Rakamin Academy X id/x partners

Data Warehouse and ETL Implementation

ID/X Partners – Data Engineer

Presented by
Rahmat Satyawan

Rahmat Satyawan

Bit.ly/rahmatst

Rakamin Academy

Education and Experience

- Education
  STTN BATAN 2014-2018
  Electronics Instrumentation – Nuclear Technophysics GPA 3.63/4.00

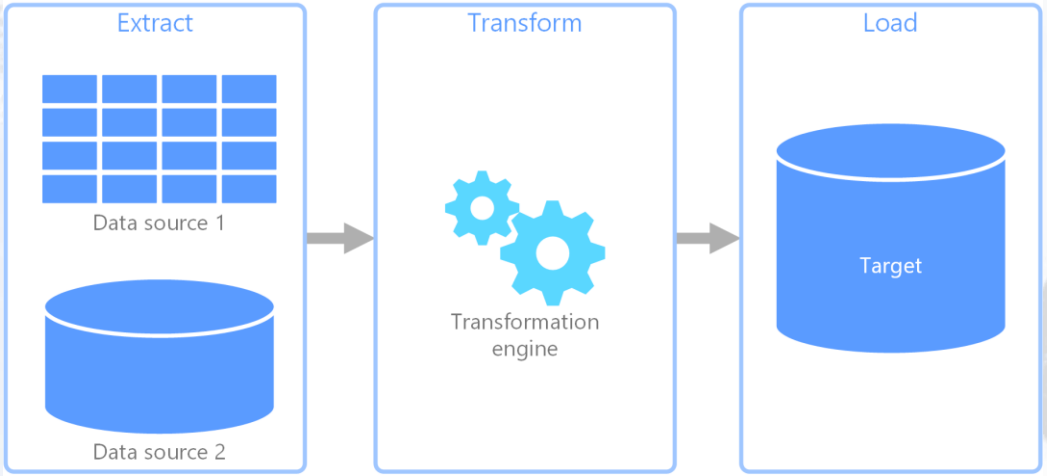
- Experience
  2+ in Nuclear industry as NDT Inspector, Engineer and Radiation Safety Officer, Currently switching to Data and Web industry

Case Study

Rakamin Academy

One of the clients of ID/X Partners, operating in the e-commerce sector, has a need to create a Data Warehouse sourced from several tables in the source database. This Data Warehouse will consist of one Fact table and several Dimension tables.

Case Study



Restore	ETL	Target
Database	Concat Column	Database
Staging.bak	Rename Column	DWH_Project
SQL Server + SSMS	**Talend Open Studio**	**Talend Open Studio + SSMS**

Prerequisite

Rakamin Academy

1. Microsoft SQL Server (Link Here)
2. SSMS SQL Server Management Studio (Link Here)
3. Talend Open Studio for Data Integration (Link Here)
4. Database Staging (Link Here)

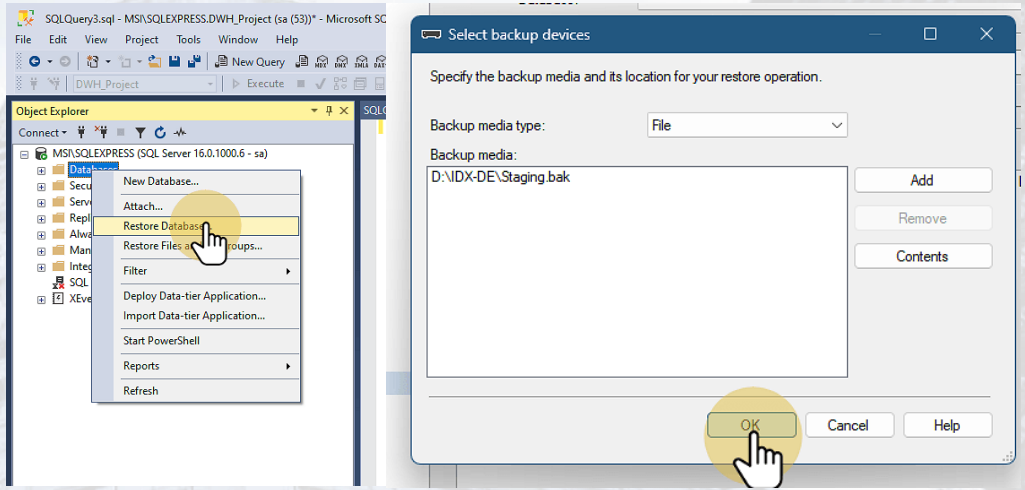
Process

Rakamin Academy

1. Import/Restore Database **Staging.bak** ke SQL Server + SSMS
2. Create New Database **DWH_Project** with table and attribute as the rule of challenge
3. Create project job ETL in Talend Open Studio from **Staging.bak** to target **DWH_Project** with detail Transformation
4. Create Store Procedure to show summary sales order base on status order in SSMS from **DWH_Project** Database that already processed from ETL

Process

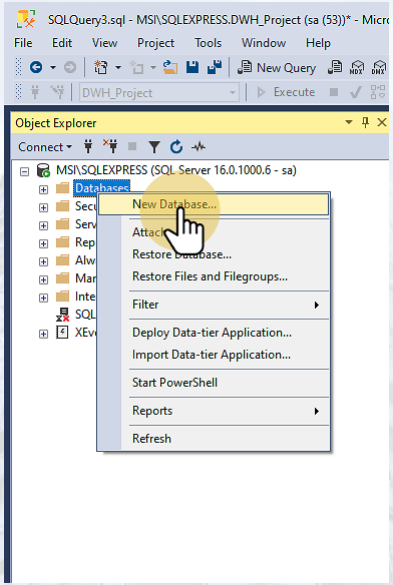
1. Import/Restore Database **Staging.bak** ke SQL Server + SSMS
    1. Right click on Database > Restore Database
    2. Click Device and select Staging.bak file in computer
    3. Click OK then if success will be notify Database restored successfully

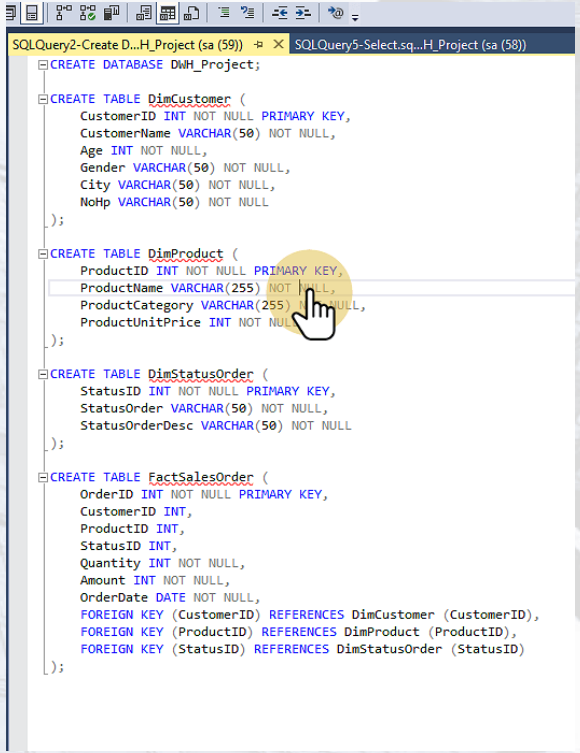


Process

2. Create New Database **DWH_Project** with table and attribute as the rule of challenge

1. Right click on Database > New Database
2. Insert Database Name with **DWH_Project**, then click OK
3. Right click on Database DWH_Project > New Query to add table and column Or you can make new database with query CREATE DATABASE in SSMS





Process

3. Create project job ETL in Talend Open Studio from **Staging.bak** to target **DWH_Project** with detail Transformation

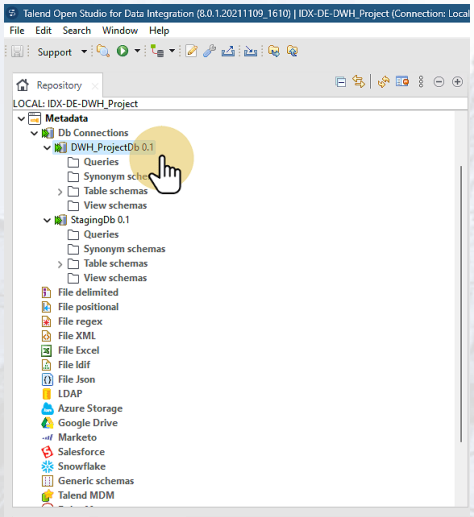
1. Import Database Staging as Input and Database DWH_Project as Output
2. Create job per table (Customer, Product, Sales Order, Sales Status)
3. Create job with
   - Input > DBInput Staging
   - Transform > tMaps
   - Output > DBOutput DWH_project

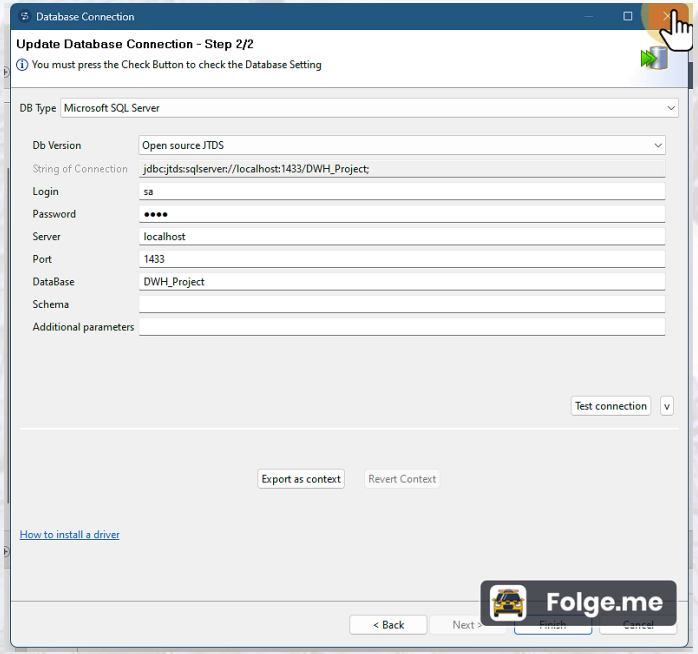
And Run all Job and see the result in SSMS with SELECT Query

Process

3. Create project job ETL in Talend Open Studio from **Staging.bak** to target **DWH_Project** with detail Transformation

    1. Import Database Staging as Input and Database DWH_Project as Output

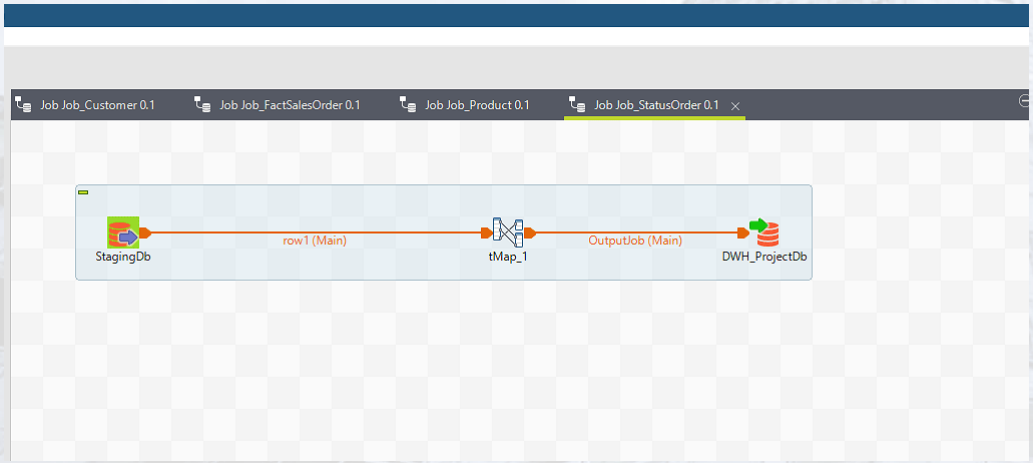


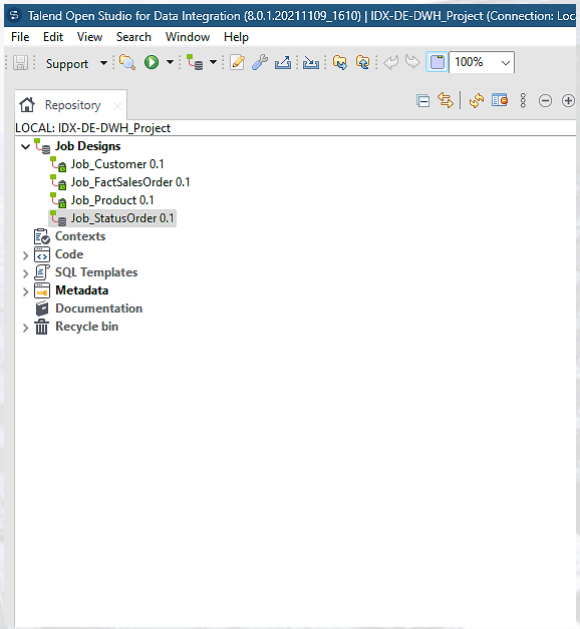


Process

3. Create project job ETL in Talend Open Studio from **Staging.bak** to target **DWH_Project** with detail Transformation

2. Create job per table (Customer, Product, Sales Order, Sales Status)



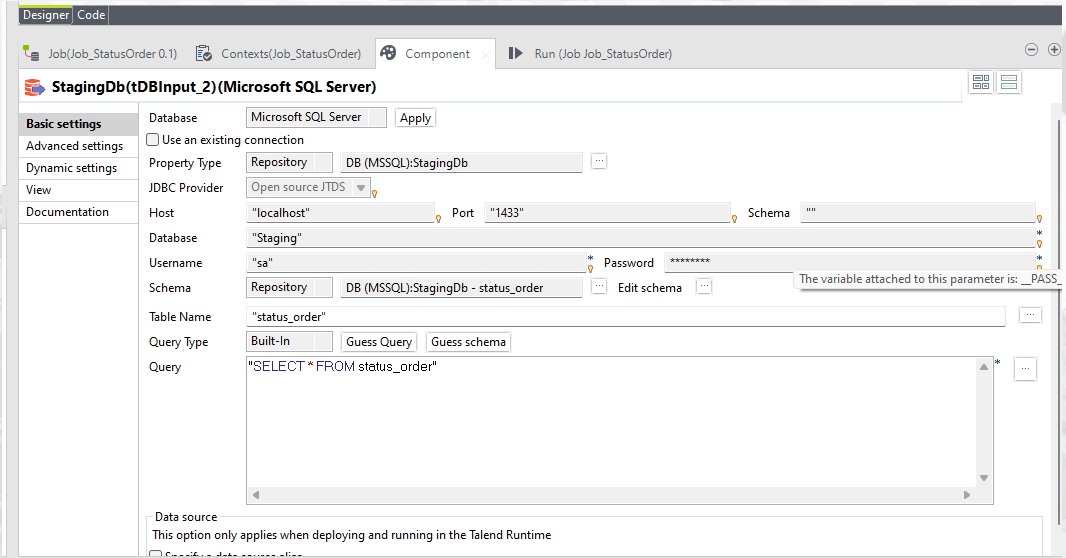


Process

3. Create project job ETL in Talend Open Studio from **Staging.bak** to target **DWH_Project** with detail Transformation

1. Create job (Input > DBInput Staging, Transform > tMaps Output > DBOutput DWH_project)

   For Input Configuration Ex. Table Status Order

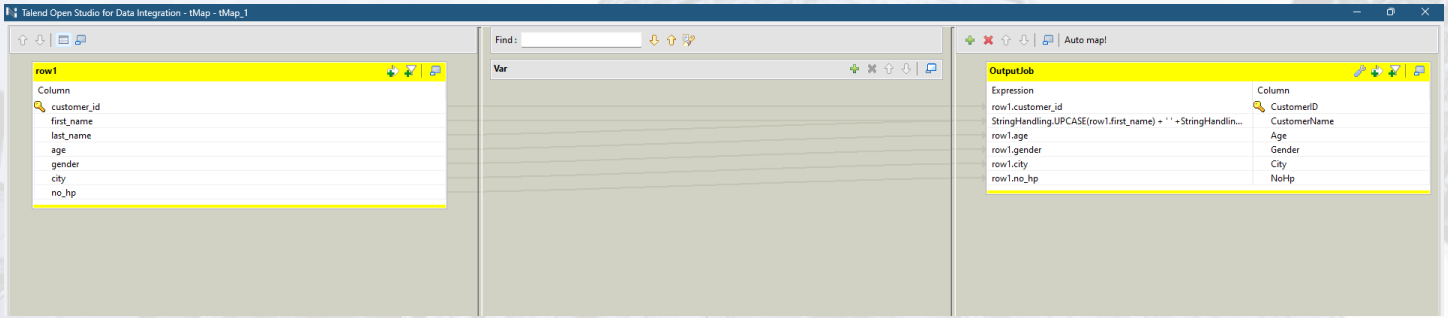


Process

3. Create project job ETL in Talend Open Studio from **Staging.bak** to target **DWH_Project** with detail Transformation

1. Create job (Input > DBInput Staging, Transform > tMaps Output > DBOutput DWH_project)

   For tMaps Configuration Ex. Table Customer



Process

3. Create project job ETL in Talend Open Studio from **Staging.bak** to target **DWH_Project** with detail Transformation

Result in SSMS

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

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

Query executed successfully.

Process

4. Create Store Procedure to show summary sales order base on status order in SSMS from DWH_Project Database that already done in ETL

1. Right click on Database DWH_Project > New Query
2. Write the Store Procedure as it is
3. Run the SP with **EXEC summary_order_status @StatusID = 2**
4. The result will be shown in table below the query

```sql
CREATE PROCEDURE summary_order_status
(@StatusID int)
AS
BEGIN
    SELECT
        f.OrderID,
        c.CustomerName,
        p.ProductName,
        f.Quantity,
        s.StatusOrder
    FROM FactSalesOrder AS f
    JOIN DimCustomer c ON f.CustomerID = c.CustomerID
    JOIN DimProduct p ON f.ProductID = p.ProductID
    JOIN DimStatusOrder s ON f.StatusID = s.StatusID
    WHERE s.StatusID = @StatusID
END

EXEC summary_order_status @StatusID = 2
```

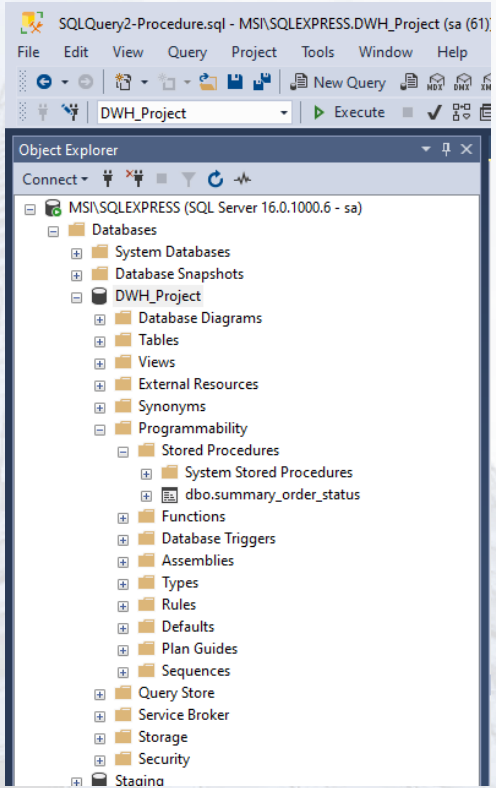
Results | Messages

	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

Process

4. Create Store Procedure to show summary sales order base on status order in SSMS from DWH_Project Database that already done in ETL

   1. You can see the Store Procedure will be saved in Database > Programmability > Stored Procedures > Your SP in this case **dbo.summary_order_status**



Result

Rakamin Academy

1. New Database DWH_Project (Backup with SSMS DWH_Project.bak) as result the ETL from staging.bak
2. The Pipeline Process ETL for Staging.bak Database with Talend include function input, tMaps, and output database SQL Server
3. Stored Procedure **dbo.summary_order_status** to know the summay from order status

Result

1. Link Video Presentation (Link Here)
2. Link Github (Link Here)

Thank You

Rakamin Academy X id/x partners