

# Membuat Data Warehouse

ID/X Partners Data Engineer  
Virtual Internship Program

Noor Kharismawan Akbar



## Noor Kharismawan Akbar

### About Me

- Fresh Graduate from Metallurgical Engineering - Institut Teknologi Bandung with cum-laude predicate.
- Also the best graduate of Purwadhika Job Connector - Data Science Online Batch 9 2023.

### Work Experience

#### **HYUNDAI MOTOR MANUFACTURING INDONESIA**

Purchasing & Project Development – Full Time

#### **ID/X PARTNERS**

Data Scientist & Engineer – Project Based Internship

#### **HOME CREDIT INDONESIA**

Data Scientist – Project Based Internship

#### **BANK MUAMALAT**

Business Intelligence Analyst & Product Business Developer – Project Based Internship



# Challenge - Task

Membuat Data Warehouse untuk kebutuhan salah satu client ID/X Partners



# Challenge - Task

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 dilakukan 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



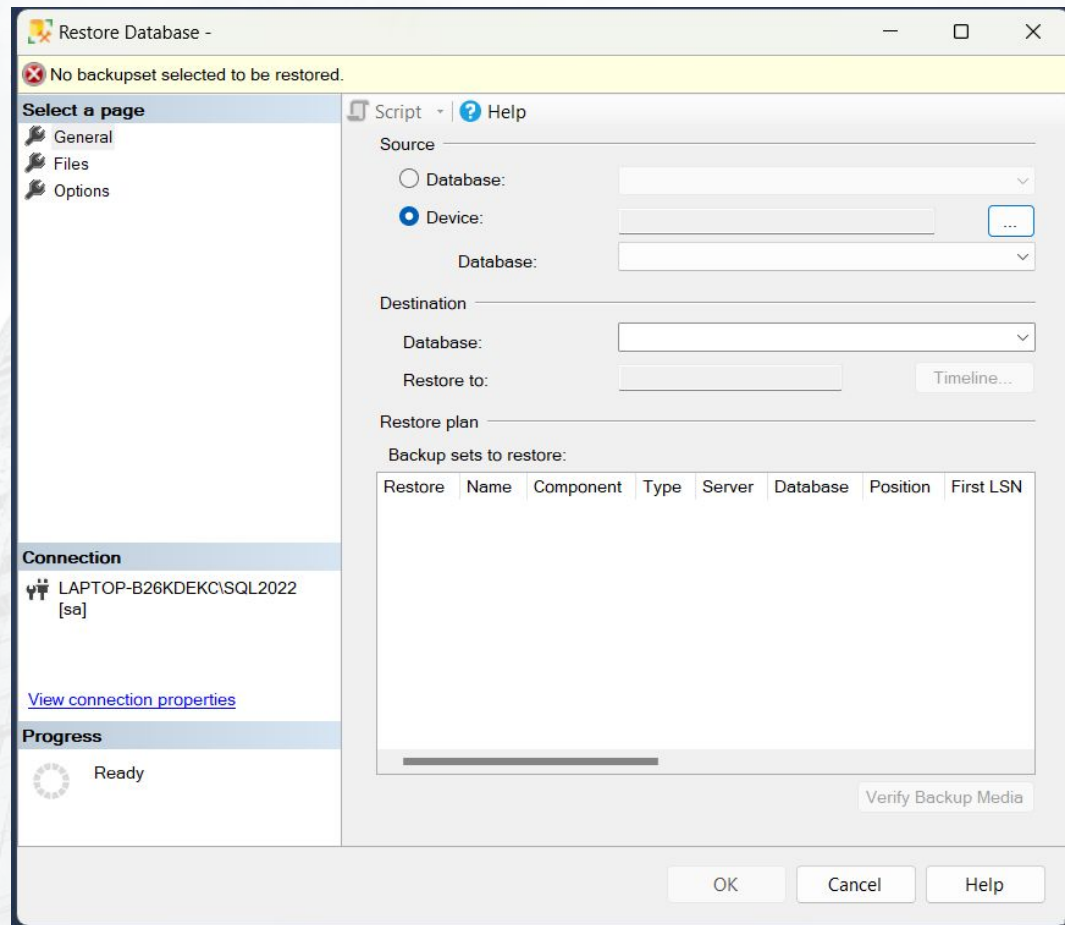
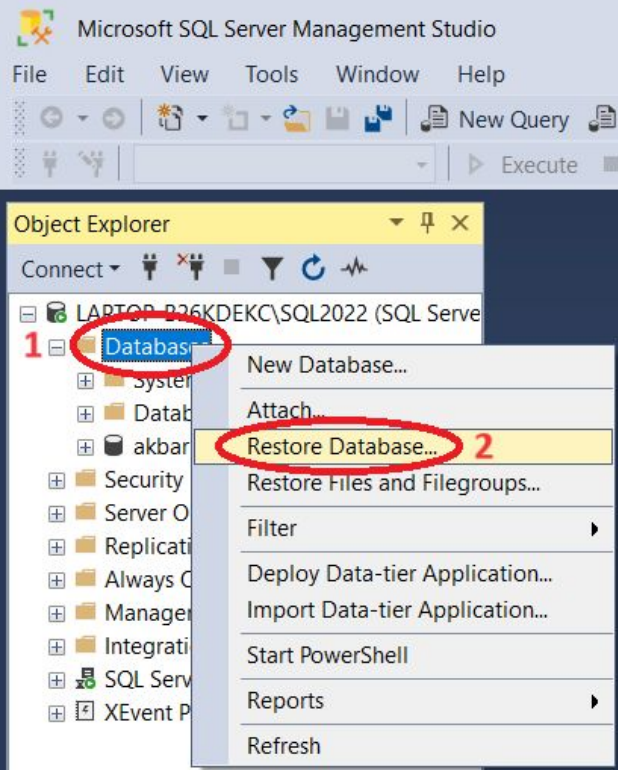
# Task - 1

## Melakukan import/restore database staging

- Menggunakan aplikasi Microsoft SQL Server Management Studio (SSMS)
- Menggunakan file **Staging.bak** yang didapat dari link : [File Restore Database Stagging](#)

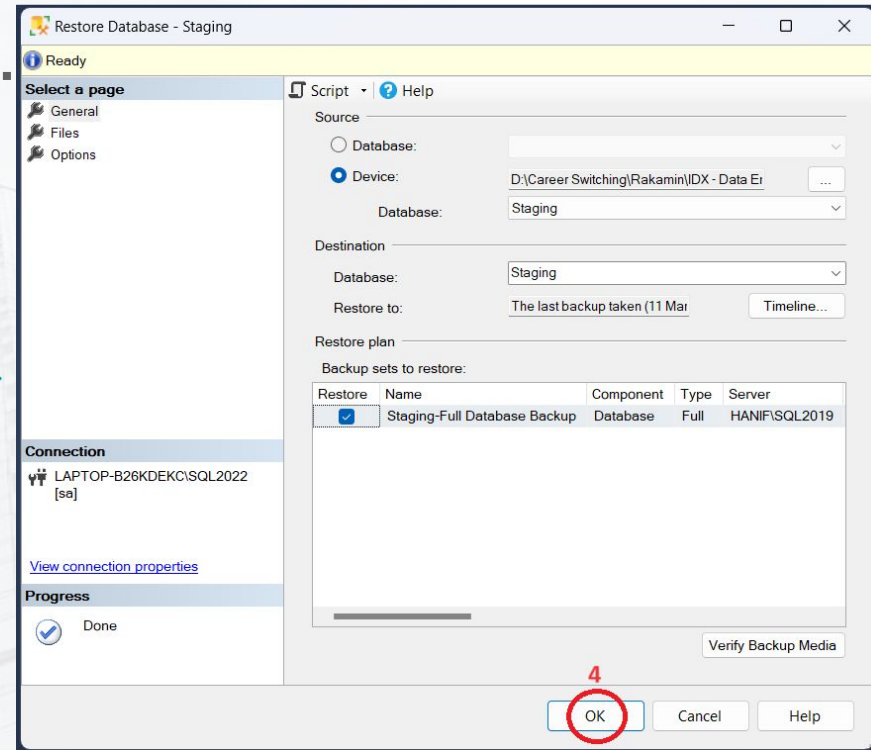
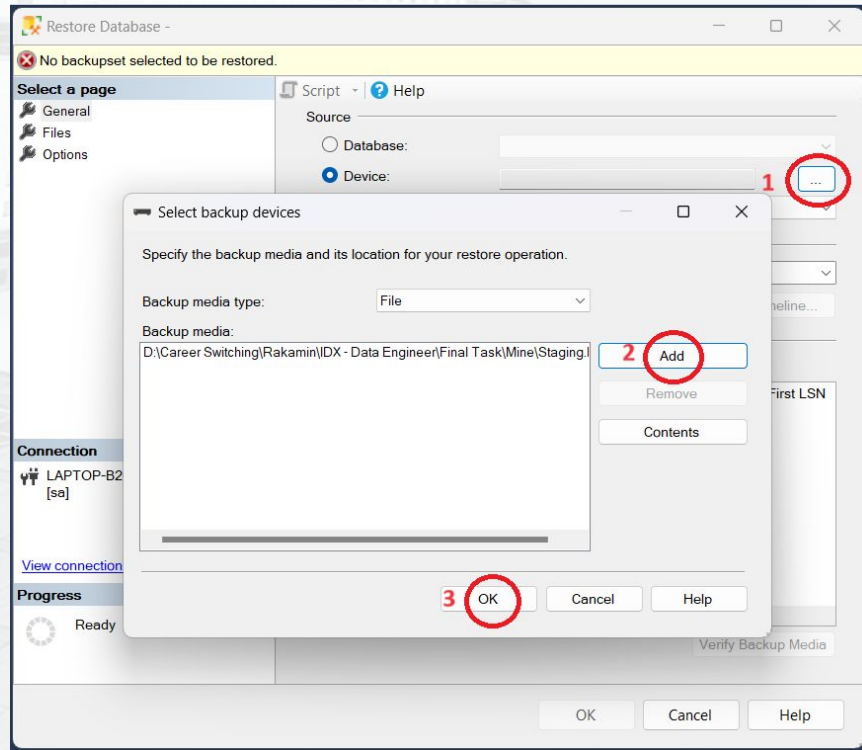
# Task 1 - Restore Database Staging

Melakukan restore database untuk melihat Database **Staging.bak** pada SSMS





# Task 1 - Restore Database Staging

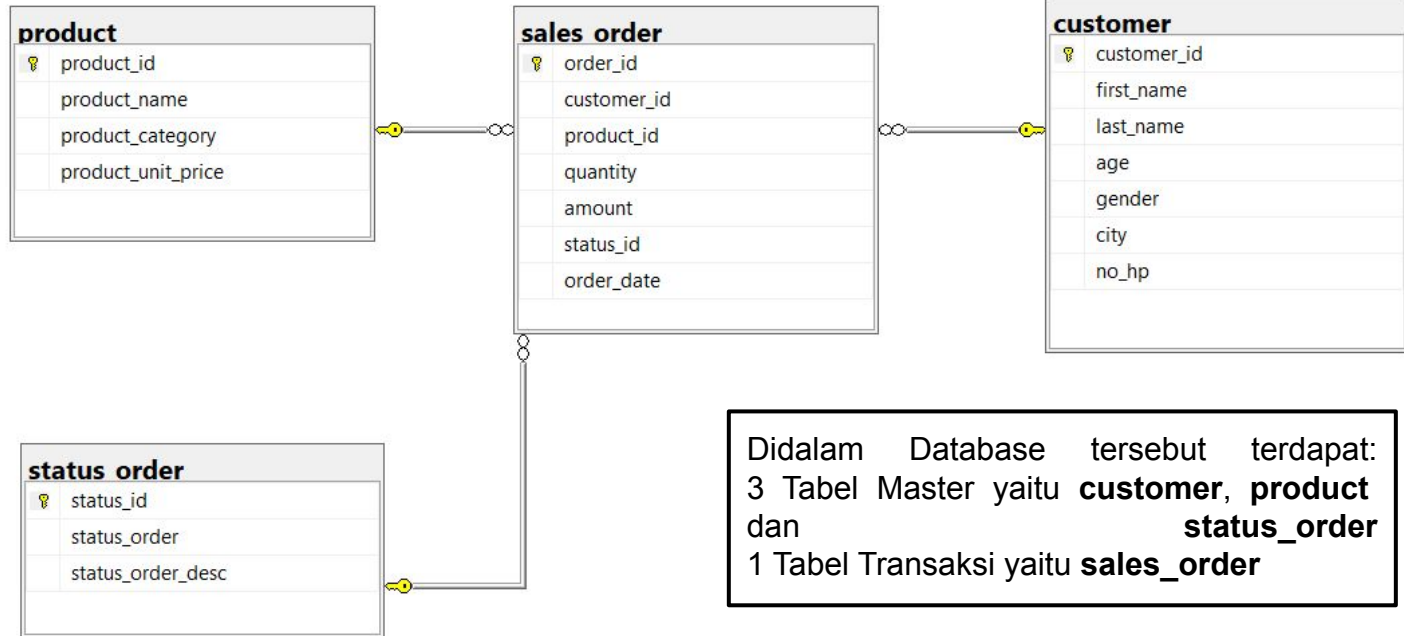


Arahkan Source ke Device sesuai tempat penyimpanan database **Staging.bak**

Pastikan File yang akan direstore betul & klik OK

# Task 1 - Restore Database Staging

LAPTOP-B26KDEKC\...iagramSalesOrder\* ➡ ✕







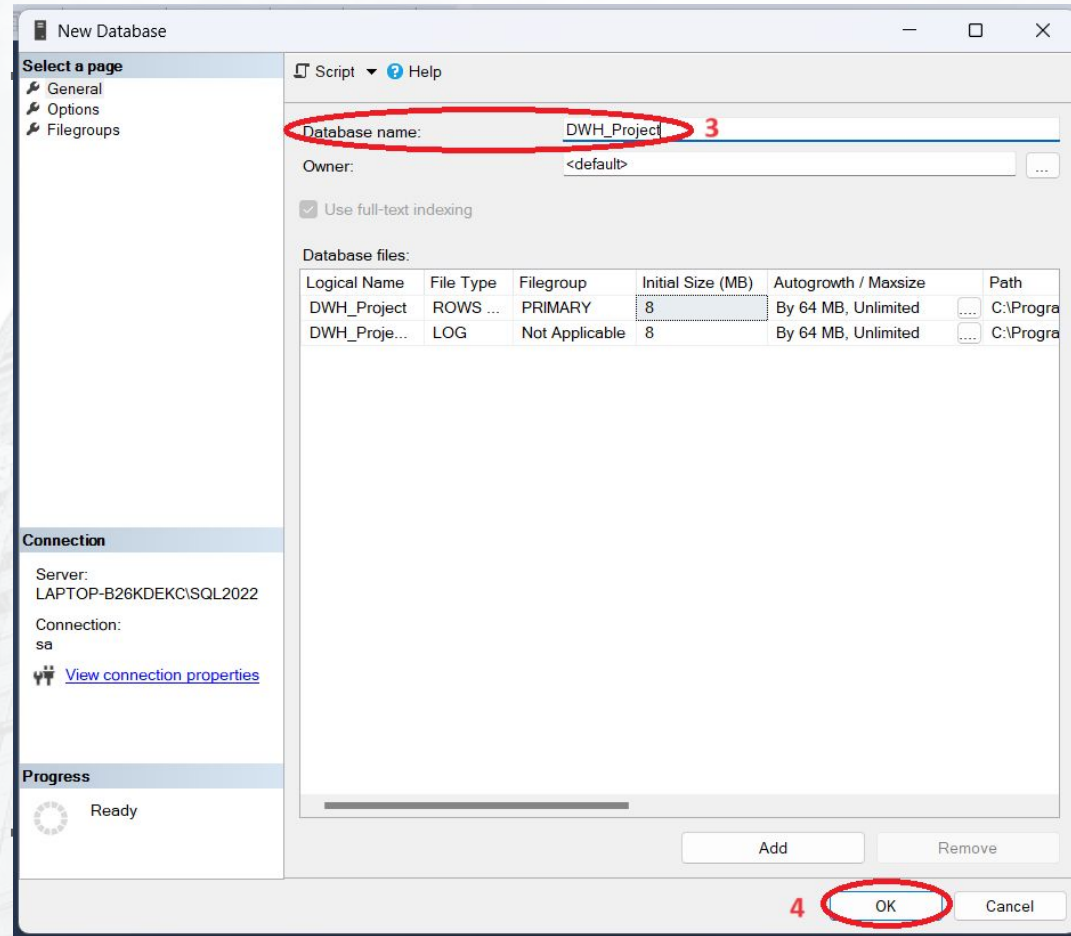
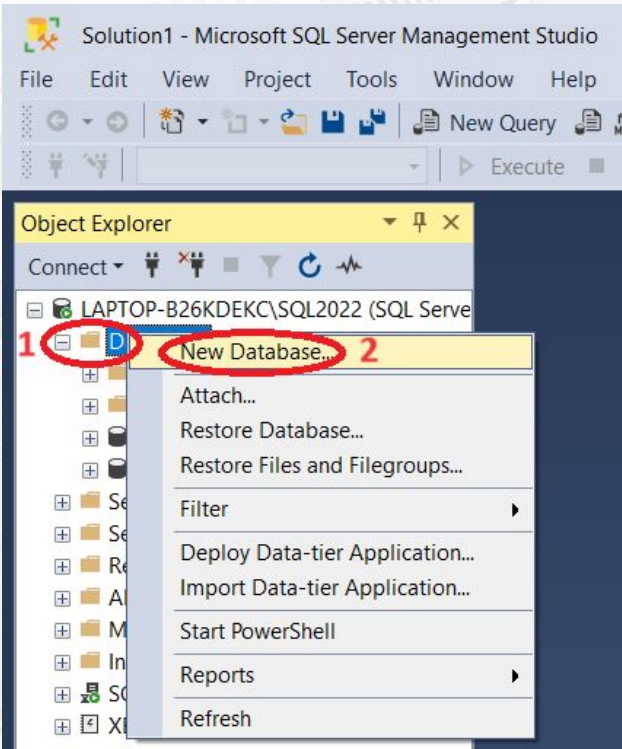
## Task - 2

**Membuat sebuah database bernama DWH\_Project**

serta membuat Tabel Fact dan Dimension dari tabel yang ada di database Staging.

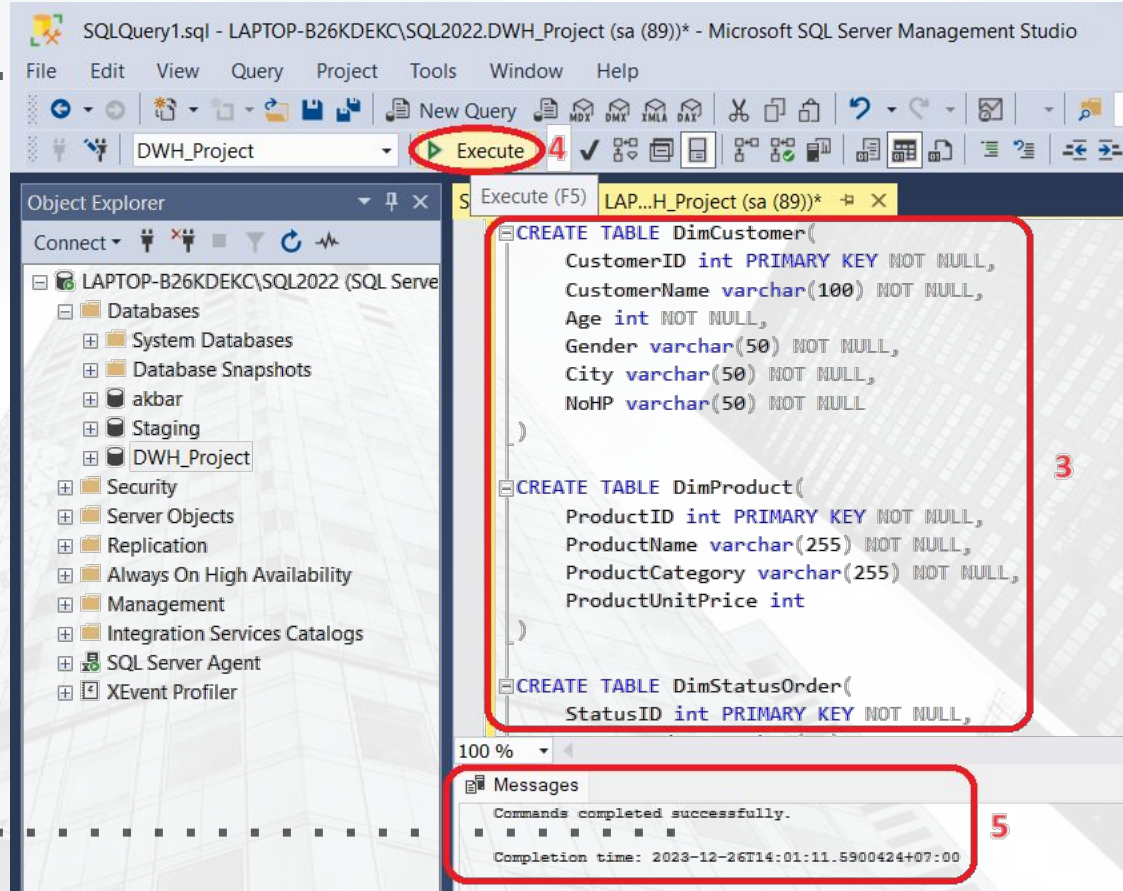
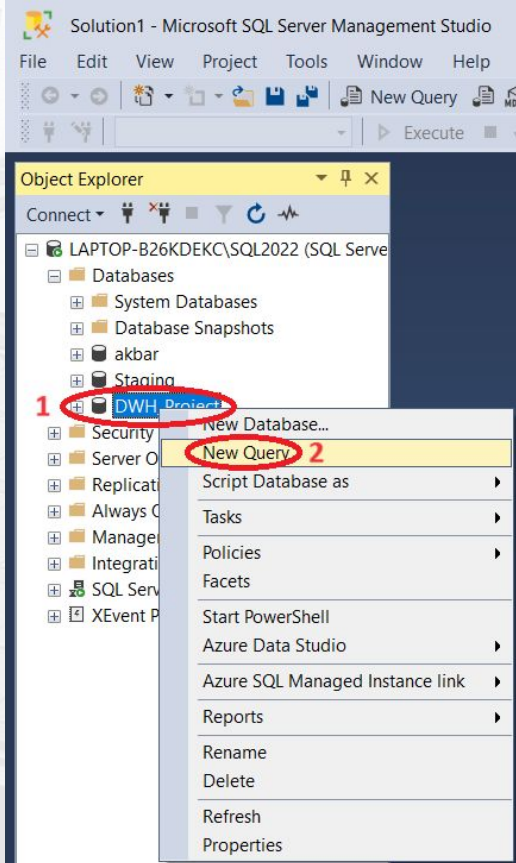
# Task 2 - Membuat Database DWH\_Project beserta Tabel

## Membuat Database DWH\_Project



# Task 2 - Membuat Database DWH\_Project beserta Tabel

## Membuat Tabel dalam DWH\_Project





# Task 2 - Membuat Database DWH\_Project beserta Tabel

**Table DimCustomer** berisi data:

- CustomerID
- CustomerName
- Age
- Gender
- City
- NoHP

**SQL Code**

```
CREATE TABLE DimCustomer(  
    CustomerID int PRIMARY KEY NOT NULL,  
    CustomerName varchar(100) NOT NULL,  
    Age int NOT NULL,  
    Gender varchar(50) NOT NULL,  
    City varchar(50) NOT NULL,  
    NoHP varchar(50) NOT NULL  
)
```

**Table DimProduct** berisi data:

- ProductID
- ProductName
- ProductCategory
- ProductUnitPrice

**SQL Code**

```
CREATE TABLE DimProduct(  
    ProductID int PRIMARY KEY NOT NULL,  
    ProductName varchar(255) NOT NULL,  
    ProductCategory varchar(255) NOT NULL,  
    ProductUnitPrice int  
)
```

# Task 2 - Membuat Database DWH\_Project beserta Tabel

## Table DimStatusOrder

berisi data:

- StatusID
- StatusOrder
- StatusOrderDesc

SQL Code

```
CREATE TABLE DimStatusOrder(  
    StatusID int PRIMARY KEY NOT NULL,  
    StatusOrder varchar(50) NOT NULL,  
    StatusOrderDesc varchar(50) NOT NULL  
)
```

## Table FactSalesOrder

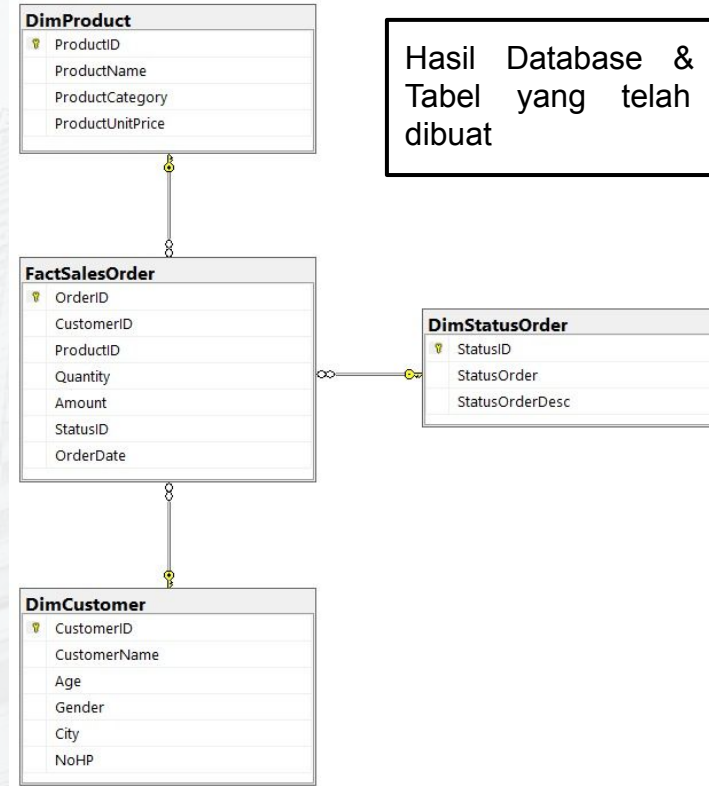
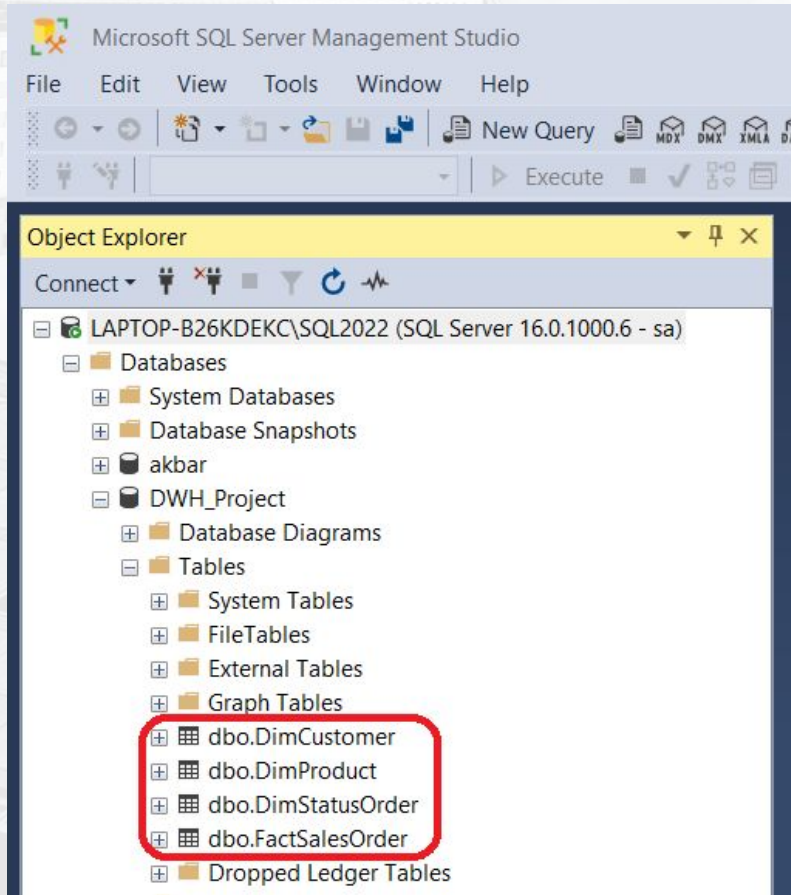
berisi data:

- OrderID
- CustomerID
- ProductID
- Quantity
- Amount
- StatusID
- OrderDate

SQL Code

```
CREATE TABLE FactSalesOrder(  
    OrderID int PRIMARY KEY NOT NULL,  
    CustomerID int FOREIGN KEY REFERENCES DimCustomer(CustomerID) NOT NULL,  
    ProductID int FOREIGN KEY REFERENCES DimProduct(ProductID) NOT NULL,  
    Quantity int NOT NULL,  
    Amount int NOT NULL,  
    StatusID int FOREIGN KEY REFERENCES DimStatusOrder(StatusID) NOT NULL,  
    OrderDate date NOT NULL  
)
```

# Task 2 - Membuat Database DWH\_Project beserta Tabel







## Task - 3

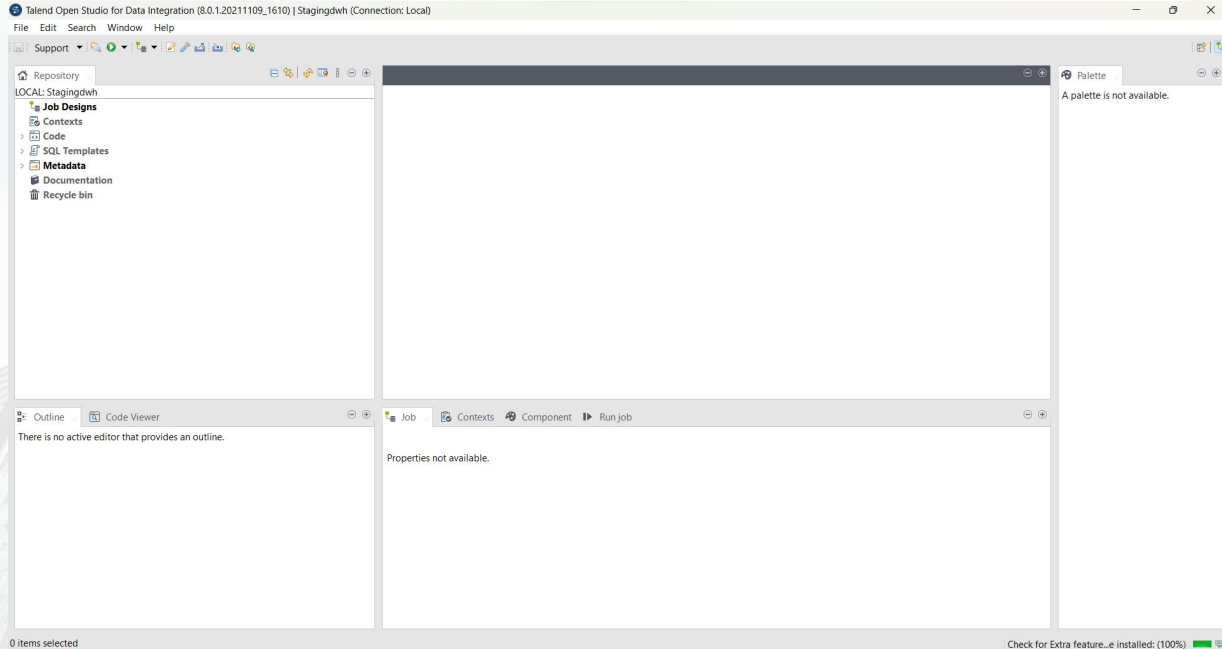
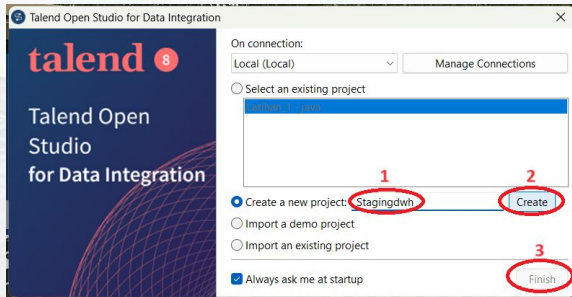
### Membuat Job ETL di Talend

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**.

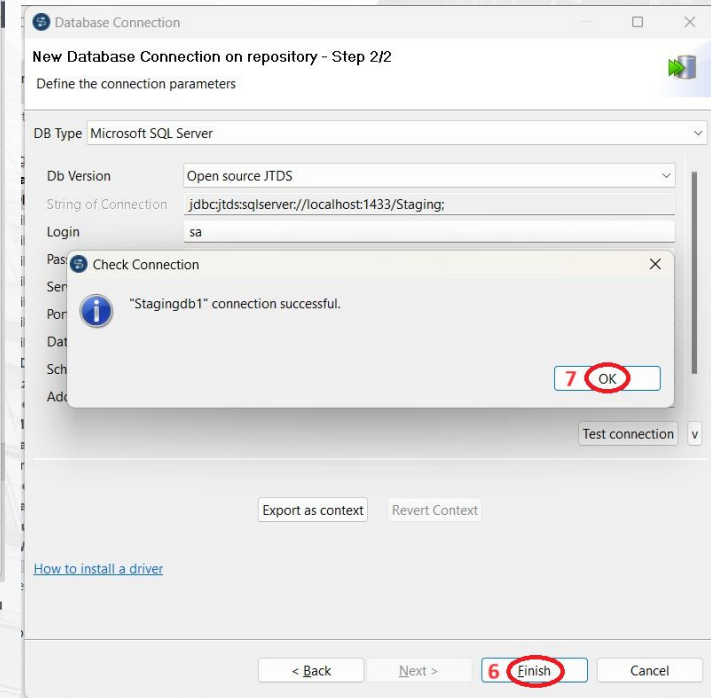
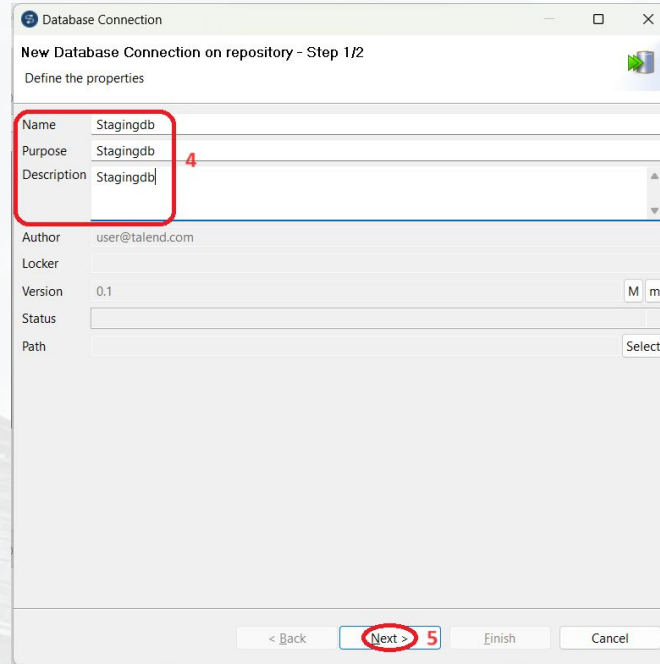
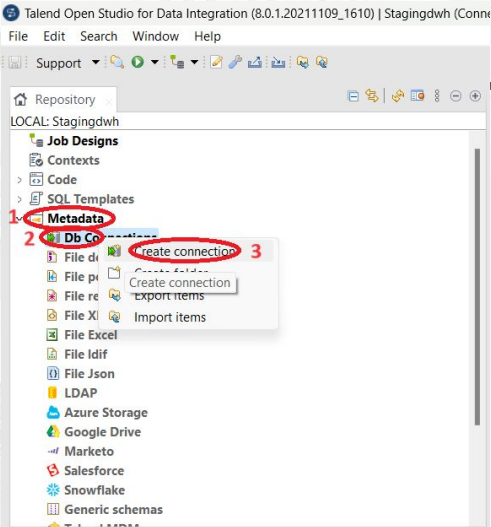
# Task 3 - Membuat Job ETL di Talend

## Membuat new project dalam Talend



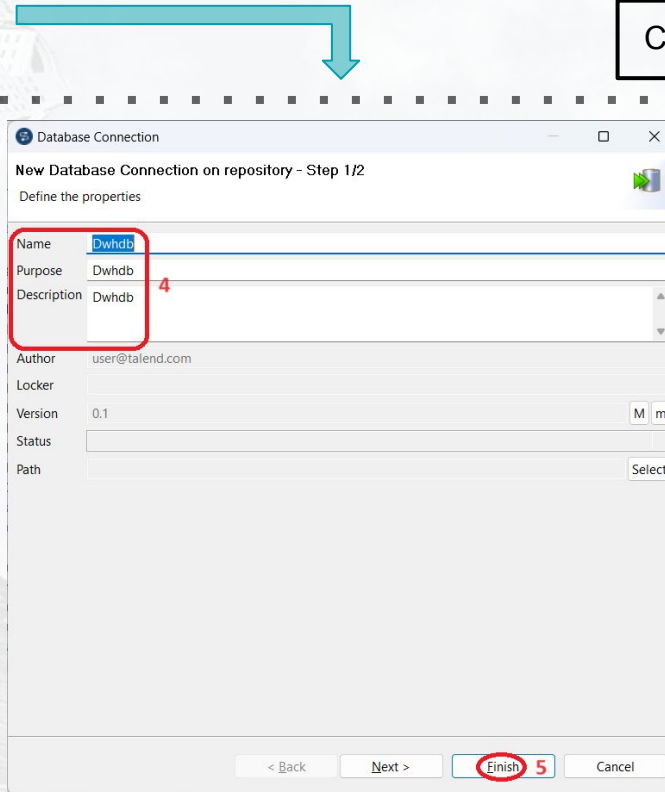
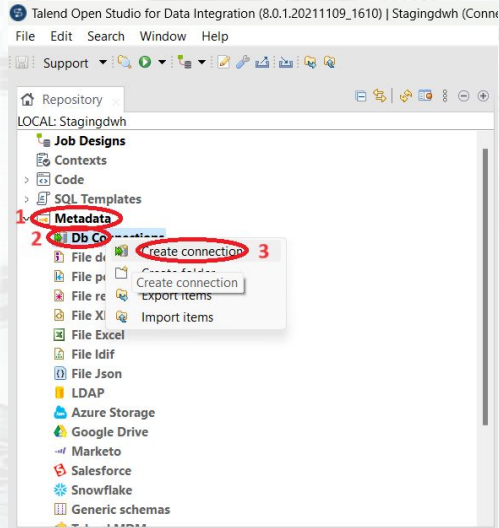
# Task 3 - Membuat Job ETL di Talend

Create Connection untuk **Stagging.db**

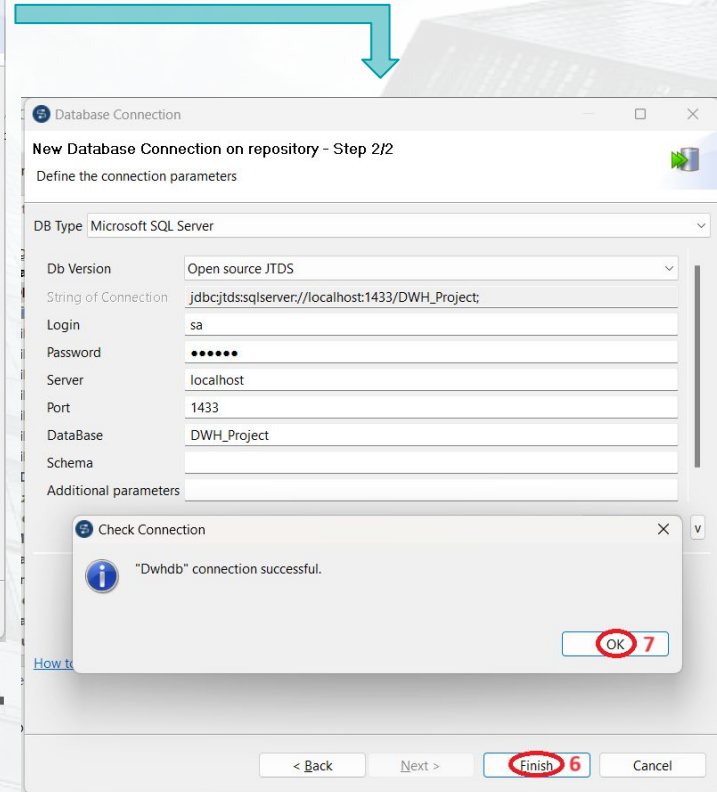




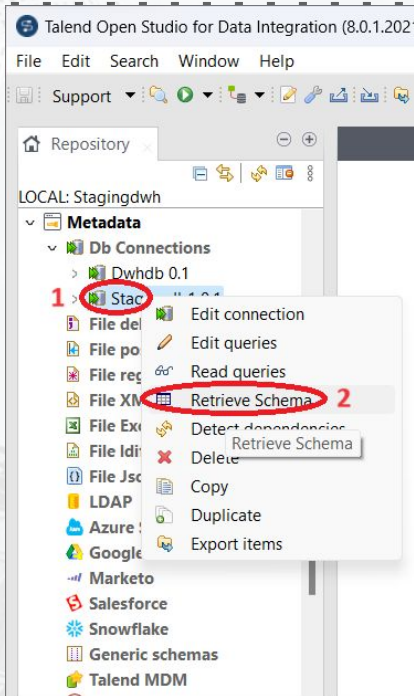
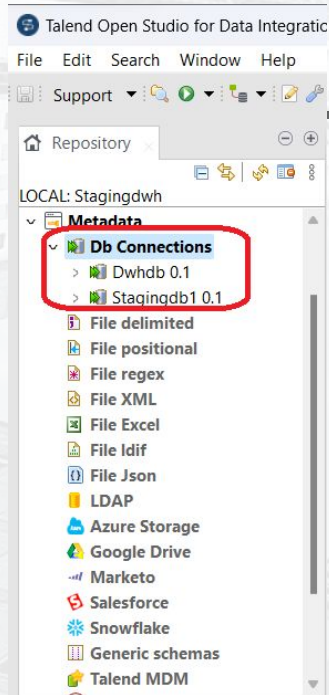
# Task 3 - Membuat Job ETL di Talend



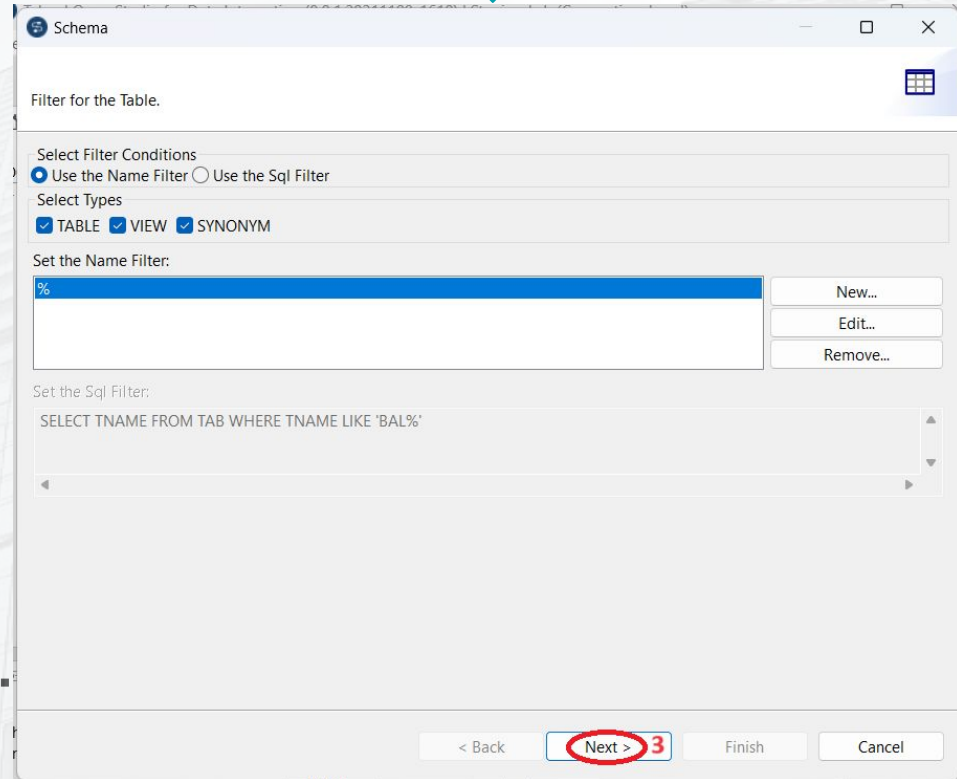
Create Connection untuk Dwhdb.db



# Task 3 - Membuat Job ETL di Talend



Retrieve Schema untuk Stagingdb



# Task 3 - Membuat Job ETL di Talend

Pilih tabel yang sesuai

Beri nama untuk schema

Schema

New Schema in connection "Stagingdb1"

Add a Schema on repository

Select Schema to create

Name Filter:

Name	Type	Column number	Creation status
> <input type="checkbox"/> db_denydatawriter	SCHEMA		
> <input type="checkbox"/> db_owner	SCHEMA		
> <input type="checkbox"/> db_securityadmin	SCHEMA		
✓ <input checked="" type="checkbox"/> dbo	SCHEMA		
<input checked="" type="checkbox"/> customer	TABLE	7	Success
<input checked="" type="checkbox"/> product	TABLE	4	Success
<input checked="" type="checkbox"/> sales_order	TABLE	7	Success
<input checked="" type="checkbox"/> status_order	TABLE	3	Success
<input type="checkbox"/> sysdiagrams	TABLE		
> <input type="checkbox"/> guest	SCHEMA		
> <input type="checkbox"/> INFORMATION_SCHEMA	SCHEMA		
> <input type="checkbox"/> sys	SCHEMA		

Select All Select None Check Connection

< Back **Next > 2** Finish Cancel



Schema

New Schema in connection "Stagingdb1"

Add a Schema on repository

Schema

- customer
- product
- sales\_order
- status\_order

Name **product 3**

Comment

Type : TABLE

Based on table product Retrieve Schema Guess Schema

Use the "Retrieve Schema" button to replace the current Schema by the table based Schema

Schema

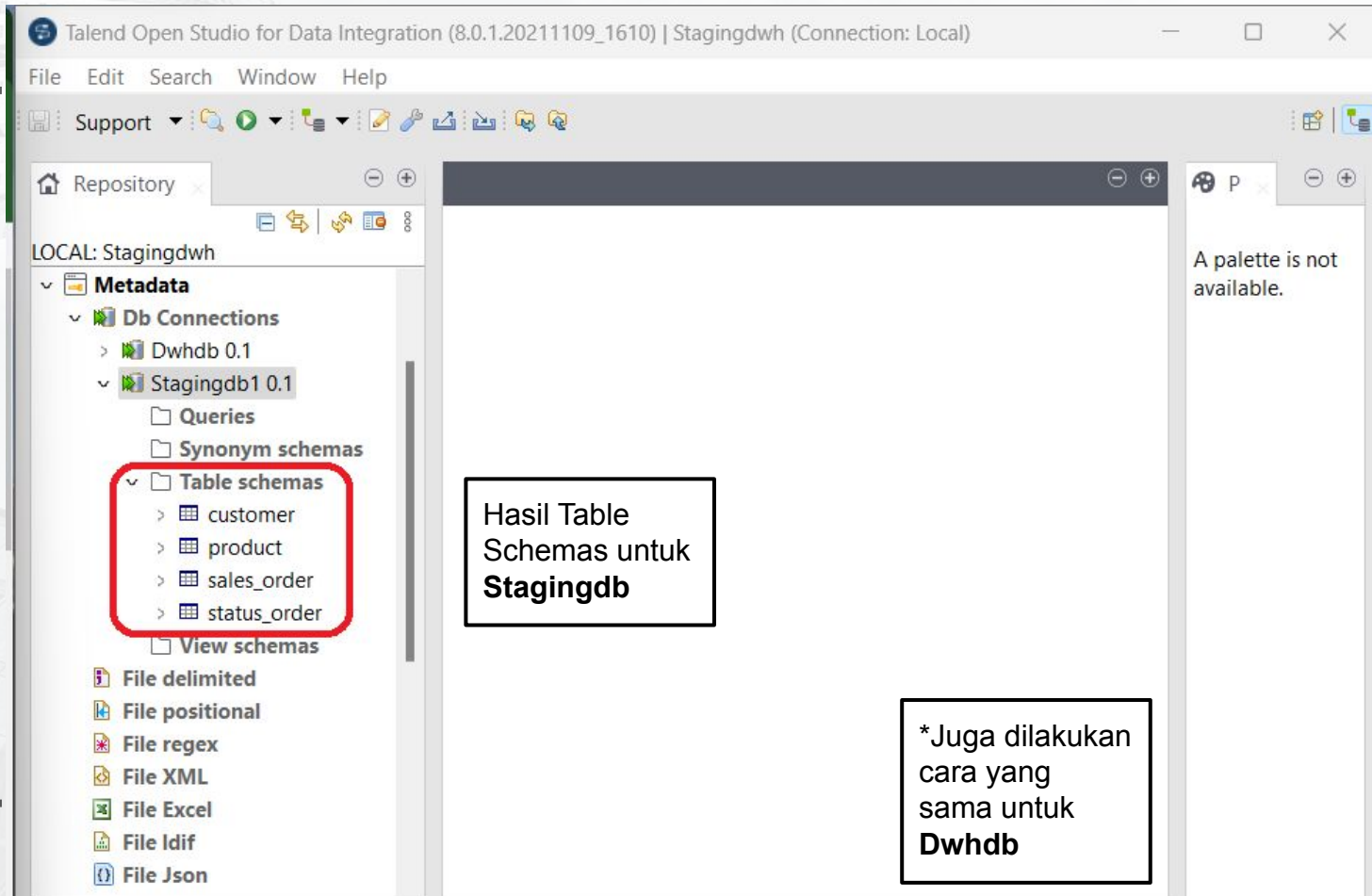
Column	Db Column	K...	DB Ty...	Type	✓	N...	Date P...	Le...	Pre...	D...
product_id	product_id	<input checked="" type="checkbox"/>	INT	int	<input type="checkbox"/>			10	0	
product_na...	product_na...	<input type="checkbox"/>	VARC...	Str...	<input type="checkbox"/>			255	0	
product_cat...	product_cat...	<input type="checkbox"/>	VARC...	Str...	<input type="checkbox"/>			255	0	
product_u...	product_u...	<input type="checkbox"/>	INT	int	<input checked="" type="checkbox"/>			10	0	

Add Schema Remove Schema

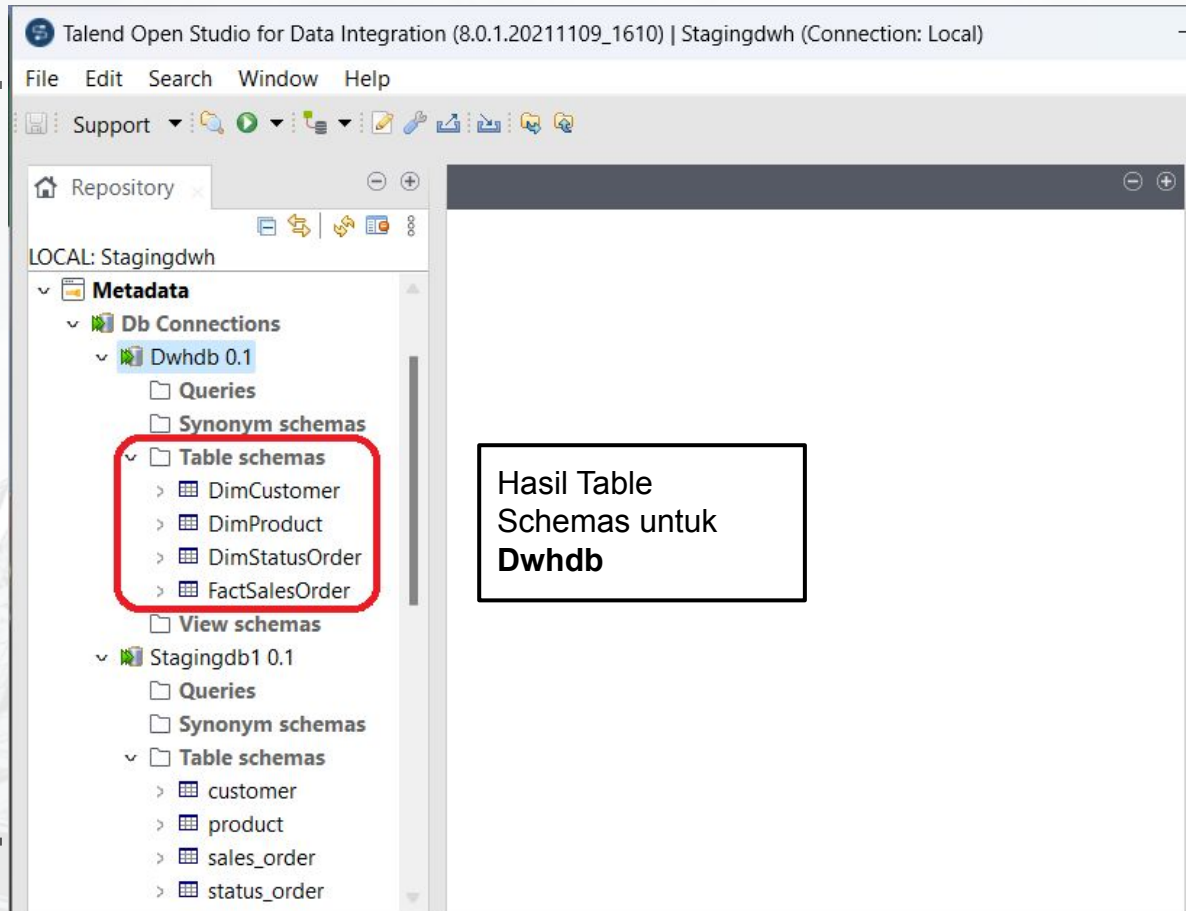
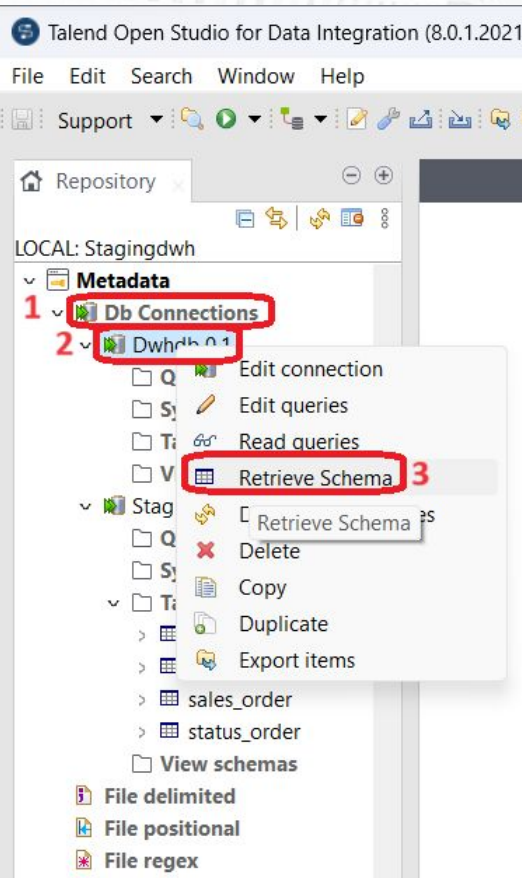
< Back Next > **Finish 4** Cancel



# Task 3 - Membuat Job ETL di Talend



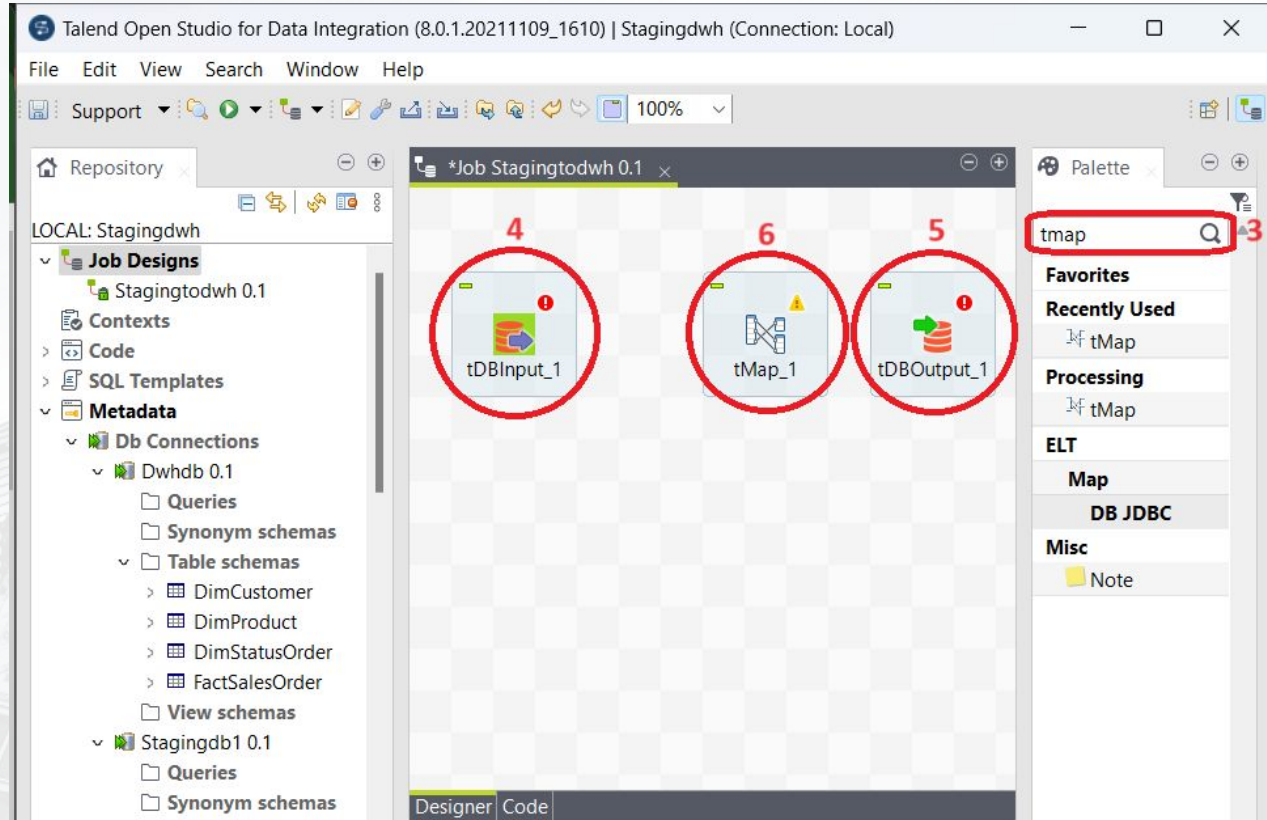
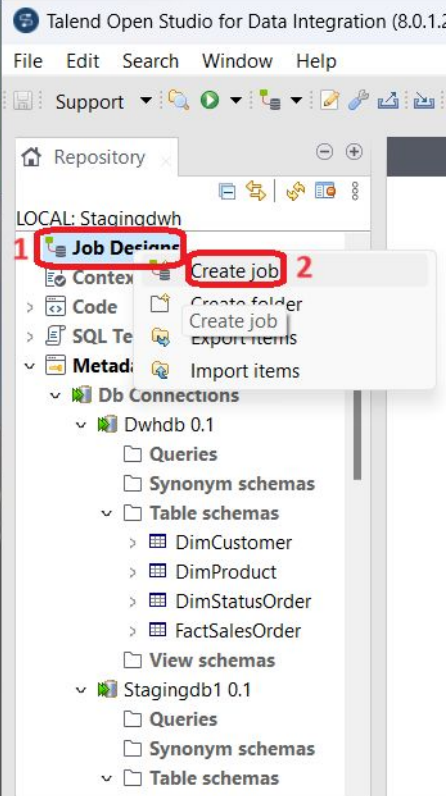
# Task 3 - Membuat Job ETL di Talend



# Task 3 - Membuat Job ETL di Talend

Membuat Job Design

Mengimport tDBInput, tMap, dan tDBOutput





# Task 3 - Membuat Job ETL di Talend

Job Stagingtodwh 0.1

Repository Content

- Metadata
  - Db Connections
    - Dwhdb 0.1
      - Stagingdb1 0.1
        - Queries
        - Synonym schemas
        - Table schemas
          - customer
          - product
          - sales\_order
          - status\_order
        - View schemas
      - File delimited
      - File positional

OK Cancel

Designer Code

Job(Stagingtodwh)

tDBInput\_1(Microsoft SQL Server)

Basic settings

Advanced settings

Dynamic settings

View

Documentation

JDBC Provider: Open source JTDS

Host: localhost Port: 1433 Schema: ""

Database: "Staging"

Username: "sa" Password: \*\*\*\*\*

Schema: Repository DB (MSSQL):Stagingdb1 - status\_or Edit schema

Table Name: "status\_order"

Query Type: Built-In Guess Query Guess schema

Query: "select id, name from employee"



tMap\_1Output

new Output name?

stagingtodwh

OK Cancel

# Task 3 - Membuat Job ETL di Talend

row1	Var	stagingtodwh
Column		Expression
customer_id		ow1.customer_id
first_name		
last_name		
age		ow1.age
gender		ow1.gender
city		ow1.city
no_hp		ow1.no_hp
		Column
		CustomerID
		CustomerName
		Age
		Gender
		City
		NoHP



Expression Builder

Expression

☒ Wrap Undo(Ctrl + Z) Clear

StringHandling.UPCASE(row1.first\_name) + ' ' + StringHandling.UPCASE(row1.last\_name)

4

Test

Test! Clear

Var Va

row1.custo... 0

row1.first\_n... nu

row1.last\_na... nu

row1.age 0

row1.gender nu

row1.city nu

row1.no\_hp nu

Add Remove

Categories

\*All

\*User Defined

DataOperation

Mathematical

Numeric

Relational

StringHandling

TalendDataGenerator

TalendDate

2

Functions

LEFT(String string, int index) : id\_String

LEN(String string) : id\_Integer - String

RIGHT(String string, int index) : id\_String

SPACE(int i) : id\_String - StringHandling

SQUOTE(String string) : id\_String - String

STR(String string, int int) : id\_String - String

TRIM(String string) : id\_String - String

UPCASE(String string) : id\_String - String

3

Help

Converts all lowercase letters

Ok 5 Cancel

# Task 3 - Membuat Job ETL di Talend

Job Stagingtodwh 0.1

8 rows in 0.02s  
100 rows/s  
row1 (Main)

8 rows in 0.3s  
26.67 rows/s  
stagingtodwh (Main)

tDBInput\_1

tMap\_1

tDBOutput\_1

1

Designer Code

Job (Stagingtodwh 0.1) Contexts (Stagingtodwh) Component Run (Job Stagingtodwh)

Job Stagingtodwh

Execution

Basic Run

Debug Run

Advanced settings

Target Exec

Memory Run

Run 2

Kill

Clear

Starting job Stagingtodwh at 06:50 27/12/2023.  
[statistics] connecting to socket on port 3525  
[statistics] connected  
[statistics] disconnected  
Job Stagingtodwh ended at 06:50 27/12/2023. [Exit code = 0]

3

**Talend:**  
Run tMap\_1

SQLQuery2.sql - LAP...H\_Project (sa (95))

```
SELECT * FROM DimCustomer;  
SELECT * FROM DimProduct;  
SELECT * FROM DimStatusOrder;  
SELECT * FROM FactSalesOrder;
```

4

100 %

Results Messages

CustomerID	CustomerName	Age	Gender	City	NoHP
1 201	BUDI SANTOSO	45	Pria	Jakarta	087645465712
2 202	AJENG SRIASIH	25	Wanita	Bogor	089045465712
3 203	BAGUS PRAKOSO	20	Pria	Depok	087905465712
4 204	LIA RAHMAWATI	31	Wanita	Bekasi	089945408712
5 205	AZMU FATI	28	Pria	Jakarta	087689765712
6 206	RIFKI MUHAMMAD	22	Pria	Depok	087645468907
7 207	BELA ADIRILIA	24	Wanita	Tangerang	087647665712
8 208	RAHMA AMELIA	18	Wanita	Bogor	087645431212

ProductID	ProductName	ProductCategory	ProductUnitPrice
1 1001	Macbook Air 2020 13 inch	Komputer & Laptop	12000000
2 1002	T-Shirt Polo Nevada	Pakaian	150000
3 1003	Blender Philips 500 watt	Elektronik	200000
4 1004	Kipas Angin Cosmos	Elektronik	120000
5 1005	HP Elitebook 840 G4	Komputer & Laptop	10000000
6 1006	Asus Zenbook 800	Komputer & Laptop	9000000
7 1007	Luciana Set Dress 2 in 1	Pakaian	300000
8 1008	Converse Cap Original	Topi	180000

StatusID	StatusOrder	StatusOrderDesc
1 1	Awaiting Payment	Menunggu Pembayaran
2 2	Awaiting Shipment	Menunggu Pengiriman
3 3	Shipped	Sedang Dikirim
4 4	Completed	Pesanan sampai tujuan
5 5	Cancelled	Pesanan dibatalkan ol...

OrderID	CustomerID	ProductID	Quantity	Amount	StatusID	OrderDate
1 1301	204	1008	2	360000	2	2022-01-06
2 1302	206	1005	1	10000000	4	2022-01-20
3 1303	201	1001	1	12000000	1	2022-02-02
4 1304	202	1002	2	300000	2	2022-02-04
5 1305	203	1003	3	600000	3	2022-03-28
6 1306	206	1006	1	9000000	4	2022-03-15

5

Query executed successfully.

**SSMS:**  
Cek hasil tMap\_1





# Task - 4

**Membuat sebuah Stored Procedure**  
untuk menampilkan summary sales order berdasarkan status pengiriman

# Task 4 - Membuat sebuah Stored Procedure

The screenshot shows the Microsoft SQL Server Management Studio interface. The title bar indicates the file is 'SQLQuery1.sql - LAPTOP-B26KDEKC\SQL2022.DWH\_Project (sa (76))\* - Microsoft SQL Server Management Studio'. The menu bar includes File, Edit, View, Query, Project, Tools, Window, and Help. The toolbar contains various icons, with the 'Execute' button (a green play icon) circled in red and labeled with a red '2'. The Object Explorer on the left shows the server hierarchy for 'LAPTOP-B26KDEKC\SQL2022 (SQL Server 16.0.1000.6 - sa)', including Databases, System Databases, Database Snapshots, akbar, DWH\_Project, Staging, Security, Server Objects, Replication, and Always On High Availability. The main query window displays the following T-SQL code:

```
CREATE PROCEDURE summary_order_status @StatusID int
AS
SELECT FactSalesOrder.OrderID,
       DimCustomer.CustomerName,
       DimProduct.ProductName,
       FactSalesOrder.Quantity,
       DimStatusOrder.StatusOrder
FROM FactSalesOrder
INNER JOIN DimCustomer ON FactSalesOrder.CustomerID = DimCustomer.CustomerID
INNER JOIN DimProduct ON FactSalesOrder.ProductID = DimProduct.ProductID
INNER JOIN DimStatusOrder ON FactSalesOrder.StatusID = DimStatusOrder.StatusID
WHERE DimStatusOrder.StatusID = @StatusID;
GO
```

The code is enclosed in a red box labeled with a red '1'.

Membuat  
Query untuk  
Stored  
Procedure

The screenshot shows the execution of the stored procedure. The command bar contains the text 'EXEC summary\_order\_status @StatusID = 2', which is circled in red and labeled with a red '3'. Below the command bar, the 'Results' tab is active, displaying a table with the following data:

	OrderID	CustomerName	ProductName	Quantity	StatusOrder
1	1301	LIA RAHMAWATI	Converse Cap Original	2	Awaiting Shipment
2	1304	AJENG SRIASIH	T-Shirt Polo Nevada	2	Awaiting Shipment
3	1307	RAHMA AMELIA	Pull & Bear T-Shirt	1	Awaiting Shipment

The results table is circled in red and labeled with a red '4'.

Mencoba  
eksekusi  
Stored  
Procedure

# **Insert Your Link Github Here**

<https://github.com/baramizzo58/VIX-IDX-Partners-Data-Engineer>



# Video Presentation Here

<https://drive.google.com/drive/folders/1awwE310vtNZPavtM4ZZWq7NBgFYg3J05?usp=sharing>

# Thank You



**Rakamin**  
Academy



**id/x**

partners