

January 2024

Online Shopping Store

Business Intelligence | Project
Report

Prepared by:
Mohamed Kadhem Jamii
Firas Amara

About Us

Our e-commerce platform is dedicated to offering a wide range of products, with a particular emphasis on skincare items. Our primary goals involve optimizing operational efficiency, maximizing profitability, and ensuring a satisfying customer experience. To achieve these objectives, we are committed to reducing order delivery times, introducing a comprehensive product rating system driven by customer reviews, analyzing order trends to identify high-demand countries, and examining the distribution of orders across genders. These strategic initiatives are aimed at refining the overall shopping experience, improving product quality across various categories, and customizing our offerings to cater to the diverse preferences of our broad customer base.

Phase 01: Data Gathering

We have four data sources with three different data types:

- Customer database as a text file
- Products database as a csv file
- Orders database as an xls file
- Order_item database as an xls file

Phase 02: Data Preparation\ ETL process\Modeling and DWH creation

Creating a Talend Studio Project

• **Customer Table:**

1. Customer Table Job Creation:

Begin by creating a job for the Customer Table named "customer":

- Navigate to the Repository, right-click on Job Designs, and select Create Standard Job.
- In the Name field, enter "customer" and click Finish.

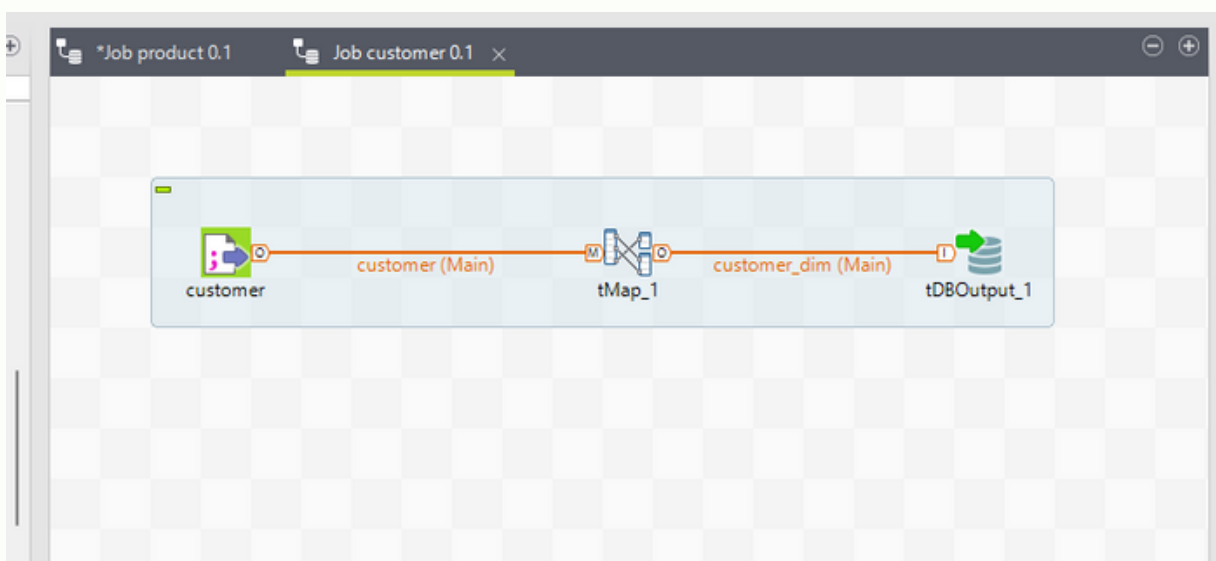
2.Data Joining using tMap Component:

- Import the customer table by dragging and dropping it into the workplace using the tFileInputDelimited component.
- Add a tMap component to the job.
- Right-click on the customer component, go to Row > Main, and click on the tMap component to establish a link, renaming it as "customer."

3.Output Configuration using tDBOutput Component:

- Drag and drop the tDBOutput component to the workplace.
- Right-click on the tMap component and link it to the tDBOutput component by selecting Row > New Output (Main), renaming the link as "customer_dim."
- Click on the tDBOutput component and set the database type to MYSQL.

By following these steps, you will efficiently create a job to handle the Customer Table, perform data joining using the tMap component, and configure the output using the tDBOutput component with specific attention to maintaining the order provided.



Entering the Customer Output table:

1.Setting up Output Table for Customer Data:

- Begin by accessing the tMap component configuration window through a double-click.
- Click on the right side to add a new output table and name it "customer_dim."

2 .Mapping Columns from Input to Output:

- Select essential columns like customer_ID, Country, Gender, Personal_ID, Customer_Name, Customer_FirstName, Customer_LastName, Birth_Date, Customer_Address, Street_Number from the "customer" input table.
- Drag and drop these chosen columns into the "customer_dim" output table, ensuring they are appropriately renamed following MYSQL server conventions.

3.Creating Address Column:

- Introduce a new column (Type: string) in the output table by using the green + icon.
- Combine the Customer_Address and Street_Number columns from the "customer" table to form the address field expression.

4.Generating Age Column:

- Add another column (Type: int) to the output table using the green + icon.
- Calculate the age by referencing the Birth_Date column from the "customer" table without delving into specific code details.

The screenshot displays the Talend Open Studio interface for mapping columns from an input table 'customer' to an output table 'customer_dim'.

customer Table Columns:

Column
Customer_ID
Country
Gender
Personal_ID
Customer_Name
Customer_FirstName
Customer_LastName
Birth_Date
Customer_Address
Street_ID
Street_Number

customer_dim Table Columns:

Column
id
country
gender
customer_name
customer_fname
customer_lname
birth_date
Var.var1
Var.var2

Var Table:

Expression	Type	Variable
customer.Customer_Address	String	var1
Mathematical.INT(TalendDate.fo...	int	var2

Schema Editor:

customer

Column	Key	Type	N..	Date Pattern (...)	Length	Precision	Defa...	Comm...
Customer_ID	<input checked="" type="checkbox"/>	Float	<input checked="" type="checkbox"/>		5	3		
Country	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		4	0		
Gender	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		3	0		
Personal_ID	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		0	0		

customer_dim

Column	Key	Type	N..	Date Pattern (...)	Length	Precision	Defa...	Comm...
id	<input checked="" type="checkbox"/>	Float	<input checked="" type="checkbox"/>		5	3		
country	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		4	0		
gender	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		3	0		
customer_name	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		33	0		

Buttons: Apply, Ok, Cancel

Connecting to the DB:

After filling the settings of the tDBOutput, and connecting to the Database.

tDBOutput_1(MySQL)

Basic settings

Database: MySQL [Apply]

Property Type: Repository DB (MYSQL):project

DB Version: Mysql 5

☐ Use an existing connection

Host: "localhost" * Port: "3306"

Database: "project"

Username: "root" * Password: "*****"

Table: "customer_dim"

Action on table: Default Action on data: Insert or update

Schema: Built-In Edit schema Sync columns

Data source

This option only applies when deploying and running in the Talend Runtime

☐ Specify a data source alias

☐ Die on error

We execute the job, then, we refresh MySQL DB to observe the changes made.

Server: 127.0.0.1 » Database: project » Table: customer_dim

Showing rows 0 - 24 (88686 total, Query took 0.0122 seconds.)

SELECT * FROM `customer_dim`

Options: 1 > >> Number of rows: 25 Filter rows: Search this table Sort by key: None

	id	country	gender	customer_name	customer_fname	customer_lname	birth_date	age	address
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	1	FR	M	Albert Collet	Albert	Collet	1940-11-24 00:00:00	83	"Square Edouard VII 1"3500101176.00
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	2	ES	F	Mercedes Martinez	Mercedes	Martinez	1955-01-15 00:00:00	68	"Edificio 2"8300101034.00
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	3	IT	M	Pier Egidio Boeris	Pier Egidio	Boeris	1970-07-01 00:00:00	53	"Via M. Di Montesole 3"4800100416.00
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	4	US	M	James Kvarniq	James	Kvarniq	0000-00-00 00:00:00	-17677	"4382 Gralyn Rd"9260106519.00
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	5	US	F	Sandrina Stephano	Sandrina	Stephano	1975-07-09 00:00:00	48	"6468 Cog Hill Cr"9260114570.00
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	6	BE	M	Rent Van Lint	Rent	Van Lint	1945-12-23 00:00:00	78	"Mispadstr 2"1900104760.00
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	7	ES	F	Julian Escorihuela Monserrate	Julian	Escorihuela Monserrate	1975-08-07 00:00:00	48	"Co. De Los Claveles 561"8300100945.00
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	8	FI	M	Aki Ivonen	Aki	Ivonen	1935-12-04 00:00:00	88	"Valimotie 5"3400100281.00
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	9	DE	F	Cornelia Krah	Cornelia	Krah	1970-02-27	53	"Kallstadterstr.

• Product Table:

1. Creating Job for Product Table - "products":

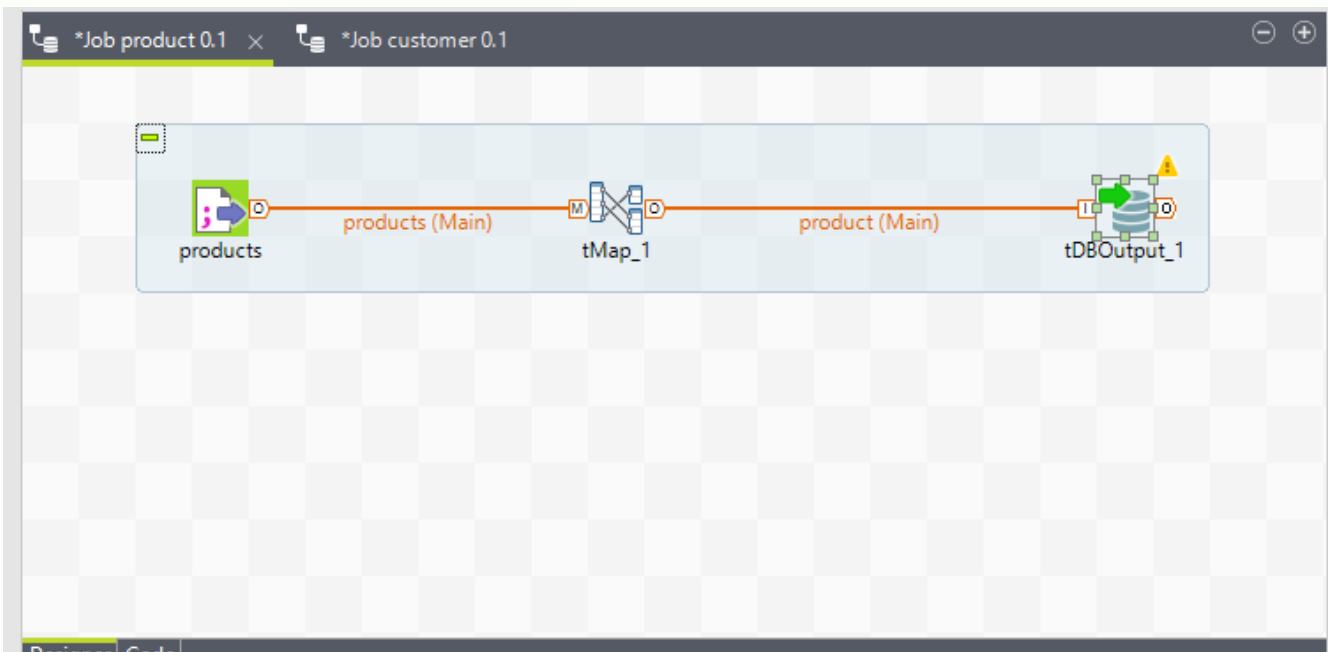
- Open the Repository, right-click on Job Designs, and select Create Standard Job.
- In the Name field, enter "products," then click Finish to create the job.

2.Data Joining with tMap Component:

- Import the "products" table by dragging and dropping it onto the workplace using the tFileInputDelimited component.
- Add a tMap component to the job.
- Right-click on the "products" component, navigate to Row > Main, and click on the tMap component to establish a link, renaming it as "products."

3.Output Configuration using tDBOutput component:

- Drag and drop the tDBOutput component onto the workplace.
- Link the tMap component to the tDBOutput component by right-clicking on the tMap, selecting Row > New Output (Main), and renaming the link as "product."
- Click on the tDBOutput component and set the database type to MYSQL.



Entering the product Output table:

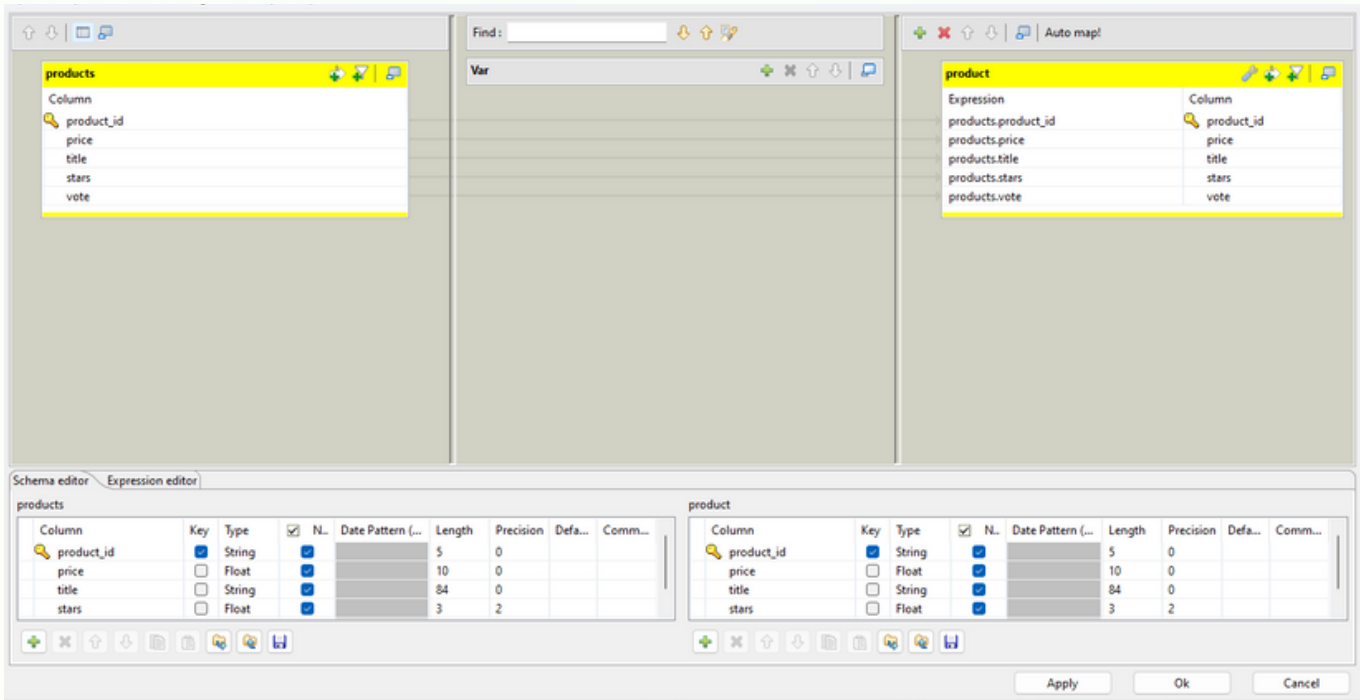
1. Configuring Output Table for Product Data:

- Begin by double-clicking the tMap component to access its configuration window.
- On the right side, click "Add output table."
- Assign the name "product" to your output table.

2. Mapping Columns from Input to Output:

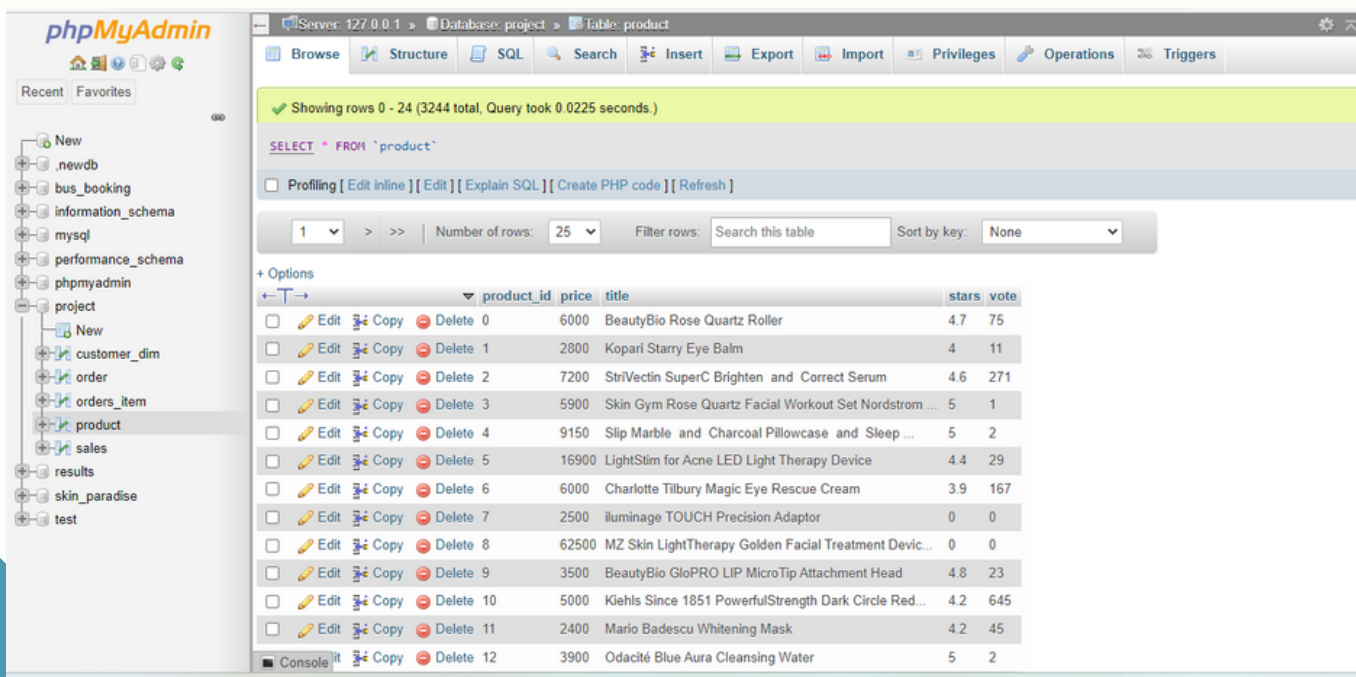
- Within the input table "products," carefully choose columns such as product_id, price, title, stars, and vote.
- Drag and drop these selected columns into the output table "product," ensuring they are renamed in accordance with the conventions used in the MYSQL server.

By following these steps, you'll effectively set up the Product Output Table within the tMap component, mapping the required columns and ensuring they are appropriately renamed as per the MYSQL server specifications.



Connecting to the DB:

After filling the settings of the tDBOutput, and connecting to the Database, we execute the job, then, we refresh MYSQL DB.



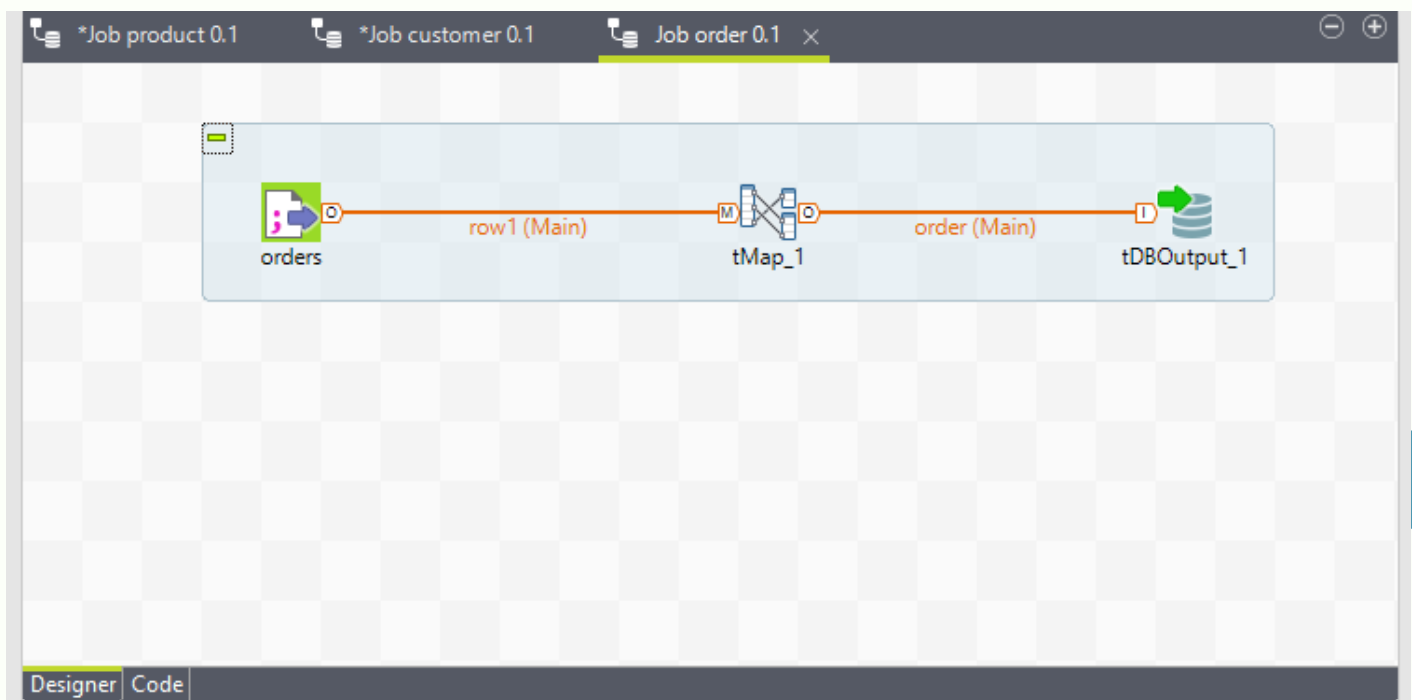
• Order Table:

1. Creating job for Order Table named order:

- In Repository, right-click Job Designs; Click Create Standard Job; In the Name field, enter a name ("orders"); Click Finish.

2. Data joining using the tMap component:

- Choose the table orders and import it by dragging-and-dropping it to the workplace using the tFileInputDelimited component.
- Add a tMap component.
- Right-click the orders component, Select Row > Main, Click on the tMap component to link the two, rename the link with "row1".
- Drag-and-drop the tDBOutput component to the workplace then right-click the tMap component to link it to the output, Select Row > *New Output* (Main) and rename the link with "order".
- Click on the tDBOutput and choose the MYSQL type of database.



Creating job for Order Table named order:

Double-click the tMap component.

You are brought to the tMap component configuration window.

On the right side of the screen, click Add output table.

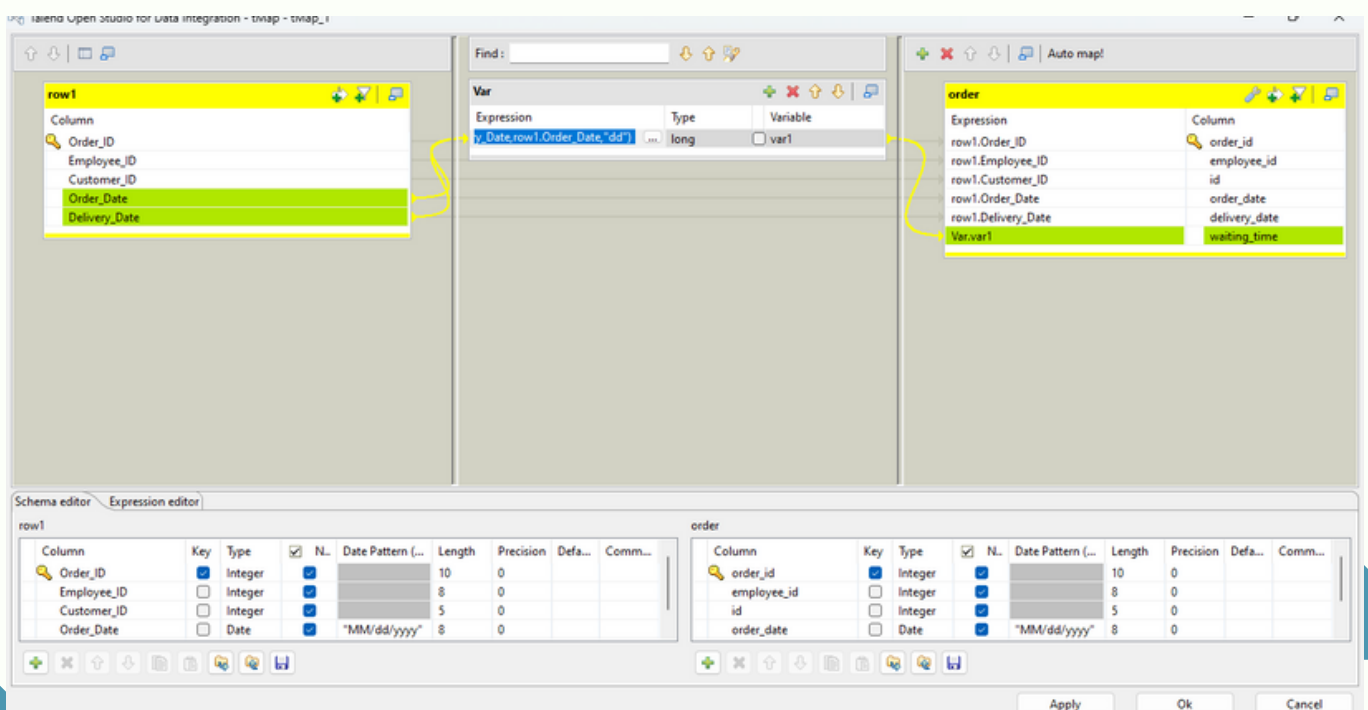
Enter a name for your output table: "order".

In input table orders, select columns Order_ID, Employee_ID, Customer_ID, Order_Date, Delivery_Date, then drag-and-drop them in output table order.

Create a waiting_time column (Type: long) in the output table by clicking on the green + icon. To fill in the field expression, we select the Order_Date and Delivery_Date columns from the orders table and drag it into the field.

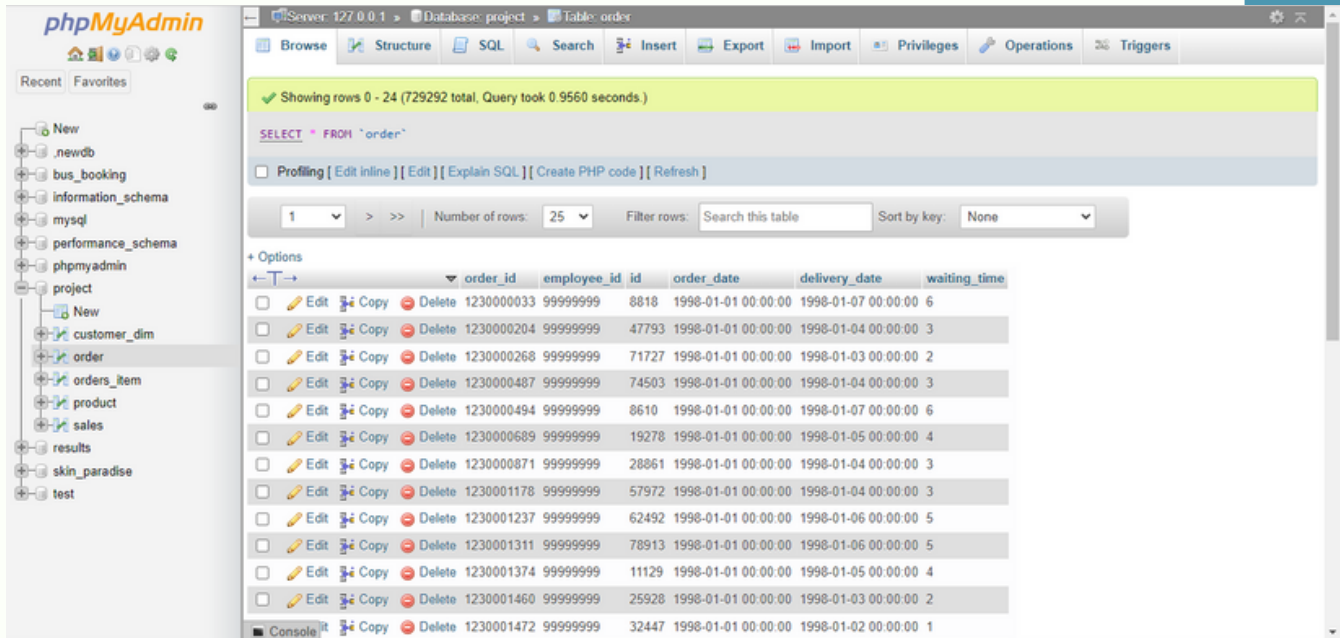
The function used for waiting_time column:

`TalendDate.diffDate(row1.Delivery_Date,row1.Order_Date,"d d")`



Connecting to the DB:

After filling the settings of the tDBOutput, and connecting to the Database, we execute the job, then, we refresh MYSQL DB to observe the changes made.



• Order_item Table:

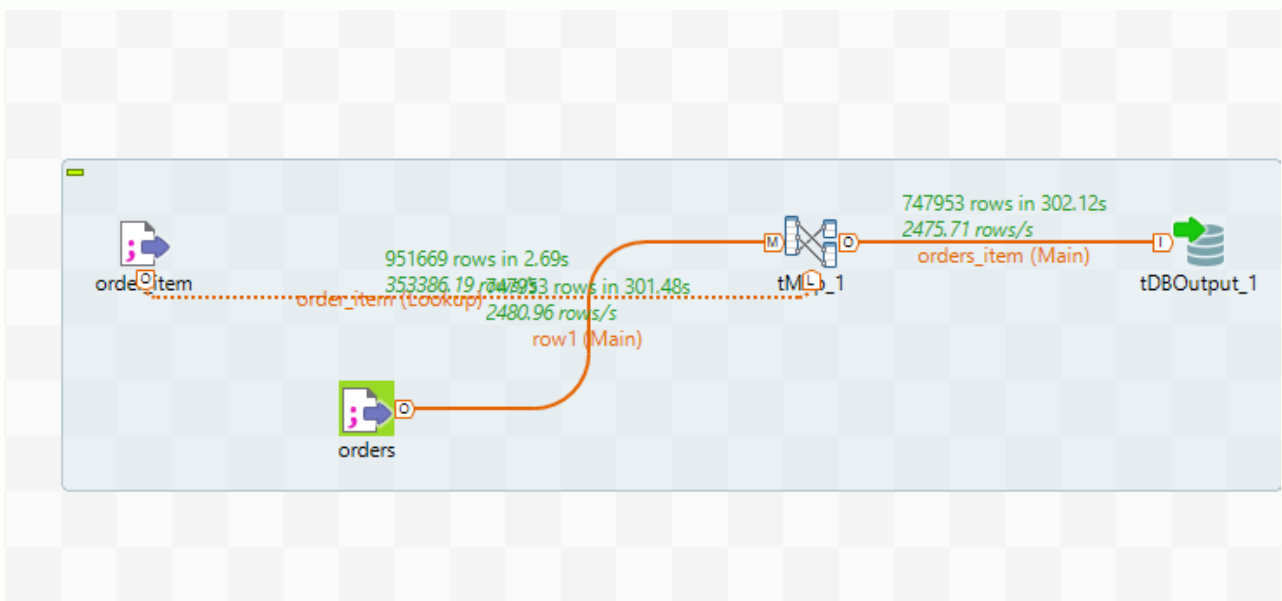
1. Creating job for Order_Item Table named order_item:

- In Repository, right-click Job Designs; Click Create Standard Job; In the Name field, enter a name ("order_item"); Click Finish.

2. Data joining using the tMap component:

- Choose the table orders and order_item and import it by dragging-and-dropping it to the workplace using the tFileInputDelimited component.
- Add a tMap component.

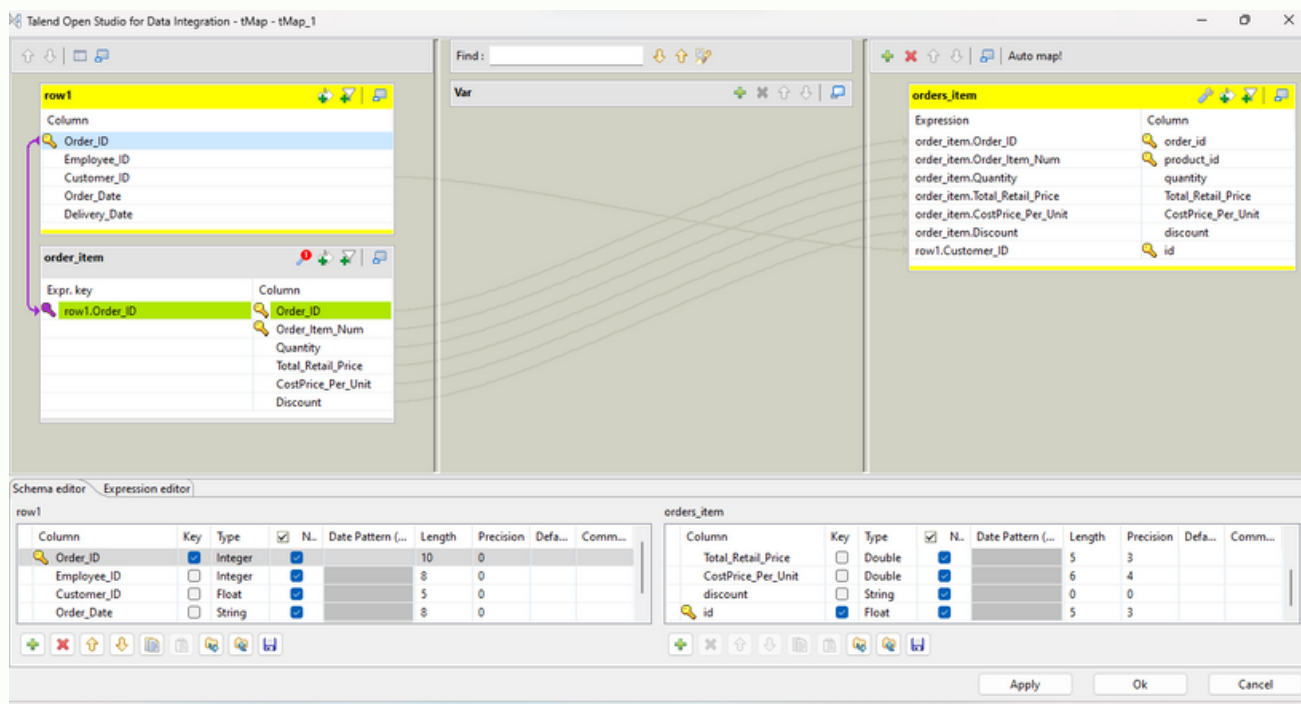
- Right-click both components , Select Row > Main, Click on the tMap component to link the two, rename the link with "order_item" and "row1".
- Drag-and-drop the tDBOutput component to the workplace , then right-click the tMap component to link it to the output
- Select Row > *New Output* (Main) and rename the link with "orders_item".
- Click on the tDBOutput and choose the MYSQL type of database.



3.Entering the Orders_Item Output table:

- Double-click the tMap component.
- You are brought to the tMap component configuration window.
 - On the right side of the screen, click Add output table.

- Enter a name for your output table: "orders_item".
- In input table order_item, select columns order_ID, Order_Item_Num, Quantity, Total_Retail_Price, CostPrice_Per_Unit, Discount then drag-and-drop them in output table orders_item and rename them as named in MYSQL server.
- In the input table order_item, drag customer_id into Order_ID to join them and drag it into the orders_item table.



Connecting to the DB:

After filling the settings of the tDBOutput, and connecting to the Database, we execute the job, then, we refresh MYSQL DB to observe the changes made.

phpMyAdmin

Server: 127.0.0.1 > Database: project > Table: orders_item

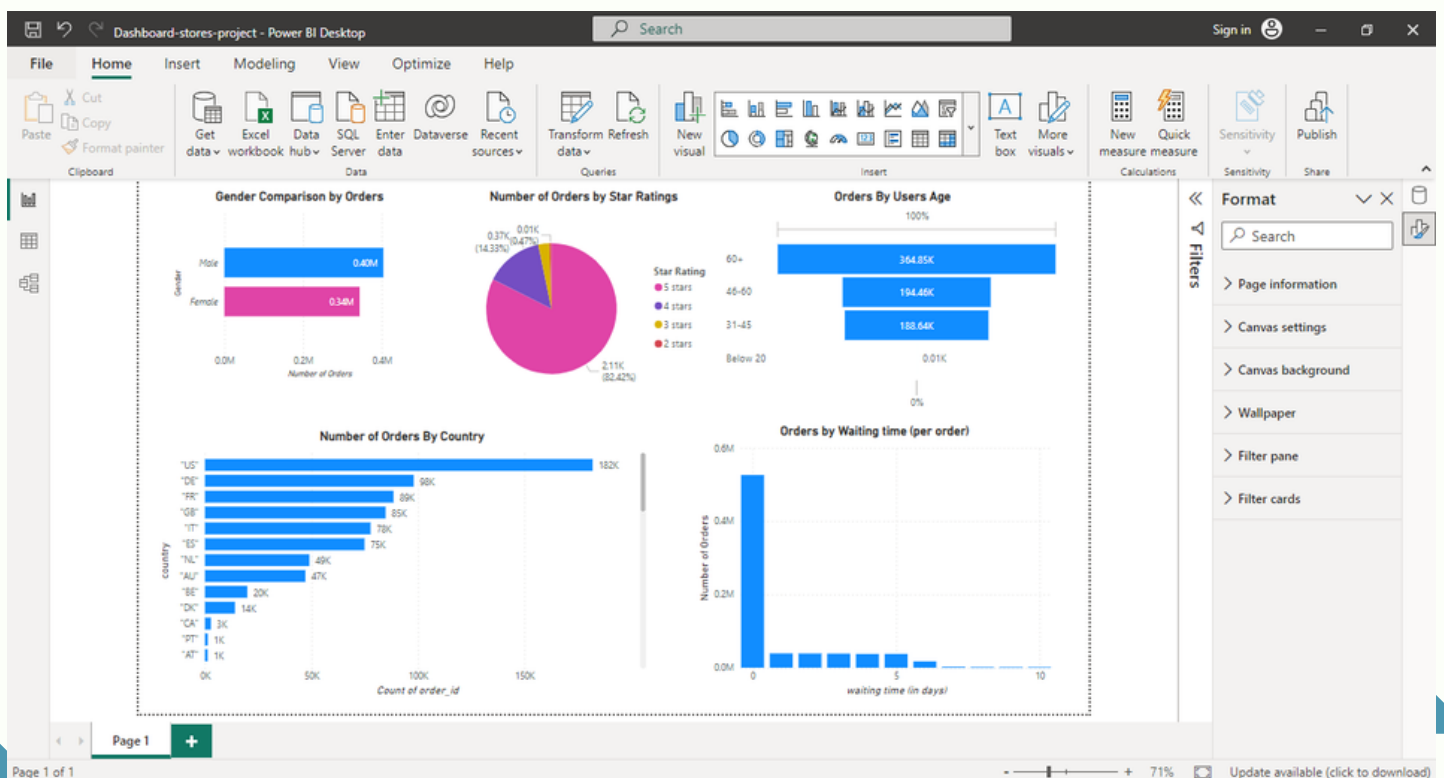
Browse Structure SQL Search Insert Export Import Privileges Operations

+ Options

				order_id	product_id	quantity	Total_Retail_Price	CostPrice_Per_Unit	discount	id
<input type="checkbox"/>	Edit	Copy	Delete	1232016491	1	2	95.98	39.9	0.5	48562
<input type="checkbox"/>	Edit	Copy	Delete	1232027771	1	2	283.98	123.75	0.5	24197
<input type="checkbox"/>	Edit	Copy	Delete	1232036481	1	2	53.98	21.75	0.5	15799
<input type="checkbox"/>	Edit	Copy	Delete	1232038165	1	2	387.98	150.2	0.5	44295
<input type="checkbox"/>	Edit	Copy	Delete	1232043658	1	2	47.98	20.5	0.5	32424
<input type="checkbox"/>	Edit	Copy	Delete	1232046908	1	4	427.96	69.25	0.4	94026
<input type="checkbox"/>	Edit	Copy	Delete	1232060228	1	2	329.98	136.75	0.5	37838
<input type="checkbox"/>	Edit	Copy	Delete	1232068060	1	2	55.98	24.7	0.5	34917
<input type="checkbox"/>	Edit	Copy	Delete	1232070316	2	2	213.98	69.25	0.4	42666
<input type="checkbox"/>	Edit	Copy	Delete	1232078779	1	2	387.98	150.2	0.5	42183
<input type="checkbox"/>	Edit	Copy	Delete	1232105475	1	2	53.98	21.75	0.5	36711
<input type="checkbox"/>	Edit	Copy	Delete	1232105737	1	2	61.98	19.95	0.3	85824
<input type="checkbox"/>	Edit	Copy	Delete	1232120221	2	2	283.98	123.75	0.5	53105
<input type="checkbox"/>	Edit	Copy	Delete	1232129222	1	2	329.98	136.75	0.5	61454
<input type="checkbox"/>	Edit	Copy	Delete	1232135582	1	2	289.98	119.25	0.5	42754
<input type="checkbox"/>	Edit	Copy	Delete	1232160485	1	2	95.98	39.9	0.5	22883
<input type="checkbox"/>	Edit	Copy	Delete	1232162677	1	2	213.98	69.25	0.4	43089
<input type="checkbox"/>	Edit	Copy	Delete	1232168989	1	4	427.96	69.25	0.4	55581
<input type="checkbox"/>	Edit	Copy	Delete	1232173239	1	2	329.98	136.75	0.5	43525

Console

Phase 03: Data Analysis



Executive Summary::

This comprehensive PowerBI Dashboard Analysis delves into the performance of our personal care products, offering invaluable insights that will guide strategic decisions. The visualizations include Gender Comparison by Orders, Number of Orders by Star Rating, Orders by Users Age, Number of Orders by Country, and Number of Orders by Waiting Time.

1. Gender Comparison by Orders (Clustered-Bar Chart):

The Clustered-Bar Chart illustrates the distribution of orders of personal care products across different genders. It is evident from the visualization that males have a slightly higher order volume (0.40M) compared to females (0.34M). This indicates a potential opportunity for targeted marketing and product customization.

2. Number of Orders by Star Rating (Pie Chart):

The Pie Chart displays the distribution of orders based on star ratings.

3. Orders by Users Age (Funnel):

The Funnel visualization showcases the distribution of orders based on user age groups.

4. Number of Orders by Country (Clustered Bar Chart):

The Clustered Bar Chart presents a comparison of order volumes across different countries.

5. Number of Orders by Waiting Time in Days (Clustered Column Chart):

The Clustered Column Chart illustrates the distribution of orders based on waiting times.

Conclusion:

In summary, the utilization of Talend Open Studio and Power BI in our business intelligence project for the online shopping store has proven to be a potent combination. Talend facilitates seamless data extraction and transformation, while Power BI provides intuitive visualization, resulting in a robust end-to-end solution. This integrated approach enhances our data-driven decision-making capabilities, enabling us to stay agile and competitive in the dynamic online retail landscape.



Thank you

