

Data Warehouse and ETL Implementation

Data Engineer

Presented by
Muhammad Naufal Farrazqi



**MUHAMMAD
NAUFAL
FARRAZQI**

ABOUT ME

I DO WHAT I LOVE AND I LOVE WHAT I DO

BACKGROUND

- **UGM**
Computer Science Student
- **INTEREST**
Machine Learning,
Robotics & Mechatronics
- **HOBBY**
Photo - Videography

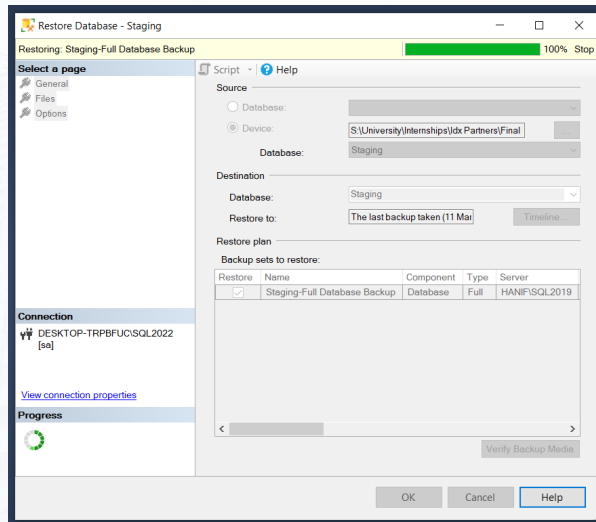
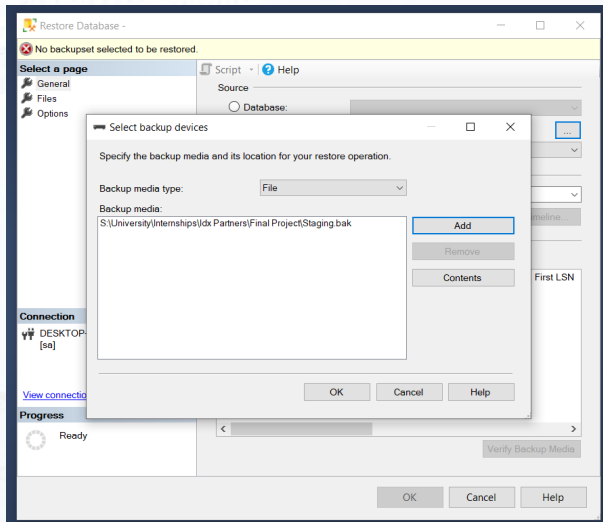
Case Study

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 . Untuk memudahkan proses pengerjaan task tersebut , bacalah bagian 'Prerequisite dan Hints' untuk informasi lebih lengkap nya.

Result

1. Melakukan Import/Restore Database Staging .

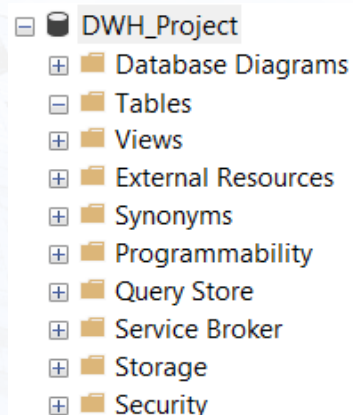
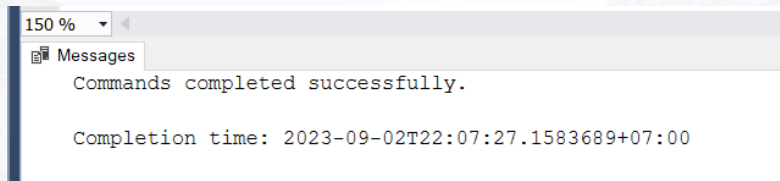
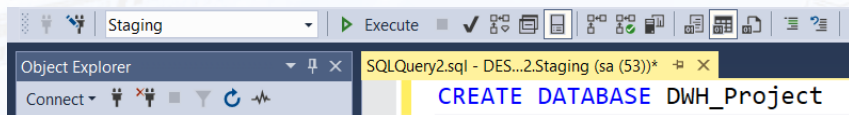


Penjelasan:

Pada SQL Server Management Studio untuk Restore Database, kita click kanan di folder “Databases” lalu pilih opsi “Restore Database”.

Result

2. Membuat sebuah Database bernama DWH_Project , serta membuat Tabel Fact dan Dimension dari tabel yang ada di database Staging .



Penjelasan: 1

Sekarang kita buat Query baru untuk menulis query CREATE DATABASE DWH_Project

Result

2. Membuat sebuah Database bernama DWH_Project , serta membuat Tabel Fact dan Dimension dari tabel yang ada di database Staging .

```
USE DWH_Project

CREATE TABLE DimCustomer (
  IDCustomer int NOT NULL,
  CustomerName varchar(50) NOT NULL,
  Age int NOT NULL,
  Gender varchar(50) NOT NULL,
  City varchar(50) NOT NULL,
  No_hp varchar(50) NOT NULL,
  CONSTRAINT PK_DimCustomer PRIMARY KEY (IDCustomer)
)

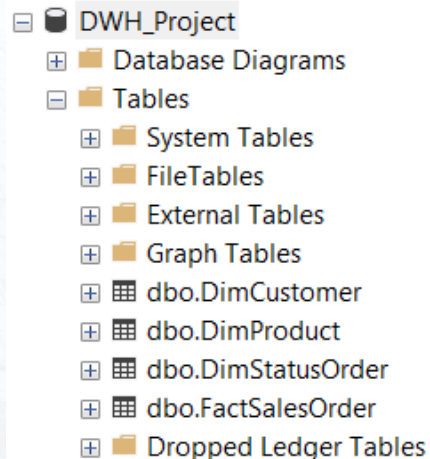
CREATE TABLE DimProduct (
  IDProduct int NOT NULL,
  Product_name varchar(255) NOT NULL,
  Product_category varchar(255) NOT NULL,
  Product_unit_price int NOT NULL,
  CONSTRAINT PK_DimProduct PRIMARY KEY (IDProduct)
)

CREATE TABLE DimStatusOrder (
  IDStatus int NOT NULL,
  Status_order varchar(50) NOT NULL,
  Status_order_desc varchar(50) NOT NULL,
  CONSTRAINT PK_DimStatusOrder PRIMARY KEY (IDStatus)
)
```

Penjelasan: 2

Untuk membuat Fact dan Dimension Table dari database Staging. Kita tulis query seperti disamping

```
CREATE TABLE FactSalesOrder (
  IDOrder int NOT NULL,
  IDCustomer int NOT NULL,
  IDProduct int NOT NULL,
  IDStatus int NOT NULL,
  Quantity int NOT NULL,
  Amount int NOT NULL,
  Order_date date NOT NULL,
  CONSTRAINT PK_FactSalesOrder PRIMARY KEY (IDOrder),
  CONSTRAINT FK_CustomerOrder FOREIGN KEY (IDCustomer) REFERENCES DimCustomer (IDCustomer),
  CONSTRAINT FK_ProductOrder FOREIGN KEY (IDProduct) REFERENCES DimProduct (IDProduct),
  CONSTRAINT FK_StatusOrder FOREIGN KEY (IDStatus) REFERENCES DimStatusOrder (IDStatus)
)
```



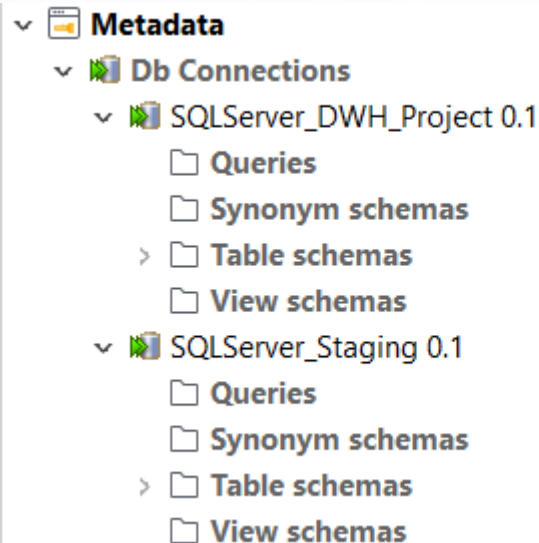
Result

3. Membuat Job ETL di aplikasi talend untuk memindahkan data dari Staging ke Data Warehouse .

Penjelasan: 1

Sekarang kita buka tools Talend lalu ke subsection “Db Connections” dari section “Metadata”. Untuk memindahkan database dari SQL Management Studio, kita click kanan “Db Connections” dan pilih opsi “Create Connection”.

Lalu kita membuat SQL Servers yang dibutuhkan, disini SQLServer_Staging dan SQLServer_DWH_Project



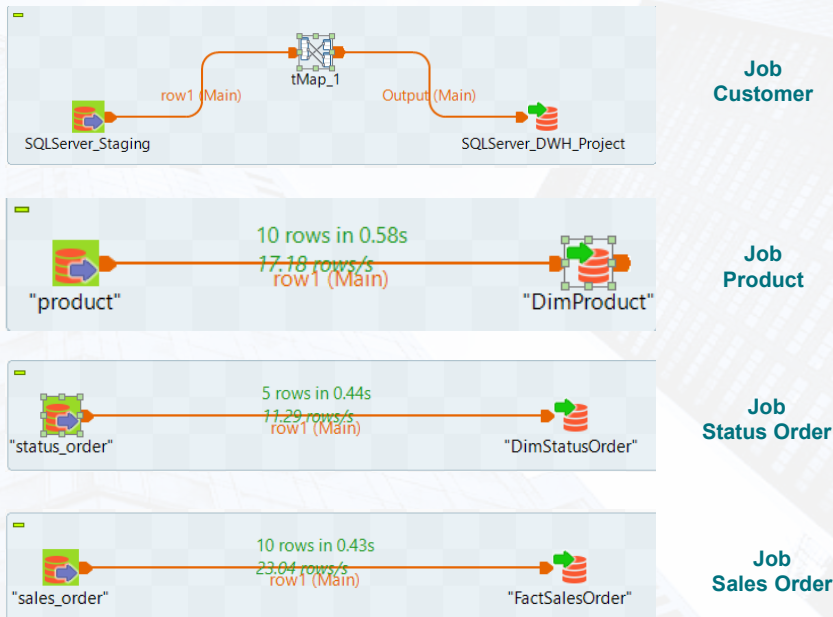
Result

3. Membuat Job ETL di aplikasi talend untuk memindahkan data dari Staging ke Data Warehouse .

Penjelasan: 2

Setelah itu untuk memuat schema database dari Management Studio. Kita klik kanan SQL Server Connection yang sudah kita buat lalu pilih opsi "Retrieve Schema". Nanti kita checklist bagian "dbo" untuk mengambil tables dari database tersebut.

Lalu, kita membuat job sesuai kebutuhan kita memindahkan database, di project ini kita buat 4 job seperti disamping:

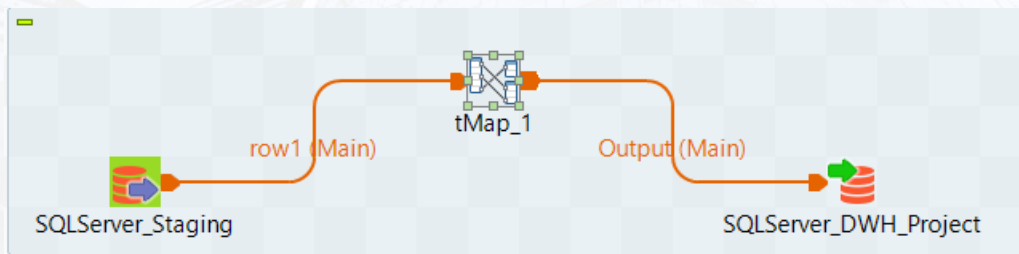


Result

3. Membuat Job ETL di aplikasi talend untuk memindahkan data dari Staging ke Data Warehouse .

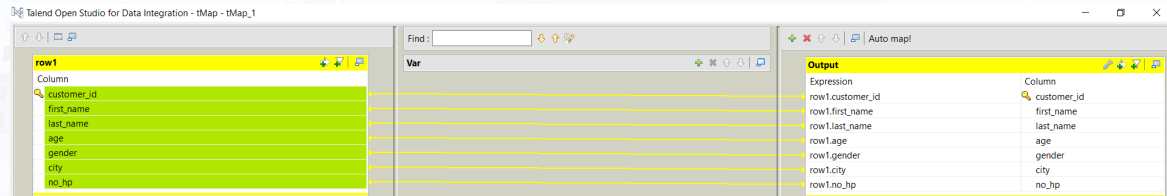
Penjelasan: 4

Lalu kita diminta untuk menggabungkan firstname dan lastname customer dan juga uppercase semua letternya. Untuk itu, kita hubungkan SQL Server Staging dan DWH Project dengan **tMap**.



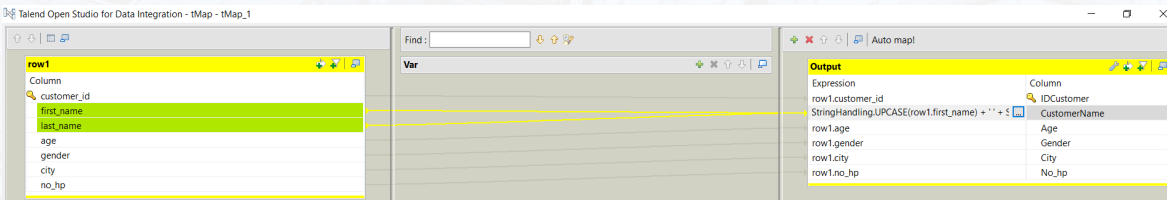
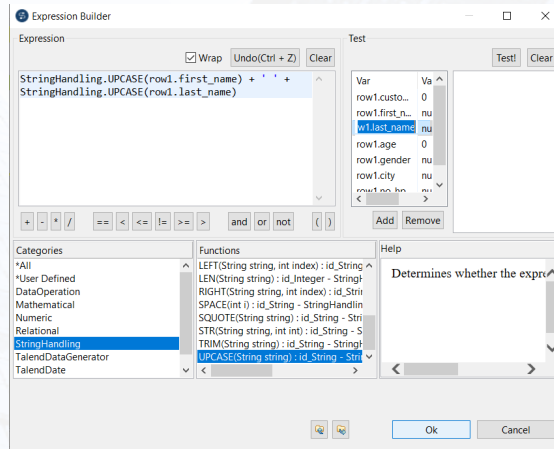
Result

3. Membuat Job ETL di aplikasi talend untuk memindahkan data dari Staging ke Data Warehouse .



Penjelasan: 5

Kita buka component map dan click "map editor". Untuk menggabungkan kita edit expression menggunakan StringHandling UPPERCASE. Kita buat dua function, satu untuk firstname yang satunya lagi untuk lastname, lalu kita gabungkan menggunakan + dan ' ' sebagai Space.



Result

3. Membuat Job ETL di aplikasi talend untuk memindahkan data dari Staging ke Data Warehouse .

	IDProduct	Product_name	Product_category	Product_unit_price
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
9	1009	Pull & Bear T-Shirt	Pakaian	250000
10	1010	Bagpack Navy Club	Tas	350000

DimProduct

	IDOrder	IDCustomer	IDProduct	IDStatus	Quantity	Amount	Order_date
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
7	1307	208	1009	1	250000	2	2022-03-09
8	1308	207	1007	2	600000	3	2022-04-28
9	1309	202	1010	1	350000	4	2022-04-22
10	1310	204	1004	2	240000	1	2022-04-25

FactSalesOrder

Penjelasan: 6

Isi dari tables di DWH Project setelah menjalankan Job ETL di tools Talend

	IDCustomer	CustomerName	Age	Gender	City	No_hp
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 ADRIILA	24	Wanita	Tangerang	087647665712
8	208	RAHMA AMELIA	18	Wanita	Bogor	087645431212

DimCustomer

	IDStatus	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

DimStatusOrder

Result

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

Penjelasan: 1

Untuk membuat stored procedure summary sales order, kita tulis Query dibawah. Setelah membuat stored procedure summary sales order, kita run/execute Stored Procedurenya sesuai input IDStatus yang ingin kita input:

```
SQLQuery3.sql - DES...H_Project (sa (63))*    SQLQuery5.sql - DES...H_Project (sa (52))*
CREATE OR ALTER PROCEDURE summary_order_status
(@StatusID varchar(50)) AS
BEGIN
    SELECT
        a.IDOrder,
        b.CustomerName,
        d.Product_name,
        a.Quantity,
        c.Status_order
    FROM FactSalesOrder a
    INNER JOIN DimCustomer b on a.IDCustomer = b.IDCustomer
    INNER JOIN DimStatusOrder c on a.IDStatus = c.IDStatus
    INNER JOIN DimProduct d on a.IDProduct = d.IDProduct
    WHERE a.IDStatus = @StatusID
END
```



SQLQuery3.sql - DES...H_Project (sa (63))* SQLQuery5.sql - DES...H_Project (sa (52))*

```
EXEC summary_order_status @StatusID = 1
EXEC summary_order_status @StatusID = 2
EXEC summary_order_status @StatusID = 3
EXEC summary_order_status @StatusID = 4
EXEC summary_order_status @StatusID = 5
```

100 %

Results Messages

	IDOrder	CustomerName	Product_name	Quantity	Status_order
1	1302	RIFKI MUHAMMAD	HP Elitebook 840 G4	10000000	Awaiting Payment
2	1303	BUDI SANTOSO	Macbook Air 2020 13 inch	12000000	Awaiting Payment
3	1306	RIFKI MUHAMMAD	Asus Zenbook 800	9000000	Awaiting Payment
4	1307	RAHMA AMELIA	Pull & Bear T-Shirt	250000	Awaiting Payment
5	1309	AJENG SRIASIH	Bagpack Navy Club	350000	Awaiting Payment

	IDOrder	CustomerName	Product_name	Quantity	Status_order
1	1301	LIA RAHMAWATI	Converse Cap Original	360000	Awaiting Shipment
2	1304	AJENG SRIASIH	T-Shirt Polo Nevada	300000	Awaiting Shipment
3	1308	BELA ADRILIA	Luciana Set Dress 2 in 1	600000	Awaiting Shipment
4	1310	LIA RAHMAWATI	Kipas Angin Cosmos	240000	Awaiting Shipment

	IDOrder	CustomerName	Product_name	Quantity	Status_order
1	1305	BAGUS PRAKOSO	Blender Philips 500 watt	600000	Shipped

	IDOrder	CustomerName	Product_name	Quantity	Status_order
--	---------	--------------	--------------	----------	--------------

	IDOrder	CustomerName	Product_name	Quantity	Status_order
--	---------	--------------	--------------	----------	--------------

Link GitHub

<https://github.com/naufarrazqi/idxFinalProject>

- Muhammad - Naufal - Farrazqi

Video Presentation

https://drive.google.com/file/d/1zuDxUcM6YJy1OSyxQNNMf7MyUxXmo7_Y/view?usp=sharing

Thank You



Rakamin
Academy



id/x partners