

BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HỒ CHÍ MINH  
KHOA CÔNG NGHỆ THÔNG TIN



**HCMUTE**

**BÁO CÁO ĐỒ ÁN CUỐI KỲ  
ĐỀ TÀI: XÂY DỰNG KHO DỮ LIỆU CHO  
KHÁCH SẠN**

Môn học: KHO DỮ LIỆU  
Mã lớp học phần: DAWH430784\_23\_2\_02  
GVHD: ThS. Nguyễn Văn Thành

Nhóm sinh viên thực hiện: Nhóm 6

Lại Văn An	21110363
Nguyễn Đức Kha	21133044
Nguyễn Nhật Trường	21110935

TP. Hồ Chí Minh, tháng 05 năm 2024

## **DANH SÁCH THÀNH VIÊN THAM GIA THỰC HIỆN ĐỀ TÀI VÀ VIẾT BÁO CÁO**

**Môn: Kho dữ liệu - HỌC KÌ II – NĂM HỌC 2023 – 2024**

<b>STT</b>	<b>HỌ VÀ TÊN</b>	<b>MSSV</b>	<b>TỶ LỆ ĐÓNG GÓP</b>
1	Nguyễn Đức Kha	21133044	100%
2	Nguyễn Nhật Trường	21110935	100%
3	Lại Văn An	21110363	100%

#### **Nhận xét của giảng viên:**

*Ngày ... tháng 05 năm 2024*

## *Giảng viên chấm điểm*

*Ths. Nguyễn Văn Thành*

## LỜI CẢM ƠN

Lời đầu tiên, nhóm chúng em xin gửi lời cảm ơn chân thành đến **Trường Đại học Sư phạm Kỹ thuật TPHCM** đã đưa môn học **Kho Dữ Liệu** vào chương trình giảng dạy. Đặc biệt, chúng em xin gửi lời cảm ơn sâu sắc đến giảng viên bộ môn – **Th.S Thầy Nguyễn Văn Thành** đã dạy dỗ, truyền đạt những kiến thức quý báu cho chúng em trong suốt thời gian học tập vừa qua. Trong thời gian tham gia lớp học Kho Dữ Liệu của Thầy, nhóm em đã có thêm cho mình nhiều kiến thức bổ ích, tinh thần học tập hiệu quả, nghiêm túc. Đây chắc chắn sẽ là những kiến thức quý báu, là hành trang để chúng em có thể vững bước sau này. Bộ môn Kho Dữ Liệu là môn học thú vị, vô cùng bổ ích và có tính thực tế cao. Đảm bảo cung cấp đủ kiến thức, gắn liền với nhu cầu thực tiễn của sinh viên. Mặc dù chúng em đã cố gắng hết sức nhưng chắc chắn bài đồ án này khó có thể tránh khỏi những thiếu sót và nhiều chỗ còn chưa chính xác, kính mong Thầy xem xét và góp ý để bài đồ án của chúng em được hoàn thiện hơn.

Nhóm chúng em xin chân thành cảm ơn!

# PHÂN CÔNG NHIỆM VỤ

Nhiệm vụ	Nhật Trưởng	Đức Kha	Văn An
Tìm kiếm tập dữ liệu	x	x	x
<b>Mô hình hóa dữ liệu vào workbook</b>	x	x	x
+Dim Date			
+ Dim Hotel			x
+ Dim Customers			
+ Fact Canceled		x	
+Fact Booking	x		
<b>Thực hiện ETL đưa dữ liệu vào các bảng dim và fact dùng SSIS</b>	x	x	x
+ Tạo và đưa dữ liệu vào các dim stages và dim			x
+ Tạo và đưa dữ liệu vào FactBooking stage và FactBooking	x		
+ Tạo và đưa dữ liệu vào FactCanceledstage và FactCanceled		x	
Nhập dữ liệu vào SSAS, tạo data source, data source view và cube		x	
<b>Tạo các phân cấp hierarchy cho các bảng dim và các attribute relationships.</b>	x	x	x
+ Hierarchy cho DimDate	x		
+ Hierarchy cho DimGeography, DimProduct			x
+ Hierarchy cho DimInventory, DimEmployee		x	

<b>Thực hiện trả lời các câu hỏi đã đặt ra cho business process: Canceled Analysis</b>	x	x	x
Số lượng huỷ phòng của Resort Hotel và City Hotel ở các tháng theo từng năm.	x		
Lượng huỷ phòng theo quốc gia của từng năm.		x	
Cho biết số lượng huỷ phòng và giá trung bình ADR qua các tháng theo từng năm của Resort Hotel và City Hotel.			x
<b>Thực hiện trả lời các câu hỏi đã đặt ra cho business process: Booking Analysis</b>	x		x
Số lượng đặt phòng qua các công ty lữ hành	x		
Thống kê số lượng đặt phòng ứng với các phương thức đặt phòng.			x
+ Trả lời câu hỏi dùng SSAS		x	
+ Trả lời câu hỏi dùng Power BI Desktop		x	
+ Trả lời câu hỏi dùng Excel Pivot	x		

# **CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI**

## **1.1. Lý do chọn đề tài**

Lý do chọn đề tài kho dữ liệu cho khách sạn xuất phát từ nhu cầu cấp thiết của ngành công nghiệp này trong việc quản lý và khai thác hiệu quả nguồn dữ liệu phong phú. Kho dữ liệu không chỉ giúp tập trung và tổ chức thông tin từ nhiều nguồn như đặt phòng, dịch vụ, và phản hồi khách hàng, mà còn hỗ trợ phân tích xu hướng và dự báo nhu cầu tương lai. Bằng cách tối ưu hóa hoạt động kinh doanh và cá nhân hóa trải nghiệm khách hàng, kho dữ liệu giúp nâng cao hiệu suất và khả năng cạnh tranh của khách sạn. Đồng thời, nó đảm bảo việc tuân thủ các quy định pháp lý về bảo mật thông tin, tạo cơ sở vững chắc cho sự đổi mới và áp dụng công nghệ tiên tiến như trí tuệ nhân tạo và học máy. Với những lợi ích vượt trội này, việc nghiên cứu và triển khai kho dữ liệu trở thành một yếu tố then chốt để thúc đẩy sự phát triển bền vững và thành công của ngành khách sạn.

## **1.2. Tổng quan về tập dữ liệu**

### **1.2.1. Nguồn dữ liệu**

- Nhóm sử dụng Tập dữ liệu Hotel Bookings được lấy từ trang web Kaggle (<https://www.kaggle.com/>), Kaggle là một nền tảng trực tuyến cho cộng đồng Machine Learning (ML) và Khoa học dữ liệu. Kaggle cho phép người dùng chia sẻ, tìm kiếm các bộ dữ liệu; tìm hiểu và xây dựng models, tương tác với những nhà khoa học và kỹ sư ML trên toàn thế giới; tham gia các cuộc thi để có cơ hội chiến thắng những giải thưởng giá trị.
- Đường dẫn tải tập dữ liệu: <https://www.kaggle.com/code/ahmedsalouh/hotel-bookings-analysis>

### 1.2.2. Mô tả chi tiết tập dữ liệu

- Tập dữ liệu ‘dataset.xlsx’ chứa thông tin về việc đặt phòng cho hai khách sạn ở Bồ Đào Nha (một khách sạn nghỉ dưỡng và một khách sạn trong thành phố) dự kiến đến trong khoảng thời gian từ ngày 1 tháng 7 năm 2015 đến ngày 31 tháng 8 năm 2017 và đối với cả hai khách sạn, thông tin giống nhau đã được thu thập như quan sát và đặt phòng khách sạn , bao gồm cả những cái đã bị hủy. Dữ liệu bao gồm các 37 cột với mô tả một số cột chính như sau:

Tên cột	Mô tả
<b>hotel</b>	Một trong các khách sạn là khách sạn nghỉ dưỡng và một khách sạn là khách sạn thành phố.
<b>is_canceled</b>	Giá trị cho biết liệu đặt phòng có bị hủy (1) hay không (0).
<b>lead_time</b>	Thời gian từ khi đặt phòng đến khi khách đến.
<b>arrival_date_year</b>	Năm của ngày đến.
<b>arrival_date_month</b>	Tháng của ngày đến với 12 danh mục: “1” đến “12”.
<b>arrival_date_week_number</b>	Số tuần của ngày đến.
<b>arrival_date_day_of_month</b>	Ngày trong tháng của ngày đến.
<b>country</b>	Quốc gia xuất xứ.
<b>market_segment</b>	Phân khúc thị trường. Trong danh mục, thuật ngữ “TA” nghĩa là “Travel Agents” và “TO” nghĩa là “Tour Operators”.
<b>reserved_room_type</b>	Mã loại phòng được đặt. Mã được trình bày thay vì tên để đảm bảo tính ẩn danh.
<b>agent</b>	ID của đại lý du lịch đã thực hiện đặt phòng.
<b>adr</b>	Giá Phòng Trung Bình Hàng Ngày (Tính bằng cách chia tổng số giao dịch lưu trú cho tổng số đêm ở).
<b>customer_type</b>	Group – khi đặt phòng liên quan đến một nhóm; Transient – khi đặt phòng không phải là một phần của nhóm hoặc hợp đồng, và không liên quan đến đặt phòng chuyển tiếp khác; Transient-party – khi đặt phòng là dạng chuyển tiếp, nhưng liên quan đến ít nhất một đặt phòng chuyển tiếp khác.

### **1.3. Giới thiệu các công cụ được sử dụng**

- Visual Studio 2019/2022 tích hợp SSIS (SQL Server Integration Services) và SSAS (SQL Server Analysis Services).
- SQL Server Management Studio 2019
- Power BI.

# CHƯƠNG 2: THIẾT KẾ XÂY DỰNG CƠ SỞ DỮ LIỆU TÁC NGHIỆP(OLAP)

## Xác định các Business Process và bảng Fact

Xây dựng Detailed Bus Matrix xác định các Business Process, bảng Fact, bảng Dim cần thiết

Business Process Name	This usually comes from the enterprise bus matrix.		Granularity	Facts	DimDate	DimCustomer	DimHotel
	Table	Type					
CusBooking	FactBooking	Transactional	one row per booking detail	ADR	x	x	x
CusCanceled	FactCanceled	Transactional	one row per canceled detail	required_car_parking_spaces total_of_special_requests	x	x	x

## 2.1. Business Process: Canceled Analysis

- Các câu hỏi đặt ra :

- + Số lượng huỷ phòng của Resort Hotel và City Hotel ở các tháng theo từng năm.
- + Lượng huỷ phòng theo quốc gia của từng năm.
- + Cho biết số lượng huỷ phòng và giá trung bình ADR qua các tháng theo từng năm của Resort Hotel và City Hotel.

## 2.2. Business Process: BookingAnalysis

- Các câu hỏi đặt ra :

- + Thống kê số lượng đặt phòng ứng với các phương thức đặt phòng.
- + Số lượng đặt phòng qua các công ty lữ hành.

## 2.3. Xây dựng các bảng Dimension

### 2.3.1. DimDate

Table Name	DimDate	Dimension	Date	Always create a table description	You can put technical information	Source System	Source Schema	Source Table	Source Field Name	Source Data type
Table Type										
Display Name										
Database Schema										
Table Description										
Comment										
Biz Filter Logic										
Size										
Generate Script?	Y									
Column Name	Display Name	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value		
Datekey	Datekey	int			PK		N			
arrival_date_year	arrival_date_year	varchar	50				N	9999		
arrival_date_month	arrival_date_month	varchar	50				N	0		
arrival_date_week_number	arrival_date_week_number	varchar	50				N	53		
arrival_date_day_of_month	arrival_date_day_of_month	varchar	50				N	0		

### 2.3.2. DimHotel.

Table Name		DimHotel								
Table Type		Dimension								
Display Name		Hotel								
Database Schema		Always create a table description								
Table Description		You can put technical information here								
Comment										
Biz Filter Logic										
Size										
Generate Script?		Y								
Column Name		Display Name	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source
HotelKey	HotelKey	HotelName	int varchar	50		PK	N	N		Source System
HotelName	HotelName									Source Schema
										Source Table
										Source Field Name
										Source Datatype

### 2.3.3. DimCustomers.

Table Name		DimCustomers								
Table Type		Dimension								
Display Name		Customers								
Database Schema		Always create a table description								
Table Description		You can put technical information here								
Comment										
Biz Filter Logic										
Size										
Generate Script?		Y								
Column Name		Display Name	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source
CustomerKey	CustomerKey	Country	int varchar	20		PK	N	N	NULL	Source System
Country	Country									Source Schema
CustomerType	CustomerType		varchar	50			N	NULL		Source Table
is_repeated_guest	is_repeated_guest		int				N	0		Source Field Name
previous_cancellations	previous_cancellations		int				N	0		Source Datatype
previous_bookings_notCanceled	previous_bookings_notCanceled		int				N	0		

## 2.4. Xây dựng các bảng Fact

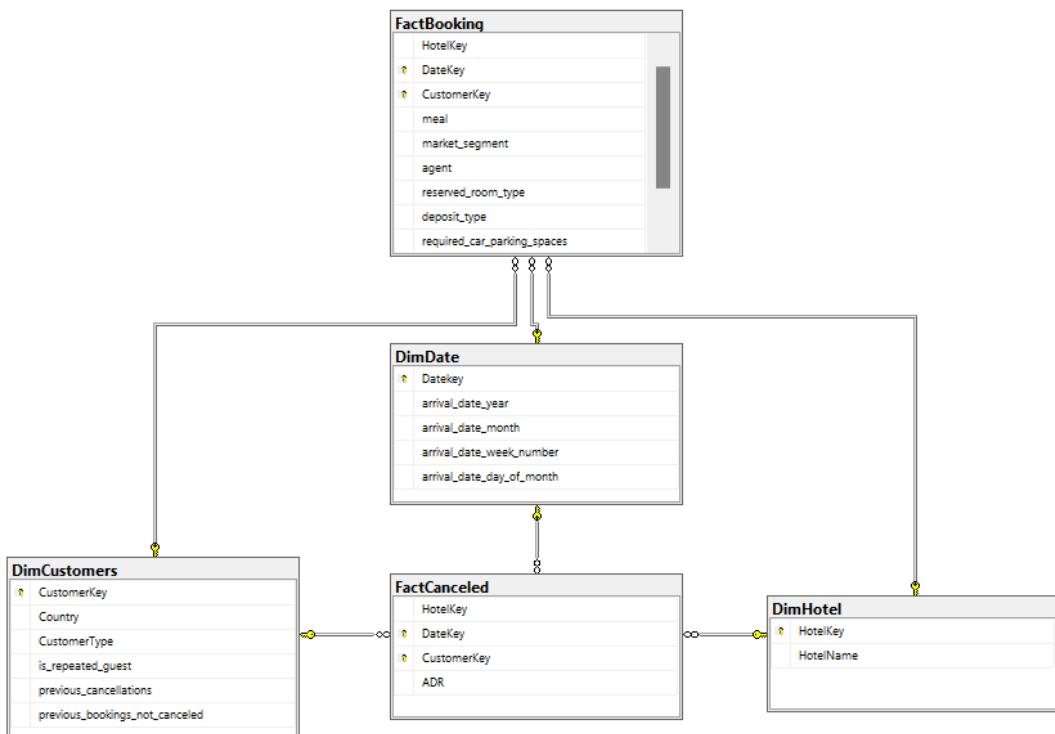
### 2.4.1. FactCanceled.

Table Name		FactCanceled								
Table Type		Fact								
Display Name		Canceled								
Database Schema		Always create a table description								
Table Description		You can put technical information here								
Comment										
Biz Filter Logic										
Size										
Generate Script?		Y								
Column Name		Display Name	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source
HotelKey	HotelKey	DateKey	int			FK	DimHotel.HotelKey	N		Source System
DateKey	DateKey	CustomerKey	int			FK, PK	DimDate.DateKey	N		Source Schema
CustomerKey	CustomerKey	ADR	float			FK, PK	DimCustomers.CustomerKey	N		Source Table
ADR	ADR									Source Field Name
										Source Datatype

### 2.4.2. FactBooking.

Table Name		FactBooking								
Table Type		Fact								
Display Name		Booking								
Database Schema		Always create a table description, when you create a fact table								
Table Description		You can put technical information here								
Comment										
Biz Filter Logic										
Size										
Generate Script?		Y								
Column Name		Display Name	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source
HotelKey	HotelKey	meal	int			FK	DimHotel.HotelKey	N		Source System
DateKey	DateKey	market_segment	int			FK, PK	DimDate.DateKey	N		Source Schema
CustomerKey	CustomerKey	agent	int			FK, PK	DimCustomers.CustomerKey	N		Source Table
meal	meal	reserved_room_type	nvarchar	255				Y		Source Field Name
market_segment	market_segment	deposit_type	nvarchar	255				N		Source Datatype
agent	agent	deposit_type	nvarchar	255				Y		
reserved_room_type	reserved_room_type	total_of_special_requests	int	255				N		
deposit_type	deposit_type	total_of_special_requests	int					Y		
total_of_special_requests	total_of_special_requests									

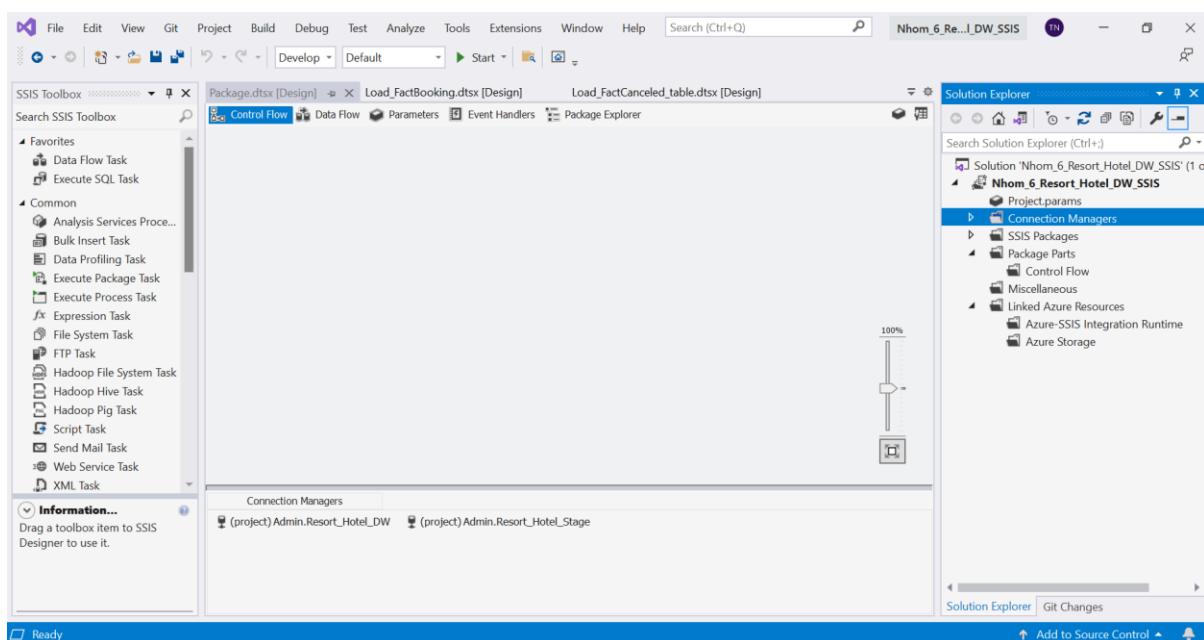
## 2.5. Star Schema (Lược đồ hình sao)



# CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS)

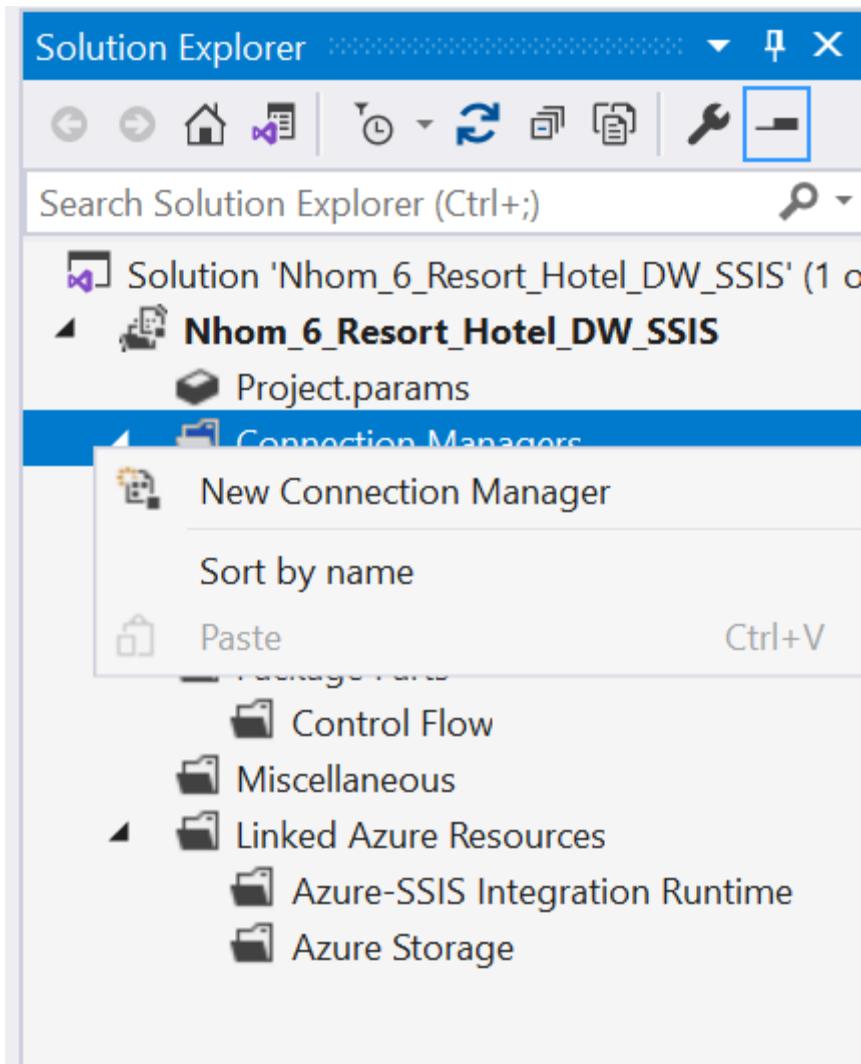
## 3.1. Tạo ETL project và solution

1. Từ Visual Studio menu, chọn File -> New -> Project.
2. Từ dialog:
  - a. Chọn Integration Services Project
  - b. Đặt tên: Nhom\_6\_Resort\_Hotel\_SSIS
3. Click OK

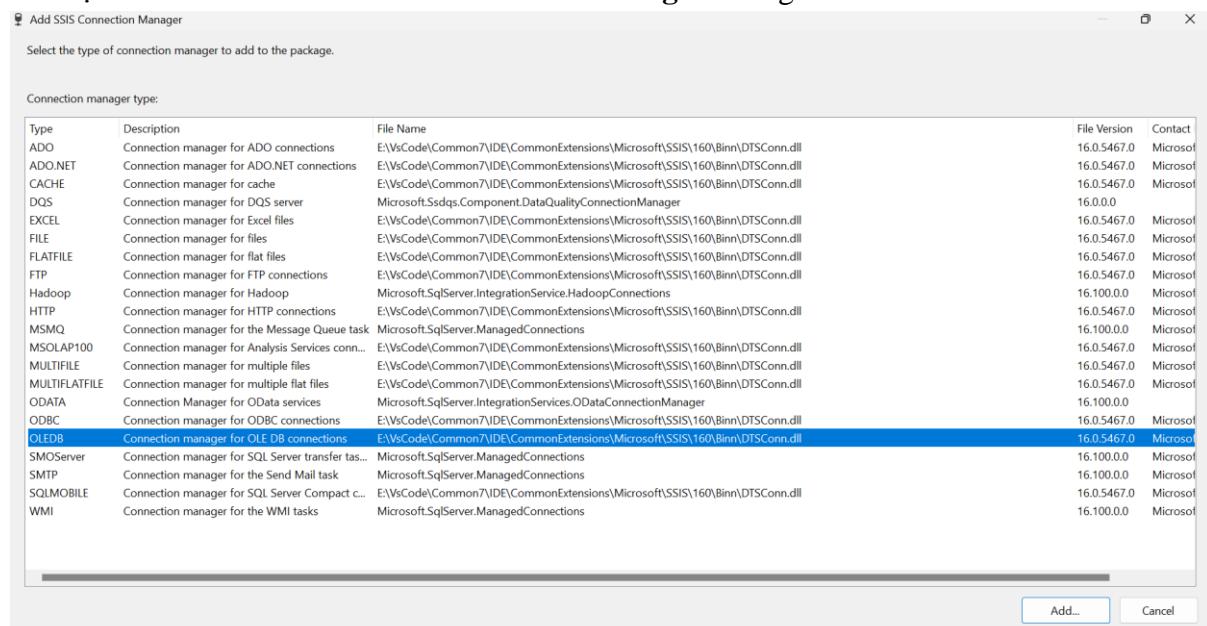


## 3.2. Thiết lập các Connection

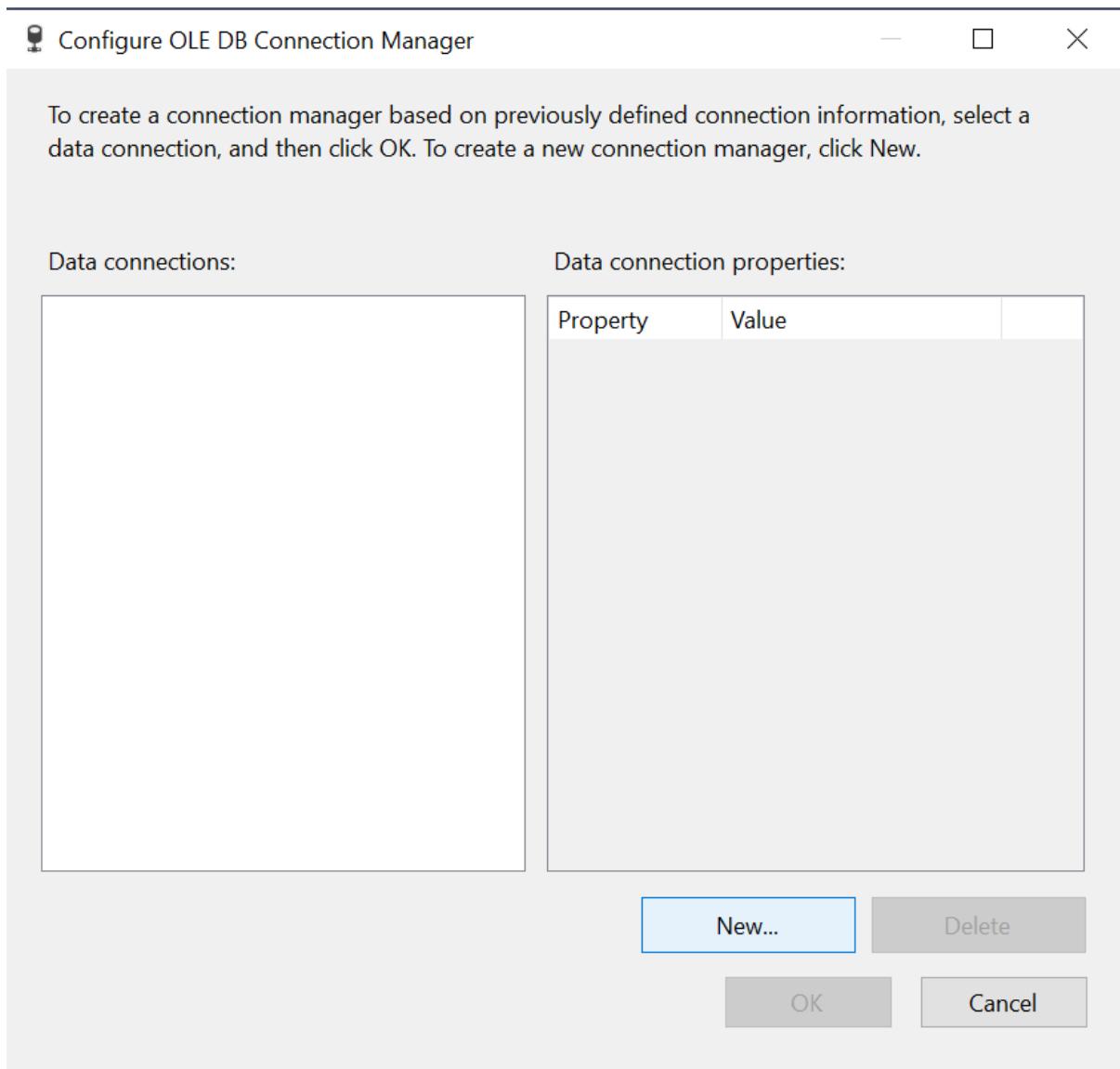
1. Tại solution explorer, right-click vào Connection Managers, chọn New Connection Manager.



## 2. Chọn OLEDB từ Add SSIS Connection Manager dialog và click Add...



## 3. Configure OLE DB Connection Manager dialog xuất hiện, click New...



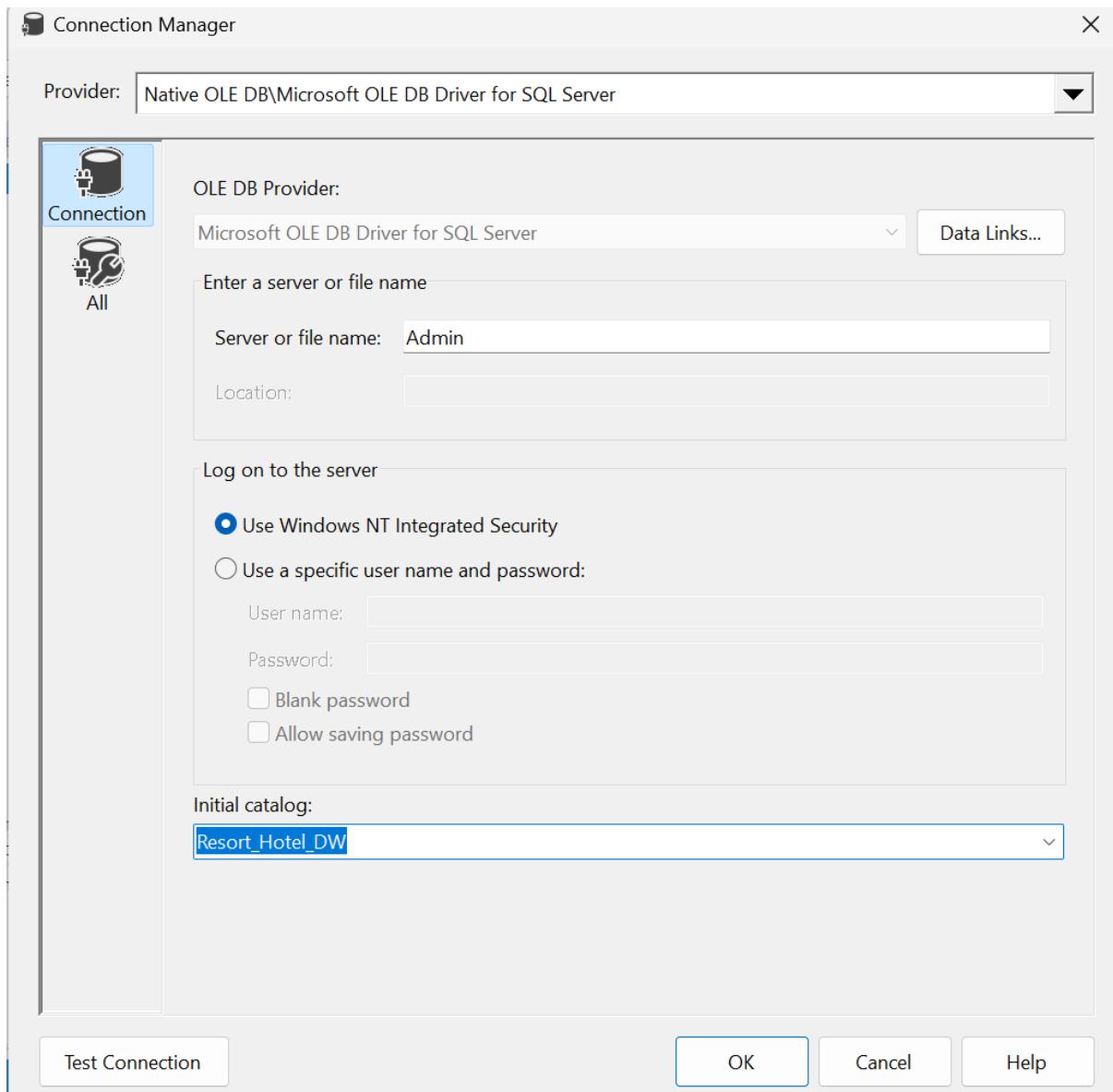
4. Đặt tên cho các thông tin như bên dưới:

Provider: **Microsoft OLE DB Driver for SQL Server**

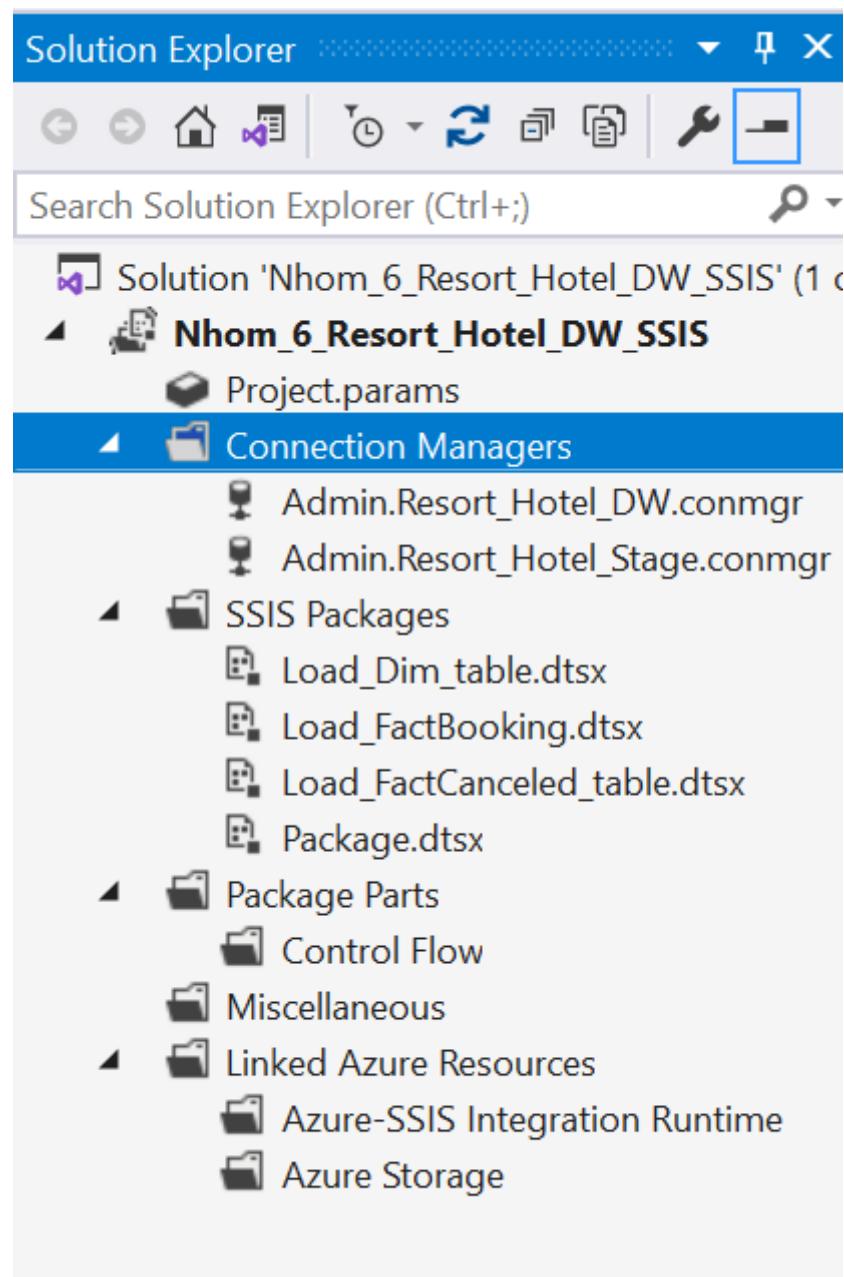
Server Name: **Admin**

Log on to the Server: **Use Windows Authentication**

Connect to a database: **Resort\_Hotel\_DW**

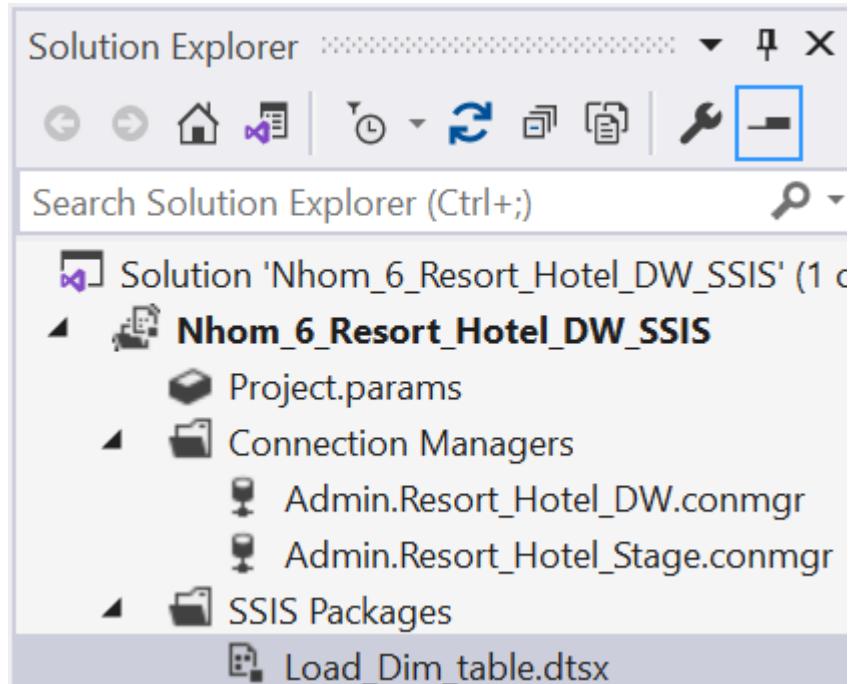


5. Lặp lại các bước trên cho **Resort\_Hotel\_Stage** source database. Sau khi tạo xong **Connection Managers** folder sẽ như bên dưới.

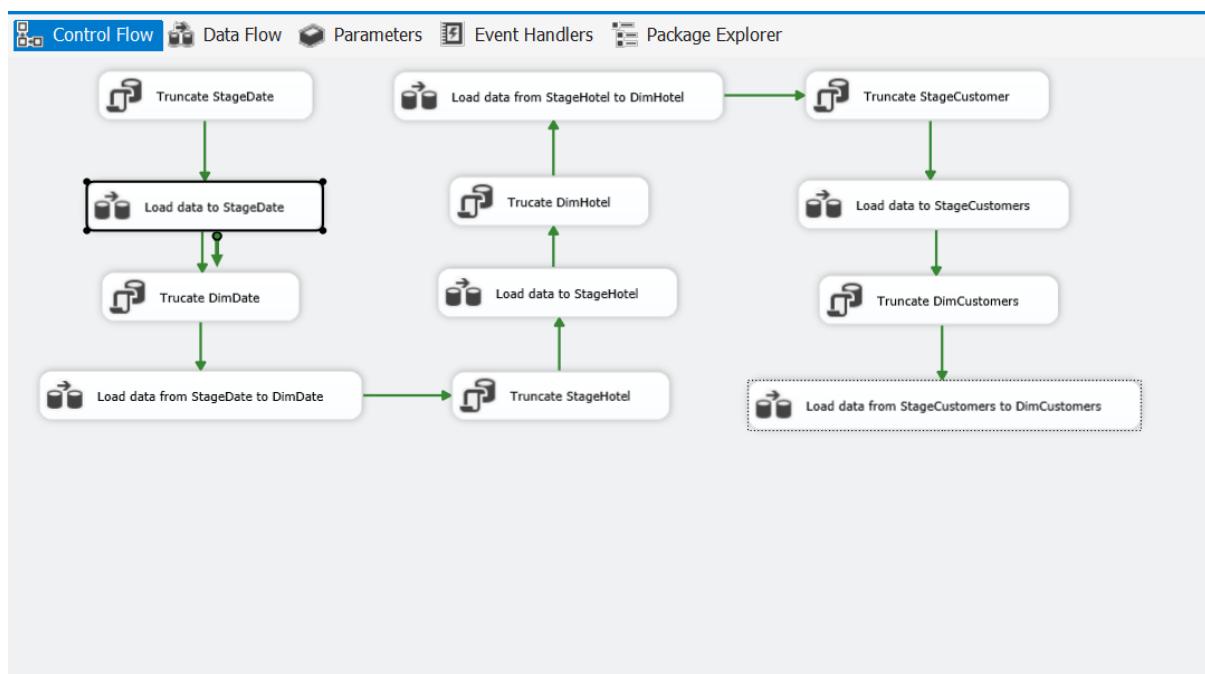


### 3.3. Import Dữ liệu vào các bảng dimension:

#### 3.3.1. Tạo Package



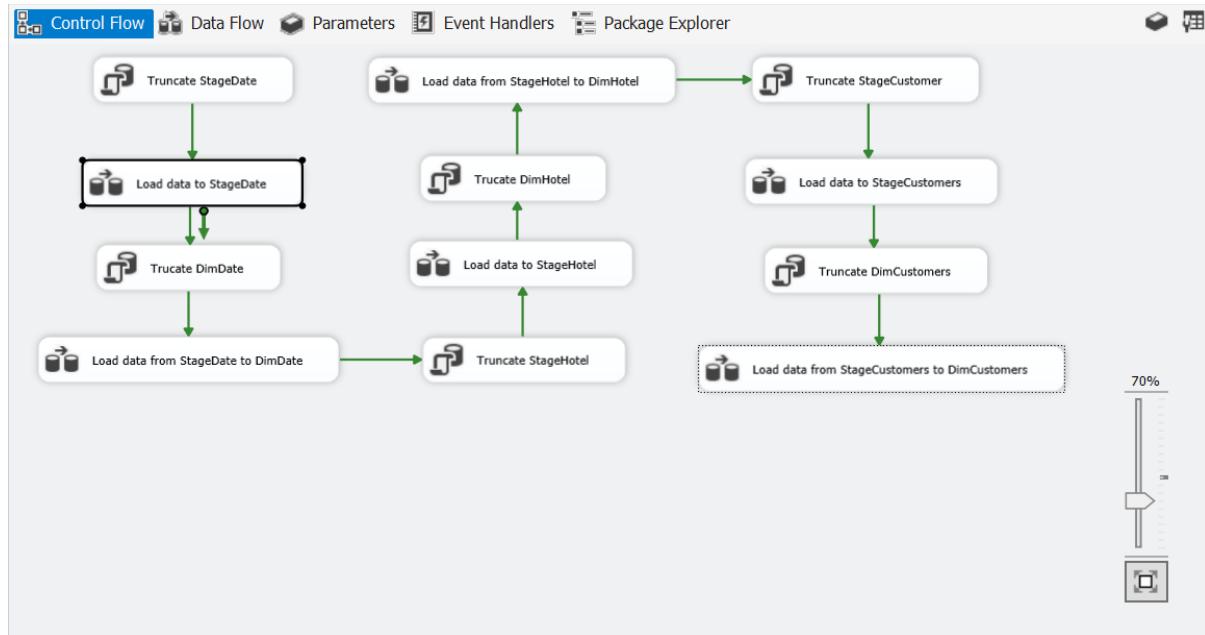
#### 3.3.2. Thiết lập Control Flow



Mô tả luồng dữ liệu:

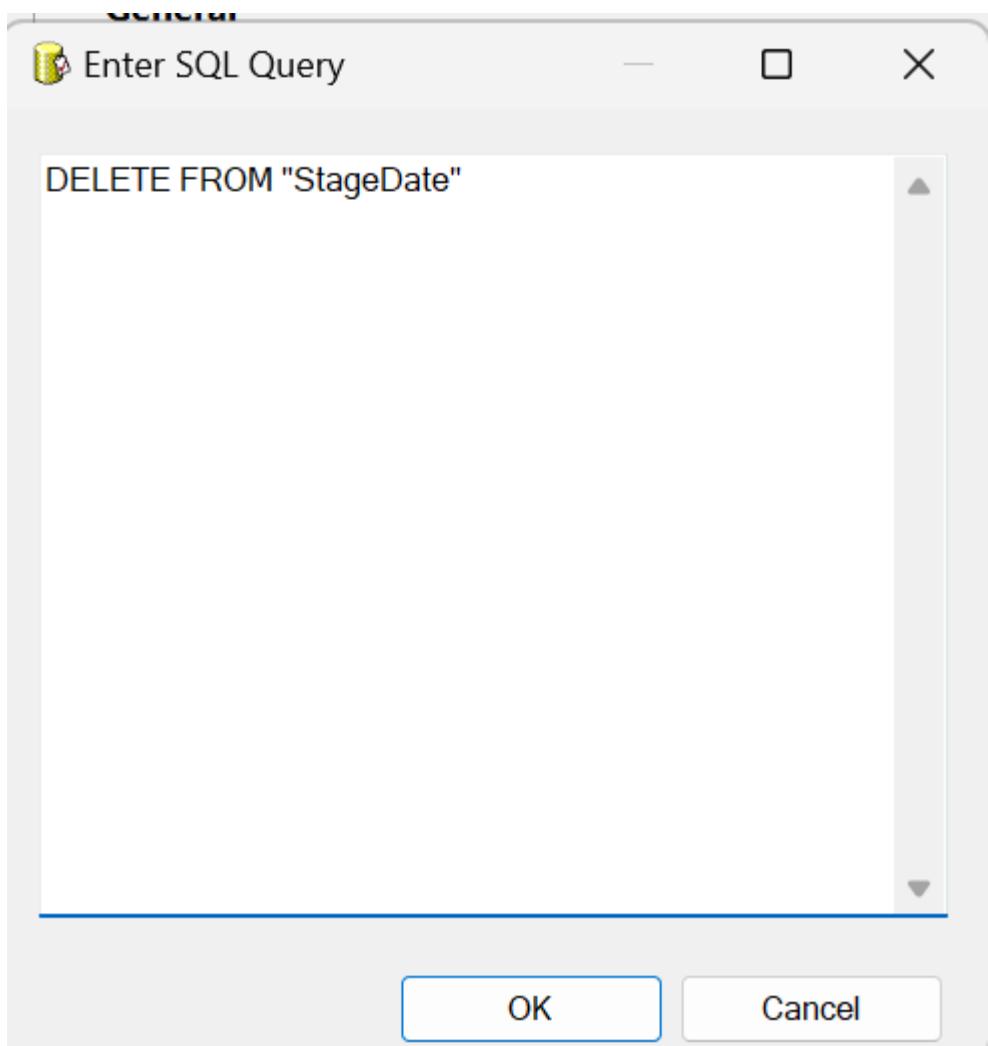
Đầu tiên cần load dữ liệu vào các bảng Stage trước thay vì load dữ liệu trước tiếp vào bảng Dim. Mục đích sử lý dữ liệu thô từ excel trước ở các bảng Stage để tránh gấp phải lỗi khi thực hiện load dữ liệu vào bảng Dim.

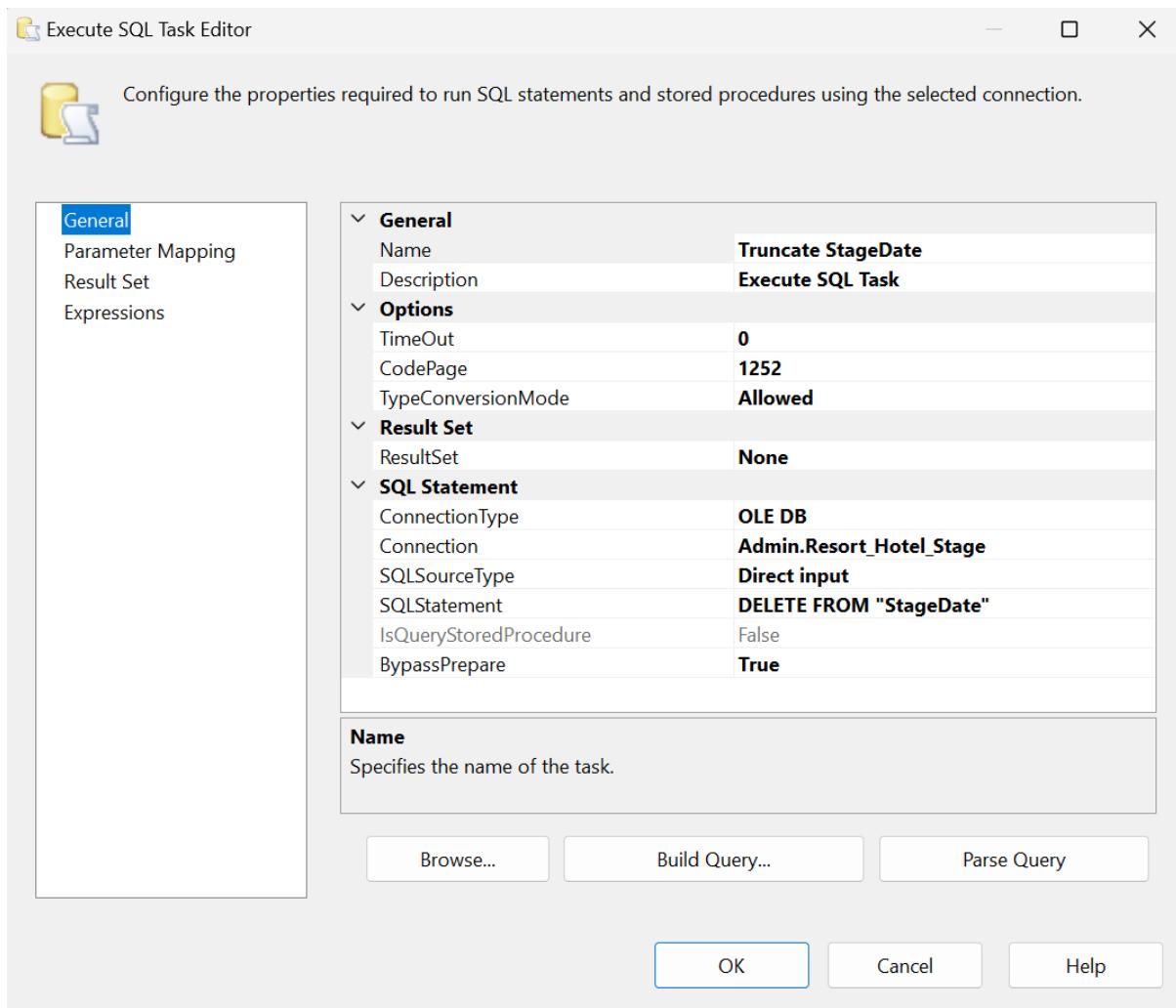
1. Truncate table cho bảng StageDate.
2. Load dữ liệu từ excel vào bảng Stage.
3. Truncate table cho bảng DimDate.
4. Load dữ liệu từ bảng StageDate và Bảng DimDate.
5. Tiếp đến là lần lược thực hiện các bước 1-4 cho việc load dữ liệu vào DimHotel và DimCustomers.



### 3.3.3. Thiết lập Truncate table cho các bảng

1. Thiết lập kết nối đến bảng cơ sở dữ liệu chứa bảng cần thực hiện. Ở đây cần thực hiện Truncate trên bảng StageDate trong cơ sở dữ liệu Resort\_Hotel\_Stage.
2. Thiết lập SQLStatement, cần thực hiện trên bảng nào thì nhập tên bảng đó vào.

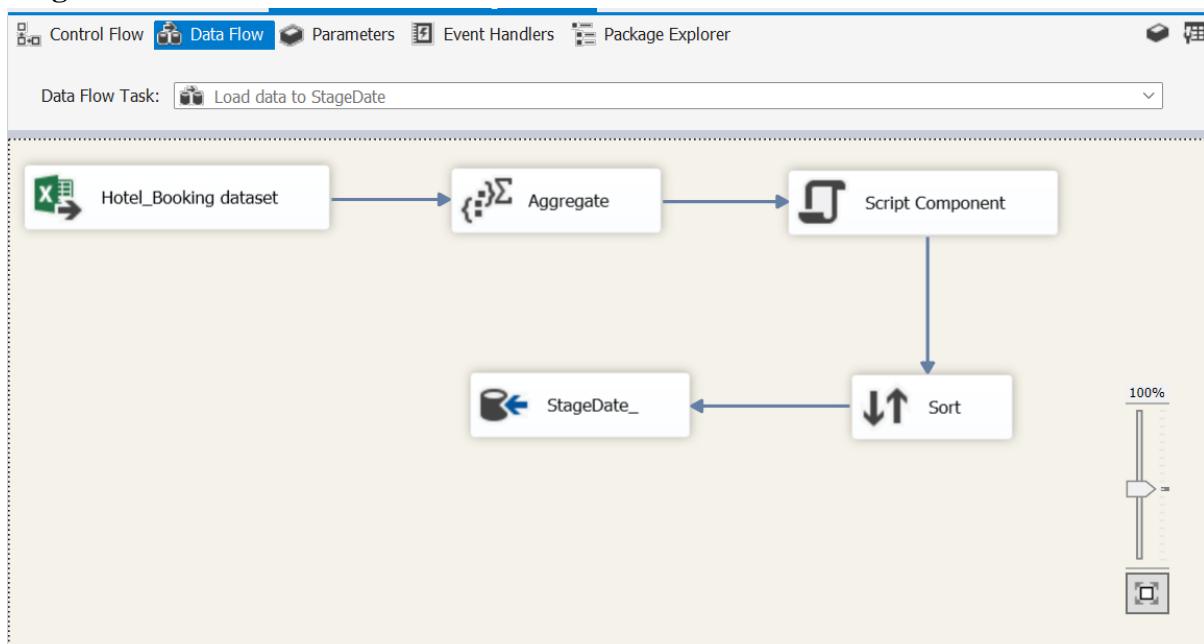




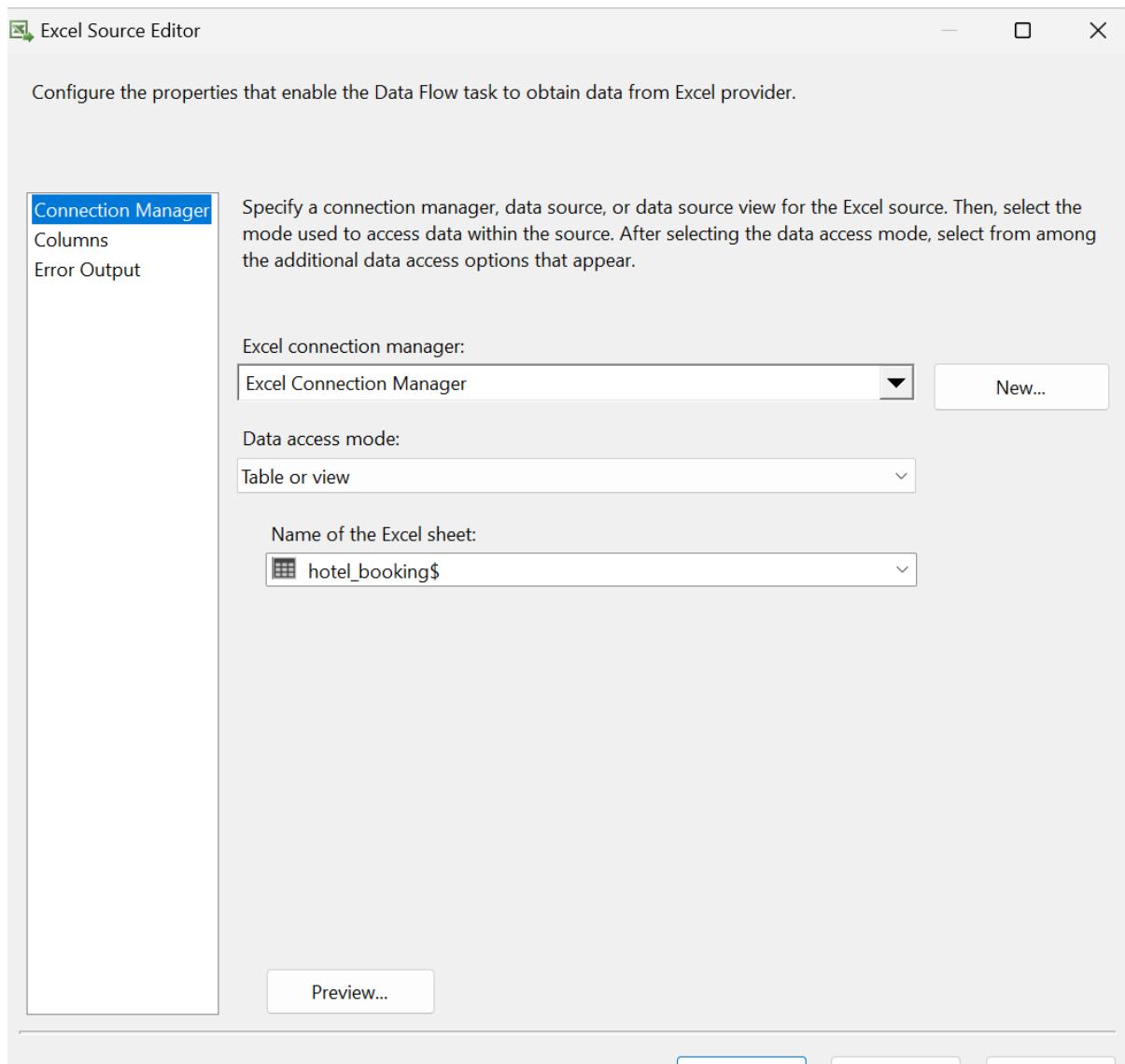
Thực hiện thiết lập cho các bảng tương tự.

### 3.3.4. Load dữ liệu từ Source vào Stage.

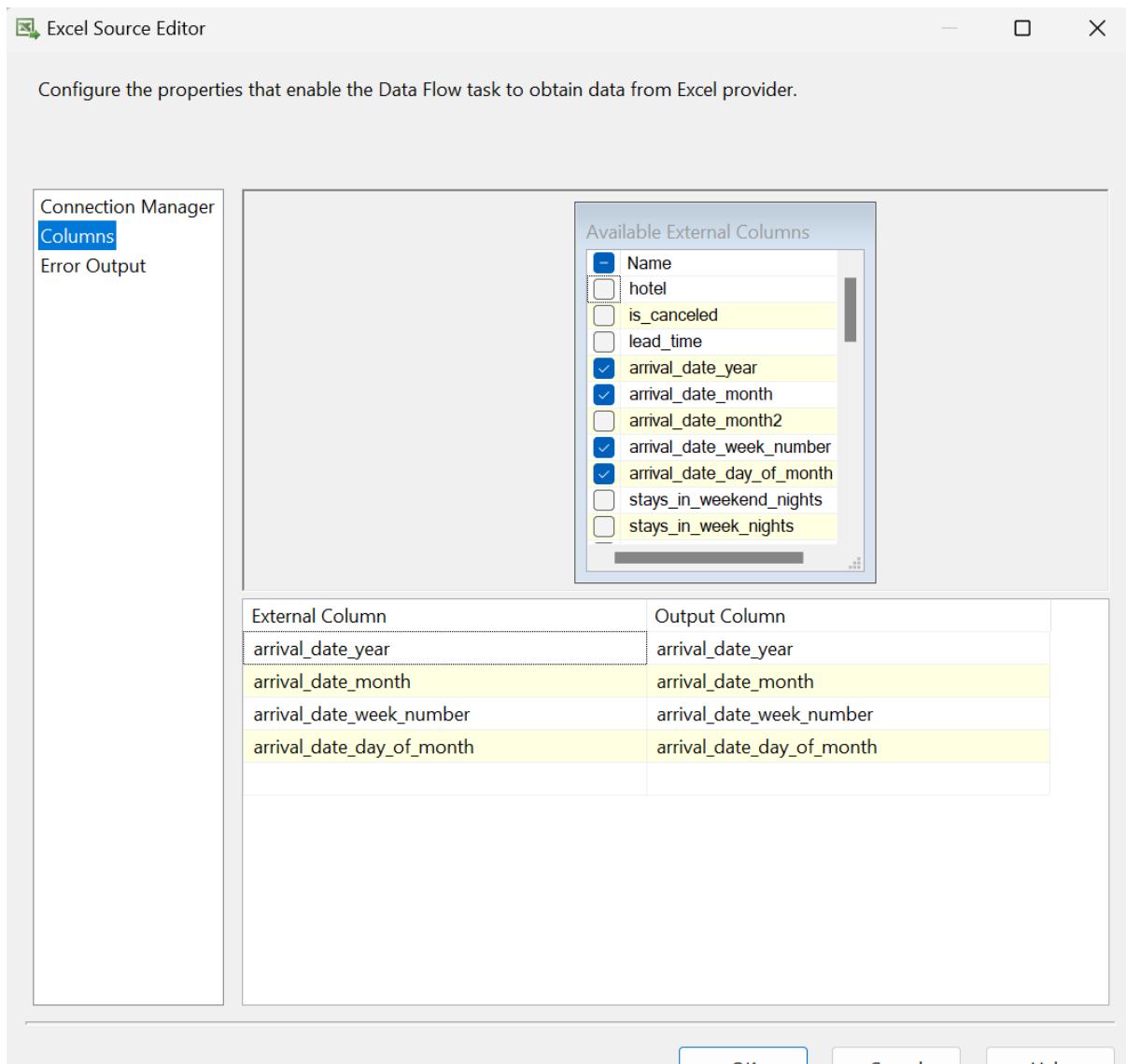
#### StageDate:



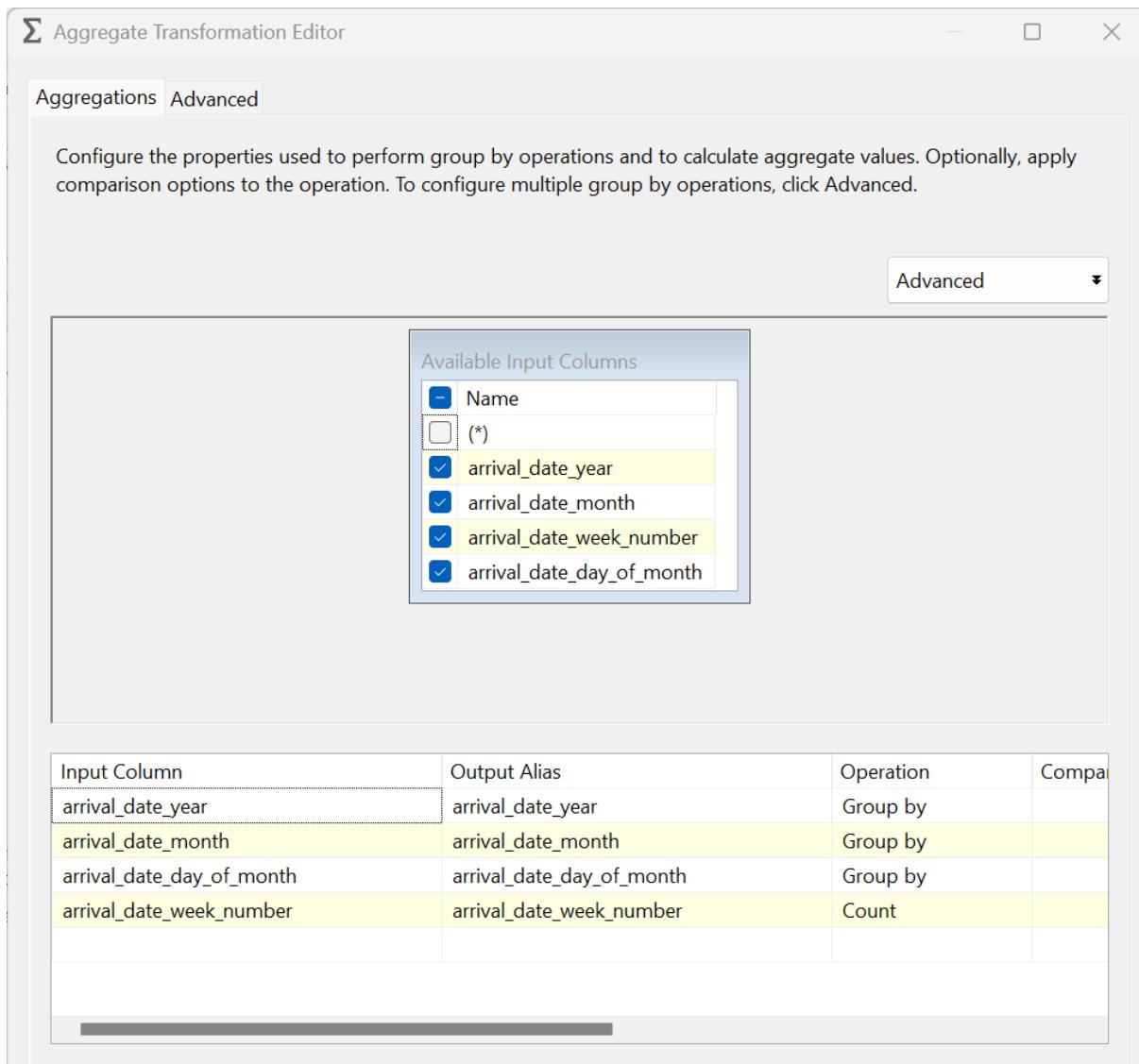
#### 1. Chọn Data Source:



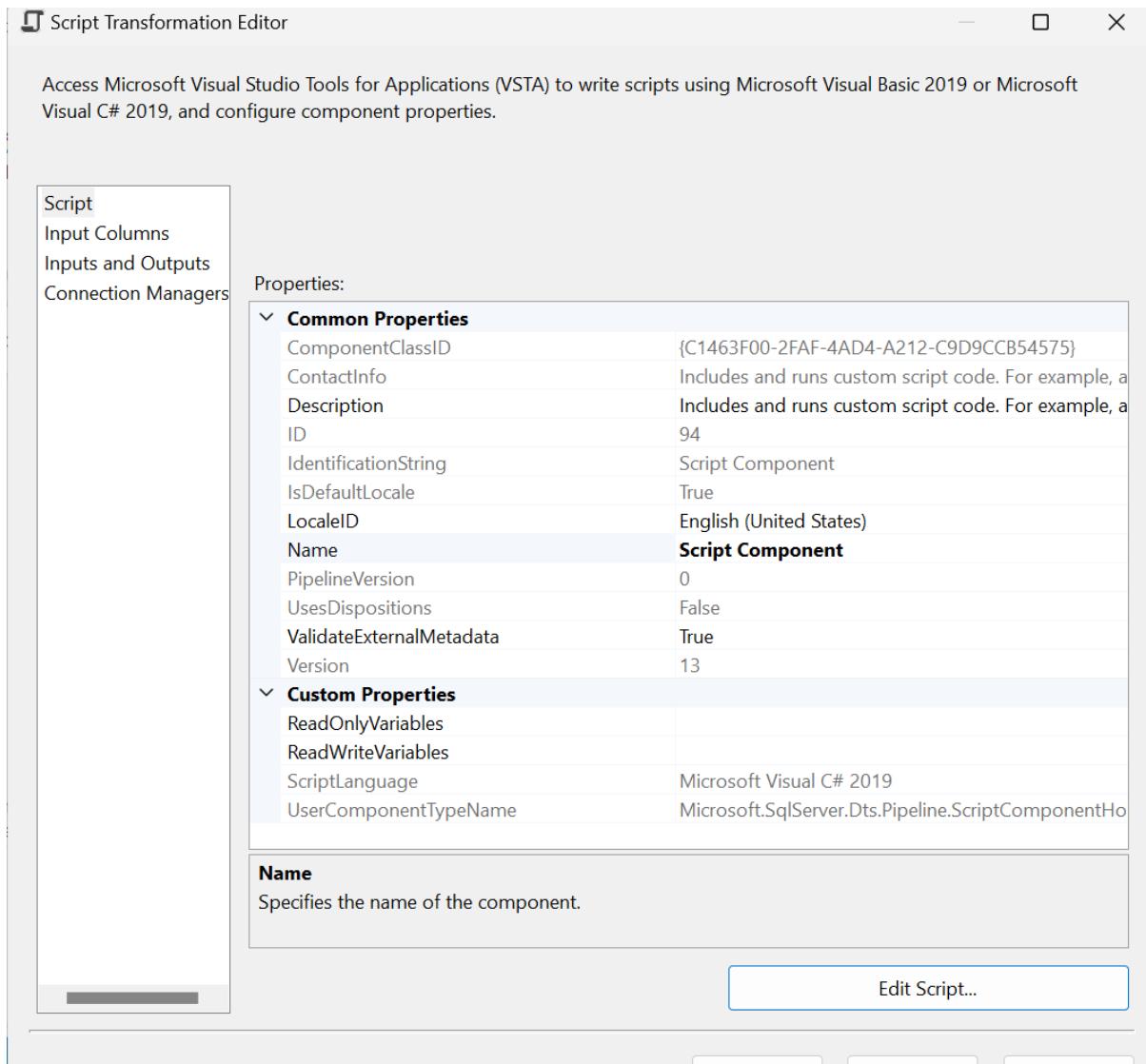
2. Chọn các trường dữ liệu cần cho bảng StageDate;



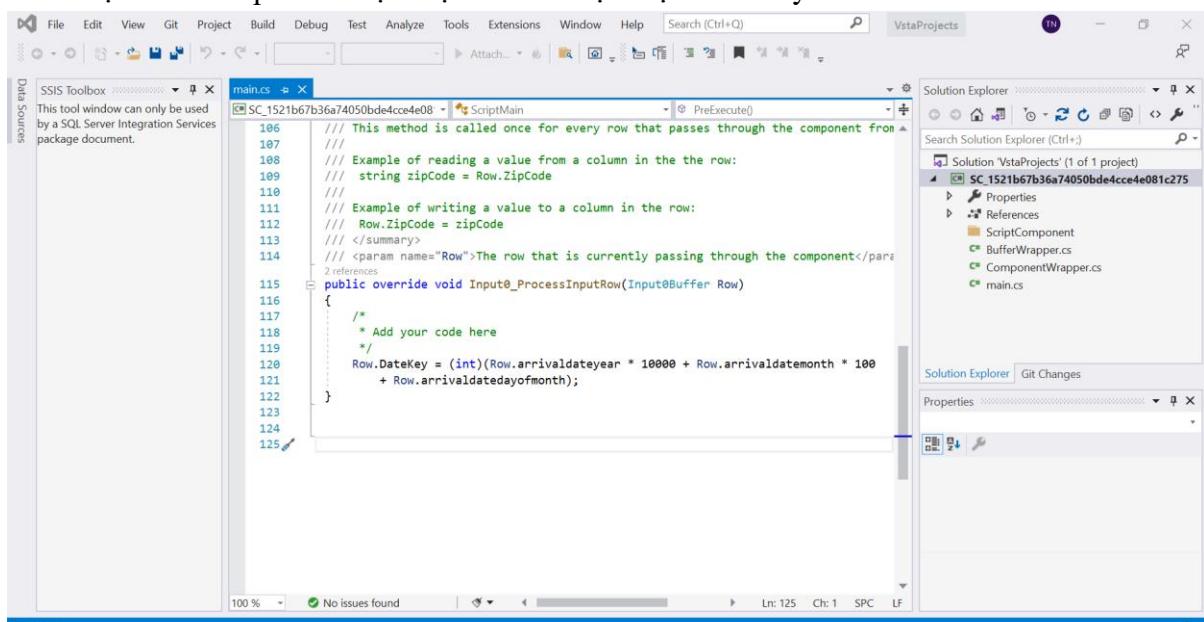
3. Thực hiện Group dữ liệu để tránh dữ liệu trùng lặp.



4. Tạo BusinessKey là DateKey:



#### 4.1. Chọn Edit Script... để thực hiện viết câu lệnh tạo DataKey.



#### 5. Sắp xếp dữ liệu.

Sort Transformation Editor

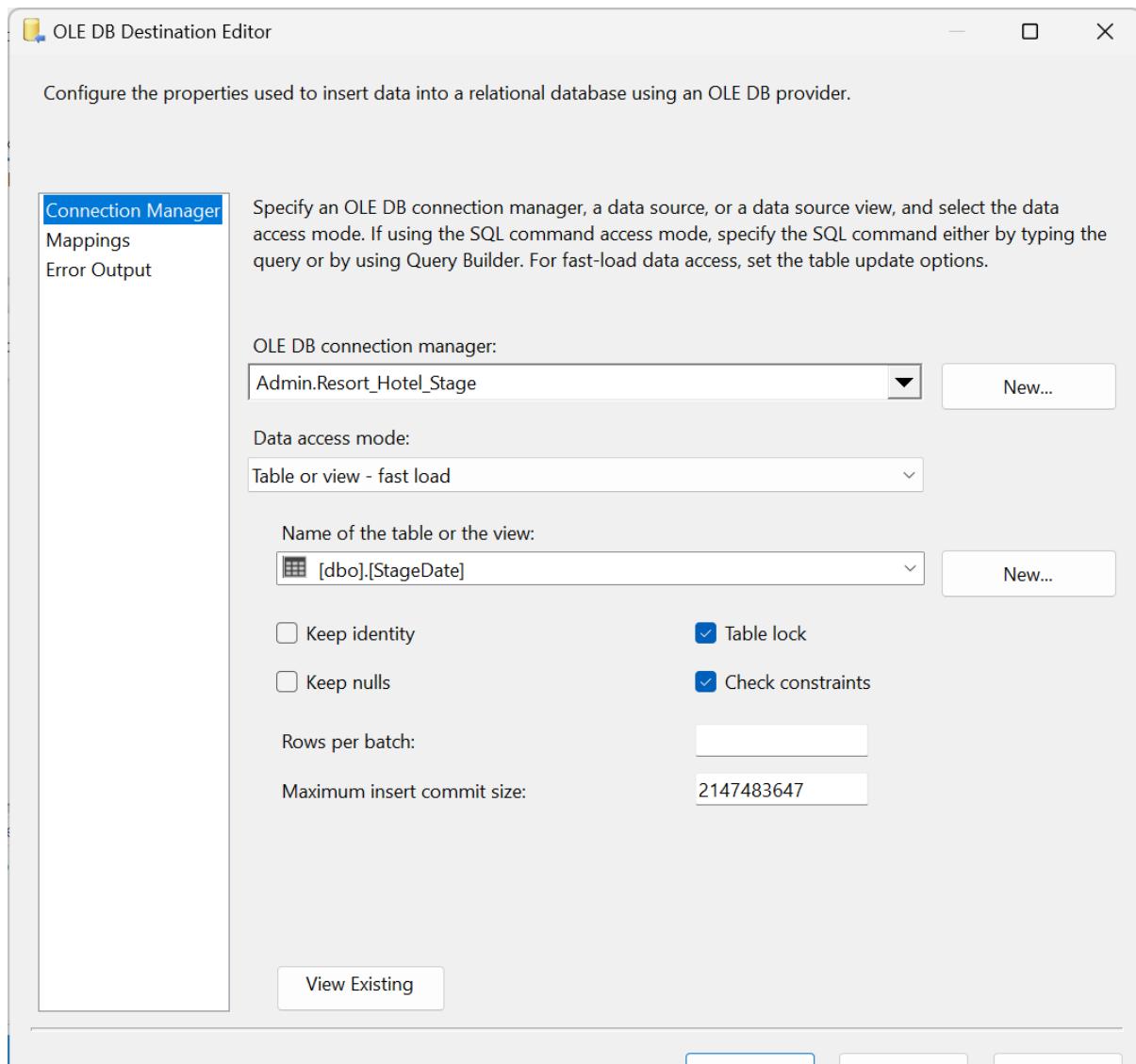
Specify the columns to sort, and set their sort type and their sort order. All nonselected columns are copied unchanged.

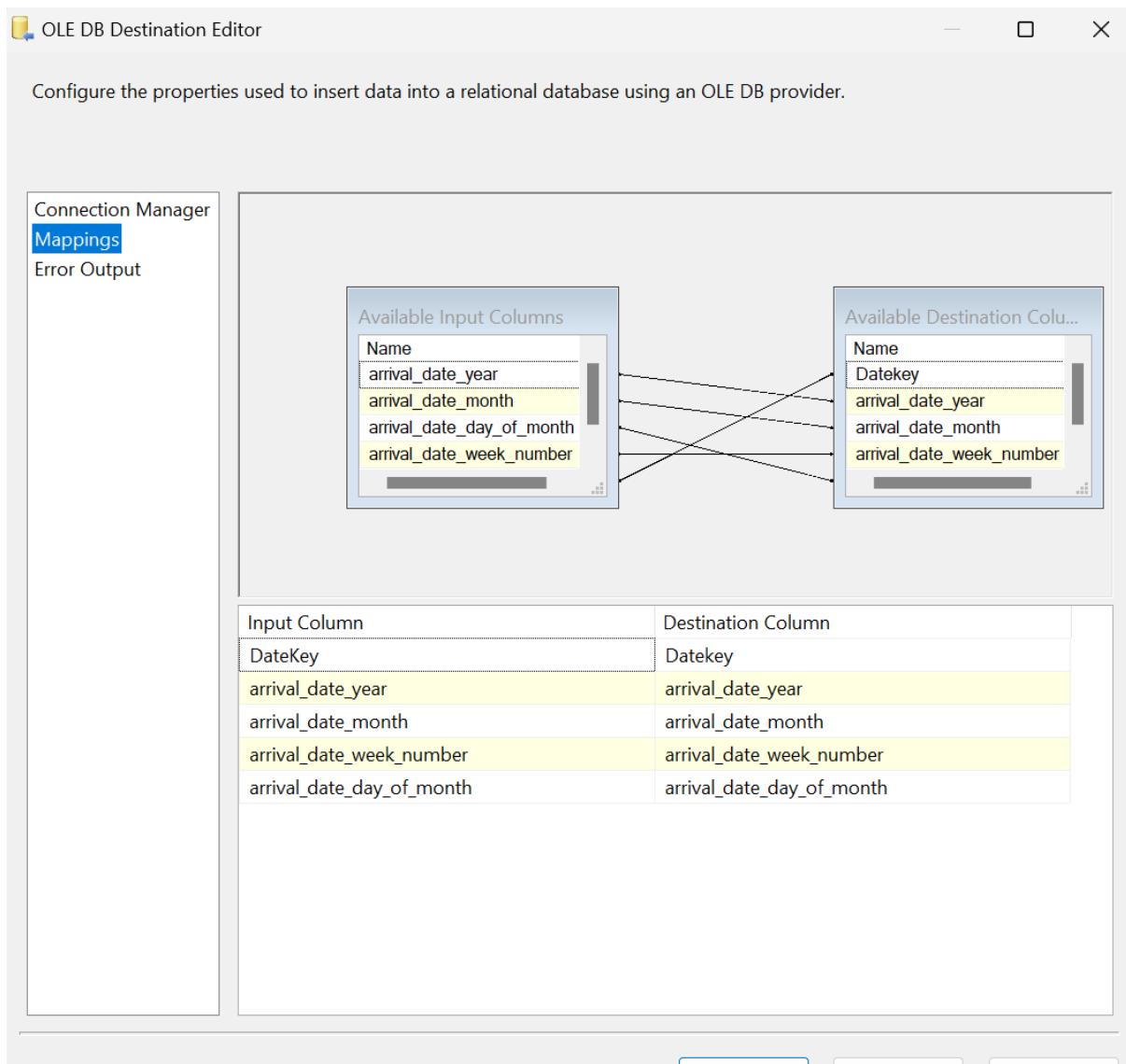
Available Input Columns

Name	Pass T...
arrival_date_year	-
arrival_date_month	-
arrival_date_day_of_month	-
arrival_date_week_number	✓

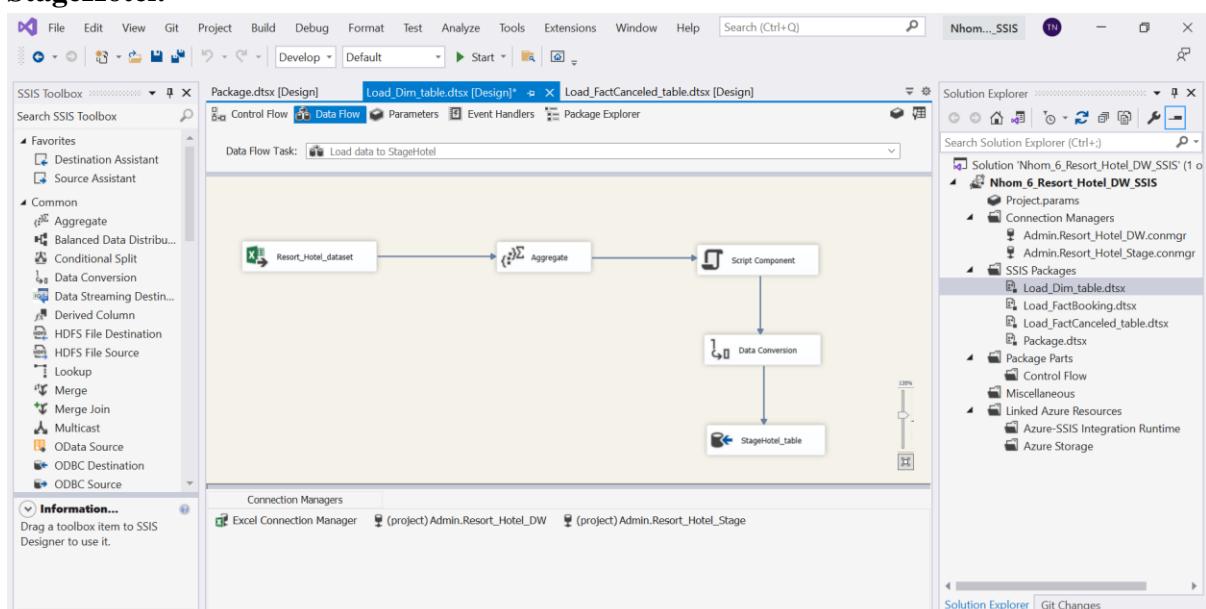
Input Column	Output Alias	Sort Type	Sort Order	Conn...
arrival_date_year	arrival_date_year	ascending	3	
arrival_date_month	arrival_date_month	ascending	2	
arrival_date_day_of_month	arrival_date_day_of_month	ascending	1	

6. Load dữ liệu vào StageDate.

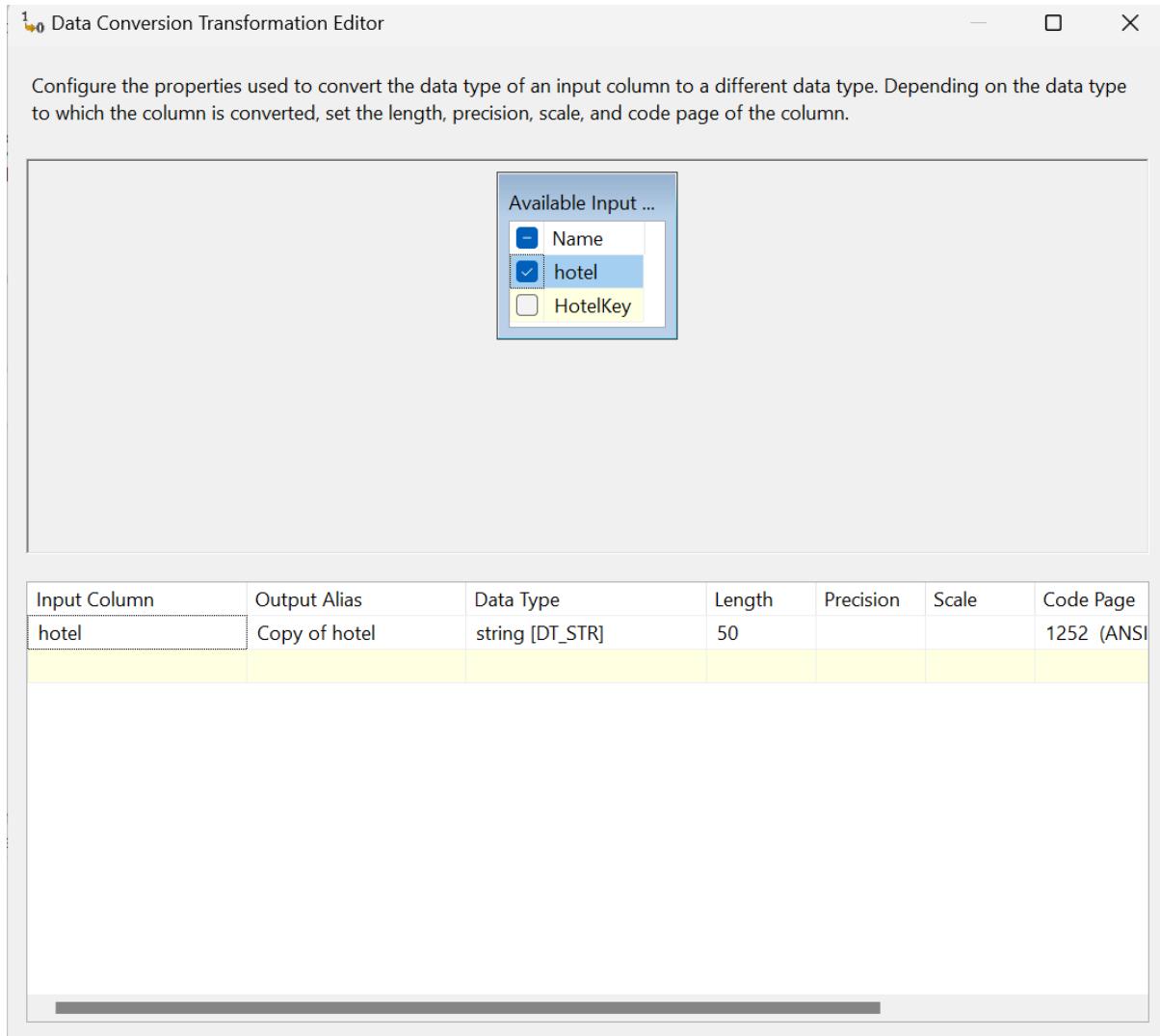




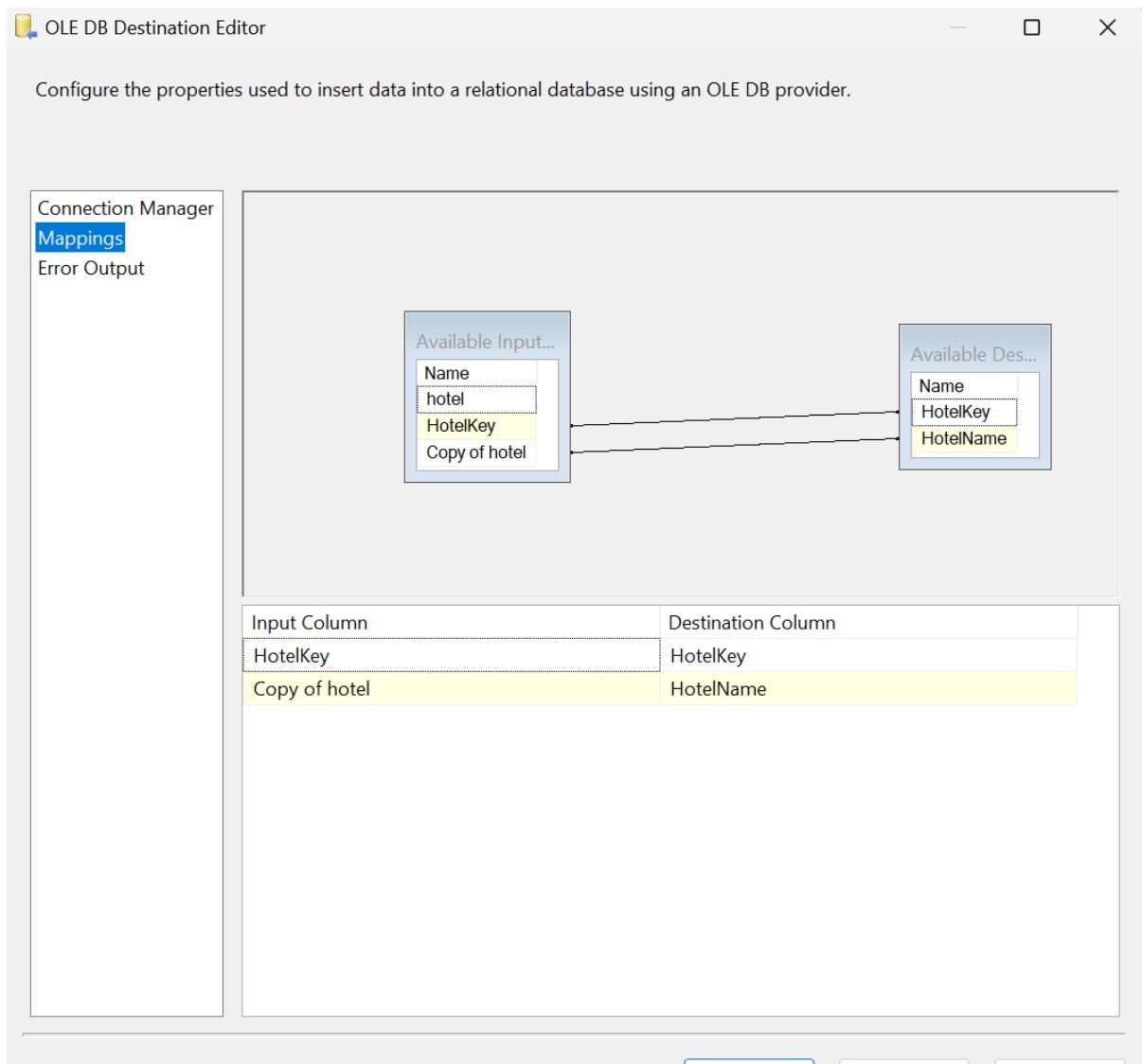
## StageHotel:



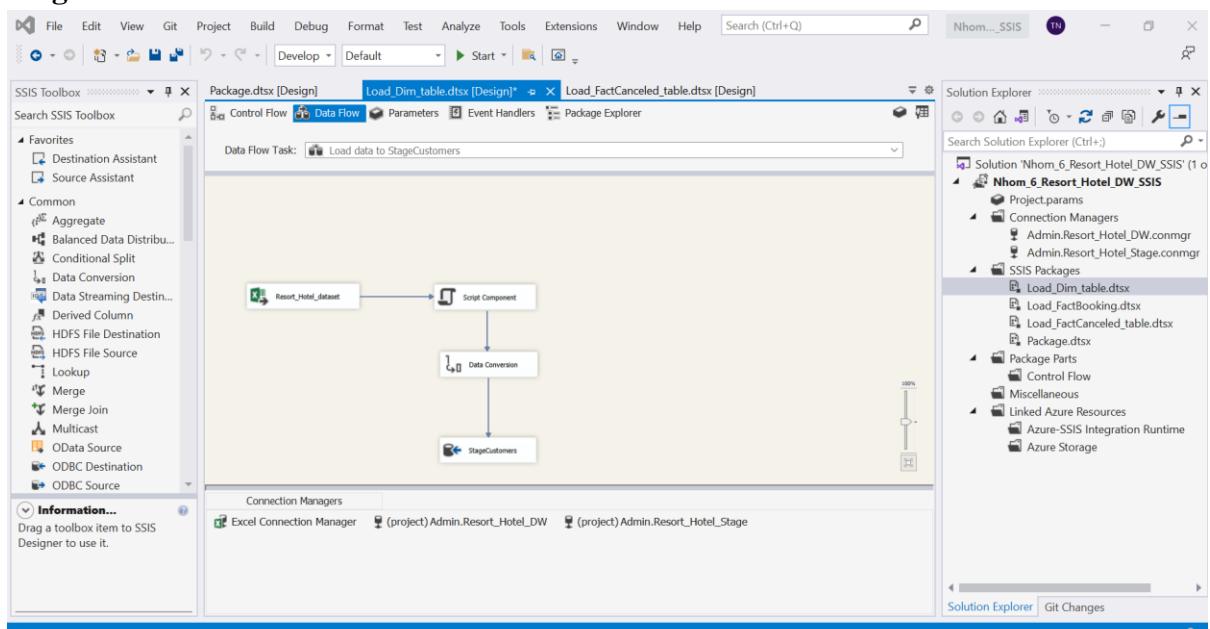
Các bước thực hiện tung tự như StageDate, tuy nhiên ở StageHotel cần ép dữ liệu của HotelName của excel từ nchar sang String cho phù hợp với kiểu dữ liệu của HotelName trong StageHotel.



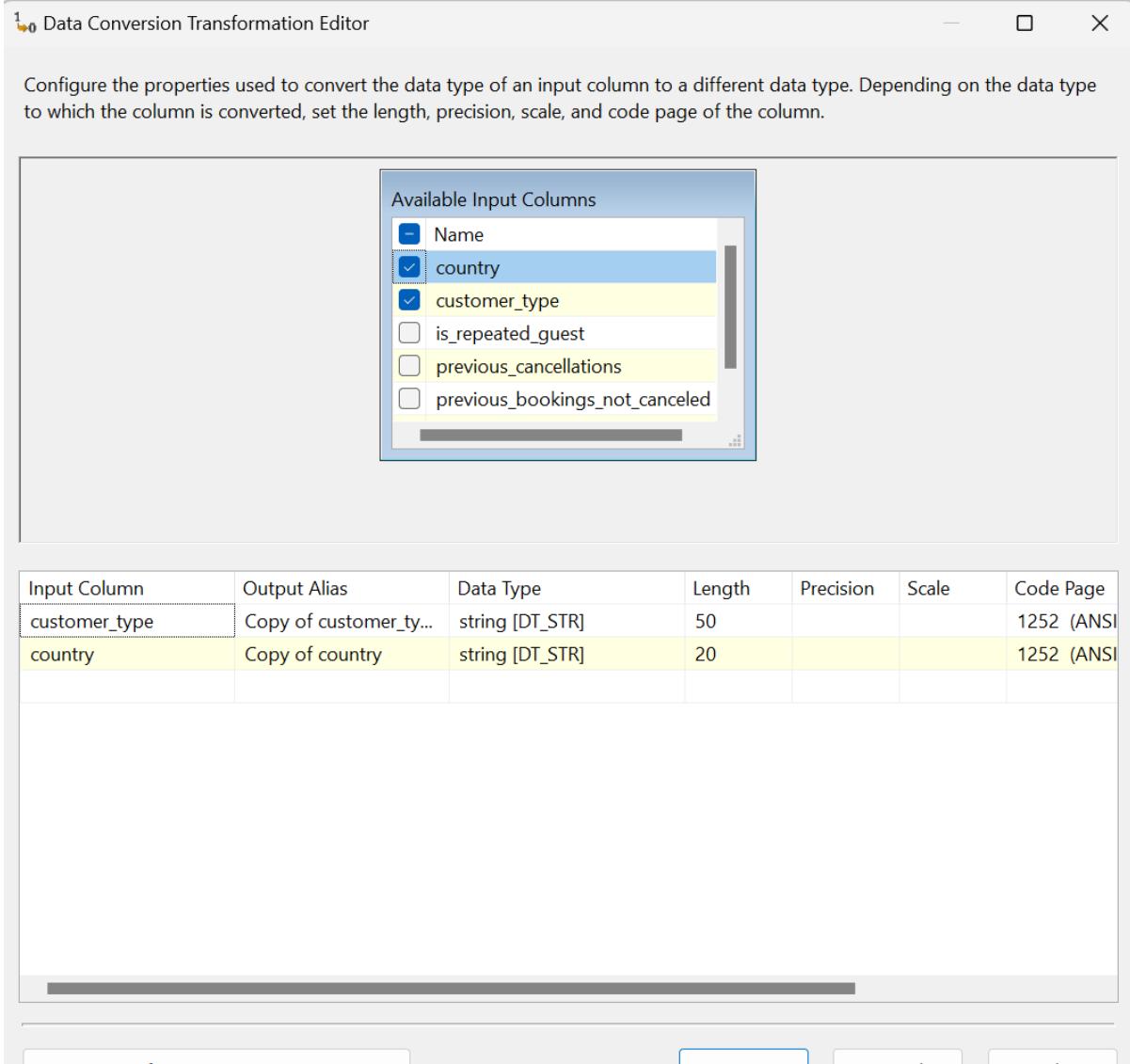
Load dữ liệu vào StageHotel.



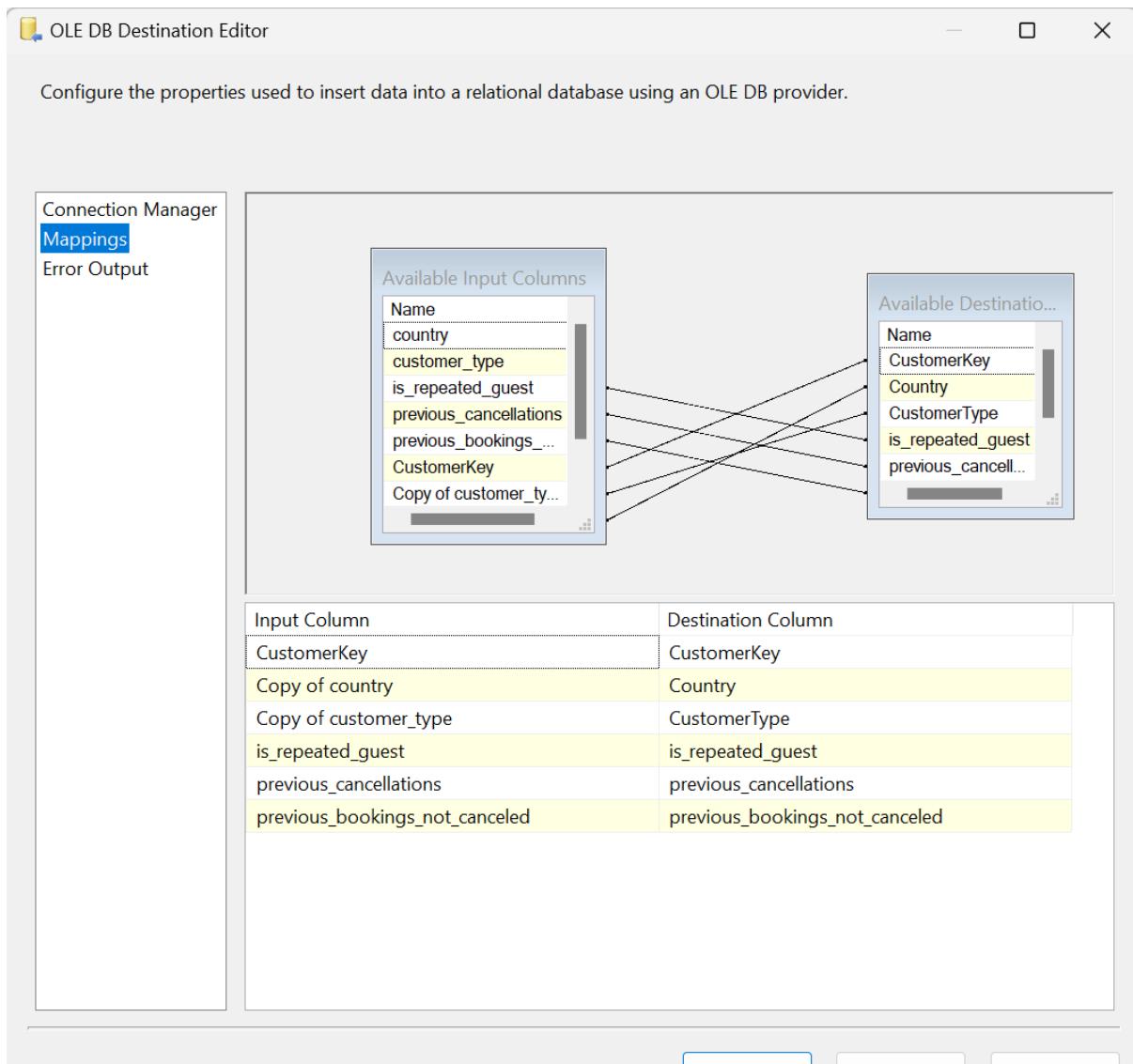
## StageCustomers:



Ép dữ liệu .

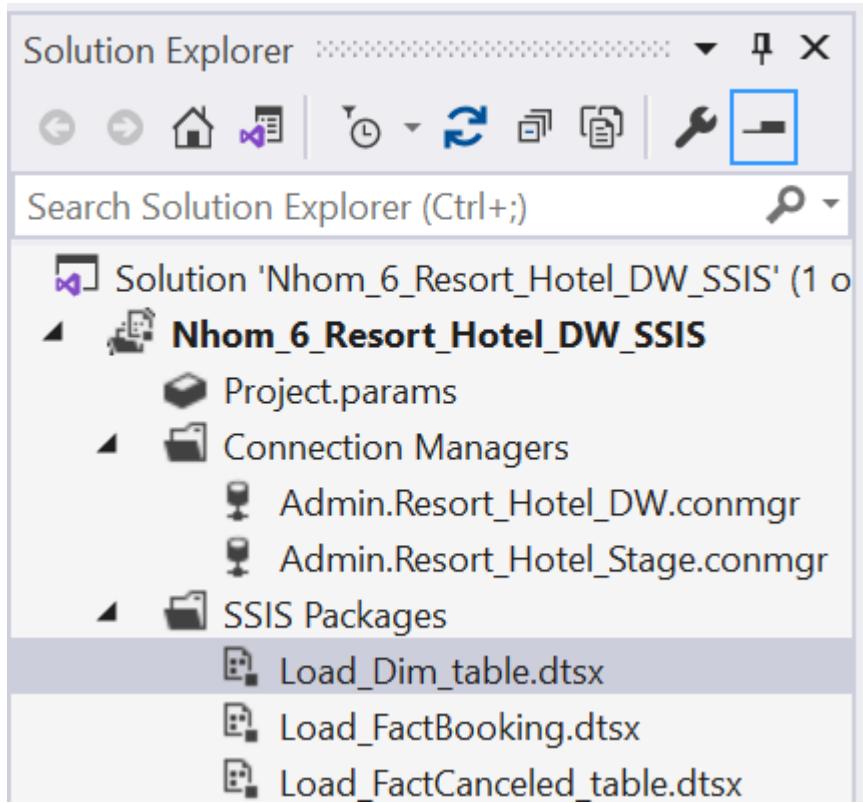


Load dữ liệu vào StageCustomers.



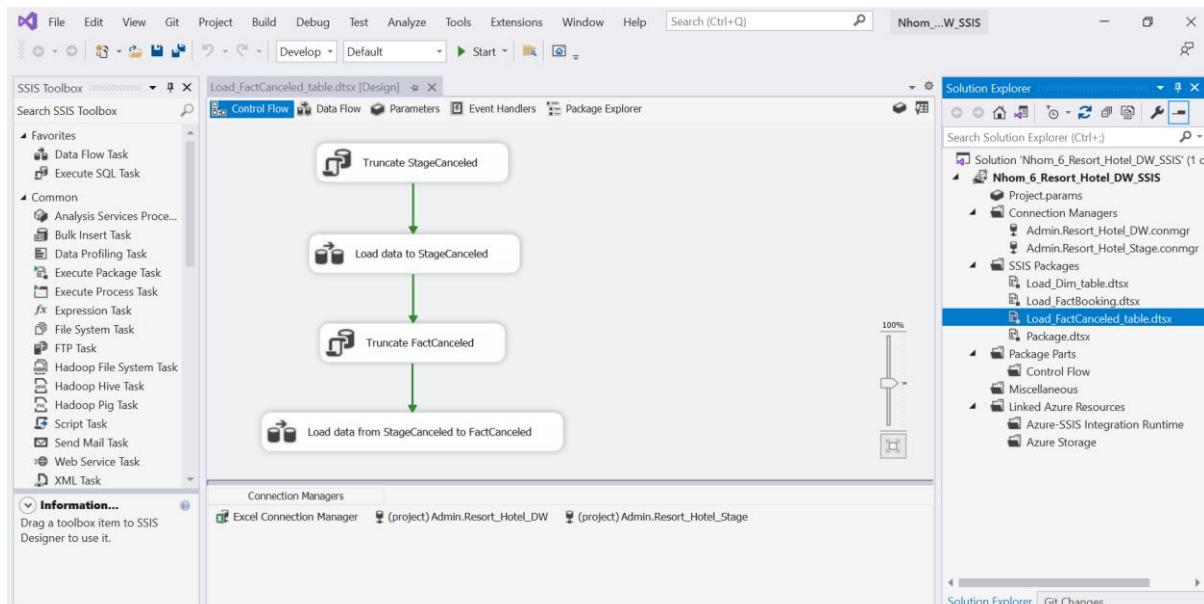
## 3.4. Import dữ liệu vào bảng các bảng fact

### 3.4.1. Tạo Package

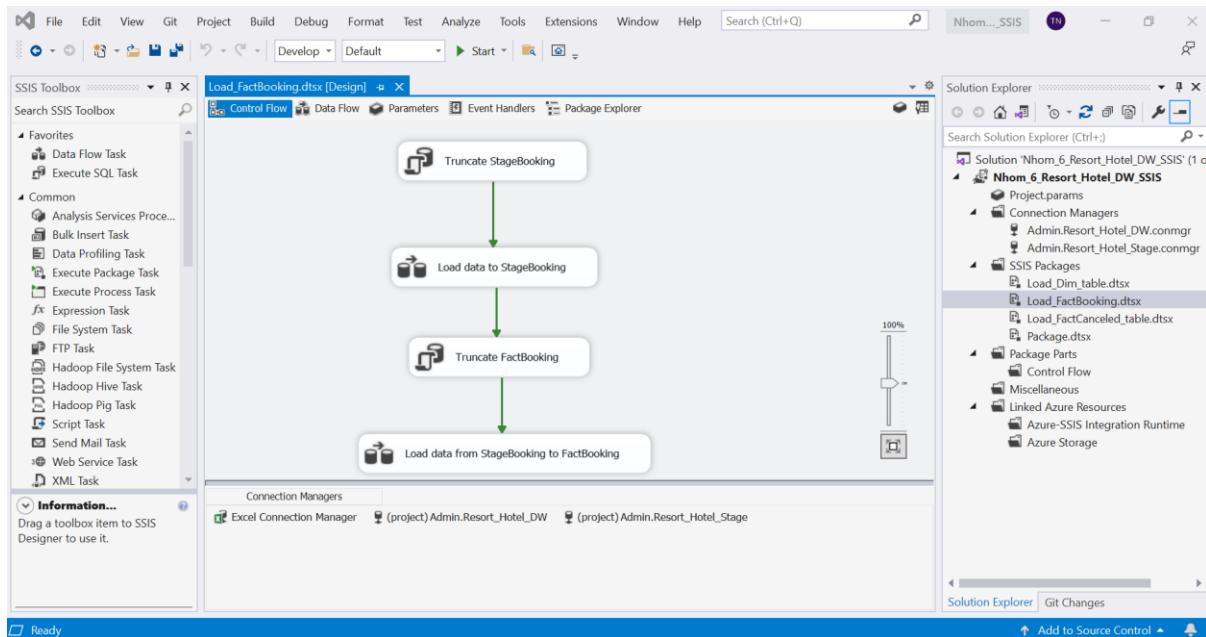


### 3.4.2. Thiết lập Control Flow

#### FactCanceled.

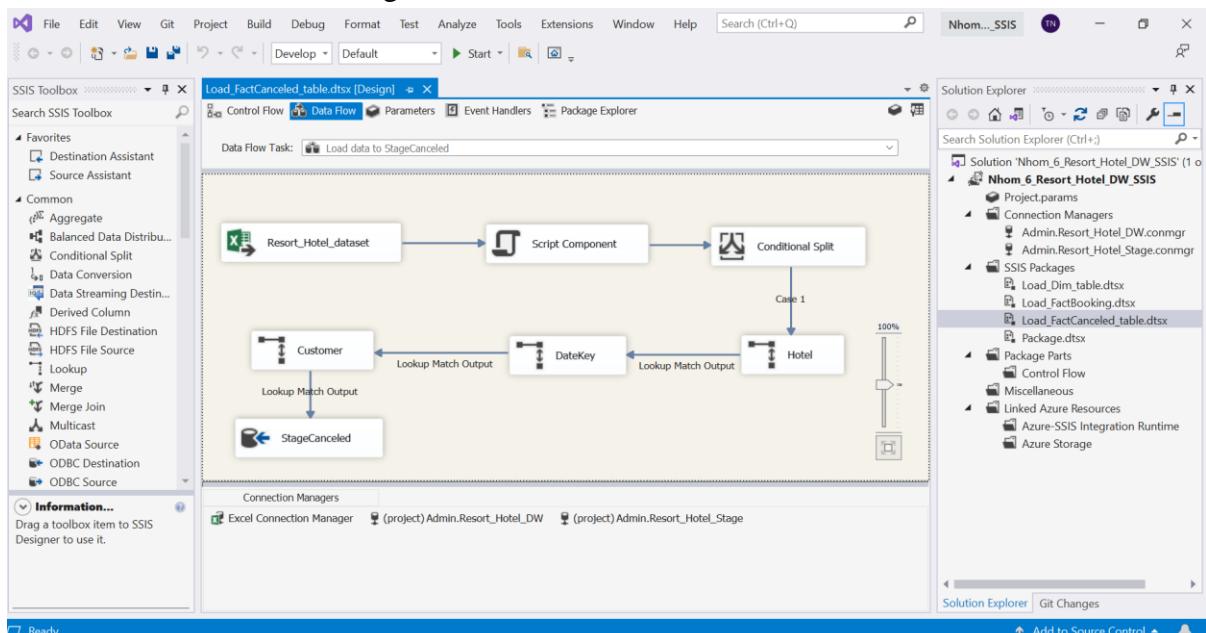


#### FactBooking.

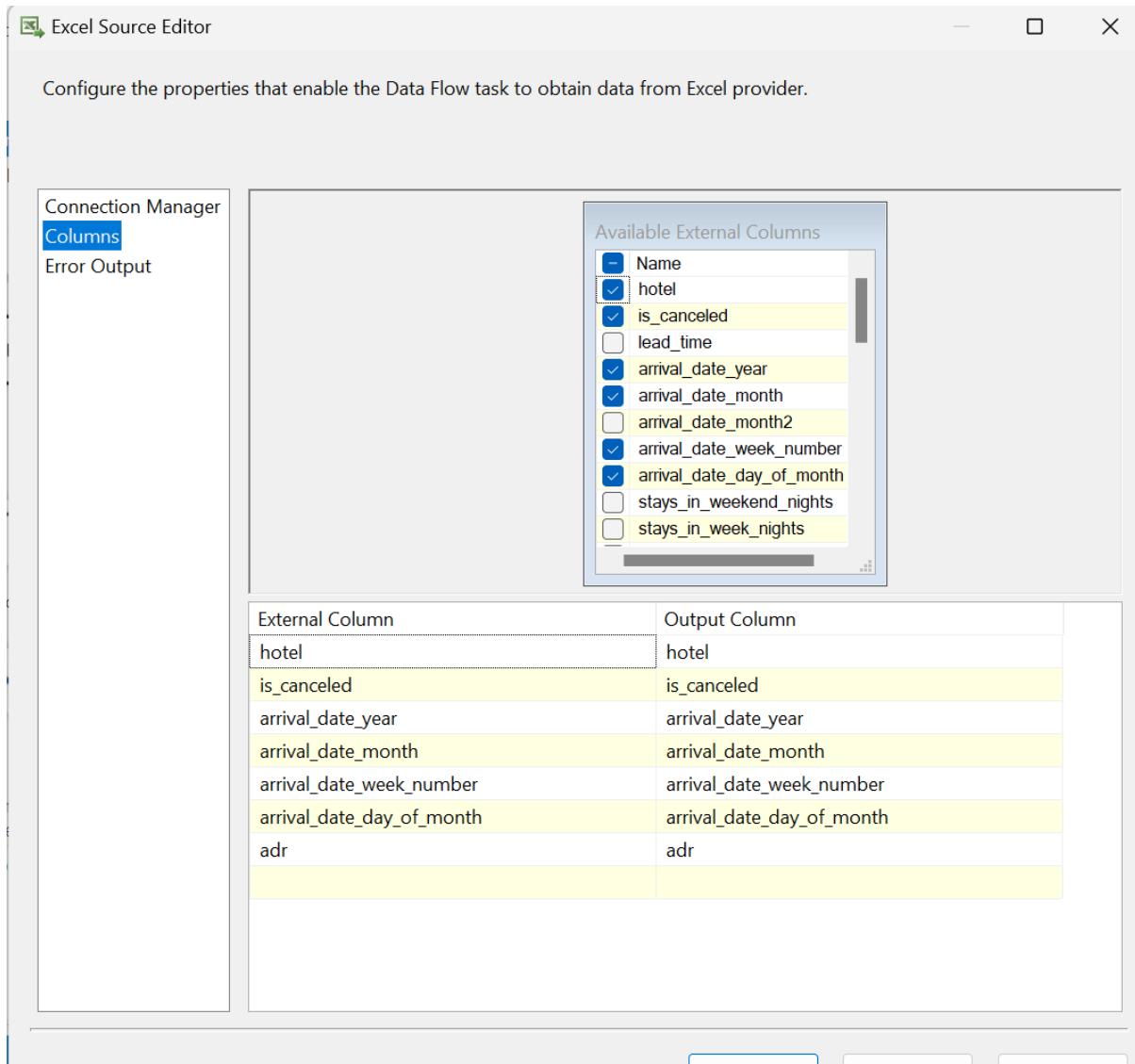


### 3.4.3. Load dữ liệu từ excel vào Stage và từ Stage vào Fact FactCanceled.

Load dữ liệu từ excel vào StageCanceled.



Chọn trường dữ liệu cần thiết.



Thực hiện tạo businessKey.

```

main.cs
=====
109     /// string zipCode = Row.ZipCode
110
111     /// Example of writing a value to a column in the row:
112     /// Row.ZipCode = zipCode
113     /// </summary>
114     /// <param name="Row">The row that is currently passing through the component</param>
115     int count = 0;
116
117     public override void Input0_ProcessInputRow(Input0Buffer Row)
118     {
119         /*
120         * Add your code here
121         */
122         count++;
123         Row.NumberKey = count;
124         Row.DateKey = (int)(Row.arrivaldateyear * 10000 +
125                             Row.arrivaldatemonth*100 + Row.arrivedatedayofmonth);
126         if(Row.hotel == "Resort Hotel")
127         {
128             Row.HotelKey = 1;
129         }
130         else
131         {
132             Row.HotelKey = 2;
133         }
134     }
135
136

```

Lọc dữ liệu theo điều kiện is-cancelled = 1.

Conditional Split Transformation Editor

Specify the conditions used to direct input rows to specific outputs. If an input row matches no condition, the row is directed to a default output.

Order	Output Name	Condition
1	Case 1	is_canceled == 1

Variables and Parameters

Mathematical Functions

String Functions

Date/Time Functions

NULL Functions

Type Casts

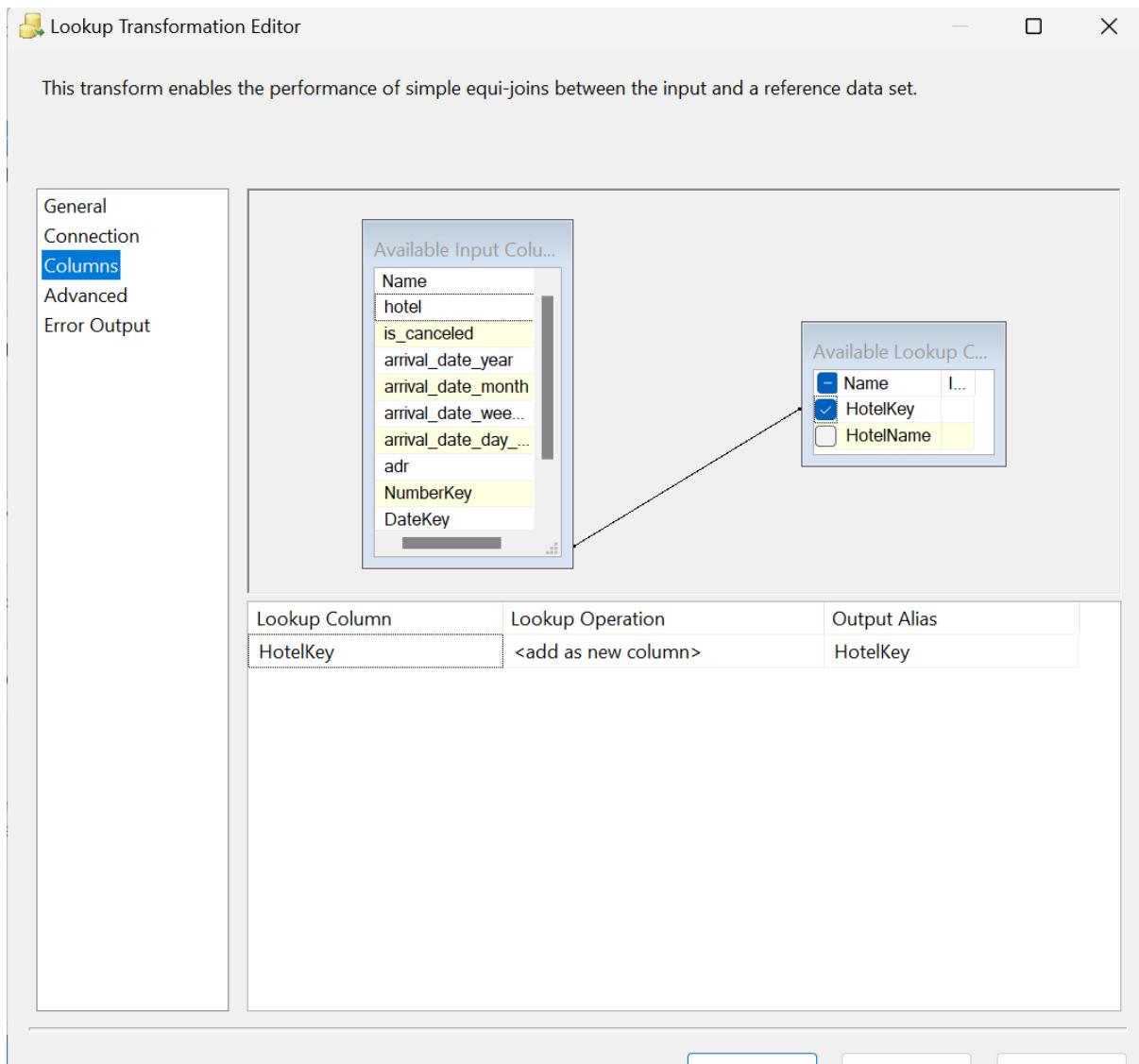
Operators

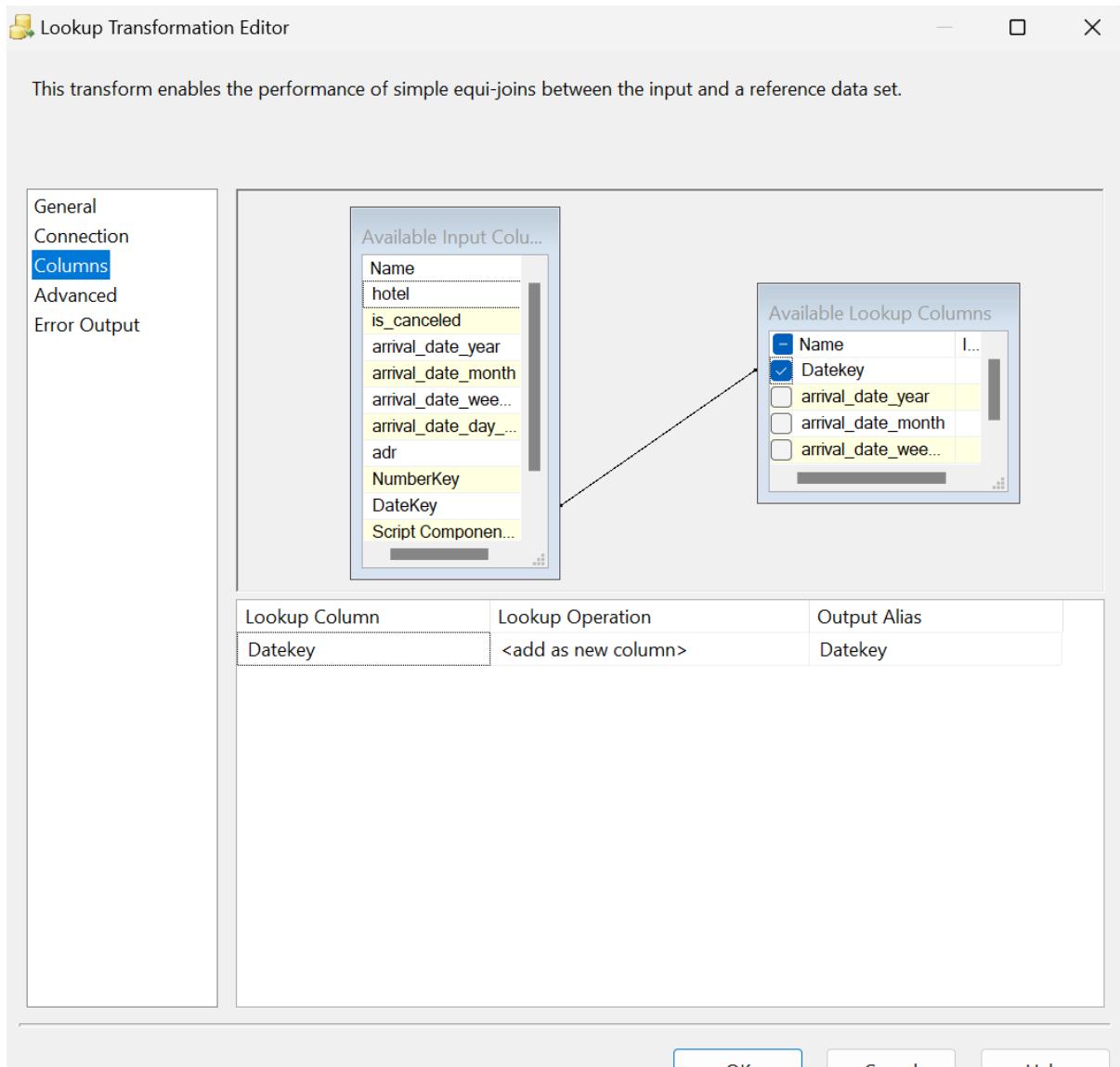
Description:

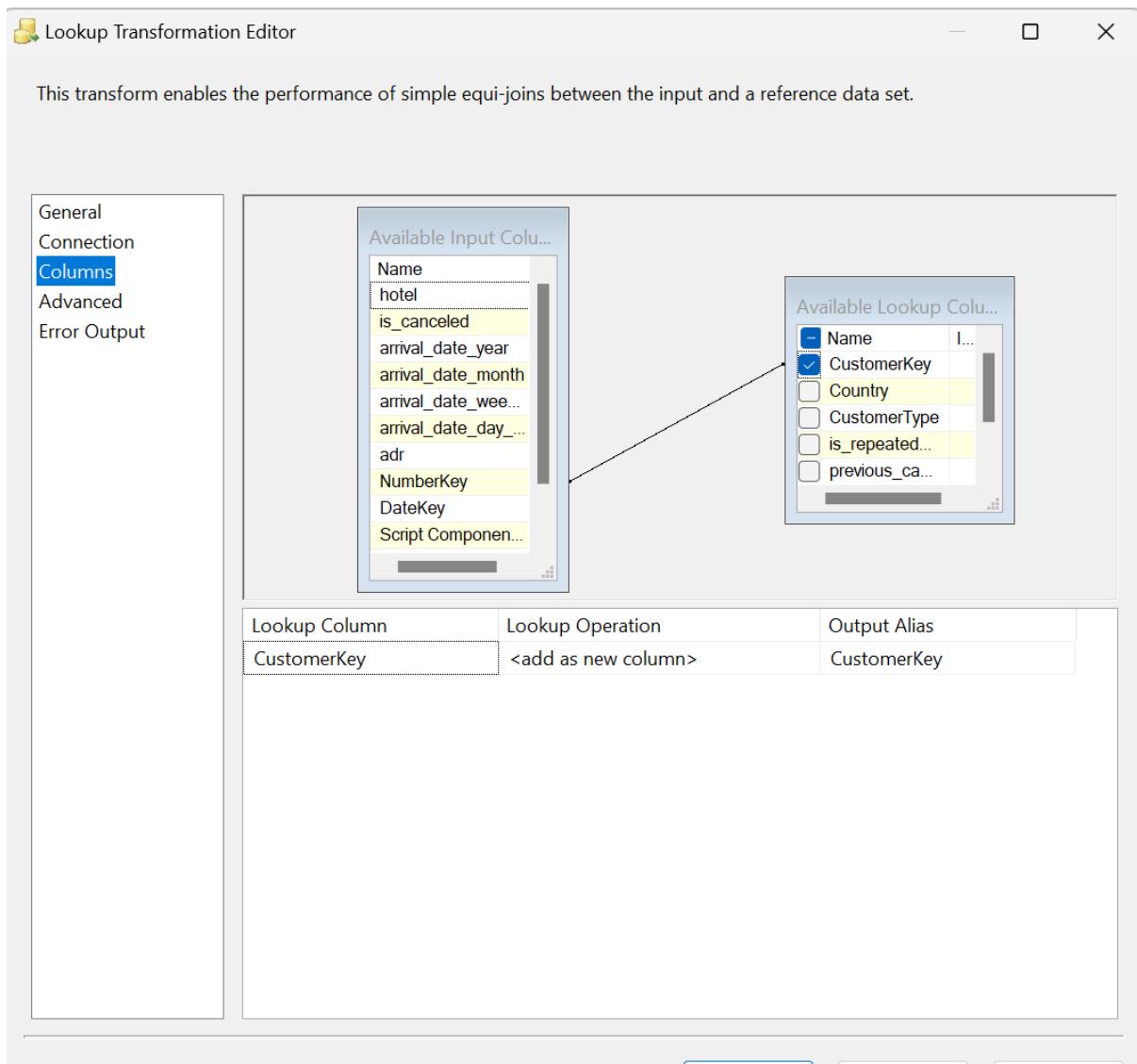
Default output name: Conditional Split Default Output

OK Cancel Help

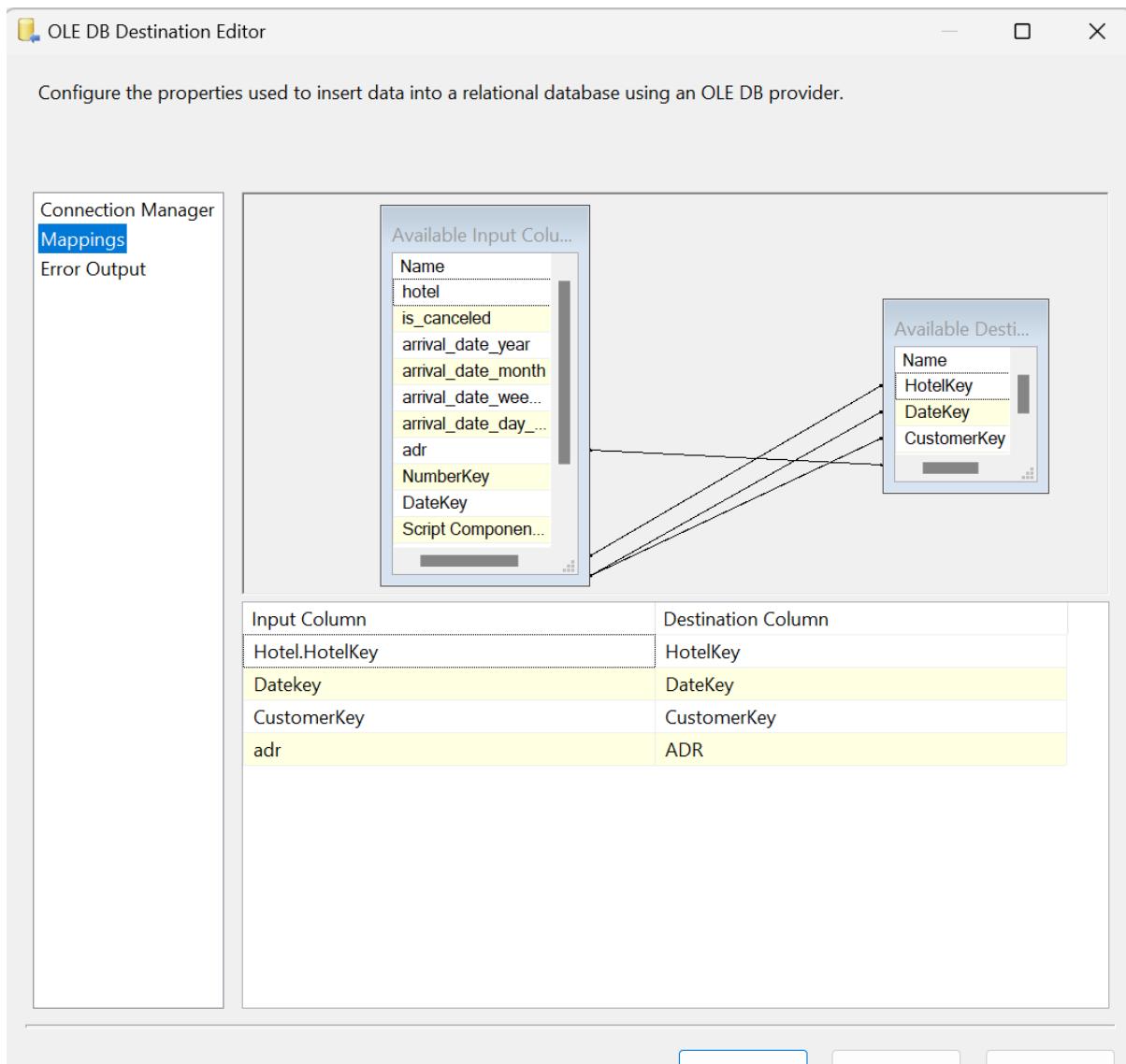
Lookup với các bảng StageHotel, StageDatekey, StageCustomers để tạo liên kết thông qua các businessKey.



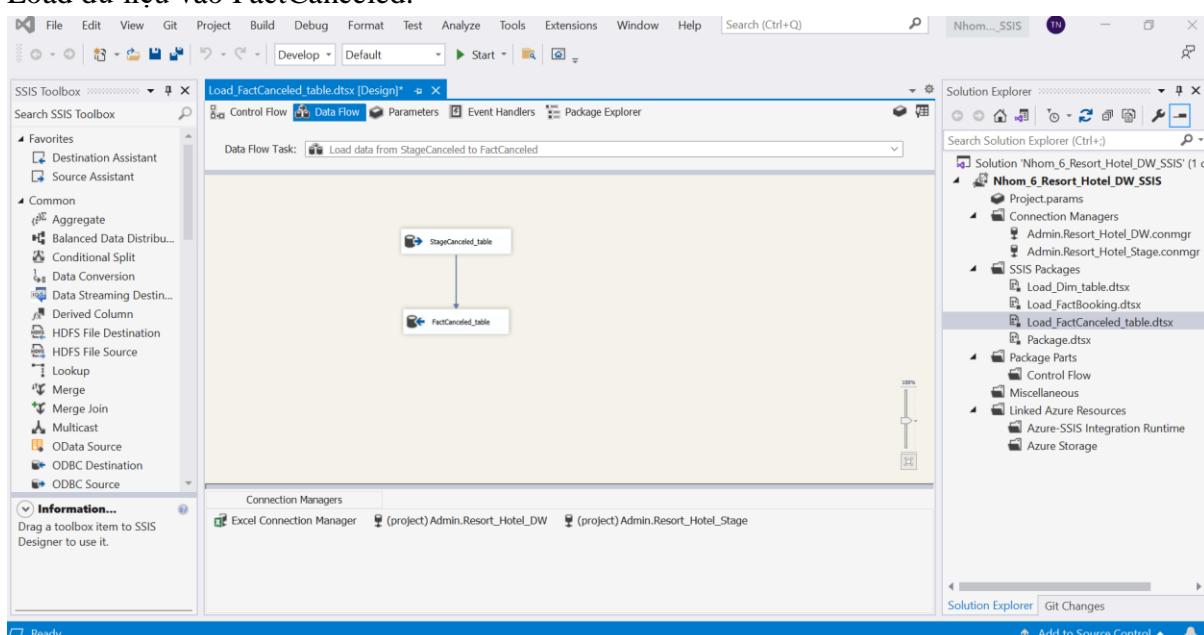




Load dữ liệu vào StageCanceled.



Load dữ liệu vào FactCanceled.



Kết quả:

100 %

Results Messages

	HotelKey	DateKey	CustomerKey	ADR
1	1	20150701	9	82
2	1	20150701	10	105.5
3	1	20150701	11	123
4	1	20150701	28	107
5	1	20150701	33	108.3
6	1	20150702	35	108.8
7	1	20150702	36	108.8
8	1	20150702	39	108.8
9	1	20150702	41	117.81
10	1	20150702	46	153
11	1	20150702	65	136.33
12	1	20150703	69	110.3
13	1	20150703	70	82
14	1	20150703	71	97
15	1	20150703	72	73.8

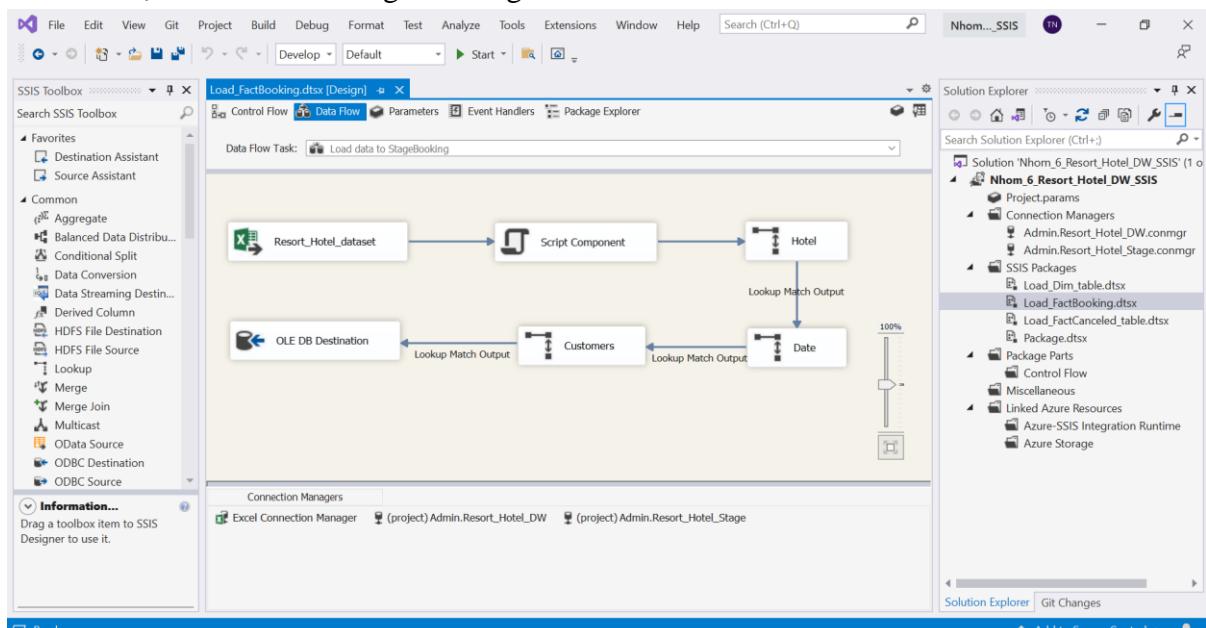
Query executed successfully.

ADMIN (16.0 RTM) ADMIN/Admin (55) Resort\_Hotel\_DW 00:00:00 44,224 rows

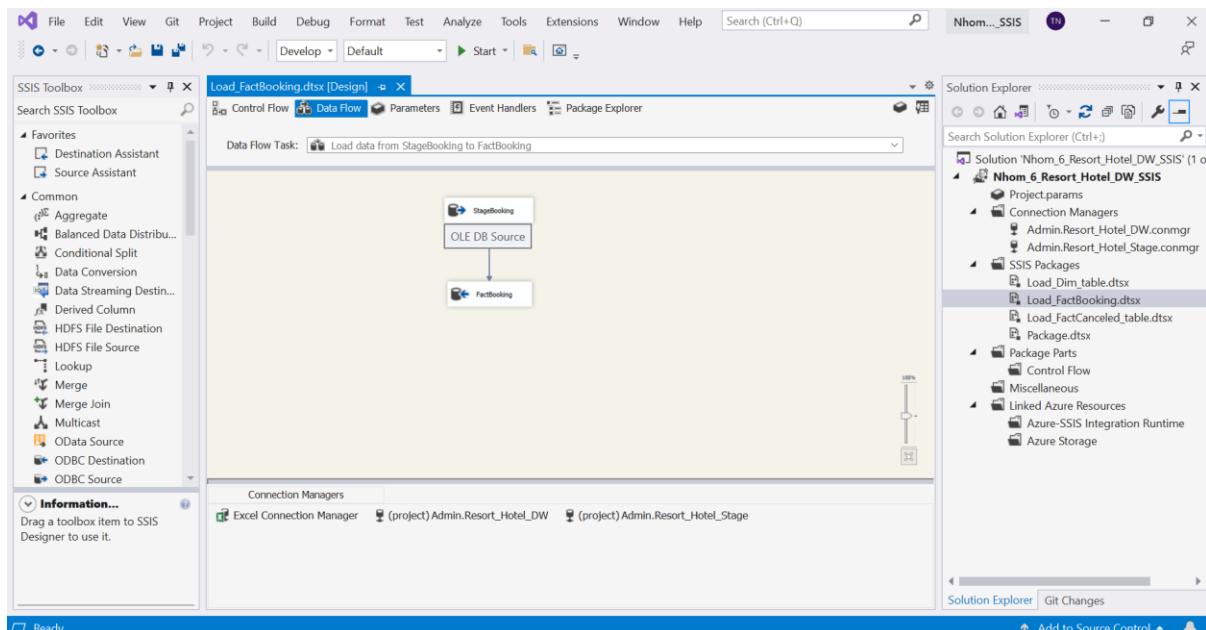
## FactBooking.

Tương tự như FactCanceled.

Load dữ liệu từ excel vào StageBooking



Load dữ liệu từ StageBooking vào DimBooking



Kết quả:

The screenshot shows the results of a database query. The message bar at the top indicates 'Query executed successfully.' and shows statistics: ADMIN (16.0 RTM) | ADMIN\Admin (75) | Resort\_Hotel\_DW | 00:00:01 | 119,390 rows. The results grid displays data from the FactBooking table, with columns including HotelKey, DateKey, CustomerKey, meal, market\_segment, agent, reserved\_room\_type, deposit\_type, required\_car\_parking\_spaces, and total\_of\_special\_requests. The data shows various booking details for different dates and customers.

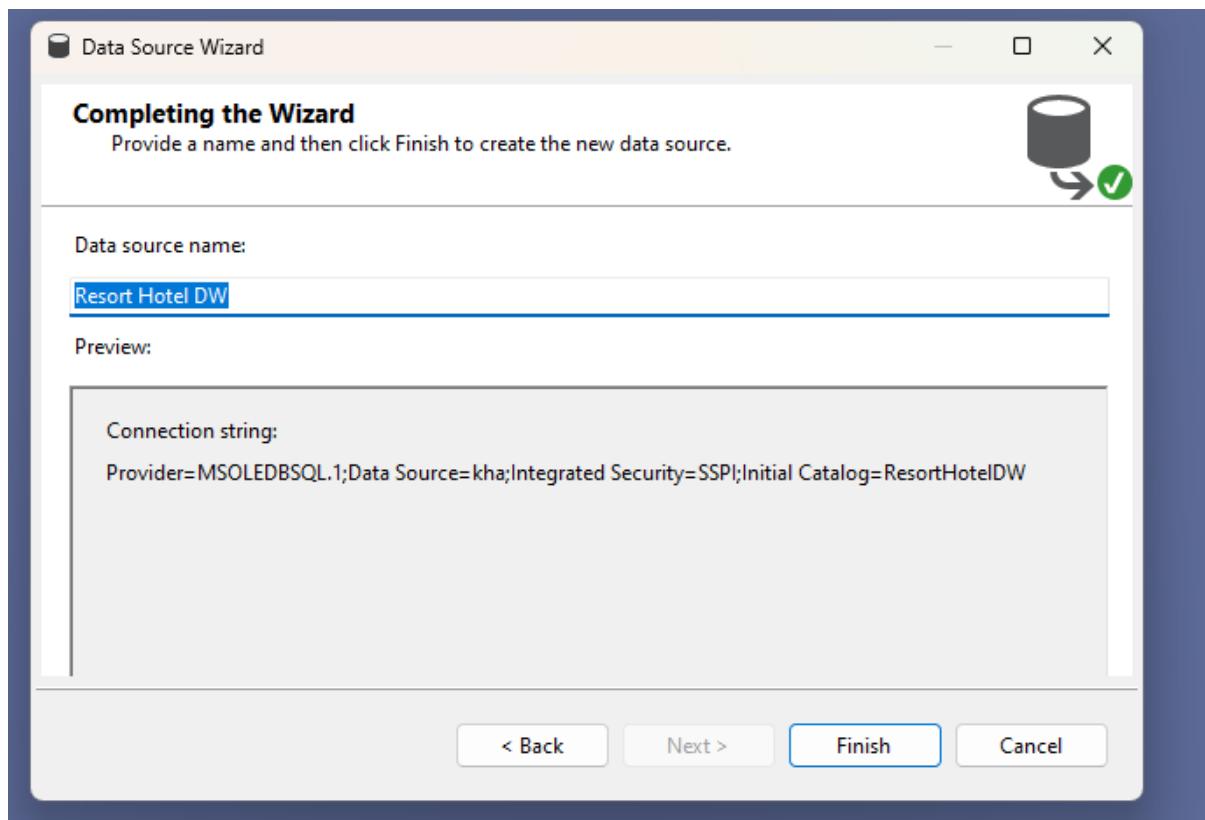
	HotelKey	DateKey	CustomerKey	meal	market_segment	agent	reserved_room_type	deposit_type	required_car_parking_spaces	total_of_special_requests
1	2	20160904	96926	BB	Offline TA/TO	86	A	No Deposit	0	0
2	2	20160909	96927	BB	Offline TA/TO	85	A	No Deposit	0	0
3	2	20160908	96928	BB	Online TA	9	A	No Deposit	0	1
4	2	20160907	96929	BB	Online TA	9	D	No Deposit	0	0
5	2	20160907	96930	BB	Online TA	9	D	No Deposit	0	0
6	2	20160910	96931	BB	Direct	NULL	G	No Deposit	0	1
7	2	20160908	96932	BB	Online TA	9	D	No Deposit	0	3
8	2	20160908	96933	BB	Online TA	9	A	No Deposit	0	1
9	2	20160905	96934	BB	Online TA	9	A	No Deposit	0	0
10	2	20160908	96935	BB	Offline TA/TO	171	A	No Deposit	0	0
11	2	20160906	96936	BB	Offline TA/TO	44	A	No Deposit	0	0

# CHƯƠNG 4: Phân tích dữ liệu

## 4.1. Quá trình xây dựng mô hình:

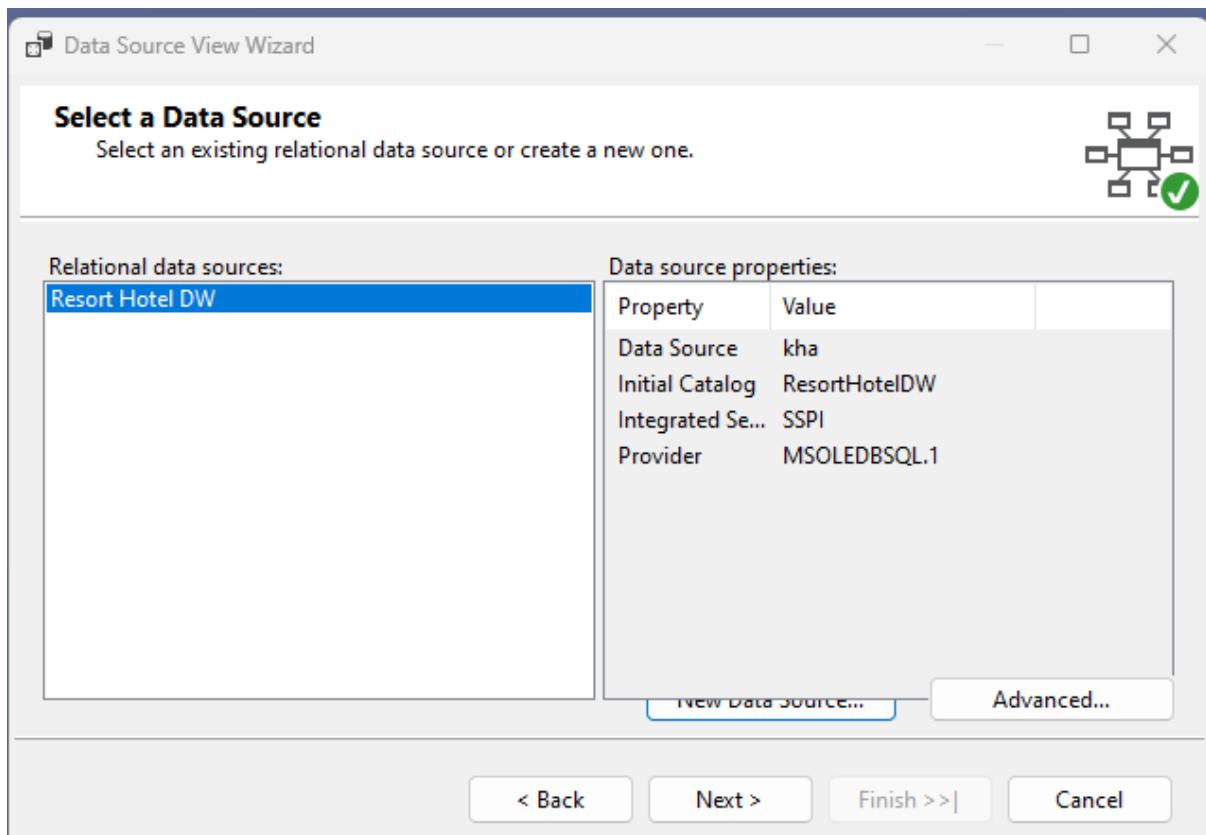
### 4.1.1. Tạo Data Source

- Right-click vào Data Sources ở khung Solution Explorer. Chon New Data Source để tạo Data Source mới. Kết nối đến Database **Resot Hotel DW**

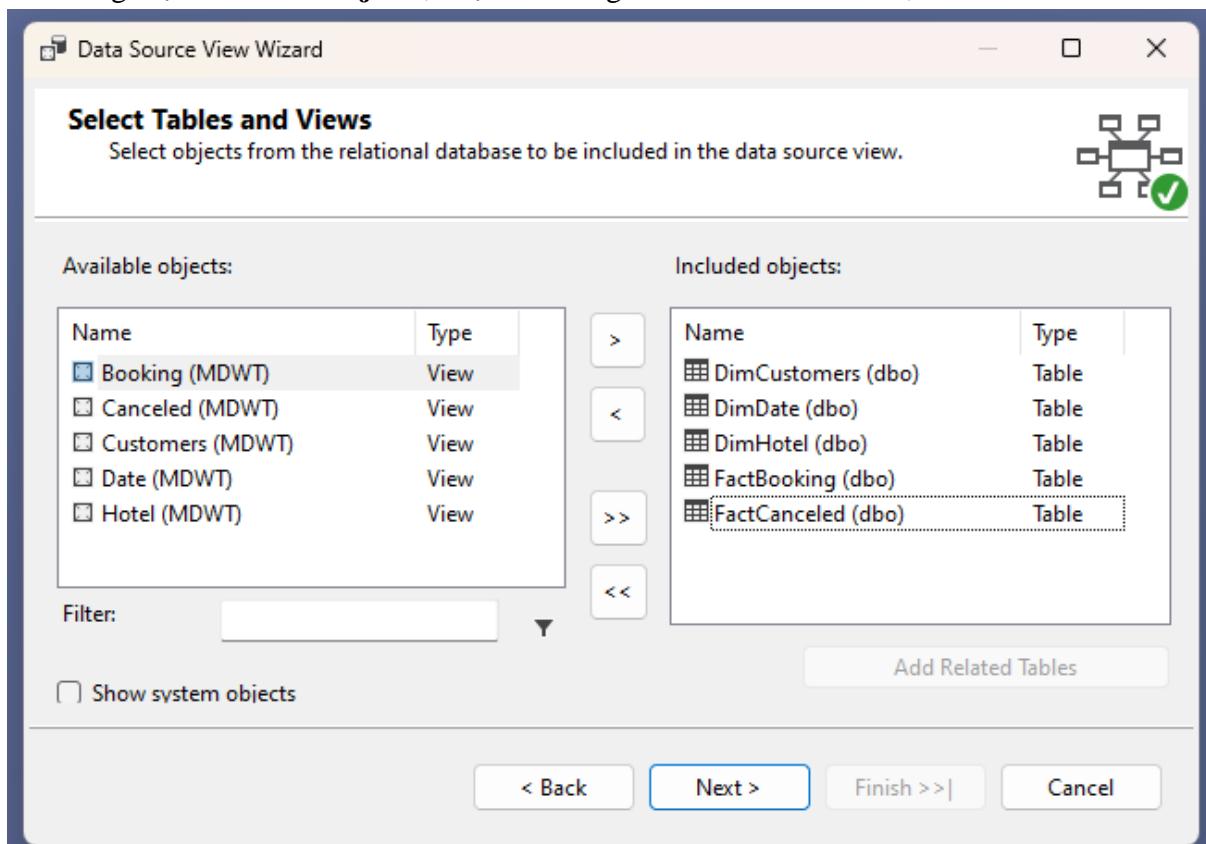


### 4.1.2. Tạo Data Source View

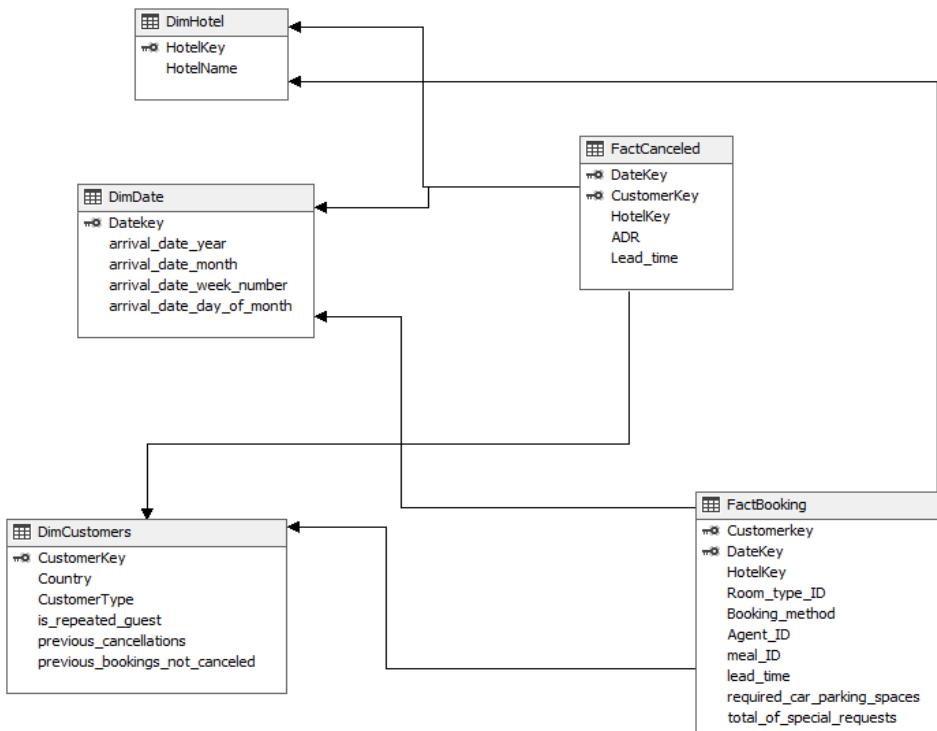
- Right-click vào Data Source Views ở khung Solution Explorer. Chon New Data Source View để tạo Data Source View mới. Chọn Data Soure là **Resot Hotel DW** và chọn Next.



- Trong mục **Included objects**, chọn các bảng Dim và Fact cần thiết, nhấn Next

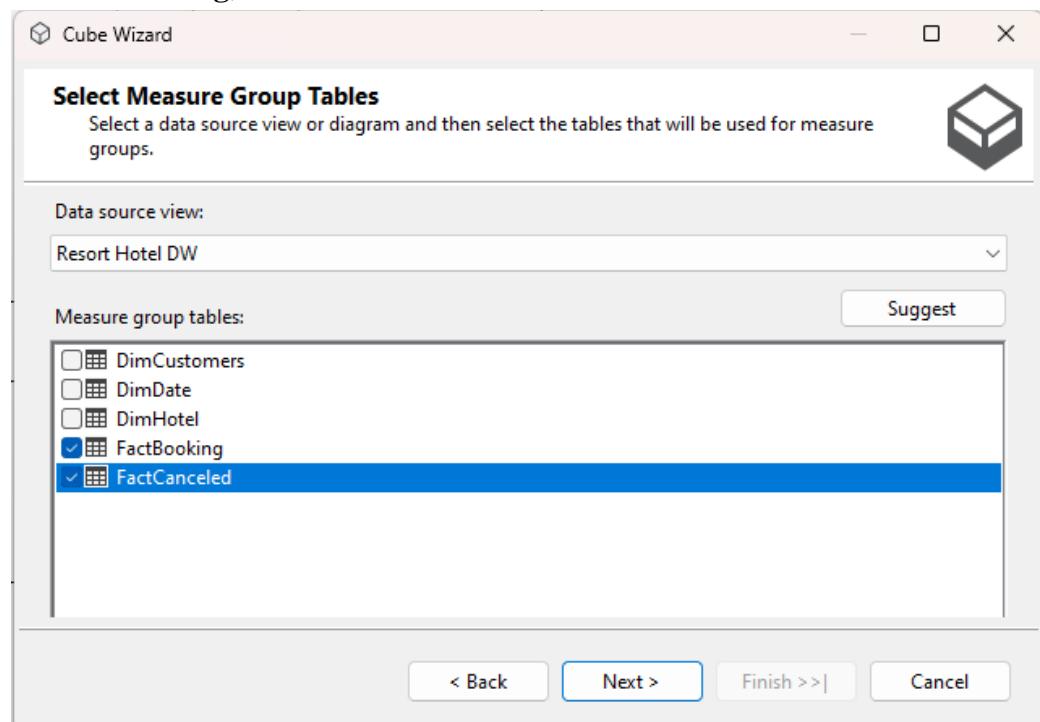


- Đặt tên cho Data Source View và nhấn Finish để hoàn tất việc tạo Data Source View
- Double-Click vào Data Source View vừa tạo để hiển thị diagram

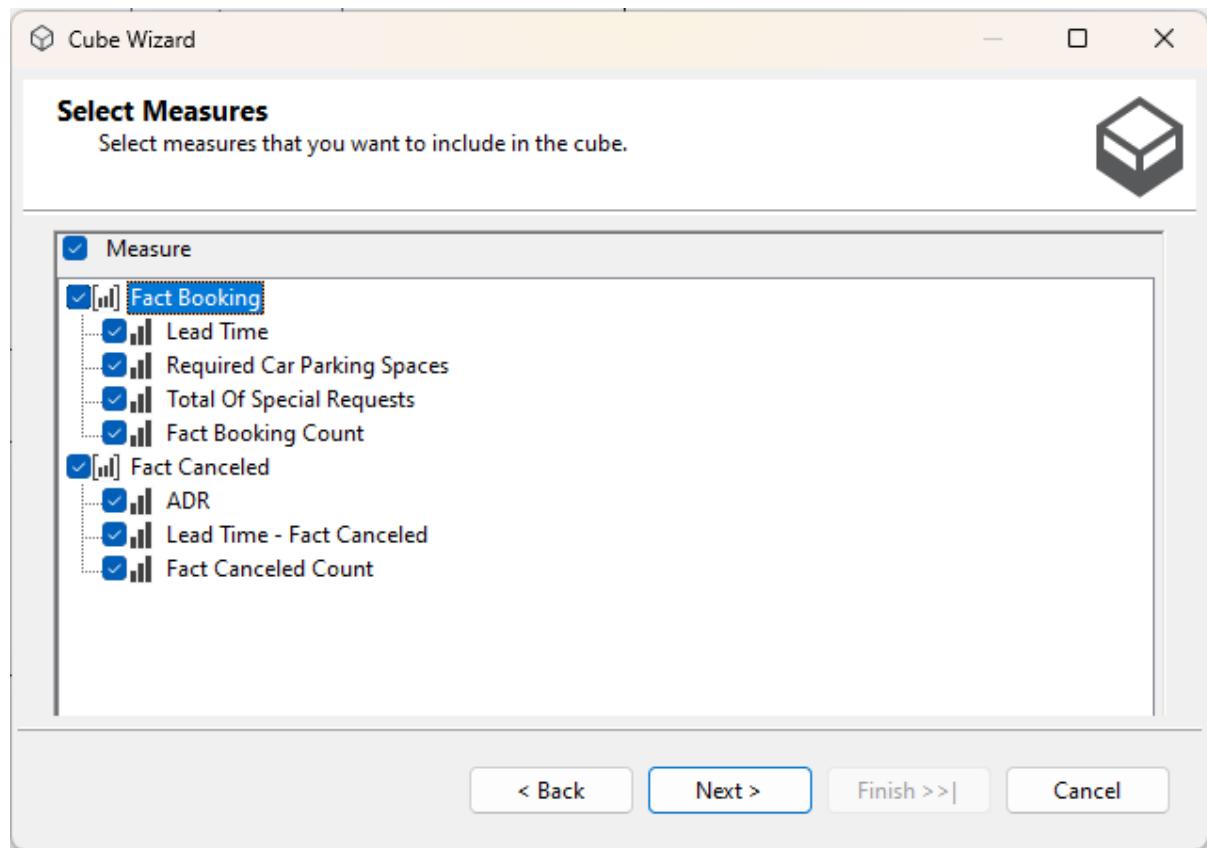


## 4.2. Quá trình xây dựng khối Cube

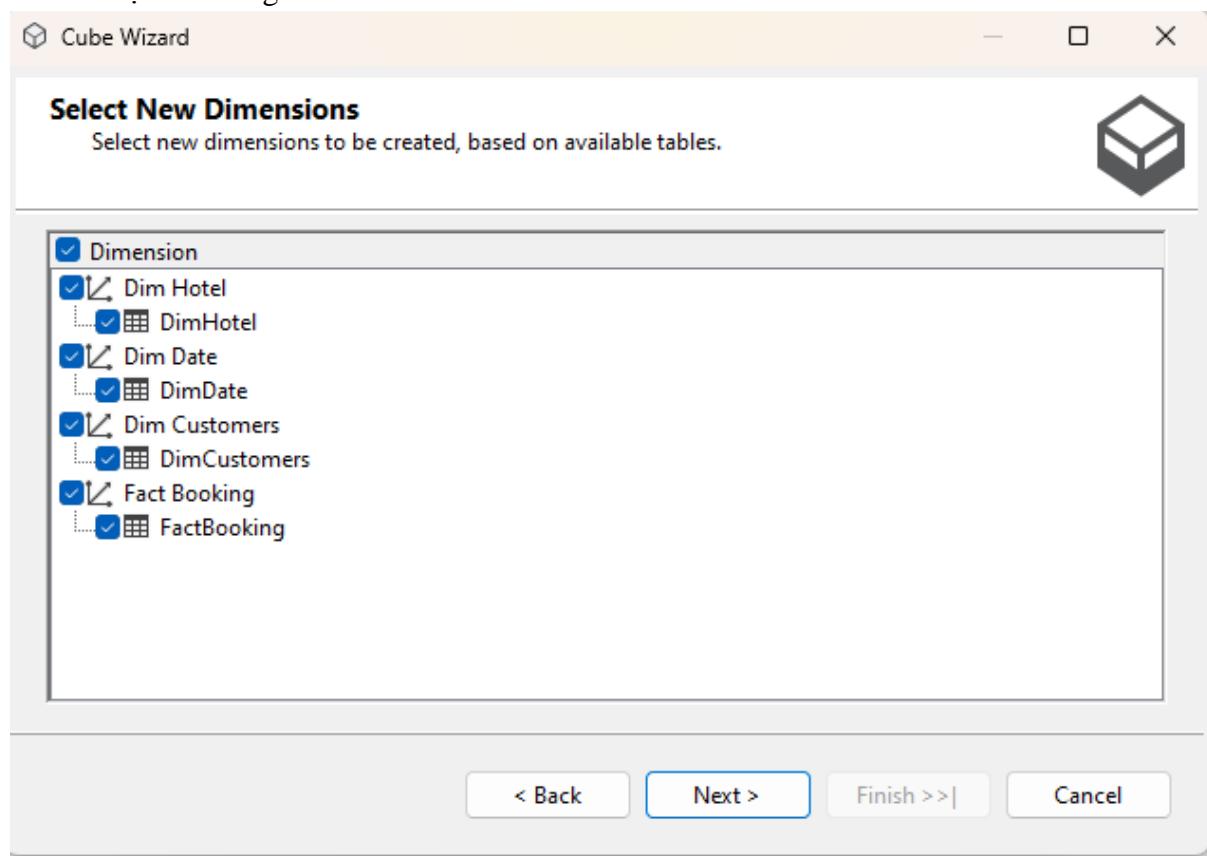
- Right-click vào **Cubes** ở khung Solution Explorer. Chon New Cube để tạo Cube mới.
- Chọn Data Source View là **Resort Hotel DW View**. Tích chọn 2 bảng **FactCanceled** và **FactBooking**, nhấn Next



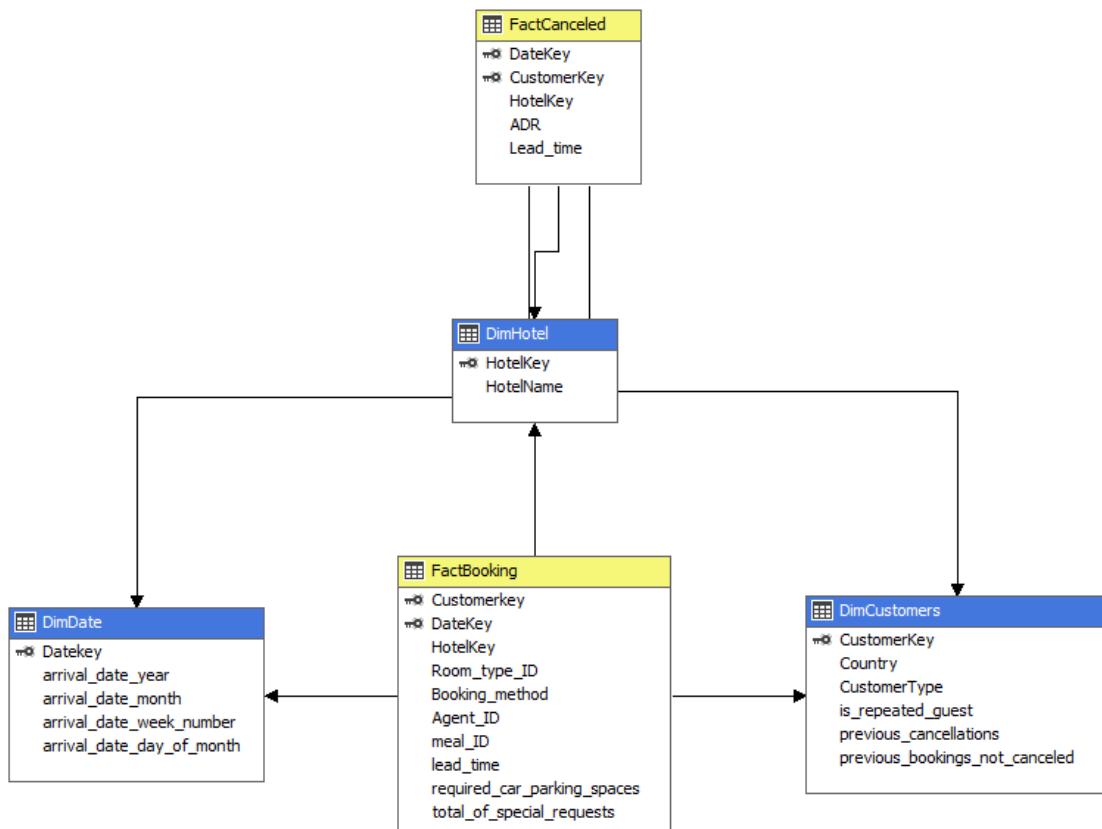
- Ở dialog Select Measures, tích chọn những measure:



- Khởi tạo các bảng Dimension:



- Đặt tên **Resort Hotel DW Cube** để hoàn tất khởi tạo.
- Các bảng Fact, Dimension và quan hệ giữa chúng sẽ được thể hiện trong Diagram.



## 4.3. Cấu hình Hierarchy

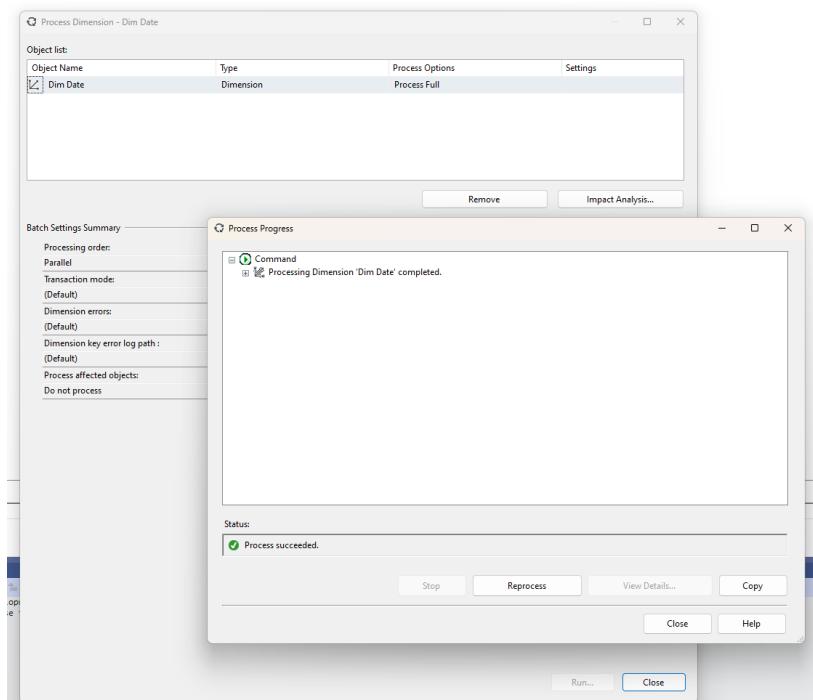
### 4.3.1. Tạo Hierarchy cho Dim Date

- Thêm thuộc tính vào bảng Dim: Tại table DimDate trong panel Data Source View, kéo thuộc tính cần thiết vào panel Attributes.
- Kéo thả 4 thuộc tính *Year*, *Month*, *Day*, *Date Key* từ Attributes vào panel Hierarchies để tạo ra Time Hierarchy
- Đặt lại tên cho các thuộc tính trong Time Hierarchy

The screenshot shows the SSAS Dimension Structure editor interface with the following details:

- Attributes** pane: Shows the attributes of the Dim Date dimension: Datekey, Day, Month, and Year.
- Hierarchies** pane: A Time Hierarchy is being created, containing levels Day, Month, and Year. A tooltip indicates: "To create a new hierarchy, drag an attribute here."
- Data Source View** pane: Shows the DimDate table with its attributes: Datekey, arrival\_date\_year, arrival\_date\_month, arrival\_date\_week\_number, and arrival\_date\_day\_of\_month.

- Process để load dữ liệu cho bảng Dim Date

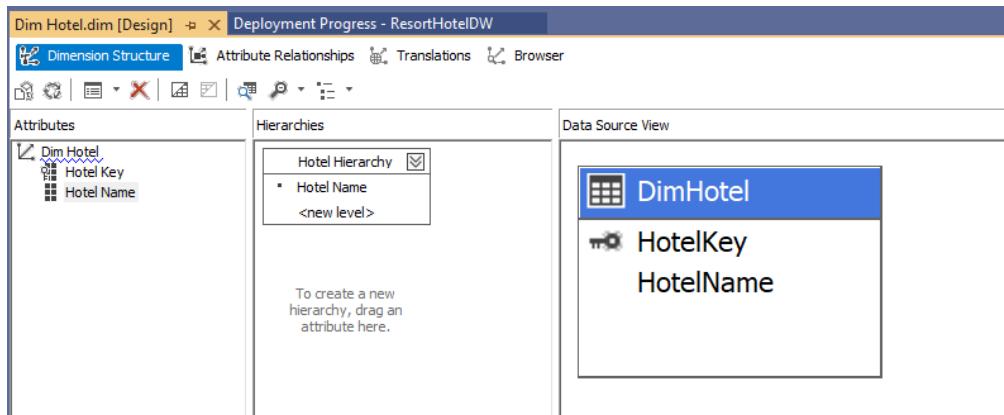


- Chuyển đến tab Browser. Ta thấy thời gian được phân cấp theo năm → tháng → ngày.

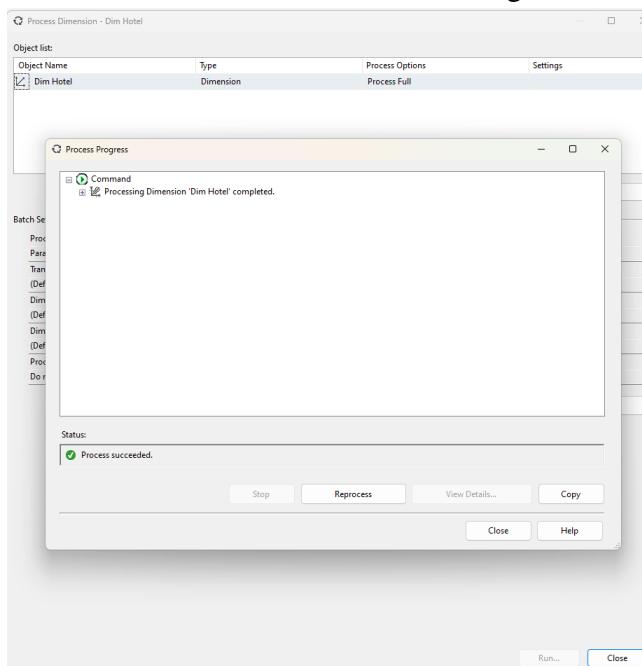
Current level:	(All)
2015	10, 11, 12, 7, 8, 9
2016	1, 10, 11, 12, 2, 3, 4, 5, 6, 7, 8, 9
2017	1, 2, 3, 4, 5, 6, 7, 8, 1, 10, 11, 12, 13, 14, 15, 16, 17, 18

#### 4.3.2. Tạo Hierarchy cho Dim Hotel:

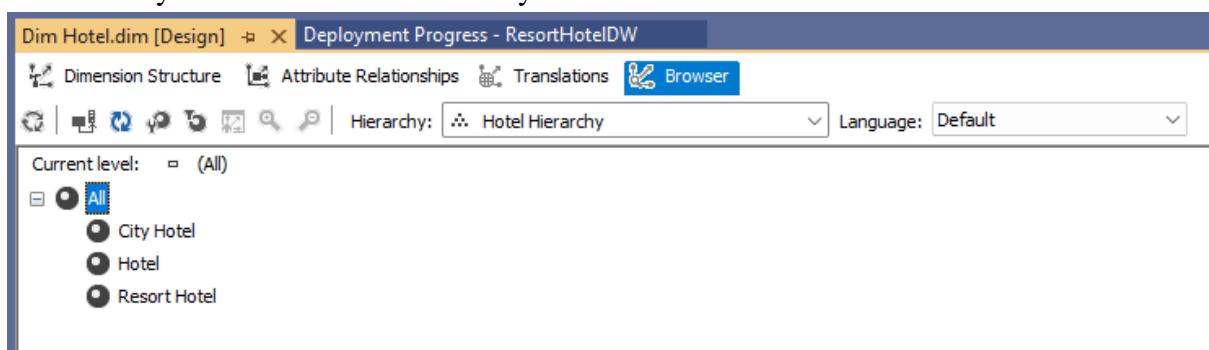
- Thêm thuộc tính vào bảng Dim: Tại table DimHotel trong panel Data Source View, kéo thuộc tính cần thiết vào panel Attributes.
- Đặt lại tên cho các thuộc tính trong Hotel Hierarchy



- Process để load dữ liệu cho bảng Dim Date



- Chuyển đến tab Browser. Ta thấy tên các hotel.

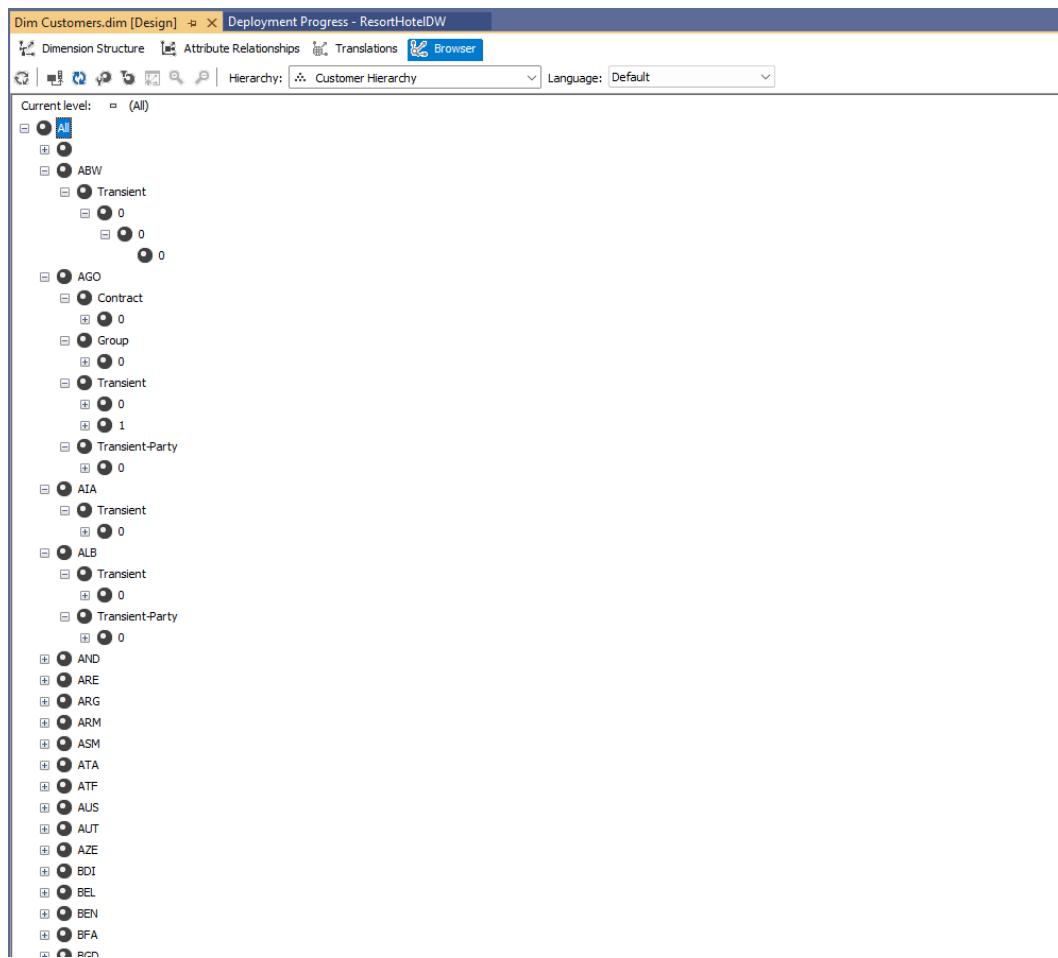


### 4.3.3. Tạo Hierarchy cho Dim Customers:

- Thêm thuộc tính vào bảng Dim: Tại table DimCustomer trong panel Data Source View, kéo thuộc tính cần thiết vào panel Attributes.
- Đặt lại tên cho các thuộc tính trong Customer Hierarchy

- Process để load dữ liệu cho bảng Dim Date

- Chuyển đến tab Browser. Ta thấy tên các thông tin của customer.



## 4.4. Thực hiện tích dũ liệu.

### 4.4.1. Câu hỏi: Số lượng huỷ phòng của Resort Hotel và City Hotel ở các tháng theo từng năm.

#### 4.4.1.1. Sử dụng công cụ SSAS

- Số lượng hủy phòng năm 2015:

The screenshot shows the SSAS Dimension Editor interface. On the left, there's a tree view of the 'Resort Hotel DW' dimension, with 'Fact Booking Count' selected under 'Measures'. The main area displays a table with two rows: 'City Hotel' with a count of 6004 and 'Resort Hotel' with a count of 2138. The top navigation bar includes 'Edit as Text', 'Import...', 'MDX', and other standard database tools.

Hotel Name	Fact Canceled Count
City Hotel	6004
Resort Hotel	2138

- Số lượng hủy phòng năm 2016:

The screenshot shows the SSAS Dimension Editor interface. The configuration is identical to the 2015 query, but the filter expression in the top right is now '{ 2016 }'. The resulting table shows the same data as the 2015 query: City Hotel with 15407 and Resort Hotel with 4930.

Hotel Name	Fact Canceled Count
City Hotel	15407
Resort Hotel	4930

- Số lượng hủy phòng năm 2017:

The screenshot shows the SSAS Dimension Editor interface. The configuration is identical to the previous years, with the filter expression set to '{ 2017 }'. The resulting table shows the same data: City Hotel with 11691 and Resort Hotel with 4054.

Hotel Name	Fact Canceled Count
City Hotel	11691
Resort Hotel	4054

#### 4.4.1.2. Sử dụng Pivot Table trong Excel

Năm 2015:

### Số lượng huỷ phòng của Resort Hotel và City Hotel theo từng năm

Count of Custer	HotelName		
Tháng	City Hotel	Resort Hotel	Tổng số lượng huỷ
10	1321	411	1732
11	301	185	486
12	668	305	973
7	939	320	1259
8	1232	366	1598
9	1543	551	2094



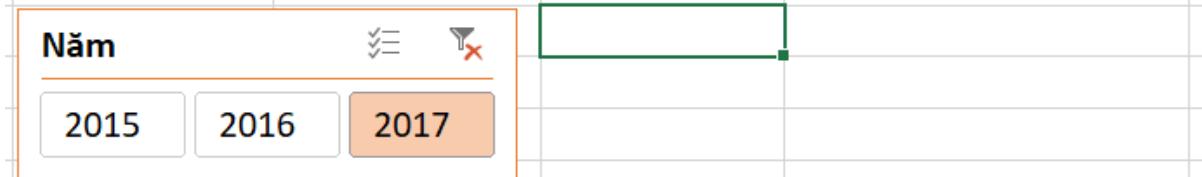
Năm 2016:

Count of Custer	HotelName		
Tháng	City Hotel	Resort Hotel	Tổng số lượng huỷ
10	1947	567	2514
11	1360	276	1636
12	1072	326	1398
7	1043	456	1499
8	1247	578	1825
9	1567	455	2022
1	438	119	557
2	930	407	1337
3	1108	369	1477
4	1539	522	2061
5	1436	479	1915
6	1720	376	2096

Năm 2017:

## Số lượng huỷ phòng của Resort Hotel và City Hotel theo từng năm

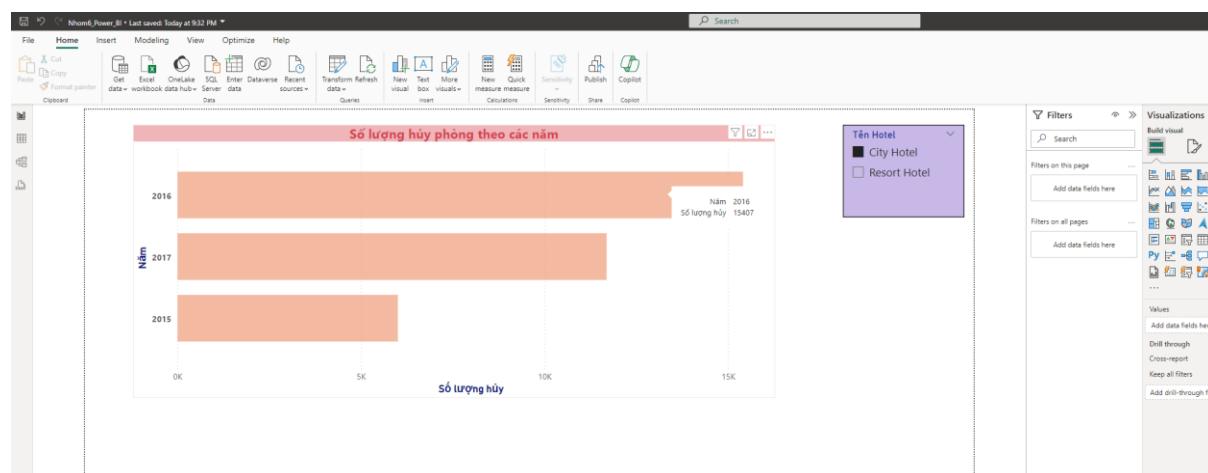
Count of Custer	HotelName		
Tháng	City Hotel	Resort Hotel	Tổng số lượng huỷ
7	1324	660	1984
8	1123	693	1816
1	1044	206	1250
2	971	388	1359
3	1278	394	1672
4	1926	537	2463
5	2217	545	2762
6	1808	631	2439



### 4.4.1.3. Sử dụng Power BI

Với các số liệu từ Pivot và Power Bi ta thấy được vào khoảng thời gian từ tháng 4 đến tháng 10 thì số lượng huỷ phòng có sự hướng tăng ở City Hotel. Còn Resort Hotel số lượng huỷ không có sự biến động lớn. Số lượng huỷ phòng ở Resort Hotel thường thấp hơn so với City Hotel.

- Số lượng huỷ phòng theo các năm, có thể tùy chọn các hotel cụ thể ( City Hotel hoặc Resort Hotel hoặc cả hai)



## 4.4.2. Câu hỏi: Lượng huỷ phòng theo quốc gia của từng năm.

### 4.4.2.1. Sử dụng công cụ SSAS

- Số lượng hủy phòng của các nước trong tất cả các năm

Country	Fact Canceled Count
ARE	43
ARG	54
ARM	2
AUS	107
AUT	230
AZE	9
BEL	474
BEN	3
BGD	9
BGR	12
BHR	4
BIH	3
BLR	9
BRA	830
CHE	428
CHL	16
CHN	462
CIV	2
CN	254
COL	23
CPV	12
CRI	1
CYP	11
CZE	37
DEU	1218
DNK	109
DOM	8
DZA	21
ECU	8
EGY	11
ESP	2177

### 4.4.2.2. Sử dụng Pivot Table trong Excel

Năm 2015:

Số lượng huỷ phòng ở theo Quốc gia của từng năm			
Quốc gia	Số lượng khách huỷ	Tỷ lệ	Năm
PRT	7582	96.23%	2015
ESP	132	0.93%	2016
ITA	125	1.27%	2017
GBR	115	0.45%	
IRL	66	0.25%	
FRA	42	0.49%	
	32	0.22%	
BEL	17	0.08%	
CN	12	0.01%	
BRA	5	0.08%	

Năm 2016:

### Số lượng huỷ phòng ở theo Quốc gia của từng năm

Quốc gia	Số lượng khách huỷ	Tỷ lệ	Năm
PRT	12405	68.00%	2015
ESP	1174	6.02%	2016
GBR	1004	5.24%	2017
FRA	903	5.79%	
ITA	662	4.42%	
DEU	583	3.70%	
BRA	381	2.37%	
IRL	317	1.46%	
CHN	251	1.70%	
USA	215	1.31%	

Năm 2017:

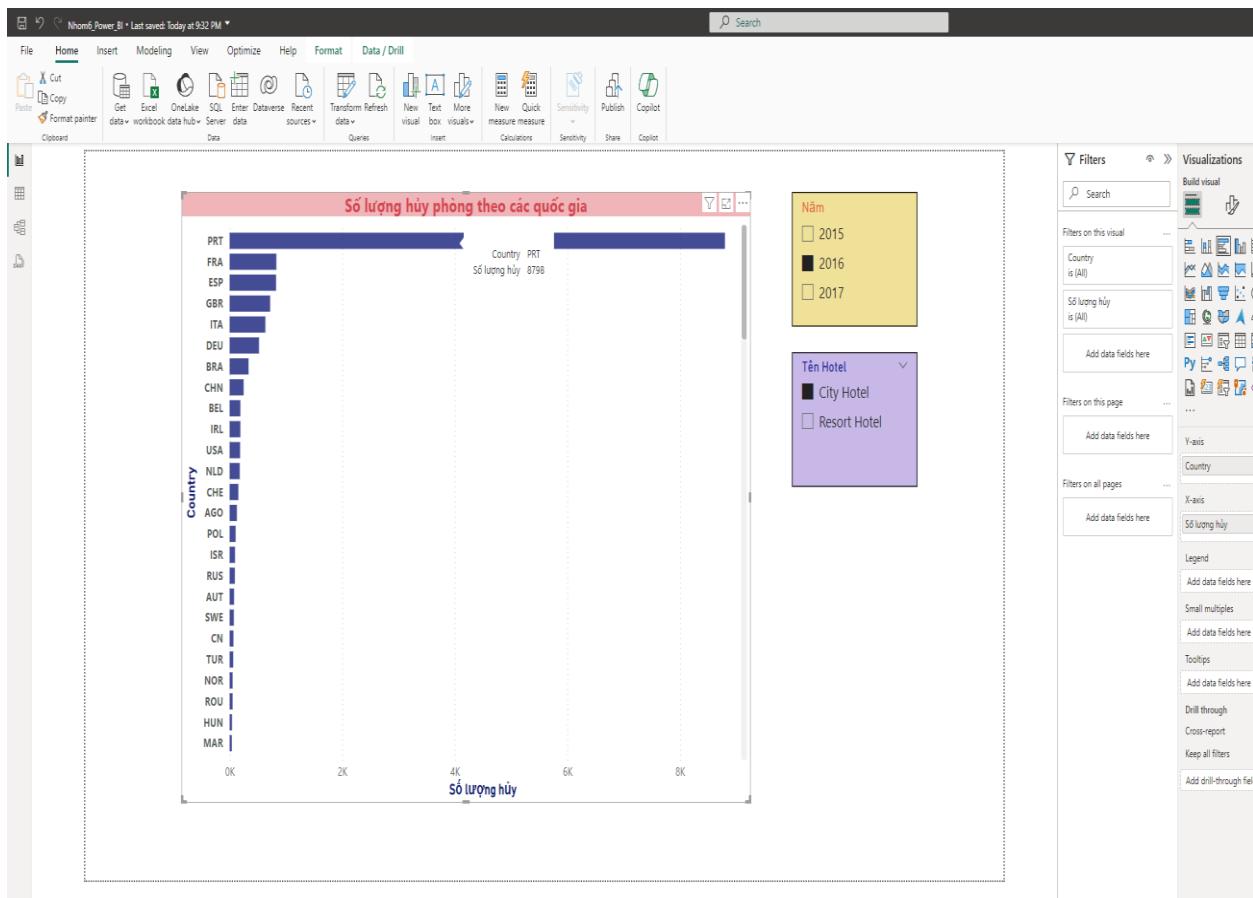
### Số lượng huỷ phòng ở theo Quốc gia của từng năm

Quốc gia	Số lượng khách huỷ	Tỷ lệ	Năm
PRT	7532	56.42%	2015
GBR	1334	8.99%	2016
FRA	989	8.47%	2017
ESP	871	4.91%	
DEU	633	5.44%	
ITA	546	4.86%	
IRL	449	2.44%	
BRA	444	3.78%	
USA	286	2.47%	
BEL	249	2.22%	

Từ số liệu và biểu đồ thì quốc giá PRT là số lượng huỷ phòng lớn. Cụ thể vào năm 2015 số lượng khách huỷ phòng lên đến 96,23% việc này có ảnh hưởng rất lớn đến việc kinh doanh của khách sạn. Tuy nhiên sau đó các năm 2016, 2017 thì số lượng khách ở PRT huỷ phòng có xu hướng giảm nhõ có biện pháp xử lý kịp thời. Tiếp đến có thấy lượng khách ở các quốc gia ESP, FRA, GBR,, chiếm phân lớn số lượng huỷ phòng. Từ đây còn có biện pháp giảm thiểu việt huỷ phòng bằng cách có chính sách kinh doanh phù hợp với khách hàng ở từng quốc gia.

#### 4.4.2.3. Sử dụng Power BI

- Số lượng hủy phòng theo các quốc gia
- Có thể tùy chọn theo Năm (2015 – 2016 – 2017 - All) , Tên Hotel ( Resort Hotel – City Hotel – All)



**4.4.3. Câu hỏi: Cho biết số lượng huỷ phòng và giá trung bình ADR qua các tháng theo từng năm của Resort Hotel và City Hotel.**

#### 4.4.3.1. Sử dụng công cụ SSAS

City Hotel số lượng hủy phòng:

The screenshot shows the SSAS Cube Design interface for the "Resort Hotel DW.cube". The left pane displays the cube's structure, including the "Fact Booking" and "Fact Canceled" measure groups. The right pane shows a table titled "Fact Canceled Count" with data for each month from 1 to 9. The table includes columns for Month and Fact Canceled Count.

Month	Fact Canceled Count
1	1482
10	3268
11	1661
12	1740
2	1901
3	2386
4	3465
5	3653
6	3528
7	3306
8	3602
9	3110

City Hotel giá trung bình ADR:

The screenshot shows the SSAS Cube Design interface for the "Resort Hotel DW.cube". The left pane displays the cube's structure, including the "Fact Booking" and "Fact Canceled" measure groups. The right pane shows a table titled "ADV\_AVERAGE" with data for each month from 1 to 9. The table includes columns for Month and ADV\_AVERAGE.

Month	ADV_AVERAGE
1	53.7896923076923
10	63.0728527607363
11	46.5266811279825
12	71.1003011093502
2	58.1460628930818
3	59.2320838794233
4	82.6212464589235
5	83.9582910156253
6	115.505635551142
7	166.232388579387
8	197.901942577886
9	86.6392345924456

Resort Hotel số lượng hủy phòng:

Cube Structure   Dimension Usage   Calculations   KPIs   Actions   Partitions   Aggregations   Perspectives   Translations   Browser

Language: English (United States)

Resort Hotel DW

Dimension   Hierarchy   Operator   Filter Expression

Dim Hotel   Hotel Hierarchy   Equal   {Resort Hotel}

Search Model   Measure Group:

Resort Hotel DW

- Measures
  - Fact Booking
    - Fact Booking Count
    - Required Car Parking Spaces
    - Total Of Special Requests
  - Fact Canceled
    - ADR
    - Fact Canceled Count
- KPIs
- Dim Customers
- Dim Date
- Dim Hotel
- Fact Booking

Month	Fact Canceled Count
1	1482
10	3268
11	1661
12	1740
2	1901
3	2386
4	3465
5	3653
6	3528
7	3306
8	3602
9	3110

Resort Hotel giá trung bình ADR:

Dimension   Hierarchy   Operator   Filter Expression

Dim Hotel   Hotel Hierarchy   Equal   {Resort Hotel}

Month   ADV\_AVERAGE

Month	ADV_AVERAGE
1	83.3413090418354
10	97.6235771113828
11	90.6165081276338
12	90.1591494252873
2	83.3237822198843
3	96.8625565800502
4	110.550822510822
5	123.13362168081
6	120.772559523809
7	103.748581367211
8	109.105805108273
9	106.426729903537

#### 4.4.3.2. Sử dụng Pivot Table trong Excel

Năm 2015:

City Hotel.

Số lượng huỷ phòng hàng tháng và tỷ lệ ADR		
Tháng	Số Lượng	Giá trung bình
10	1321	85.09
11	301	80.23
12	668	86.58
7	939	67.98
8	1232	73.67
9	1543	97.88

HotelName

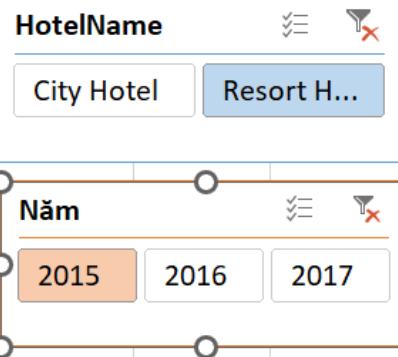
City Hotel   Resort H...

Năm

2015   2016   2017

Resort Hotel.

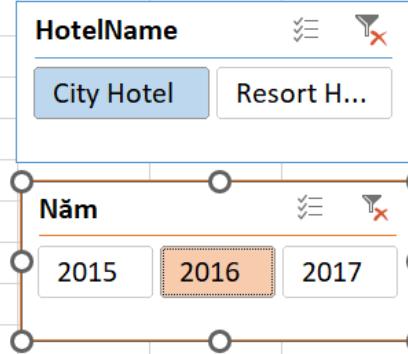
V	W	X	Y	Z	AA	AB
Số lượng huỷ phòng hàng tháng và tỷ lệ ADR						
<b>Tháng</b>	<b>Số Lượng</b>	<b>Giá trung bình</b>				
10	411	54.36				
11	185	46.95				
12	305	64.74				
7	320	129.68				
8	366	162.13				
9	551	67.00				



Năm 2016:

**City Hotel.**

V	W	X	Y	Z	AA	AB
Số lượng huỷ phòng hàng tháng và tỷ lệ ADR						
<b>Tháng</b>	<b>Số Lượng</b>	<b>Giá trung bình</b>				
1	438	78.50				
10	1947	106.13				
11	1360	92.91				
12	1072	92.39				
2	930	76.72				
3	1108	99.06				
4	1539	99.75				
5	1436	108.15				
6	1720	111.36				
7	1043	106.40				
8	1247	113.88				
9	1567	114.84				



**Resort Hotel.**

V	W	X	Y	Z	AA	AB
Số lượng huỷ phòng hàng tháng và tỷ lệ ADR						
<b>Tháng</b>	<b>Số Lượng</b>	<b>Giá trung bình</b>				
1	119	57.05				
10	567	69.39				
11	276	46.24				
12	326	77.05				
2	407	63.24				
3	369	61.97				
4	522	72.73				
5	479	72.47				
6	376	110.11				
7	456	160.19				
8	578	198.39				
9	455	110.42				

Năm 2017:

City Hotel.

V	W	X	Y	Z	AA	AB
Số lượng huỷ phòng hàng tháng và tỷ lệ ADR						
<b>Tháng</b>	<b>Số Lượng</b>	<b>Giá trung bình</b>				
1	1044	85.37				
2	971	89.65				
3	1278	94.96				
4	1926	119.18				
5	2217	132.84				
6	1808	129.72				
7	1324	127.03				
8	1123	142.69				

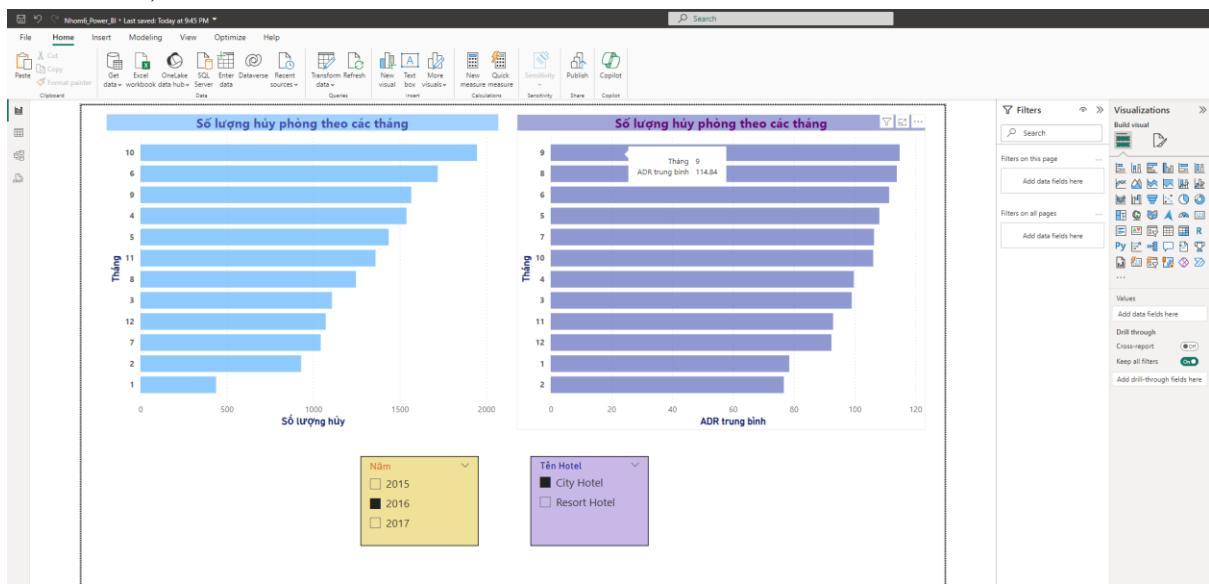
Resort

V	W	X	Y	Z	AA	AB
Số lượng huỷ phòng hàng tháng và tỷ lệ ADR						
<b>Tháng</b>	<b>Số Lượng</b>	<b>Giá trung bình</b>				
1	206	51.91				
2	388	52.80				
3	394	56.67				
4	537	92.23				
5	545	94.05				
6	631	118.79				
7	660	188.13				
8	693	216.39				

Với ADR trung bình từng tháng và đi kèm số lượng huỷ phòng. Có thể thấy lý do huỷ phòng cũng có bị ảnh hưởng một phần do ADR. Những tháng có lượng huỷ phòng cao thì cũng có phần đi kèm do ADR cao. Và giá ADR có vẻ là tăng theo các lên từ tháng 1 đến tháng 12.

#### 4.4.3.3. Sử dụng Power BI

- Số lượng huỷ phòng và giá trung bình ADR
- Có thể tùy chọn theo Năm (2015 – 2016 – 2017 - All) , Tên Hotel ( Resort Hotel – City Hotel – All)



#### 4.4.4. Câu hỏi: Số lượng đặt phòng qua các công ty lữ hành

##### 4.4.4.1. Sử dụng công cụ SSAS

- Số lượng đặt phòng theo từng Agent:

Agent	Fact Booking Count
1	7191
10	260
103	21
104	53
105	14
106	2
107	2
11	395
110	12
111	16
112	15
114	1
115	225
117	1
118	69
119	304
12	578
121	37
122	2
126	14
127	45
128	23
129	14
13	82
132	143
133	56

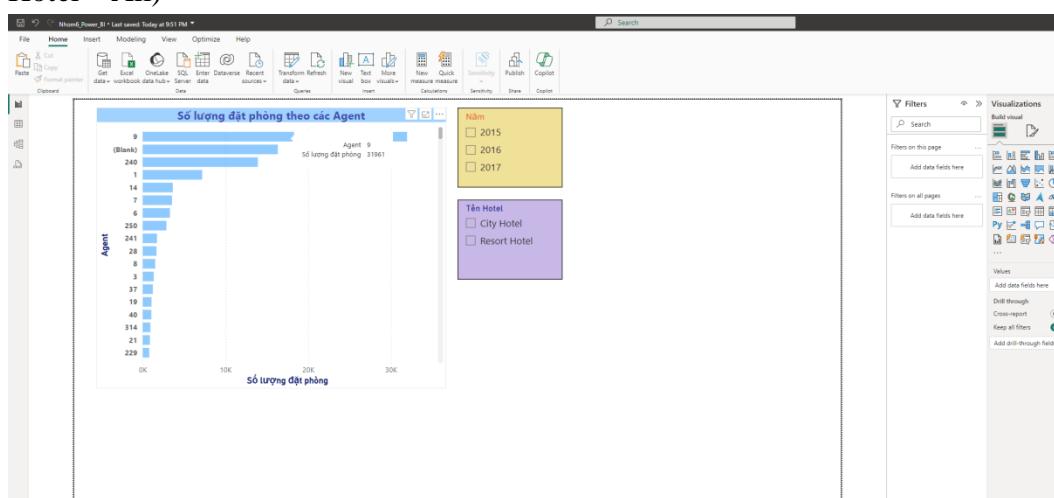
#### 4.4.4.2 . Sử dụng Pivot table

Số lượng đặt phòng qua Công ty lữ hành			
agent	Số lượng đặt phòng	Tỷ lệ	DataBar
(blank)	16340	18.97%	
1	7191	8.35%	
14	3640	4.23%	
240	13922	16.16%	
241	1721	2.00%	
250	2870	3.33%	
28	1666	1.93%	
6	3290	3.82%	
7	3539	4.11%	
9	31961	37.10%	

Đây là dữ liệu top 10 agent có số lượng đặt phòng lớn. Với những lượt đặt phòng không qua agent thì là blank. Với Agent 9 số lượng đặt phòng lớn nhất chiếm 37.10% và Agent 240 là 16.16%. Qua đây ta sẽ có các chính sách hợp tác với các công ty lữ hành để tạo mối hợp tác tốt và có các khuyến mãi giúp đẩy kinh doanh của cả hai bên.

#### 4.4.4.3. Sử dụng Power BI

- Số lượng đặt phòng qua các công ty lữ hành.
- Có thể tùy chọn theo Năm (2015 – 2016 – 2017 - All) , Tên Hotel ( Resort Hotel – City Hotel – All)



#### 4.4.5. Câu hỏi: Thống kê số lượng đặt phòng ứng với các phương thức đặt phòng.

##### 4.4.5.1. Sử dụng công cụ SSAS

- Số lượng đặt phòng theo phương thức:

Market Segment	Fact Booking Count
Aviation	237
Complementary	743
Corporate	5295
Direct	12606
Groups	19811
Offline TA/TO	24219
Online TA	56477
Undefined	2

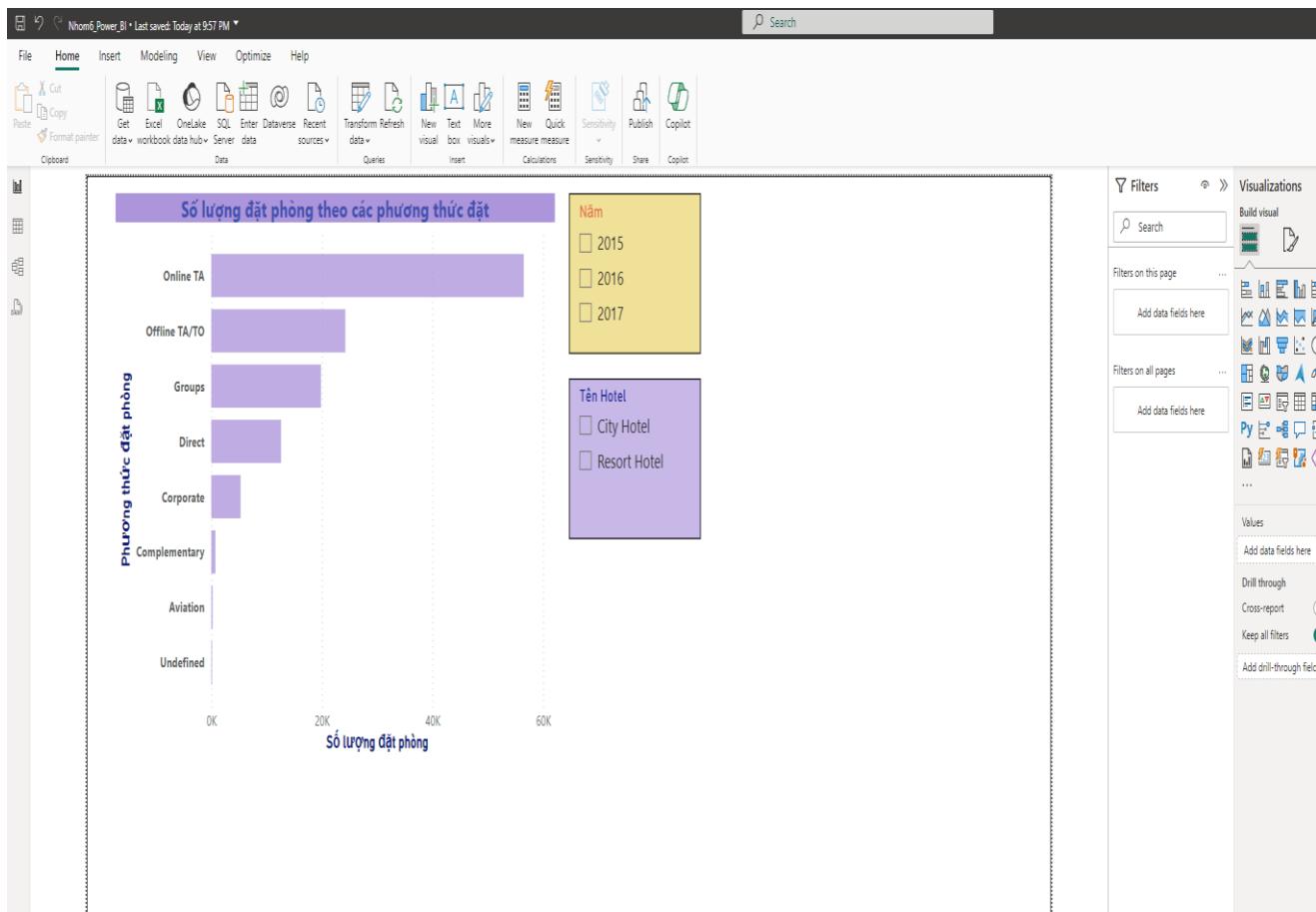
##### 4.4.5.2. Sử dụng Pivot Table trong Excel

Phương thức đặt phòng	
Phương thức đặt phòng	Số lượng
Aviation	237
Complementary	743
Corporate	5295
Direct	12606
Groups	19811
Offline TA/TO	24219
Online TA	56477
Undefined	2

Với phương thức Online TA thì có số lượng rất lớn, với phương thức Aviation thì rất ít khách hàng sử dụng đặt phòng bằng phương thức này.

##### 4.4.5.3. Sử dụng Power BI

- Số lượng đặt phòng theo các phương thức đặt.
- Có thể tùy chọn theo Năm (2015 – 2016 – 2017 - All) , Tên Hotel ( Resort Hotel – City Hotel – All)



# **CHƯƠNG 5: KẾT LUẬN**

## **5.1. Kết quả đạt được**

Thực hiện xây dựng ‘*KHO DỮ LIỆU CHO KHÁCH SẠN*’ giúp chúng em hiểu sâu hơn về kiến thức môn học Kho Dữ liệu và thành thạo hơn trong việc sử dụng các công cụ hỗ trợ quá trình ETL đưa dữ liệu vào kho dữ liệu như công cụ SSIS, SSAS, các công cụ để trực quan hóa dữ liệu như Excel Pivot, Power BI Desktop. Nâng cao kỹ năng nhìn nhận và tiếp thu, xử lý thông tin đầu vào, đặt những câu hỏi tiền đề và thảo luận hướng đi để giải quyết và trả lời câu hỏi.

## **5.2. 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.