA
<input type="checkbox"/>		2	Andrew Fuller	Vice President Sales	New York	USA
<input type="checkbox"/>		3	Janet Leverling	Sales Representative	New York	USA
<input type="checkbox"/>		4	Margaret Peacock	Sales Representative	New York	USA
<input type="checkbox"/>		5	Steven Buchanan	Sales Manager	London	UK
<input type="checkbox"/>		6	Michael Suyama	Sales Representative	London	UK
<input type="checkbox"/>		7	Robert King	Sales Representative	London	UK
<input type="checkbox"/>		8	Laura Callahan	Sales Manager	New York	USA
<input type="checkbox"/>		9	Anne Dodsworth	Sales Representative	London	UK

4. Hasil Eksekusi Dim_Shipper

Eksekusi pipeline dim_shipper berjalan dengan sukses. Data shipper berhasil dimuat ke tabel dim_shipper pada data warehouse dan telah melalui proses transformasi serta validasi data sehingga siap digunakan dalam analisis penjualan.



Execution Results

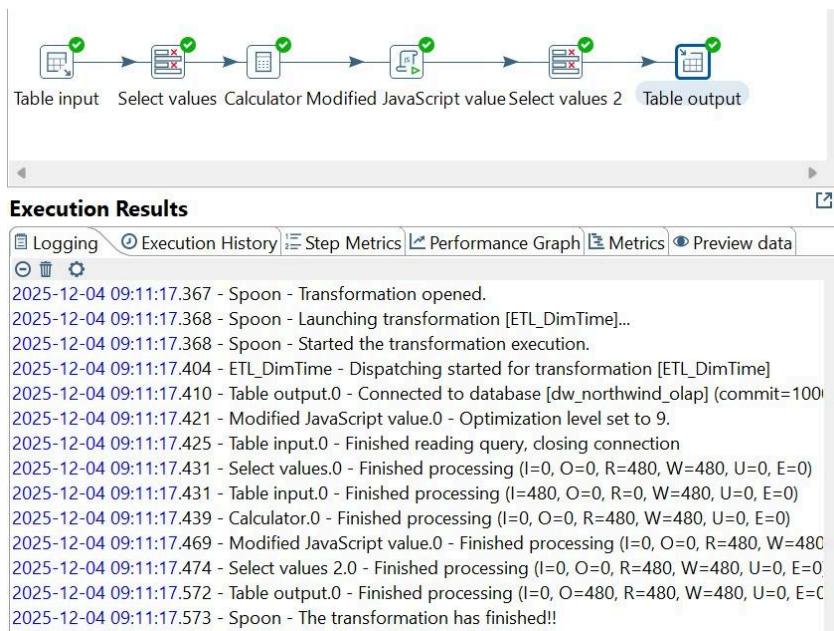
Logging Execution History Step Metrics Performance Graph Metrics Preview data

2025-12-03 18:34:18.992 - Spoon - Transformation opened.
2025-12-03 18:34:18.992 - Spoon - Launching transformation [ETL_DimShipper]...
2025-12-03 18:34:18.992 - Spoon - Started the transformation execution.
2025-12-03 18:34:19.029 - ETL_DimShipper - Dispatching started for transformation [ETL_DimShipper]
2025-12-03 18:34:19.036 - Table output.0 - Connected to database [dw_northwind.olap] (commit=1000)
2025-12-03 18:34:19.051 - Table input.0 - Finished reading query, closing connection
2025-12-03 18:34:19.054 - Select values.0 - Finished processing (I=0, O=0, R=3, W=3, U=0, E=0)
2025-12-03 18:34:19.054 - Table input.0 - Finished processing (I=3, O=0, R=0, W=3, U=0, E=0)
2025-12-03 18:34:19.056 - Filter rows.0 - Finished processing (I=0, O=0, R=3, W=3, U=0, E=0)
2025-12-03 18:34:19.068 - Table output.0 - Finished processing (I=0, O=3, R=3, W=3, U=0, E=0)
2025-12-03 18:34:19.074 - Spoon - The transformation has finished!!

		shipper_id	shipper_name
<input type="checkbox"/>	Edit Copy Delete	1	Speedy Express
<input type="checkbox"/>	Edit Copy Delete	2	United Package
<input type="checkbox"/>	Edit Copy Delete	3	Federal Shipping

5. Hasil Eksekusi Dim_Time

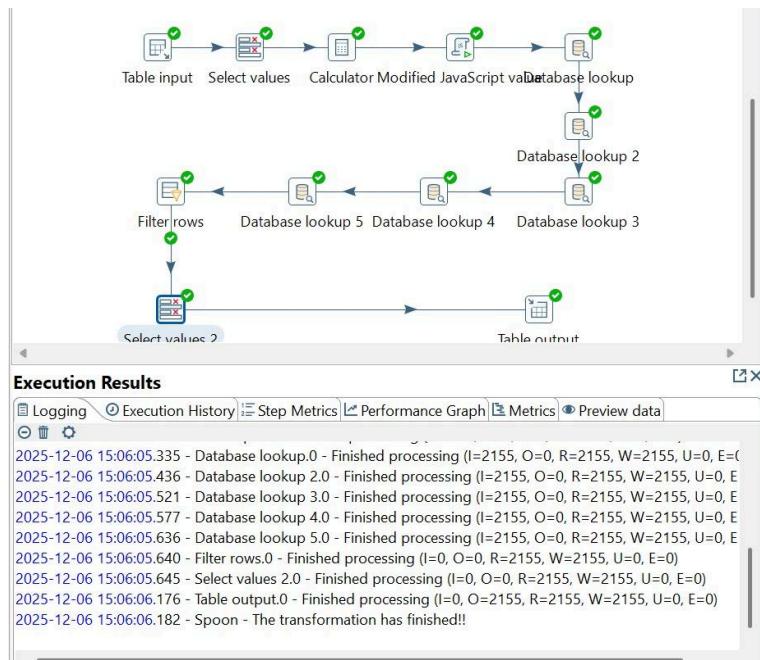
Transformation dim_time berhasil dijalankan dengan menghasilkan data dimensi waktu yang lengkap. Seluruh atribut waktu seperti tanggal, bulan, tahun, kuartal, dan hari berhasil dibentuk dan dimuat ke tabel dim_time. Data ini menjadi referensi utama dalam analisis berbasis waktu pada data warehouse.



	time_id	full_date	day	month	month_name	quarter	year	weekday
Delete	20130704	2013-07-04	4	7	July	3	2013	Thursday
Delete	20130705	2013-07-05	5	7	July	3	2013	Friday
Delete	20130708	2013-07-08	8	7	July	3	2013	Monday
Delete	20130709	2013-07-09	9	7	July	3	2013	Tuesday
Delete	20130710	2013-07-10	10	7	July	3	2013	Wednesday
Delete	20130711	2013-07-11	11	7	July	3	2013	Thursday
Delete	20130712	2013-07-12	12	7	July	3	2013	Friday
Delete	20130715	2013-07-15	15	7	July	3	2013	Monday
Delete	20130716	2013-07-16	16	7	July	3	2013	Tuesday
Delete	20130717	2013-07-17	17	7	July	3	2013	Wednesday
Delete	20130718	2013-07-18	18	7	July	3	2013	Thursday
Delete	20130719	2013-07-19	19	7	Juiy	3	2013	Friday
Delete	20130722	2013-07-22	22	7	July	3	2013	Monday
Delete	20130723	2013-07-23	23	7	July	3	2013	Tuesday
Delete	20130724	2013-07-24	24	7	July	3	2013	Wednesday
Delete	20130725	2013-07-25	25	7	July	3	2013	Thursday
Delete	20130726	2013-07-26	26	7	July	3	2013	Friday
Delete	20130729	2013-07-29	29	7	July	3	2013	Monday
Delete	20130730	2013-07-30	30	7	July	3	2013	Tuesday
Delete	20130731	2013-07-31	31	7	July	3	2013	Wednesday
Delete	20130801	2013-08-01	1	8	August	3	2013	Thursday
Delete	20130802	2013-08-02	2	8	August	3	2013	Friday
Delete	20130805	2013-08-05	5	8	August	3	2013	Monday
Delete	20130806	2013-08-06	6	8	August	3	2013	Tuesday
Delete	20130807	2013-08-07	7	8	August	3	2013	Wednesday

6. Hasil Eksekusi Fact_Sales

Hasil eksekusi ETL fact_sales menunjukkan bahwa data transaksi penjualan berhasil dimuat ke tabel fakta pada data warehouse. Proses perhitungan measure penjualan dan validasi foreign key terhadap seluruh tabel dimensi berjalan dengan baik. Seluruh data yang tersimpan memiliki keterkaitan yang valid dengan tabel dimensi sehingga siap digunakan dalam analisis OLAP.



	<input type="button" value="← T →"/>	<input type="button" value="▼ fact_id"/>	order_id	customer_id	employee_id	product_id	shipper_id	time_id	unit_price	quantity	discount	total_sales	
<input type="checkbox"/>				1	10248	VINET	5	11	3 20130704	14.00	12	0.00	168.00
<input type="checkbox"/>				2	10248	VINET	5	42	3 20130704	9.80	10	0.00	98.00
<input type="checkbox"/>				3	10248	VINET	5	72	3 20130704	34.80	5	0.00	174.00
<input type="checkbox"/>				4	10249	TOMSP	6	14	1 20130705	18.60	9	0.00	167.40
<input type="checkbox"/>				5	10249	TOMSP	6	51	1 20130705	42.40	40	0.00	1696.00
<input type="checkbox"/>				6	10250	HANAR	4	41	2 20130708	7.70	10	0.00	77.00
<input type="checkbox"/>				7	10250	HANAR	4	51	2 20130708	42.40	35	0.15	1261.40
<input type="checkbox"/>				8	10250	HANAR	4	65	2 20130708	16.80	15	0.15	214.20
<input type="checkbox"/>				9	10251	VICTE	3	22	1 20130708	16.80	6	0.05	95.76
<input type="checkbox"/>				10	10251	VICTE	3	57	1 20130708	15.60	15	0.05	222.30
<input type="checkbox"/>				11	10251	VICTE	3	65	1 20130708	16.80	20	0.00	336.00
<input type="checkbox"/>				12	10252	SUPRD	4	20	2 20130709	64.80	40	0.05	2462.40
<input type="checkbox"/>				13	10252	SUPRD	4	33	2 20130709	2.00	25	0.05	47.50
<input type="checkbox"/>				14	10252	SUPRD	4	60	2 20130709	27.20	40	0.00	1088.00
<input type="checkbox"/>				15	10253	HANAR	3	31	2 20130710	10.00	20	0.00	200.00
<input type="checkbox"/>				16	10253	HANAR	3	39	2 20130710	14.40	42	0.00	604.80
<input type="checkbox"/>				17	10253	HANAR	3	49	2 20130710	16.00	40	0.00	640.00
<input type="checkbox"/>				18	10254	CHOPS	5	24	2 20130711	3.60	15	0.15	45.90
<input type="checkbox"/>				19	10254	CHOPS	5	55	2 20130711	19.20	21	0.15	342.72
<input type="checkbox"/>				20	10254	CHOPS	5	74	2 20130711	8.00	21	0.00	168.00
<input type="checkbox"/>				21	10255	RICSU	9	2	3 20130712	15.20	20	0.00	304.00
<input type="checkbox"/>				22	10255	RICSU	9	16	3 20130712	13.90	35	0.00	486.50
<input type="checkbox"/>				23	10255	RICSU	9	36	3 20130712	15.20	25	0.00	380.00
<input type="checkbox"/>				24	10255	RICSU	9	59	3 20130712	44.00	30	0.00	1320.00
<input type="checkbox"/>				25	10256	WELLI	3	53	2 20130715	26.20	15	0.00	393.00

5.6 Penyimpanan File .etl

Pada proyek ini, hasil implementasi pipeline ETL disimpan dalam format file bawaan Pentaho Data Integration. File ETL terdiri dari :

- Transformation (.ktr), yaitu file yang berisi rangkaian step ETL untuk masing-masing tabel dimensi dan tabel fakta.

Seluruh file ETL disimpan dalam satu folder proyek agar mudah dikelola, dipindahkan, dan dieksekusi ulang (re-run). Penyimpanan file dalam struktur folder yang rapi juga memudahkan proses dokumentasi dan evaluasi hasil implementasi.

Daftar file ETL yang disimpan :

- dim_product.ktr
- dim_customer.ktr
- dim_employee.ktr
- dim_shipper.ktr
- dim_time.ktr
- fact_sales.ktr

BAB 6

HASIL ANALISIS DAN VISUALISASI

6.1 Data Warehouse Final

6.1.1 Tabel Dimensi

Tabel dimensi berfungsi sebagai data referensi untuk memberikan konteks terhadap data transaksi penjualan. Tabel dimensi digunakan untuk melakukan analisis berdasarkan atribut tertentu seperti produk, pelanggan, waktu, karyawan, dan shipper. Tabel dimensi yang digunakan pada data warehouse ini adalah :

1. dim_product

Menyimpan informasi produk seperti identitas produk, nama produk, kategori, dan harga satuan.

		product_id	product_name	category	unit_price
<input type="checkbox"/>	  	1	Chai	Beverages	18.00
<input type="checkbox"/>	  	2	Chang	Beverages	19.00
<input type="checkbox"/>	  	3	Aniseed Syrup	Condiments	10.00
<input type="checkbox"/>	  	4	Chef Anton's Cajun Seasoning	Condiments	22.00
<input type="checkbox"/>	  	5	Chef Anton's Gumbo Mix	Condiments	21.35
<input type="checkbox"/>	  	6	Grandma's Boysenberry Spread	Condiments	25.00
<input type="checkbox"/>	  	7	Uncle Bob's Organic Dried Pears	Produce	30.00
<input type="checkbox"/>	  	8	Northwoods Cranberry Sauce	Condiments	40.00
<input type="checkbox"/>	  	9	Mishi Kobe Niku	Meat & Poultry	97.00
<input type="checkbox"/>	  	10	Ikura	Seafood	31.00
<input type="checkbox"/>	  	11	Queso Cabrales	Dairy Products	21.00
<input type="checkbox"/>	  	12	Queso Manchego La Pastora	Dairy Products	38.00
<input type="checkbox"/>	  	13	Konbu	Seafood	6.00
<input type="checkbox"/>	  	14	Tofu	Produce	23.25
<input type="checkbox"/>	  	15	Genen Shouyu	Condiments	15.50
<input type="checkbox"/>	  	16	Pavlova	Confections	17.45
<input type="checkbox"/>	  	17	Alice Mutton	Meat & Poultry	39.00
<input type="checkbox"/>	  	18	Carmarvon Tigers	Seafood	62.50
<input type="checkbox"/>	  	19	Teatime Chocolate Biscuits	Confections	9.20
<input type="checkbox"/>	  	20	Sir Rodney's Marmalade	Confections	81.00
<input type="checkbox"/>	  	21	Sir Rodney's Scones	Confections	10.00
<input type="checkbox"/>	  	22	Gustaf's Knackebrod	Grains & Cereals	21.00
<input type="checkbox"/>	  	23	Tunnbröd	Grains & Cereals	9.00
<input type="checkbox"/>	  	24	Guarana Fantastica	Beverages	4.50
<input type="checkbox"/>	  	25	NuNuCa Nuß-Nougat-Creme	Confections	14.00

2. dim_customer

Menyimpan informasi pelanggan seperti nama perusahaan, kontak pelanggan, kota, dan negara.

	<input type="button" value="←"/> <input type="button" value="→"/>	<input type="button" value="customer_id"/>	company_name	contact_name	contact_title	city	country
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	ALFKI	Alfreds Futterkiste	Maria Anders	Sales Representative Berlin Germany
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	ANATR	Ana Trujillo Emparedados y helados	Ana Trujillo	Owner Mexico City Mexico
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	ANTON	Antonio Moreno Taquería	Antonio Moreno	Owner Mexico City Mexico
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	AROUT	Around the Horn	Thomas Hardy	Sales Representative London UK
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	BERGS	Berglunds snabbköp	Christina Berglund	Order Administrator Luleå Sweden
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	BLAUS	Blauer See Delikatessen	Hanna Moos	Sales Representative Mannheim Germany
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	BLONP	Blondesdösl père et fils	Frédérique Citeaux	Marketing Manager Strasbourg France
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	BOLID	Bólido Comidas preparadas	Martín Sommer	Owner Madrid Spain
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	BONAP	Bon app'	Laurence Lebihan	Owner Marseille France
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	BOTTM	Bottom-Dollar Markets	Elizabeth Lincoln	Accounting Manager Tsawassen Canada
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	BSBEV	B's Beverages	Victoria Ashworth	Sales Representative London UK
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	CACTU	Cactus Comidas para llevar	Patricia Simpson	Sales Agent Buenos Aires Argentina
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	CENTC	Centro comercial Moctezuma	Francisco Chang	Marketing Manager Mexico City Mexico
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	CHOPS	Chop-suey Chinese	Yang Wang	Owner Bern Switzerland
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	COMMI	Comércio Mineiro	Pedro Afonso	Sales Associate São Paulo Brazil
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	CONSH	Consolidated Holdings	Elizabeth Brown	Sales Representative London UK
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	DRACD	Drachenblut Delikatessen	Sven Ottlieb	Order Administrator Aachen Germany
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	DUMON	Du monde entier	Janine Labrune	Owner Nantes France
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	EASTC	Eastern Connection	Ann Devon	Sales Agent London UK
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	ERNSH	Ernst Handel	Roland Mendel	Sales Manager Graz Austria
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	FAMIA	Familia Arquibaldo	Aria Cruz	Marketing Assistant São Paulo Brazil
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	FISSA	FISSA Fabrica Inter. Salchichas S.A.	Diego Roel	Accounting Manager Madrid Spain
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	FOLIG	Folies gourmandes	Martine Rancé	Assistant Sales Agent Lille France
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	FOLKO	Folk och fa HB	Maria Larsson	Owner Bräcke Sweden
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	FRANK	Frankenversand	Peter Franken	Marketing Manager München Germany

3. dim_employee

Menyimpan informasi karyawan yang menangani transaksi, seperti nama karyawan, jabatan, dan lokasi.

	<input type="button" value="←"/> <input type="button" value="→"/>	<input type="button" value="employee_id"/>	employee_name	title	city	country
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	1 Nancy Davolio	Sales Representative	New York USA
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	2 Andrew Fuller	Vice President Sales	New York USA
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	3 Janet Leverling	Sales Representative	New York USA
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	4 Margaret Peacock	Sales Representative	New York USA
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	5 Steven Buchanan	Sales Manager	London UK
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	6 Michael Suyama	Sales Representative	London UK
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	7 Robert King	Sales Representative	London UK
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	8 Laura Callahan	Sales Manager	New York USA
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	9 Anne Dodsworth	Sales Representative	London UK

4. dim_shipper

Menyimpan informasi perusahaan pengiriman (shipper), seperti identitas dan nama shipper.

	<input type="button" value="←"/> <input type="button" value="→"/>	<input type="button" value="shipper_id"/>	shipper_name	
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	1 Speedy Express
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	2 United Package
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	3 Federal Shipping

5. dim_time

Menyimpan informasi waktu (tanggal, hari, bulan, tahun, kuartal, nama bulan, dan nama hari) untuk mendukung analisis berbasis waktu.

	time_id	full_date	day	month	month_name	quarter	year	weekday
Delete	20130704	2013-07-04	4	7	July	3	2013	Thursday
Delete	20130705	2013-07-05	5	7	July	3	2013	Friday
Delete	20130708	2013-07-08	8	7	July	3	2013	Monday
Delete	20130709	2013-07-09	9	7	July	3	2013	Tuesday
Delete	20130710	2013-07-10	10	7	July	3	2013	Wednesday
Delete	20130711	2013-07-11	11	7	July	3	2013	Thursday
Delete	20130712	2013-07-12	12	7	July	3	2013	Friday
Delete	20130715	2013-07-15	15	7	July	3	2013	Monday
Delete	20130716	2013-07-16	16	7	July	3	2013	Tuesday
Delete	20130717	2013-07-17	17	7	July	3	2013	Wednesday
Delete	20130718	2013-07-18	18	7	July	3	2013	Thursday
Delete	20130719	2013-07-19	19	7	July	3	2013	Friday
Delete	20130722	2013-07-22	22	7	July	3	2013	Monday
Delete	20130723	2013-07-23	23	7	July	3	2013	Tuesday
Delete	20130724	2013-07-24	24	7	July	3	2013	Wednesday
Delete	20130725	2013-07-25	25	7	July	3	2013	Thursday
Delete	20130726	2013-07-26	26	7	July	3	2013	Friday
Delete	20130729	2013-07-29	29	7	July	3	2013	Monday
Delete	20130730	2013-07-30	30	7	July	3	2013	Tuesday
Delete	20130731	2013-07-31	31	7	July	3	2013	Wednesday
Delete	20130801	2013-08-01	1	8	August	3	2013	Thursday
Delete	20130802	2013-08-02	2	8	August	3	2013	Friday
Delete	20130805	2013-08-05	5	8	August	3	2013	Monday
Delete	20130806	2013-08-06	6	8	August	3	2013	Tuesday
Delete	20130807	2013-08-07	7	8	August	3	2013	Wednesday

Tabel-tabel dimensi tersebut terhubung dengan tabel fakta melalui atribut key sehingga membentuk struktur star schema yang memudahkan analisis OLAP.

6.1.2 Tabel Fakta

Tabel fakta pada data warehouse ini adalah **fact_sales**. Tabel ini menyimpan data transaksi penjualan sebagai pusat analisis. Setiap baris pada tabel fact_sales merepresentasikan **satu produk pada satu transaksi pemesanan (order detail)**.

Tabel fact_sales terdiri dari dua jenis atribut utama yaitu :

- **Foreign Key** yang menghubungkan ke tabel dimensi, yaitu: order_id, customer_id, employee_id, product_id, shipper_id, dan time_id.
- **Measure** yang digunakan untuk perhitungan analisis, yaitu: unit_price, quantity, discount, dan total_sales.

Dengan struktur tersebut, fact_sales memungkinkan analisis penjualan berdasarkan dimensi waktu, produk, pelanggan, karyawan, serta shipper.

	<input type="button" value="← T →"/>	<input type="button" value="▼"/>	fact_id	order_id	customer_id	employee_id	product_id	shipper_id	time_id	unit_price	quantity	discount	total_sales	
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	1	10248	VINET	5	11	3	20130704	14.00	12	0.00	168.00
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	2	10248	VINET	5	42	3	20130704	9.80	10	0.00	98.00
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	3	10248	VINET	5	72	3	20130704	34.80	5	0.00	174.00
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	4	10249	TOMSP	6	14	1	20130705	18.60	9	0.00	167.40
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	5	10249	TOMSP	6	51	1	20130705	42.40	40	0.00	1696.00
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	6	10250	HANAR	4	41	2	20130708	7.70	10	0.00	77.00
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	7	10250	HANAR	4	51	2	20130708	42.40	35	0.15	1261.40
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	8	10250	HANAR	4	65	2	20130708	16.80	15	0.15	214.20
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	9	10251	VICTE	3	22	1	20130708	16.80	6	0.05	95.76
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	10	10251	VICTE	3	57	1	20130708	15.60	15	0.05	222.30
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	11	10251	VICTE	3	65	1	20130708	16.80	20	0.00	336.00
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	12	10252	SUPRD	4	20	2	20130709	64.80	40	0.05	2462.40
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	13	10252	SUPRD	4	33	2	20130709	2.00	25	0.05	47.50
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	14	10252	SUPRD	4	60	2	20130709	27.20	40	0.00	1088.00
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	15	10253	HANAR	3	31	2	20130710	10.00	20	0.00	200.00
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	16	10253	HANAR	3	39	2	20130710	14.40	42	0.00	604.80
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	17	10253	HANAR	3	49	2	20130710	16.00	40	0.00	640.00
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	18	10254	CHOPS	5	24	2	20130711	3.60	15	0.15	45.90
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	19	10254	CHOPS	5	55	2	20130711	19.20	21	0.15	342.72
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	20	10254	CHOPS	5	74	2	20130711	8.00	21	0.00	168.00
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	21	10255	RICSU	9	2	3	20130712	15.20	20	0.00	304.00
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	22	10255	RICSU	9	16	3	20130712	13.90	35	0.00	486.50
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	23	10255	RICSU	9	36	3	20130712	15.20	25	0.00	380.00
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	24	10255	RICSU	9	59	3	20130712	44.00	30	0.00	1320.00
<input type="checkbox"/>	<input type="button" value="Edit"/>	<input type="button" value="Copy"/>	<input type="button" value="Delete"/>	25	10256	WELLI	3	53	2	20130715	26.20	15	0.00	393.00

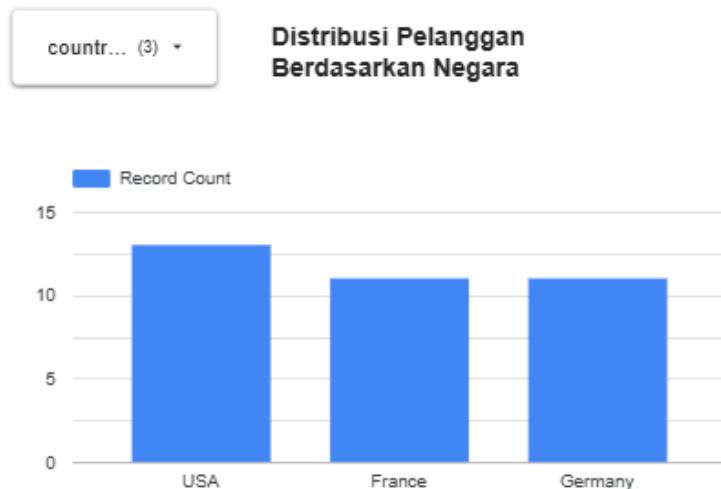
6.1.3 ANALISIS KPI

KPI	NILAI	Implikasi Strategis
Konsentrasi Pasar Inti	USA, France, Germany adalah 3 negara teratas (Total sekitar 36% dari basis pelanggan 91)	Mengonfirmasi tiga negara ini sebagai fokus utama untuk investasi dan dukungan penjualan lokal.
Kebutuhan Jangkauan Luas	67% pelanggan berada di 'Lainnya' (kota tersebar)	Menuntut adopsi strategi pemasaran digital yang fleksibel dan tidak terbatas pada geografi ketat untuk menjangkau mayoritas pasar.
Fokus Regional	London, Mexico City, São Paulo sebagai kota utama	Membutuhkan inisiatif regional yang ditargetkan (misalnya, roadshow atau acara lokal) untuk memanfaatkan kepadatan yang ada.

6.1.4 Visualisasi

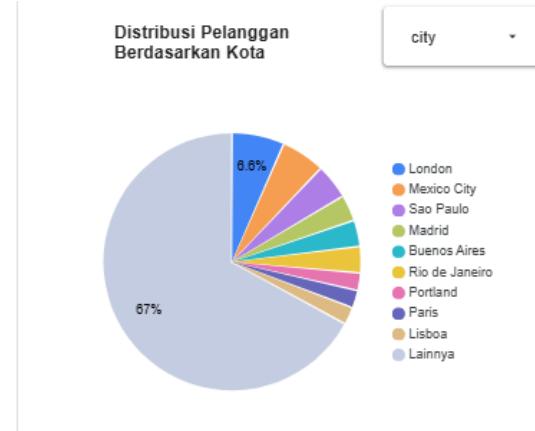
Dashboard Pelanggan ini berfungsi sebagai peta strategis untuk mengalokasikan sumber daya dan mengukur jangkauan pasar. Analisis studi kasus berfokus pada tiga komponen utama: **prioritas pasar nasional, kepadatan pelanggan perkotaan, dan aktivasi kontak operasional.**

1. Analisis Prioritas Pasar Nasional



Grafik batang menunjukkan bahwa basis pelanggan saat ini terkonsentrasi di tiga pasar inti: Amerika Serikat (USA) memimpin, diikuti oleh France dan Germany yang memiliki volume pelanggan yang identik. Konsentrasi tinggi di ketiga negara ini mengindikasikan bahwa mereka harus diklasifikasikan sebagai pasar Tingkat 1 dan menerima prioritas tertinggi dalam alokasi sumber daya. Tim Penjualan dan Pemasaran harus mengalokasikan Manajer Akun senior dan menyesuaikan materi kampanye agar sesuai dengan bahasa dan konteks bisnis ketiga negara ini untuk memaksimalkan hasil dari konsentrasi pelanggan yang ada.

2. Analisis Kepadatan Pelanggan Perkotaan



Analisis diagram lingkaran mengungkapkan wawasan penting tentang pola distribusi: meskipun kota-kota besar seperti London, Mexico City, dan São Paulo menyumbang persentase signifikan, 67% pelanggan tersebar di berbagai lokasi yang dikelompokkan ke dalam kategori 'Lainnya'. Distribusi yang sangat luas ini menunjukkan bahwa strategi *go-to-market* tidak boleh semata-mata bergantung pada interaksi tatap muka di pusat kota metropolitan. Sebaliknya, strategi pemasaran digital yang luas dan fleksibel sangat diperlukan untuk menjangkau mayoritas pelanggan yang tersebar luas, sambil tetap mengadakan kegiatan atau *event* yang ditargetkan di kota-kota utama yang disorot oleh diagram lingkaran.

3. Pemanfaatan data operasional dan kontak

Rincian Data Kontak						
contact_name	company_name	contact_title	country	city	country	
1. Ana Trujillo	Ana Trujillo Emparedados y Helados	Owner	Mexico	Mexico City	1	
2. Ann Devon	Eastern Connection	Sales Agent	UK	London	1	
3. Antonio Moreno	Antonio Moreno Taquería	Owner	Mexico	Mexico City	1	
4. Elizabeth Brown	Consolidated Holdings	Sales Representative	UK	London	1	
5. Francisco Chang	Centro comercial Moctezuma	Marketing Manager	Mexico	Mexico City	1	
6. Guillermo Fernández	Pericles Comidas clásicas	Sales Representative	Mexico	Mexico City	1	
7. Hari Kumar	Seven Seas Imports	Sales Manager	UK	London	1	
8. Miguel Angel Paolino	Tortuga Restaurante	Owner	Mexico	Mexico City	1	
9. Simon Crowther	North/South	Sales Associate	UK	London	1	
10. Thomas Hardy	Around the Horn	Sales Representative	UK	London	1	

Tabel detail di bagian bawah adalah alat operasional yang penting, memberikan rincian untuk aktivitas *follow-up* harian dan segmentasi lanjutan. Data ini memungkinkan tim Pemasaran untuk memfilter kontak secara cepat, misalnya, menarik daftar semua 'Owner' untuk komunikasi tingkat eksekutif, atau memfilter semua 'Sales Manager' di Brazil untuk penawaran kemitraan. Dengan menghubungkan nama kontak dan jabatan dengan lokasi geografis (Negara dan Kota), tim dapat memastikan bahwa setiap interaksi (email, panggilan, atau pertemuan) sangat relevan dan disesuaikan dengan konteks profesional dan lokasi pelanggan.

BAB 7

KESIMPULAN DAN SARAN

7.1 Kesimpulan

Berdasarkan implementasi Rancang Bangun ETL dan Data Warehouse menggunakan *Dataset Retail Kaggle* (Northwind Traders), dapat disimpulkan bahwa proyek ini telah berhasil merancang dan mengimplementasikan Data Warehouse berbasis Star Schema sederhana. Struktur yang digunakan terdiri dari satu Tabel Fakta utama, yaitu fact_sales, serta lima Tabel Dimensi pendukung, meliputi dim_customer, dim_employee, dim_product, dim_shipper, dan dim_time, dengan fact_sales sebagai pusat analisis yang memuat metrik penjualan seperti total_sales dan quantity. Selain itu, tahapan ETL (*Extract, Transform, Load*) juga telah berhasil diimplementasikan sesuai rencana menggunakan Pentaho Data Integration (PDI/Spoon). Proses *Transformasi* mencakup *Data Cleaning*, pembentukan atribut turunan (time_id, total_sales), dan *Integrasi* untuk memvalidasi keterkaitan *Foreign Key*. Sebagai hasil akhirnya, Data Warehouse yang terbentuk bersifat bersih, konsisten, dan granular, sehingga siap mendukung analisis KPI, termasuk identifikasi pasar inti (USA, France, Germany) dan analisis kepadatan pelanggan perkotaan (London, Mexico City, São Paulo), dengan fleksibilitas agregasi yang dilakukan pada tahap *query* atau *dashboard*.

7.2 Saran

Untuk pengembangan proyek *Business Intelligence* ini di masa depan, terdapat beberapa saran yang dapat dipertimbangkan, di antaranya adalah penggunaan Slowly Changing Dimensions (SCD) Tipe 2 untuk dimensi-dimensi kritis seperti dim_product atau dim_customer, mengingat dimensi saat ini dimuat sebagai Tipe 1. Adopsi SCD Tipe 2 ini penting agar sistem dapat melacak dan menganalisis data historis secara akurat. Selain itu, disarankan pula untuk membangun Job (.kjb) di Pentaho untuk mengotomatisasi urutan eksekusi seluruh *Transformation* (.ktr), yang bertujuan untuk mempermudah proses eksekusi ulang (*re-run*) dan memastikan urutan *load* dimensi ke fakta selalu benar. Terakhir, mengingat temuan adanya konsentrasi pasar di kota-kota tertentu, visualisasi dapat ditingkatkan dengan menggunakan Peta (Map Chart) untuk menampilkan kepadatan penjualan secara visual di kota-kota fokus (London, Mexico City, São Paulo) dan membandingkannya dengan kategori 'Lainnya' yang tersebar luas.