

E-commerce Data Warehouse ETL Process

ID/X Partners Data Engineer
Project Based Internship VIX

Presented by
Wan Aufa Azis



Wan Aufa Azis

Undergraduate
Student at ITB

About Me

A third-year student majoring in Information Systems and Technology at ITB. Interested in data and technology. Experienced in working on data analytics, data processing, and data visualization projects in teams or individually. Possesses adequate analytical thinking, leadership, and communication skills.

Experience and Organizations

 Data Science Academy COMPFEST by Fasilkom UI - Top 15 Finalist

Jul – Aug 2022

 Pemodelan Basis Data Course Student Assistant - 1 of 6 assistants

Jul 2023 – Present

 Study in Bangkit Academy led by Google, Tokopedia, Gojek, & Traveloka - Machine Learning Path

Jan 2024 – Present

Case Study

Salah satu client dari ID/X Partners yang bergerak di bidang e-commerce memiliki kebutuhan untuk membuat sebuah Data Warehouse yang berasal dari beberapa tabel dari database sumber. Data Warehouse ini nantinya terdiri dari satu tabel Fact dan beberapa tabel Dimension.

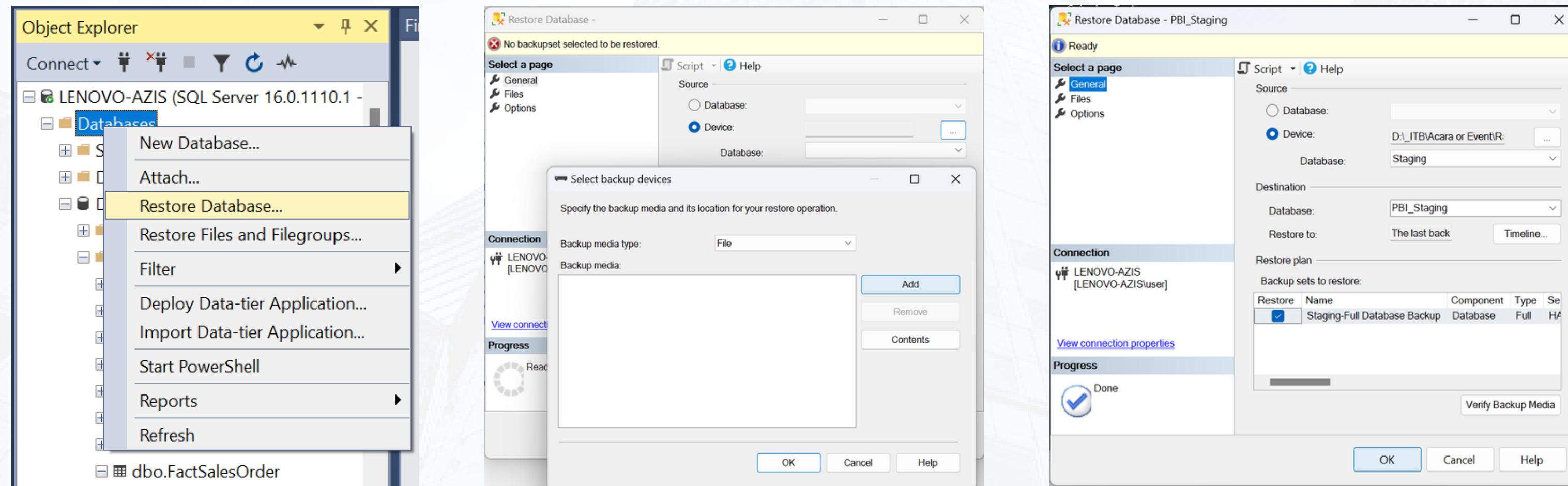
Sebagai Data Engineer, ada beberapa task yang perlu anda lakukan yaitu :

- 1. Melakukan Import/Restore Database Staging.**
- 2. Membuat sebuah Database bernama DWH_Project**, serta membuat Tabel Fact dan Dimension dari tabel yang ada di database Staging.
- 3. Membuat Job ETL di aplikasi talend untuk memindahkan data dari Staging ke Data Warehouse.** Khusus untuk Tabel DimCustomer, lakukan transformasi data dengan merubah data dari kolom FirstName dan LastName menjadi huruf kapital semua, lalu gabungkan kedua kolom tersebut menjadi satu kolom yang bernama CustomerName.
- 4. Membuat Store Procedure (SP)** untuk menampilkan summary sales order berdasarkan status pengiriman.

My Result

1. Import database staging

Download dan simpan file Staging.bak yang sudah diberikan, kemudian masuk ke menu Restore Database, upload file Staging.bak, dan lakukan restore file dengan menyesuaikan nama database target yang diinginkan



2. Membuat database DWH_Project

CREATE database baru dengan nama DWH_Project yang akan menjadi data warehouse. CREATE TABLE dan sesuaikan attribute dengan database Staging. Modifikasi sedikit pada customer dengan menggabungkan first_name dan last_name menjadi CustomerName.

```
Final Task.sql - not connected ✎ X
-- Create Table DWH_Project
CREATE TABLE DimCustomer (
    CustomerID INT,
    CustomerName VARCHAR(50) NOT NULL,
    Age INT NOT NULL,
    Gender VARCHAR(50) NOT NULL,
    City VARCHAR(50) NOT NULL,
    NoHP VARCHAR(50) NOT NULL,
    CONSTRAINT PK_DimCustomer PRIMARY KEY (CustomerID)
);

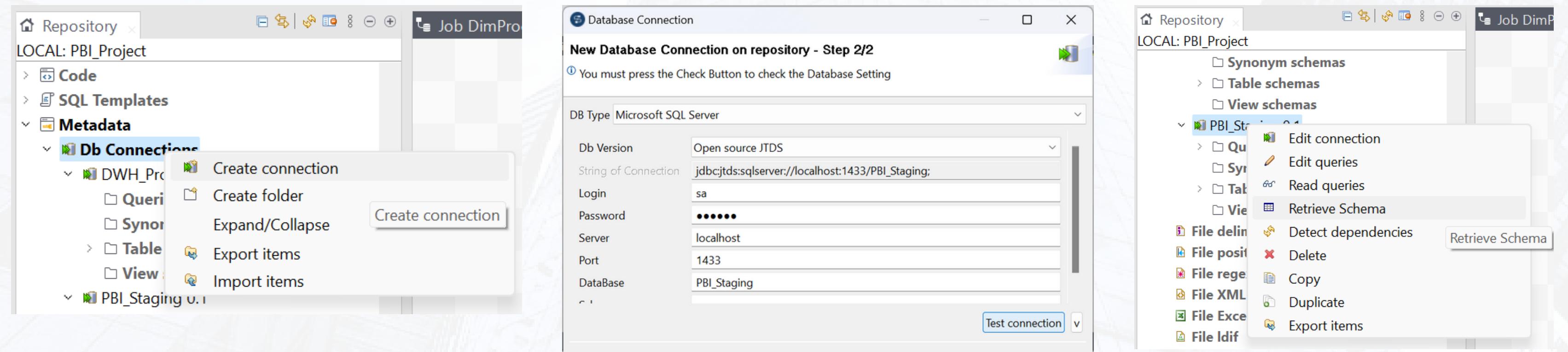
CREATE TABLE DimProduct (
    ProductID INT,
    ProductName VARCHAR(255) NOT NULL,
    ProductCategory VARCHAR(255) NOT NULL,
    ProductUnitPrice INT,
    CONSTRAINT PK_DimProduct PRIMARY KEY (ProductID)
);
```

```
Final Task.sql - not connected ✎ X
CREATE TABLE DimStatusOrder (
    StatusID INT,
    StatusOrder VARCHAR(50) NOT NULL,
    StatusOrderDesc VARCHAR(50) NOT NULL,
    CONSTRAINT PK_DimStatusOrder PRIMARY KEY (StatusID)
);

CREATE TABLE FactSalesOrder (
    OrderID INT,
    CustomerID INT NOT NULL,
    ProductID INT NOT NULL,
    Quantity INT NOT NULL,
    Amount INT NOT NULL,
    StatusID INT NOT NULL,
    OrderDate DATE NOT NULL,
    CONSTRAINT PK_FactSalesOrder PRIMARY KEY (OrderID),
    CONSTRAINT FK_CustomerOrder FOREIGN KEY (CustomerID) REFERENCES DimCustomer(CustomerID),
    CONSTRAINT FK_ProductOrder FOREIGN KEY (ProductID) REFERENCES DimProduct(ProductID),
    CONSTRAINT FK_StatusOrder FOREIGN KEY (StatusID) REFERENCES DimStatusOrder(StatusID)
);
```

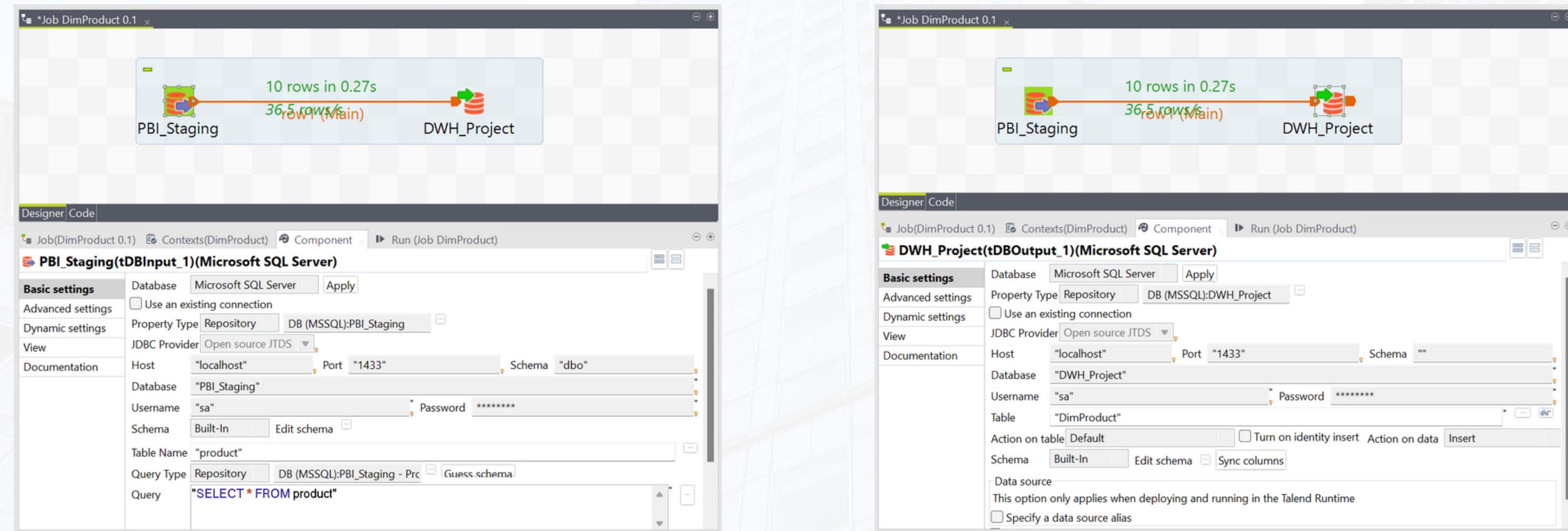
3. Buat job ETL di Talend

Buat job di Talend untuk memindahkan semua data dari Staging ke DWH_Project yg telah kita buat. Pertama, buat *metadata db connections* dari setiap database untuk memudahkan pemanggilan nantinya. Isi semua data yang diminta, lalu *test connection* hingga berhasil. Kemudian retrieve schema dan pilih semua schema dbo atau *database owner*



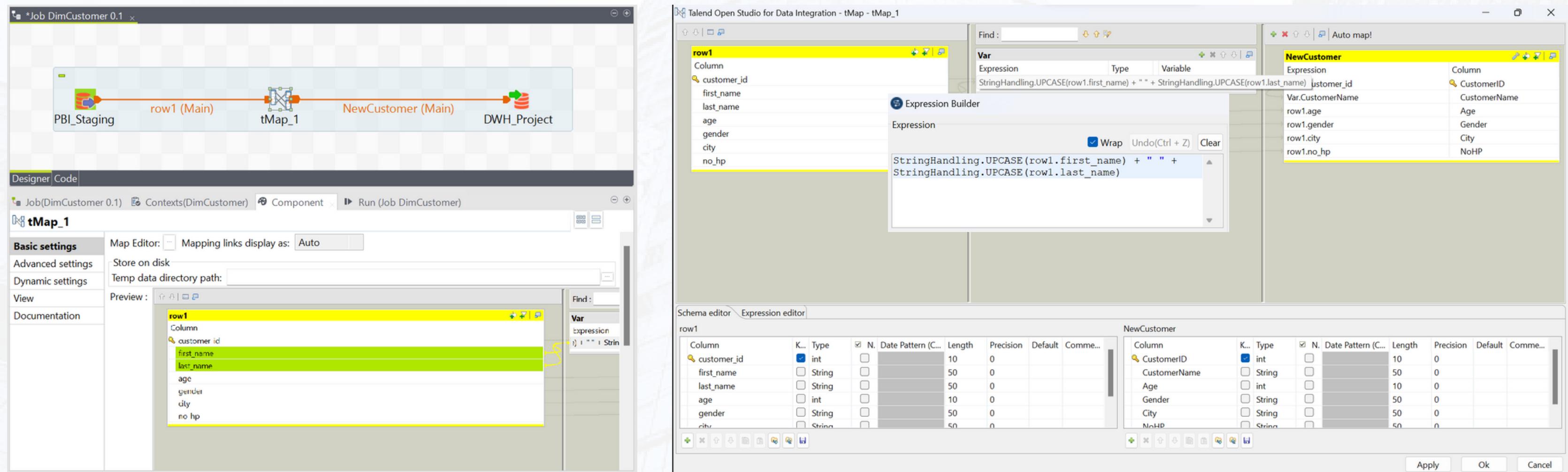
3. Buat job ETL di Talend

Jika sudah, kita mulai memindahkan data dari dimention table terlebih dahulu untuk menjaga aturan key. CREATE NEW JOB, lalu drag and hubungkan 2 metadata yang telah kita buat. Staging sebagai *input* dan DWH_Project sebagai *output*.



3. Buat job ETL di Talend

Untuk table customer sedikit berbeda, karena terdapat transformasi data. Gunakan komponen tMap untuk mengubah struktur data sesuai permintaan. Lalu jalankan semua job dan check data apakah berhasil masuk di MSSQL



4. Membuat stored procedure

Berikut stored procedure yang dapat menampilkan summary sales order berdasarkan status pengirimannya sebagai filter. Contoh penggunaan untuk mengecek sales order yang sudah completed

```
Final Task.sql - not connected ✎ X

-- Stored_Procedure
CREATE PROCEDURE summary_order_status
    @StatusID INT --filter param
AS
BEGIN
    SELECT
        OrderID,
        CustomerName,
        ProductName,
        Quantity,
        StatusOrder
    FROM FactSalesOrder fact
        JOIN DimCustomer cust ON fact.CustomerID=cust.CustomerID
        JOIN DimProduct prd ON fact.ProductID=prd.ProductID
        JOIN DimStatusOrder stat ON fact.StatusID=stat.StatusID
    WHERE fact.StatusID=@StatusID
END;
```

```
EXEC summary_order_status @StatusID=4;
```

100 %

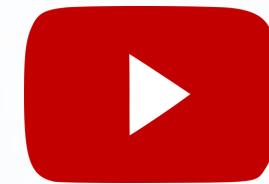
	OrderID	CustomerName	ProductName	Quantity	StatusOrder
1	1302	RIFKI MUHAMMAD	HP Elitebook 840 G4	1	Completed
2	1306	RIFKI MUHAMMAD	Asus Zenbook 800	1	Completed
3	1309	AJENG SRIASIH	Bagpack Navy Club	1	Completed

Project Link



github.com/WanAzis/ETL-Process_PBI-Project.git

Video Presentation



<https://youtu.be/UpT8Kkr5rmk>

Thank You

