

BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT  
THÀNH PHỐ HỒ CHÍ MINH

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM  
Độc lập – Tự do – Hạnh phúc

KHOA CÔNG NGHỆ THÔNG TIN



**HCMUTE**

**ĐỒ ÁN MÔN HỌC**  
**MÔN: KHO DỮ LIỆU**  
**MÃ MÔN HỌC: DAWH430784**

**XÂY DỰNG KHO DỮ LIỆU TỪ HỆ THỐNG THƯƠNG MẠI  
ĐIỆN TỬ OLIST TẠI BRAZIL**

**Nhóm sinh viên thực hiện:** Nhóm 5  
**Học kỳ:** II  
**Năm học:** 2024 – 2025

**GVHD: Nguyễn Văn Thành**

*Thành phố Hồ Chí Minh, tháng 5 năm 2025*

## PHÂN CÔNG NHIỆM VỤ

**Nhóm: 05**

**Lớp: DAWH430784**

Nhiệm vụ	Trần Khánh Dương <b>22133011</b>	Nguyễn Đinh Hồng Phúc <b>22133041</b>	Lê Văn Sang <b>22133047</b>	Dương Thành Sơn <b>22133048</b> (Nhóm trưởng)
Tìm kiếm dữ liệu				X
Xử lý dữ liệu đầu vào	X	X		X
Xác định business process		X	X	
Xác định Dim và Fact	ProductSellerPerformance (3)	Sales&Revenue Analysis (1) OrderFulfillment &DeliveryPerformance (2)	Sales&Revenue Analysis (1) OrderFulfillment &DeliveryPerformance (2)	ProductSellerPerformance (3)
Staging và load các dim, fact bằng T-SQL	X	X	X	X
Staging và load các dim, fact bằng SSIS		X	X	X
Nhập dữ liệu vào SSAS, tạo Data Cube				X
Đặt câu hỏi	Đặt câu hỏi cho trực quan	Đặt câu hỏi cho phân tích	Đặt câu hỏi cho phân tích	
Trả lời bằng Pivot Table				X
Trực quan hóa dữ liệu bằng PowerBI	X			

## MỤC LỤC

<b>1. Giới thiệu chung.....</b>	<b>1</b>
<b>2. Xử lý dữ liệu đầu vào .....</b>	<b>2</b>
<b>2.1. Chuyển dữ liệu nguồn thành file CSV .....</b>	<b>2</b>
<b>2.2. Load dữ liệu gốc vào cơ sở dữ liệu trong SQL Server.....</b>	<b>3</b>
<b>3. Xây dựng kho dữ liệu .....</b>	<b>5</b>
<b>3.1. Star shema.....</b>	<b>6</b>
<b>3.2. Cấu trúc triển khai các Business Process.....</b>	<b>9</b>
<b>4. Tải dữ liệu (ETL) .....</b>	<b>15</b>
<b>4.1. Load dữ liệu sử dụng T-SQL .....</b>	<b>16</b>
<b>4.1.1. Load dữ liệu vào các Dimension .....</b>	<b>16</b>
<b>4.1.2. Load dữ liệu vào các Fact.....</b>	<b>26</b>
<b>4.2. Load bằng công cụ SSIS .....</b>	<b>29</b>
<b>4.2.1. Kết Nối .....</b>	<b>29</b>
<b>4.2.2. Dim Date Time .....</b>	<b>30</b>
<b>4.2.3. Business Process: Order Fulfillment &amp; Delivery Performance.....</b>	<b>31</b>
<b>4.2.4. Business Process : Sales &amp; Revenue Analysis .....</b>	<b>34</b>
<b>4.2.5. Business Process: Product Seller Peformance .....</b>	<b>37</b>
<b>5. Phân tích Dữ liệu bằng SSAS .....</b>	<b>38</b>
<b>5.1. Tạo Data Sources và Data Sources View kết nối với OlistDW .....</b>	<b>38</b>
<b>5.2. Tạo Cubes .....</b>	<b>39</b>
<b>5.2.1. Tạo Cube FactProductSellerPerformance .....</b>	<b>39</b>
<b>5.2.2. Tạo Cube FactSales .....</b>	<b>40</b>
<b>5.2.3. Tạo Cube FactOrderFullfillment .....</b>	<b>41</b>
<b>5.3. Thiết lập Measures và Dimensions .....</b>	<b>42</b>
<b>5.4. Triển khai và Xử lý Cube .....</b>	<b>43</b>
<b>6. Trả lời câu hỏi bằng Pivot Table .....</b>	<b>45</b>
<b>6.1. Kết nối Excel với Cube SSAS.....</b>	<b>45</b>
<b>6.2. Trả lời các câu hỏi phân tích .....</b>	<b>47</b>
<b>6.2.1. Business Process : Order Fulfillment &amp; Delivery Performance.....</b>	<b>47</b>

<b>6.2.2. Business Process : Sales &amp; Revenue Analysis .....</b>	<b>49</b>
<b>6.2.3. Business Process: Product Seller Peformance .....</b>	<b>53</b>
<b>7. Trực quan hóa bằng Power BI .....</b>	<b>57</b>
<b>    7.1. Thực hiện phân tích dữ liệu cho Business process: Sales &amp; Revenue Analysis .....</b>	<b>57</b>
<b>        7.1.1. Câu hỏi: Cho biết số lượng đơn đặt hàng, doanh thu qua các năm 2017, 2018 .....</b>	<b>58</b>
<b>        7.1.2. Câu hỏi: Cho biết sự tăng giảm doanh thu qua các tháng (2017, 2018) dựa theo thị trường/khu vực .....</b>	<b>59</b>
<b>    7.2. Thực hiện phân tích dữ liệu cho Business process: Order Fulfillment.....</b>	<b>60</b>
<b>        7.2.1. Câu hỏi: Cho biết thời gian giao hàng trung bình qua các tháng của năm 2017 và 2018 .....</b>	<b>60</b>
<b>        7.2.2. Câu hỏi: Cho biết tỉ lệ đơn giao đúng hạn và trễ hạn qua các tháng của năm 2017 và 2018 .....</b>	<b>61</b>
<b>    7.3. Thực hiện phân tích dữ liệu cho Business process: Product &amp; Seller Performance.....</b>	<b>62</b>
<b>        7.3.1. Câu hỏi: Cho biết 5 danh mục sản phẩm bán chạy và đem theo doanh thu nhiều nhất trong năm 2018 .....</b>	<b>63</b>
<b>        7.3.2. Câu hỏi: Cho biết 5 danh mục sản phẩm nhận phản hồi tệ nhất trong năm 2018 .....</b>	<b>64</b>
<b>        7.3.3. Câu hỏi: Cho biết 3 Seller bán được nhiều đơn hàng và đem lại doanh thu nhiều nhất trong năm 2018 .....</b>	<b>65</b>
<b>8. Kết luận .....</b>	<b>66</b>
<b>    8.1. Kết quả đạt được .....</b>	<b>66</b>
<b>    8.2. Những hạn chế .....</b>	<b>66</b>
<b>9. Tài liệu tham khảo .....</b>	<b>66</b>

## 1. Giới thiệu chung

- Lý do chọn đề tài

Trong thời đại công nghệ số hiện nay, thương mại điện tử (e-commerce) đang phát triển mạnh mẽ và trở thành xu hướng tất yếu trong hoạt động kinh doanh toàn cầu. Việc khai thác và phân tích dữ liệu từ các nền tảng thương mại điện tử giúp các doanh nghiệp hiểu rõ hơn về hành vi tiêu dùng, xu hướng mua sắm, hiệu suất bán hàng, cũng như đánh giá chất lượng dịch vụ. Từ đó, họ có thể đưa ra các chiến lược kinh doanh hiệu quả, tối ưu hóa trải nghiệm khách hàng và tăng khả năng cạnh tranh.

Xuất phát từ nhu cầu thực tiễn đó, nhóm em chọn đề tài "**Xây dựng kho dữ liệu từ hệ thống thương mại điện tử Olist tại Brazil**". Đây là một bài toán thực tế, giàu tiềm năng phân tích, với dữ liệu phong phú và đa chiều, bao gồm thông tin về khách hàng, đơn hàng, sản phẩm, nhà bán hàng, phương thức vận chuyển, đánh giá dịch vụ, thanh toán,...

- Nguồn dữ liệu

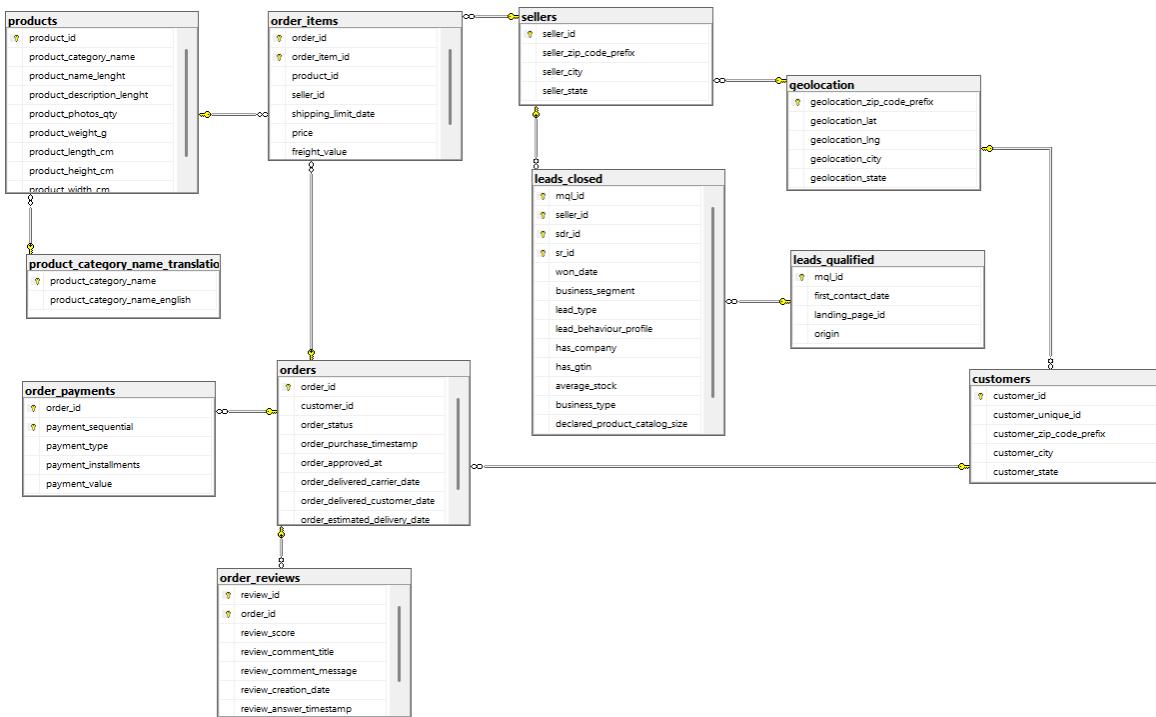
Tập dữ liệu được sử dụng trong đồ án được lấy từ Kaggle - E-commerce dataset by Olist (SQLite), bao gồm hơn 100.000 đơn hàng được thực hiện từ năm 2016 đến 2018. Đây là một nguồn dữ liệu thực tế, đã được xử lý sơ bộ, phù hợp với mục tiêu xây dựng kho dữ liệu và khai thác thông tin nhằm hỗ trợ ra quyết định trong lĩnh vực thương mại điện tử.

Đường dẫn kaggle: <https://www.kaggle.com/datasets/terencicp/e-commerce-dataset-by-olist-as-an-sqlite-database>

- Tập dữ liệu gồm các bảng và thông tin từng bảng

Tên bảng	Số dòng	Số cột
customers	99441	5
geolocation	1000163	5
leads_closed	842	14
leads_qualified	8000	4

order_items	112650	7
order_payments	103886	5
order_reviews	99224	7
orders	99441	8
product_category_name_translation	71	2
products	32951	9
sellers	3095	4



## 2. Xử lý dữ liệu đầu vào

### 2.1. Chuyển dữ liệu nguồn thành file CSV

- Dữ liệu: Là 1 file dạng olist.sqlite
- Công cụ thực hiện (Python, SQLite) để xuất file CSV từ các bảng trong tập olist.sqlite

```
import sqlite3
import csv

conn = sqlite3.connect("olist.sqlite")
cursor = conn.cursor()

cursor.execute("SELECT * FROM customers")
rows = cursor.fetchall()

# Lấy tên cột
columns = [description[0] for description in cursor.description]

# Ghi ra file CSV
with open("customers.csv", "w", newline='', encoding="utf-8") as f:
    writer = csv.writer(f)
    writer.writerow(columns)
    writer.writerows(rows)

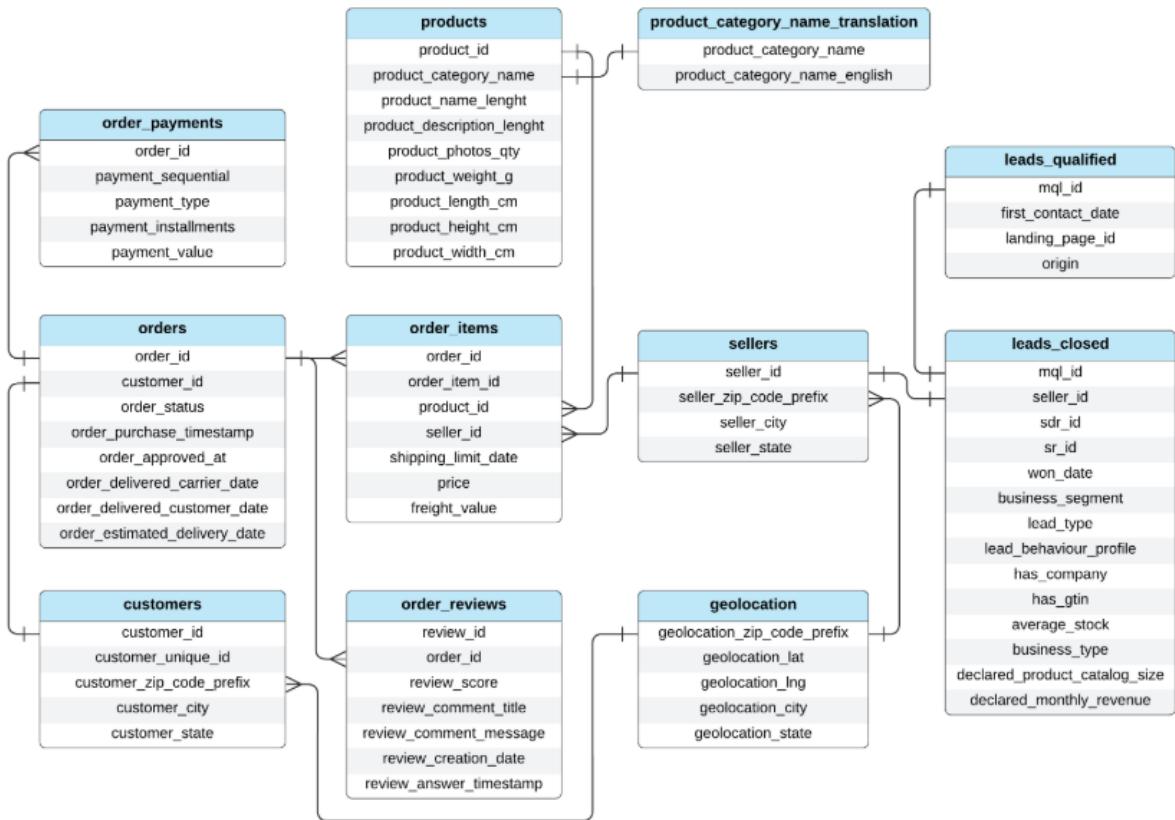
conn.close()
```

## 2.2. Load dữ liệu gốc vào cơ sở dữ liệu trong SQL Server

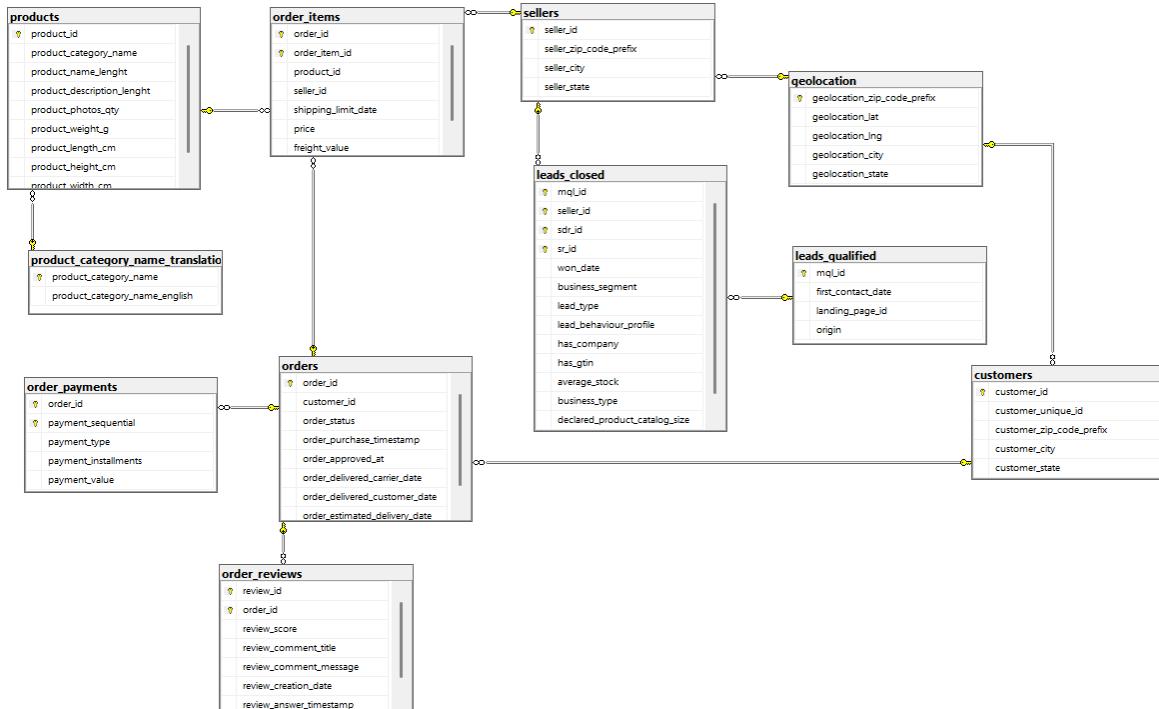
### 2.2.1. Tạo cơ sở dữ liệu từ trong SQL Server dựa trên dữ liệu nguồn

- Lược đồ SQL Lite của dữ liệu nguồn

## DATABASE SCHEMA



- Lược đồ trong SQL Server được tạo lại



- Thực hiện load dữ liệu từ CSV vào database và xem kết quả

```

SQLQuery1.sql - (local).OlistDB (sa (55)) - Microsoft SQL Server Management Studio
File Edit View Query Project Tools Window Help
Object Explorer
Connect ▾
(OQL Server 16.0.1000.6 - sa)
Databases
System Databases
Database Snapshots
DW Configuration
DW Diagnostic
DW Queue
NorthwindDW
OlistDB
Database Diagrams
Tables
Views
Synonyms
Programmability
Query Store
Service Broker
Storage
Security
OlistDW
OlistDWStage
ProductSellerPerformanceDWStage
PTDL
Store Sales
Security
Server Objects
Replication
Always On High Availability
Management
Integration Services Catalogs
SQL Server Agent (Agent XPs disabled)
XEvent Profiler
Results Messages
table_name total_rows
orders 9941
order_items 112550
order_payments 103886
order_reviews 95224
products 32952
product_category_name_translation 75
sellers 3557
customers 9941
geolocation 19177

```

Query executed successfully.

(local) (16.0 RTM) | sa (55) | OlistDB | 00:00:00 | 9 rows

### 3. Xây dựng kho dữ liệu

Olist – một nền tảng thương mại điện tử tổng hợp tại Brazil – cung cấp hệ sinh thái bán hàng cho các nhà cung cấp thông qua việc kết nối họ với nhiều marketplace khác nhau. Do đó, lượng dữ liệu phát sinh từ các giao dịch, đánh giá, thanh toán, vận chuyển... là rất lớn và phong phú, đòi hỏi một hệ thống lưu trữ và phân tích hiệu quả.

Tuy nhiên, dữ liệu gốc của Olist hiện được tổ chức dưới dạng các bảng transactional rời rạc như orders, order\_items, customers, sellers, order\_reviews, order\_payments,... Dạng dữ liệu này tuy phù hợp với hệ thống vận hành (OLTP), nhưng lại gây khó khăn cho các phân tích tổng hợp (OLAP) do thiếu tính tổ chức và khó truy vấn hiệu suất cao.

Chính vì vậy, việc xây dựng kho dữ liệu (Data Warehouse) cho Olist là cần thiết. Kho dữ liệu giúp tổ chức dữ liệu theo mô hình Star Schema, trong đó các bảng fact trung tâm kết nối với các bảng dimension mô tả, tạo điều kiện cho việc phân tích dễ dàng, trực quan và hiệu quả hơn.

Dựa trên dữ liệu Olist đã được cung cấp, đồ án này lựa chọn ba business process đặc trưng đại diện cho ba khía cạnh quan trọng trong hoạt động thương mại điện tử để xây dựng mô hình kho dữ liệu:

- Order Fulfillment & Delivery Performance:** phân tích hiệu suất giao hàng để đánh giá độ hài lòng khách hàng và hiệu quả vận hành.

- **Sales & Revenue Analysis:** đánh giá doanh thu theo sản phẩm, thời gian, seller, và chi phí logistics.
- **Product & Seller Performance:** theo dõi chất lượng sản phẩm và hiệu suất bán hàng để hỗ trợ chiến lược kinh doanh.

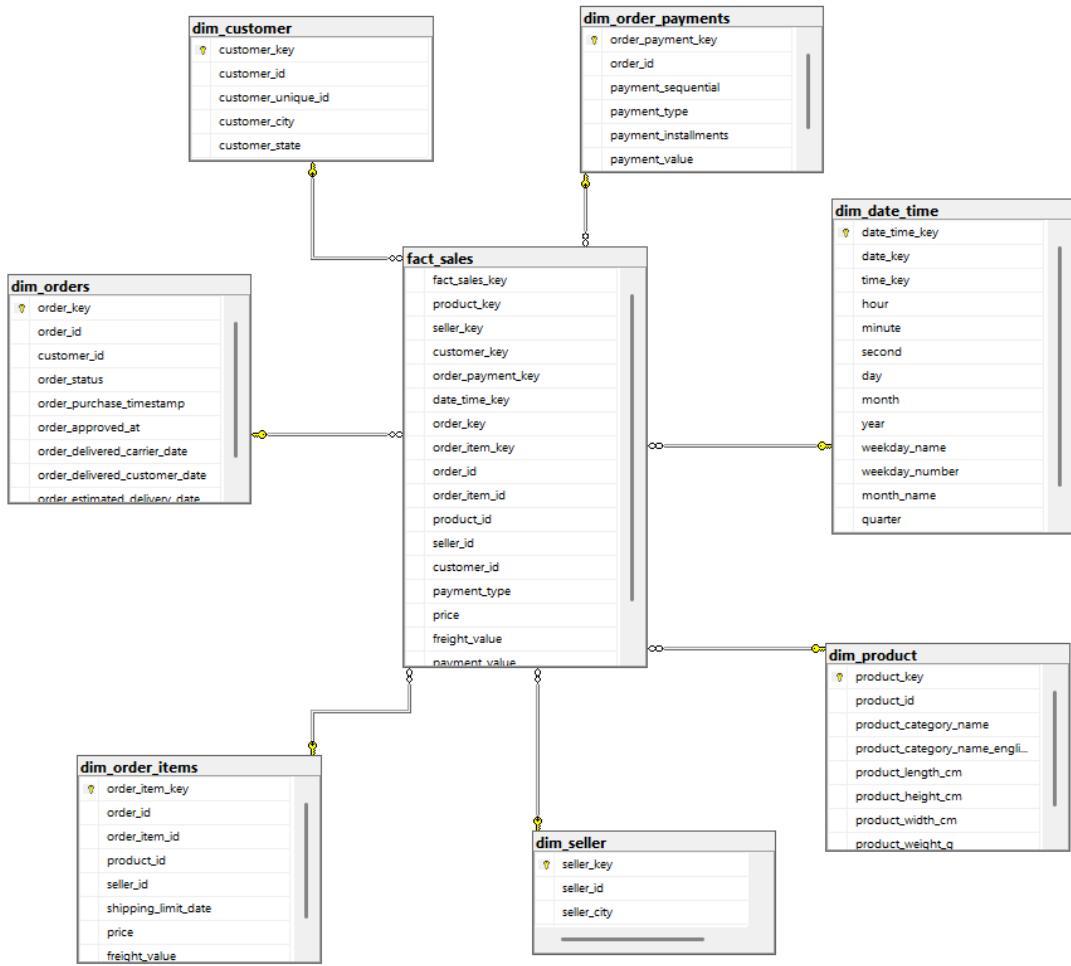
Mỗi business process đều được thiết kế tương ứng với một bảng fact chứa dữ liệu định lượng, cùng các bảng dimension chứa dữ liệu mô tả như khách hàng, sản phẩm, nhà bán hàng, thời gian, phương thức thanh toán,... Từ đó, hệ thống có thể trả lời được nhiều câu hỏi nghiệp vụ cụ thể như:

- Thời gian giao hàng trung bình là bao lâu?
- Doanh thu theo loại sản phẩm và theo thời gian ra sao? Seller nào tạo doanh thu cao nhất?
- Sản phẩm nào bị đánh giá thấp nhất? Seller nào thường xuyên bị khách hàng chê trách?

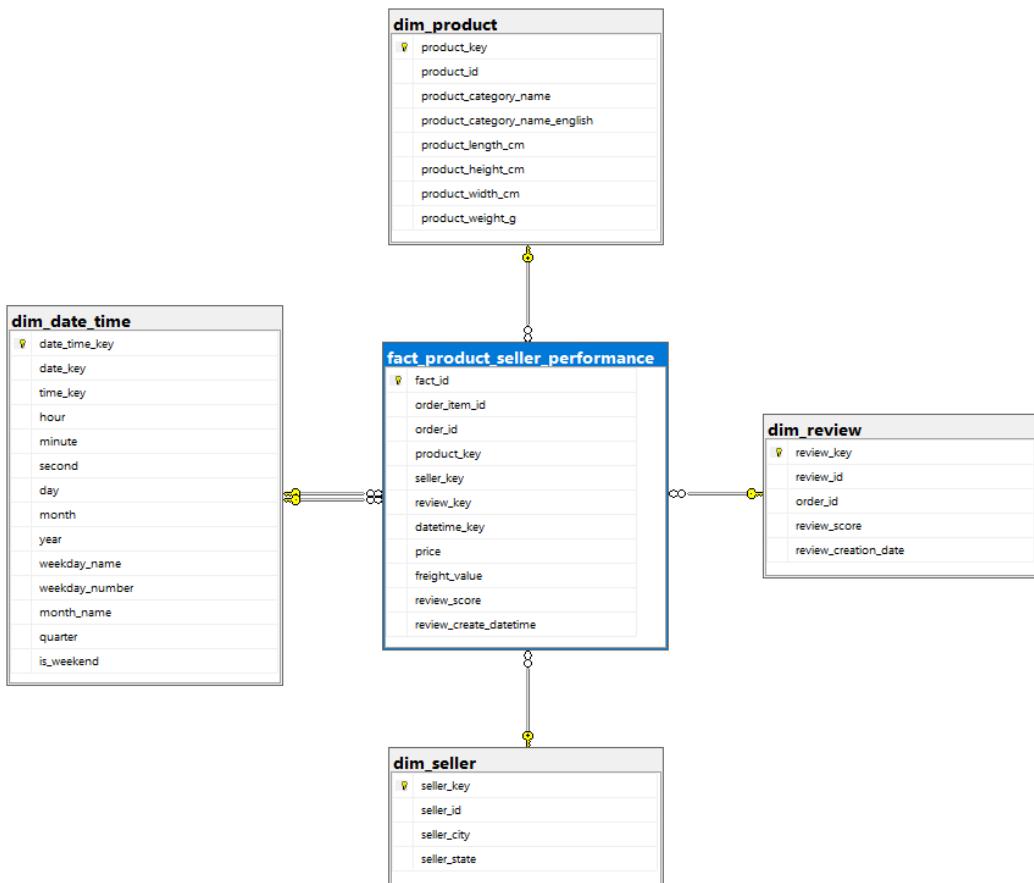
Việc triển khai kho dữ liệu giúp đơn giản hóa truy vấn, hỗ trợ trực quan hóa dữ liệu và tạo nền tảng vững chắc cho các hệ thống phân tích BI (Business Intelligence), từ đó giúp doanh nghiệp đưa ra quyết định nhanh chóng và chính xác hơn.

### 3.1. Star schema

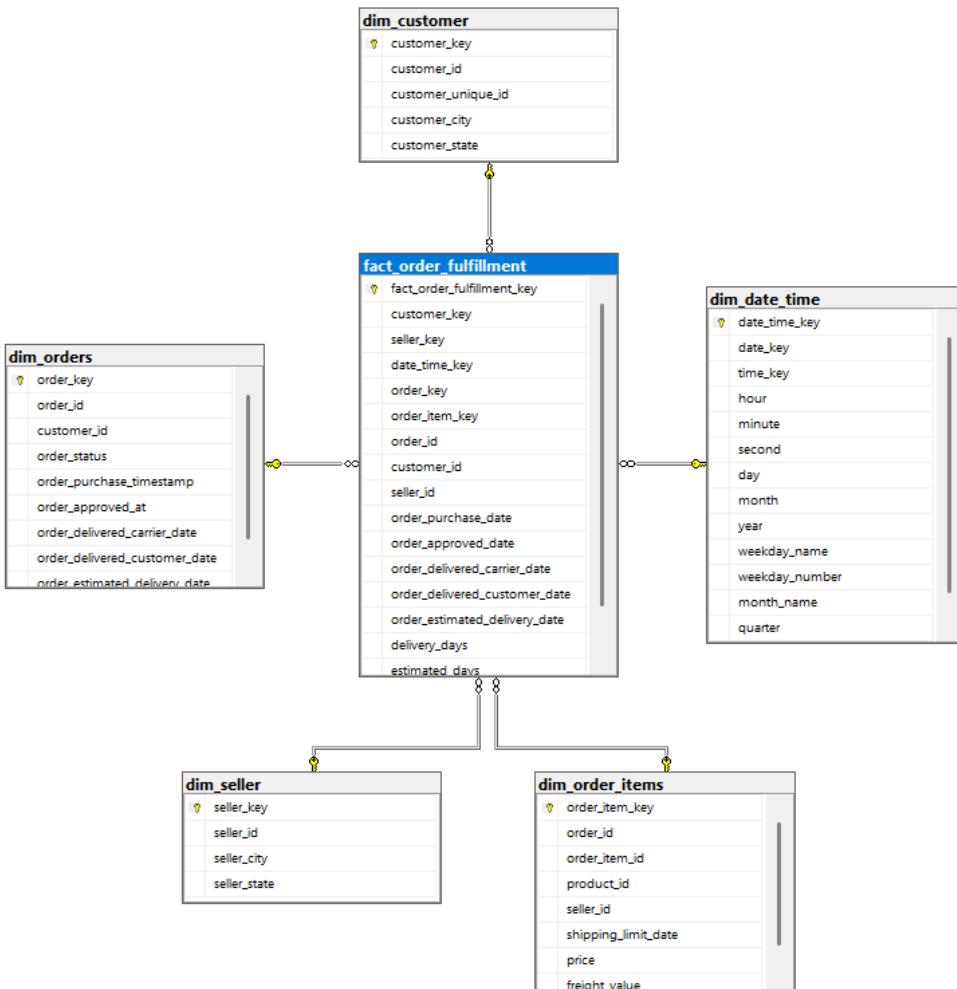
- Thiết kế cho Fact Sales



- Thiết kế cho Fact Product Seller Performance



- Thiết kế cho Fact Product Order Fullfillment



## 3.2. Cấu trúc triển khai các Business Process

### 3.2.1. Dim Customer

	Column Name	Data Type	Allow Nulls
!	customer_key	int	<input type="checkbox"/>
	customer_id	nvarchar(100)	<input checked="" type="checkbox"/>
	customer_unique_id	nvarchar(100)	<input checked="" type="checkbox"/>
	customer_city	nvarchar(100)	<input checked="" type="checkbox"/>
	customer_state	nvarchar(100)	<input checked="" type="checkbox"/>

### 3.2.2. Dim Seller

	Column Name	Data Type	Allow Nulls
Y	seller_key	int	<input type="checkbox"/>
	seller_id	nvarchar(100)	<input checked="" type="checkbox"/>
	seller_city	nvarchar(100)	<input checked="" type="checkbox"/>
	seller_state	nvarchar(100)	<input checked="" type="checkbox"/>

### 3.2.3. Dim Orders

	Column Name	Data Type	Allow Nulls
Y	order_key	int	<input type="checkbox"/>
	order_id	nvarchar(50)	<input checked="" type="checkbox"/>
	customer_id	nvarchar(50)	<input checked="" type="checkbox"/>
	order_status	nvarchar(50)	<input checked="" type="checkbox"/>
	order_purchase_timestamp	datetime2(7)	<input checked="" type="checkbox"/>
	order_approved_at	datetime2(7)	<input checked="" type="checkbox"/>
	order_delivered_carrier_date	datetime2(7)	<input checked="" type="checkbox"/>
	order_delivered_customer_date	datetime2(7)	<input checked="" type="checkbox"/>
	order_estimated_delivery_date	datetime2(7)	<input checked="" type="checkbox"/>
			<input type="checkbox"/>

### 3.2.4. Dim Order Items

	Column Name	Data Type	Allow Nulls
Y	order_item_key	int	<input type="checkbox"/>
	order_id	nvarchar(50)	<input checked="" type="checkbox"/>
	order_item_id	int	<input checked="" type="checkbox"/>
	product_id	nvarchar(50)	<input checked="" type="checkbox"/>
	seller_id	nvarchar(50)	<input checked="" type="checkbox"/>
	shipping_limit_date	datetime2(7)	<input checked="" type="checkbox"/>
	price	float	<input checked="" type="checkbox"/>
	freight_value	float	<input checked="" type="checkbox"/>
			<input type="checkbox"/>

### 3.2.5. Dim Product

	Column Name	Data Type	Allow Nulls
!	product_key	int	<input type="checkbox"/>
	product_id	nvarchar(100)	<input checked="" type="checkbox"/>
	product_category_name	nvarchar(100)	<input checked="" type="checkbox"/>
	product_category_name_english	nvarchar(100)	<input checked="" type="checkbox"/>
	product_length_cm	float	<input checked="" type="checkbox"/>
	product_height_cm	float	<input checked="" type="checkbox"/>
	product_width_cm	float	<input checked="" type="checkbox"/>
	product_weight_g	float	<input checked="" type="checkbox"/>
▶			<input type="checkbox"/>

### 3.2.6. Dim Order Payments

	Column Name	Data Type	Allow Nulls
!	order_payment_key	int	<input type="checkbox"/>
	order_id	nvarchar(50)	<input checked="" type="checkbox"/>
	payment_sequential	int	<input checked="" type="checkbox"/>
	payment_type	nvarchar(50)	<input checked="" type="checkbox"/>
	payment_installments	int	<input checked="" type="checkbox"/>
	payment_value	float	<input checked="" type="checkbox"/>
▶			<input type="checkbox"/>

### 3.2.7. Dim Date Time

	Column Name	Data Type	Allow Nulls
!	date_time_key	varchar(20)	<input type="checkbox"/>
	date_key	int	<input checked="" type="checkbox"/>
	time_key	int	<input checked="" type="checkbox"/>
	hour	int	<input checked="" type="checkbox"/>
	minute	int	<input checked="" type="checkbox"/>
	second	int	<input checked="" type="checkbox"/>
	day	int	<input checked="" type="checkbox"/>
	month	int	<input checked="" type="checkbox"/>
	year	int	<input checked="" type="checkbox"/>
	weekday_name	varchar(20)	<input checked="" type="checkbox"/>
	weekday_number	int	<input checked="" type="checkbox"/>
	month_name	varchar(20)	<input checked="" type="checkbox"/>
	quarter	int	<input checked="" type="checkbox"/>
	is_weekend	bit	<input checked="" type="checkbox"/>
▶			<input type="checkbox"/>

### 3.2.8. Dim Review

	Column Name	Data Type	Allow Nulls
!	review_key	int	<input type="checkbox"/>
	review_id	nvarchar(50)	<input checked="" type="checkbox"/>
	order_id	nvarchar(50)	<input checked="" type="checkbox"/>
	review_score	tinyint	<input checked="" type="checkbox"/>
	review_creation_date	datetime	<input checked="" type="checkbox"/>
▶			<input type="checkbox"/>

### 3.2.9. Fact Order Fulfillment

DESKTOP-RM3I2NJ.O...\_order\_fulfillment ✎ X OlistDW\_Create.sql...l).OlistDW (sa (51))

	Column Name	Data Type	Allow Nulls
▼	fact_order_fulfillment_key	int	<input type="checkbox"/>
	customer_key	int	<input checked="" type="checkbox"/>
	seller_key	int	<input checked="" type="checkbox"/>
	date_time_key	varchar(20)	<input checked="" type="checkbox"/>
	order_key	int	<input checked="" type="checkbox"/>
	order_item_key	int	<input checked="" type="checkbox"/>
	order_id	nvarchar(100)	<input checked="" type="checkbox"/>
	customer_id	nvarchar(100)	<input checked="" type="checkbox"/>
	seller_id	nvarchar(100)	<input checked="" type="checkbox"/>
	order_purchase_date	datetime	<input checked="" type="checkbox"/>
	order_approved_date	datetime	<input checked="" type="checkbox"/>
	order_delivered_carrier_date	datetime	<input checked="" type="checkbox"/>
	order_delivered_customer_date	datetime	<input checked="" type="checkbox"/>
	order_estimated_delivery_date	datetime	<input checked="" type="checkbox"/>
	delivery_days	int	<input checked="" type="checkbox"/>
	estimated_days	int	<input checked="" type="checkbox"/>
▶	late_delivery	bit	<input checked="" type="checkbox"/>
			<input type="checkbox"/>

### 3.2.10. Fact Sales

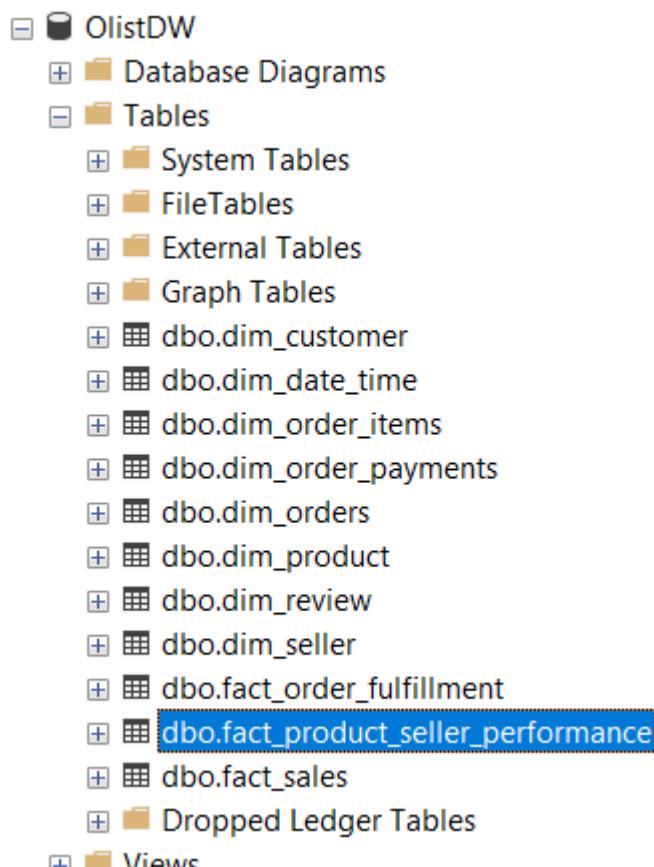
Column Name	Data Type	Allow Nulls
fact_sales_key	int	<input type="checkbox"/>
product_key	int	<input checked="" type="checkbox"/>
seller_key	int	<input checked="" type="checkbox"/>
customer_key	int	<input checked="" type="checkbox"/>
order_payment_key	int	<input checked="" type="checkbox"/>
date_time_key	varchar(20)	<input checked="" type="checkbox"/>
order_key	int	<input checked="" type="checkbox"/>
order_item_key	int	<input checked="" type="checkbox"/>
order_id	nvarchar(100)	<input checked="" type="checkbox"/>
order_item_id	int	<input checked="" type="checkbox"/>
product_id	nvarchar(100)	<input checked="" type="checkbox"/>
seller_id	nvarchar(100)	<input checked="" type="checkbox"/>
customer_id	nvarchar(100)	<input checked="" type="checkbox"/>
payment_type	nvarchar(100)	<input checked="" type="checkbox"/>
price	float	<input checked="" type="checkbox"/>
freight_value	float	<input checked="" type="checkbox"/>
payment_value	float	<input checked="" type="checkbox"/>
order_purchase_date	datetime	<input checked="" type="checkbox"/>
		<input type="checkbox"/>

### 3.2.11 Fact Product Seller Performance

DESKTOP-RM3I2NJ....eller\_performance X OlistDW\_Create.sql...l).OlistDW (sa (51))

	Column Name	Data Type	Allow Nulls
!	fact_id	int	<input type="checkbox"/>
	order_item_id	tinyint	<input checked="" type="checkbox"/>
	order_id	nvarchar(50)	<input checked="" type="checkbox"/>
	product_key	int	<input checked="" type="checkbox"/>
	seller_key	int	<input checked="" type="checkbox"/>
	review_key	int	<input checked="" type="checkbox"/>
	datetime_key	varchar(20)	<input checked="" type="checkbox"/>
	price	float	<input checked="" type="checkbox"/>
	freight_value	float	<input checked="" type="checkbox"/>
	review_score	tinyint	<input checked="" type="checkbox"/>
	review_create_datetime	varchar(20)	<input checked="" type="checkbox"/>
			<input type="checkbox"/>

- Cấu trúc tổng hợp



#### 4. Tải dữ liệu (ETL)

## 4.1. Load dữ liệu sử dụng T-SQL

### 4.1.1. Load dữ liệu vào các Dimension

#### 4.1.1.1. Dim Customer

##### Tạo Dim Customer Stage

```
SELECT * INTO stg_customers FROM OlistDB.dbo.customers;
```

The screenshot shows the Microsoft SQL Server Management Studio interface. In the Object Explorer, the connection is set to '(Local).OlistDWStage (sa (76))'. The current database is 'OlistDWStage'. In the center pane, a query window displays the following T-SQL code:

```
SELECT TOP (1000) [customer_id]
      ,[customer_unique_id]
      ,[customer_zip_code_prefix]
      ,[customer_city]
      ,[customer_state]
  FROM [OlistDWStage].[dbo].[stg_customers]
```

The results grid below the code shows 1,000 rows of data. The columns are:

	customer_id	customer_unique_id	customer_zip_code_prefix	customer_city	customer_state
1	194de7c2d9e090cd938308c02b7292f	b53d4d73e92d7e206140c084e2a18f	49020	serra	SE
2	194e584217a35719442520937ade041	3835ff159db5cafeae611d894776c	29168	serra	ES
3	19491be65f13e353ba4853be363c2e8f5	a2bc6f336fe9e7e9bd3e9df9db5a2e8f1	21931	rio de janeiro	RJ
4	19500cb0ea00f4fb75153c4e959fc5	017ae1dedebc1749a9e520e3783c3c	4896	seo paulo	SP
5	195133a4d12b2d6f18c84493e48c46	a30990e6b91907ac3e5e57b7e2747407	19911	ourinhos	SP
6	1951367310f1a124270b26e0fa8c0c136	b8e55e9e8e83653a5d5e2529aef3e	39917	bandeira	MG
7	195487592f0efed573852b1ed7463b8	4322456fc8d8d00164e8ba9f45211	1210	seo paulo	SP
8	1954e4e89170a036dd2639a05d776d009	23078901259abed9d0e476bedc0e7f	20240	rio de janeiro	RJ
9	19557e0c50a69c2f2e2ee48628ff1	9e41be9e930a292cfdb98715fe991c7c	12246	seo jose dos campos	SP
10	195723b01cad0e0552c292693650	391ab062710d6d04e7e4d398e71d5	22831	rio de janeiro	RJ
11	195712b24e354c275e5f353d6d4	cdf846a8e572240e84089cb949017	6436	benueri	SP
12	1958a70e6c1d652a5f1a65959e0a0628	9635cc2dc37d7d2d4e28765b40592	73755	planaltina	GO
13	1958ecc55e91b363b6c567114e4b9b	40869b13e99e716ffd9b9d4f3e110	72460	brasilia	DF
14	1958999fc054e50319e5a0744d2f24	b0b057559e5c4e99e41a21c4e0d87780	4012	seo paulo	SP
15	195e363446d04a4e2a4a79be3d705	df492b430477feee52b63813d1ed2a06	35650	pitangui	MG
16	195b6b760d9946814d621fc6e21af5	446e005d35d0e162996e6515c312c292	16201	birigui	SP
17	195ca01e416510a6d65fd44a50516	b5b2852d888ad0bd26a3bwe07352	923	santo andre	SP
18	195d4fc0b177e04e0b2c2a23c49c18c	ee2559f37746b533e1d124150c0df36	28400	seo fielis	RJ
19	195f0d0a556a92f25ab1a1ccc017439	2e49e3bbef7e297ee0ff9d9f9c2456c	21832	rio de janeiro	RJ
20	196052c199e9d668e99780c3a38a77	c5506a0e00c0ce49d92e0e1769096b	27257	volta redonda	RJ
21	1961e41144d48610f2e2a4b5070e	7e68a5e4e404c180b1dew1ed000803	5749	seo paulo	SP
22	19625deff53dd7e3142b18597469e8	7df0fe86f96248577ca3f634de44b0c	70233	brasilia	DF
23	19630b653062711790f0d00435d66	7df0fe86f96248577ca3f634de44b0c	60442	fortaleze	CE
24	19643397a548926984180e95e17	#1d81c3678c2b8a865d50c7e89d4	17690	bastos	SP

At the bottom of the results grid, it says 'Query executed successfully.'

##### Load vào dim\_customer

```
INSERT INTO dbo.dim_customer (customer_id, customer_unique_id,
customer_city, customer_state)
```

```
SELECT DISTINCT customer_id, customer_unique_id, customer_city,
customer_state
```

```
FROM OlistDWStage.dbo.stg_customers;
```

```

SELECT TOP (1000) [customer_key]
      ,[customer_id]
      ,[customer_unique_id]
      ,[customer_city]
      ,[customer_state]
  FROM [olistdw].[dbo].[dim_customer]
  
```

#### 4.1.1.2. Dim Seller

##### Tạo Dim Seller Stage

`SELECT * INTO stg_sellers FROM OlistDB.dbo.sellers;`

```

SELECT TOP (1000) [seller_id]
      ,[seller_zip_code_prefix]
      ,[seller_city]
      ,[seller_state]
  FROM [olistdwstage].[dbo].[stg_sellers]
  
```

##### Load vào dim\_seller

`INSERT INTO dbo.dim_seller (seller_id, seller_city, seller_state)`  
`SELECT DISTINCT seller_id, seller_city, seller_state`  
`FROM OlistDWStage.dbo.stg_sellers;`

The screenshot shows the Microsoft SQL Server Management Studio interface. In the Object Explorer, the 'OlistDW' database is selected. In the center pane, a query result grid is displayed for a SELECT statement. The columns are 'seller\_key', 'seller\_id', 'seller\_city', and 'seller\_state'. The data shows 1000 rows of seller information from the 'dim\_seller' table.

seller_key	seller_id	seller_city	seller_state
1	00065520beccb8785e2cf78355eb9b68	Porto Alegre	RJ
2	0015a52c2e000fa5aa3e2e5b532	santo andre	SP
3	001cc57ea9e17fb1c9efffb1094831	carieica	ES
4	001e6ad4d69e905060d95994fb1e14f	seo goncalo	RJ
5	0021009778cabb431b7e1020ff7eb4ff	franca	SP
6	003554e2dce176b555535a4d555ec8	goenia	GO
7	0044c1c087a3c30c522d4c4c07416	ibiriba	SP
8	00720abef5b0055907595bb454533b	guerinhos	SP
9	008a53eaffb15b192e5f1a63bcccfe11c8	seo paulo	SP
10	008b143d12632bae99e0e865ed52825	belo horizonte	MG
11	00ee6302b4b5bc5e2660cd833c361cc	seo paulo	SP
12	00fc707eaaad2e1347e683cd2df10	maringa	PR
13	010543af2bd80ea422851e79a3b7c7540	seo paulo	SP
14	010de0602d774602cd1b3f567b7709e	seo bernardo do campo	SP
15	011b0eaba873862ae9e9a7c32bb531d1	pompeia	SP
16	012664ac46ef5a19678d16e8b683d325	curitiba	PR
17	013900e83eace745d3e7614cab5b1a	curitiba	PR
18	013d10060864e4c0275b0338ed064	São Paulo	SP
19	014a0e79dd340e0e33872e7e85666	timoteo	MG
20	014d9a685fd572726679eadd0e0709a5	maringa	PR
21	0176f73cc119505677b52b1e53ea9	ibiriba	SP
22	01bccd254e01430ce9791b69b02e47	ureuca	GO
23	01c97ebeb5dca5c52891c0ed1c37b0012	catanduva	SP
24	01d7e3d21494c41fb860342e7141a1	francisco bellarmino	PR
25	01e2d54b0ff807407ffdb9d9b8a1e17d923	mogi das cruzes	SP
26	016d07212124329bec32490a8e0809	ribearia preto / seo pa...	

Query executed successfully.

#### 4.1.1.3. Dim Date

##### Load dim\_date\_time

```
BULK INSERT dim_date_time
FROM
'C:\Users\Asus\Desktop\Data_Engineer\Kho_du_lieu\project\olist_csv_export\
DimDateTime.csv'
WITH (
    FORMAT = 'CSV',
    FIRSTROW = 2,
    FIELDTERMINATOR = ',',
    ROWTERMINATOR = '\n',
    TABLOCK,
    ERRORFILE = 'C:\Users\Asus\Desktop\DimDateTime_ErrorRows.log'
);
```

```

SELECT TOP (1000) [date_time_key]
      ,[date_key]
      ,[time_key]
      ,[hour]
      ,[minute]
      ,[second]
      ,[day]
      ,[month]
      ,[year]
      ,[weekday_name]
      ,[weekday_number]
      ,[month_name]
      ,[quarter]
      ,[is_weekend]
  FROM [olistDW].[dbo].[dim_date_time]

```

	date_time_key	date_key	time_key	hour	minute	second	day	month	year	weekday_name	weekday_number	month_name	quarter	is_weekend
1	2016-09-04 21:00:00	20160904	2100	21	0	0	4	9	2016	Sunday	7	September	3	1
2	2016-09-04 21:01:00	20160904	2101	21	1	0	4	9	2016	Sunday	7	September	3	1
3	2016-09-04 21:02:00	20160904	2102	21	2	0	4	9	2016	Sunday	7	September	3	1
4	2016-09-04 21:03:00	20160904	2103	21	3	0	4	9	2016	Sunday	7	September	3	1
5	2016-09-04 21:04:00	20160904	2104	21	4	0	4	9	2016	Sunday	7	September	3	1
6	2016-09-04 21:05:00	20160904	2105	21	5	0	4	9	2016	Sunday	7	September	3	1
7	2016-09-04 21:06:00	20160904	2106	21	6	0	4	9	2016	Sunday	7	September	3	1
8	2016-09-04 21:07:00	20160904	2107	21	7	0	4	9	2016	Sunday	7	September	3	1
9	2016-09-04 21:08:00	20160904	2108	21	8	0	4	9	2016	Sunday	7	September	3	1
10	2016-09-04 21:09:00	20160904	2109	21	9	0	4	9	2016	Sunday	7	September	3	1
11	2016-09-04 21:10:00	20160904	2110	21	10	0	4	9	2016	Sunday	7	September	3	1
12	2016-09-04 21:11:00	20160904	2111	21	11	0	4	9	2016	Sunday	7	September	3	1
13	2016-09-04 21:12:00	20160904	2112	21	12	0	4	9	2016	Sunday	7	September	3	1
14	2016-09-04 21:13:00	20160904	2113	21	13	0	4	9	2016	Sunday	7	September	3	1
15	2016-09-04 21:14:00	20160904	2114	21	14	0	4	9	2016	Sunday	7	September	3	1
16	2016-09-04 21:15:00	20160904	2115	21	15	0	4	9	2016	Sunday	7	September	3	1

#### 4.1.1.4. Dim Product

##### Tạo Dim Product Stage

SELECT \* INTO stg\_products FROM OlistDB.dbo.products;

```

SELECT TOP (1000) [product_id]
      ,[product_category_name]
      ,[product_name_length]
      ,[product_description_length]
      ,[product_photos_qty]
      ,[product_weight_g]
      ,[product_length_cm]
      ,[product_height_cm]
      ,[product_width_cm]
  FROM [olistDWStage].[dbo].[stg_products]

```

	product_id	product_category_name	product_name_length	product_description_length	product_photos_qty	product_weight_g	product_length_cm
1	727e49b05b9a9ef165ef97ebc1ebf9	esporte_lezer	1440	200	2	200	17
2	7281018d2a0311873a8588a75ec	esporte_lezer	442	2	200	17	
3	728160ce1791ed10254854407981	beleza_saude	910	1	650	50	
4	7285e977b56a25203636b2407725e	NULL	NULL	NULL	400	25	
5	7289e8d55e2c96fdde278e257a3d	beleza_saude	940	2	1100	31	
6	728a3cd3c5437c72784d4728edc4	cool_stuff	482	3	200	22	
7	728b05c296264ed3982354261654a	movies_escritorio	1200	1	9000	53	
8	728c17e07e1608631149d261792c4	casa_conforto	216	3	2300	40	
9	728cfefeca3d9054807d7ea94e04	movies_decoracao	357	1	400	30	
10	7295743fcf8049-951783372250fd1	market_place	365	3	10800	100	
11	7297344273110893a12683259717	instrumentos_musicais	647	1	5300	30	
12	7298345180980e002340e2b34df	cool_stuff	873	3	1400	18	
13	729a3e918a4303d97e441ae0e43	movies_cozinha_rea_de_servicio_jenter_e_jardim	444	1	1700	30	
14	729a75e5cb3f785e67edbf84acbc0e	religos_presentes	621	1	350	23	
15	729b654c7d397c89770ad7ab	informatica_acessorios	714	1	3550	30	
16	729c55b45093d18b660b0ec17ca5d	telefonia	702	6	250	18	
17	729d36954326a93a0da562571943	sinalizacao_e_seguranca	1174	3	200	19	
18	729e4de4ffed0e16489fb611c715	fashion_bolsas_e_acessorios	219	4	200	18	
19	72a137b5d56a182cde242493f11843	consoles_games	411	3	600	24	
20	72a137b5d56a182cde242493f11843	utilidades_domesticas	565	3	1100	30	

##### Load vào dim\_product

INSERT INTO dbo.dim\_product (product\_id, product\_category\_name, product\_category\_name\_english, product\_length\_cm, product\_height\_cm, product\_width\_cm, product\_weight\_g)

)

SELECT

p.product\_id,

```

p.product_category_name,
t.product_category_name_english,
p.product_length_cm, p.product_height_cm, p.product_width_cm,
p.product_weight_g
FROM OlistDWStage.dbo.stg_products p
JOIN OlistDWStage.dbo.stg_product_category_translation t ON
p.product_category_name = t.product_category_name;

```

product_key	product_id	product_category_name	product_category_name_english	product_length_cm	product_height_cm	product_width_cm	product_weight_g
1	eb1bb871be49118a98b1cb12e	case_constructao	home_construction	17	14	16	383
2	eb0972c17b2503d4971a1659e8024d	beleza_saudade	health_beauty	22	3	15	200
3	eb1a593d269670408ee90a8754e	movies_decoracao	furniture_decor	70	5	15	7050
4	eb0e99c5c701994cfec246941e37070	utilidades_domesticas	housewares	30	2	13	300
5	ebec0fe5e83056835e11359cf11	telefonio	telephony	30	5	12	200
6	eb0d703e86703291e5e586bd040392	consoles_games	console_games	25	5	15	200
7	eb0f72c24910909a76d245ac	esportes_lazer	sports_leisure	30	30	30	8000
8	eb142016e25450911c52105f0	utilidades_domesticas	housewares	24	24	24	1000
9	eb11a6d2505a11ef0222080177	brinquedos	toys	38	38	26	1874
10	eb1a13032249e8e01785e8d85417985	case_constructao	home_construction	50	80	50	30000
11	eb24ee072330cc0f68f44a74185d716	canape_banho	bed_bath_table	44	6	33	483
12	ebf0d915ea3923498e430e96d59ec	canape_banho	bed_bath_table	32	5	30	1150
13	ebf7d98c166c02e672-573e6bd30458	canape_banho	bed_bath_table	44	6	33	526
14	ebfb0933e46109a411311073d4ff73ba	canape_banho	bed_bath_table	38	6	29	900
15	ebfb94c2e4532cd50064af9fb699	utilidades_domesticas	housewares	28	20	28	700
16	ebfb9edc00e4d469138a3c110f6a85	canape_banho	bed_bath_table	50	8	41	1300
17	ebfb003d9005e4c45-217728413755	bebes	baby	45	11	40	1200
18	ebfb7e1e04d54375b373405a114c32	utilidades_domesticas	housewares	19	11	18	417
19	ec019a7e1eb-20272e3076517c7c6b	brinquedos	toys	65	7	45	10550
20	ec0245d330128f1a18f6e01fd3d8d32	cool_stuff	cool_stuff	45	25	18	2275
21	ec037c3b94465515942793c7ce665a	case_constructao	home_construction	40	13	25	2600
22	ec037c44fa09b6119c2e20e7b798284	canape_banho	bed_bath_table	45	15	35	5900
23	ec1314c23c3f6210wad4fa625ca095	automotivo	auto	32	16	16	600

#### 4.1.1.5. Dim Review

##### Tạo Dim Review Stage

```

SELECT
    review_id,
    order_id,
    review_score,
    review_creation_date
INTO dbo.stg_dim_review
FROM OlistDB.dbo.order_reviews;

```

The screenshot shows the Microsoft SQL Server Management Studio interface. The title bar reads "SQLQuery11.sql - (local)\OlistDWStage (sa (119)) - Microsoft SQL Server Management Studio". The main window has two panes: "Object Explorer" on the left and "Results" on the right. In the "Object Explorer", under "OlistDWStage", there are several database objects like "dbo.dim\_product", "dbo.dim\_review", etc. In the "Results" pane, a query is being run:

```

SELECT TOP (1000) [review_id]
      ,[order_id]
      ,[review_score]
      ,[review_creation_date]
  FROM [OlistDWStage].[dbo].[stg_dim_review]

```

The results grid shows 1,000 rows of data from the "stg\_dim\_review" table. The columns are "review\_id", "order\_id", "review\_score", and "review\_creation\_date". The first few rows of data are:

review_id	order_id	review_score	review_creation_date	
1	d1cda00d072e981cb8f02592b04985b	2	2017-02-18 00:00:00.0000000	
2	80223039e0b07973e9ea60159b52e41a	1	2018-03-29 00:00:00.0000000	
3	b1e0ef0b73514fb6cb2c29e9fb9e0e39	3	2017-12-07 00:00:00.0000000	
4	45e9e7503e8d5e599fb52ec02d7e3fa	1	2018-08-10 00:00:00.0000000	
5	6ca0f2050a10daw5daeb514d026ce34	5	2017-12-02 00:00:00.0000000	
6	029905e7597e7ab050ee7171c13	0412781fb7e2094dd4dbd0520305e294c	4	2018-09-25 00:00:00.0000000
7	f5be452a845539d25445feac063e4c21	1	2018-07-31 00:00:00.0000000	
8	ebed4e942dd0df19c7813557a89c62	1	2017-12-17 00:00:00.0000000	
9	9999da1c307ba370919864fb26444	f5c5f556ab6cefe934426ebe2a130b	1	2018-08-21 00:00:00.0000000
10	20999ef8e27c1cffe5007b433a563	68572b2e2ee176d1f7c9afe7ed15e1	5	2017-09-23 00:00:00.0000000
11	20999e4e035a0093bb542765d7851b	8bc1152e0e15117808a07ad5e50539fc4	4	2017-11-24 00:00:00.0000000
12	209a70531123b4e0c9294ed74d31b	de2d003fb6f4533a828b658c0d9d4fd0	5	2017-05-25 00:00:00.0000000
13	209a90c0d2d2fe7502c2b78e674dd9f9	809c33617442e48ed9b9253aa7397ac	1	2017-12-17 00:00:00.0000000
14	209a9eab0739e6b62b62b1644fc336	b705e7b594407b7b733f39e6b11596e	3	2017-09-15 00:00:00.0000000
15	209c7b9a7e6a855fafa7b8616351e	c84591b9e9b68655ba109e92324de8c	1	2017-12-22 00:00:00.0000000
16	209f52b5fb19e3262675c42d91d8372	5e23dfeefc1eac7cabbcb259e94e2	5	2017-11-17 00:00:00.0000000
17	209f00207e174ab4b4747959e16192	0b30a59b6e985f7983a00753d959n	1	2017-10-11 00:00:00.0000000
18	209f088b76738444893bafa70102	b6f6adacc3a2d545e6b2080403d99	5	2017-03-07 00:00:00.0000000
19	20a16c116551671e99c531c0e61b14	657499524125d2d4e3c6d1031680	4	2018-02-07 00:00:00.0000000
20	20a21820d427e775eae9e0c0cf9e6c44	ae93695a6780a86b7356743d5bcf7	5	2017-08-08 00:00:00.0000000
21	20a2467426a28b3893707a4be800	4eac8362e4d9737df50755b8884bd	1	2018-03-21 00:00:00.0000000
22	20a2788c0-a7364300a2e4b60ca	1b6c5ab6653a9bca12206879e94fe4e	5	2018-04-06 00:00:00.0000000
23	20a2d3340f96c83090ebf33246734	2c274f00603339907c238d0d9f166d	4	2017-06-06 00:00:00.0000000
24	20a545e7b41fa7c3a0d32133a436cd	2b9c1a63342d6262cb6d11189e65f	5	2018-03-01 00:00:00.0000000
25	20a5600b776bb51800e161e3794ee	9855472a25a51c2e8b310d2125c171511	5	2017-09-16 00:00:00.0000000
26	20a56bbdc123a493d7769dcbc350d	42be6028bdc8516e5b40059b0a3	5	2018-04-14 00:00:00.0000000

At the bottom of the results grid, a green checkmark icon indicates "Query executed successfully."

## Load dim\_review

```

INSERT INTO dim_review (
    review_id,
    order_id,
    review_score,
    review_creation_date
)
SELECT
    review_id,
    order_id,
    review_score,
    review_creation_date
FROM OlistDWStage.dbo.stg_dim_review;

```

```

SQLQuery12.sql - (local).OlistDW (sa (71)) - Microsoft SQL Server Management Studio
File Edit View Query Project Tools Window Help
Object Explorer
Connect
FileTables
Tables
Graph Tables
dbo.dim_customer
dbo.dim_date_time
dbo.dim_order_items
dbo.dim_order_payments
dbo.dim_orders
dbo.dim_product
dbo.dim_review
dbo.dim_seller
dbo.dim_fulfillment
dbo.fact_product_seller_performance
dbo.fact_sales
Dropped Ledger Tables
Views
External Resources
Synonyms
Programmability
Query Store
Service Broker
Storage
Security
OlistDWStage
Database Diagrams
Tables
System Tables
FileTables
External Tables
Graph Tables
dbo.dim_review
SELECT TOP (1000) [review_key]
,[review_id]
,[order_id]
,[review_score]
,[review_creation_date]
FROM [olistDW].[dbo].[dim_review]

```

review_key	review_id	order_id	review_score	review_creation_date
1	5ea259f8db2357776c3d327a125f	75b807e20f82b171c29a7e140e883	5	2018-05-12 00:00:00.000
2	5ea259f8db2357776c3d327a125f	e57d5707ebe95772890c99330e7e21	5	2018-05-16 00:00:00.000
3	5ea333ff40e053525cfa5de2901	3315e3e770d0eab3c0e7e431fe0420d	5	2018-07-12 00:00:00.000
4	5ea3e4910653e0038814b6e74b782	9c0779514556beb3e347b9eb0934390	1	2017-11-24 00:00:00.000
5	5ea4edc3a7c4859e73f922a364ef8b	33456500c4acada9e036959e7f9f4	4	2018-01-20 00:00:00.000
6	5ea54ewb8a70b67cefb138e803c0	1476194dd3d7c57e304b3e6af526b	2	2018-03-30 00:00:00.000
7	5ea54ewb8a70b67cefb138e803c0	2768684ef03d4a4e07ed8d2e908c	1	2017-06-08 00:00:00.000
8	5ea58764e21d2f4545867739e2c2d	41229e7db11ff4cc3e4a4e49907	5	2017-09-14 00:00:00.000
9	5ea5c35d8d990a2455cab2868e78	b467a92cd0e5f5e6a44becc1418423	4	2018-01-20 00:00:00.000
10	5ea69e30559686144ddfb530327ef1	343abed8770e0b747df81a6073235	5	2017-08-29 00:00:00.000
11	5768c9fc9334e0317c4ca54fb	fe02391540ed4dd5d5b711f8703e	5	2017-05-20 00:00:00.000
12	5780d417cfc58a74d14ef69569	4495b7542262020518de210231e	4	2017-09-24 00:00:00.000
13	58a8cf79688fc50642e6e72124d	39b675444287eb7c9a354e4cab	2	2018-08-04 00:00:00.000
14	58ad8fb263020569e60504c04254	edd9f97bd3da8e3330534fb902792	4	2017-08-08 00:00:00.000
15	5aa9f1149490b06398419840d8b7	c8c7e8563c295b0e3ee5fe0a1b0c	1	2018-02-07 00:00:00.000
16	5aa9f708e2b5c7abd0fe02c2738	c738688e1adff662d0e001fb1ba75d	3	2017-08-01 00:00:00.000
17	5acc04d0c7c3e561e93c11921b26	7f86602d0d5230065698a14095fb	5	2017-11-02 00:00:00.000
18	5ad6f30559686144ddfb530327ef1	9761e998e3d3309a1083346a5e5f11	5	2018-04-25 00:00:00.000
19	5ad6fd827c7e07d0ebe9e944fb62c	d36515b9f9d93a374b19794789206	5	2017-07-14 00:00:00.000
20	5ad921e17cfc58a74d14ef69569	f10305517d52fbb61e933d381c9d5	4	2017-05-13 00:00:00.000
21	5adeed853c8d3c8e6605a8dc76d	06b3987f5a599832e9ab13b0498f	4	2017-10-28 00:00:00.000
22	5ae27601999ebccfc6767c5ec759	432916af4f581e9b636960e208d676	5	2017-12-16 00:00:00.000
23	5ae3cb375b399520c956904a805d	1e7eadaa05b5b40252614679424991	1	2018-01-23 00:00:00.000
24	5ae60917b986b51721fb2494a3c349	410f7995035b4027006761234498c	5	2018-04-12 00:00:00.000
25	5ae620573e44c7e33112eef0025	25ce2a2625372056220ebe496f593	5	2018-03-25 00:00:00.000

Query executed successfully.

#### 4.1.1.6. Dim Orders

##### Tạo Dim Orders Stage

```
SELECT * INTO stg_orders FROM OlistDB.dbo.orders;
```

```

SQLQuery13.sql - (local).OlistDWStage (sa (68)) - Microsoft SQL Server Management Studio
File Edit View Query Project Tools Window Help
Object Explorer
Connect
Tables
System Tables
External Tables
Graph Tables
dbo.stg_customers
dbo.stg_dim_review
dbo.stg_order_items_mix
dbo.stg_order_payments
dbo.stg_order_reviews
dbo.stg_orders
dbo.stg_product_category_translation
dbo.stg_products
dbo.stg_sellers
Dropped Ledger Tables
Views
External Resources
Synonyms
Programmability
Query Store
Service Broker
Storage
Security
OlistDWStage
Database Diagrams
Tables
System Tables
FileTables
External Tables
Graph Tables
dbo.stg_order_items
SELECT TOP (1000) [order_id]
,[customer_id]
,[order_status]
,[order_purchase_timestamp]
,[order_approved_at]
,[order_delivered_carrier_date]
,[order_delivered_customer_date]
,[order_estimated_delivery_date]
FROM [OlistDWStage].[dbo].[stg_orders]

```

order_id	customer_id	order_status	order_purchase_timestamp	order_approved_at	order_delivered_carrier_date	order_delivered_customer_date	order_deliver
1	11c931dc4ed4a62a90dd1248a7673	73d31093311c1a1b2148993979e7	delivered	2018-01-24 20:56:37.0000000	2018-01-26 21:49:00.0000000	2018-01-26 21:49:00.0000000	2018-01-26 21:49:00.0000000
2	11c971579819d94a3b70811b3e74	7681bc1b6e4588e761d7c184052984	delivered	2018-07-25 16:09.1000000	2018-07-26 03:25.16.0000000	2018-07-26 11:33.00.0000000	2018-07-30 1:
3	11c971505db128db326d471715bd5d	5691c0364024569792b50d686dc	delivered	2018-02-12 10:17.11.0000000	2018-02-15 03:51.05.0000000	2018-02-19 23:42:34.0000000	2018-05-13 0
4	11cb97108fb8cfc7ef093d68e5b6c	#f93a140565309beed1ab0c939146	delivered	2018-02-26 15:07.43.0000000	2018-02-27 04:31.11.0000000	2018-02-27 17:33.02.0000000	2018-09-10 1
5	11cc52169384d794c10ab040b120	#f82b32d5ed45961864102014064	delivered	2017-10-24 21:36:39.0000000	2017-10-24 21:50.03.0000000	2017-10-24 20:52:50.00.0000000	2017-11-07 1:
6	11cd24b7e1b4316b03156c03156e	198ce6d4081b571e762e5cc07	delivered	2018-02-26 18:56.04.0000000	2018-02-09 02:50.11.0000000	2018-02-15 17:29:09.0000000	2018-02-23 1
7	11cf686b51d6e060001ad7c5eb08	#f06986880e591dfe953672e6f597	delivered	2018-01-16 06:25:36.0000000	2018-01-16 06:36-19.0000000	2018-01-16 12:29:02.0000000	2018-01-07 0:
8	11cfb7eff39a26998a94e7832b16c	3499bb8409c2d9e2e403d94d415	delivered	2018-08-20 22:23:51.0000000	2018-08-20 22:35:14.0000000	2018-08-22 10:06.00.0000000	2018-08-29 1:
9	11d0085fc1ff1c2e630c93e29e790	#f926473d2d526909b1e282959809434	delivered	2018-02-28 21:49:44.0000000	2018-02-28 22:27.47.0000000	2018-03-01 20:42:39.0000000	2018-09-21 1
10	11d1d2a4d11b0de56991a2869251a533739a	4e8767a3db2151930c15cf68f50	delivered	2018-02-09 13:48.09.0000000	2018-02-09 14:09.03.0000000	2018-02-15 20:07:54.0000000	2018-02-21 2:
11	11d2e411b0de56991a2869251a533739a	#f9172289991927e3a29727	delivered	2017-03-08 17:14.18.0000000	2017-03-08 17:14.18.0000000	2017-03-08 10:21:30.0000000	2017-09-21 1
12	11d4140c9e0924574743744	#f00859327ea16e296590459865e	delivered	2018-01-03 11:53.11.0000000	2018-01-03 12:07.13.0000000	2018-01-03 20:18:35.0000000	2018-01-09 1:
13	11d4183301049a7539e0232d3a21e5cb	61b0c07d93a520959a32ad1e5cb5b	delivered	2017-07-23 15:39.56.0000000	2017-07-23 15:50.13.0000000	2017-07-24 18:49:33.0000000	2017-07-25 2:
14	11d5b3b0e903259852121a3e2128	#f8a585373652cd37726286b58	delivered	2017-04-13 12:05.15.0000000	2017-04-14 13:01.54.0000000	2017-04-14 13:01.54.0000000	2017-04-14 13:01.54.0000000
15	11d5b610e903259852121a3e2128	#f105610e903259852121a3e2128	delivered	2017-12-18 17:05.53.0000000	2017-12-18 17:13.30.0000000	2017-12-19 22:45.61.0000000	2018-01-12 2:
16	11d6810ab0d982e6d7b4a374a	#c512e2baab0779fe64b20422b20	delivered	2017-04-04 16:33.30.0000000	2017-04-04 16:45.21.0000000	2017-04-04 10:41.27.0000000	2017-04-10 1:
17	11d71704c92126d7e279987e727987b	#f06210380609b762957e6f0d1	delivered	2017-06-10 21:12.00.0000000	2017-06-10 21:23.35.12.0000000	2017-06-12 09:35.50.0000000	2017-07-03 1:
18	11d8e2b1f6e6c1854049846852404	#f05cde0103804371130309651b2b	delivered	2017-11-26 12:08.00.0000000	2017-12-06 12:25.00.0000000	2017-11-28 19:11.00.0000000	2017-12-05 1:
19	11d916b104e2a0d7eab7e0a71e7457798	#f0194532c1e791304984589	delivered	2018-01-30 12:05.00.0000000	2018-01-30 13:05.00.0000000	2018-02-02 17:14.51.0000000	2018-02-02 17:
20	11d96939eaef7e748531a7e757798	#f92210e9ef0d5e5c1b77eef	delivered	2018-04-17 10:59.29.0000000	2018-04-17 15:55.58.0000000	2018-04-24 22:12.27.0000000	2018-04-24 22:12.27.0000000
21	11d6868cfcf7f13e1e9c2b682969	16287652c8a434bdc86b8653441861	delivered	2017-12-04 11:26.11.0000000	2017-12-14 16:31.35.0000000	2017-12-14 14:17.01.0000000	2017-12-14 14:

Query executed successfully.

##### Load vào dim\_orders

```

INSERT INTO dbo.dim_orders (
    order_id, customer_id, order_status,
    order_purchase_timestamp, order_approved_at,
    order_delivered_carrier_date, order_delivered_customer_date,
    order_estimated_delivery_date
)
SELECT

```

```

order_id,
customer_id,
order_status,
order_purchase_timestamp,
order_approved_at,
order_delivered_carrier_date,
order_delivered_customer_date,
order_estimated_delivery_date

```

FROM OlistDWStage.dbo.stg\_orders;

```

SQLQuery15.sql - (Local).OlistDW (sa (68)) - Microsoft SQL Server Management Studio
File Edit View Query Project Tools Window Help
File New Query File Open Execute
SQLQuery15.sql (L).OlistDW (sa (68)) OlistDW_Load.sql (L).OlistDW (sa (70))
SELECT TOP (1000) [order_key]
      ,[order_id]
      ,[customer_id]
      ,[order_status]
      ,[order_purchase_timestamp]
      ,[order_approved_at]
      ,[order_delivered_carrier_date]
      ,[order_delivered_customer_date]
      ,[order_estimated_delivery_date]
  FROM [olistDW].[dbo].[dim_orders]

110 %
Object Explorer Results Messages
order_key order_id customer_id order_status order_purchase_timestamp order_approved_at order_delivered_carrier_date
1 11c931d6ce4ef462a90dd1248e9076716 73d1093311fc1a1eb2148fb93979ef8 delivered 2018-01-24 20:56:37.0000000 2018-01-24 21:18:22.0000000 2018-01-26 21:49.0000000
2 11c971575881770811b3e74 768f1bc64588a7617d7c9144052984 delivered 2018-07-25 16:09.10.0000000 2018-07-26 03:25.16.0000000 2018-07-26 11:33.00.000000
3 11c971505d128b3b46471c8d0ff 56f1c0364002a5e9ff2b500066d4 delivered 2018-02-12 10:17.11.0000000 2018-02-15 03:51.05.0000000 2018-02-19 23:42.34.000000
4 11c9714045053098ed18a956393146 f38263d3ed5459b1864102014064 delivered 2018-02-26 15:07.43.0000000 2018-02-27 04:31.11.0000000 2018-02-27 21:17.33.02.000000
5 11cc65219638e4d794c1d0d040e120 f38263d3ed5459b1864102014064 delivered 2017-10-24 21:36.39.0000000 2017-10-24 21:50.03.0000000 2017-10-26 20:52.50.000000
6 11cd24674b9463f1b6a03156e80cb0 198c6ed4081b5713a7625c5c07 delivered 2018-02-08 18:56.04.0000000 2018-02-09 02:50.11.0000000 2018-02-15 17:29.09.000000
7 11c686bd51a5e060001ad7c3ae0950e e8098868a591dfef8f562d5857 delivered 2018-01-16 06:25.36.0000000 2018-01-16 06:19.0000000 2018-01-18 12:29.02.000000
8 11cf7b7f639426988934ee7832f16c 34998b609-02d9e2a4ec03604a15 delivered 2018-09-20 22:23.51.0000000 2018-09-22 10:06.00.0000000 2
9 11d0005cda1f6c32aa530cb92ea970 19e2647d2b3262669bae28295850434 delivered 2018-02-28 21:49.44.0000000 2018-02-28 22:27.47.0000000 2
10 11d1d2c41418db528262561a553739 4a8767e3d0b2511930c15cfef8ff delivered 2018-09-09 13:48.09.0000000 2018-09-15 20:07.54.0000000 2
11 11d2e2111b1bde559991e1ccbe40e0e7 07436d501e712289991218e3e29727 delivered 2017-09-08 17:14.18.0000000 2017-03-08 17:14.18.0000000 2
12 11d4103c3e0a0292956e43741493347 608e69327ea16e2895094596a05 delivered 2018-01-03 11:53.11.0000000 2018-01-03 20:18.35.0000000 2
13 11d4183301045e75d359c3b36894eb 61b0df47d3a50ccac2ed1e5cf58 delivered 2017-09-23 15:39.56.0000000 2017-07-23 15:39.56.0000000 2
14 11d585b9a0e32d7e6eb9a148374 8fa588327ea16e2895094596a05 delivered 2017-04-14 13:01.54.0000000 2017-04-18 12:23.27.0000000 2
15 11d5dfdef070419ab5ab55e5a55c 841e167ea97dc14040511edf85220 delivered 2017-18-17 03:53.0000000 2017-12-18 17:13.30.0000000 2
16 11d6810b0d8982d3e2d67b4de374a c5b17d2b9d9799fe8d820422b28c6 delivered 2017-04-16 16:45.21.0000000 2017-04-16 10:41.27.0000000 2
17 11d7140d68b215627779876e27e932 a5943e1c908eb9a3b8523678c6e1 delivered 2017-01-10 23:12.16.0000000 2017-06-10 23:35.12.0000000 2
18 11d8e2bcb681c83049964592462402 d05cdaf1030d037113c70396951b2e08 delivered 2017-11-26 12:25.00.0000000 2017-11-28 19:11.30.0000000 2
19 11d91606942add7e6eb9a148374 5914a5332c1e79119134d6fa5c4589 delivered 2018-01-30 13:40.53.0000000 2018-01-30 13:51.29.0000000 2
20 11d9619a6af7f478ee17ed757750 fd925126e9d4da5c5c1b71ef2301 delivered 2018-04-13 10:39.29.0000000 2018-04-17 05:55.36.0000000 2018-04-24 23:32.17.0000000 2

```

#### 4.1.1.7. Dim Orders Items

##### Tạo Dim Orders Items Stage

SELECT \* INTO stg\_order\_items FROM OlistDB.dbo.order\_items;

```

SQLQuery16.sql - (Local).OlistDWStage (sa (68)) - Microsoft SQL Server Management Studio
File Edit View Query Project Tools Window Help
File New Query File Open Execute
SQLQuery16.sql (L).OlistDWStage (sa (68)) OlistDW_Load.sql (L).OlistDW (sa (70))
SELECT TOP (1000) [order_id]
      ,[order_item_id]
      ,[product_id]
      ,[seller_id]
      ,[shipping_limit_date]
      ,[price]
      ,[freight_value]
  FROM [olistDWStage].[dbo].[stg_order_items]

110 %
Object Explorer Results Messages
order_id order_item_id product_id seller_id shipping_limit_date price freight_value
1 320e058336505d41a1e8331793d24a 1 000420000000050964a17e949e079 0051e13d97139e498fb125e1a5951b 2017-12-04 14:56:24.0000000 159.899993696484 21.79999804
2 320e058336505d41a1e8331793d24a 1 368e627700942b7011e6e023807e3729 1659020178ef1db949020352100 2017-12-04 03:16.0000000 49 13.39999806
3 320e058336505d41a1e8331793d24a 1 368e627700942b7011e6e023807e3729 1659020178ef1db949020352100 2017-12-04 03:25.16.0000000 49 13.39999806
4 320e058336505d41a1e8331793d24a 1 446b37509502a3b28d8f77-692989 b2479944a1b80-08d8-1b4d5c84346 2018-05-10 20:55.10.0000000 24.899997711182 18.22999847
5 320e058336505d41a1e8331793d24a 2 ee3d777a143e998e2099572d596c2 1 63085b2496d5-2d7b632bb59a1403a 56decfa20747a92ff87e997d4f8c2 2017-06-07 21:02:32.0000000 368 58.81999894
6 320e058336505d41a1e8331793d24a 1 013eaa3e93b6-4e11ff-4e4d-8e5c-3c1 640a31b7d1e78714e9a1b23a9904004c 640a31b7d1e78714e9a1b23a9904004c 2018-05-28 02:15.36.0000000 69.8999978737095 14.57999982
7 320e058336505d41a1e8331793d24a 1 d08847e02a0203e13c6d3374 7ee4334b125ef02d8a2b318029194 2017-05-09 08:30.19.0000000 87.8999978737095 10.8999981
8 320e058336505d41a1e8331793d24a 1 448e04b09-03d5-4e88-3e0b-a0ba2039 0ea22ctfrdc759fb9a4b1c9c16043 2017-06-12 14:30.26.0000000 24.899996158303 16.7999984
9 320e058336505d41a1e8331793d24a 1 448e04b09-03d5-4e88-3e0b-a0ba2039 bdb4tvefc277925b9e160992027c76 9556e0216e6b617e5c0517800e9 2017-09-26 15:10.21.0000000 85 16.21999831
10 320e058336505d41a1e8331793d24a 1 32ee863a1273bb63a03bceda769f0 1 0ca27d49fb1049b63e0ba18a4d137 e06999c04eac2 10779ef1cd3c3d19 2017-12-29 02:12.40.0000000 155 16.850000381
11 320e058336505d41a1e8331793d24a 1 32ee863a1273bb63a03bceda769f0 1 0ca27d49fb1049b63e0ba18a4d137 e06999c04eac2 10779ef1cd3c3d19 2017-12-29 02:12.40.0000000 155 16.850000381
12 320e058336505d41a1e8331793d24a 1 689b01074243e4ee72949fe2d42e22 3d871de0142e0907814e2b9d1733cb1 2017-06-05 22:50.16.0000000 68 15.22999847
13 320e058336505d41a1e8331793d24a 1 32ee863007584467647a1 170767e2b2 1 e03072009565fd9bf8d7d7d9e4b0 6a6058816e1779f2d8e5b458f6 2018-03-23 15:50.58.0000000 16.899996185303 12.78999861
14 320e058336505d41a1e8331793d24a 1 b00ae73288344d1e8e5-3e789d 99494b371e2813e2-2dfe0f6edc2 2018-01-18 02:08.41.0000000 69.9000015259789 11.77999847
15 320e058336505d41a1e8331793d24a 1 b3799476d7d27c3a4-42d236e2b2 23c3d8ebaf4e259a30fd8d958a13 2018-07-06 18:30.37.0000000 74.9000015258789 15.61999888
16 320e058336505d41a1e8331793d24a 1 5c3aeac078a3b3b733262e7e2cc 056b4e4a5b2c5d2c7842547d6e9b51 2018-05-15 11:51.47.0000000 159.900005493164 14.560000416
17 320e058336505d41a1e8331793d24a 2 5c3aeac078a3b3b733262e7e2cc 056b4e4a5b2c5d2c7842547d6e9b51 2018-05-15 11:51.47.0000000 159.900005493164 14.560000416
18 320e058336505d41a1e8331793d24a 1 fc144989e59d9eb61b902c1b84f1 2 25fc5916360744a97b143e2b353531 2017-11-12 03:50.31.0000000 149.899995171188 18.22999823
19 320e058336505d41a1e8331793d24a 2 fc144989e59d9eb61b902c1b84f1 2 25fc5916360744a97b143e2b353531 2017-11-12 03:50.31.0000000 149.899995171188 18.22999823
20 320e058336505d41a1e8331793d24a 3 fc144989e59d9eb61b902c1b84f1 2 25fc5916360744a97b143e2b353531 2017-11-12 03:50.31.0000000 149.899995171188 18.22999823
21 320e058336505d41a1e8331793d24a 1 ae9077e47-13906700ee9f84759def 2646bbe62d492e49047a4589b8d2 2018-08-06 13:10.23.0000000 31.200007629395 7.48000019
22 32113e68183197d6671e2d37823bd 1 e0e247176507078fc312c16904d74 86b42a50976166466d60c84822545 2018-08-14 17:49.0000000 110 20.81999694

```

## Load vào dim\_order\_items

```
INSERT INTO dbo.dim_order_items (
    order_id, order_item_id, product_id, seller_id,
    shipping_limit_date, price, freight_value
)
SELECT
    order_id,
    order_item_id,
    product_id,
    seller_id,
    shipping_limit_date,
    price,
    freight_value
FROM OlistDWStage.dbo.stg_order_items;
```

The screenshot shows the Microsoft SQL Server Management Studio interface. A query window titled "SQLQuery18.sql - (local).OlistDW (sa (84)) - Microsoft SQL Server Management Studio" is open. The query is:

```
SELECT TOP (1000) [order_item_key]
    ,[order_id]
    ,[order_item_id]
    ,[product_id]
    ,[seller_id]
    ,[shipping_limit_date]
    ,[price]
    ,[freight_value]
FROM [olistDW].[dbo].[dim_order_items]
```

The results pane displays the data being inserted. The columns are:

order_item_key	order_id	order_item_id	product_id	seller_id	shipping_limit_date	price
1	4122675696367772e11035e75391a	1	cecd09725da5ed01471d9a505e73989a	4d6b651bd7684afcffabdf08d12e5a	2016-05-15 10:14:44.0000000	69.000001525878
2	0426e46bb0a9fb67a417100fbfcxa71	1	55b0a296824029e9a209eb3d200e69d0	6c750c241cc3cd2fd834bd469b34e073	2017-10-17 05:14:47.0000000	37.5
3	0426769bb583d5d28b688247d5809	1	a30c310dd4ad467282787e075cbcd	b1d19ed5164d45d118d57819e1891f	2017-12-04 10:38:20.0000000	37.000001525878
4	0435262e3f3b02bb7523906d5774a8	1	b6c89bc5fa5dc01a5d0d97ff06ec6cb	6d652c528db3b8be2ea044946b811f8	2017-11-29 19:15:27.0000000	105.5999847412
5	0435262e3f3b02bb7523906d5774a8	2	b6c89bc5fa5dc01a5d0d97ff06ec6cb	6d652c528db3b8be2ea044946b811f8	2017-11-29 19:15:27.0000000	105.5999847412
6	0455363b0204ee1450c35392d5de5	1	b273f8b24b6e11a16ebf1d0a34ed	d919fb7d941e83d64a0053cf7d7e4f	2018-04-26 19:31:37.0000000	14.10000031469
7	046cd8e6a85ade2864cd56811fb9b	1	b6e8e5e1259a9fe445bc1771404709	65522040332627w0c4949a17436694	2018-05-18 11:58:42.0000000	13
8	046cd8e6a85ade2864cd56811fb9b	2	b6e8e5e1259a9fe445bc1771404709	65522040332627w0c4949a17436694	2018-05-18 11:58:42.0000000	139
9	046cd8e6a85ade2864cd56811fb9b	3	b6e8e5e1259a9fe445bc1771404709	65522040332627w0c4949a17436694	2018-05-18 11:58:42.0000000	139
10	0472c235c53b5e43c042ee73b1290	1	7d046599f780f07991e8b6999e5c5d	f01c13f2614d705c1749cb52fec0d94	2018-03-12 03:55:36.0000000	124.900001525878
11	0476342e4b3d7b33bf86e0899d685	1	8b8702416a4ec6b596700596ce372	8b2b0d966340356767a22d357b53368c	2017-09-04 20:44:19.0000000	15.47999842236
12	0487787045a20403546af3fc8d81	1	2ff995eae9d63a157a07b3664e3d3	8bb68ee4e47e7d09f57979c0b38324d	2018-07-24 16:31:45.0000000	72.989997863769
13	048997e7d1fc0179c1a374b5cc01a52	1	575900a8fbfd3d374b0be7b94fa31	30a29535b4830991d09e4a4fc84bb	2017-11-30 22:37:31.0000000	49.000001525878
14	04917d186906d8147122234788026	1	31c242e87390e8777de0dfbabc182d	4-039b9d4c11ew2cb5c8c-9ef942d0	2017-05-24 17:05:11.0000000	128.899993939648
15	0492792b196e9a030225742e0751a5c	1	8ed094b6078c5686bb10leads3t75d	7c671448b009e098a365ca8d010ab	2018-05-29 09:14:41.0000000	179.899993939648
16	0494352a5c97302053f3fb2b14eb4	1	bff0dc0a1907b0ac504ec12999f4d	10250e0244d7041d5fb5b6550e0fba	2018-06-21 09:31:50.0000000	215
17	0494352a5c97302053f3fb2b14eb4	2	bff0dc0a1907b0ac504ec12999f4d	10250e0244d7041d5fb5b6550e0fba	2018-06-21 09:31:50.0000000	215
18	049679f66493d9363d06545c2325	1	ad21c6ff97ed9d7a3c7351bca2e0	execcfbcff9904a2d94a0593d9eef2b	2018-05-18 17:21:28.0000000	47.5999847412
19	049692062589698900dd566cc3763	1	2b4609894a2e1874494203496e318	c419e0505a3c5b877189a18327556a	2018-04-11 05:30:19.0000000	79.899997863769
20	04b476993a3c0c124b25eab6e4	1	a5f1d485d52259a2e182b1b50156e6	5dc0ca129747e792ff87a997dc4f8c8	2017-05-31 13:10:17.0000000	259.899993939648
21	04ecdcc723f16238835a21291b7d0	1	aeeads71d1592edff30127249662b	bf2d7a966e91cfa2fb25d76589f0d3	2018-02-10 10:30:47.0000000	99

At the bottom of the results pane, it says "Query executed successfully." The status bar at the bottom right shows "(local) (16.0 RTM) sa (84) OlistDW 00:00:00 1.000 rows".

### 4.1.1.8. Dim Orders Payment

#### Tạo Orders Payment Stage

```
SELECT * INTO stg_order_payments FROM OlistDB.dbo.order_payments;
```

```

SQLQuery19.sql - (local)\OlistDWStage (sa (69)) - Microsoft SQL Server Management Studio
File Edit View Query Project Tools Window Help
0035246a40f520710769010f752
Object Explorer
SQLQuery19.sql - (local)\OlistDWStage (sa (69)) 0035246a40f520710769010f752
SELECT TOP (1000) [order_id]
    ,[payment_sequential]
    ,[payment_type]
    ,[payment_installments]
    ,[payment_value]
FROM [olistDWStage].[dbo].[stg_order_payments]

```

order_id	payment_sequential	payment_type	payment_installments	payment_value
1	1	credit_card	10	158.97995727539
2	1	credit_card	5	58.450007620395
3	1	credit_card	1	47.899999399484
4	1	credit_card	6	328.170013427734
5	1	credit_card	9	98.2200012207031
6	1	credit_card	8	239.899999399484
7	1	boleto	1	136.75
8	1	credit_card	1	49.659999474121
9	1	credit_card	1	106.25
10	1	credit_card	4	40.409999474121
11	1	credit_card	7	76.389999399484
12	1	credit_card	1	4.2899996185303
13	1	voucher	1	25
14	1	credit_card	10	368.200012207031
15	1	credit_card	2	57.4300003051758
16	1	credit_card	4	96.550003051758
17	1	boleto	1	445.190002441406
18	1	credit_card	3	156.639999399484
19	1	credit_card	2	104.870002746882
20	1	credit_card	8	161.860000910352
21	1	credit_card	1	152.190002441406
22	1	credit_card	1	28.620000392334
23	1	voucher	1	9.7700004770367
24	1	voucher	1	88.0100021362205
25	1	boleto	1	88.1500012526789
...	...	...	...	...
997	1	credit_card	4	213.899999399484

(local) (16.0 RTM) sa (69) OlistDWStage 00:00:00 1.000 rows

## Load vào dim\_order\_payments

```

INSERT INTO dim_order_payments (
    order_id,
    payment_sequential,
    payment_type,
    payment_installments,
    payment_value
)
SELECT
    order_id,
    payment_sequential,
    payment_type,
    payment_installments,
    payment_value
FROM OlistDWStage.dbo.stg_order_payments;

```

```

SELECT TOP (1000) [order_payment_key]
      ,[order_id]
      ,[payment_sequential]
      ,[payment_type]
      ,[payment_installments]
      ,[payment_value]
   FROM [OlistDW].[dbo].[dim_order_payments]
  
```

order_payment_key	order_id	payment_sequential	payment_type	payment_installments	payment_value
1	4075332589365776d8e39e8807e6	1	credit_card	10	158.97995727539
2	407840763884412747e4c12c762a8	1	credit_card	5	58.4500007829395
3	407850d8c0e80f6ed2900605085cd3b49	1	credit_card	1	47.3899993896484
4	4078d8c769902585ff4fd539e8157	1	credit_card	6	328.17001347734
5	40790d8d2221468e6bb1c5dccc892d8	1	credit_card	9	98.2200012207031
6	4079ff24a51b505325e2d29fb9954d9	1	credit_card	8	230.89999389648
7	407945a9e6cb1a7d6c35232a4996478	1	boleto	1	136.75
8	4078c6773b340c8ecbf863e0768	1	credit_card	1	49.6599998474121
9	407c915b14e1ef7b6ab0227272cd9	1	credit_card	1	106.25
10	407c9b02c7575d6f771c2b39e6c26914	1	credit_card	4	40.4099998474121
11	407c9b02c7575d6f771c2b39e6c26914	2	credit_card	7	76.3899993896484
12	407cd5618b8939e00b5727593a5e	1	credit_card	1	4.28999996185303
13	407cd5618b8939e00b5727593a5e	2	voucher	1	25
14	407cda1a1714d90e355c574d7b49	1	credit_card	10	368.200012207031
15	407d3134dca95d544c41169993c3c6	1	credit_card	2	57.4300003051758
16	407ee2ab7f5c2026801124b7b7d7d	1	credit_card	4	96.5500003051758
17	407fe4674ea7829e324cfw7e7e16	1	boleto	1	445.190002441406
18	40804972099603061f2798d1338d7e1	1	credit_card	3	156.639999389648
19	4082016610c0d8b53aaed0116a068c4b3	1	credit_card	2	104.870002746582
20	4082a9e70f0cded54c17e25e2a50	1	credit_card	8	161.860000610352
21	4083e01516138337c09bc24bfe2f64	1	credit_card	1	152.190002441406
22	4083b3de383063296da5c588ddde6	1	credit_card	1	28.620000839234
23	4086b3a5550b0effc5c94eb7b116fe7e	1	credit_card	1	9.77000045776367
24	4086b3a5550b0effc5c94eb7b116fe7e	2	voucher	1	86.0100021962305

Query executed successfully.

## 4.1.2. Load dữ liệu vào các Fact

### 4.1.2.1. Fact Order Fulfillment

```

INSERT INTO dbo.fact_order_fulfillment (
    customer_key, seller_key, date_time_key,
    order_key, order_item_key, order_id,
    customer_id, seller_id,
    order_purchase_date, order_approved_date,
    order_delivered_carrier_date, order_delivered_customer_date,
    order_estimated_delivery_date,
    delivery_days, estimated_days, late_delivery
)
SELECT
    dc.customer_key, ds.seller_key, d.date_time_key,
    o.order_key, oi.order_item_key, o.order_id,
    o.customer_id, oi.seller_id,
    o.order_purchase_timestamp, o.order_approved_at,
    o.order_delivered_carrier_date, o.order_delivered_customer_date,
    o.order_estimated_delivery_date,
    DATEDIFF(DAY, o.order_purchase_timestamp,
    o.order_delivered_customer_date),
    DATEDIFF(DAY, o.order_purchase_timestamp,
    o.order_estimated_delivery_date),
    CASE
        WHEN
            (o.order_delivered_customer_date - o.order_purchase_timestamp) > 0
            AND
            (o.order_estimated_delivery_date - o.order_purchase_timestamp) > 0
            AND
            (o.order_delivered_customer_date - o.order_estimated_delivery_date) > 0
        THEN 1
        ELSE 0
    END AS late_delivery

```

```

WHEN o.order_delivered_customer_date >
o.order_estimated_delivery_date THEN 1
ELSE 0
END
FROM OlistDW.dbo.dim_orders o
JOIN OlistDW.dbo.dim_order_items oi ON o.order_id = oi.order_id
JOIN OlistDW.dbo.dim_customer dc   ON dc.customer_id = o.customer_id
JOIN OlistDW.dbo.dim_seller ds    ON ds.seller_id = oi.seller_id
JOIN OlistDW.dbo.dim_date_time d   ON d.date_time_key =
FORMAT(o.order_purchase_timestamp, 'yyyy-MM-dd HH:mm') + ':00';

```

fact_order_fulfillment_key	customer_key	seller_key	date_time_key	order_key	order_item_key	order_id	customer_id	seller_id
1	37904	1128	2016-10-03 09:44:00	40305	45126	3b697a20d9e427646d92567910af6d57	355077884019f7f60a031856bd7262b8	522620ddc18e61
2	47922	3339	2016-10-03 16:59:00	94702	86426	b65b2c0d14d807e245451d119d9	7ec40b22510f6be1b08921d3d9e3d8	096769d23495e
3	33839	3293	2016-10-03 21:01:00	77623	31046	65d1e226d0eb8dc4266542522d14	70fc57ee0292675927697e03ad3f5	ed49d67c7ade4
4	51164	953	2016-10-03 21:13:00	49762	65546	a41c8759fb7eab36e07e0382d44465	6f989332712d3222b6571b1cf5835ce	45d33f715e24d1
5	73146	2858	2016-10-03 22:06:00	78382	78373	d207c272675637bfe0002eef00818	b6c41b9e97e9759572d32628d1ab7e7	cca3071e5e9bb5
6	29936	2502	2016-10-03 22:31:00	84999	74932	cdb3b8574d2b4261296502690c...	7812fc0bcf5e0065d31e1b5b00017d8e	b49c500298407f
7	88915	1821	2016-10-03 22:44:00	58754	77645	ae8e80e4b03c5e4b9a0c072c164b181	e695984334d1d53b6d685266998ba12	817855db56a3
8	98285	2858	2016-10-03 22:51:00	56062	107692	ef1b29b591d31d57c0d733746d0d83c9	dc607d958d6115d5d4df2a70a6d3c4	cca3071e5e9bb5
9	13536	714	2016-10-04 09:06:00	4918	1561	0a0537a5eew9e7e9c2b1fa531944227	4f3f7780222eef22399e52d2c47edf3	3481as57cd91f9
10	96998	1016	2016-10-04 09:16:00	4993	24907	f1217ea612f6c7c4255c5b6e931c8	b3e98t20037593c5c6991919356d	4b1eadfd791bd
11	64130	1265	2016-10-04 09:59:00	9888	100498	ed8c7b13e2b56c70ce0c74231e1da88	da08a29935ca5e46100e3bca93d8	5b179e9e8cc7a7
12	47131	1363	2016-10-04 10:16:00	57645	98877	c3d8a402b60ffew2a5f7cb411173c	5720a15d22c09e2634c71c080e841102	6244ace8de4aa
13	85739	2152	2016-10-04 10:41:00	26587	30068	46046ad6ew222a292596d53922fe8	e972ab28655607763e0ca275161c90	99ee0ac9e6046c
14	96298	1672	2016-10-04 11:03:00	21617	51505	8060826995cde21e155085e060463	9e241a9965210e222859a94c033162	77530e977257e

#### 4.1.2.2. Fact Sales

```

INSERT INTO dbo.fact_sales (
product_key, seller_key, customer_key,
order_payment_key, date_time_key,
order_key, order_item_key,
order_id, order_item_id, product_id, seller_id, customer_id,
payment_type, price, freight_value, payment_value, order_purchase_date
)
SELECT
product_key, seller_key, customer_key,
p.order_payment_key, d.date_time_key,
o.order_key, oi.order_item_key,
oi.order_id, oi.order_item_id, oi.product_id, oi.seller_id, o.customer_id,
p.payment_type, oi.price, oi.freight_value, p.payment_value,
o.order_purchase_timestamp
FROM OlistDW.dbo.dim_order_items oi

```

```

JOIN OlistDW.dbo.dim_orders o ON oi.order_id = o.order_id
LEFT JOIN (
    SELECT
        order_payment_key, order_id,
        MAX(payment_type) AS payment_type,
        SUM(payment_value) AS payment_value
    FROM OlistDW.dbo.dim_order_payments
    GROUP BY order_id, order_payment_key
) p ON oi.order_id = p.order_id
JOIN OlistDW.dbo.dim_product dp ON dp.product_id = oi.product_id
JOIN OlistDW.dbo.dim_seller ds ON ds.seller_id = oi.seller_id
JOIN OlistDW.dbo.dim_customer dc ON dc.customer_id = o.customer_id
JOIN OlistDW.dbo.dim_date_time d
    ON d.date_time_key = FORMAT(o.order_purchase_timestamp, 'yyyy-MM-
dd HH:mm') + ':00';

```

fact_sales_key	product_key	seller_key	customer_key	order_payment_key	date_time_key	order_key	order_item_key	order_item_id	order_id	product_id	
1	1	1006	534	48152	2016-04-06 09:25:00	45576	68581	96c5337e09e989eb5c13d0470d7f2c0	1	eb08b5f59715e	
2	2	1006	30026	26064	2017-10-06 08:55:00	324	8954	24e5920553fb9b74f98e0917b092061	3	eb08b5f59715e	
3	3	1	1006	30026	2017-10-06 08:55:00	324	8952	24e5920553fb9b74f98e0917b092061	1	eb08b5f59715e	
4	4	1	1006	30026	2017-10-06 08:55:00	324	8953	24e5920553fb9b74f98e0917b092061	2	eb08b5f59715e	
5	5	2772	16289	82468	2018-05-22 11:59:00	53020	89564	a89d4e0b364d149e98e276ded1777f7	1	eb08b72c17b25c	
6	6	2772	16289	82470	2018-05-22 11:59:00	53020	89564	a89d4e0b364d149e98e276ded1777f7	1	eb08b72c17b25c	
7	7	2	2772	16289	82468	2018-05-22 11:59:00	53020	89565	a89d4e0b364d149e98e276ded1777f7	2	eb08b72c17b25c
8	8	2	2772	16289	82470	2018-05-22 11:59:00	53020	89565	a89d4e0b364d149e98e276ded1777f7	2	eb08b72c17b25c
9	9	2	2772	16289	82469	2018-05-22 11:59:00	53020	89564	a89d4e0b364d149e98e276ded1777f7	1	eb08b72c17b25c
10	10	2	2772	16289	82469	2018-05-22 11:59:00	53020	89565	a89d4e0b364d149e98e276ded1777f7	2	eb08b72c17b25c
11	11	3	2016	9200	7254	2017-12-26 13:27:00	4960	1609	0x20000000000000000000000000000000	2	eb09e59ad4289
12	12	3	2016	9200	7254	2017-12-26 13:27:00	4960	1609	0x20000000000000000000000000000000	3	eb09e59ad4289
13	13	4	1769	82908	38731	2018-07-23 19:04:00	64490	62438	9e2fbff9b51437c915ed549e475ewc2c	1	eb09e55c5c7019
14	14	5	2352	90838	70574	2017-02-23 14:53:00	56363	73549	b2475330370722208ec2ca573d475f	1	eb09e5fb5a305

#### 4.1.2.3. Fact Product Seller & Performance

```

insert into OlistDW.dbo.fact_product_seller_performance
(order_item_id, order_id, price, freight_value, review_score, product_key,
seller_key, review_key, datetime_key, review_create_datetime)
select
    o.order_item_id,
    o.order_id,
    o.price,
    o.freight_value,
    r.review_score,
    p.product_key,

```

```

s.seller_key,
r.review_key,
FORMAT(o.order_purchase_timestamp, 'yyyy-MM-dd HH:mm') + ':00'
AS datetime_key,
FORMAT(r.review_creation_date, 'yyyy-MM-dd HH:mm') + ':00' AS
review_create_datetime
from OlistDWStage.dbo.stg_order_items_mix o
join OlistDW.dbo.dim_product p on o.product_id = p.product_id
join OlistDW.dbo.dim_seller s on o.seller_id = s.seller_id
join OlistDW.dbo.dim_review r on o.order_id = r.order_id;

```

SQLQuery23.sql - (local).OlistDW (sa (71)) - Microsoft SQL Server Management Studio

File Edit View Project Tools Window Help

Quick Launch (Ctrl+Q)

Object Explorer

Connect to... New Query Execute

0035246a40f520710769010f752

SQLQuery23.sql - (local).OlistDW (sa (71)) - OlistDW (sa (70))

SELECT TOP (1000) [fact\_id]  
,[order\_item\_id]  
,[order\_id]  
,[product\_key]  
,[seller\_key]  
,[review\_key]  
,[datetime\_key]  
,[price]  
,[freight\_value]  
,[review\_score]  
,[review\_create\_datetime]  
FROM [OlistDW].[dbo].[fact\_product\_seller\_performance]

Result Messages

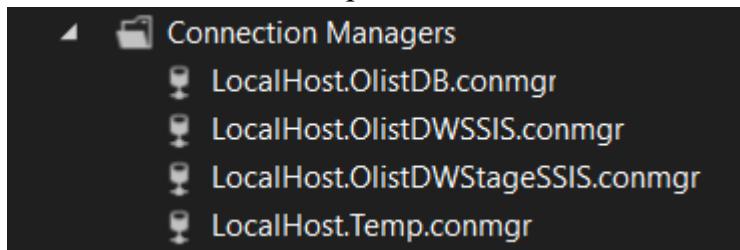
fact_id	order_item_id	order_id	product_key	seller_key	review_key	datetime_key	price	freight_value	review_score	review_create_datetime			
1	1	9e7c381ca5d262d599ff9947902d	21757	2670	57923	2017-11-26 23:13:00	239	17430003051758	4	2017-12-20 00:00:00			
2	2	1	9e7668ccbc7394795d9c059751c	18435	3010	185888	2016-07-08 19:40:00	10	18229999542263	5	2018-07-22 00:00:00		
3	3	1	9e766313c199b23d8d07c24437813d	13248	1404	19588	2016-07-25 19:02:00	46	1543999950385	1	2018-07-31 00:00:00		
4	4	1	9e775217e3994115a2cd09e18ad2e	26305	1155	23998	2017-09-04 14:28:00	1299	900002441406	38	50998321533	5	2017-09-12 00:00:00
5	5	1	9e762f65c0969e4a1996722d2e91	7652	3318	86772	2017-07-22 22:42:00	22	989999771182	14	1000003814897	5	2017-10-14 00:00:00
6	6	2	9e762f65c0969e4a1996722d2e91	7652	3318	86772	2017-07-22 22:42:00	22	989999771182	14	1000003814897	5	2017-10-14 00:00:00
7	7	1	9e740431e199c1161900d0e7d	21951	1816	33652	2017-04-12 21:35:00	29	899999771182	24	400000152879	4	2017-04-28 00:00:00
8	8	1	9e760717a242523442098174166	22478	3244	26559	2017-06-16 20:08:00	119	900001525879	21	499999771182	4	2017-06-29 00:00:00
9	9	1	9e760717a242523442098174166	1502	1180	44220	2017-09-29 16:30:00	99	699999694222	15	4499999992651	5	2017-10-12 00:00:00
10	10	1	9e760717a242523442098174166	5467	1189	80954	2017-06-22 22:23:00	29	989999771182	10	309999999201789	5	2018-08-08 00:00:00
11	11	1	9e760717a242523442098174166	16001	5749	44480	2017-06-17 10:00:00	50	900001525879	18	65999994941211	4	2017-12-06 00:00:00
12	12	1	9e760717a242523442098174166	111600	30930	1404	07522	2016-06-16 10:30:00	37	099999841211	5	2018-06-05 00:00:00	
13	13	1	9e760717a242523442098174166	21674	1872	NUL	2017-01-13 16:52:00	158	99000048164	33	5899999494242	NUL	NUL
14	14	1	9e760717a242523442098174166	8911	2974	30611	2017-06-17 14:18:00	38	899999396404	19	770000457837	5	2017-08-24 00:00:00
15	15	1	9e760717a242523442098174166	20987	2246	NUL	2016-03-07 06:51:00	149	4039999811884	NUL	NUL	NUL	
16	16	1	9e0308a2702d7597ca1bf0d12090	11598	1943	45294	2016-08-04 14:40:00	57	9900016784686	23	5599994659424	4	2018-08-16 00:00:00
17	17	2	9e0308a2702d7597ca1bf0d12090	31160	1943	45294	2016-08-04 14:40:00	57	9900016784686	23	5599994659424	4	2018-08-16 00:00:00
18	18	1	9e013e49b077.01577ec08e16207d7b5	4291	3277	11827	2017-04-07 07:00:00	189	180599994659424	5	2017-04-09 00:00:00		
19	19	1	9e014ecf83c3e5fe8ee780ba38e	16394	1950	83763	2018-02-15 10:03:00	38	900001525879	8	72000025702861	4	2018-02-21 00:00:00

Query executed successfully.

#### 4.2. Load bằng công cụ SSIS

#### 4.2.1. Kết Nối

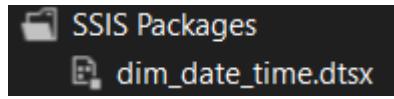
- Tiến hành thiết lập các kết nối:



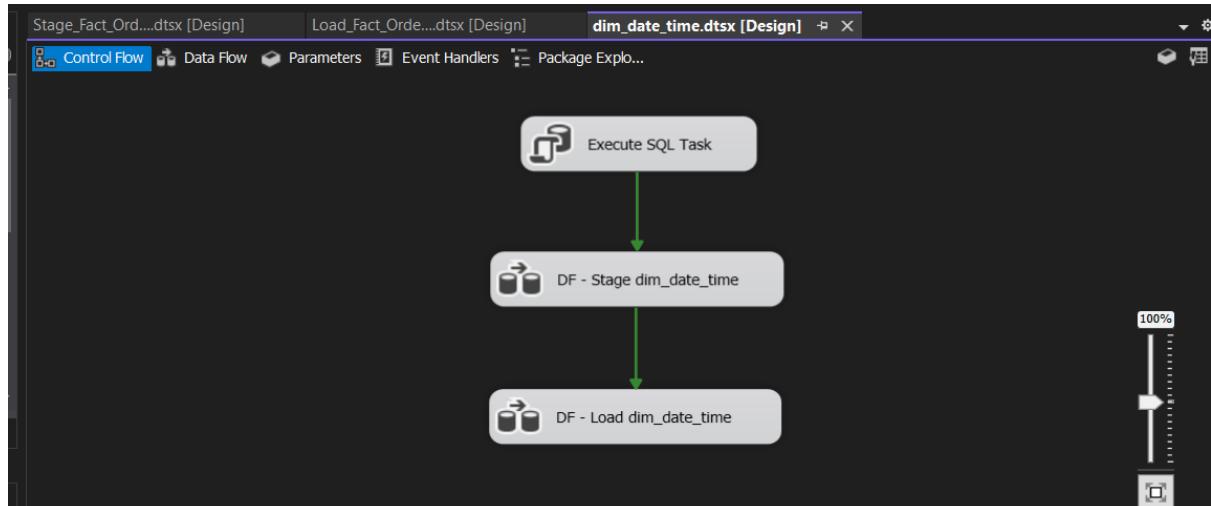
Với OlistDB chứa dữ liệu gốc, OlistDWStageSSIS dùng làm database cho staging, Temp chứa dữ liệu cho dim\_date\_time, và OlistDWSSIS là data warehouse sau khi load cuối cùng.

#### 4.2.2. Dim Date Time

Tạo package:

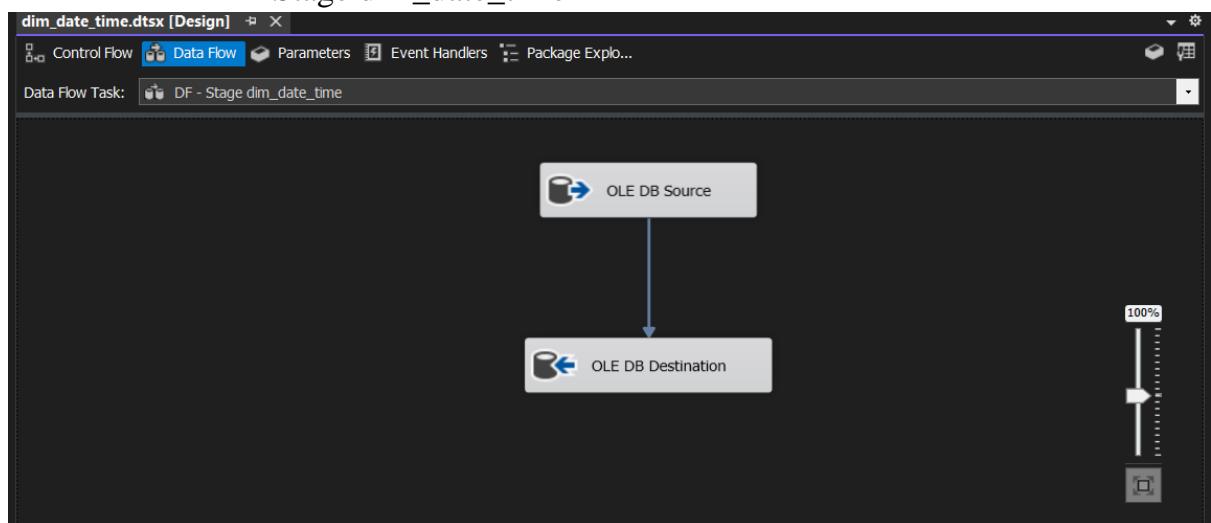


Control Flow:

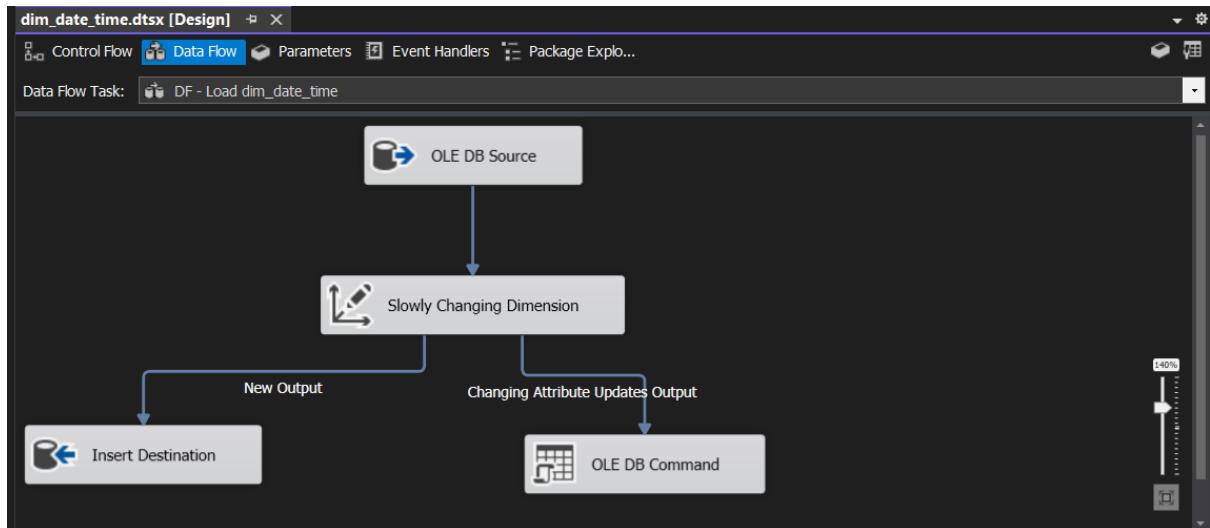


Các Data Flow:

- DF - Stage dim\_date\_time



- DF - Load dim\_date\_time



Kết quả cuối cùng:

The screenshot shows the SSIS environment with the Object Explorer on the left and the Results pane on the right. The Results pane displays the contents of the 'dim\_date\_time' table, which includes columns such as date\_time, date\_key, time\_key, hour, minute, second, day, month, year, weekday\_n, weekday\_n, month\_n, quarter, and is\_we.

date_time	date_key	time_key	hour	minute	second	day	month	year	weekday_n	weekday_n	month_n	quarter	is_we
2016-09-04 21:00:00	2100	21	0	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:01:00	2101	21	1	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:02:00	2102	21	2	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:03:00	2103	21	3	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:04:00	2104	21	4	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:05:00	2105	21	5	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:06:00	2106	21	6	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:07:00	2107	21	7	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:08:00	2108	21	8	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:09:00	2109	21	9	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:10:00	2110	21	10	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:11:00	2111	21	11	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:12:00	2112	21	12	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:13:00	2113	21	13	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:14:00	2114	21	14	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:15:00	2115	21	15	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:16:00	2116	21	16	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:17:00	2117	21	17	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:18:00	2118	21	18	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:19:00	2119	21	19	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:20:00	2120	21	20	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:21:00	2121	21	21	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:22:00	2122	21	22	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:23:00	2123	21	23	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:24:00	2124	21	24	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:25:00	2125	21	25	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:26:00	2126	21	26	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:27:00	2127	21	27	0	4	9	2016	Sunday	7	September	3	True	
2016-09-04 21:28:00	2128	21	28	0	4	9	2016	Sunday	7	September	3	True	

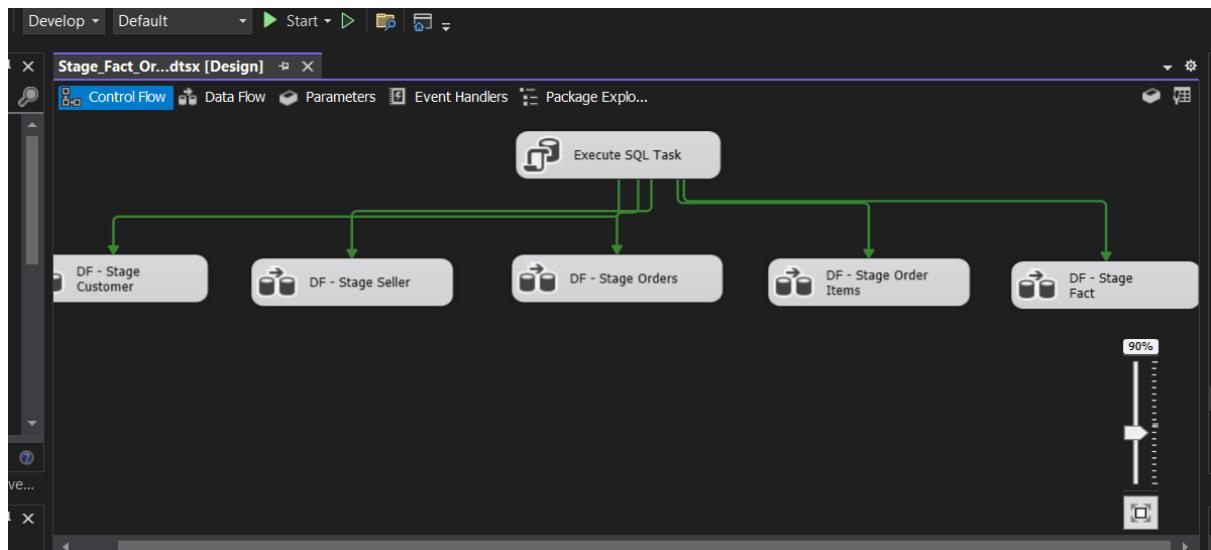
#### 4.2.3. Business Process: Order Fulfillment & Delivery Performance

##### 4.2.3.1. Tạo các Data Flow Stage cần thiết cho Order Fulfillment & Delivery Performance

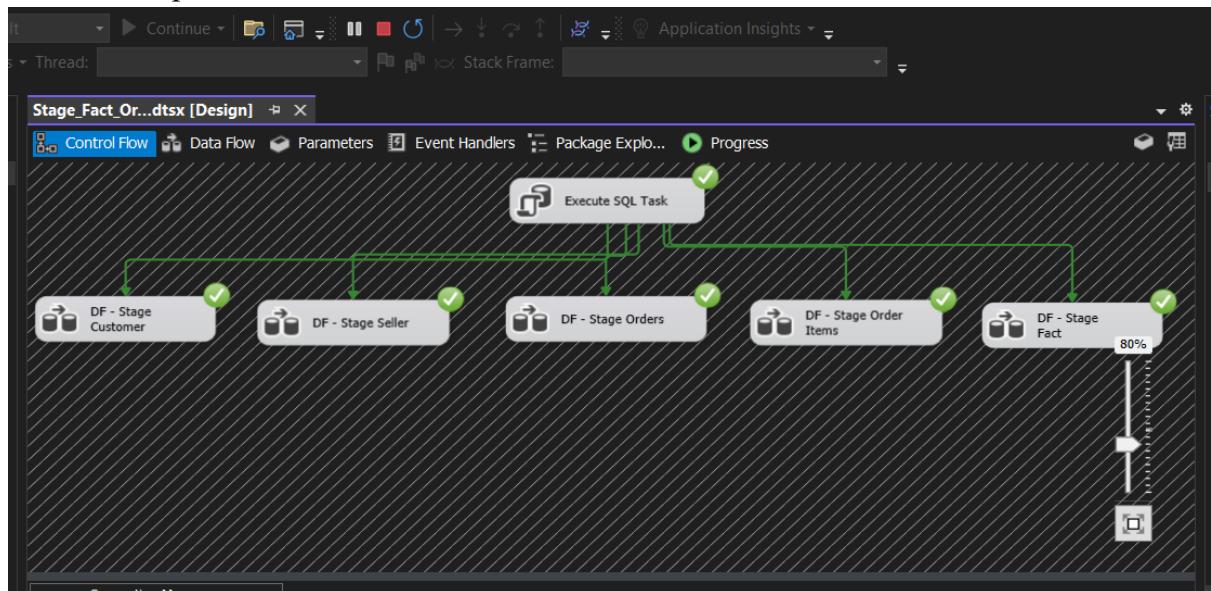
- Sử dụng lại dim\_date\_time đã staging và load trước đó.
- Tạo package:

Stage\_Fact\_Order\_Fulfillment.dtsx

- Tạo các Control Flow Staging cần thiết:



- Kết quả thực thi:

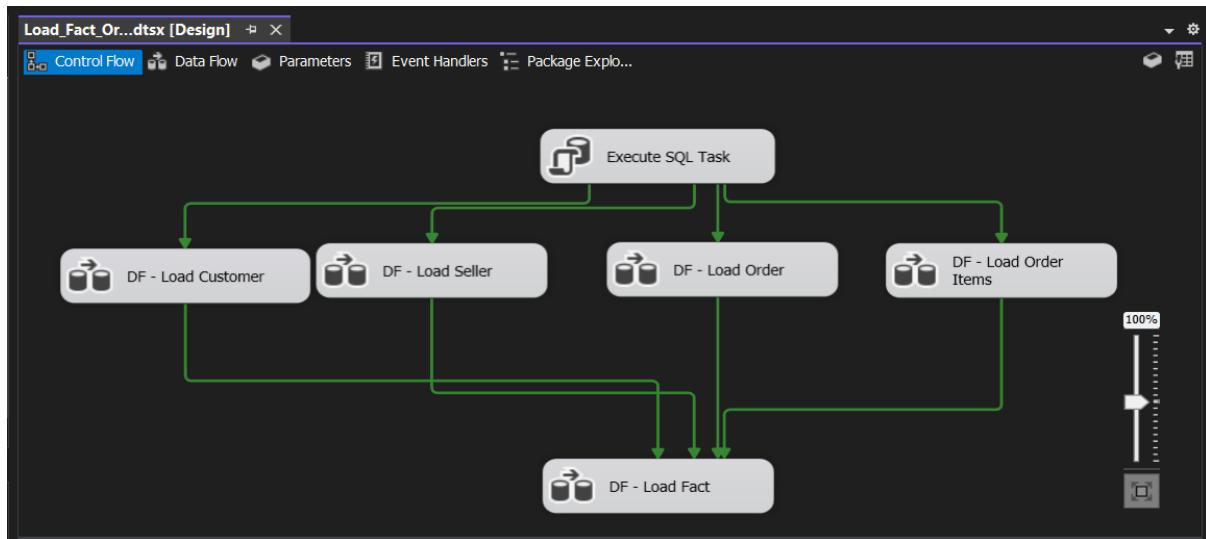


#### 4.2.3.2. Tạo các Data Flow load cần thiết cho Order Fulfillment & Delivery Performance

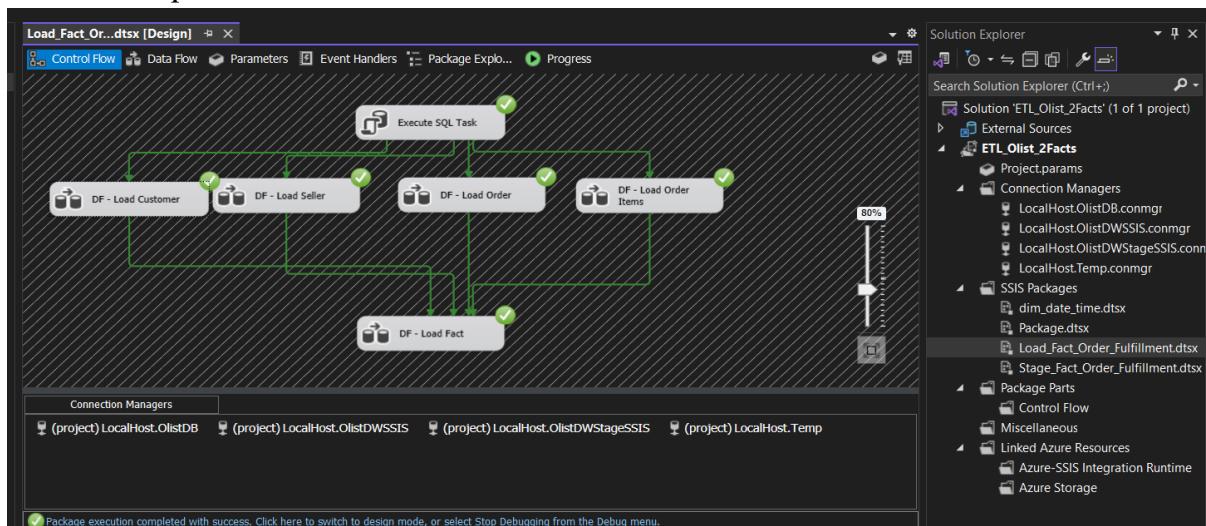
- Tạo package:

**Load\_Fact\_Order\_Fulfillment.dtsx**

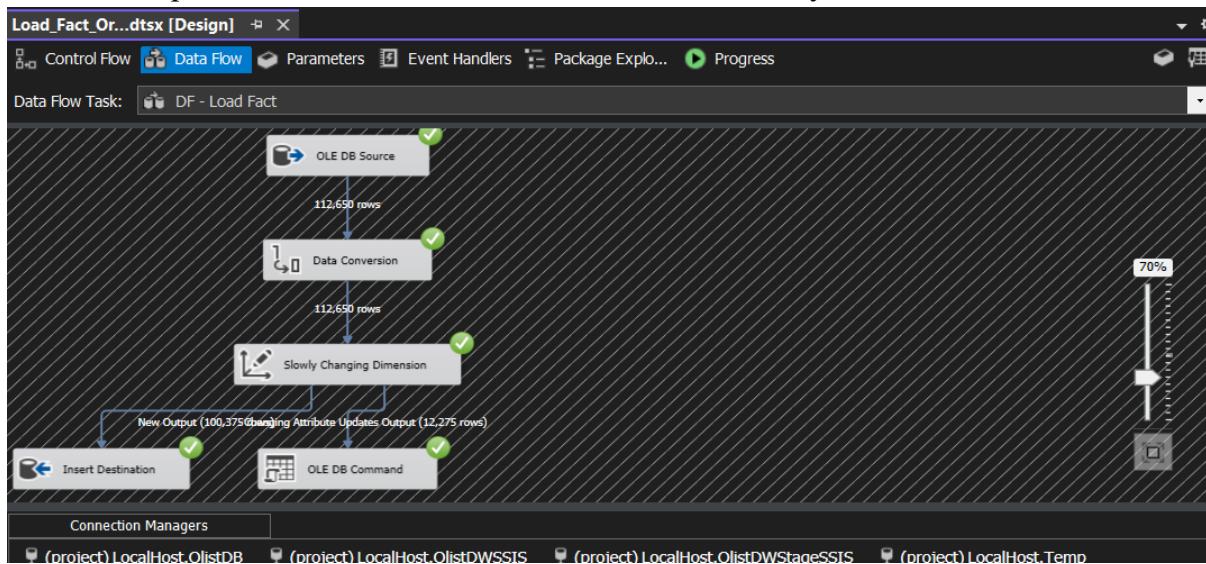
- Control Flow:

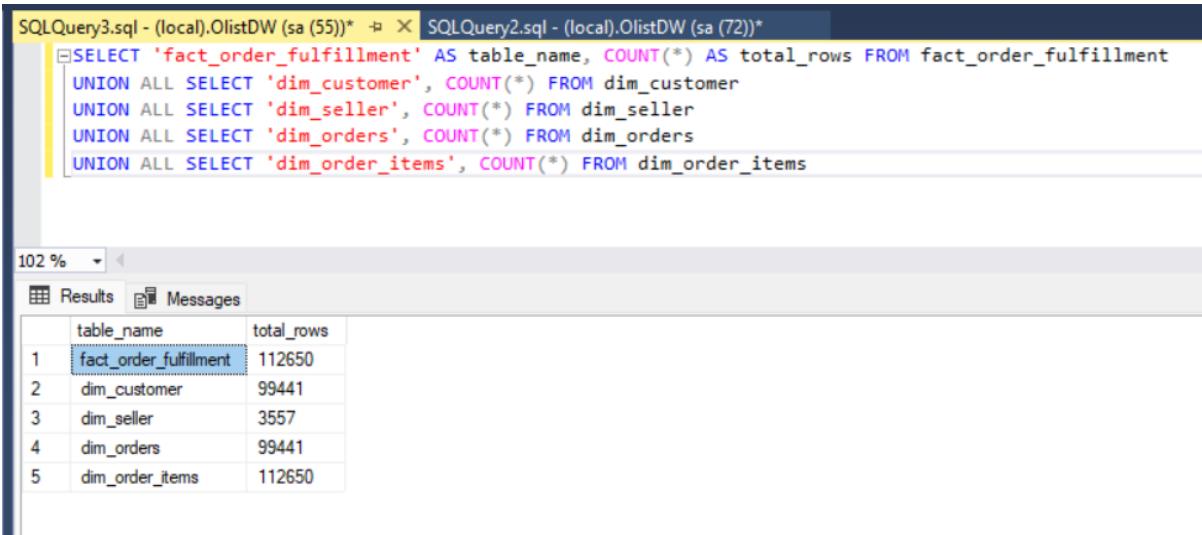


- Kết quả thực thi:



- Kết quả load cho Fact Order Fulfillment & Delivery Performance





The screenshot shows a SQL Server Management Studio window with two tabs: 'SQLQuery3.sql - (local).OlistDW (sa (55))' and 'SQLQuery2.sql - (local).OlistDW (sa (72))'. The 'SQLQuery3.sql' tab contains a query that counts rows across five tables: fact\_order\_fulfillment, dim\_customer, dim\_seller, dim\_orders, and dim\_order\_items. The results grid shows the following data:

	table_name	total_rows
1	fact_order_fulfillment	112650
2	dim_customer	99441
3	dim_seller	3557
4	dim_orders	99441
5	dim_order_items	112650

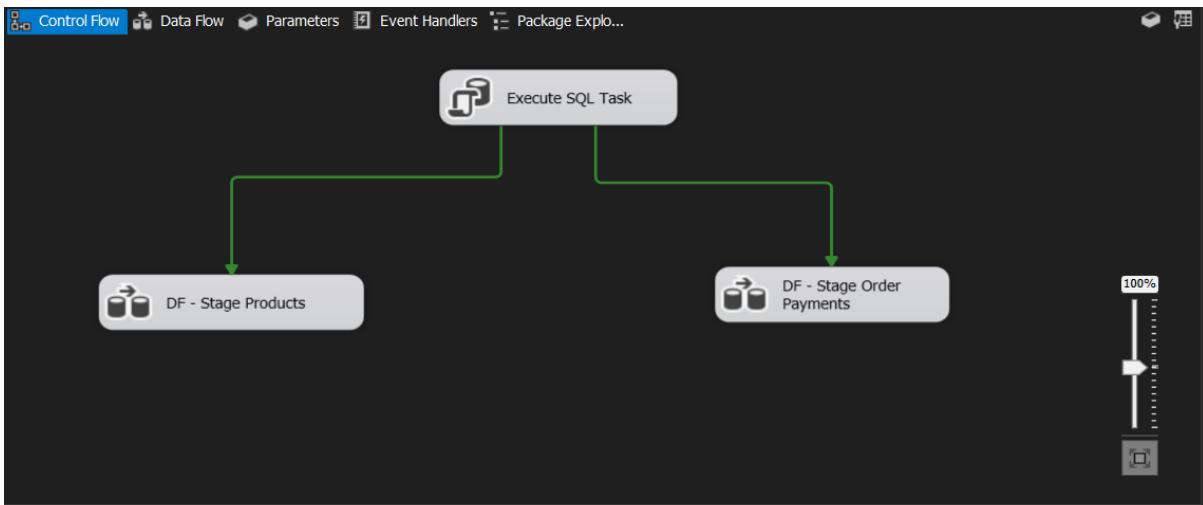
#### 4.2.4. Business Process : Sales & Revenue Analysis

##### 4.2.4.1. Tạo các Data Flow Stage cần thiết cho Sales & Revenue Analysis

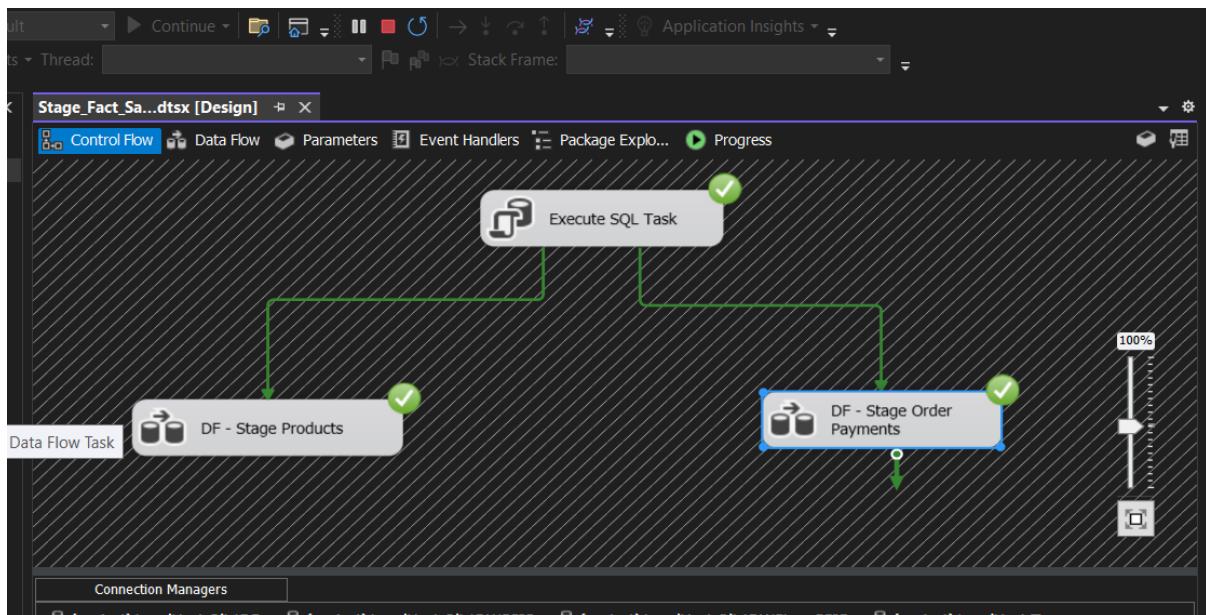
- Sử dụng lại các dim như customer, seller, orders, order items, date
- Tạo package:

 Stage\_Fact\_Sales.dtsx

- Control Flow:



- Kết quả thực thi:

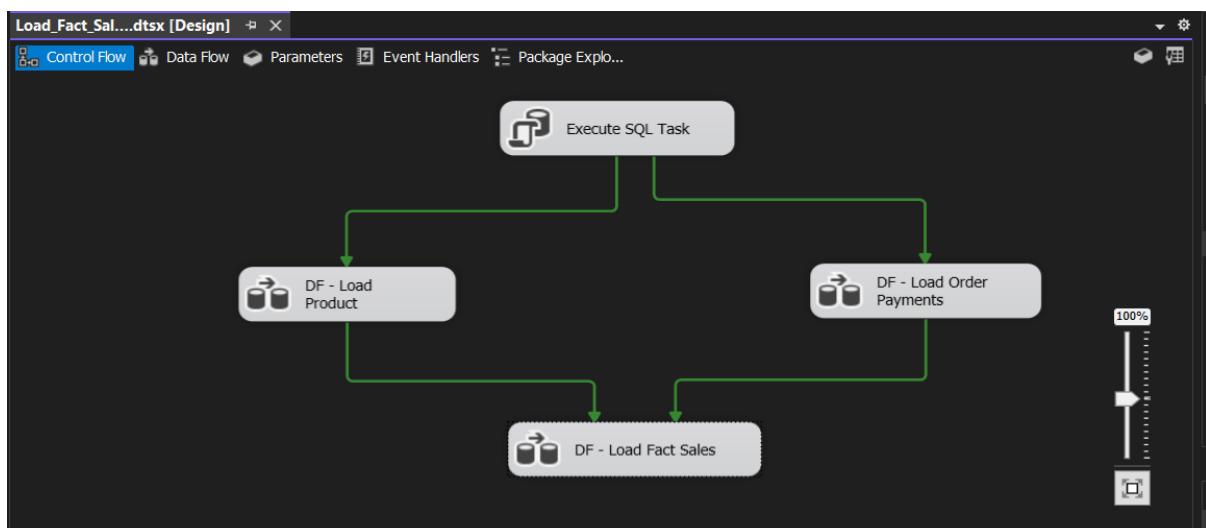


#### 4.2.4.2. Tạo các Data Flow load cần thiết cho Sales & Revenue Analysis

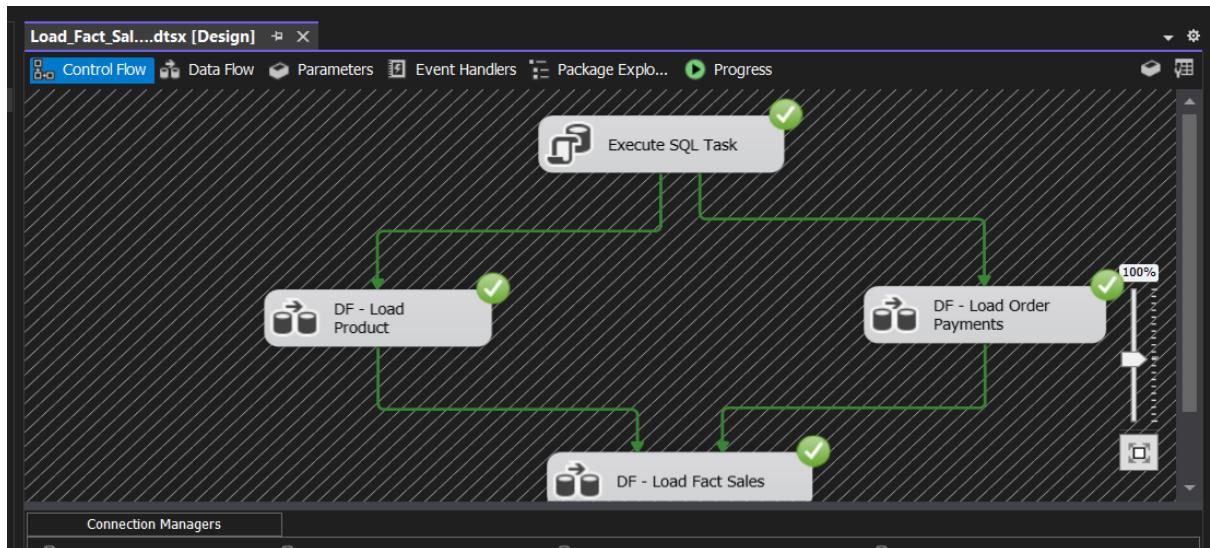
- Tạo package:

**Load\_Fact\_Sales.dtsx**

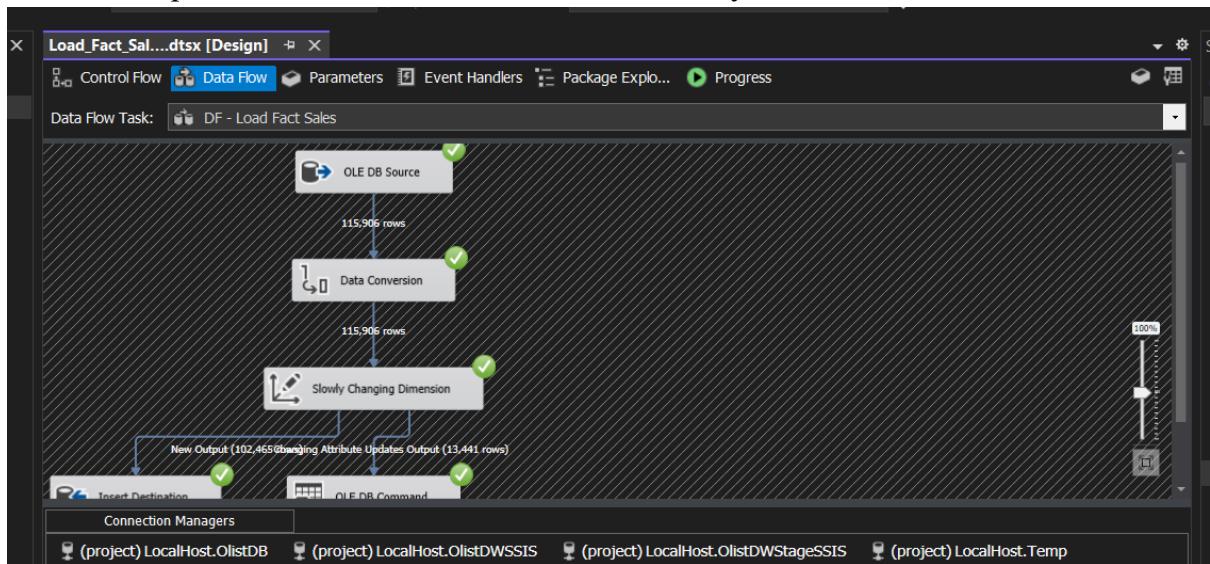
- Control Flow:



- Kết quả thực thi:



- Kết quả load cho Fact Sales & Revenue Analysis



SQLQuery3.sql - (local).OlistDW (sa (55))\*

SQLQuery2.sql - (local).OlistDW (sa (72))\*

```

SELECT 'fact_sales' AS table_name, COUNT(*) AS total_rows FROM fact_sales
UNION ALL SELECT 'dim_customer', COUNT(*) FROM dim_customer
UNION ALL SELECT 'dim_seller', COUNT(*) FROM dim_seller
UNION ALL SELECT 'dim_product', COUNT(*) FROM dim_product
UNION ALL SELECT 'dim_orders', COUNT(*) FROM dim_orders
UNION ALL SELECT 'dim_order_items', COUNT(*) FROM dim_order_items
UNION ALL SELECT 'dim_order_payments', COUNT(*) FROM dim_order_payments

```

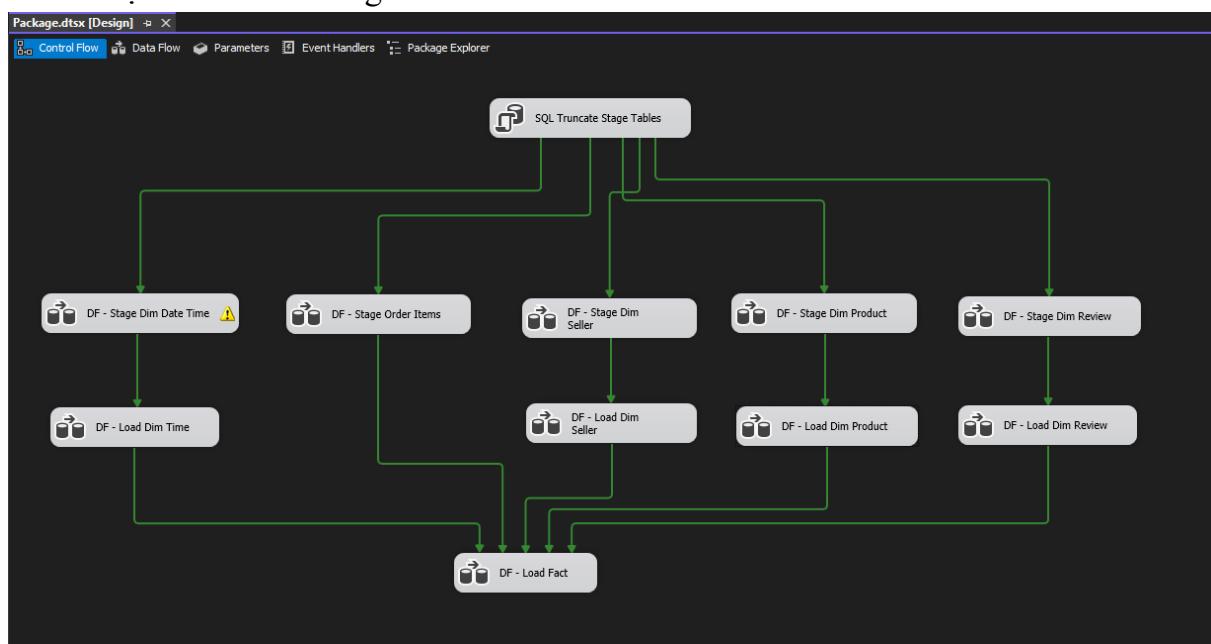
102 %

Results Messages

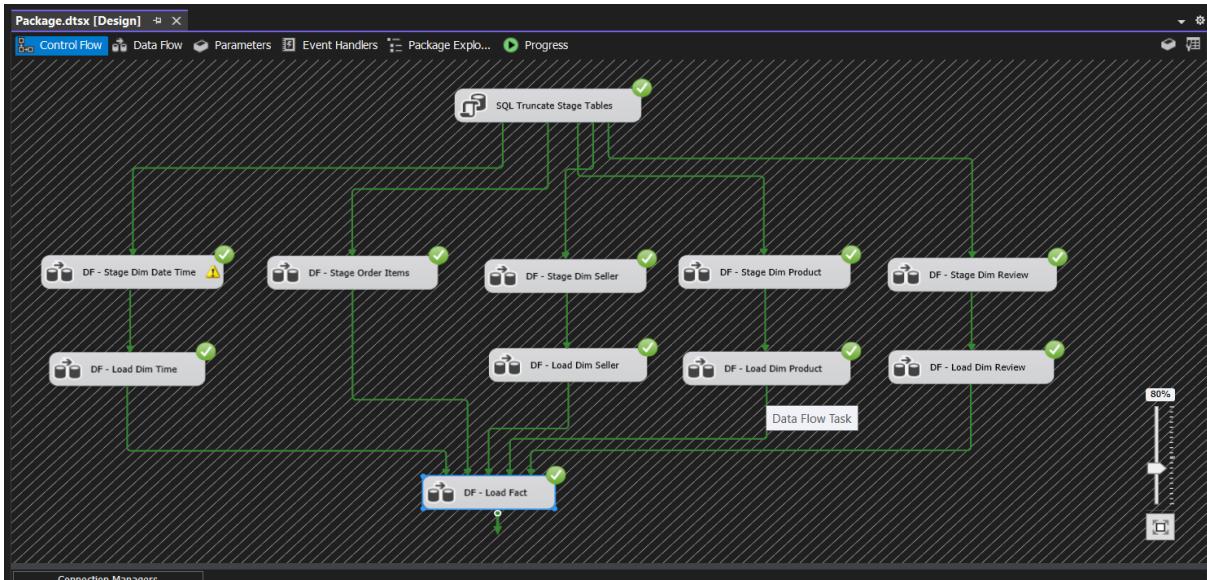
	table_name	total_rows
1	fact_sales	115906
2	dim_customer	99441
3	dim_seller	3557
4	dim_product	32342
5	dim_orders	99441
6	dim_order_items	112650
7	dim_order_payments	103886

#### 4.2.5. Business Process: Product Seller Performance

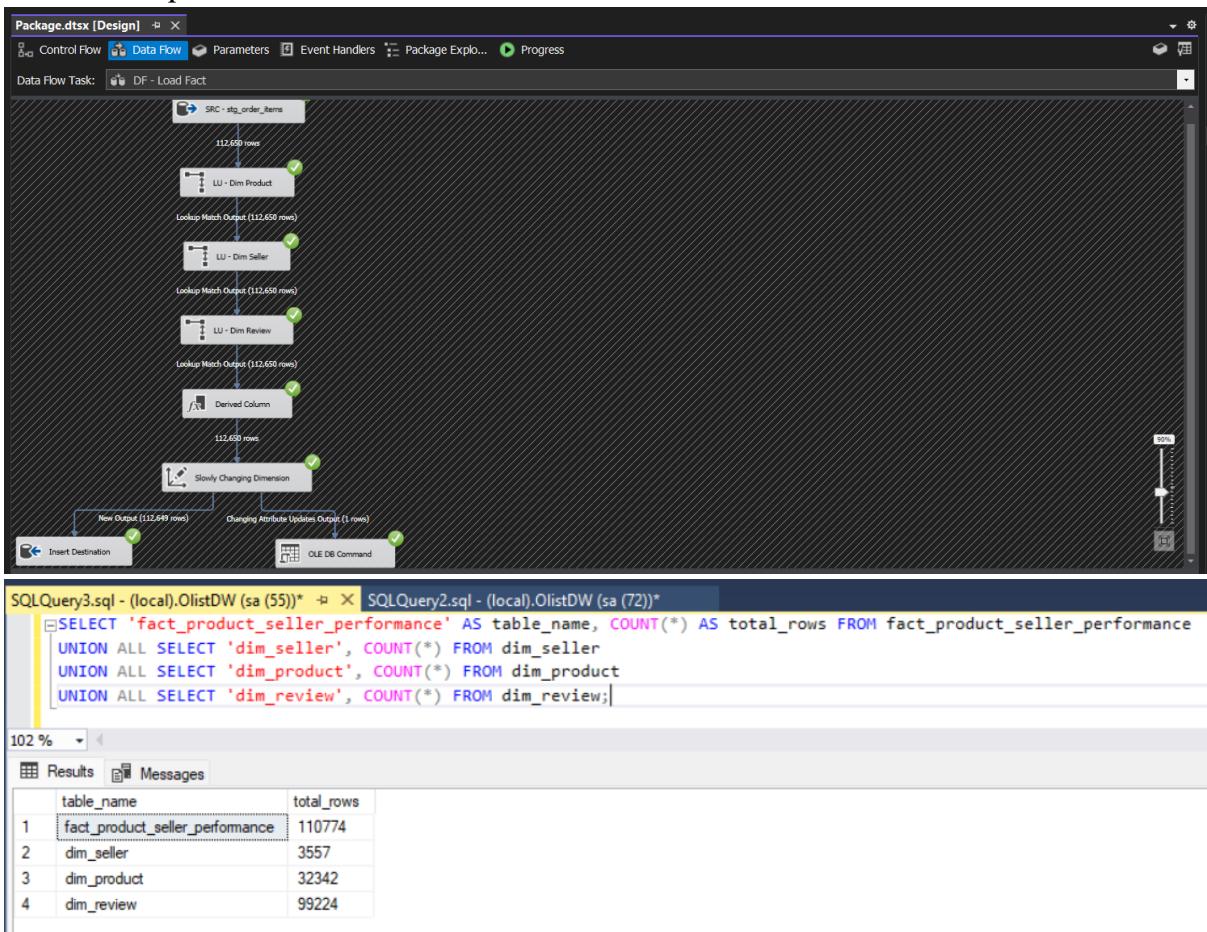
- Tạo Data Flow Stage và Load cần thiết cho Product Seller Performance



- Kết quả thực thi:



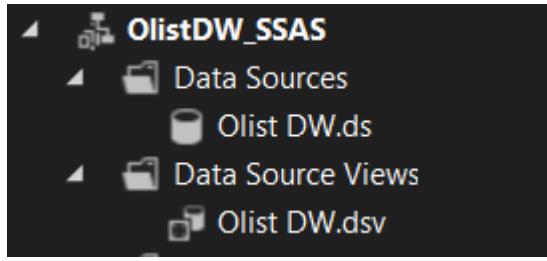
- Kết quả load cho Fact Product Seller Performance



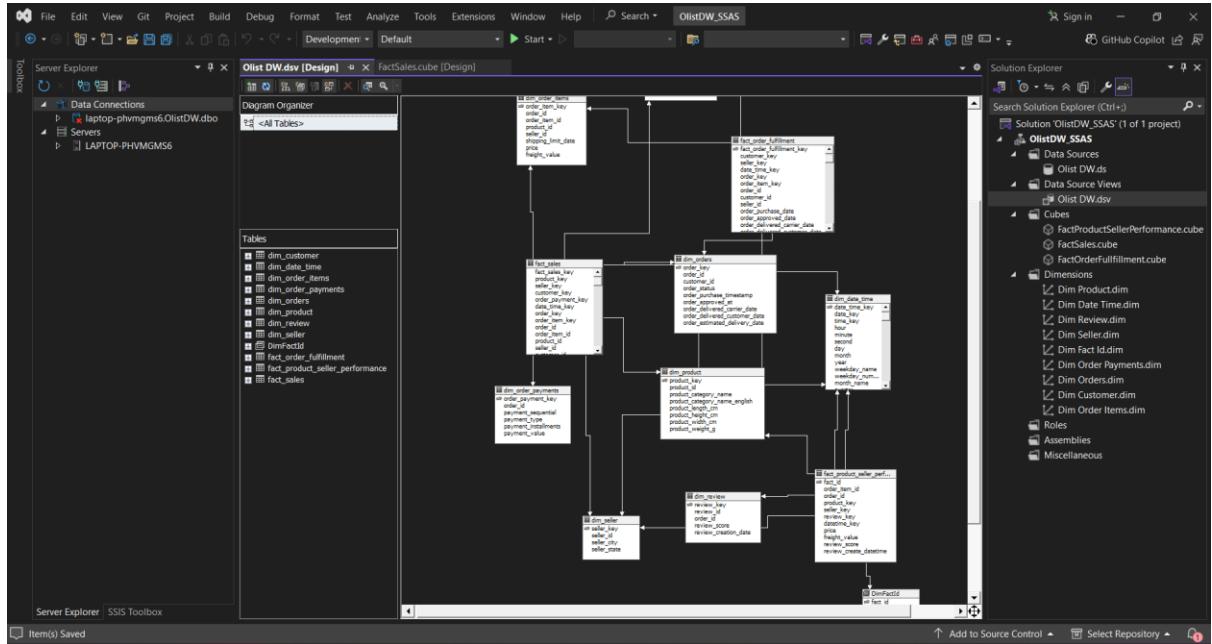
## 5. Phân tích Dữ liệu bằng SSAS

### 5.1. Tạo Data Sources và Data Sources View kết nối với OlistDW

- Tạo kết nối



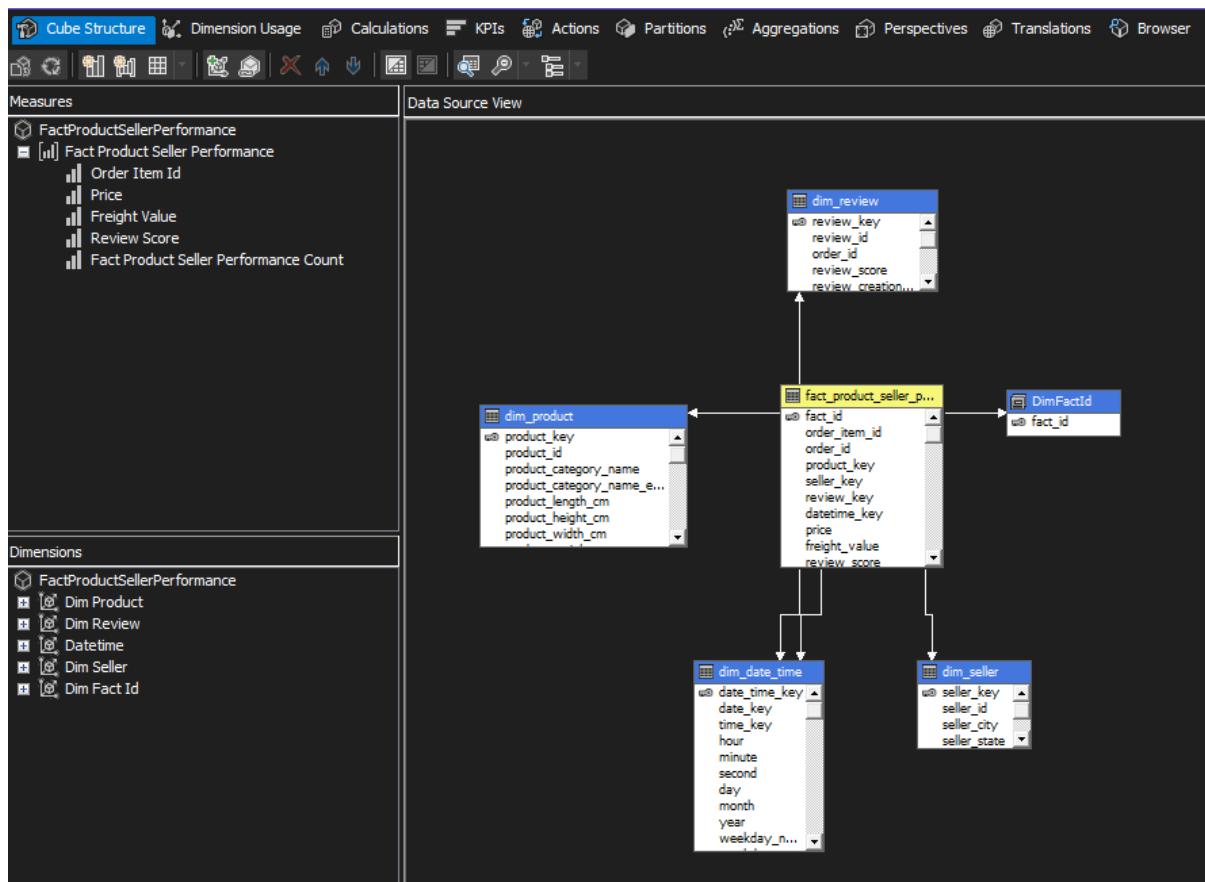
- Kết nối với OlistDW



## 5.2. Tạo Cubes

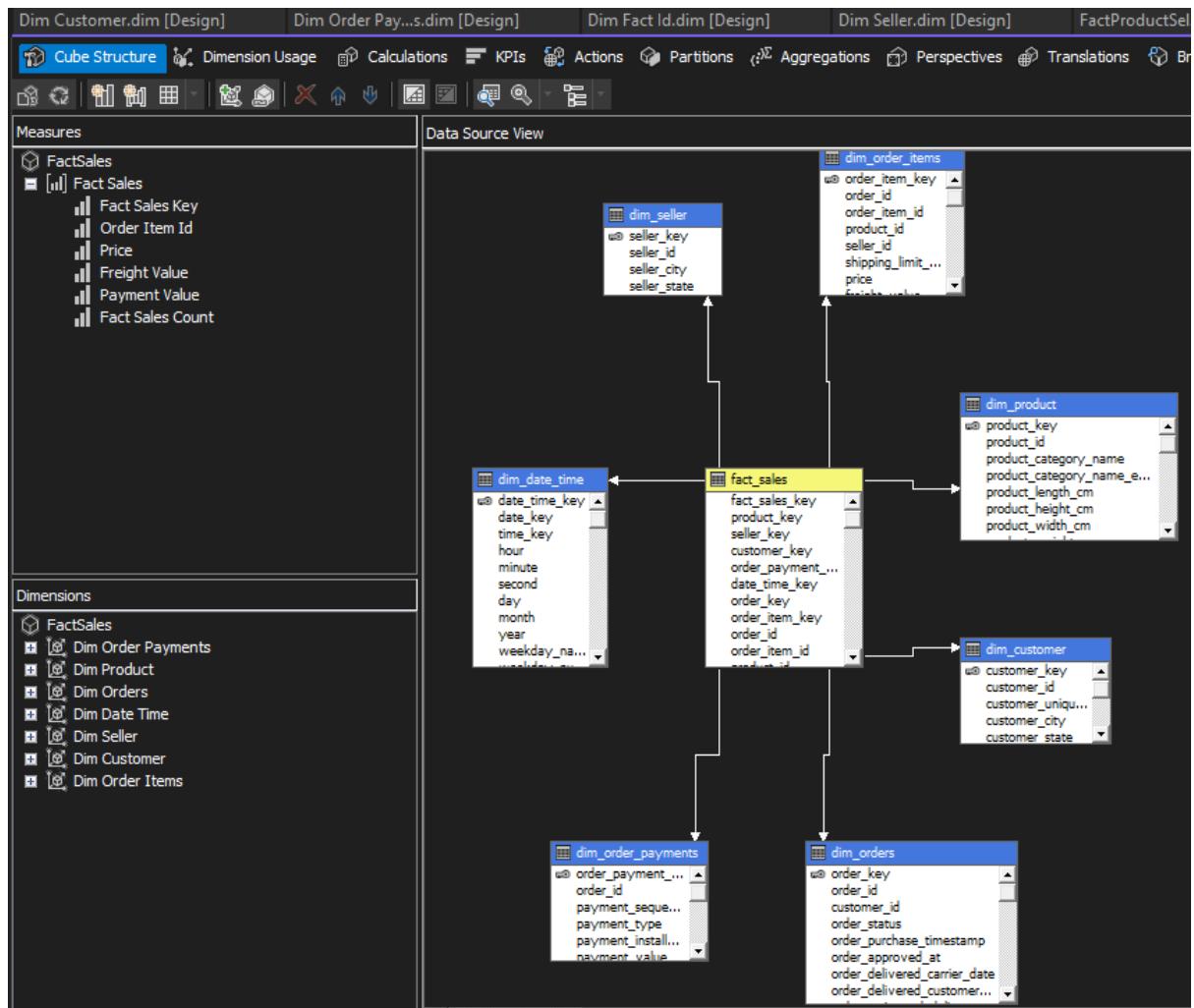
### 5.2.1. Tạo Cube FactProductSellerPerformance

- Chọn các Dimensions và Fact làm Measures đúng cho phân tích



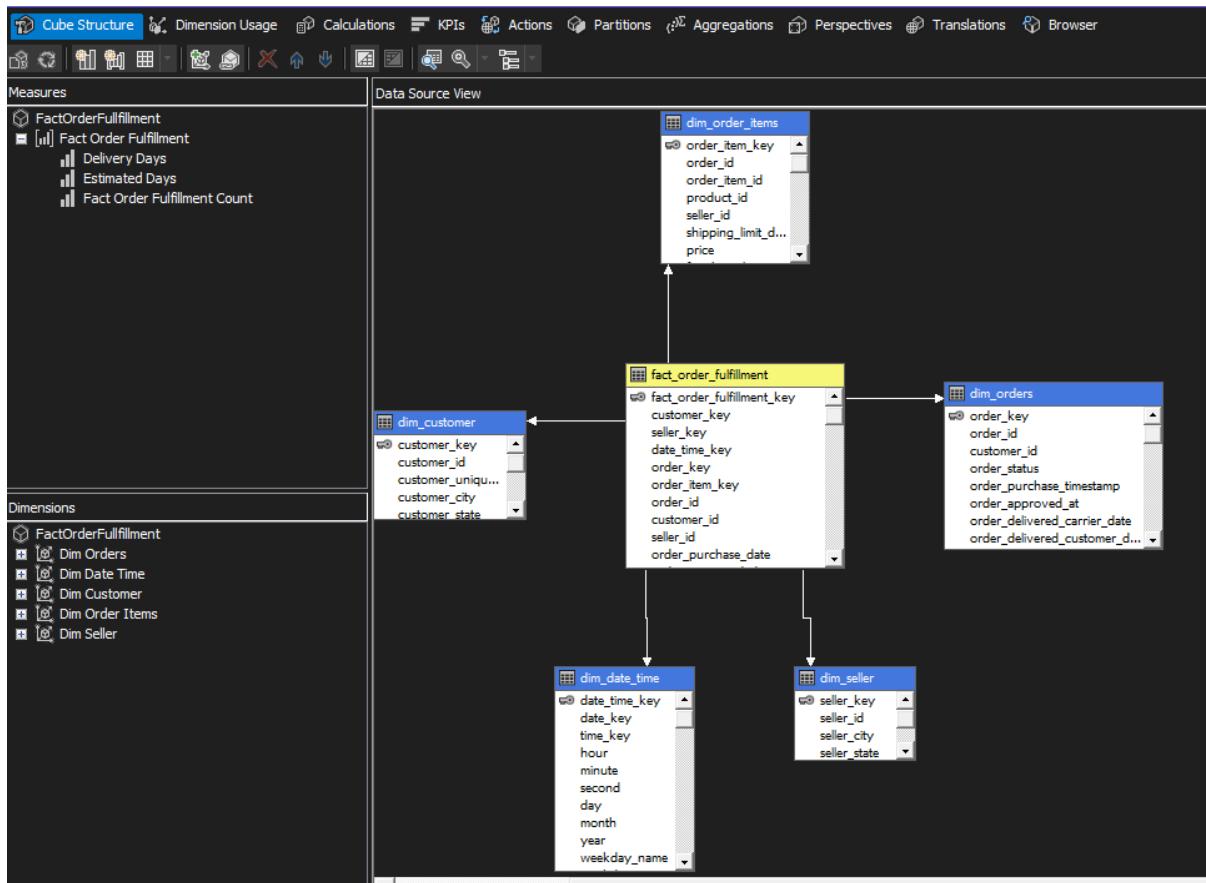
### 5.2.2. Tạo Cube FactSales

- Chọn các Dimensions và Fact làm Measures đúng cho phân tích



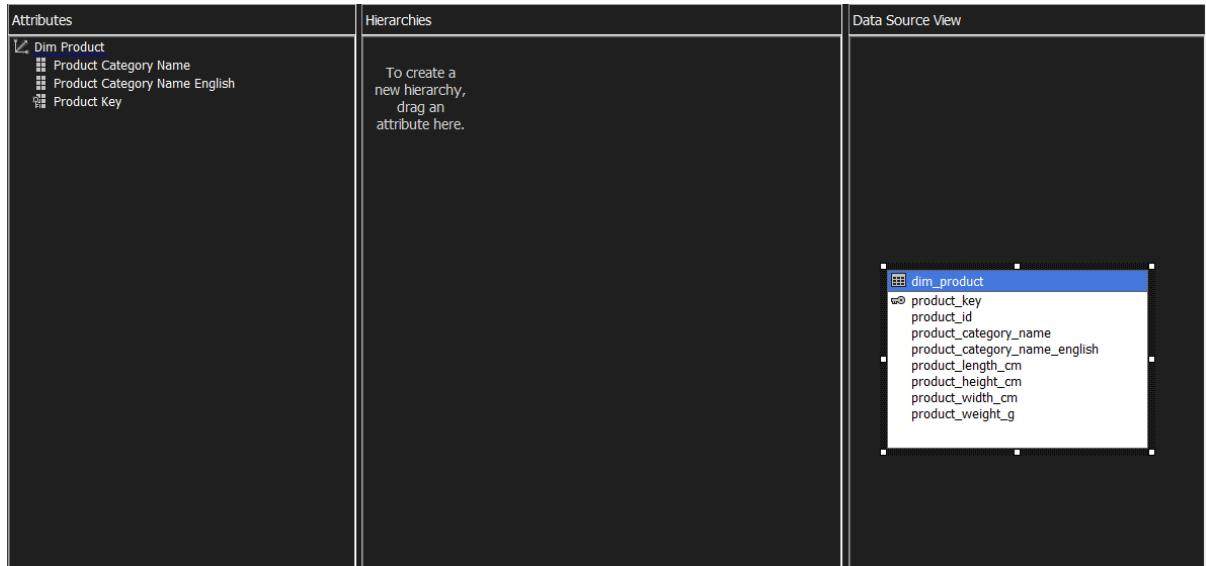
### 5.2.3. Tạo Cube FactOrderFullfillment

- Chọn các Dimensions và Fact làm Measures đúng cho phân tích



### 5.3. Thiết lập Measures và Dimensions

- Thêm thuộc tính hiển thị cho Dim Product



- Thêm thuộc tính và phân cấp cho Dim Date Time

The screenshot shows the Analysis Studio interface with three main panes:

- Attributes** pane: Shows the selected dimension "Dim Date Time" with its attributes: Date Time Key, Day, Month, and Year.
- Hierarchies** pane: A dialog box titled "Hierarchy" with the instruction "To create a new hierarchy, drag an attribute here." It lists the attributes: Year, Month, Day, and <new level>.
- Data Source View** pane: A list of attributes from the "dim\_date\_time" table: date\_time\_key, date\_key, time\_key, hour, minute, second, day, month, year, weekday\_name, weekday\_number, month\_name, and quarter.

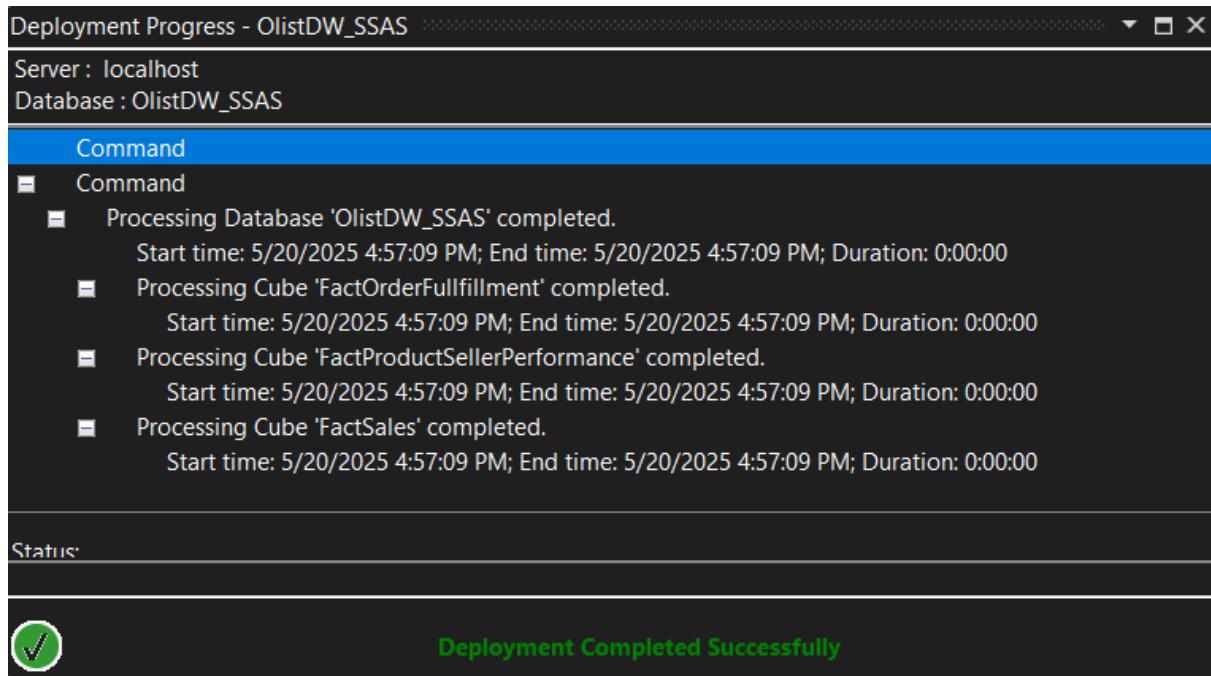
- Thêm thuộc tính hiển thị cho Dim Customer

The screenshot shows the Analysis Studio interface with three main panes:

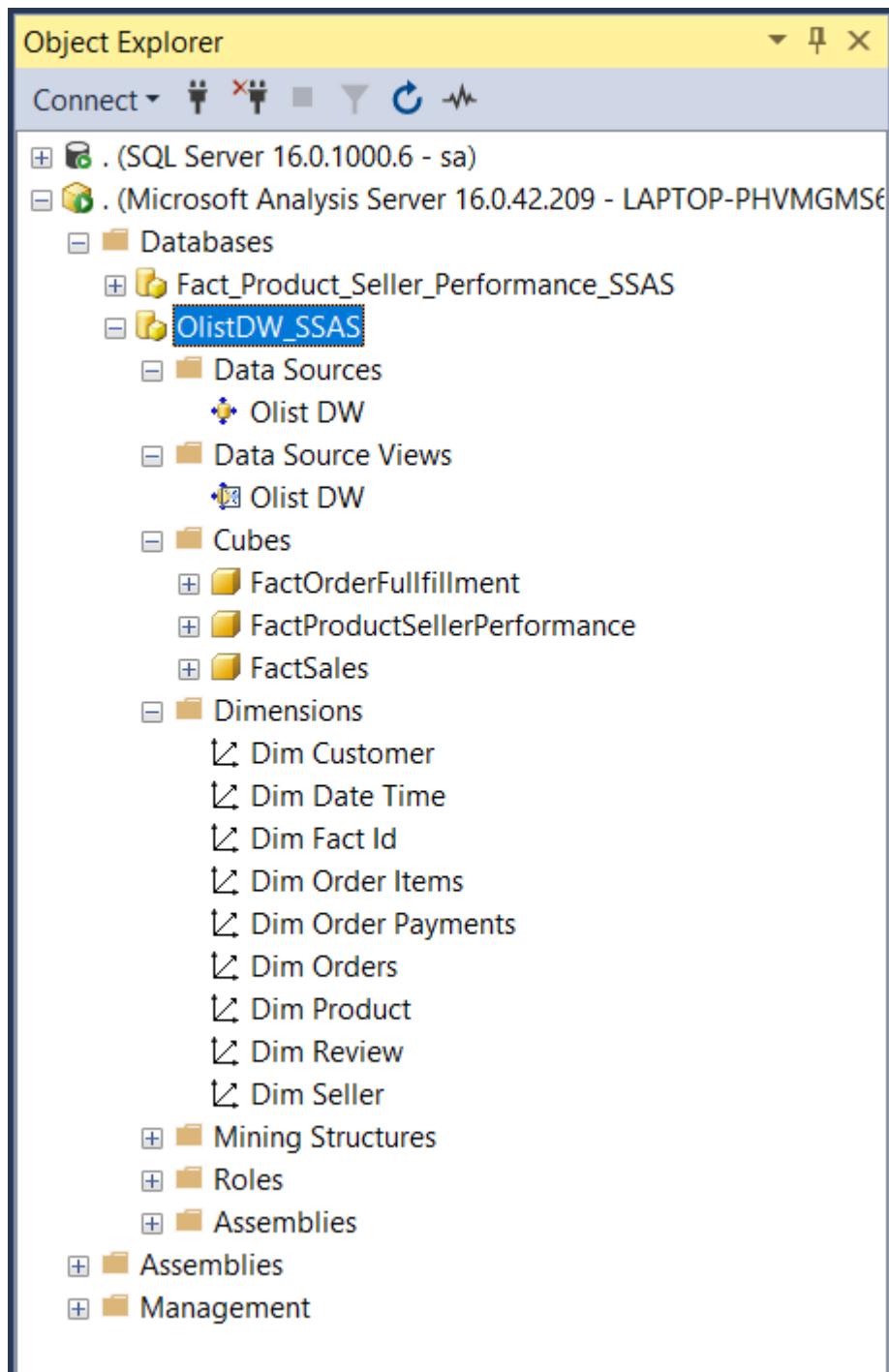
- Attributes** pane: Shows the selected dimension "Dim Customer" with its attributes: Customer City, Customer Key, and Customer State.
- Hierarchies** pane: A dialog box with the instruction "To create a new hierarchy, drag an attribute here." It lists the attributes: Customer City, Customer Key, and Customer State.
- Data Source View** pane: A list of attributes from the "dim\_customer" table: customer\_key, customer\_id, customer\_unique\_id, customer\_city, and customer\_state.

#### 5.4. Triển khai và Xử lý Cube

- Thực hiện Deploy Cube lên Analysis Services.



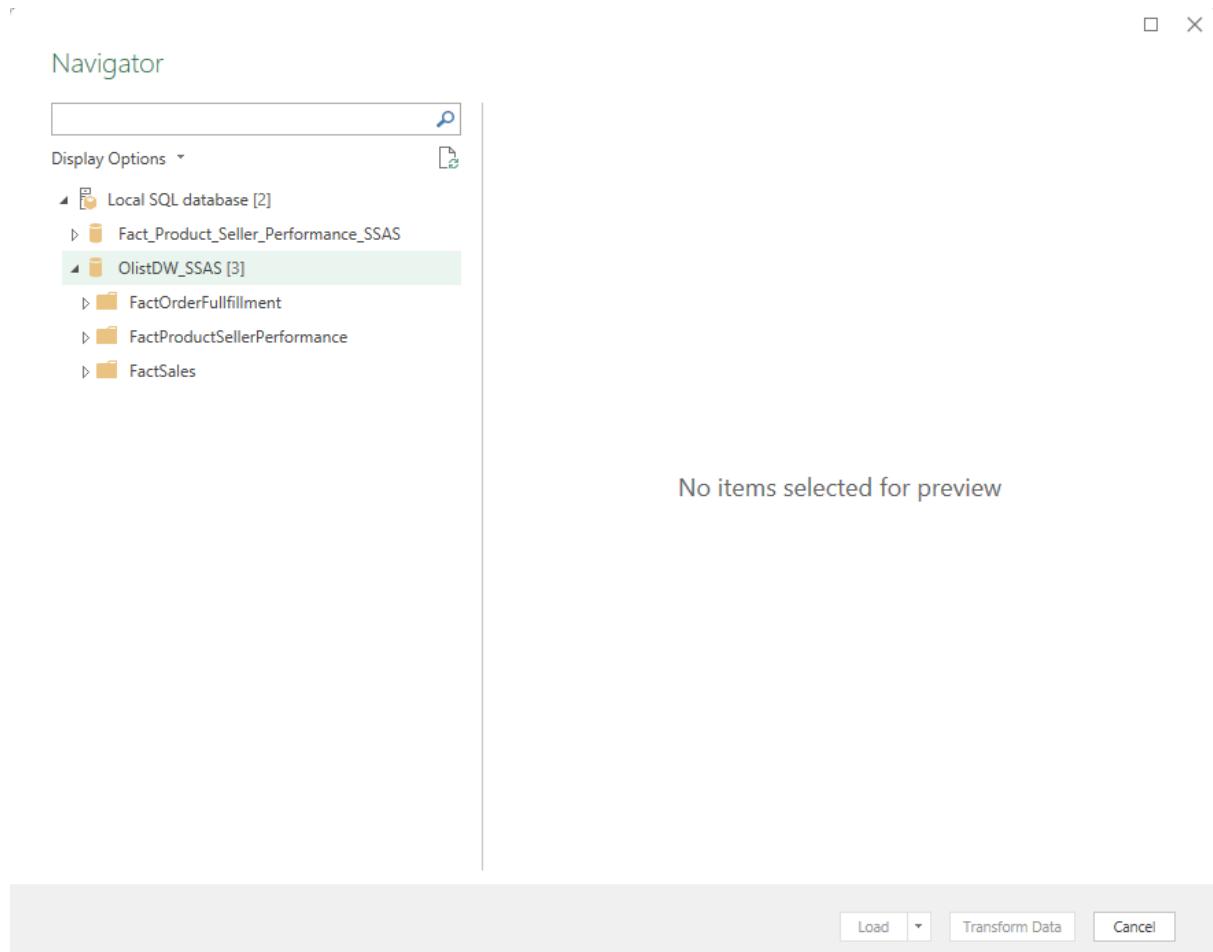
- Kiểm tra kết quả



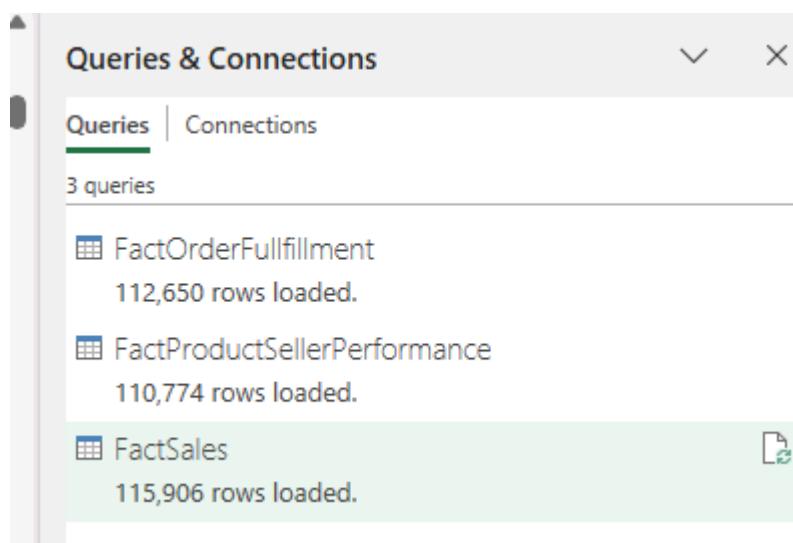
## 6. Trả lời câu hỏi bằng Pivot Table

### 6.1. Kết nối Excel với Cube SSAS

- Kết nối với Analysis Services sẽ có OlistDW\_SSAS



- Chọn các Fact để load



- Kết quả load thành công

The screenshot shows a Microsoft Excel spreadsheet titled 'C:\DW.xlsx - Excel'. The ribbon menu is visible with tabs like File, Home, Insert, Draw, Page Layout, Formulas, Data, Review, View, Help, Table Design, and Query. The Data tab is selected, showing various tools such as Get & Transform Data, Refresh All, Sort, Advanced Filter, Text to Columns, Flash Fill, Remove Duplicates, Data Validation, Consolidate, Data Model, What-If Analysis, Forecast, Group, Ungroup, Subtotal, Hide Detail, and Data Analysis. A 'Queries & Connections' pane on the right lists three queries: FactOrderFulfillment (112,630 rows loaded), FactProductSellerPerformance (110,774 rows loaded), and FactSales (115,906 rows loaded). The main table area contains data from two dimensions: 'Dim Customer.Customer City' and 'Dim Date Time'. The columns include Customer ID, Customer Name, State, and various date/time dimensions. The data spans from 2017-03-02 to 2018-08-01.

## 6.2. Trả lời các câu hỏi phân tích

### 6.2.1. Business Process : Order Fulfillment & Delivery Performance

Câu hỏi 1: Thời gian giao hàng trung bình là bao lâu?

PivotTable Fields

Choose fields to add to report:

Search

- Dim Date Time.Day
- Dim Date Time.Month
- Dim Date Time.Year
- Dim Date Time.Year.1
- Dim Date Time.Month.1
- Dim Date Time.Day.1
- Dim Order Items.Order Item Key
- Dim Orders.Order Key
- Dim Seller.Seller Key
- Delivery Days
- Estimated Days
- Fact Order Fulfillment Count

More Tables...

Drag fields between areas below:

Filters	Columns

Rows	Values
	Thời gian giao hàng trung bình
	12.14059476

Defer Layout Update

Câu hỏi 2: Có sự khác biệt về thời gian giao hàng giữa các bang/khu vực không?

PivotTable Fields

Choose fields to add to report:

Search

Quốc Gia - City	Thời gian giao hàng trung bình
<b>AC</b>	<b>20.45652174</b>
brasileia	30
cruzeiro do sul	23.66666667
epitaciolandia	13
manoel urbano	12
porto acre	29
rio branco	20.21794872
senador guiomard	13.66666667
xapuri	24.5
<b>AL</b>	<b>23.51126126</b>
agua branca	37
anadia	14
arapiraca	25.02941176
atalaia	11
barra de santo antonio	20
barra de sao miguel	24
batalha	16
belem	12.66666667
boca da mata	19.5
cacimbinhas	0
cajueiro	27
campo alegre	38
canapi	39
coite do noia	12
colonia leopoldina	15.33333333
coruripe	22
delmiro gouveia	20.16666667
dois riachos	33
feliz deserto	21
girau do ponciano	14
ibateguara	15.5
igaci	38.5
igreja nova	28
inhapi	25
jequia da praia	17
junqueiro	15.5
lagoa da canoa	46
luziapolis	15

Drag fields between areas below:

Filters

Rows

Dim Customer.Custo...

Dim Customer.Custo...

Defer Layout Update

### 6.2.2. Business Process : Sales & Revenue Analysis

Câu hỏi 3: Doanh thu theo loại sản phẩm?

**PivotTable Fields**

Choose fields to add to report:

Search

Dim Product.Product Category Name English

Dim Product.Product Key

Dim Seller.Seller Key

Fact Sales Count

Fact Sales Key

Freight Value

Order Item Id

Payment Value

Drag fields between areas below:

Filters	Columns

Rows	Values
Dim Product.Product ...	Tổng doanh thu

**Tên doanh mục** **Tổng doanh thu**

Tên doanh mục	Tổng doanh thu
bed_bath_table	1966848.531
sports_leisure	1818581.319
furniture_decor	1615782.392
health_beauty	1516819.591
housewares	1257322.589
auto	1158419.529
computers_accessories	1149881.34
toys	828189.4191
watches_gifts	777665.66
telephony	643785.3612
fashion_bags_accessories	548994.2707
stationery	533334.3797
cool_stuff	510969.8397
baby	501165.9794
perfumery	478052.5196
pet_shop	464569.6203
garden_tools	436310.3994
electronics	368538.7297
home_appliances	292221.5906
construction_tools_construction	230028.3202
luggage_accessories	225401.2696
consoles_games	206799.7504
home_construction	188209.0201
books_general_interest	167933.0492
musical_instruments	156673.5505
small_appliances	156434.0001
office_furniture	149560.1179
air_conditioning	134364.7798
signaling_and_security	84321.0797
agro_industry_and_commerce	81812.4594
fixed_telephony	75034.56988
fashion_shoes	73118.76005
costruction_tools_garden	68669.81017
home_comfort	66758.42046
food_drink	66202.88038
	65567.85003

**Câu hỏi 4: Doanh thu theo thời gian?**

**PivotTable Fields**

Choose fields to add to report:

Search

Dim Date Time.Day

- Dim Date Time.Month
- Dim Date Time.Year
- Dim Date Time.Year.1
- Dim Date Time.Month.1
- Dim Date Time.Day.1
- Dim Order Items.Order Item Key
- Dim Order Payments.Order Payment Key
- Dim Orders.Order Key
- Dim Product.Product Category Name
- Dim Product.Product Category Name English
- Dim Product.Product Key
- Dim Seller.Seller Key

Drag fields between areas below:

Filters	Columns

Rows	Values
Dim Date Time.Year	Doanh thu
Dim Date Time.Month	

Defer Layout Update

**Thời gian**  **Doanh thu**

Year	Month	Value
2016		74185.80008
	10	73818.66009
	12	19.62000084
	9	347.519989
2017		9012543.547
	1	185733.92
	10	991931.4986
	11	1565721.399
	12	1022777.73
	2	333177.0892
	3	517484.6499
	4	493728.1594
	5	710249.1701
	6	592189.1202
	7	732200.3797
	8	862173.4807
	9	1005176.95
2018		10968603.65
	1	1374723.229
	2	1288277.798
	3	1451515.86
	4	1486213.402
	5	1501509.591
	6	1294651.031
	7	1345179.657
	8	1226366.622
	9	166.4600067
	<b>Grand Total</b>	<b>20055333</b>

**Câu hỏi 5: Chi phí vận chuyển chiếm bao nhiêu % giá trị đơn hàng?**

**PivotTable Fields**

Choose fields to add to report:

Search

Dim Seller.Seller Key  
 Fact Sales Count  
 Fact Sales Key  
 Freight Value  
 Order Item Id  
 Payment Value

Drag fields between areas below:

<b>Filters</b>	<b>Columns</b>
	$\sum$ Values
<b>Rows</b>	$\Sigma$ Values
	Tổng phí vận chuyển
	Tổng đơn hàng

Tổng phí vận chuyển	Tổng đơn hàng	Tỉ lệ
2326818.323	20055333	11.602%

**Câu hỏi 6: Seller nào có doanh thu cao nhất?**

PivotTable Fields

Choose fields to add to report:

Search

Mã Seller Key

Tổng doanh thu

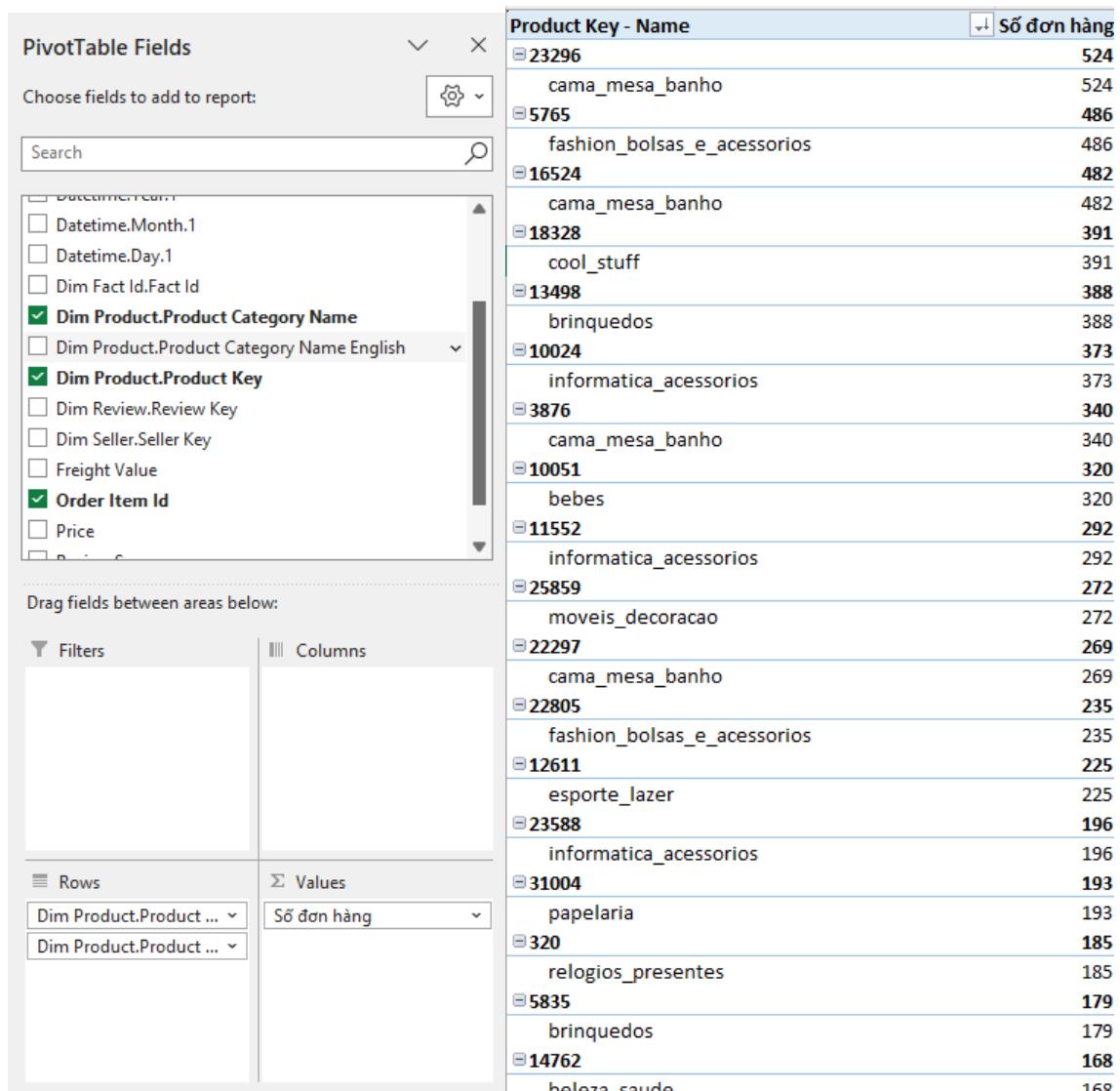
1748 507166.9073  
 222 308222.0398  
 1005 301245.2698  
 424 290253.4201  
 1155 284903.0805  
 3041 272219.3193  
 977 264166.1209  
 2094 236322.3005  
 3474 206513.2299  
 1784 185134.2097  
 1404 179657.749  
 1716 169030.8  
 505 157373.9199  
 2246 154356.9102  
 2488 145319.0401  
 1359 145267.9499  
 2854 141309.5801  
 962 137611.23  
 1307 132974.4205  
 835 131982.15  
 1756 129169.9798  
 2858 107125.3899  
 1873 94322.4899  
 3441 88248.62947  
 3298 86381.31959  
 3092 85852.25027  
 2862 84993.28008  
 1378 83496.82942  
 1672 83399.38963

Drag fields between areas below:

Filters	Columns
Rows	Values
Dim Seller.Seller Key	Tổng doanh thu

### 6.2.3. Business Process: Product Seller Performance

Câu hỏi 7: Sản phẩm nào bán chạy nhất?



Câu hỏi 8: Sản phẩm nào bị đánh giá thấp?

PivotTable Fields

Choose fields to add to report:

Search

Datetime.Day.1  
 Dim Fact Id.Fact Id  
 Dim Product.Product Category Name  
 Dim Product.Product Category Name English  
 Dim Product.Product Key  
 Dim Review.Review Key  
 Dim Seller.Seller Key  
 Freight Value  
 Order Item Id  
 Price  
 Review Score  
[More Tables...](#)

Drag fields between areas below:

Filters	Columns
<input type="checkbox"/> Rows <a href="#">Dim Product.Product ...</a> <a href="#">Dim Product.Product ...</a>	<input type="checkbox"/> Values <b>Đánh giá trung bình</b>

Product Key - Name

Product Key	Đánh giá trung bình
11245	1
moveis_decoracao	1
1241	1
eletroportateis	1
12406	1
papelaria	1
12400	1
cool_stuff	1
11298	1
musica	1
10054	1
cama_mesa_banho	1
1130	1
construcao_ferramentas_seguranca	1
10078	1
informatica_acessorios	1
11306	1
automotivo	1
10099	1
construcao_ferramentas_construcao	1
11319	1
moveis_decoracao	1
10107	1
moveis_decoracao	1
11330	1
beleza_saude	1
10114	1
beleza_saude	1
11337	1
esporte_lazer	1
10126	1
cama_mesa_banho	1
1136	1
informatica_acessorios	1
1017	1
perfumaria	1

Câu hỏi 9: Seller nào có đánh giá thấp nhất?

PivotTable Fields

Choose fields to add to report:

Search

<input type="checkbox"/> Datetime.Day.1
<input type="checkbox"/> Dim Fact Id.Fact Id
<input type="checkbox"/> Dim Product.Product Category Name
<input type="checkbox"/> Dim Product.Product Category Name English
<input type="checkbox"/> Dim Product.Product Key
<input type="checkbox"/> Dim Review.Review Key
<input checked="" type="checkbox"/> Dim Seller.Seller Key
<input type="checkbox"/> Freight Value
<input type="checkbox"/> Order Item Id
<input type="checkbox"/> Price
<input checked="" type="checkbox"/> Review Score
<a href="#">More Tables...</a>

Drag fields between areas below:

Filters	Columns

Rows	Σ Values
Dim Seller.Seller Key	Đánh giá trung bình

Product Key - Name Đánh giá trung bình

Product Key	Đánh giá trung bình
1119	1
1104	1
1087	1
1259	1
1026	1
1230	1
1125	1
1071	1
1182	1
1069	1
1203	1
1023	1
1226	1
1408	1
1418	1
1386	1
1290	1
1535	1
14	1
1344	1
1287	1
1357	1
1427	1
1379	1
1531	1
144	1
1356	1
1440	1
1306	1
1513	1
1358	1
1712	1
1864	1
1710	1

Câu hỏi 10: Seller nào có nhiều đơn hàng nhất?

**PivotTable Fields**

Choose fields to add to report:

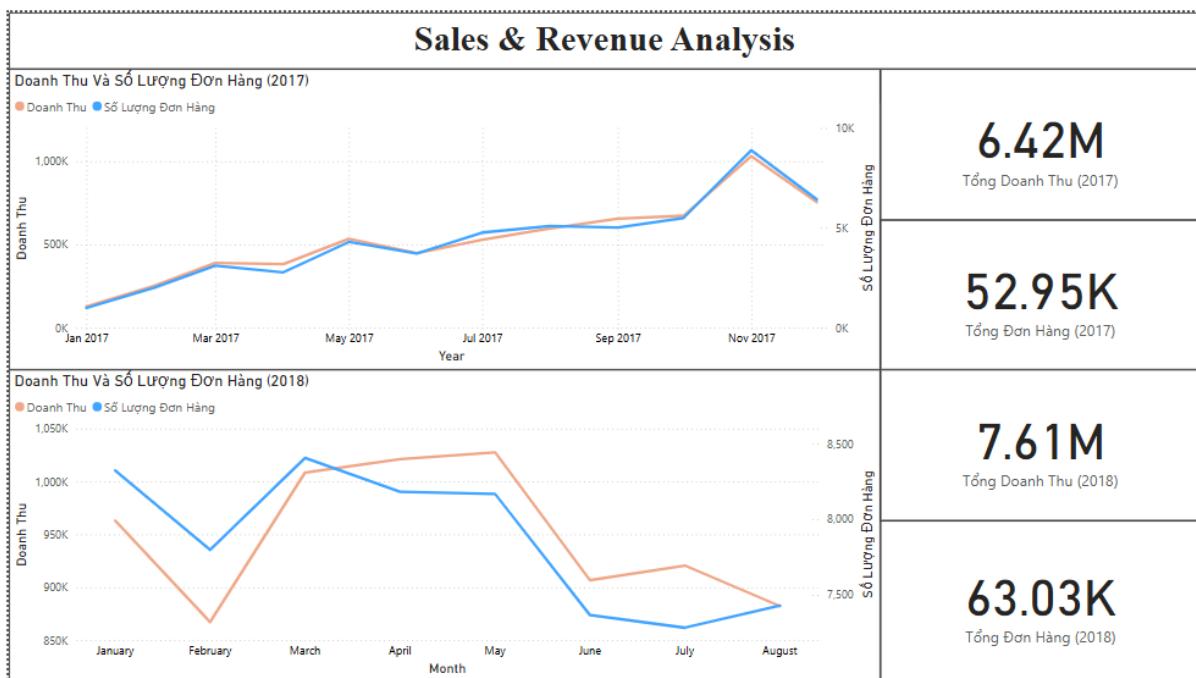
Seller Key	Số đơn hàng
1404	2020
1005	1984
424	1932
2854	1811
3041	1568
2094	1489
222	1431
1748	1367
3259	1197
1716	1166
977	1148
835	1136
1950	1014
2858	831
1359	790
2246	767
3240	747
3455	719
2930	649
767	621
3474	582
1756	574
3010	561
328	554
248	554
337	524
1876	522
Dim Seller.Seller Key	1167
	520
	962
	496
	1873
	442
	1078
	432
	3054
	429
	1818
	429
	453
	428

Drag fields between areas below:

Filters	Columns
Rows	Values
Dim Seller.Seller Key	Số đơn hàng

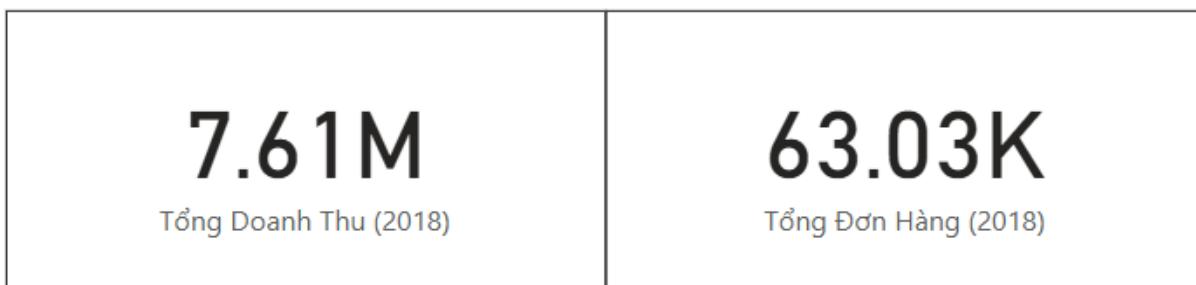
## 7. Trực quan hóa bằng Power BI

### 7.1. Thực hiện phân tích dữ liệu cho Business process: Sales & Revenue Analysis

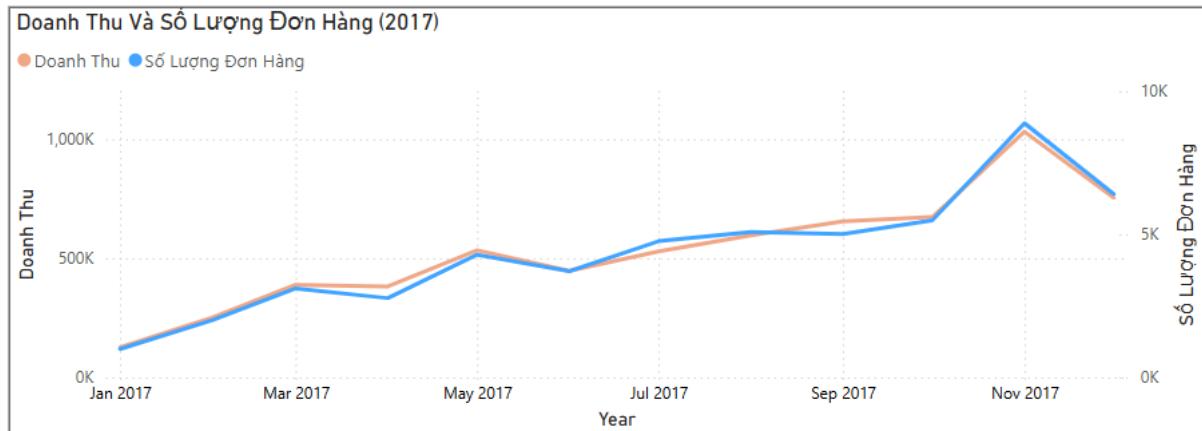


### 7.1.1. Câu hỏi: Cho biết số lượng đơn đặt hàng, doanh thu qua các năm 2017, 2018

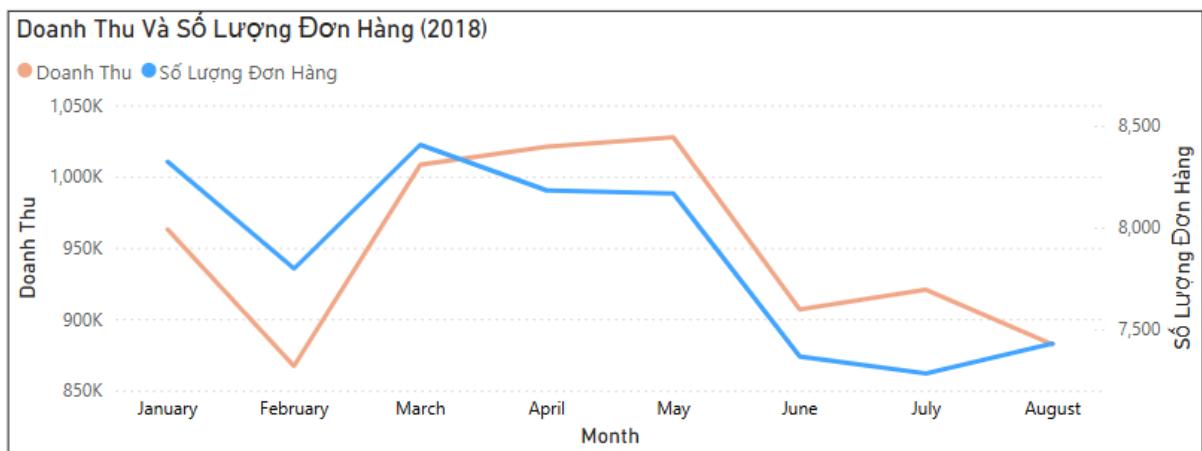
Từ DashBoard, Ta có thể xem được lợi nhuận và số đơn đặt hàng năm 2017 lần lượt là: 6,42 triệu đô doanh thu và 52,950 đơn đặt hàng và năm 2018 là: 7,61 triệu đô doanh thu và 63,030 đơn đặt hàng



### 7.1.2. Câu hỏi: Cho biết sự tăng giảm doanh thu qua các tháng (2017, 2018) dựa theo thị trường/khu vực

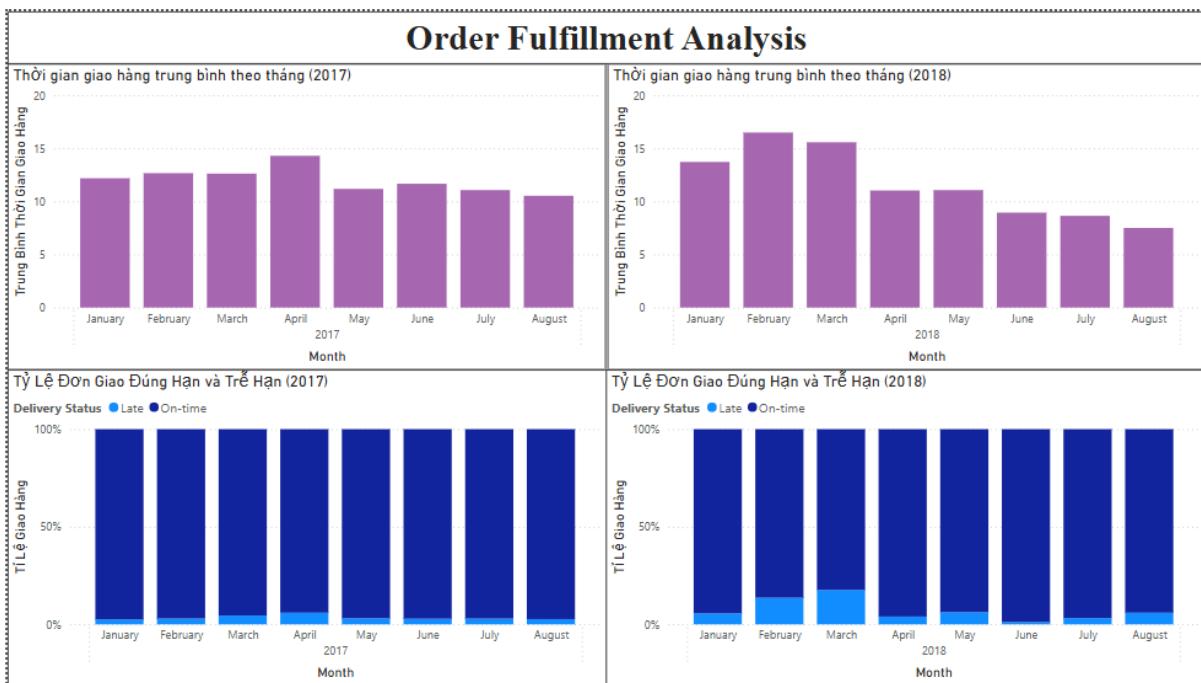


Biểu đồ 2017 cho thấy xu hướng tăng trưởng ổn định của cả doanh thu và số lượng đơn hàng trong năm, đặc biệt là tăng mạnh vào tháng 11, cho thấy cao điểm bán hàng cuối năm.

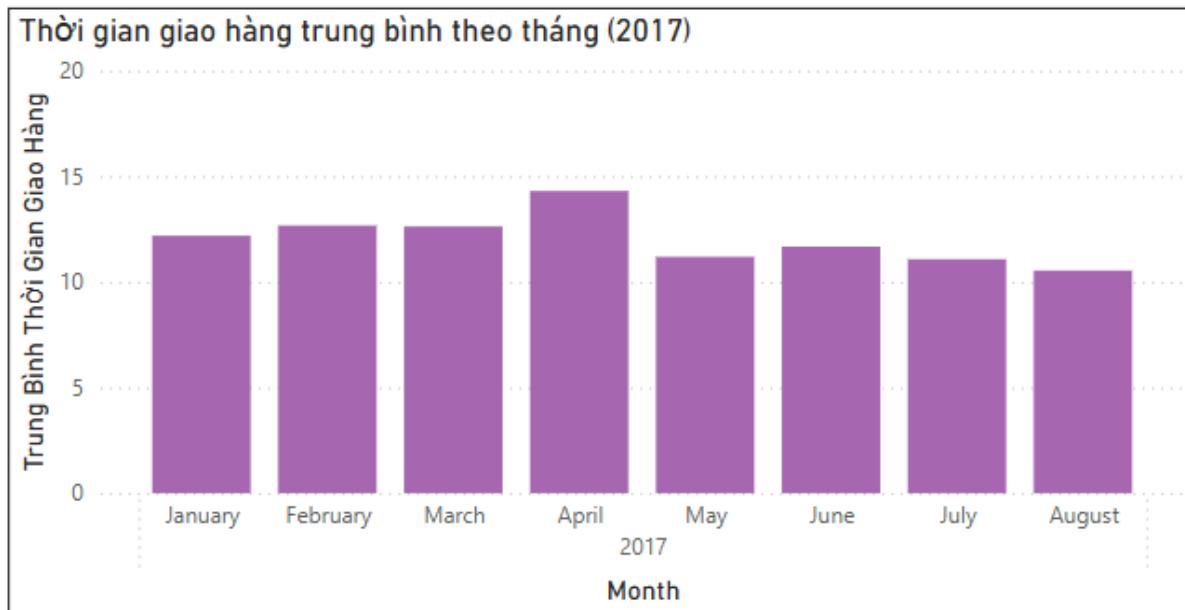


Biểu đồ 2018 lại thể hiện sự biến động rõ rệt giữa các tháng. Mặc dù doanh thu và số đơn hàng có xu hướng tăng cùng nhau trong quý I, nhưng từ tháng 6 trở đi đều sụt giảm, phản ánh sự chững lại trong hoạt động kinh doanh.

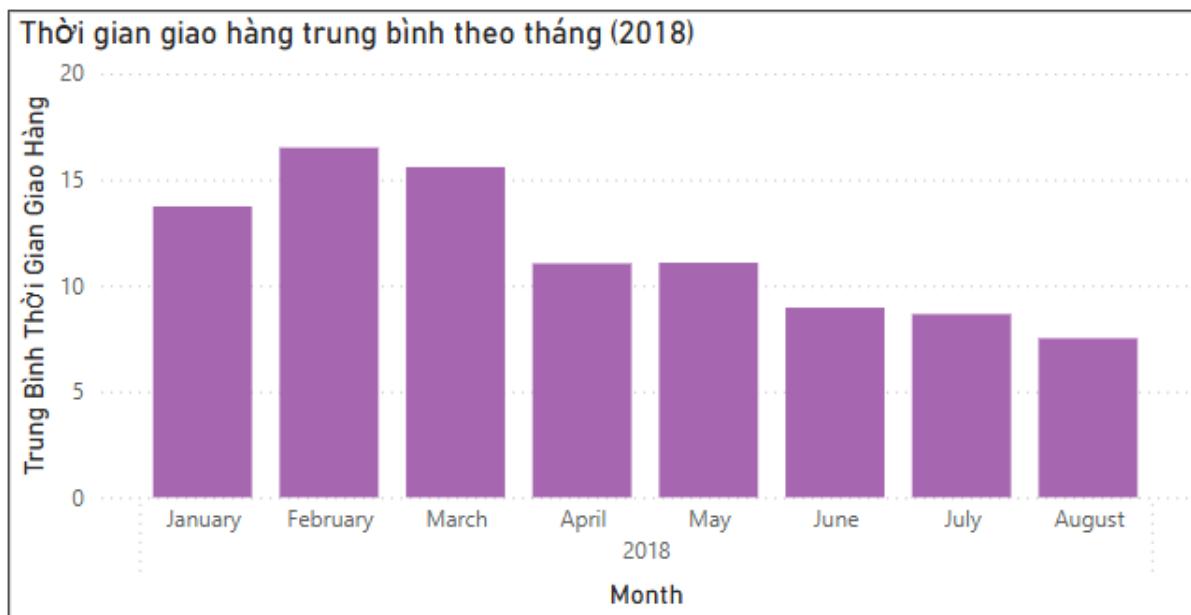
## 7.2. Thực hiện phân tích dữ liệu cho Business process: Order Fulfillment



### 7.2.1. Câu hỏi: Cho biết thời gian giao hàng trung bình qua các tháng của năm 2017 và 2018

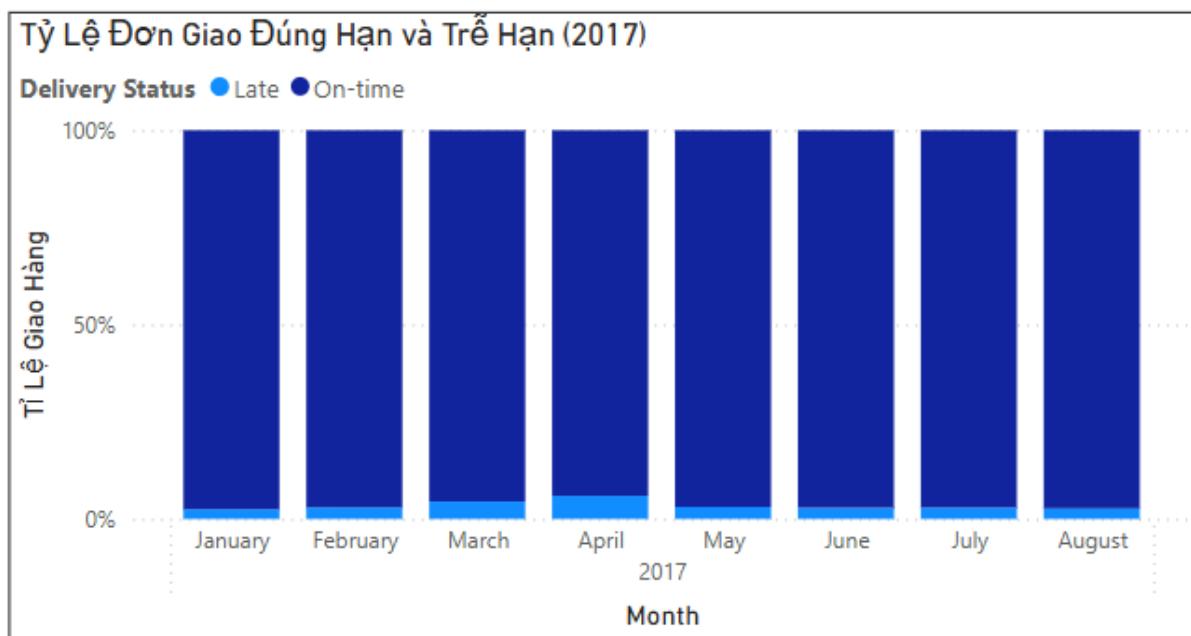


Biểu đồ 2017: Thời gian giao hàng trung bình duy trì ở mức tương đối ổn định (khoảng 15–17 ngày), với xu hướng giảm nhẹ từ tháng 4 đến tháng 8, cho thấy sự cải thiện dần trong nửa cuối năm.

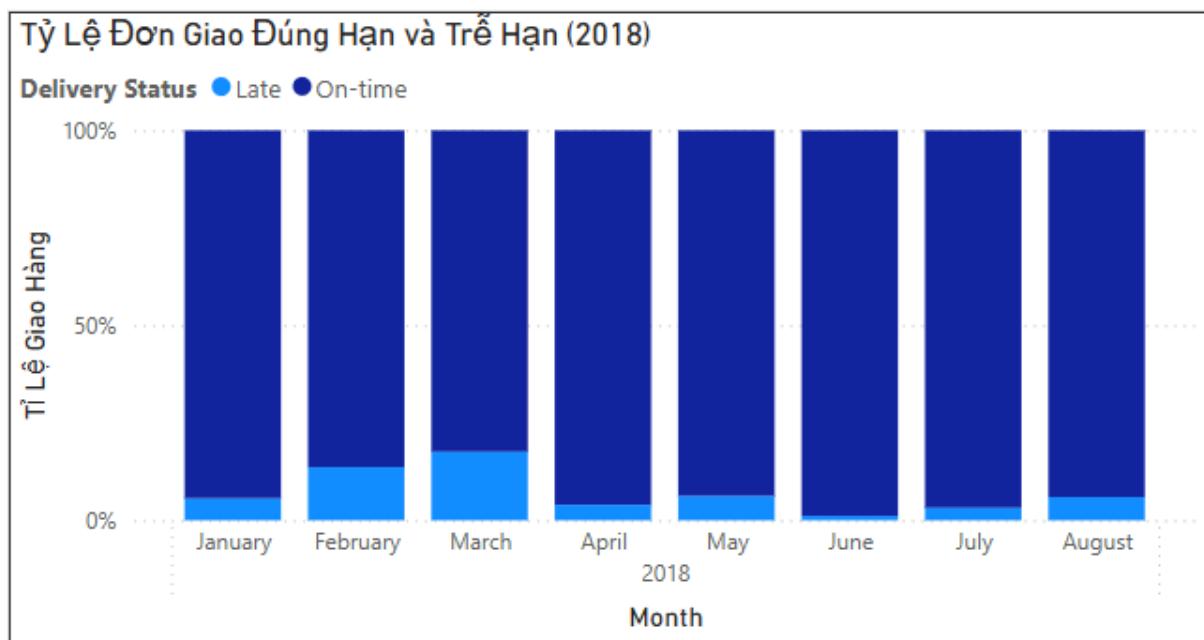


Biểu đồ 2018: Đầu năm có thời gian giao hàng cao hơn, đặc biệt tháng 2 vượt 20 ngày – thể hiện hiệu suất thấp. Tuy nhiên, từ tháng 4 trở đi, thời gian giao giảm rõ rệt, cho thấy hiệu quả giao hàng được cải thiện rõ rệt qua các tháng.

#### 7.2.2. Câu hỏi: Cho biết tỉ lệ đơn giao đúng hạn và trễ hạn qua các tháng của năm 2017 và 2018



Biểu đồ 2017: Tỷ lệ đơn giao đúng hạn duy trì rất cao qua các tháng, gần như tuyệt đối. Điều này cho thấy hiệu quả vận hành và cam kết giao hàng được đảm bảo ổn định trong suốt năm.

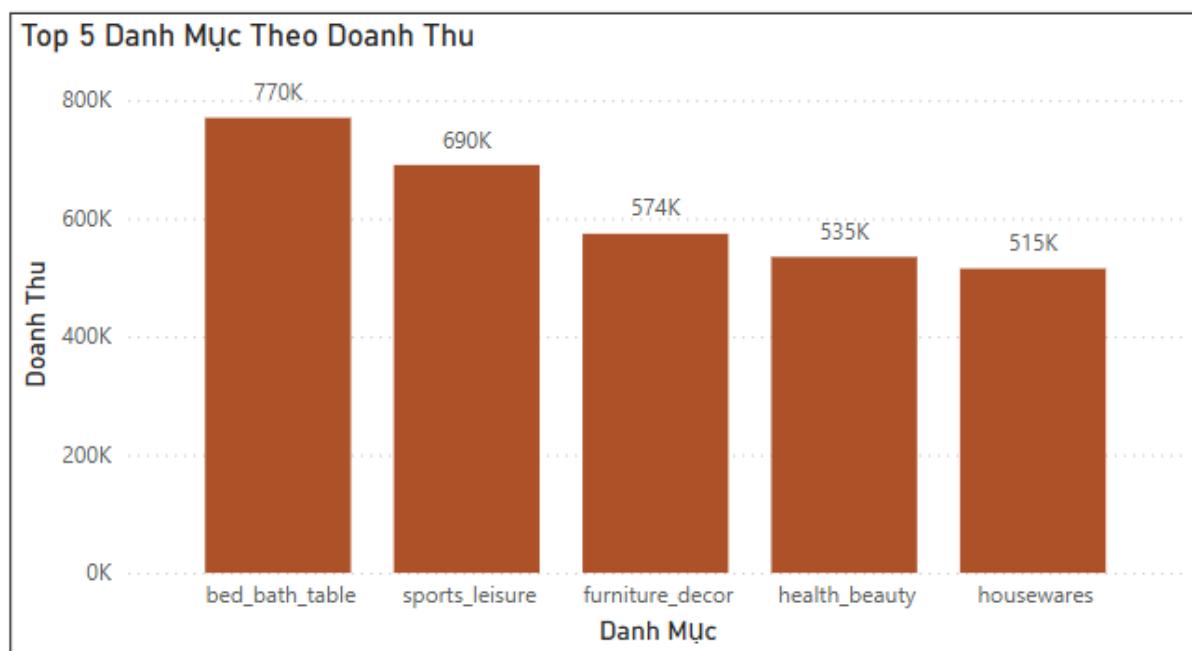
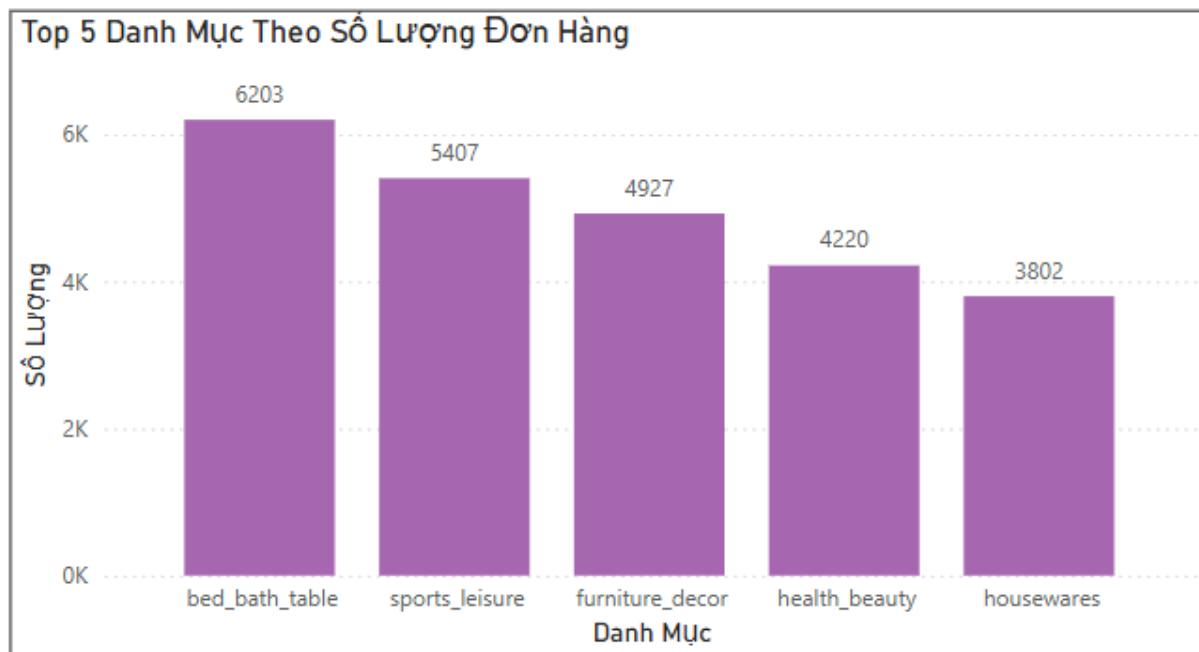


Biểu đồ 2018: Đầu năm (đặc biệt là tháng 2 và 3) ghi nhận tỷ lệ giao trễ cao hơn đáng kể so với các tháng còn lại. Tuy nhiên, từ tháng 4 trở đi, tỷ lệ giao đúng hạn được cải thiện rõ rệt và duy trì ở mức gần 100%, tương đương với năm 2017

### 7.3. Thực hiện phân tích dữ liệu cho Business process: Product & Seller Performance



### 7.3.1. Câu hỏi: Cho biết 5 danh mục sản phẩm bán chạy và đem theo doanh thu nhiều nhất trong năm 2018

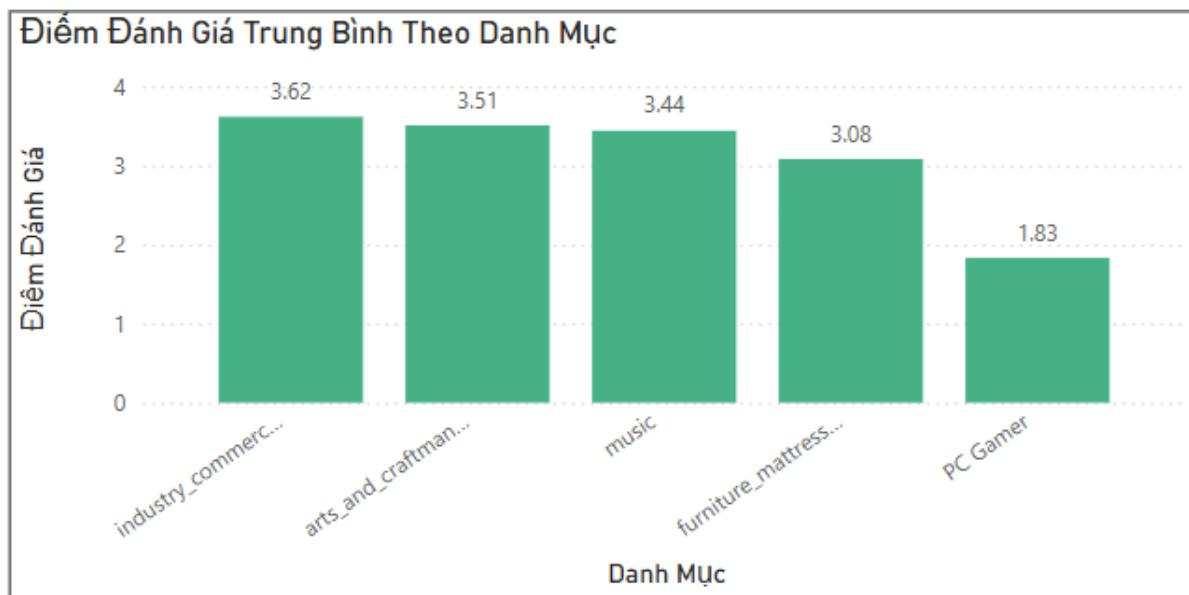


Hai biểu đồ trên cho thấy Top 5 danh mục sản phẩm bán chạy nhất và mang lại doanh thu cao nhất trong năm 2018. Có thể nhận thấy:

Danh mục "bed\_bath\_table" dẫn đầu cả về số lượng đơn hàng (6203 đơn) và doanh thu (770K), cho thấy đây là mặt hàng có sức mua lớn và giá trị cao. Các danh mục còn lại như sports\_leisure, furniture\_decor, health\_beauty, và housewares cũng nằm trong Top 5 ở cả hai biểu đồ, chứng tỏ chúng có sức tiêu thụ ổn định và đóng góp đáng kể vào doanh thu. Mặc dù thứ tự có sự thay đổi nhẹ giữa hai biểu đồ, nhưng nhìn chung, các

danh mục có số lượng đơn hàng cao cũng đồng thời đem lại doanh thu lớn, cho thấy hiệu quả kinh doanh đồng đều.

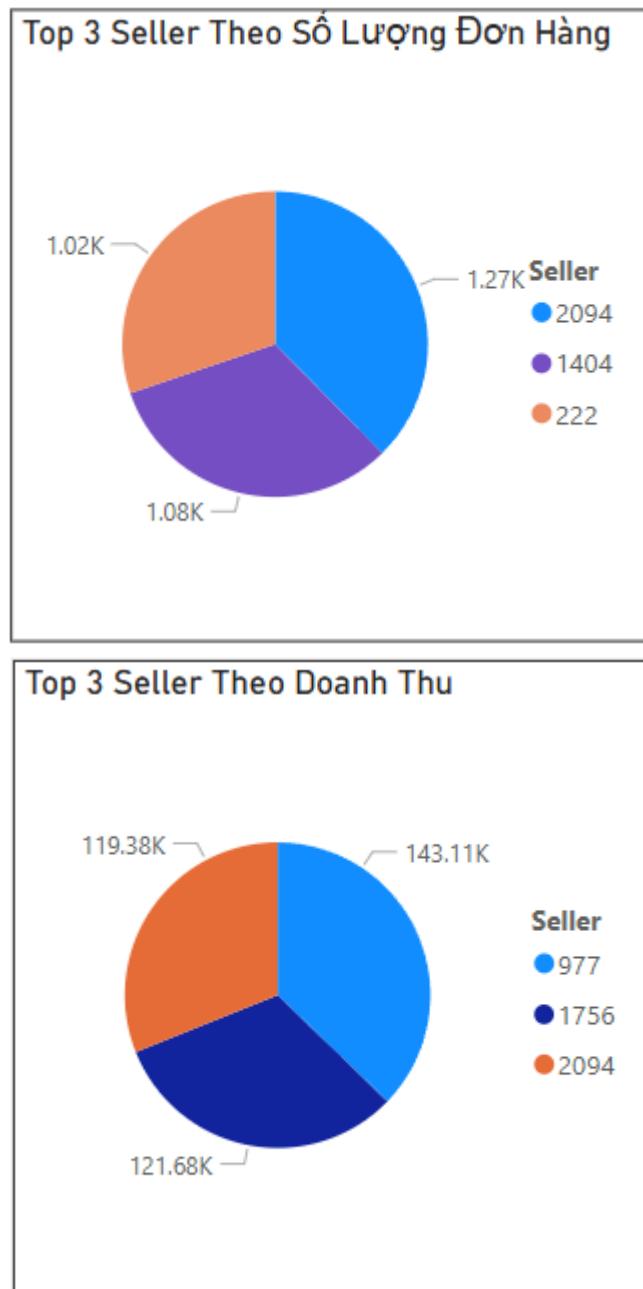
### 7.3.2. Câu hỏi: Cho biết 5 danh mục sản phẩm nhận phản hồi tệ nhất trong năm 2018



Biểu đồ thể hiện điểm đánh giá trung bình của khách hàng theo từng danh mục sản phẩm. Có thể nhận thấy:

Danh mục "industry\_commerce" có điểm đánh giá cao nhất với 3.62, cho thấy mức độ hài lòng của khách hàng khá tốt đối với nhóm sản phẩm này. Các danh mục arts\_and\_crafts, music, và furniture\_matters có điểm đánh giá trung bình ở mức khá, dao động từ 3.08 đến 3.51. Danh mục "PC Gamer" có điểm đánh giá thấp nhất, chỉ đạt 1.83, phản ánh có thể tồn tại các vấn đề về chất lượng, dịch vụ hoặc kỳ vọng không được đáp ứng đúng với nhóm sản phẩm này. Từ biểu đồ này, doanh nghiệp có thể xác định được các danh mục cần cải thiện chất lượng hoặc nâng cao trải nghiệm khách hàng, đặc biệt là với nhóm "PC Gamer".

### 7.3.3. Câu hỏi: Cho biết 3 Seller bán được nhiều đơn hàng và đem lại doanh thu nhiều nhất trong năm 2018



Hai biểu đồ tròn trên thể hiện Top 3 người bán hàng (Seller) theo số lượng đơn hàng và doanh thu:

Seller 2094 dẫn đầu về số lượng đơn hàng (1.27K) và cũng đứng Top 1 về doanh thu (143.11K), cho thấy hiệu quả bán hàng cao cả về số lượng lẫn giá trị. Seller 1404 có số lượng đơn hàng lớn (1.08K) nhưng không xuất hiện trong Top 3 doanh thu, cho thấy giá trị trung bình mỗi đơn hàng có thể thấp. Ngược lại, Seller 977 không nằm trong Top 3 về số đơn hàng nhưng đạt doanh thu cao thứ 2, cho thấy các đơn hàng của người

bán này có giá trị cao. Seller 222 xuất hiện trong cả hai biểu đồ, cho thấy sự ổn định về mặt hiệu suất kinh doanh.

Từ đây có thể đánh giá được hiệu quả của từng Seller không chỉ dựa vào số lượng đơn hàng mà còn cần xem xét giá trị đơn hàng để đưa ra chiến lược hỗ trợ phù hợp.

## **8. Kết luận**

### **8.1. Kết quả đạt được**

Dự án “Xây dựng kho dữ liệu từ hệ thống thương mại điện tử OList tại Brazil” đã mang lại cho nhóm chúng em nhiều trải nghiệm quý báu, đồng thời giúp củng cố và nâng cao kiến thức chuyên môn về lĩnh vực kho dữ liệu một cách toàn diện hơn. Thông qua quá trình thực hiện, chúng em đã có cơ hội áp dụng những khái niệm lý thuyết đã học vào thực tế, từ việc xác định các quy trình nghiệp vụ, thiết kế mô hình dữ liệu đa chiều đến triển khai quy trình ETL bằng công cụ SSIS, xây dựng các khối phân tích với SSAS, và trực quan hóa dữ liệu thông qua Power BI Desktop.

Không chỉ dừng lại ở việc vận dụng kỹ thuật, dự án còn giúp nhóm rèn luyện tư duy phân tích, khả năng tiếp cận và xử lý dữ liệu đầu vào một cách logic và có hệ thống. Chúng em học được cách đặt ra các câu hỏi định hướng để khai thác dữ liệu hiệu quả hơn, cũng như cách làm việc nhóm, thảo luận, và đưa ra các quyết định kỹ thuật trong quá trình xây dựng kho dữ liệu. Ngoài ra, chúng em cũng tích lũy thêm kinh nghiệm thực tế trong việc xử lý các vấn đề phát sinh khi thao tác với dữ liệu thô, từ đó nâng cao khả năng thích ứng và giải quyết vấn đề trong môi trường làm việc thực tiễn.

### **8.2. Những hạn chế**

Trong quá trình thực hiện xây dựng kho dữ liệu, nhóm chúng em không tránh khỏi gặp một số những khó khăn vì còn hạn chế về mặt kiến thức. Tập dữ liệu còn nhiều dữ kiện chưa được khai thác hết. Những thực hiện của chúng em chưa đạt đến mức nâng cao trong tổng thể quá trình xây dựng một kho dữ liệu hoàn chỉnh.

## **9. Tài liệu tham khảo**

[1]. Tài liệu các file PDF hướng dẫn về Data Integration với SQL Server, SSIS, SSAS của GVHD Ths. Nguyễn Văn Thành.

[2] Inmon, W. H., “Building the Data Warehouse,” Wiley, 2005.

[3] Kimball, R., Ross, M., “The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling,” Wiley, 2013.

- [4] Golfarelli, M., Rizzi, S., “Data Warehouse Design: Modern Principles and Methodologies,” McGraw-Hill, 2009.
- [5] IBM, “IBM Data Warehousing and Business Intelligence,” IBM Redbooks, 2010.
- [6] Microsoft, “Data Warehousing Guide,” Microsoft Learn, 2023. [Online]. Available: <https://learn.microsoft.com/en-us/sql relational-databases>
- [7] Oracle, “Data Warehousing Concepts,” Oracle Documentation, 2022. [Online]. Available: <https://docs.oracle.com/en/database/oracle database/19/dwhsg/index.html>
- [8] Google Cloud, “Introduction to Data Warehousing,” Google Cloud Documentation, 2023. [Online]. Available: <https://cloud.google.com/learn/what-is-a-data-warehouse>