LAPORAN PRAKTIKUM DATA WAREHOUSE



Disusun Oleh:

Aldo Febriansyah	2341760146
Bayu Triwibowo	2341760073
Gegas Anugrah Derajat	2341760140
M Hamdan Ubaidilah	2341760190
Qusnul Diah Mawanti	2341760035

PROGRAM STUDI D-IV SISTEM INFORMASI BISNIS

JURUSAN TEKNOLOGI INFORMASI

POLITEKNIK NEGERI MALANG

2025



Jurusan Teknologi Informasi – Politeknik Negeri

Malang

Mata Kuliah: Data Warehouse

Kelas: 2A – SIB Laporan: UAS

Juni 2025

A. Pengertian dan Tujuan ETL

ETL adalah proses **Extract – Transform – Load**, yaitu teknik untuk mengambil data dari sumber (Extract), membersihkan dan mengubah sesuai kebutuhan (Transform), lalu menyimpannya ke dalam database warehouse (Load).

Referensi: Kimball, Ralph. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 2013.

Tujuan ETL

- Mengintegrasikan data dari berbagai sumber.
- Meningkatkan kualitas data (konsistensi dan kebersihan).
- Menyediakan data siap pakai untuk keperluan analitik dan pelaporan.

B. Studi Kasus Pengelolaan Data

Studi kasus dipilih dari berbagai sumber terbuka seperti <u>Kaggle</u>, <u>data.gov</u>, dan <u>mockaroo.com</u> dengan kriteria ketersediaan atribut yang cukup, mudah diolah, dan dapat digunakan untuk simulasi ETL dan warehouse.

No	Kasus Data	Atribut Utama	
1	Data Karyawan	ID karyawan, nama, jabatan, departemen, gaji, tanggal masuk	
2	Data Penjualan	ID transaksi, tanggal, produk, jumlah, harga, pelanggan	
3	Data Inventaris	ID barang, nama barang, kategori, jumlah stok, lokasi gudang	
4	Data Pelanggan	ID pelanggan, nama, alamat, histori pembelian, kategori pelanggan	

Studi Kasus Dipilih:

Pengelolaan Data Karyawan & Kinerjanya

Alasan Pemilihan:

- Atribut yang tersedia cukup lengkap: identitas karyawan, jabatan, tanggal masuk, performa.
- Relevan untuk pembangunan data mart HR dan analisis KPI.
- Sesuai untuk perancangan skema bintang (star schema).
- Dapat digunakan untuk membuat laporan historis dan prediktif performa pegawai.

C. Data Dummy yang Digunakan

Sumber Data

Dataset asli diunduh dari Kaggle Employee Performance and Productivity Data

Penyesuaian

Dataset diubah dan disesuaikan, kemudian disimpan dengan nama **employee_db.csv** untuk digunakan dalam proses ETL.

Format File

CSV (Comma Separated Values)

Atribut Utama yang Digunakan:

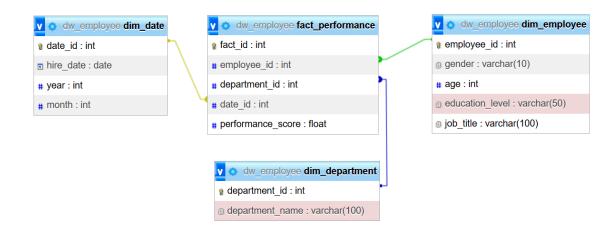
- o employee_id
- o employee_name
- o hire date
- position
- department_name
- o performance score

Keterangan

Dataset ini mencakup data produktivitas, divisi kerja, dan karakteristik karyawan yang dapat dimanfaatkan untuk membangun *data warehouse* dan analisis performa.

D. Rancangan Skema Bintang

Skema bintang adalah model data dalam data warehouse yang menyederhanakan hubungan antara **tabel fakta** dan **tabel dimensi**. Pada kasus ini, kita merancang data warehouse untuk **analisis performa karyawan**, dengan satu tabel fakta utama dan tiga tabel dimensi. Tiga tabel dimensi (dim_employee, dim_department, dim_date) terhubung langsung ke fact_performance menggunakan **foreign key**.



a. Tabel Fakta: fact_performance

Tabel ini mencatat **hasil evaluasi kinerja karyawan** dan merekamnya bersama informasi waktu dan unit organisasi.

Atribut	Tipe Data	Keterangan	
fact_id	INT (PK)	Primary Key (otomatis), sebagai identifikasi unik setiap record.	
employee_id	INT (FK)	Foreign Key ke dim_employee, mengacu ke data karyawan.	
department_id	INT (FK)	Foreign Key ke dim_department, mengacu ke departemen.	
date_id	INT (FK)	Foreign Key ke dim_date, mengacu ke waktu penilaian.	
performance_score	FLOAT	Nilai penilaian kinerja karyawan, bisa dalam skala 0–100.	

b. Tabel Dimensi

• dim_employee

Dimensi ini menjelaskan **profil karyawan**, yang digunakan untuk analisis berdasarkan karakteristik individu.

Atribut	Tipe Data	Keterangan	
fact_id	INT (PK)	Primary Key (otomatis), sebagai identifikasi unik setiap record.	
employee_id	INT (FK)	Foreign Key ke dim_employee, mengacu ke data karyawan.	
department_id	INT (FK)	Foreign Key ke dim_department, mengacu ke departemen.	
date_id	INT (FK)	Foreign Key ke dim_date, mengacu ke waktu penilaian.	
performance_score	FLOAT	Nilai penilaian kinerja karyawan, bisa dalam skala 0–100.	

• dim_department

Digunakan untuk mengetahui **divisi atau unit kerja** karyawan berada, serta menganalisis performa berdasarkan departemen.

Atribut	Tipe Data	Keterangan
department_id	INT (PK)	Primary Key, ID unik untuk departemen
department_name	VARCHAR(100)	Nama departemen

• dim_date

Digunakan untuk **analisis berbasis waktu**, seperti tren kinerja tahunan atau bulanan.

Atribut	Tipe Data	Keterangan	
date_id	INT (PK)	ID unik tanggal	
hire_date	DATE	Tanggal masuk kerja (optional)	
year	INT	Tahun evaluasi/peristiwa	
month	INT	Bulan evaluasi/peristiwa	

E. Implementasi Proyek ETL

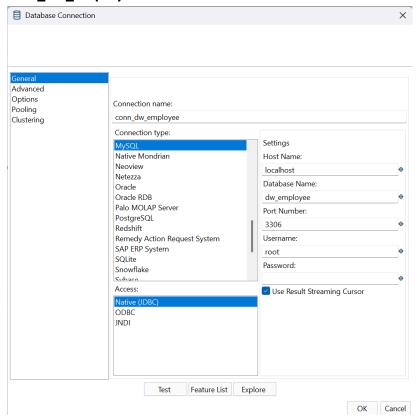
1. Tool:

• Pentaho Data Integration (PDI) / Spoon

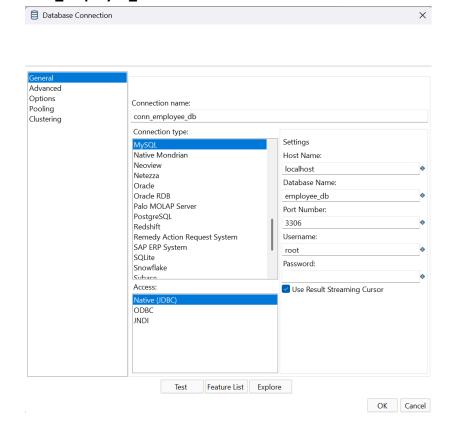
• Format file pipeline: .ktr

2. Connection

• conn_dw_employee



• conn_employee_db



3. Langkah - Langkah ETL

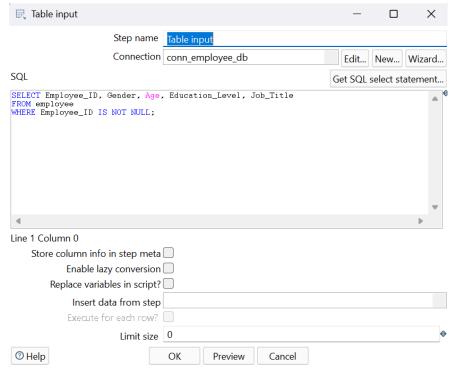
a. dim_employee

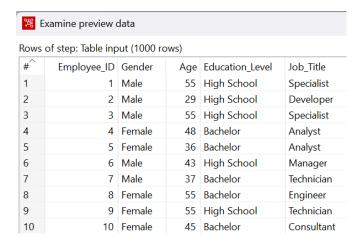




table input

Langkah ini digunakan untuk mengambil data dari database sumber employee_db. Kolom yang diambil meliputi employee_id, gender, age, education_level, dan job_title. Data ini merupakan data mentah yang akan diproses lebih lanjut.





select values

Digunakan untuk memilih hanya kolom yang dibutuhkan, serta mengganti nama kolom agar sesuai dengan struktur tabel dim_employee. Kolom-kolom yang tidak relevan akan dihapus pada tahap in

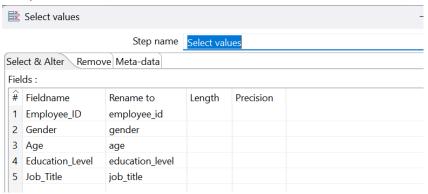
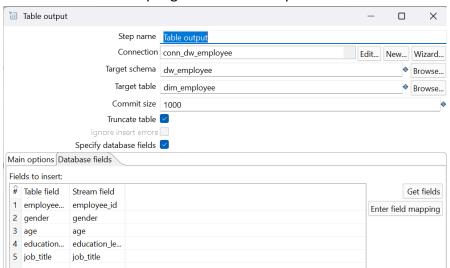
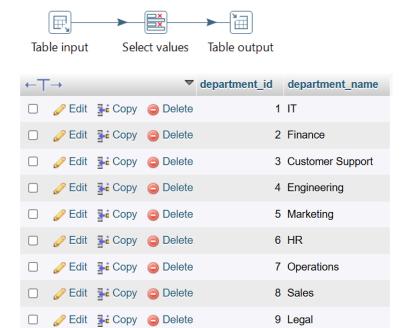


table output

Menyimpan hasil transformasi ke dalam tabel dim_employee di data warehouse dw_employee. Proses ini memastikan bahwa data telah sesuai struktur dan siap digunakan untuk keperluan analisis.

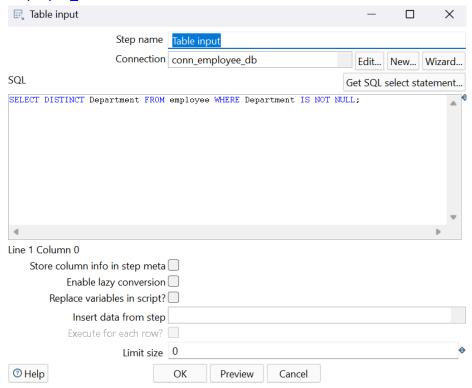


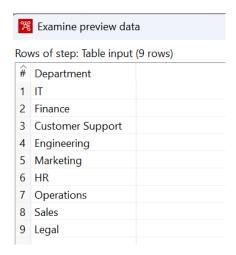
b. dim_department



• table Input

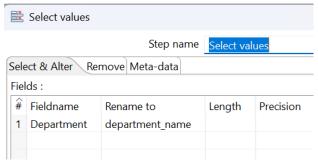
Langkah ini mengambil data departemen dari data sumber, yang besar berasal dari kolom Department yang ada dari database sumber employee_db.





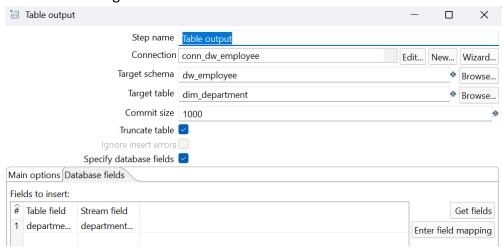
select values

Tahap ini melakukan rename pada kolom Department menjadi department_name.

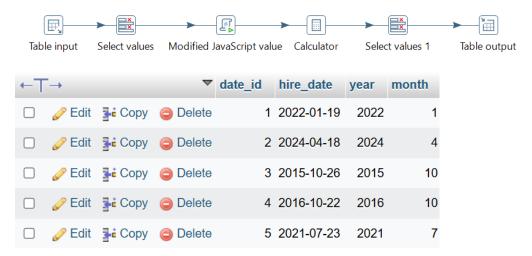


• table output

Data hasil transformasi disimpan ke dalam tabel dim_department pada database dw_employee, memastikan referensi departemen tersedia untuk digunakan oleh tabel fakta.

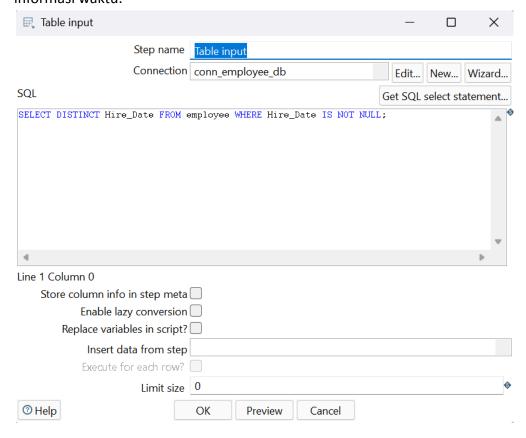


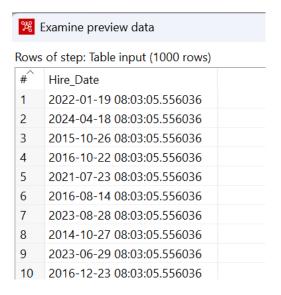
c. dim_date



• table input

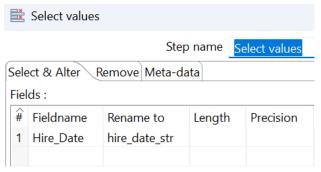
Mengambil data tanggal pada Hire_Date dari database yang berisi informasi waktu.





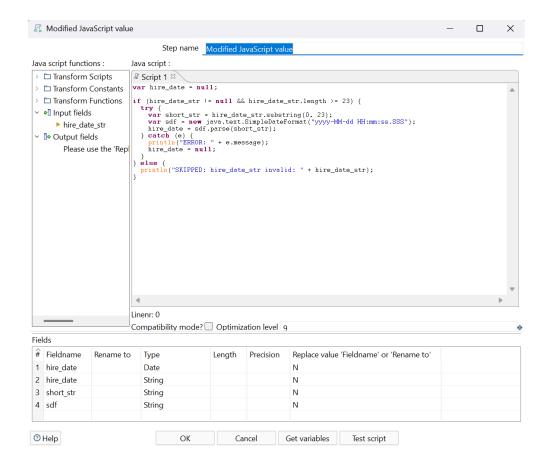
select values

Langkah ini digunakan untuk mengambil kolom Hire_Date dari data sumber dan mengganti namanya menjadi hire_date_str. Penggantian nama ini dilakukan agar data tanggal yang awalnya berupa tipe bawaan dapat diproses sebagai string pada langkah selanjutnya. Hal ini penting karena transformasi menggunakan JavaScript membutuhkan data dalam format string untuk parsing manual.



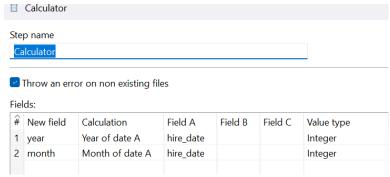
modified javaScript value

Pada langkah ini, kolom hire_date_str yang sebelumnya sudah disiapkan akan diproses dengan skrip JavaScript untuk mengkonversinya menjadi tanggal dengan format yyyy-MM-dd HH:mm:ss.SSS.



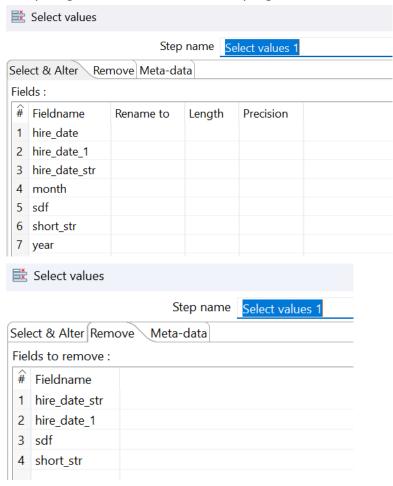
calculator

Langkah Calculator digunakan untuk menghasilkan kolom year dan month dari field hire_date. Fungsi ini secara otomatis mengekstrak tahun dan bulan dari data tanggal untuk melengkapi atribut pada tabel dim_date.



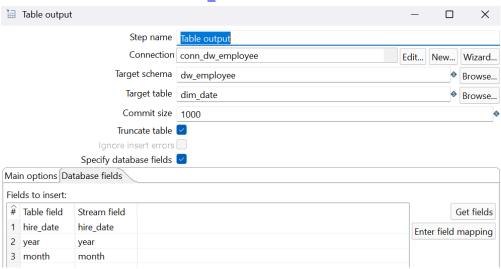
• select values 1

Menyaring dan memilih kolom akhir yang akan dimuat ke tabel.

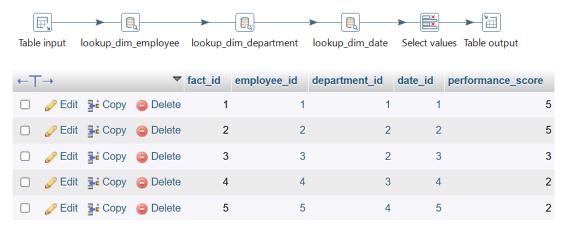


• table output

Memuat hasil akhir ke tabel dim_date

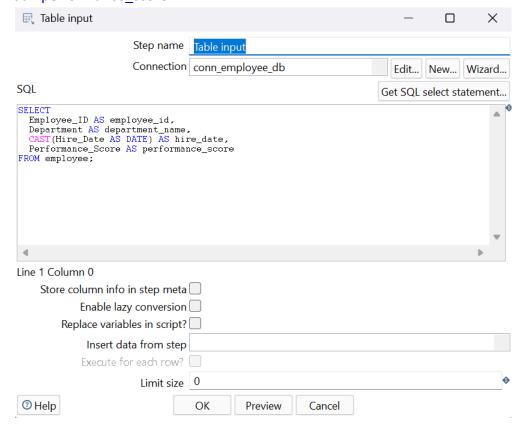


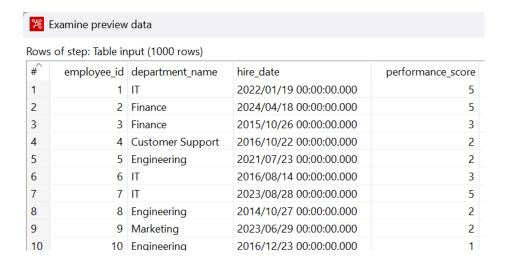
d. fact_performance



• table input

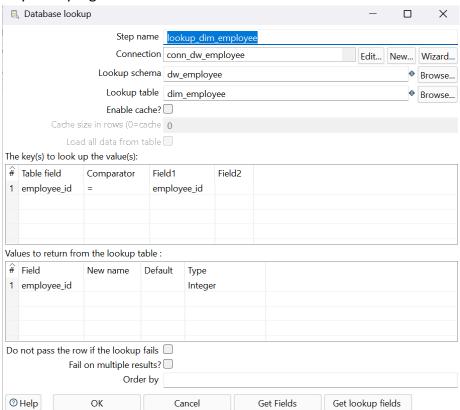
Membaca data performa karyawan dari data sumber seperti. Kolom yang dibaca mencakup employee_id, department_name, hire_date, dan performance score.





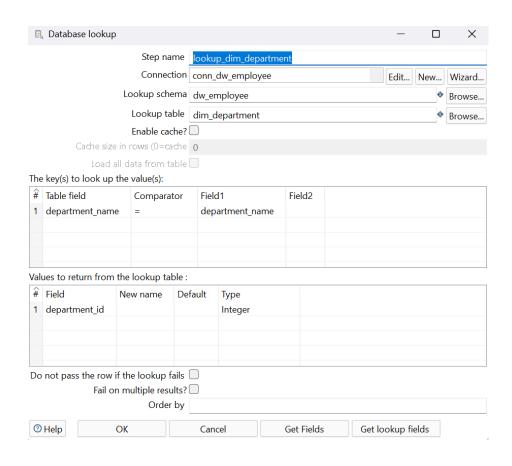
lookup_dim_employee

Mencari employee_id dari dim_employee untuk memperoleh ID karyawan yang sudah dimuat di dimensi.



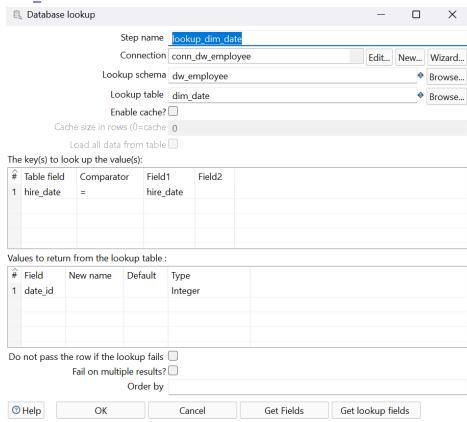
lookup_dim_department

Melakukan pencocokan department_name dengan tabel dim_department untuk mendapatkan department_id.



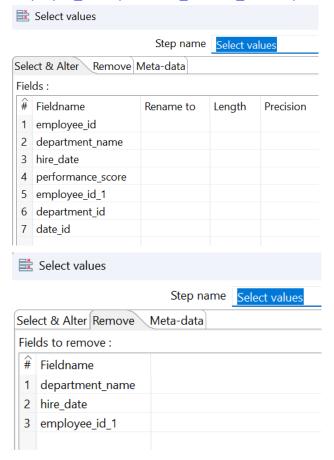
• lookup dim date

Menghubungkan hire_date dengan dim_date untuk menghasilkan date id.



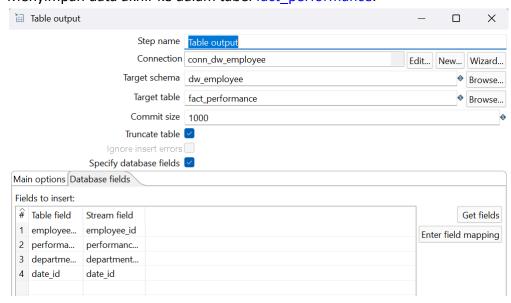
select values

Menyaring dan menyusun kembali kolom menjadi format akhir: employee_id, department_id, date_id, dan performance_score.



• table output

Menyimpan data akhir ke dalam tabel fact_performance.



e. job

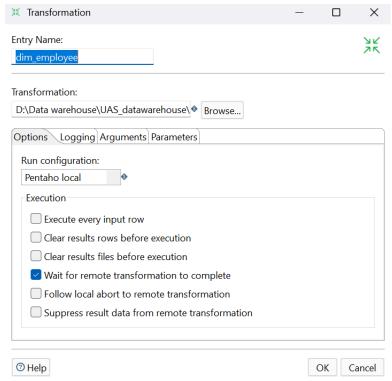


start

Langkah awal yang menandai dimulainya proses ETL. Job akan mulai mengeksekusi transformasi secara berurutan.

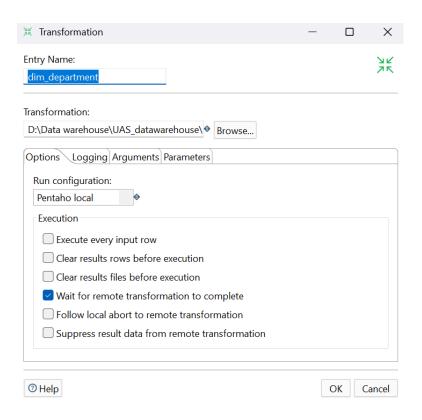
• dim_employee

Transformasi pertama yang dijalankan. Memuat data karyawan dari sumber ke tabel dimensi dim_employee.



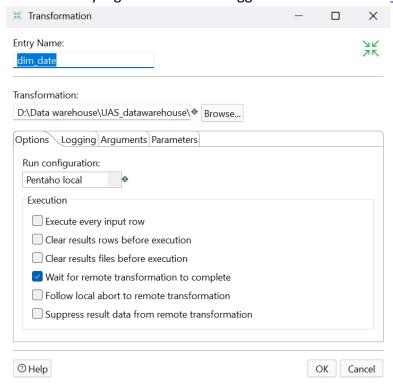
dim_department

Menjalankan transformasi untuk memuat data departemen ke dalam tabel dim_department.



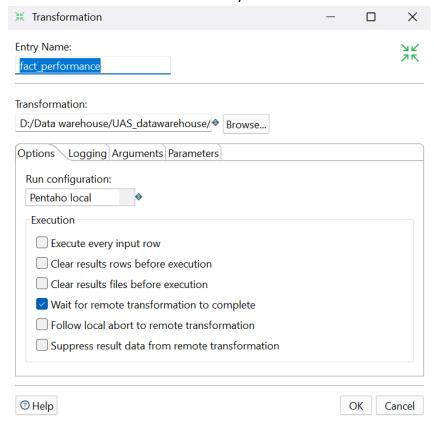
• dim_date

Transformasi yang memuat data tanggal ke dalam tabel dim_date.



• fact_performance

Setelah seluruh dimensi selesai, transformasi ini memuat data performa ke tabel fact_performance, dengan menggunakan referensi dari semua tabel dimensi sebelumnya.



success

Langkah akhir yang menunjukkan bahwa job ETL telah selesai dijalankan.

4. Analisis KPI

a. Rata-rata Skor Kinerja per Departemen per Tahun

KPI ini adalah rata-rata skor kinerja per departemen per tahun, sangat berguna untuk mengevaluasi efektivitas departemen dari waktu ke waktu, mendukung pengambilan keputusan manajemen berbasis data.

```
1 SELECT
2    d.department_name,
3    dt.year,
4    ROUND(AVG(f.performance_score), 2) AS avg_score
5 FROM fact_performance f
6 JOIN dim_department d ON f.department_id = d.department_id
7 JOIN dim_date dt ON f.date_id = dt.date_id
6 GROUP BY d.department_name, dt.year
9 ORDER BY dt.year, d.department_name;
```

department_name	year 🔺 1	avg score
Customer Support	2014	2.92
Engineering	2014	3
Finance	2014	2.8
HR	2014	2.92
IT	2014	3.05
Legal	2014	2.98
Marketing	2014	3.1
Operations	2014	3.02
Sales	2014	2.98
Customer Support	2015	2.99
Engineering	2015	3.03
Finance	2015	2.96
HR	2015	3.02
IT	2015	3.02
Legal	2015	2.98
Marketing	2015	2.91
Operations	2015	3.04
Sales	2015	3.04
Customer Support	2016	3.02
Engineering	2016	3.04
Finance	2016	2.97
HR	2016	2.95
IT	2016	3.02
Legal	2016	2.99
Marketing	2016	2.98

b. Jumlah Pegawai per Departemen per Tahun

KPI ini digunakan untuk mengetahui jumlah karyawan yang terlibat dalam penilaian kinerja per departemen setiap tahun. Sangat berguna untuk melacak pertumbuhan, perpindahan, atau penurunan jumlah tenaga kerja di setiap divisi.

```
1 SELECT
2    d.department_name,
3    dt.year,
4    COUNT(DISTINCT f.employee_id) AS total_employees
5 FROM fact_performance f
6 JOIN dim_department d ON f.department_id = d.department_id
7 JOIN dim_date dt ON f.date_id = dt.date_id
8 GROUP BY d.department_name, dt.year
9 ORDER BY dt.year, d.department_name;
```

department_name	year 🔺 1	total_employees
Customer Support	2014	348
Engineering	2014	329
Finance	2014	354
HR	2014	335
IT	2014	321
Legal	2014	365
Marketing	2014	381
Operations	2014	325
Sales	2014	339
Customer Support	2015	1116
Engineering	2015	1104
Finance	2015	1095
HR	2015	1051
IT	2015	1137
Legal	2015	1093
Marketing	2015	1108
Operations	2015	1143
Sales	2015	1039
Customer Support	2016	1103
Engineering	2016	1102
Finance	2016	1090
HR	2016	1057
IT	2016	1077
Legal	2016	1084
Marketing	2016	1106