

Data Warehouse & ETL Implementation

ID/X Partners Data Engineer Virtual Internship Program

Presented by
Muhammad Alvero Johansyah

Experience & Projects:

- Computer Lab Assistant on computing field in Department of Mathematics, Universitas Indonesia
- Assistant Lecturer of Numerical Method in Department of Mathematics, Universitas Indonesia
- [Project: Classification of Cardiotocography to Model the Diagnosis of Problems in the Fetus using the Artificial Neural Network \(ANN\) Model with the Keras Library and Multilayer Perceptron \(MLP\) and the Decision Tree Model.](#)
- [Project: Handling Missing Value on Online Administration Universities form using Mean, Median, & KNN Method.](#)

<http://www.linkedin.com/in/muhammadalvero>



Muhammad Alvero Johansyah

About Me:

A third-year Undergraduate Mathematics Student at the Universitas Indonesia with a passion for data science. Experienced as a computer laboratory assistant in the Department of Mathematics, I have a basic understanding and experience in the field of data science with a few project experience. I have a keen interest in Programming Language Scripting and SQL Operation, and I am enthusiastic about applying these skills to solve complex data problems.

Case Study



Rakamin
Academy



id/x

partners

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 harus dilakukan:

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*.**
4. **Membuat *Store Procedure* (SP) untuk menampilkan summary sales order berdasarkan status pengiriman.**

Tools yang digunakan adalah SQL Server Management Studio (SMSS) dan Talend. Selain itu, ada File staging.bak yang merupakan file backup database yang terlampir.

Result of the Task & Completion of the Case Study

Link Folder:

<https://drive.google.com/drive/folders/1UbQgCPNWntIxd1qKAUbqDwkkiWDhxxxk?usp=sharing>



Rakamin
Academy



id/x

partners

Task 1

Melakukan Import/Restore Database Staging.

Restore database - localhost

General Files Options

Source

Restore from: Database

Database: Staging

Destination

Target database: Staging

Restore to: The last backup taken (Saturday, 11 March 2023 07:53:33)

Restore plan

Backup sets to restore

R...	Name	Type	Comp...	Server	Datab...	Position	First L...	Last L...	Full LSN	Check...	Start ...	Finish ...	Size	User N...	Expirat...	Id
<input checked="" type="checkbox"/>	Staging...	Database	Full	HANIF...	Staging	1	37000...	37000...	0	37000...	11/03/2...	11/03/2...	3889008	sa		5b148e...

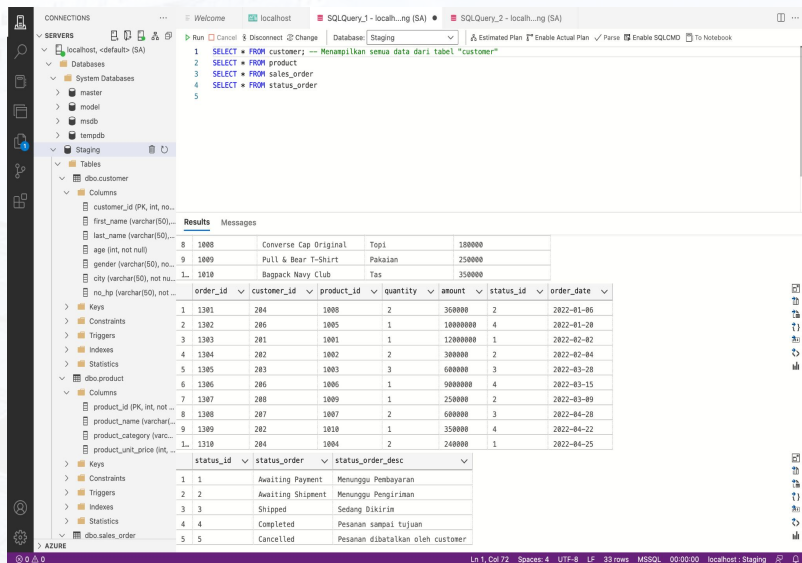
CONNECTIONS

SERVERS

- localhost, <default> (SA)
 - Databases
 - System Databases
 - master
 - Tables
 - System Tables
 - dbo.MSreplication_optio...
 - dbo.spt_fallback_db
 - dbo.spt_fallback_dev
 - dbo.spt_fallback_usg
 - dbo.spt_monitor
 - Views
 - Synonyms
 - Programmability
 - External Resources
 - Service Broker
 - Storage
 - Security
 - model
 - msdb
 - tempdb
 - Security
 - Server Objects

Task 1

Melakukan Import/Restore Database Staging.

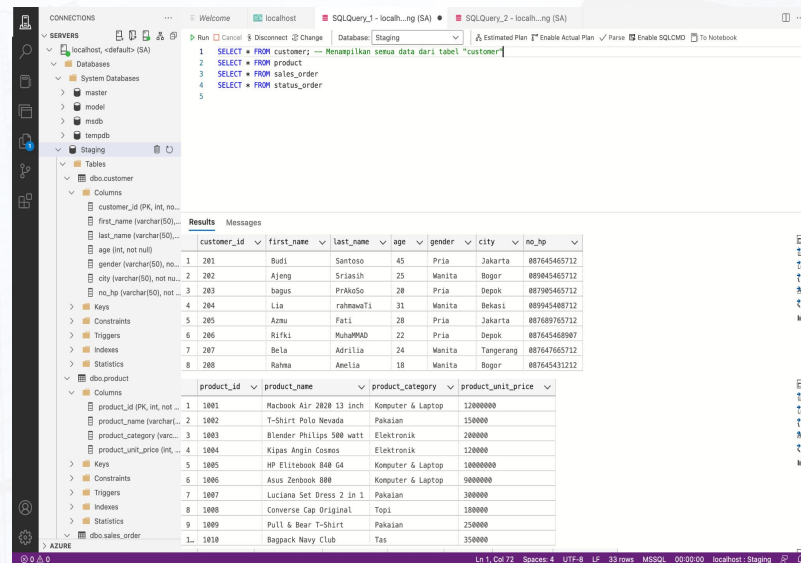


SQLQuery_1 - localh...ng (SA) • SQLQuery_2 - localh...ng (SA)

```
1 SELECT * FROM customer; -- Menampilkan semua data dari tabel "customer"
```

order_id	customer_id	product_id	quantity	amount	status_id	order_date	
1	1301	284	1008	2	350000	2	2022-01-05
2	1302	286	1005	1	10000000	4	2022-01-08
3	1303	201	1001	1	12000000	1	2022-02-02
4	1304	282	1002	2	300000	2	2022-02-04
5	1305	283	1003	3	600000	3	2022-03-28

status_id	status_order	status_order_desc
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 oleh customer



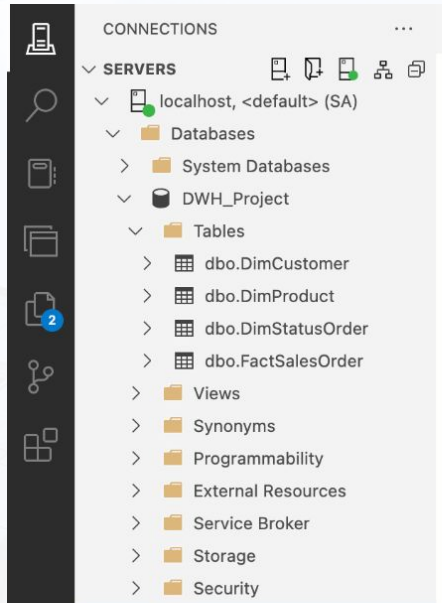
SQLQuery_1 - localh...ng (SA) • SQLQuery_2 - localh...ng (SA)

```
1 SELECT * FROM customer; -- Menampilkan semua data dari tabel "customer"
```

product_id	product_name	product_category	product_unit_price
1	Macbook Air 2020 13 inch	Komputer & Laptop	12000000
2	T-Shirt Polo Nevada	Pakaian	150000
3	Blender Philips 500 watt	Elektronik	200000
4	Kipas Angin Cosmes	Elektronik	120000
5	HP Elitebook 840 G4	Komputer & Laptop	10000000
6	Asus Zenbook 880	Komputer & Laptop	9000000
7	Luciana Set Dress 2 in 1	Pakaian	300000
8	Converse Cap Original	Pakaian	100000
9	Pull & Bear T-Shirt	Pakaian	250000
10	Bagpack Navy Club	Tas	350000

Task 2

Membuat sebuah database bernama DWH_Project, serta membuat tabel fact dan dimension dari tabel yang ada di database staging.



```
Run Cancel Disconnect Change Database: DWH_Project
1  -- Membuat database DWH_Project
2  CREATE DATABASE DWH_Project;
3
4  -- Menggunakan database yang telah dibuat
5  USE DWH_Project;
6
7  CREATE TABLE DimCustomer (
8      CustomerID INT PRIMARY KEY,
9      FirstName NVARCHAR(255),
10     LastName NVARCHAR(255),
11     Age INT,
12     Gender NVARCHAR(1),
13     City NVARCHAR(255),
14     NoHP NVARCHAR(20)
15 );
16
17 CREATE TABLE DimProduct (
18     ProductID INT PRIMARY KEY,
19     ProductName NVARCHAR(255),
20     ProductCategory NVARCHAR(255),
21     ProductUnitPrice DECIMAL(18, 2)
22 );
23
24 CREATE TABLE DimStatusOrder (
25     StatusID INT PRIMARY KEY,
26     StatusOrder NVARCHAR(255),
27     StatusOrderDesc NVARCHAR(255)
28 );
29
30 CREATE TABLE FactSalesOrder (
31     OrderID INT PRIMARY KEY,
32     CustomerID INT,
33     ProductID INT,
34     Quantity INT,
35     Amount DECIMAL(18, 2),
36     StatusID INT,
37     OrderDate DATE,
38     FOREIGN KEY (CustomerID) REFERENCES DimCustomer(CustomerID),
39     FOREIGN KEY (ProductID) REFERENCES DimProduct(ProductID),
40     FOREIGN KEY (StatusID) REFERENCES DimStatusOrder(StatusID)
41 );
```