

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



BÁO CÁO CUỐI KỲ
MÔN: KHO DỮ LIỆU

**ĐỀ TÀI: XÂY DỰNG KHO DỮ LIỆU PHÂN TÍCH
NHẬP XUẤT VIỆN CỦA BỆNH NHÂN**

GVHD: ThS. NGUYỄN VĂN THÀNH

SINH VIÊN THỰC HIỆN (NHÓM 9):

- | | |
|----------------------|----------|
| 1. Mai Hồng Hải | 22133014 |
| 2. Vy Gia Nghi | 22133037 |
| 3. Trần Tuyên Trường | 20133108 |

Thành phố Hồ Chí Minh, tháng 05 năm 2025

PHÂN CÔNG NHIỆM VỤ

Nhóm: 09

Lớp: DAWH430784

Nhiệm vụ	Hồng Hải	Gia Nghi	Tuyên Trường
Tìm kiếm dữ liệu			X
Xử lý dữ liệu	X	X	X
Xác định business process	Admission Report	Diagnosis Report	
Xác định bảng Dim	DimDate, DimPerson	DimPersonHealth, DimComorbidity	DimDiagnosis, DimHabit
Xác định bảng Fact		FactAdmission	FactDiagnosis
Đẩy dữ liệu từ CSV – SQL Server	X		
Tạo nguồn kết nối dữ liệu	X		
Staging, load các Dim & Fact	X		
Nhập dữ liệu vào SSAS, tạo Data Cube			X
Phân tích bằng SSAS			X
Đặt 10 câu hỏi		X	
Trả lời bằng SSAS			X
Trả lời bằng Pivot Table		X	
Tạo báo cáo Power BI		X	

NHẬN XÉT CỦA GIẢNG VIÊN

ĐIỂM CỦA GV

MỤC LỤC

CHƯƠNG 1: TỔNG QUAN ĐỀ TÀI.....	1
1.1. Lý do chọn đề tài.....	1
1.2. Tổng quan tập dữ liệu	1
1.2.1. Nguồn gốc và cách thu thập dữ liệu	1
1.2.2. Mô tả các cột dữ liệu	1
1.3. Mục tiêu và câu hỏi phân tích chính	5
1.4. Các công cụ được sử dụng	6
CHƯƠNG 2: THIẾT KẾ DATA WAREHOUSE	7
2.1. Xác định Business Process và yêu cầu nghiệp vụ	7
2.2. Thiết kế bảng Fact và Dimension	8
2.2.1. Danh sách các bảng Dimension.....	8
2.2.2. Thiết kế bảng Fact	12
2.3. Thiết kế mô hình dữ liệu (Star/Galaxy Schema)	13
CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO DỮ LIỆU (SSIS).....	14
3.1. Tạo Solution và Project ETL	14
3.2. Thiết lập Connection và Cấu hình nguồn dữ liệu	14
3.3. Import Dữ liệu vào Database Gốc	17
3.3.1. Tạo Package Import.....	17
3.3.2. Thiết lập Control Flow	18
3.3.3. Load dữ liệu từ File nguồn vào Database.....	19
3.4. Tích hợp Dữ liệu vào các bảng Dimension	23
3.4.1. Tạo Package xử lý Dimension.....	23
3.4.2. Stage các Dimension	24
3.4.3. Load dữ liệu vào các bảng Dimension trong Data Warehouse	37
3.5. Tích hợp Dữ liệu vào các bảng Fact	49

3.5.1. Tạo Package xử lý Fact	49
3.5.2. Stage các Fact	50
3.5.3. Load dữ liệu vào các bảng Fact	59
3.6. Tổng hợp quy trình ETL với SSIS	62
CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU VỚI SSAS	64
4.1. Xây dựng mô hình OLAP trong SSAS	64
4.1.1. Tạo Project SSAS	64
4.1.2. Tạo Data Source	65
4.1.3. Tạo Data Source View.....	69
4.2. Tạo và cấu hình Cube	72
4.2.1. Khởi tạo Cube.....	72
4.2.2. Cập nhật thuộc tính cho Dimension	75
4.3. Tạo Hierarchy (Phân cấp thuộc tính).....	76
4.3.1. Phân cấp thời gian – DimDate.....	76
4.3.2. Phân cấp bệnh lý – DimDiagnosis.....	82
4.3.3. Phân cấp liên quan tới bệnh đồng mắc – DimComorbidity	83
4.4. Xây dựng & phân tích qua Excel Pivot Table	84
4.4.1. Kết nối SSAS với Excel	84
4.4.2. Trả lời các câu hỏi nghiệp vụ từ các bảng Pivot	86
4.5. Xây dựng báo cáo qua Power BI	96
CHƯƠNG 5: KẾT LUẬN.....	105
5.1. Kết quả đạt được	105
5.2. Hạn chế	105
Tài liệu tham khảo	107

CHƯƠNG 1: TỔNG QUAN ĐỀ TÀI

1.1. Lý do chọn đề tài

Nhóm chúng em chọn chủ đề "Xây dựng kho dữ liệu về việc nhập và xuất viện của bệnh nhân" bằng cách sử dụng tập dữ liệu Hospital Admissions Data bởi vì nhóm chúng em tin rằng đây là một chủ đề quan trọng có thể có tác động đáng kể đến ngành chăm sóc sức khỏe. Kho dữ liệu là một trung tâm lưu trữ dữ liệu được sử dụng cho phân tích và báo cáo. Bằng cách xây dựng một kho dữ liệu cho dữ liệu nhập viện và xuất viện, bệnh viện có thể có những thông tin quý giá về hoạt động của mình để cải thiện chăm sóc bệnh nhân, giảm chi phí và tăng hiệu quả.

Ví dụ, bệnh viện có thể sử dụng kho dữ liệu để:

- Nhận biết xu hướng trong việc nhập viện và xuất viện của bệnh nhân.
- Phân tích hiệu quả của các kế hoạch điều trị khác nhau.
- Nhận ra những lĩnh vực chi phí có thể giảm.
- Cải thiện hiệu quả của hoạt động bệnh viện.

Nhóm tin rằng việc xây dựng kho dữ liệu cho dữ liệu nhập viện và xuất viện bệnh viện là một bước quan trọng trong việc cải thiện ngành chăm sóc sức khỏe. Bằng cách có được những thông tin quý giá về hoạt động của mình, bệnh viện có thể cải thiện chăm sóc bệnh nhân, giảm chi phí và tăng hiệu quả.

1.2. Tổng quan tập dữ liệu

1.2.1. Nguồn gốc và cách thu thập dữ liệu

Nguồn dữ liệu được thu thập từ Kaggle - một nền tảng trực tuyến cho các nhà khoa học dữ liệu và các chuyên gia về máy học, AI và phân tích dữ liệu, nền tảng cung cấp các tài nguyên để các nhà khoa học dữ liệu và các chuyên gia có thể tìm kiếm các tập dữ liệu, tham gia vào các cuộc thi, học hỏi từ những chuyên gia và đồng nghiệp, và tìm kiếm việc làm trong lĩnh vực dữ liệu.

Nguồn dữ liệu có thể được tải theo đường dẫn sau:

<https://www.kaggle.com/datasets/ashishsahani/hospital-admissions-data/data>

1.2.2. Mô tả các cột dữ liệu

Tập dữ liệu Hospital Admissions Data gồm có 3 tệp, nhưng để thực hiện đúng chủ đề nhóm chúng em chỉ trích xuất dữ liệu từ tệp "HDHI Admission data.csv" với mô tả các cột dữ liệu như sau:

Bảng mô tả HDHI Admission data

STT	Tên hiển thị	Tên đầy đủ	Kiểu dữ liệu	Giải thích
1	SNO	Serial Number	Int	
2	MRD No.	Admission Number	Int	
3	D.O.A	Date of Admission	Date	Ngày nhập viện
4	D.O.D	Date of Discharge	Date	Ngày xuất viện
5	AGE	AGE	Int	Tuổi
6	GENDER	GENDER	Varchar(1)	Giới tính
7	RURAL	RURAL(R) /Urban(U)	Varchar(1)	Nơi ở
8	TYPE OF ADMISSION EMERGENCY/OPD	TYPE OF ADMISSION EMERGENCY/OPD	Varchar(1)	Loại tiếp nhận
9	month year	month year	Varchar(1 0)	
10	DURATION OF STAY	DURATION OF STAY	Int	Thời gian trong viện
11	duration of intensive unit stay	duration of intensive unit stay	Int	Thời gian trong phòng hồi sức tích cực
12	OUTCOME	OUTCOME	Varchar(5 0)	
13	SMOKING	SMOKING	Bit	Có hút thuốc
14	ALCOHOL	ALCOHOL	Bit	Có sử dụng rượu bia
15	DM	Diabetes Mellitus	Bit	Có bị tiểu đường
16	HTN	Hypertension	Bit	Có bị cao huyết áp
17	CAD	Coronary Artery Disease	Bit	Có bị bệnh động mạch vành

18	PRIOR CMP	CARDIOMYOPATHY	Bit	Có tiền sử bệnh đau tim
19	CKD	CHRONIC KIDNEY DISEASE	Bit	Có bị bệnh thận mạn tính
20	HB	Haemoglobin	Float	Chỉ số haemoglobin
21	TLC	TOTAL LEUKOCYTES COUNT	Float	Chỉ số bạch cầu
22	PLATELETS	PLATELETS	Int	Chỉ số tiểu cầu
23	GLUCOSE	GLUCOSE	Int	Chỉ số đường trong máu
24	UREA	UREA	Int	Chỉ số u-rê
25	CREATININE	CREATININE	Float	Chỉ số creatinine
26	BNP	B-TYPE NATRIURETIC PEPTIDE	Int	Chỉ số peptide natriuretic loại B
27	RAISED CARDIAC ENZYMES	RAISED CARDIAC ENZYMES	Bit	Có tăng enzym tim
28	EF	Ejection Fraction	Int	Chỉ số phân suất tổng máu
29	SEVERE ANAEMIA	SEVERE ANAEMIA	Bit	Bị thiếu máu nặng
30	ANAEMIA	ANAEMIA	Bit	Bị thiếu máu
31	STABLE ANGINA	STABLE ANGINA	Bit	Có bị đau thắt ngực ổn định
32	ACS	Acute coronary Syndrome	Bit	Có bị hội chứng mạch vành cấp
33	STEMI	ST ELEVATION MYOCARDIAL INFARCTION	Bit	Có bị nhồi máu cơ tim có ST chênh

34	ATYPICAL CHEST PAIN	ATYPICAL CHEST PAIN	Bit	Có bị đau ngực không điển hình
35	HEART FAILURE	HEART FAILURE	Bit	Có bị suy tim
36	HFREF	HEART FAILURE WITH REDUCED EJECTION FRACTION	Bit	Có bị suy tim giảm phân suất tống máu
37	HFNEF	HEART FAILURE WITH NORMAL EJECTION FRACTION	Bit	Có bị suy tim với phân suất bình thường
38	VALVULAR	Valvular Heart Disease	Bit	Có bị hở van tim
39	CHB	Complete Heart Block	Bit	Có bị block tim
40	SSS	Sick sinus syndrome	Bit	Có bị hội chứng nút xoang
41	AKI	ACUTE KIDNEY INJURY	Bit	Có bị tổn thương thận cấp tính
42	CVA INFRACT	Cerebrovascular Accident INFRACT	Bit	Có bị tai biến mạch máu não
43	CVA BLEED	Cerebrovascular Accident BLEED	Bit	Có bị tai biến mạch máu não xuất huyết
44	AF	Atrial Fibrillation	Bit	Có bị rung tâm nhĩ
45	VT	Ventricular Tachycardia	Bit	Có bị nhịp nhanh tâm thất
46	PSVT	PAROXYSMAL SUPRA VENTRICULAR TACHYCARDIA	Bit	Có bị rối loạn nhịp tim
47	CONGENITAL	Congenital Heart Disease	Bit	Có bị bệnh tim bẩm sinh
48	UTI	Urinary tract infection	Bit	Có bị nhiễm trùng đường tiết

				niêu
49	NEURO CARDIOGENIC SYNCOPE	NEURO CARDIOGENIC SYNCOPE	Bit	Có bị ngất do thần kinh phế vị
50	ORTHOSTATIC	ORTHOSTATIC	Bit	Có bị chán thương chỉnh hình
51	INFECTIVE ENDOCARDITIS	INFECTIVE ENDOCARDITIS	Bit	Có bị viêm nội tâm mạc
52	DVT	Deep venous thrombosis	Bit	Có bị huyết khối tĩnh mạch sâu
53	CARDIOGENIC SHOCK	CARDIOGENIC SHOCK	Bit	Có bị sốc tim
54	SHOCK	SHOCK	Bit	Có bị sốc
55	PULMONARY EMBOLISM	PULMONARY EMBOLISM	bit	Có bị thuyên tắc phổi
56	CHEST INFECTION	CHEST INFECTION	Bit	Có bị nhiễm trùng đường phổi

1.3. Mục tiêu và câu hỏi phân tích chính

Mục tiêu chính của dự án này là xây dựng một cơ sở dữ liệu kho dữ liệu từ tập dữ liệu Hospital Admissions Data, cho phép thực hiện các phân tích nghiệp vụ sâu sắc về quá trình nhập và xuất viện của bệnh nhân.

Các câu hỏi phân tích chính mà dự án hướng đến trả lời bao gồm:

- Xu hướng nhập viện và xuất viện thay đổi như thế nào theo thời gian (tháng, năm)?
- Có sự khác biệt nào về thời gian nằm viện giữa các nhóm bệnh nhân khác nhau (theo tuổi, giới tính, loại nhập viện, tiền sử bệnh lý)?
- Những yếu tố nào có mối tương quan với thời gian nằm viện kéo dài hoặc tỷ lệ tái nhập viện?
- Phân bố các bệnh lý đồng mắc phổ biến ở bệnh nhân nhập viện tim mạch là gì?
- Có sự khác biệt nào về kết quả xét nghiệm (HB, TLC, Platelets, Glucose, v.v.) giữa các nhóm bệnh nhân có kết quả điều trị khác nhau?

- Làm thế nào các yếu tố nhân khẩu học và tiền sử bệnh lý ảnh hưởng đến nguy cơ mắc các bệnh tim mạch cụ thể (ví dụ: CAD, STEMI, Heart Failure)?

1.4. Các công cụ được sử dụng

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

CHƯƠNG 2: THIẾT KẾ DATA WAREHOUSE

Trong chương này, nhóm trình bày quá trình xác định yêu cầu nghiệp vụ, thiết kế bảng Fact và Dimension, đồng thời xây dựng mô hình dữ liệu phù hợp để phục vụ phân tích trong môi trường Data Warehouse. Các bước thiết kế đều hướng đến việc hỗ trợ báo cáo và khai thác dữ liệu hiệu quả thông qua quá trình ETL và các công cụ BI.

2.1. Xác định Business Process và yêu cầu nghiệp vụ

Để thiết kế kho dữ liệu phù hợp, trước tiên cần xác định các quy trình nghiệp vụ (Business Process) chính trong bệnh viện có nhu cầu phân tích cao. Hai quy trình trọng tâm được xác định gồm:

Instructions!											
BusinessProcessName	FactTable	FactGrainType	Granularity	Facts	StartDate	EndDate	Anaemia	Habit	Health	PersonGroup	Diagnosis
AdmissionReport	admission_fact	Transaction	One row per patient admission	Admission Type, Outcome	X	X	X	X	X	X	X
DiagnosisReport	diagnosis_fact	Transaction	One row per patient diagnosis	Diagnosis Type, Severity, Diagnosis Status	X	X	X	X	X	X	X

Business Process 1: Admission Report

Các nghiệp vụ liên quan đến nhập viện, xuất viện và đặc điểm bệnh nhân.

Các yêu cầu khai thác chính:

- Thống kê số lượng bệnh nhân nhập/xuất viện theo ngày
- Phân tích bệnh nhân theo thói quen (hút thuốc, uống rượu, v.v.)
- Phân tích bệnh nhân theo tiền sử bệnh lý (béo phì, tiểu đường, thiếu máu...)
- Phân loại bệnh nhân theo mức độ thiếu máu

Table Name		Admission_fact	Home Page						
Table Type	Fact	Admission_fact	Home Page						
Display Name	Admission_fact	dbo							
Database Schema		Admission_fact	Sourced from Covid19						
Table Description			One row per patient admission report details						
Comment			Y						
Size									
Generate Script?									
Column Name		Display Name	Description	Example Values	Display Folder	ETL Rules	Comments		
AnaemiaKey	AnaemiaKey	Key to Anaemia	1, 2, 3	key	Surrogate key pipeline lookup for dimension key lookup using business key				
PersonKey	PersonKey	Key to Person	1, 2, 3	key	Surrogate key pipeline lookup for dimension key lookup using business key				
PersonHealthKey	PersonHealthKey	Key to PersonalHealth	1, 2, 3	key	Surrogate key pipeline lookup for dimension key lookup using business key				
HabitKey	HabitKey	Key to Habit	1, 2, 3	key	Surrogate key pipeline lookup for dimension key lookup using business key				
AdmissionDateKey	AdmissionDateKey	FK to role-play dimension for DimDate	20041123	key	convert datetime to int of format YYYYMMDD				
DischargeDateKey	DischargeDateKey	FK to role-play dimension for DimDate	20041123	key	convert datetime to int of format YYYYMMDD				
MRD No	MRD No	Admission number	234735	Amounts					
TypeOfAdmission	TypeOfAdmission	Type of Admission	E,O	Amounts					
Outcome	Outcome	Outcome	Discharge, Expiry	Amounts					
InsertAuditKey	Insert Audit Key	Key to Audit dimension for row insertion	1, 2, 3	key	Standard auditing				
UpdateAuditKey	Update Audit Key	Key to Audit dimension for row update	1, 2, 3	key	Standard auditing				

Business Process 2: Diagnosis Report

Liên quan đến chẩn đoán bệnh và nguyên nhân bệnh lý. Các yêu cầu phân tích bao gồm:

- Thống kê sốc tim theo giới tính
- Phân tích nguyên nhân gây đột quy
- Số lượng bệnh nhân có rối loạn nhịp tim
- Số lượng ca nhiễm trùng theo thời gian
- Phân tích các vấn đề về mạch máu: thuyên tắc tĩnh mạch sâu, thuyên tắc phổi...
- Số ca nhồi máu cơ tim, đau thắt ngực liên tục và các bệnh đau ngực khác

Table Name	Diagnosis_fact	Home Page					
Table Type	Fact						
Display Name	Diagnosis_fact						
Database Schema	dbo						
Table Description	Diagnosis_fact						
Comment	Sourced from Covid19						
Biz Filter Logic							
Size	One row per patient diagnosis details						
Generate Script?	Y						
Column Name	Display Name	Description	Example Values	Display Folder	ETL Rules	Comments	
DiagnosisReportID	DiagnosisReportID	The natural key for the fact table, if any (eg order number)	1, 2, 3	Exclude from cube			
AnaemiaKey	AnaemiaKey	Key to Anaemia	1, 2, 3	key	Surrogate key pipeline lookup for dimension key lookup using business key		
PersonKey	PersonKey	Key to Person	1, 2, 3	key	Surrogate key pipeline lookup for dimension key lookup using business key		
PersonHealthKey	PersonHealthKey	Key to PersonalHealth	1, 2, 3	key	Surrogate key pipeline lookup for dimension key lookup using business key		
DiagnosisKey	DiagnosisKey	Key to Diagnosis	1, 2, 3	key	Surrogate key pipeline lookup for dimension key lookup using business key		
AdmissionDateKey	AdmissionDateKey	FK to role-play dimension for DimDate	20041123	key	convert datetime to int of format YYYYMMDD		
DischargeDateKey	DischargeDateKey	FK to role-play dimension for DimDate	20041123	key	convert datetime to int of format YYYYMMDD		
MRD No	MRD No	Diagnosis number	234735	Amounts			
InsertAuditKey	Insert Audit Key	Key to Audit dimension for row insertion	1, 2, 3	key	Standard auditing		
UpdateAuditKey	Update Audit Key	Key to Audit dimension for row update	1, 2, 3	key	Standard auditing		

Các yêu cầu phân tích trên cho thấy cần một cấu trúc Data Warehouse linh hoạt, hỗ trợ đa chiều nhằm phục vụ các báo cáo y tế chi tiết và kịp thời.

2.2. Thiết kế bảng Fact và Dimension

Dựa trên yêu cầu phân tích, nhóm tiến hành thiết kế các bảng Dimension và Fact nhằm lưu trữ dữ liệu lịch sử một cách chuẩn hóa và dễ khai thác.

2.2.1. Danh sách các bảng Dimension

Tên bảng	Mô tả
DimPerson	Thông tin bệnh nhân (ID, tuổi, giới tính, ở thành thị hay nông thôn...)
DimComorbidity	Thông tin về tình trạng các bệnh đồng mắc của bệnh nhân
DimPersonHealth	Tiền sử bệnh lý: béo phì, tiểu đường, cao huyết áp...
DimDiagnosis	Kết quả chẩn đoán của bệnh nhân tại mỗi lần nhập viện
DimHabit	Thói quen sinh hoạt: hút thuốc, uống rượu, ăn uống, thể dục...
DimDate	Thời gian nhập viện, xuất viện, chẩn đoán...

Chi tiết các bảng:

- DimPerson:

Instructions!		Attribute / Fact Name	Description	Alternate Names	Sample Values													
Dimension / Fact Table	Table Name	DimPerson	Age	The patient's age	35													
			Gender	The patient's gender	M/F													
			Rural	Whether the patient is living in rural or urban area	RURAL(R) /Urban(U)													
1	Table Name	DimPerson	Home Page															
2	Table Type	Dimension																
3	Display Name	DimPerson																
4	Database Schema	dbo																
5	Table Description	Person Dimension																
6	Comment																	
7	Biz Filter Logic																	
8	Size	One row for each person																
9	Generate Script?	Y																
10																		
11	Column Name	Display Name	Description	Unknown Member	Example Values	SCD Type	Display Folder	ETL Rules	Comments	Target	datatype	size	Precision	Key?	FK To	NULL?	Default Value	
12	PersonKey	PersonKey	Surrogate primary key	-1	1,2,3..	key					int			PK ID	N			
13	PersonID	PersonID	Business key from source system (aka natural key)	-1	1,2,3..	key					int				N			
14	Age	Age	The patient's age	25	2						int				N			
15	Gender	Gender	The patient's gender	M/F	2						nchar		1					
16	Rural	Rural	Whether the patient is living in rural or urban	RURAL(R) /Urban(U)	2						nchar		1		N			
17	RowIsCurrent	Row Is Current	Is this the current row for this member (Y/N)?	1	Y,0	nls	Exclude from cube	Standard SCD-2			bit				N			
18	RowStartDate	Row Start Date	When did this row become valid for this member?	12/31/9999	12/4/2011	nls	Exclude from cube	Standard SCD-2			datetime				Y			
19	RowEndDate	Row End Date	When did this row become invalid? (if current row)	12/31/9999	1/1/1998, 12/31/9999	nls	Exclude from cube	Standard SCD-2			datetime				N			
20	RowChangeReason	Row Change Reason	Why did this row change last?	NA		nls	Exclude from cube	Standard SCD-2			nvarchar	200					12/31/9999	
21	InsertAuditKey	Insert Audit Key	What process loaded this row?	-1		nls	Exclude from cube	Standard Audit dim			int			FK	DimAudit.AuditKey	Y		
22	UpdateAuditKey	Update Audit Key	What process most recently updated this row?	-1		nls	Exclude from cube	Standard Audit dim			int			FK	DimAudit.AuditKey	Y		
23																		
24																		
25																		
26																		
27																		
28																		
29							Comments											
30							This is the template for a dimension table. Use the "Move or Copy Sheet" choice on the Edit menu to make a copy of this sheet before you make any changes to it. Be sure to check the "Create a Copy" box in the dialog box.											
31																		
32																		
33																		
34																		
35																		
36							Usage notes:											
37																		
38																		
39																		
40																		
41																		
42																		
43																		
44																		
45																		
46																		
47																		
48																		
49																		
50																		
51																		
52																		
53																		
54																		
55																		
56																		
57																		
58																		
59																		
60																		
61																		
62																		
63																		
64																		
65																		
66																		
67																		
68																		
69																		
70																		

- DimDiagnosis:

Instructions!	Attribute / Fact Name	Description	Alternate Names	Sample Values
DimDiagnosis	STABLE ANGINA	STABLE ANGINA (Đau thắt ngực)		1/0
	ACS	Acute coronary Syndrome (Hội chứng mạch vành cấp)		1/0
	STEMI	ST ELEVATION MYOCARDIAL INFARCTION (Nhồi máu cơ tim)		1/0
	ATYPICAL CHEST PAIN	ATYPICAL CHEST PAIN (Đau ngực không điển hình)		1/0
	HEART FAILURE	HEART FAILURE (Suy tim)		1/0
	HFREF	HEART FAILURE WITH REDUCED EJECTION FRACTION (Suy tim giảm phân suất tổng máu)		1/0
	HFNEF	HEART FAILURE WITH NORMAL EJECTION FRACTION (Suy tim phân suất tổng máu bình thường)		1/0
	VALVULAR	Valvular Heart Disease (Hở van tim)		1/0
	CHB	Complete Heart Block (Block nhĩ thất)		1/0
	SSS	Sick sinus syndrome (Hội chứng nút xoang)		1/0
	AKI	ACUTE KIDNEY INJURY (Tổn thương thận cấp tính)		1/0
	CVA INFRACT	Cerebrovascular Accident INFRACT (Tai biến mạch máu não)		1/0
	CVA BLEED	Cerebrovascular Accident BLEED (Tai biến mạch máu não xuất huyết)		1/0
	AF	Atrial Fibrillation (Rung tâm nhĩ)		1/0
	VT	Ventricular Tachycardia (Nhịp nhanh tâm thất)		1/0
	PSVT	PAROXYSMAL SUPRA VENTRICULAR TACHYCARDIA (Rối loạn nhịp tim)		1/0
	CONGENITAL	Congenital Heart Disease (Tim bẩm sinh)		1/0
	UTI	Urinary tract infection (Nhiễm trùng đường tiết niệu)		1/0
	NEURO CARDIOGENIC SYNCOP	NEURO CARDIOGENIC SYNCOP (Ngất do dây thần kinh phế vị)		1/0
	ORTHOSTATIC	ORTHOSTATIC (Chấn thương chỉnh hình)		1/0
	INFECTIVE ENDOCARDITIS	INFECTIVE ENDOCARDITIS (Viêm nội tâm mạc)		1/0
	DVT	Deep venous thrombosis (Huyết khối tĩnh mạch sâu)		1/0
	CARDIOGENIC SHOCK	CARDIOGENIC SHOCK (Sốc tim)		1/0
	SHOCK	SHOCK (Sốc)		1/0
	PULMONARY EMBOLISM	PULMONARY EMBOLISM (Thuyên tắc phổi)		1/0
	CHEST INFECTION	CHEST INFECTION (Nhiễm trùng đường phổi)		1/0

- DimHabit:

Instructions!				
Dimension / Fact Table	Attribute / Fact Name	Description	Alternate Names	Sample Values
DimHabit	Smoking	Whether the patient smokes or not		1/0
	Alcohol	Whether the patient uses alcohol or not		1/0

- DimDate:

Instructions!				
Dimension / Fact Table	Attribute / Fact Name	Description	Alternate Names	Sample Values
DimDate	D.O.A	Date of Admission		1/4/2017
	D.O.D	Date of Discharge/Death		3/4/2017
	Month-Year	Admission Month and Year		17-Aug

Table Name	DimDate	Home Page									
Table Type	Dimension										
Display Name	DimDate										
Database Schema	dbo										
Table Description	Date dimension										
Comment	Sourced from Hospital										
Biz Filter Logic											
Size	365 days a year										
Generate Script?	Y										
Column Name	Display Name	Description	Unknown Member	Example Values	SCD Type	Display Folder	ETL Rules	Comments	Datatype	Size	Pr
DateKey	DateKey	Surrogate primary key	-1	20041123		In the form: yyyyymmdd			int		
Date	Date	Full date as a SQL date	0	3/15/2022 10:30		Day			smalldatetime		
DayOfWeek	DayOfWeek	Number of the day of week; Sunday = 1	0	1, 7	1	Day			tinyint		
DayName	DayName	Day name of week eg Monday	Unk date	Sunday	1	Day			varchar	9	
DayOfMonth	DayOfMonth	Number of the day in the month	0	1, 31	1	Day			tinyint		
DayOfYear	DayOfYear	Number of the day in the year	0	1, 365	1	Day			smallint		
WeekOfYear	WeekOfYear	Week of year, 1..53	0	1..52 or 53	1	Calendar			tinyint		
MonthName	MonthName	Month name, eg January	Unk month	November	1	Calendar			varchar	9	
MonthOfYear	MonthOfYear	Month of year, 1..12	0	1, 2, ..., 12	1	Calendar			tinyint		
Quarter	Quarter	Calendar quarter, 1..4	0	1, 2, 3, 4	1	Calendar			tinyint		
QuarterName	QuarterName	Quarter name eg. First	First	First	1	Calendar			varchar	10	
Year	Year	Calendar year, eg 2004							smallint		
IsWeekday	IsWeekday	Is today a weekday							char	7	

2.2.2. Thiết kế bảng Fact

Bảng Fact nhập - xuất viện (FactAdmission)

Mô tả: Bảng này lưu trữ thông tin chi tiết về từng lần nhập và xuất viện của bệnh nhân. Nó ghi lại các sự kiện chính liên quan đến quá trình điều trị tại bệnh viện, bao gồm thông tin cơ bản về lần nhập viện, kết quả và thời gian.

Các thuộc tính chính:

- MRD No: (Có thể hiểu là AdmissionID - Mã số nhập viện). Đây có thể là khóa chính của bảng FactAdmission, định danh duy nhất cho mỗi lần nhập viện.
- TypeOfAdmission: Loại hình nhập viện (ví dụ: 'E' cho Cấp cứu, 'O' cho Ngoại trú/OPD).
- Outcome: Kết quả cuối cùng của lần nhập viện (ví dụ: DISCHARGE, EXPIRY, DAMA).
- D.O.A: (AdmissionDate) - Ngày nhập viện.
- D.O.D: (DischargeDate/DeathDate) - Ngày xuất viện hoặc ngày tử vong.
- Month Year: Tháng và năm nhập viện (có thể được sử dụng cho mục đích phân tích nhanh theo thời gian, nhưng nên liên kết đến bảng DimDate để có khả năng phân tích chi tiết hơn).

Instructions!	Attribute / Fact Name	Description	Alternate Names	Sample Values
Dimension / Fact Table	MRD No	Admission number		234735
FactAdmission	TypeOfAdmission	Type of Admission		E (Emergency)/ O (OPD)
	Outcome	Outcome		DISCHARGE/EXPIRY/DAMA

Bảng Fact chẩn đoán (FactDiagnosis)

Mô tả: Bảng này lưu trữ thông tin chi tiết về các chẩn đoán được đưa ra trong quá trình nhập viện của bệnh nhân. Một lần nhập viện có thể có nhiều chẩn đoán khác nhau.

Các thuộc tính chính:

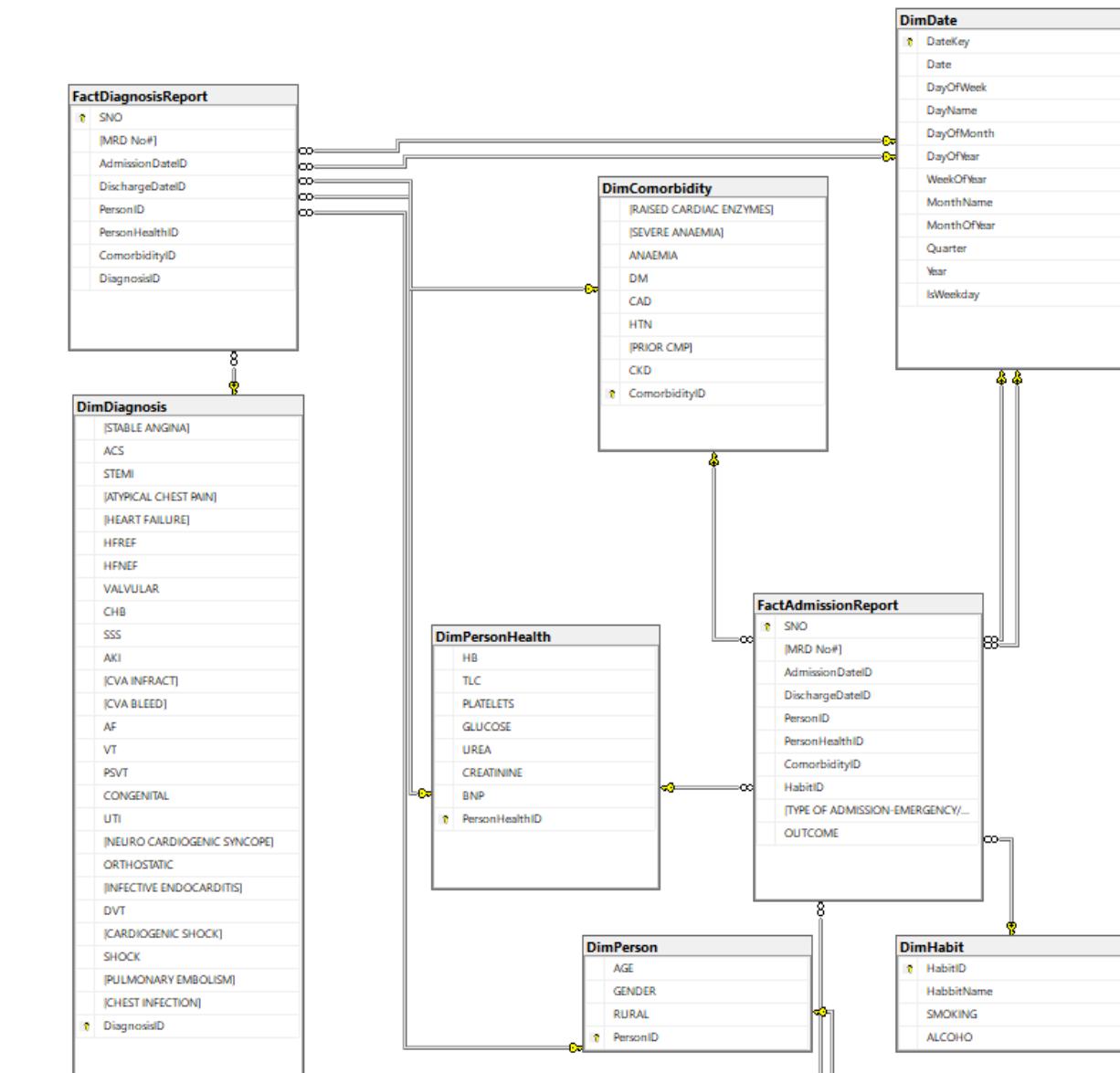
- DiagnosisType: Tên loại chẩn đoán (ví dụ: STEMI, HEART FAILURE, ACS).
- DiagnosisStatus: Trạng thái của chẩn đoán (ví dụ: Confirmed, Suspected).
- Severity: Mức độ nghiêm trọng của chẩn đoán (ví dụ: Nhẹ / Vừa / Nặng).

Instructions!	Dimension / Fact Table	Attribute / Fact Name	Description	Alternate Names	Sample Values
FactDiagnosis	DiagnosisType	Tên loại chẩn đoán		STEMI / HEART FAILURE / ACS	
	DiagnosisStatus	Trạng thái: Confirmed / Suspected		Confirmed	
	Severity	Nhẹ / Vừa / Nặng (nếu có thể xác định)		Moderate	

2.3. Thiết kế mô hình dữ liệu (Star/Galaxy Schema)

Dựa trên hai quy trình nghiệp vụ chính (Admission Report và Diagnosis Report), nhóm chọn **Galaxy Schema** để biểu diễn mối quan hệ giữa các bảng Fact và các bảng Dimension.

Sơ đồ Galaxy Schema giúp giảm dư thừa, tối ưu hóa lưu trữ và hỗ trợ truy vấn phân tích phức tạp.



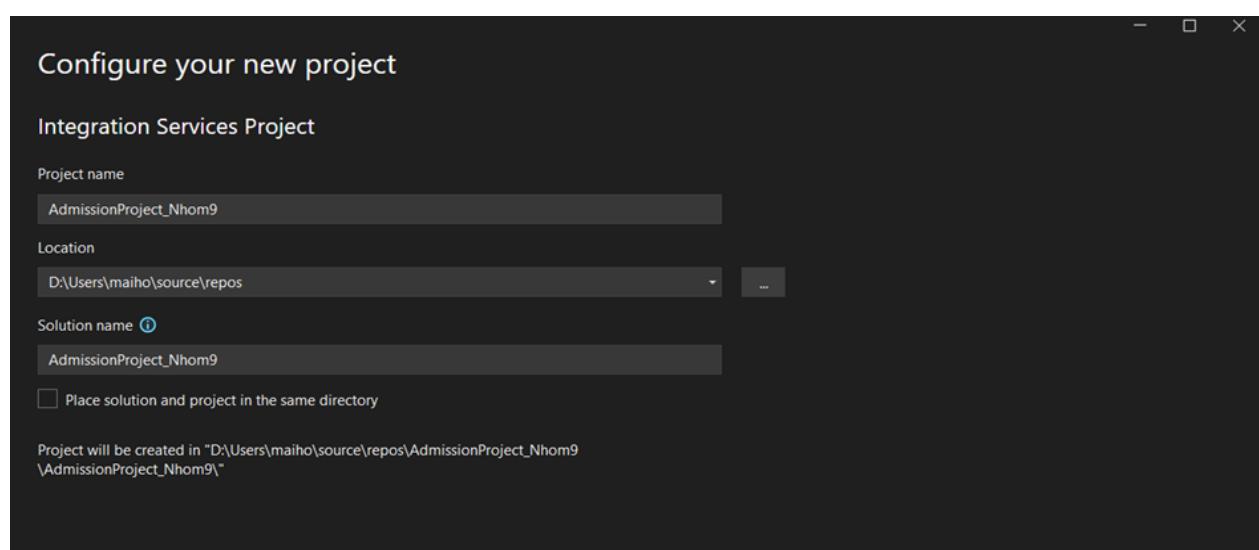
CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO DỮ LIỆU (SSIS)

Chương này trình bày quá trình xây dựng các gói ETL sử dụng công cụ SQL Server Integration Services (SSIS), nhằm trích xuất, chuyển đổi và tải dữ liệu từ hệ thống nguồn vào Data Warehouse theo đúng mô hình đã thiết kế.

3.1. Tạo Solution và Project ETL

Mở Visual Studio -> Chọn File -> New -> Project

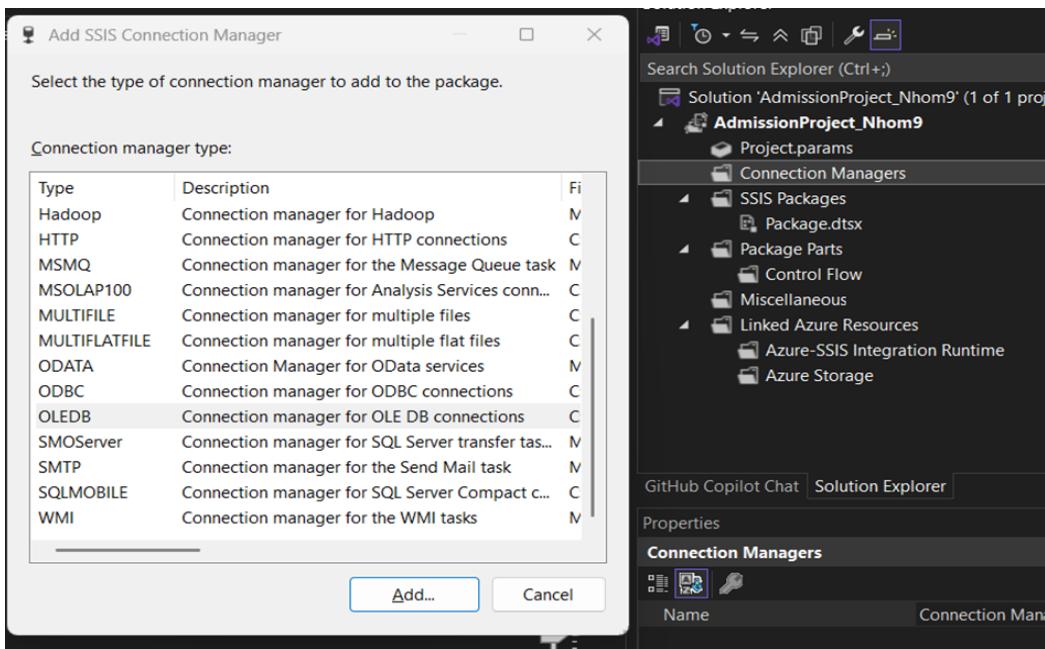
Chọn Integration Services Project → Next, đặt tên cho project là “AdmissionProject_Nhom9” và click Create



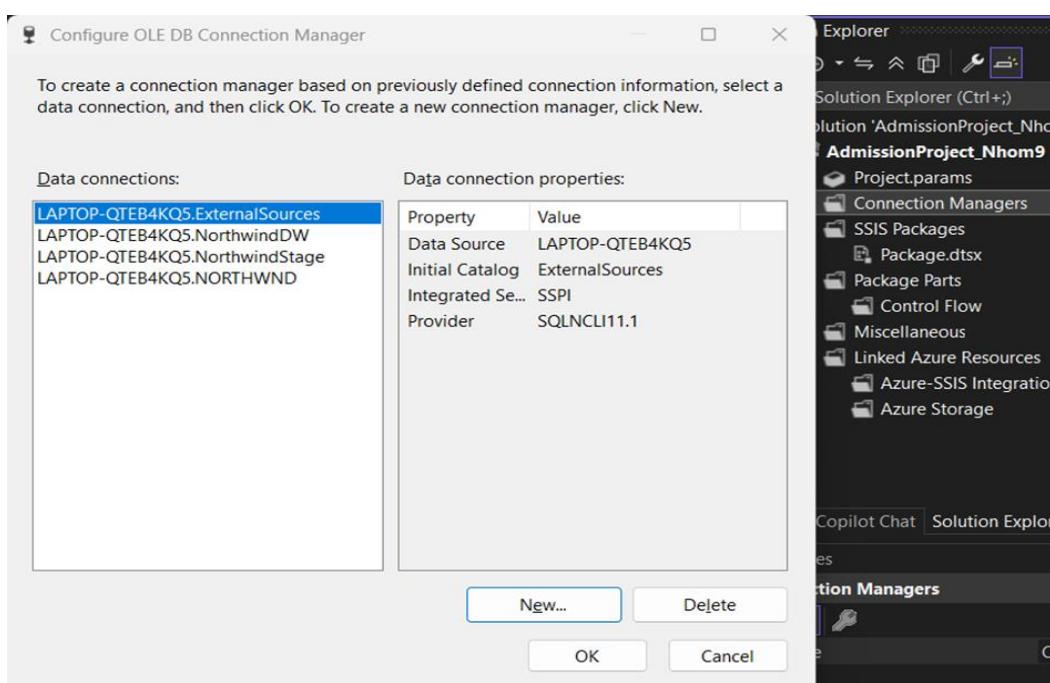
3.2. Thiết lập Connection và Cấu hình nguồn dữ liệu

Để có thể truy xuất dữ liệu cho quá trình ETL, ta cần thiết lập các connection tới nguồn (source) và đích (target). Ta có thể tạo các connection này trong mỗi package. Tuy nhiên, để có thể sử dụng lại các connection trong các package khác nhau, ta sẽ tạo các connection một cách toàn cục (globally).

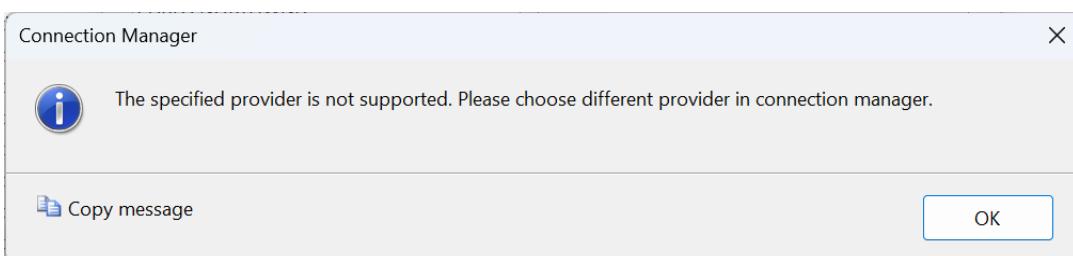
1. Mở Connection Managers: Trong cửa sổ Solution Explorer của dự án SSIS, tìm đến thư mục Connection Managers.
2. Tạo Connection Manager mới: Nhấn chuột phải vào thư mục Connection Managers và chọn New Connection Manager...
3. Chọn loại kết nối OLE DB: Trong hộp thoại Add SSIS Connection Manager, chọn OLEDB từ danh sách các loại kết nối có sẵn. Sau đó, nhấp vào nút Add...



4. Cấu hình OLE DB Connection Manager: Hộp thoại Configure OLE DB Connection Manager sẽ xuất hiện. Nhấp vào nút New... để tạo một cấu hình kết nối mới.

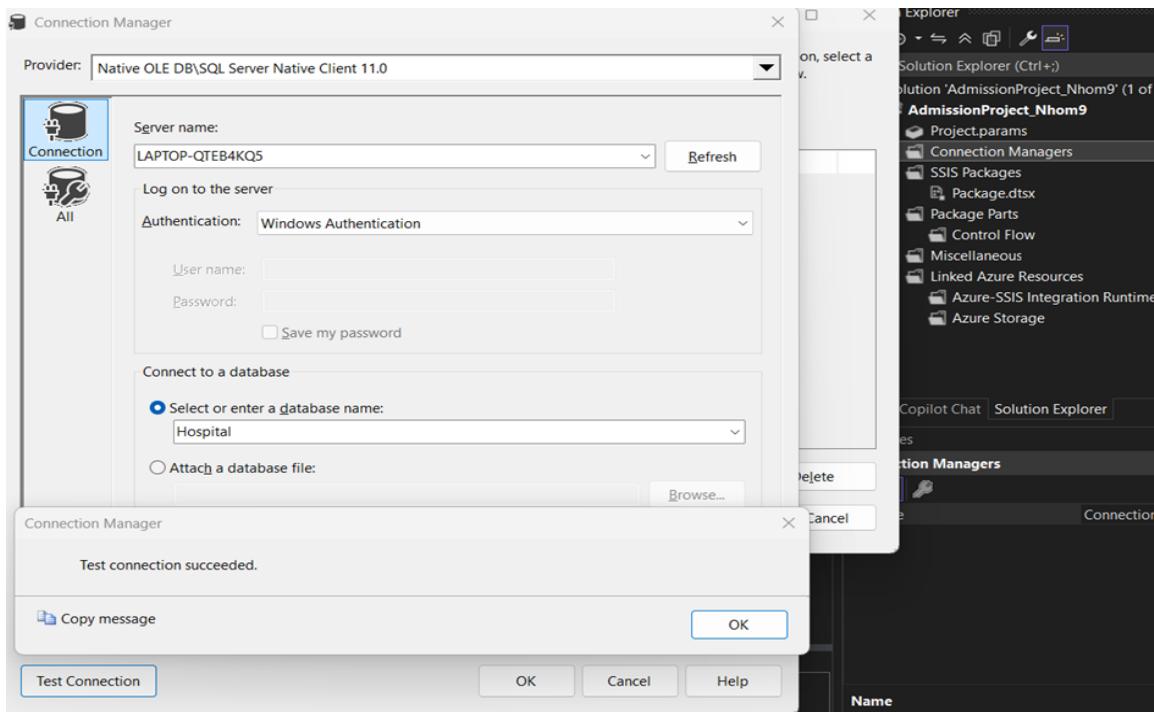


Connection Manager diaglog xuất hiện và bấm OK.

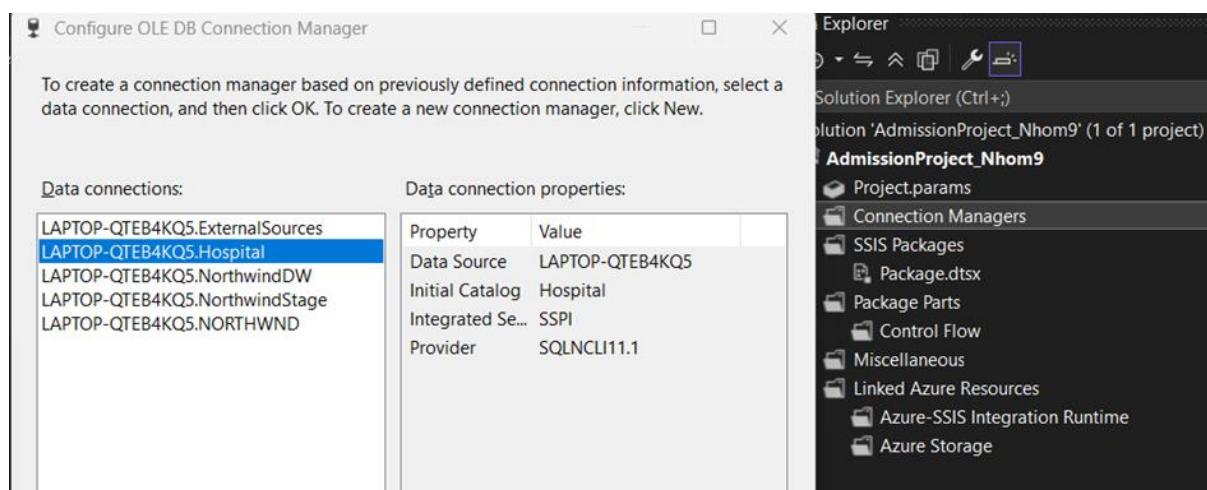


5. Nhập thông tin kết nối: Trong hộp thoại cấu hình kết nối mới, hãy điền các thông tin sau:

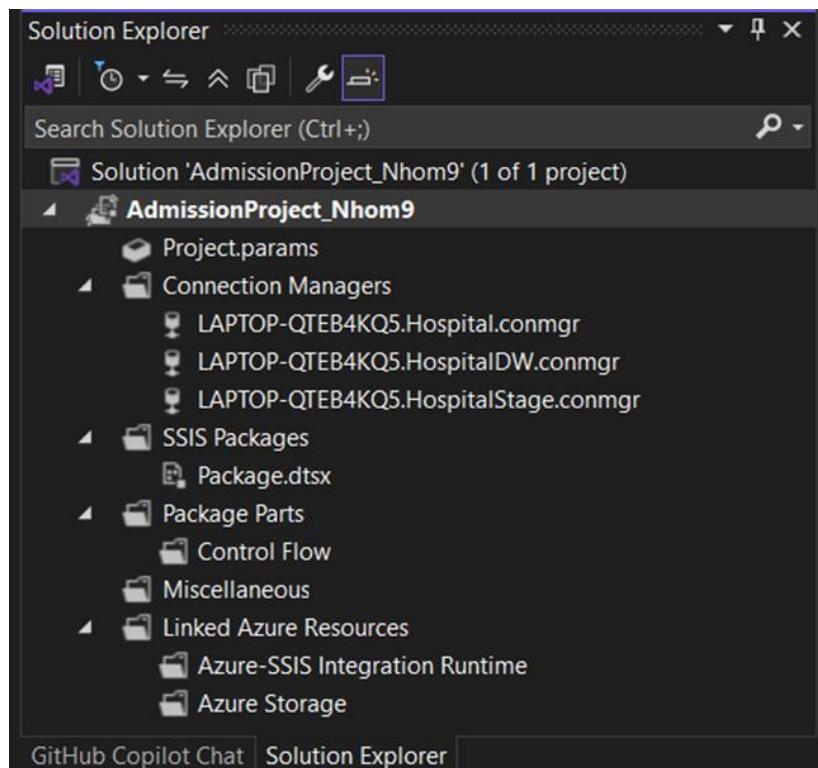
- Provider: Chọn Microsoft OLE DB Driver for SQL Server từ danh sách.
- Server Name: Nhập tên Server SQL Server mà bạn đang sử dụng.
- Log on to the Server: Chọn Use Windows Authentication (nếu bạn sử dụng xác thực Windows) hoặc nhập thông tin đăng nhập SQL Server (nếu bạn sử dụng SQL Server Authentication).
- Connect to a database: Chọn tên database nguồn mà bạn muốn kết nối từ danh sách thả xuông.



6. Kiểm tra kết nối: Nhấp vào nút Test Connection để đảm bảo rằng các thông tin bạn đã nhập là chính xác và kết nối đến cơ sở dữ liệu thành công. Nếu kết nối thành công, một thông báo xác nhận sẽ xuất hiện. Nhấp OK để đóng hộp thoại kiểm tra kết nối



7. Lưu cấu hình kết nối: Trở lại hộp thoại Configure OLE DB Connection Manager, bạn sẽ thấy tên kết nối hiển thị theo định dạng <Tên Server của bạn>. <Tên database của bạn>. Nhấp OK để lưu cấu hình kết nối này và thêm nó vào project của bạn.



3.3. Import Dữ liệu vào Database Gốc

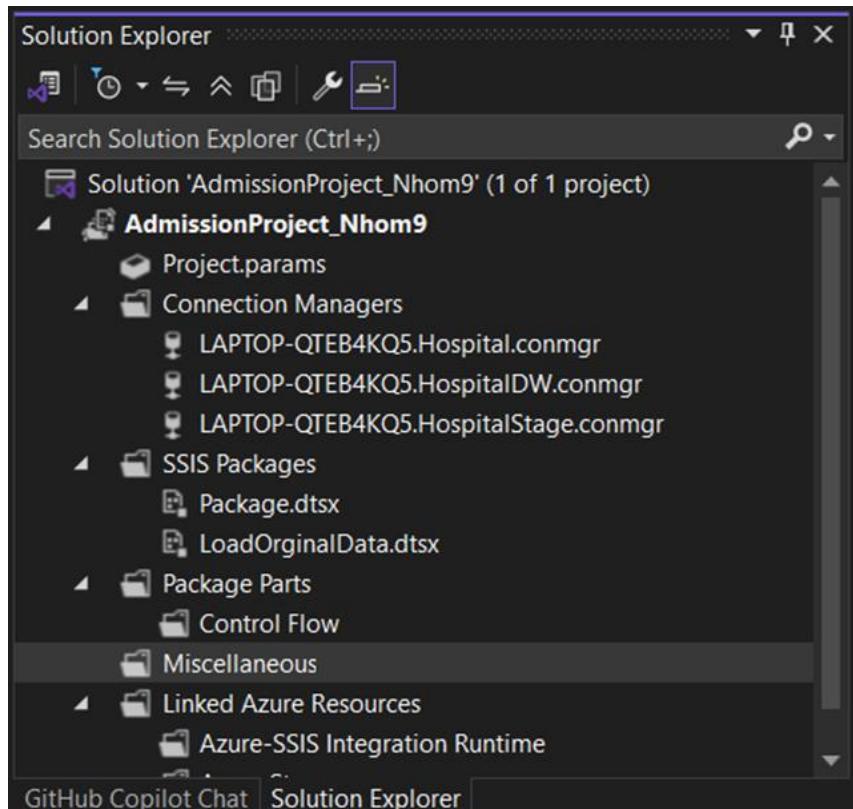
3.3.1. Tạo Package Import

Mục tiêu:

Tạo một package SSIS mới để import dữ liệu từ file Excel vào CSDL gốc (raw database).

Các bước thực hiện:

1. Mở Solution Explorer trong SSDT (SQL Server Data Tools).
2. Click chuột phải vào thư mục SSIS Packages, chọn New SSIS Package.
 - Một package mới sẽ được tạo tự động với tên mặc định là Package1.dtsx.
3. Click chuột phải vào Package1.dtsx, chọn Rename.
4. Đổi tên thành: LoadOriginalData.dtsx.
5. Lúc này trong Solution Explorer bạn sẽ có:
 - Package.dtsx
 - LoadOriginalData.dtsx



6. Double-click vào LoadOriginalData.dtsx để mở giao diện thiết kế

3.3.2. Thiết lập Control Flow

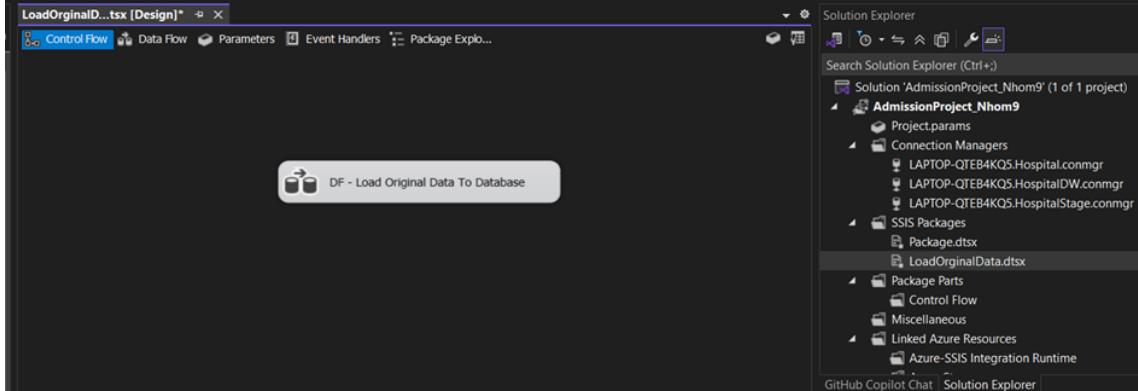
Mục tiêu:

Tạo một Control Flow chứa một Data Flow Task để load dữ liệu.

Các bước thực hiện:

1. Trong cửa sổ SSIS Toolbox, kéo Data Flow Task vào Control Flow surface (mặt thiết kế).
2. Đổi tên Data Flow Task thành:

DF - Load Original Data To Database



3.3.3. Load dữ liệu từ File nguồn vào Database

Mục tiêu:

Thiết kế luồng dữ liệu để:

- Đọc dữ liệu từ Excel
- Chia nhánh điều kiện nếu cần
- Thực hiện xử lý (nếu có)
- Load vào bảng trong Database

Các bước thực hiện:

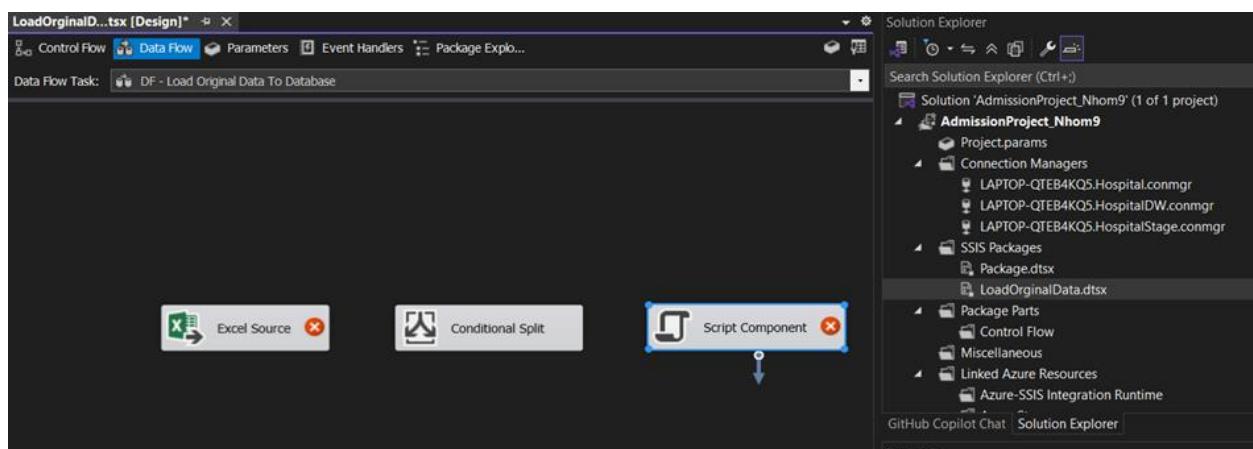
Bước 1: Mở Data Flow Design

Double-click vào task DF - Load Original Data To Database → mở Data Flow tab.

Bước 2: Kéo các thành phần vào Data Flow

Trong SSIS Toolbox, kéo các thành phần sau vào vùng thiết kế:

- Excel Source
- Conditional Split
- Script Component
- + Khi thả Script Component, một hộp thoại sẽ xuất hiện → chọn Transformation, nhấn OK.
- Destination Assistant



Bước 3: Kết nối các thành phần

Kết nối các thành phần bằng cách kéo mũi tên từ output của thành phần trước sang input của thành phần sau:

Excel Source → Conditional Split → Script Component → Destination Assistant

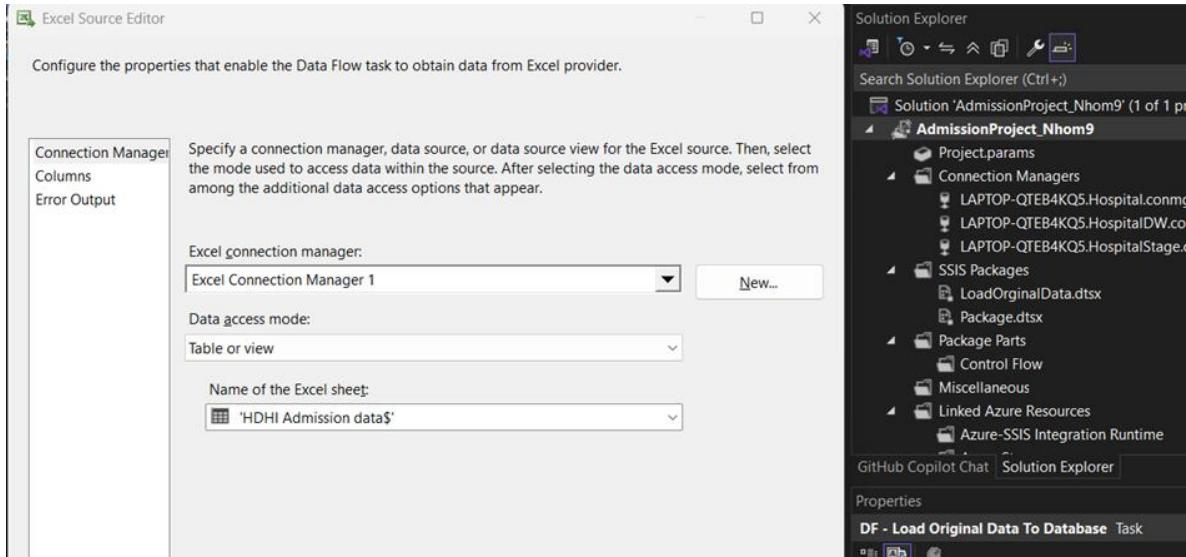
Bước 4: Cấu hình Excel Source

1. Double-click vào Excel Source.

2. Trong Excel Connection Manager:

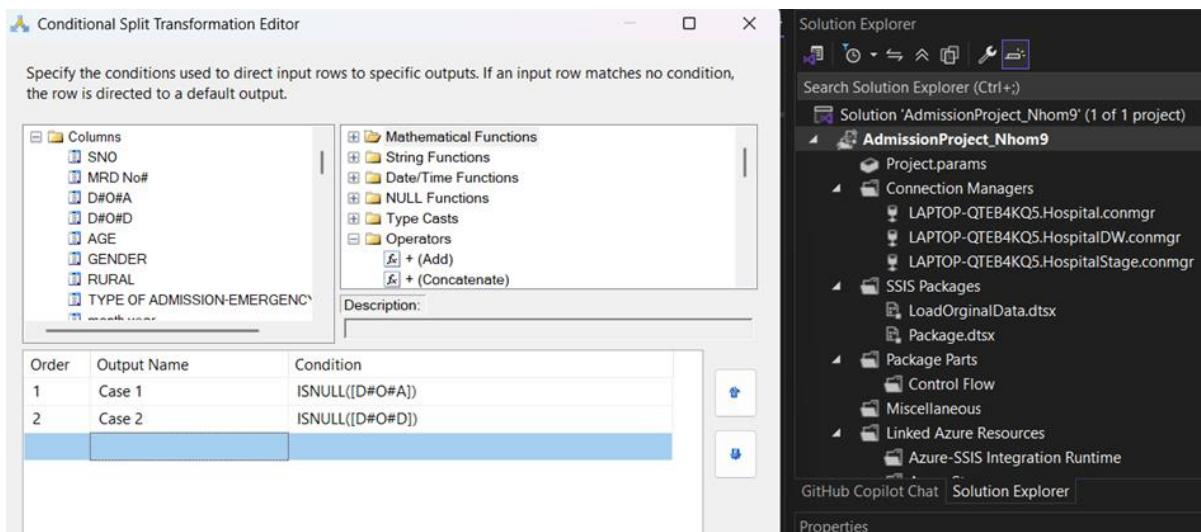
- Chọn hoặc tạo mới kết nối tới file Excel chứa dữ liệu gốc.
- Chọn Sheet chứa dữ liệu (ví dụ: Sheet1\$).

3. Nhấn OK để hoàn tất.



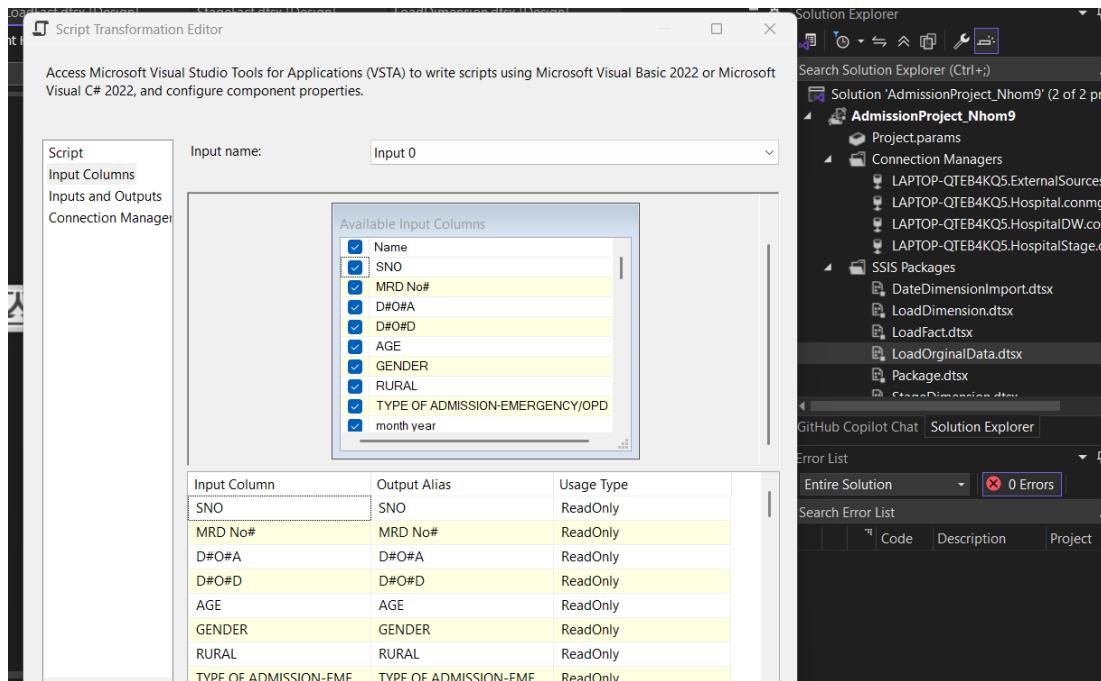
Bước 5: Cấu hình Conditional Split

1. Double-click vào Conditional Split.
2. Định nghĩa các điều kiện phân nhánh (ví dụ phân biệt các dòng hợp lệ/không hợp lệ).
3. Đặt tên các output tương ứng.



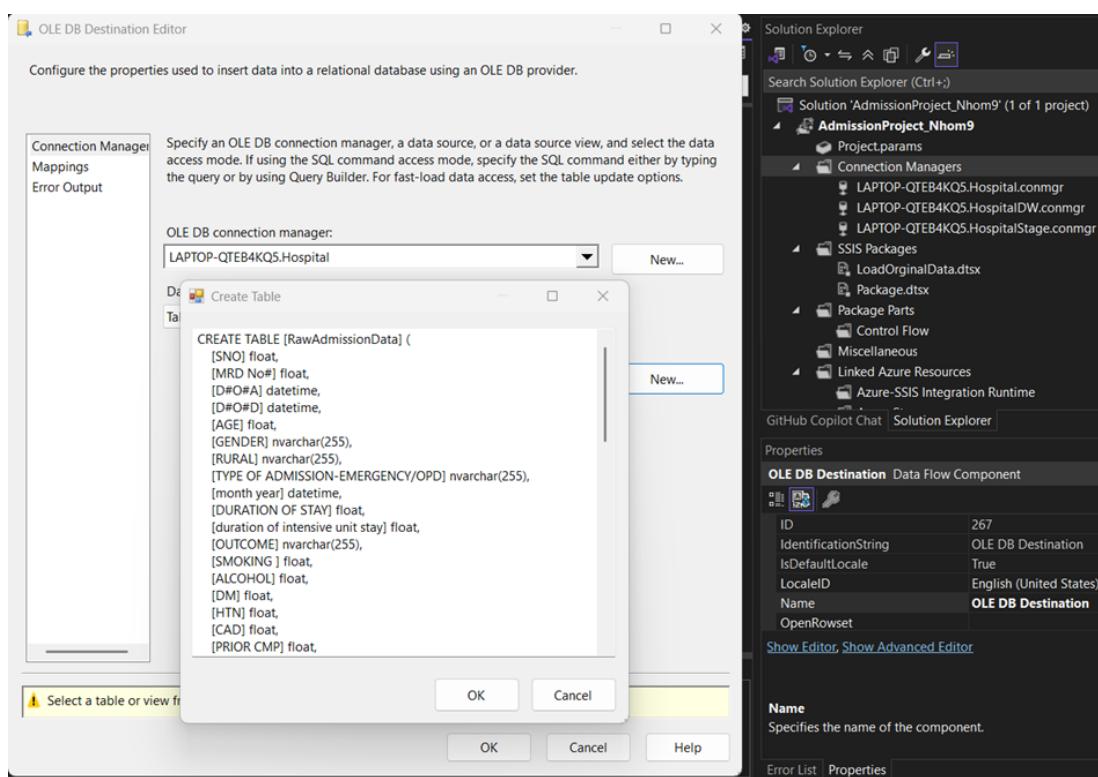
Bước 6: Cấu hình Script Component

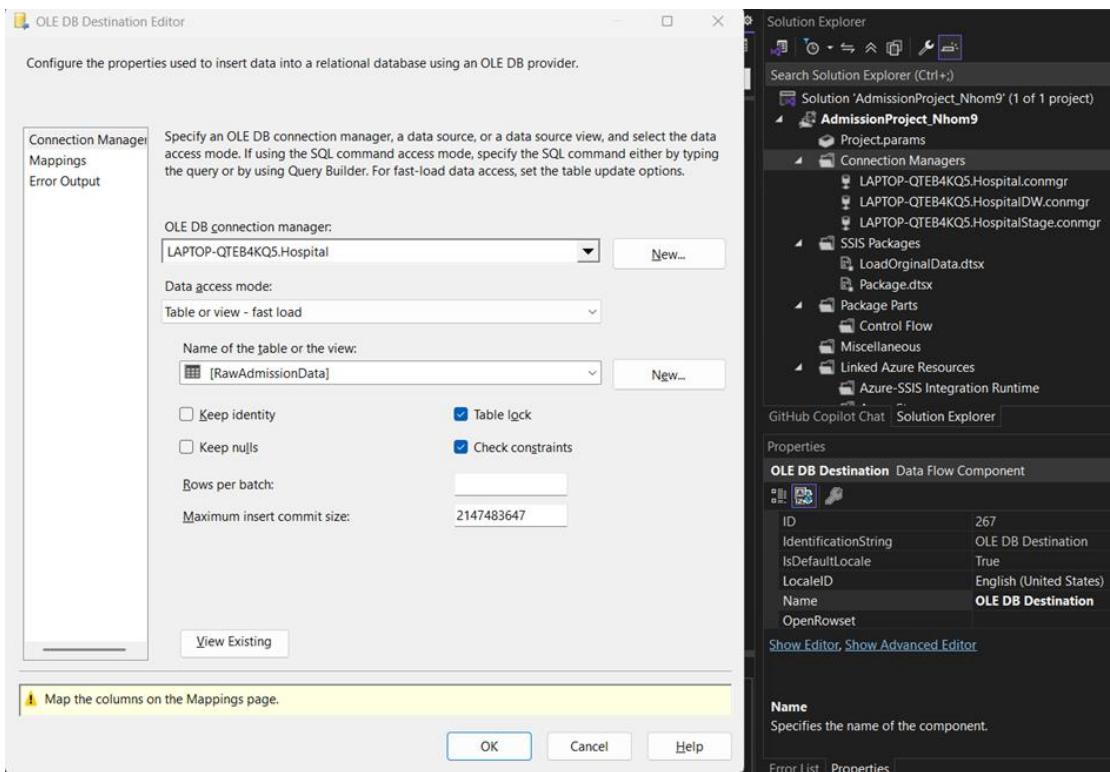
Double-click vào Script Component. Nhấn OK để đóng.



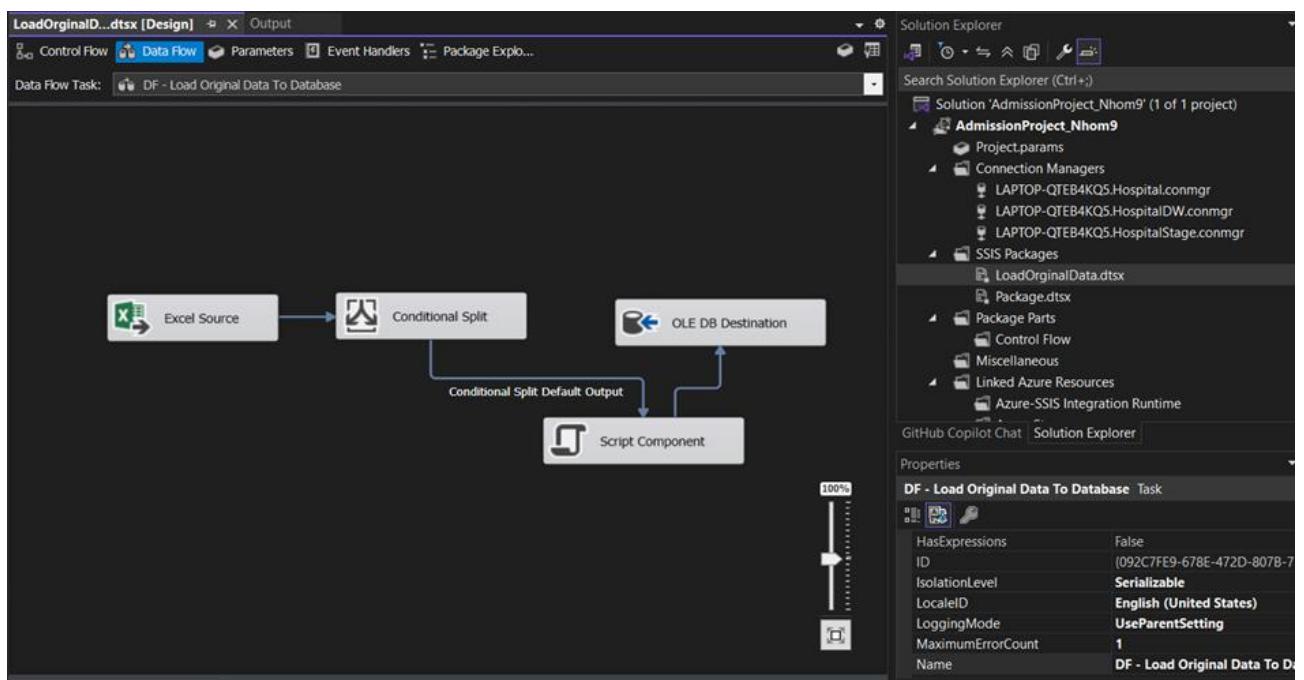
Bước 7: Cấu hình Destination Assistant

1. Double-click vào Destination Assistant.
2. Trong cửa sổ kết nối: Chọn kết nối tới Database gốc hoặc tạo mới.
3. Khi được hỏi chọn bảng đích, nhấn New... để tạo bảng mới.
 - Cửa sổ Create Table xuất hiện.
 - Đổi tên bảng mặc định thành: RawAdmissionData
 - Có thể sửa lại định nghĩa bảng nếu cần (các kiểu dữ liệu,...)

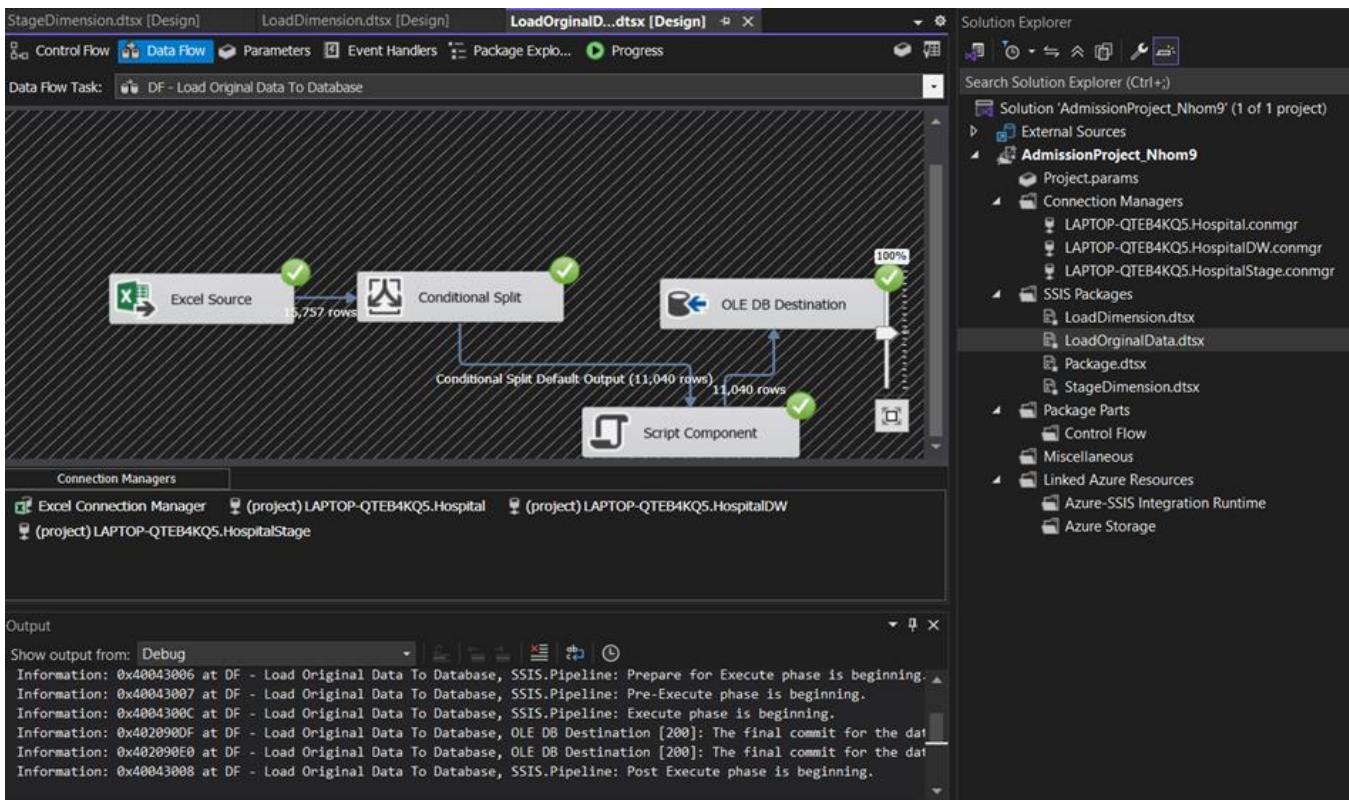




4. Nhấn OK để hoàn tất việc tạo bảng.



5. Chạy start để import data vào database gốc



Kết quả:

The screenshot shows the Object Explorer and a query results grid in SSMS. The query is:

```
SELECT TOP (1000) [SNO]
      ,[MRD No#]
      ,[DHOHA]
      ,[DHOHD]
      ,[AGE]
      ,[GENDER]
      ,[RURAL]
      ,[TYPE OF ADMISSION-EMERGENCY/OPD]
      ,[month year]
      ,[DURATION OF STAY]
      ,[duration of intensive unit stay]
      ,[OUTCOME]
      ,[SMOKING ]
      ,[ALCOHOL]
      ,[DM]
      ,[HTN]
      ,[CAD]
      ,[PRIOR CMP]
```

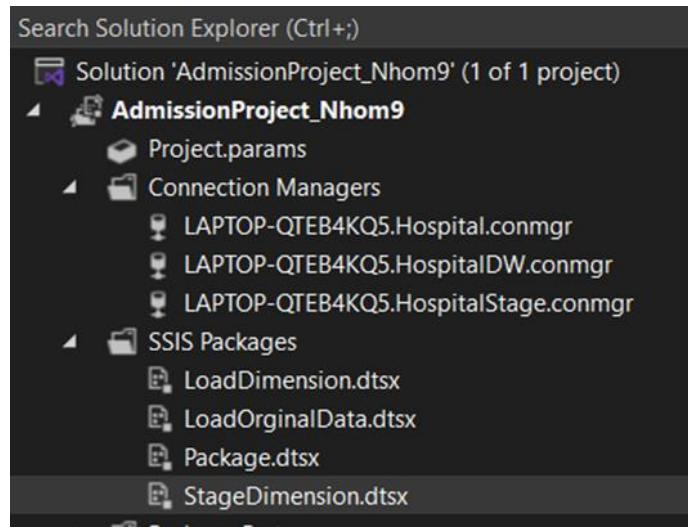
The results grid shows 15 rows of data from the StageDimension table, with columns including SNO, MRD No#, DHOHA, DHOHD, AGE, GENDER, RURAL, TYPE OF ADMISSION-EMERGENCY/OPD, month year, DURATION OF STAY, and duration of ir.

3.4. Tích hợp Dữ liệu vào các bảng Dimension

3.4.1. Tạo Package xử lý Dimension

Tạo 2 SSIS packages:

1. StageDimension.dtsx – Dùng để load dữ liệu từ nguồn vào các bảng staging dimension.
2. LoadDimension.dtsx – Dùng để load dữ liệu từ bảng staging vào các bảng dimension chính thức trong Data Warehouse.

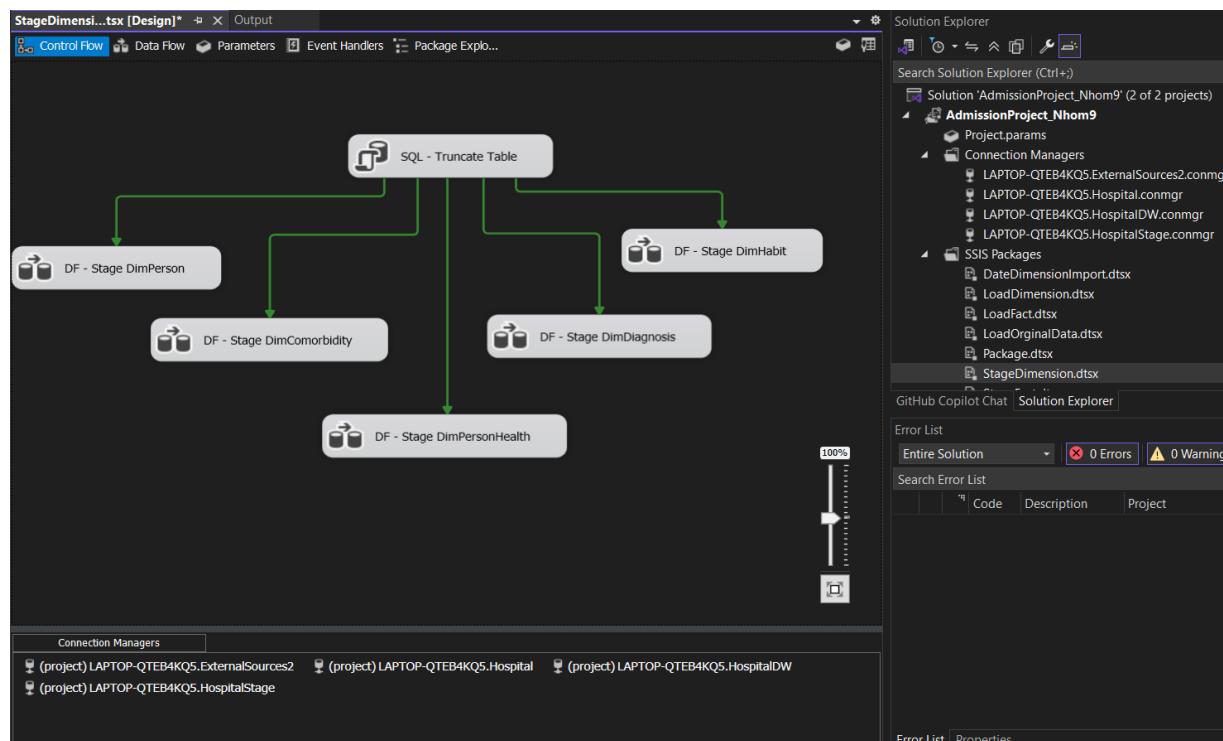


3.4.2. Stage các Dimension

- Click vào **StageDimension.dtsx** và tiến hành thêm vào 1 **Execution SQL Task** và 6 **Data Flow Task** như sau:
- Kéo 1 **Execute SQL Task** → dùng để làm sạch dữ liệu staging (truncate hoặc delete trước khi load).
- Kéo thêm 5 **Data Flow Task** → mỗi task để stage một bảng dimension.

Đổi tên các task: SQL - Truncate Table ,DF - Stage DimPerson, DF - Stage

DimComorbidity , DF - Stage DimPersonHealth, DF - Stage DimDiagnosis, DF - Stage DimHabit

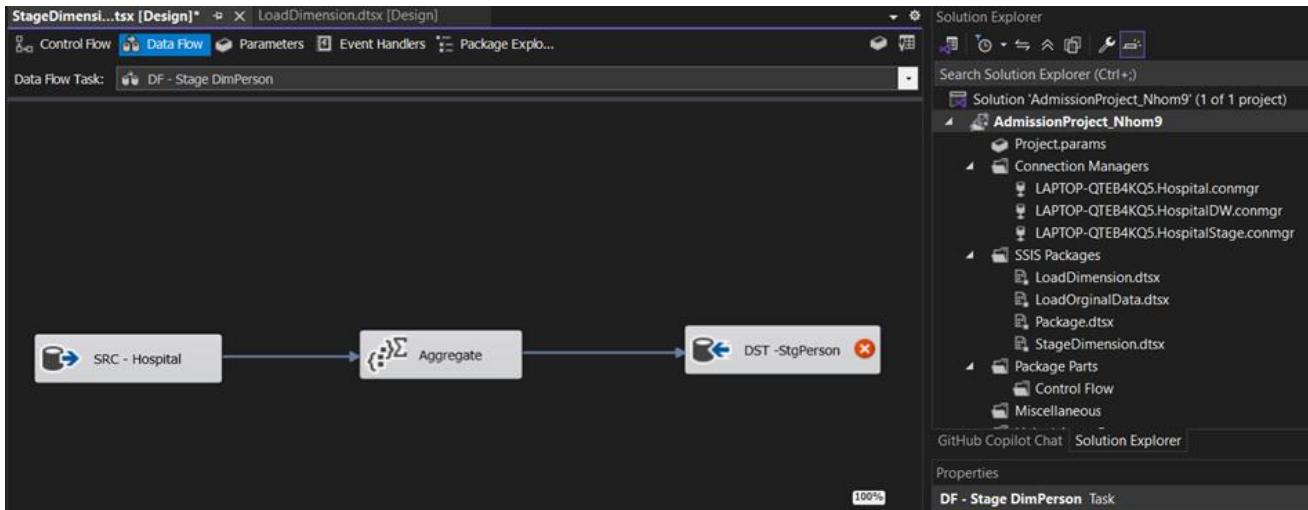


Tiến hành stage các Dimension sau:

StgPerson

1. Click vào DF - Stage DimPerson và thêm Source Assistant, Aggregate và Destination Assistant

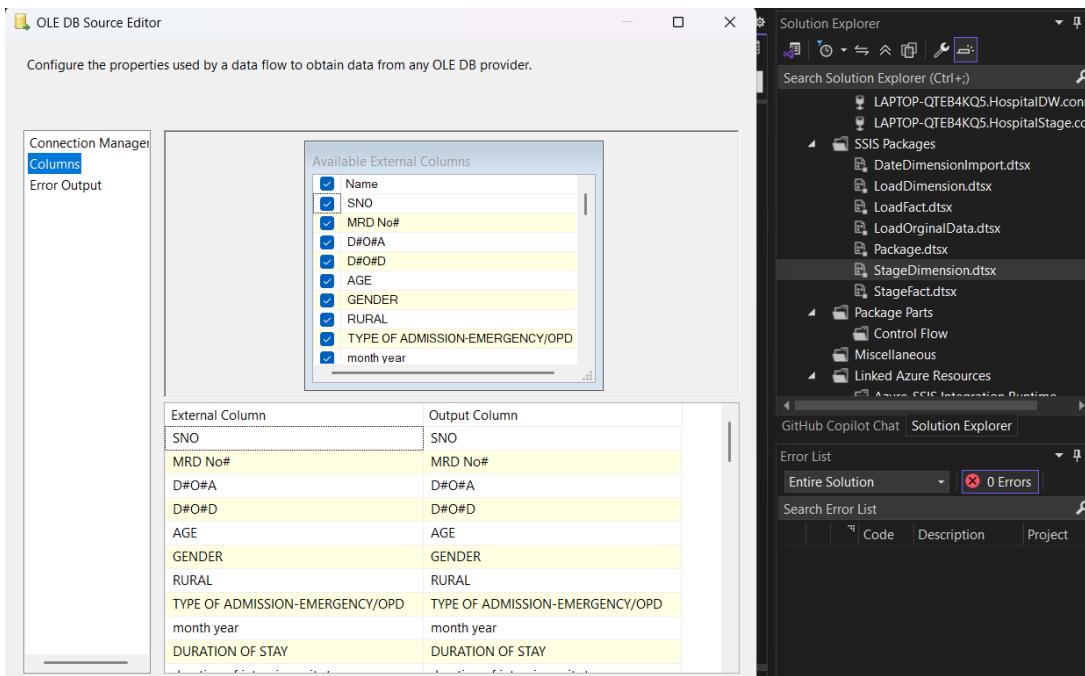
2. Tiến hành đổi tên và ta có được hình như sau:



3. Cấu hình chi tiết:

- Cấu hình SRC - Hospital
 - + Connection Manager: chọn đến CSDL nguồn (OLTP).
 - + Data access mode: Table or View.
 - + Name of the table or view:[dbo] [RawAdmissionData]
 - + Columns: tick các cột bạn cần stage (FirstName, LastName, BirthDate,...)

SNO	MRD No#	D#0#A	D#0#D	AGE	GENDER
1	234735	1/4/20...	3/4/20...	81	M
2	234696	1/4/20...	5/4/20...	65	M
3	234882	1/4/20...	3/4/20...	53	M
4	234635	1/4/20...	8/4/20...	67	F
6	234675	1/4/20...	10/4/20...	44	M
7	234563	1/4/20...	6/4/20...	56	F
9	67070	1/4/20...	3/4/20...	65	F
10	153218	1/4/20...	3/4/20...	59	M
12	232597	1/4/20...	2/4/20...	64	M
14	233403	1/4/20...	2/4/20...	44	M
15	86443	1/4/20...	4/4/20...	62	M
16	413903	1/4/20...	5/4/20...	50	M
17	413903	4/1/20...	4/3/20...	54	M
18	234658	1/4/20...	2/4/20...	58	F
19	219007	1/4/20...	8/4/20...	58	F
20	277983	1/4/20...	6/4/20...	57	M
21	62180	1/4/20...	2/4/20...	35	F
22	323820	1/4/20...	4/4/20...	52	M

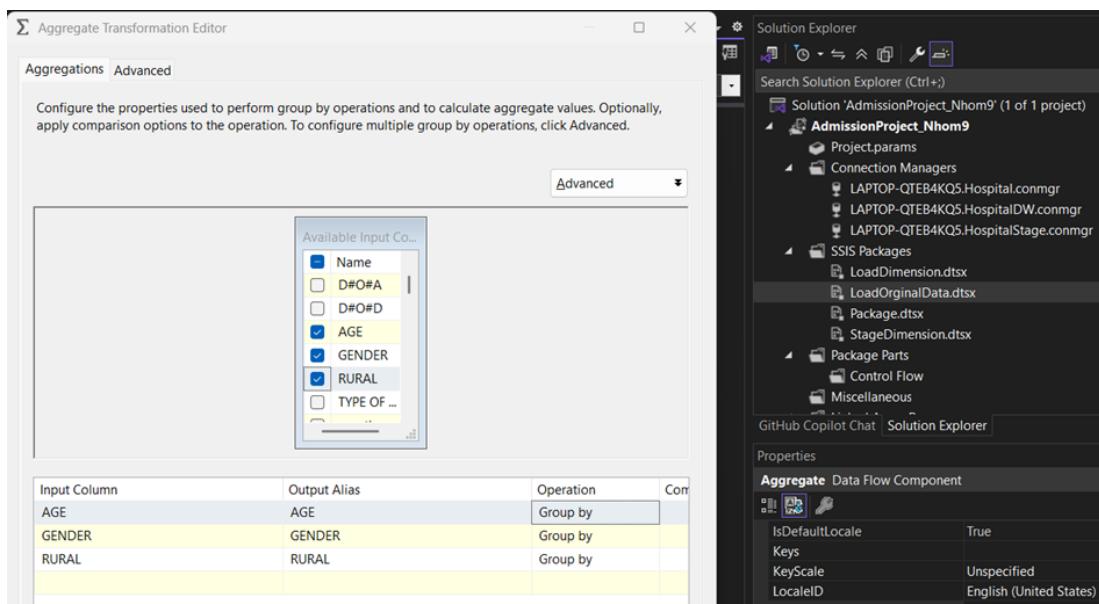


- Cấu hình Aggregate:

Mở cửa sổ Aggregate Editor:

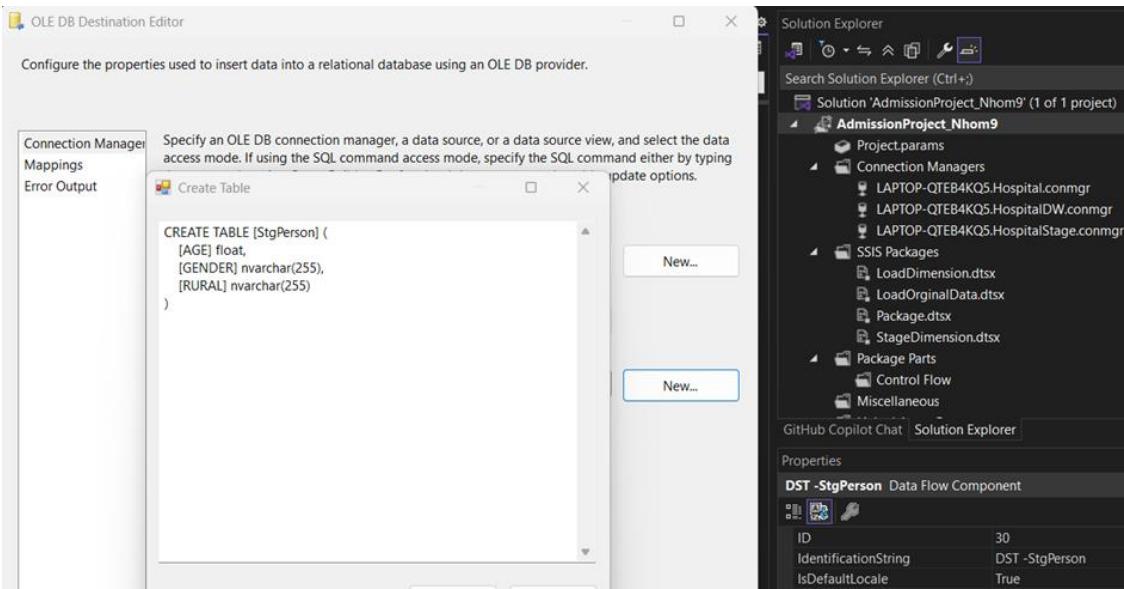
- + Với mỗi cột cần giữ lại, chọn: Group by.

Mục đích: đảm bảo dữ liệu không bị trùng (nếu nguồn có trùng lặp)



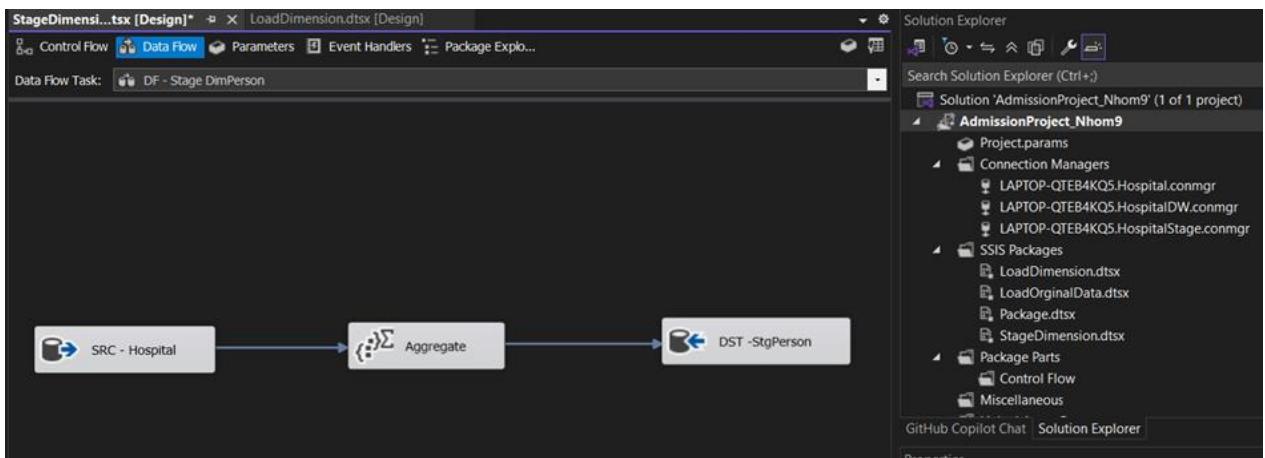
- Cấu hình DST - StgPerson

- + Connection Manager: trỏ tới CSDL staging (Data Warehouse).
- + Data access mode: Table or View - Fast Load.
- + Table: chọn StgDimPerson.
- + Mappings: ánh xạ các cột đầu ra từ Script Component vào bảng staging.



4. Kết quả cuối cùng

Luồng dữ liệu của bạn sẽ như sau:

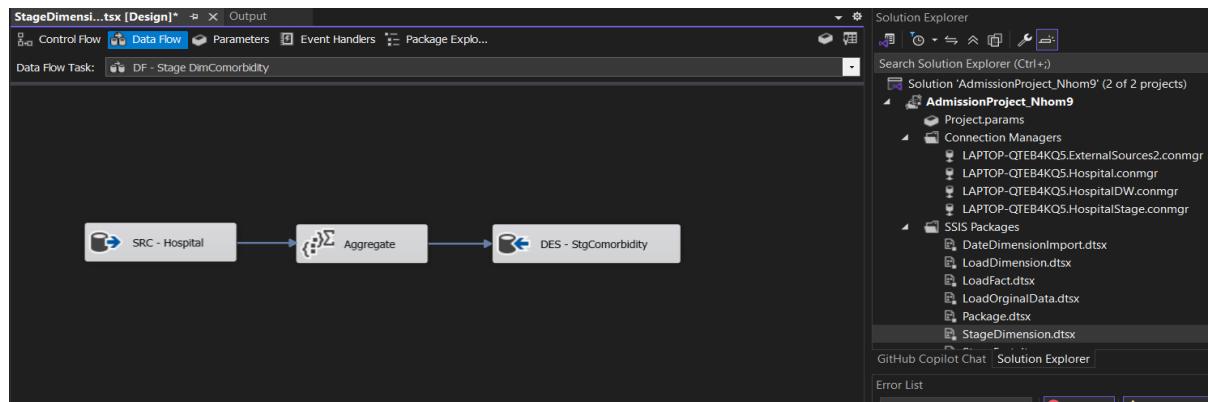


Thực hiện các bước tương tự như StgPerson cho StgComorbidity, StgDiagnosis, StgPersonHealth, StgHabit.

Ta có được các thiết lập như sau:

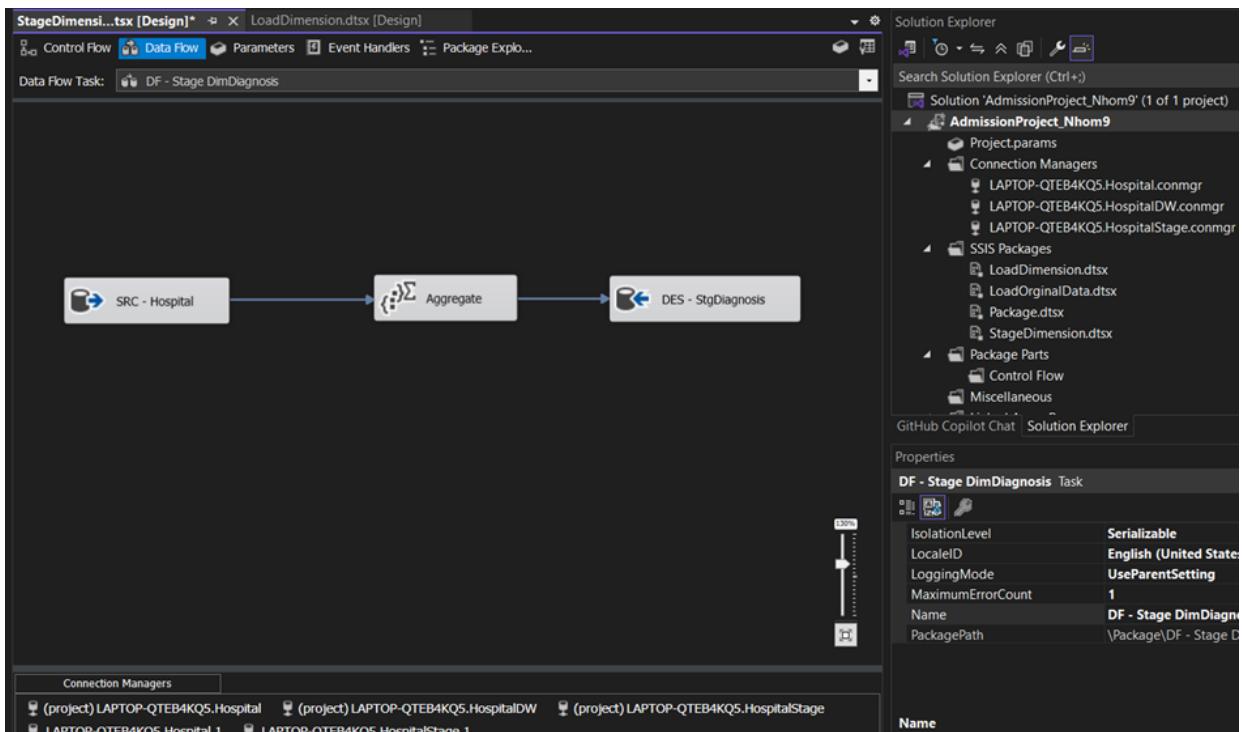
StgComorbidity:

- Luồng dữ liệu sẽ như sau:



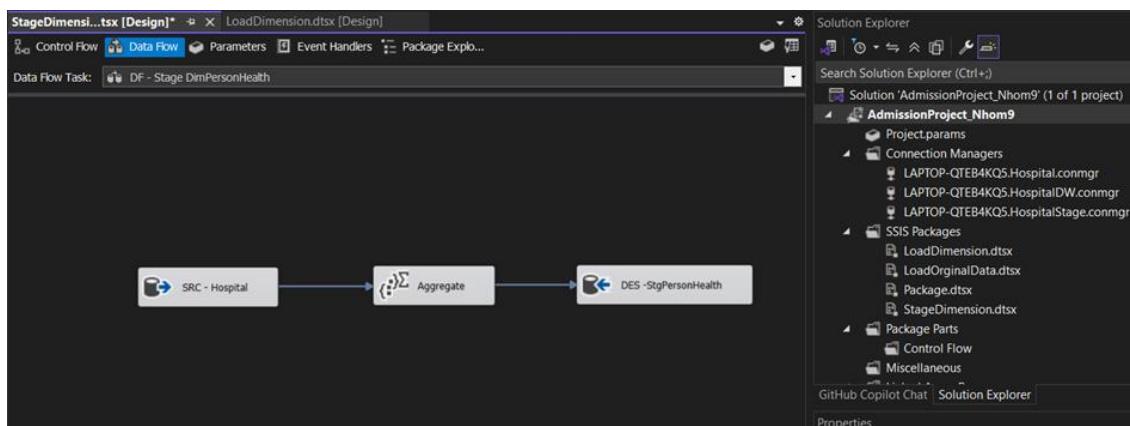
StgDiagnosis:

- Luồng dữ liệu sẽ như sau:



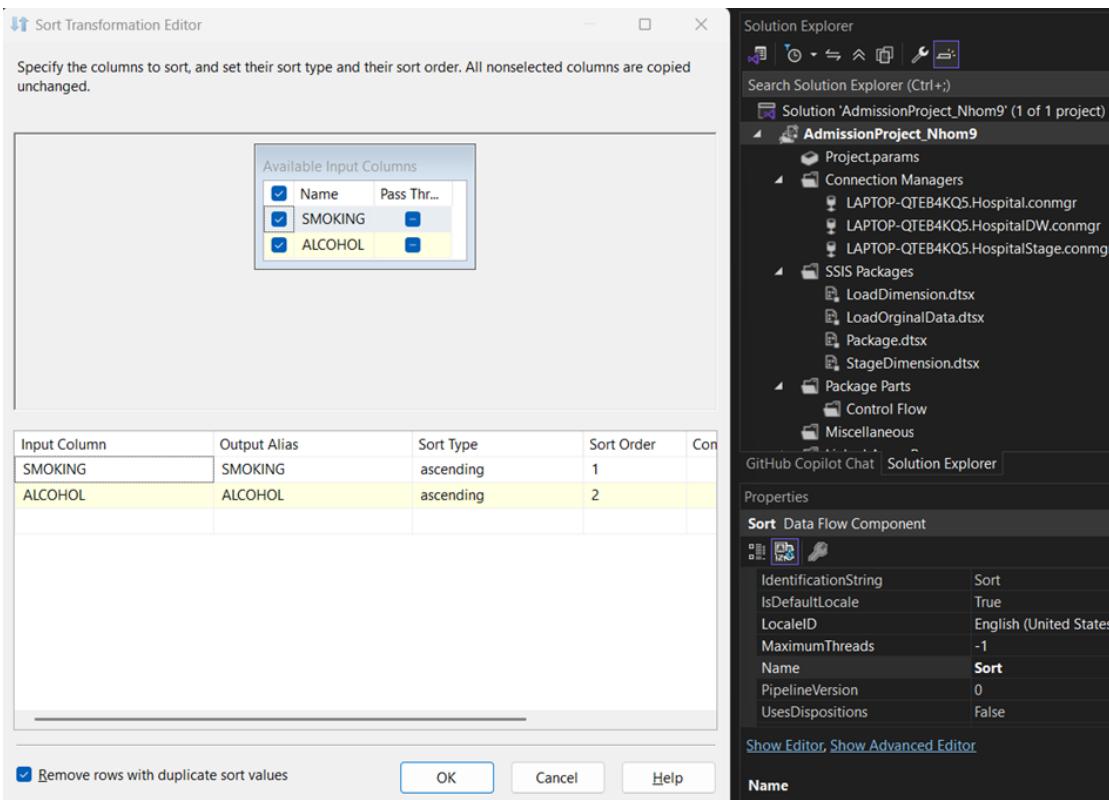
StgPersonHealth:

- Luồng dữ liệu sẽ như sau:

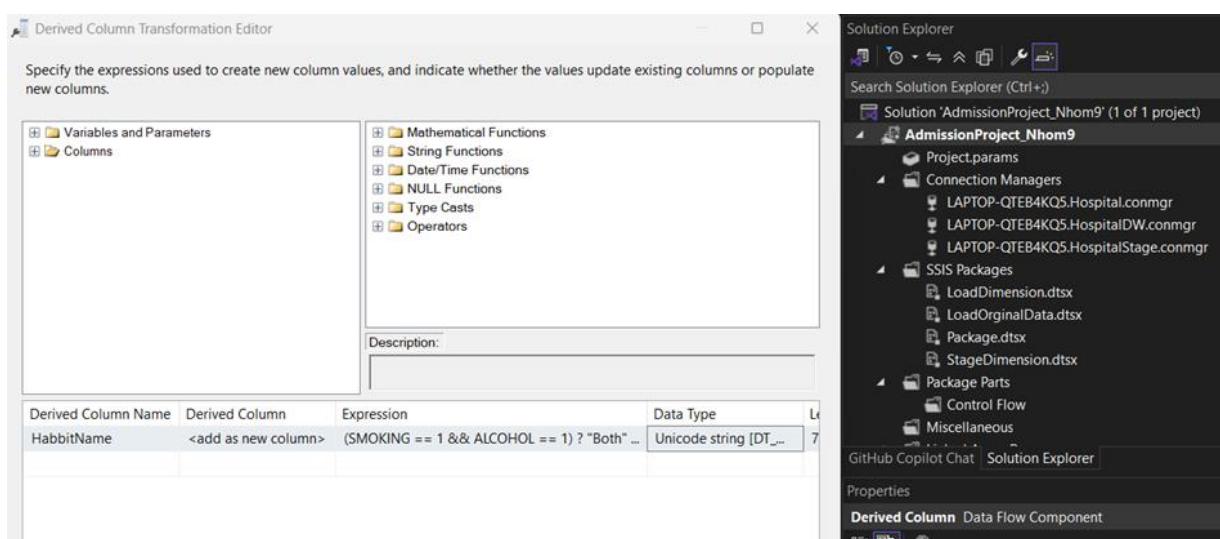


StgHabit: Bổ sung thêm Sort và Derived Column để xét điều kiện tạo cột HabitName;

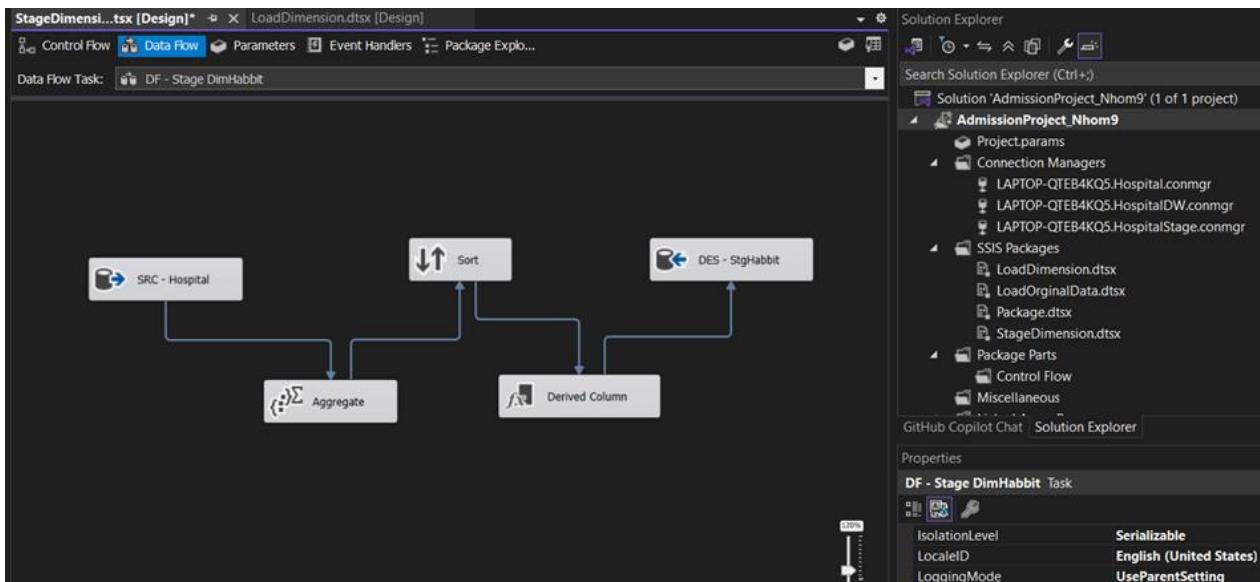
- Cấu hình Sort:
 - + Trong Sort Editor:
Chọn sắp xếp theo SMOKING, ALCOHOL



- Cấu hình Derived Column:
- + Mở cửa sổ Derived Column Transformation Editor.
- + Thêm một cột mới:
Derived Column Name: HabbitName
Expression: (SMOKING == 1 && ALCOHOL == 1) ? "Both" : (SMOKING == 1 && ALCOHOL == 0) ? "Smoke" : (SMOKING == 0 && ALCOHOL == 1) ? "Alcohol" : "None"

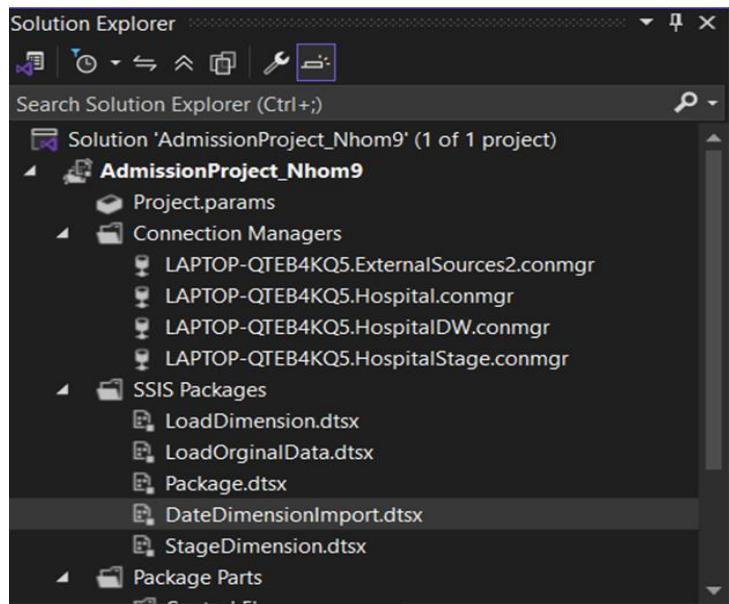


- Luồng dữ liệu sẽ như sau:



StgDate:

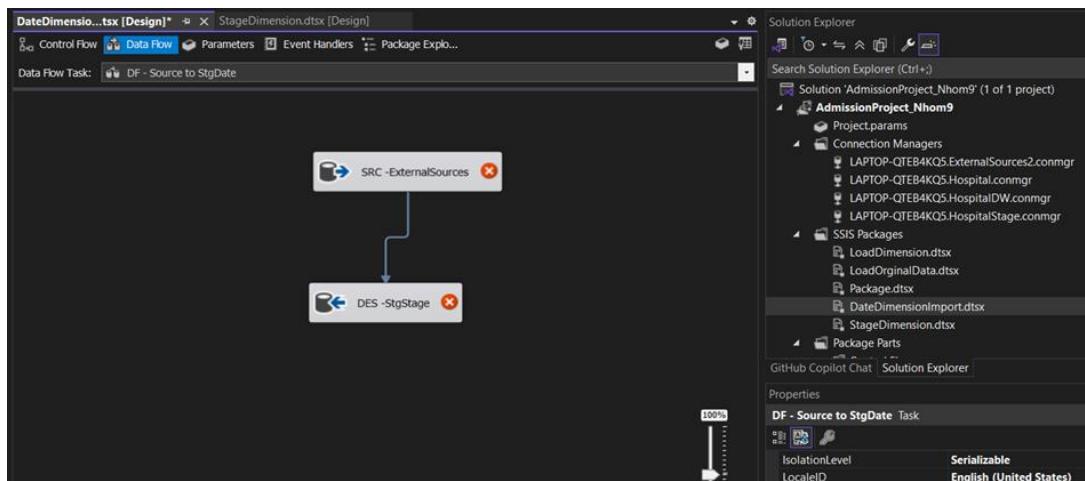
1. Ta tạo một package mới mang tên **DateDimensionImport** đồng thời ta tạo thêm một kết nối mới tới **External Sources** để lấy dữ liệu thời gian.



2. Nhấp chọn **DateDimensionImport** và kéo thả 1 **Execute SQL task** và 2 **Data flow tasks** vào **surface design** và tiến hành đặt tên cho các task. Sau khi hoàn thành ta sẽ có kết quả như hình sau:

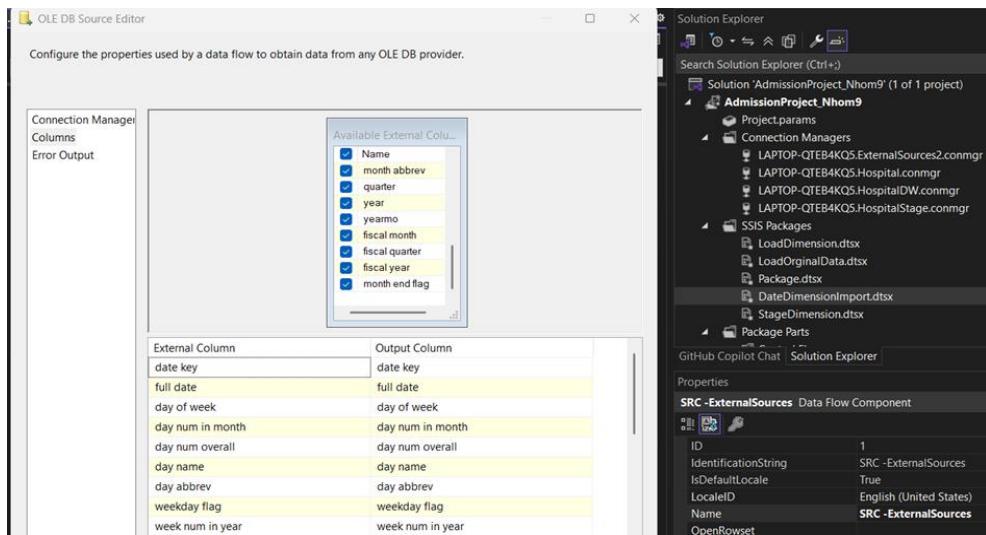


3. Double-click **DF - Source to StgDate** task để mở nó trong Data Flow design surface và thêm vào 1 **Source Assitant** và 1 Destination Assitant. Tiến hành đặt tên và ta sẽ có kết quả như hình sau:



- Cấu hình Source Assistant:

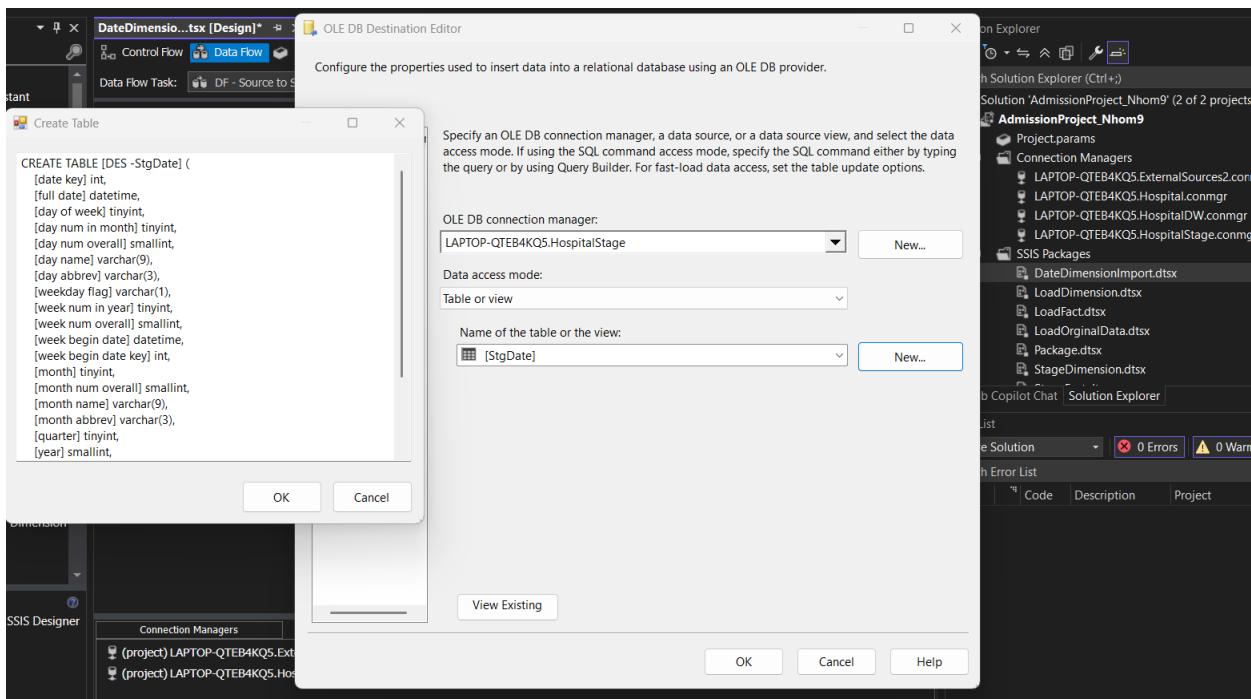
date key	full date	day of ...	day nu... 1	day nu... 2	day name Sunday
201701...	1/1/20...	7	1	1	Sunday
201701...	1/2/20...	1	2	2	Monday
201701...	1/3/20...	2	3	3	Tuesday
201701...	1/4/20...	3	4	4	Wednesday
201701...	1/5/20...	4	5	5	Thursday
201701...	1/6/20...	5	6	6	Friday
201701...	1/7/20...	6	7	7	Saturday
201701...	1/8/20...	7	8	8	Sunday
201701...	1/9/20...	1	9	9	Monday
201701...	1/10/20...	2	10	10	Tuesday
201701...	1/11/20...	3	11	11	Wednesday
201701...	1/12/20...	4	12	12	Thursday
201701...	1/13/20...	5	13	13	Friday
201701...	1/14/20...	6	14	14	Saturday
201701...	1/15/20...	7	15	15	Sunday



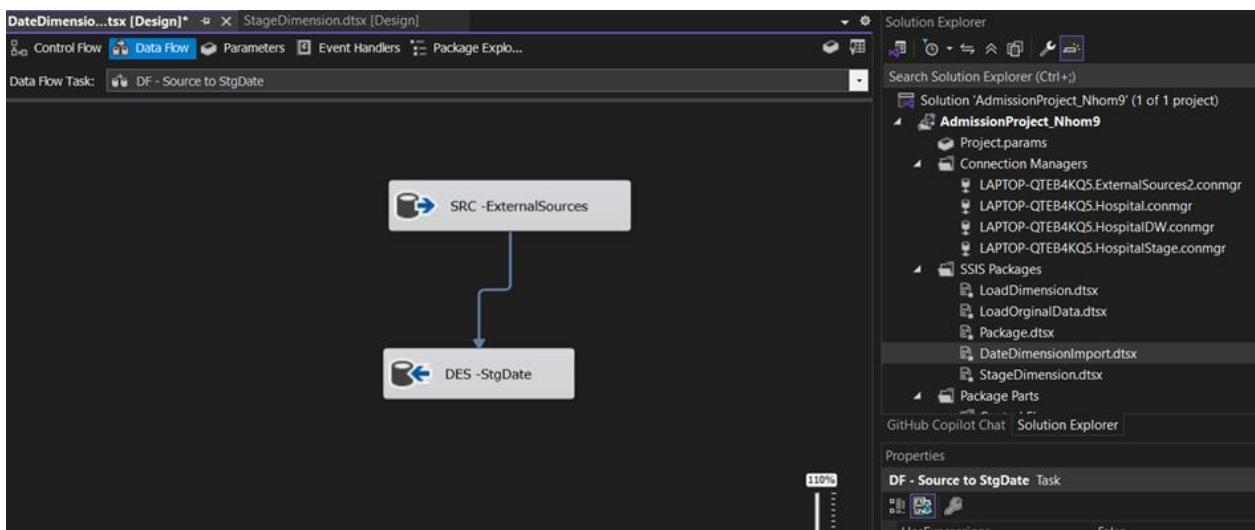
Dữ liệu của bảng được lấy từ file Excel **DateStage.xls** với nội dung dữ liệu như sau

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	date key	full date	day of week	num ir	day num o	day name	day abbrev	weekday f	week num	week num	begin week	begin week	month	month	month	month abt	quarter	year	yearmo	fiscal mon	fiscal quar	fiscal year	month end
2	20170101	01-01-17	7	1	1	Sunday	Sun	N	1	1	42730	20161226	1	1	January	Jan	1	2017	201701	7	3	2017 N	
3	20170102	02-01-17	1	2	2	Monday	Mon	Y	2	2	42737	20170102	1	1	January	Jan	1	2017	201701	7	3	2017 N	
4	20170103	03-01-17	2	3	3	Tuesday	Tue	Y	2	2	42737	20170102	1	1	January	Jan	1	2017	201701	7	3	2017 N	
5	20170104	04-01-17	3	4	4	Wednesday	Wed	Y	2	2	42737	20170102	1	1	January	Jan	1	2017	201701	7	3	2017 N	
6	20170105	05-01-17	4	5	5	Thursday	Thu	Y	2	2	42737	20170102	1	1	January	Jan	1	2017	201701	7	3	2017 N	
7	20170106	06-01-17	5	6	6	Friday	Fri	Y	2	2	42737	20170102	1	1	January	Jan	1	2017	201701	7	3	2017 N	
8	20170107	07-01-17	6	7	7	Saturday	Sat	N	2	2	42737	20170102	1	1	January	Jan	1	2017	201701	7	3	2017 N	
9	20170108	08-01-17	7	8	8	Sunday	Sun	N	2	2	42737	20170102	1	1	January	Jan	1	2017	201701	7	3	2017 N	
10	20170109	09-01-17	1	9	9	Monday	Mon	Y	3	3	42744	20170109	1	1	January	Jan	1	2017	201701	7	3	2017 N	
11	20170110	10-01-17	2	10	10	Tuesday	Tue	Y	3	3	42744	20170109	1	1	January	Jan	1	2017	201701	7	3	2017 N	
12	20170111	11-01-17	3	11	11	Wednesday	Wed	Y	3	3	42744	20170109	1	1	January	Jan	1	2017	201701	7	3	2017 N	
13	20170112	12-01-17	4	12	12	Thursday	Thu	Y	3	3	42744	20170109	1	1	January	Jan	1	2017	201701	7	3	2017 N	
14	20170113	13-01-17	5	13	13	Friday	Fri	Y	3	3	42744	20170109	1	1	January	Jan	1	2017	201701	7	3	2017 N	
15	20170114	14-01-17	6	14	14	Saturday	Sat	N	3	3	42744	20170109	1	1	January	Jan	1	2017	201701	7	3	2017 N	
16	20170115	15-01-17	7	15	15	Sunday	Sun	N	3	3	42744	20170109	1	1	January	Jan	1	2017	201701	7	3	2017 N	
17	20170116	16-01-17	1	16	16	Monday	Mon	Y	4	4	42751	20170116	1	1	January	Jan	1	2017	201701	7	3	2017 N	
18	20170117	17-01-17	2	17	17	Tuesday	Tue	Y	4	4	42751	20170116	1	1	January	Jan	1	2017	201701	7	3	2017 N	
19	20170118	18-01-17	3	18	18	Wednesday	Wed	Y	4	4	42751	20170116	1	1	January	Jan	1	2017	201701	7	3	2017 N	
20	20170119	19-01-17	4	19	19	Thursday	Thu	Y	4	4	42751	20170116	1	1	January	Jan	1	2017	201701	7	3	2017 N	
21	20170120	20-01-17	5	20	20	Friday	Fri	Y	4	4	42751	20170116	1	1	January	Jan	1	2017	201701	7	3	2017 N	
22	20170121	21-01-17	6	21	21	Saturday	Sat	N	4	4	42751	20170116	1	1	January	Jan	1	2017	201701	7	3	2017 N	
23	20170122	22-01-17	7	22	22	Sunday	Sun	N	4	4	42751	20170116	1	1	January	Jan	1	2017	201701	7	3	2017 N	
24	20170123	23-01-17	1	23	23	Monday	Mon	Y	5	5	42758	20170123	1	1	January	Jan	1	2017	201701	7	3	2017 N	
25	20170124	24-01-17	2	24	24	Tuesday	Tue	Y	5	5	42758	20170123	1	1	January	Jan	1	2017	201701	7	3	2017 N	
26	20170125	25-01-17	3	25	25	Wednesday	Wed	Y	5	5	42758	20170123	1	1	January	Jan	1	2017	201701	7	3	2017 N	
27	20170126	26-01-17	4	26	26	Thursday	Thu	Y	5	5	42758	20170123	1	1	January	Jan	1	2017	201701	7	3	2017 N	
28	20170127	27-01-17	5	27	27	Friday	Fri	Y	5	5	42758	20170123	1	1	January	Jan	1	2017	201701	7	3	2017 N	

- Cấu hình Destination Assistant:



4. Data Flow Task của ta sẽ trông như sau:



Cấu hình Execute SQL Task:

1. Vào StageDimension.dtsx → SQL - Truncate Table → SQLStatement điền:

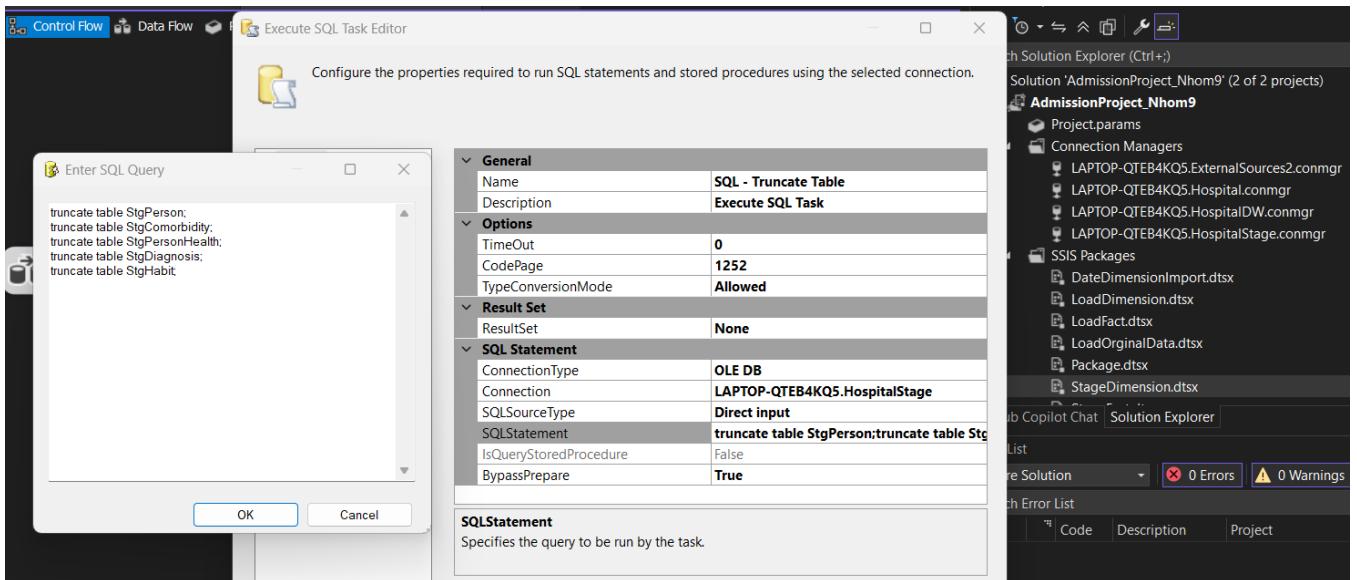
truncate table StgPerson;

truncate table StgComorbidity;

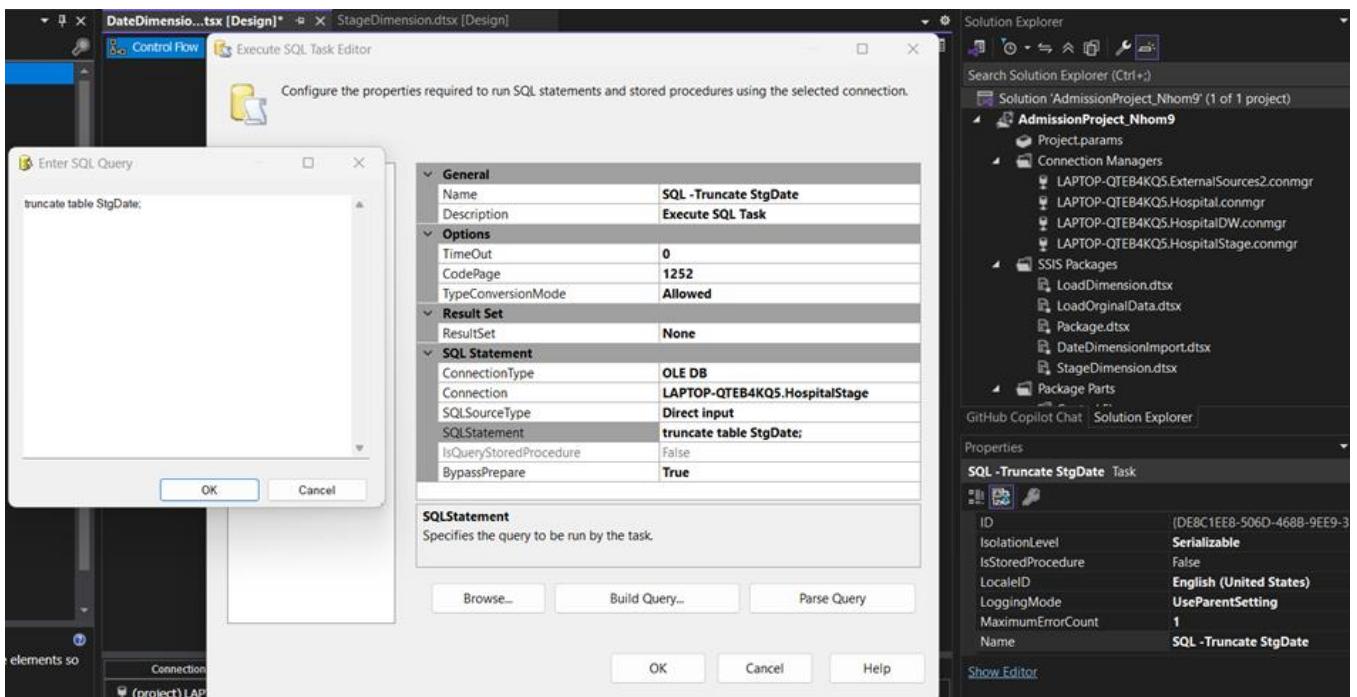
truncate table StgPersonHealth;

truncate table StgDiagnosis;

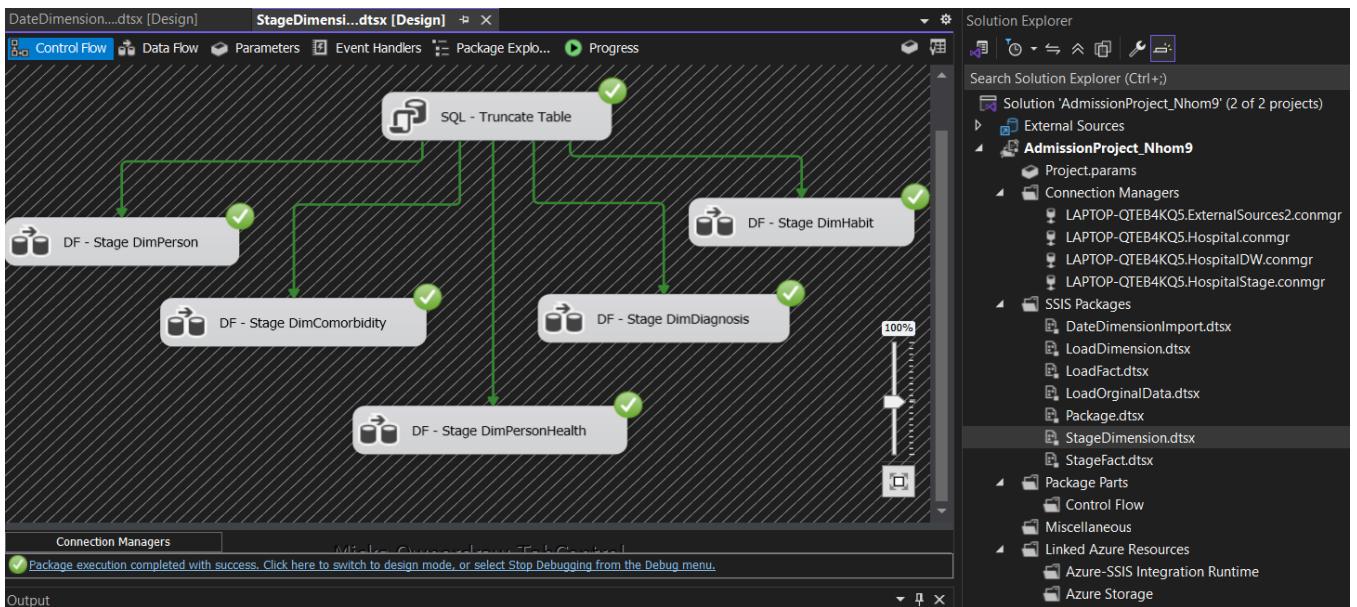
truncate table StgHabit;



- Sau đó vào DateDimensionImport.dtsx điền **truncate table StgDate;** vào SQLStatement trong SQL- Truncate StgDate để Xóa toàn bộ dữ liệu cũ trong bảng Staging StgDate trước khi nạp dữ liệu mới vào

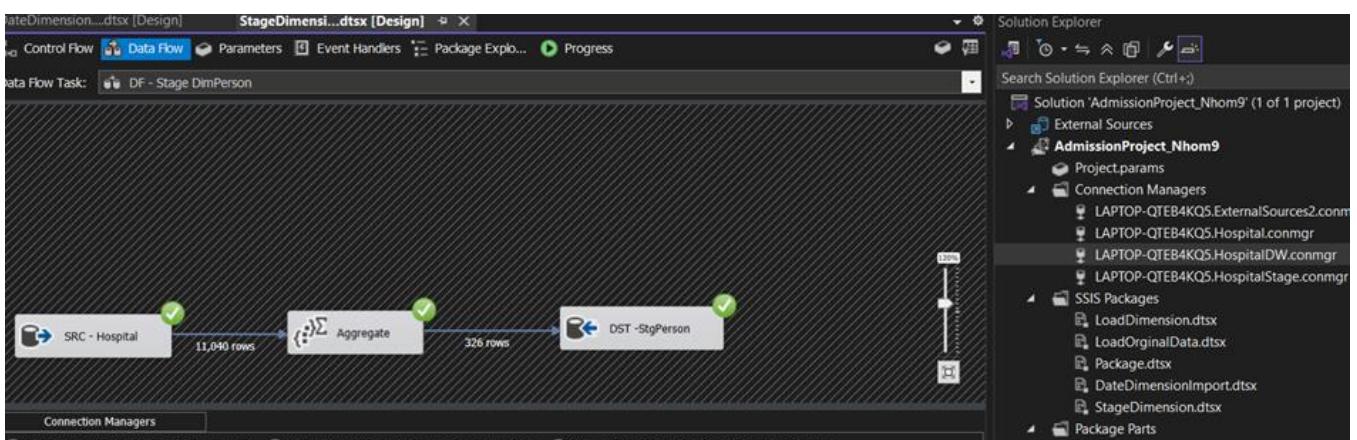


Tiến hành chạy Stage dữ liệu

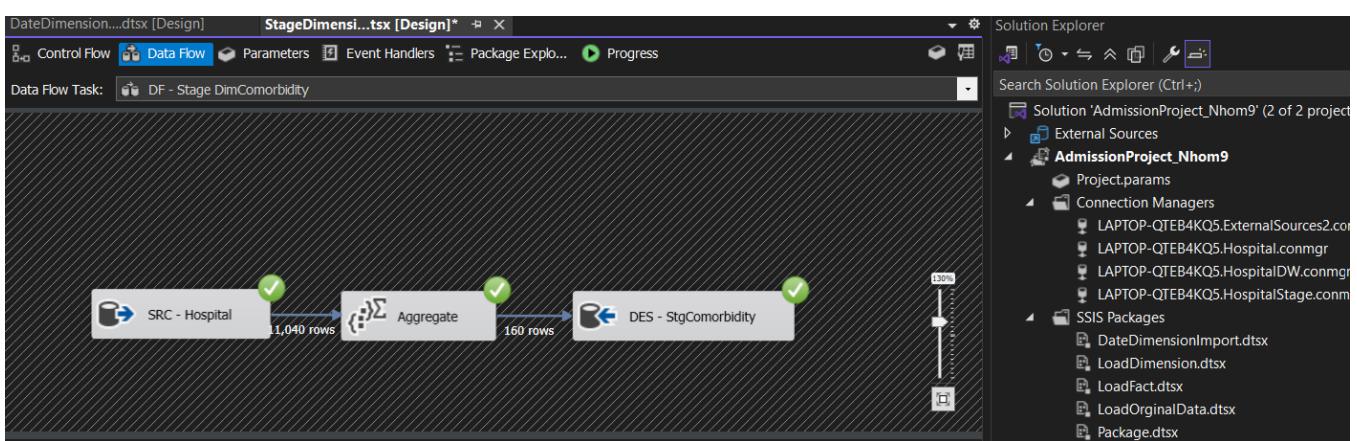


Chi tiết các Stg:

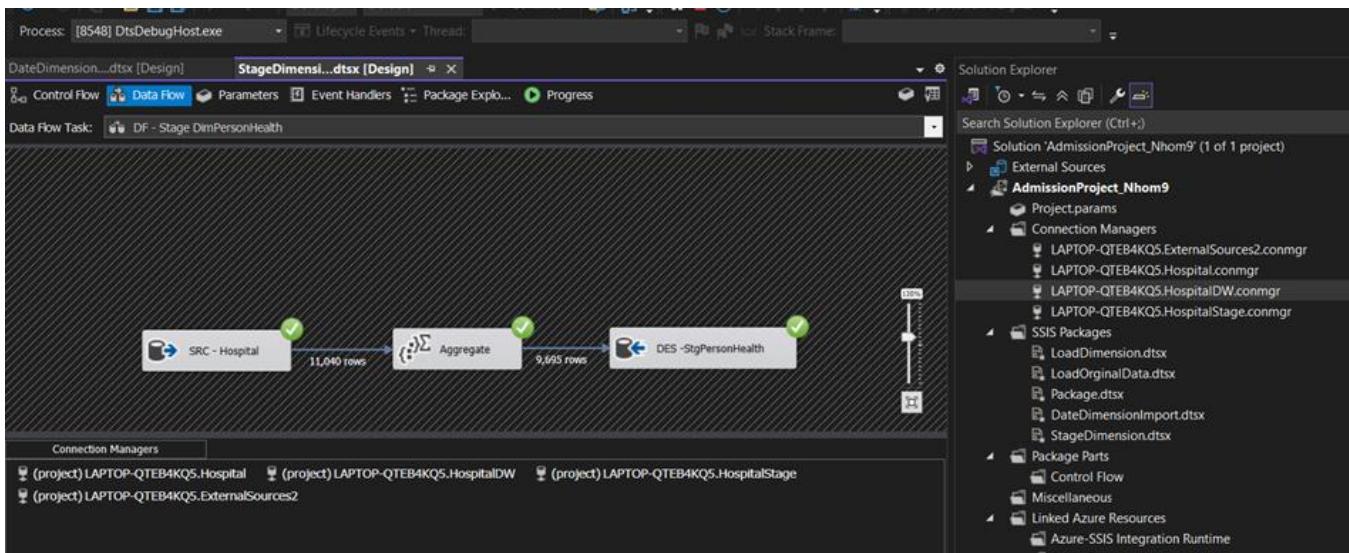
- StgPerson:



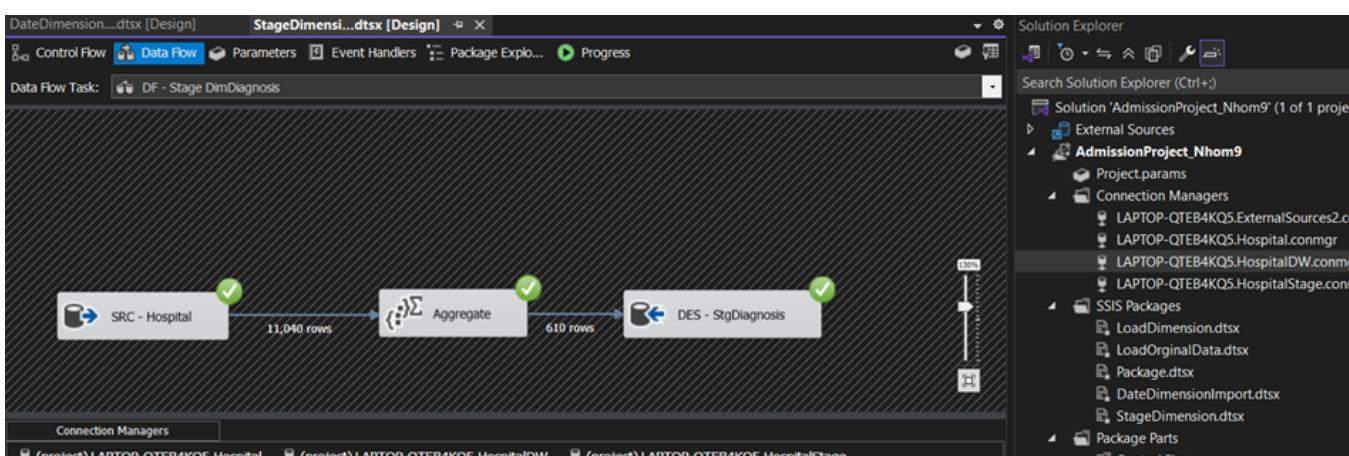
- StgComorbidity:



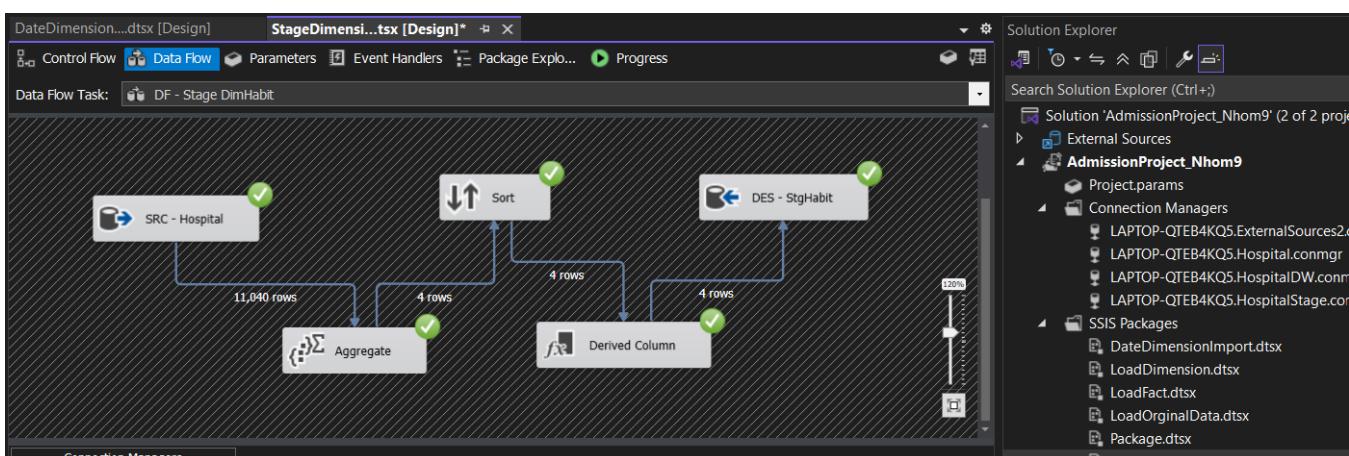
- StgPersonHealth:



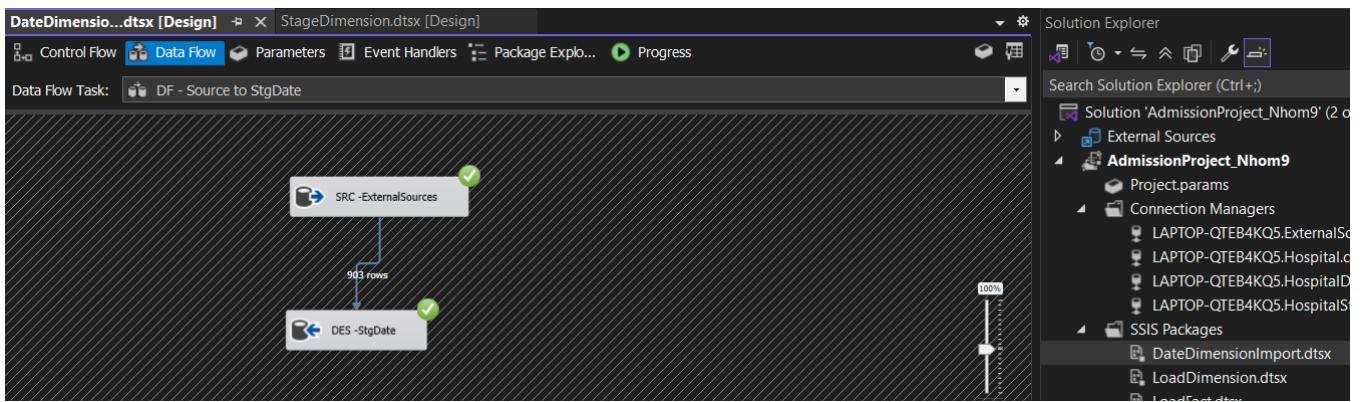
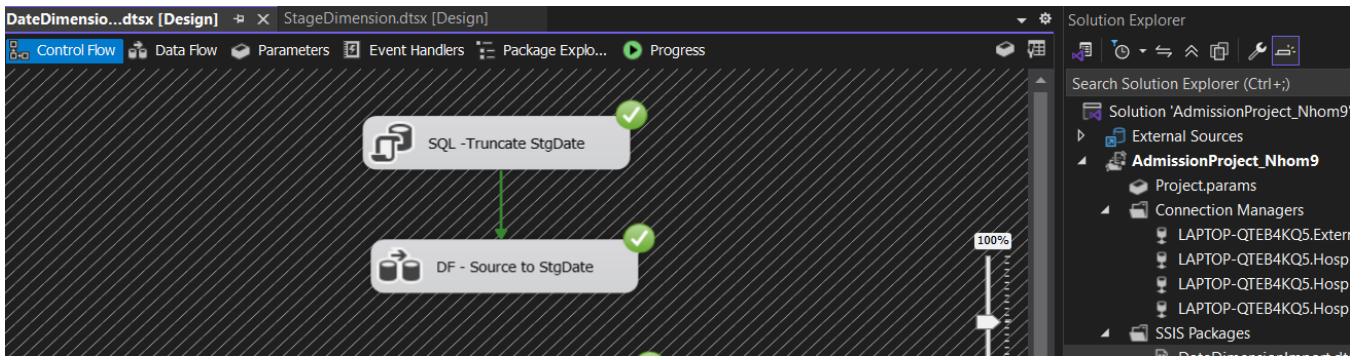
- StgDiagnosis:



- StgHabit:



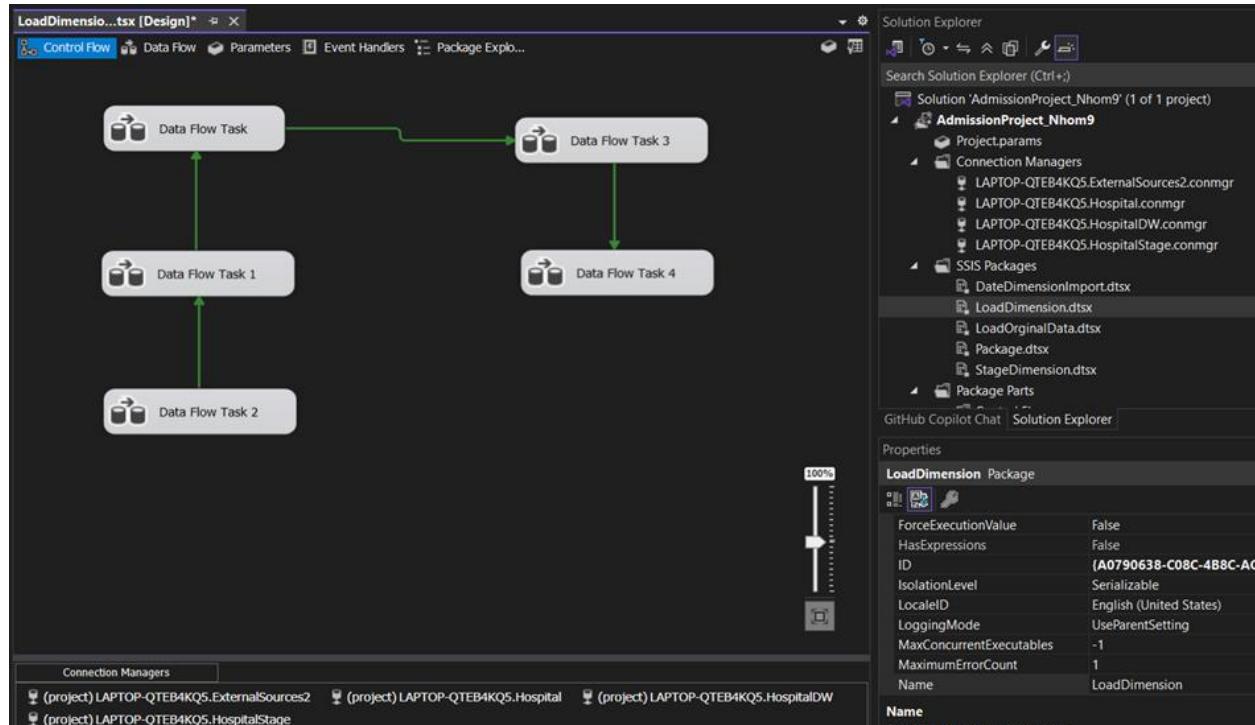
Sau khi hoàn thành thì click qua DateDimensionImport.dtsx rồi run:



3.4.3. Load dữ liệu vào các bảng Dimension trong Data Warehouse

- Thiết kế luồng điều khiển (Control Flow) trong LoadDimension.dtsx:

Bắt đầu bằng cách thêm các "Data Flow Task" vào Control Flow của package LoadDimension.dtsx



- Đổi tên các Data Flow Task:

Để dễ quản lý và hiểu rõ chức năng của từng tác vụ, bạn đã đổi tên các "Data Flow

"Task" ban đầu thành các tên có ý nghĩa hơn, phản ánh việc tải dữ liệu vào các bảng Dimension cụ thể.

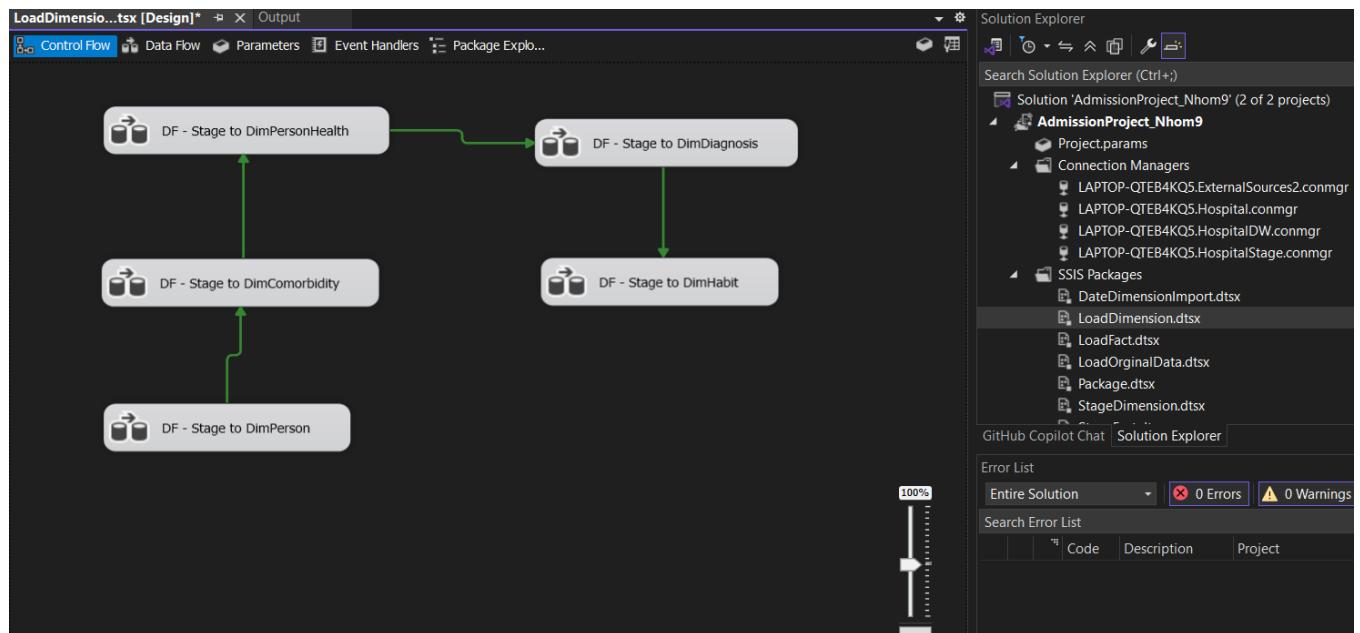
DF - Stage to DimPersonHealth

DF - Stage to DimDiagnosis

DF - Stage to DimComorbidity

DF - Stage to DimHabit

DF - Stage to DimPerson



3. Tiến hành load các Dimension sau:

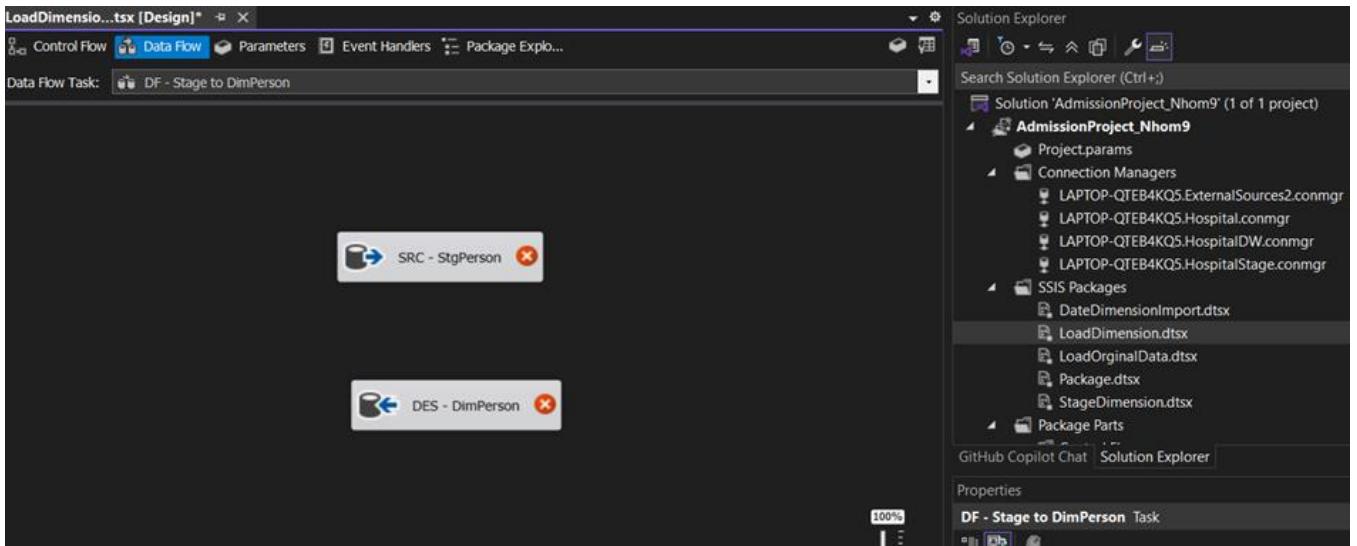
Bước tiếp theo là đi sâu vào cấu hình chi tiết của từng "Data Flow Task" để thực hiện việc tải dữ liệu từ các bảng "Stage" (dữ liệu tạm thời) vào các bảng "Dimension" tương ứng trong Data Warehouse (HospitalDW).

DimPerson:

1. Click vào **DF - Stage to DimPerson** và thêm **Source Assistant** và **Destination Assistant**.

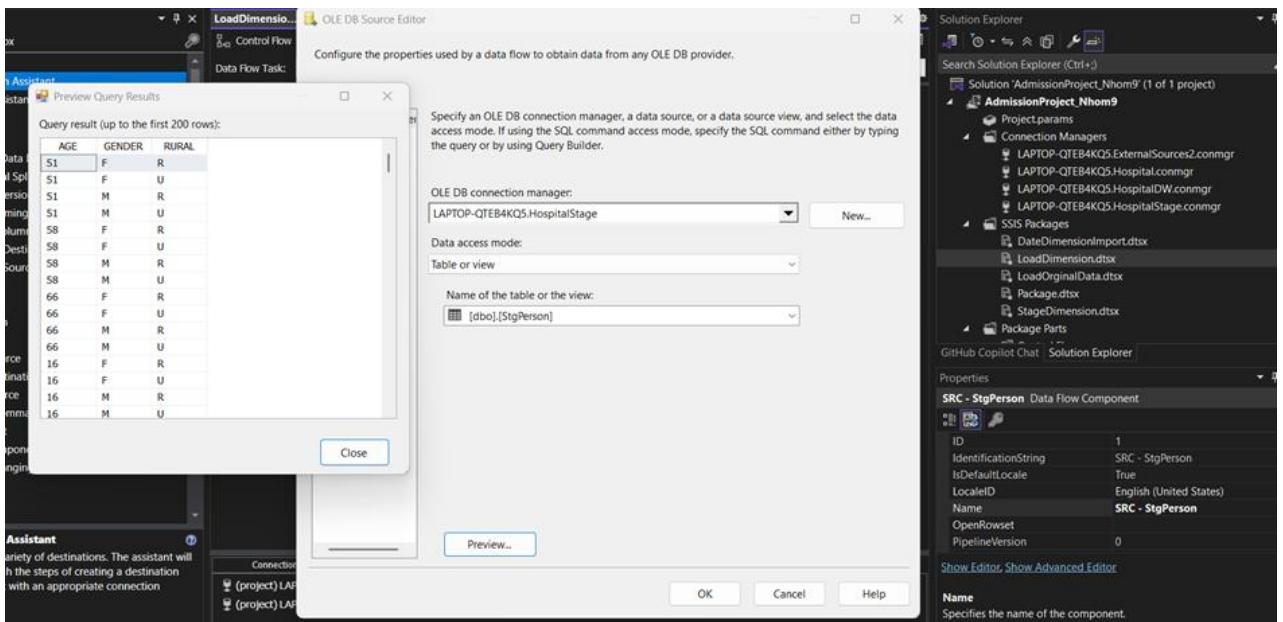
Assistant.

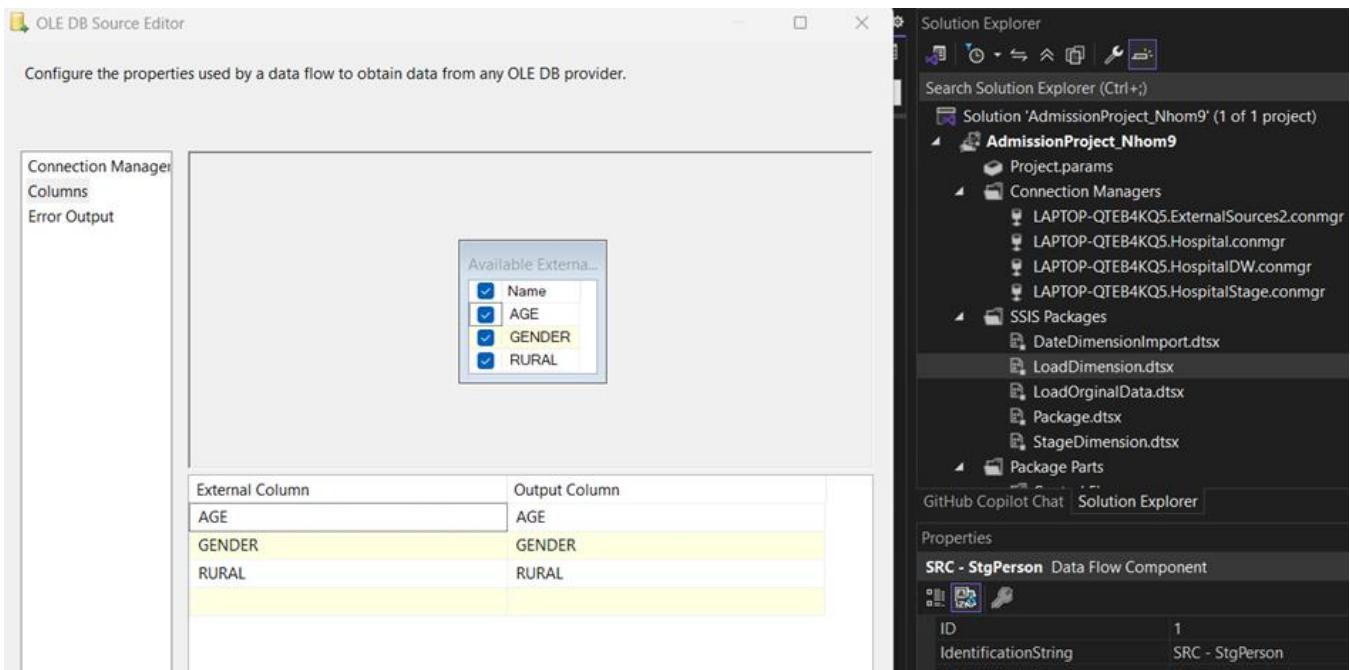
2. Tiến hành đổi tên và ta có được hình như sau:



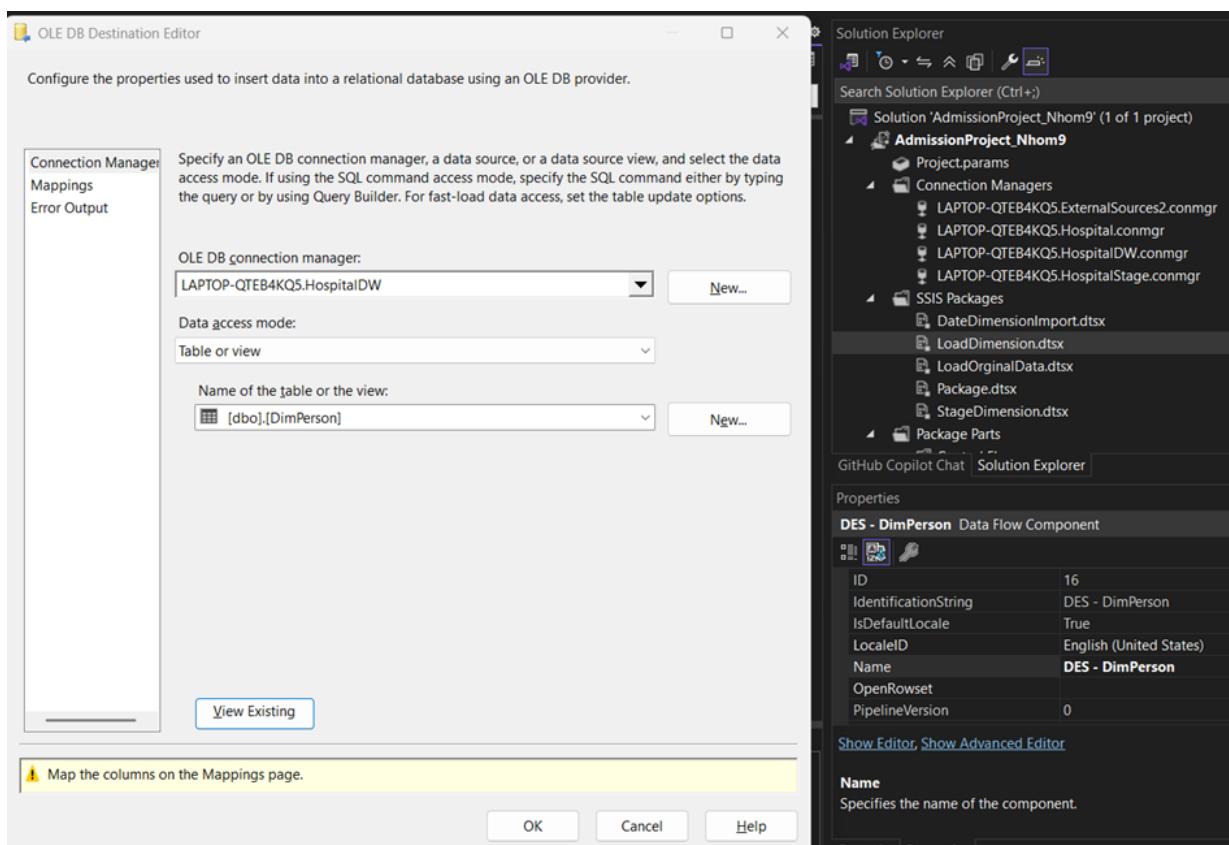
3. Tiến hành cấu hình các task:

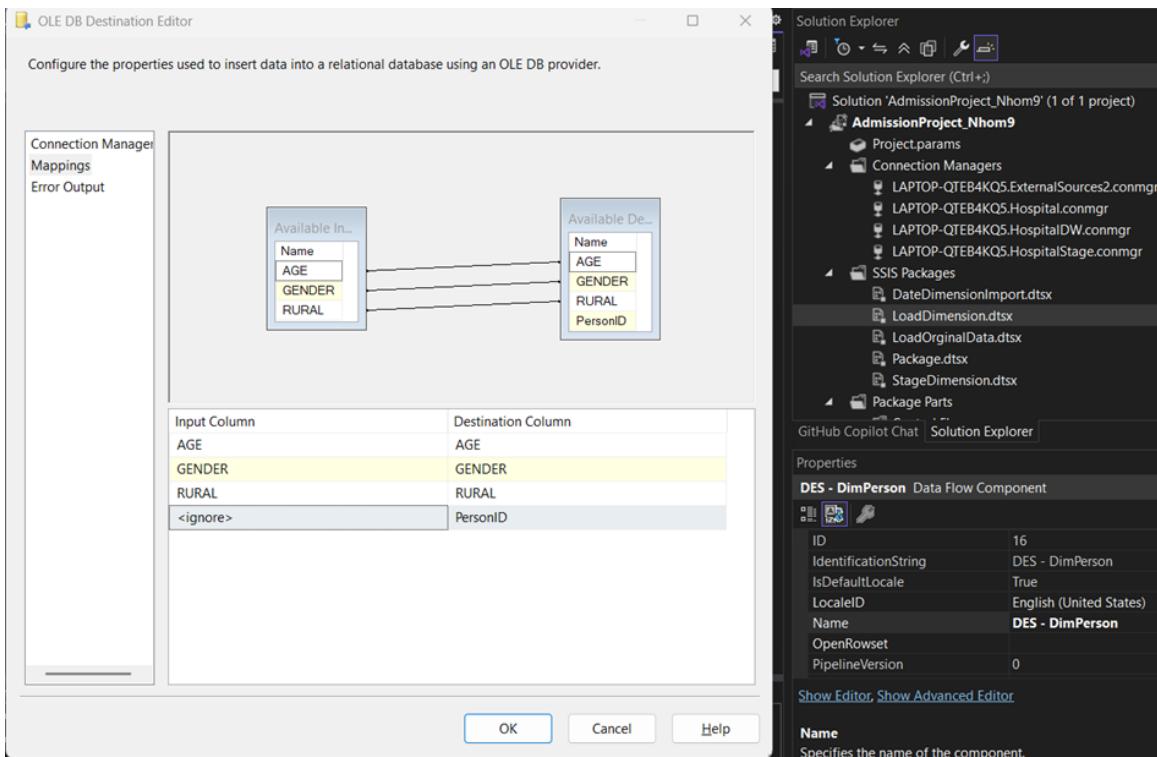
- Cấu hình Source Assistant:



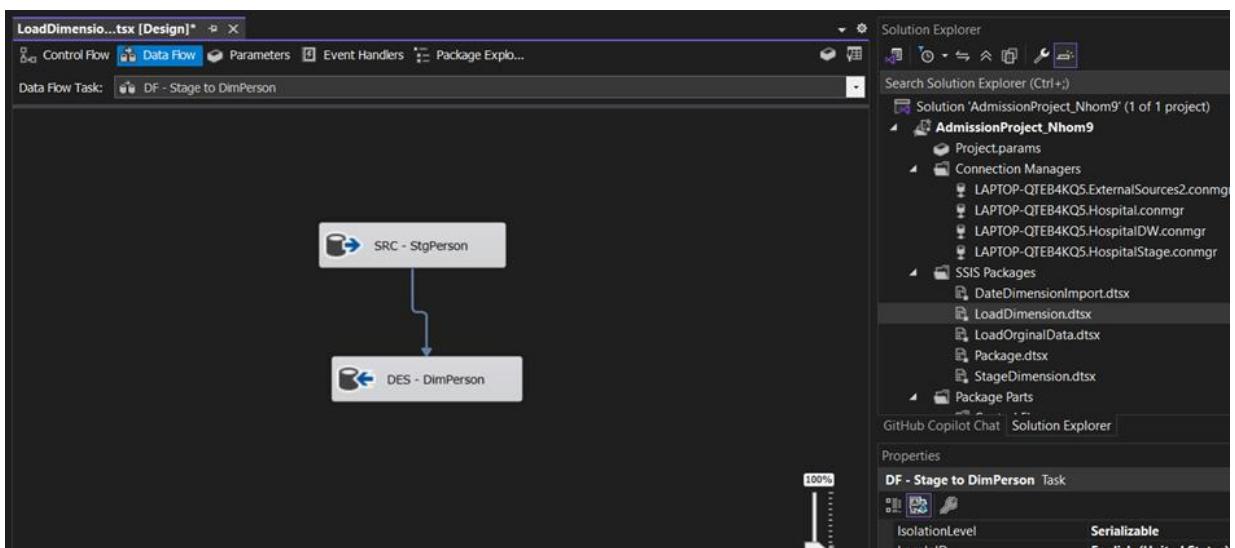


- Cấu hình Destination Assistant:



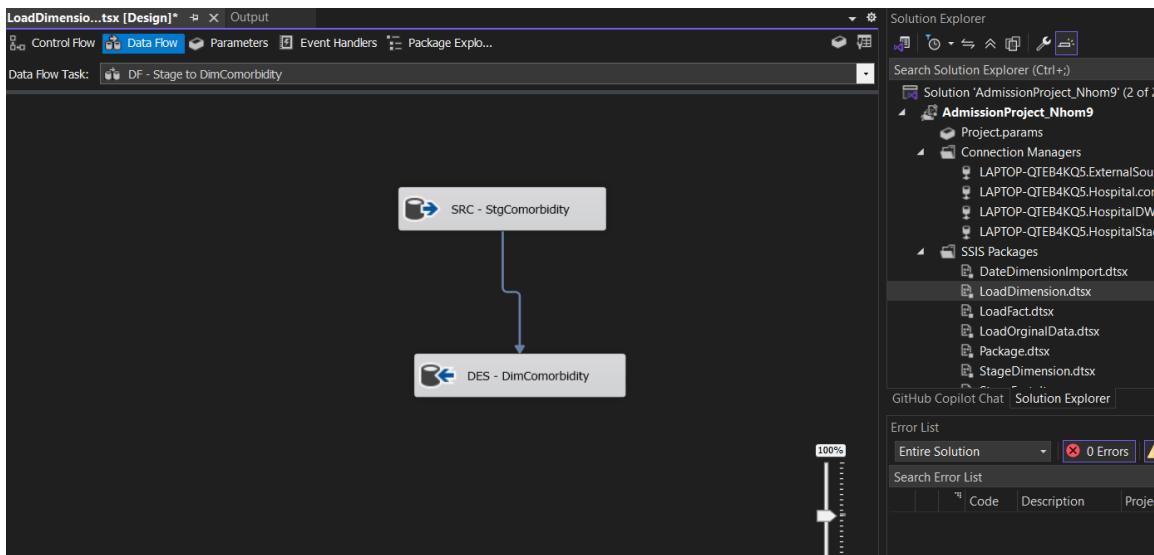


4. Data Flow Task của ta sẽ trông như sau:



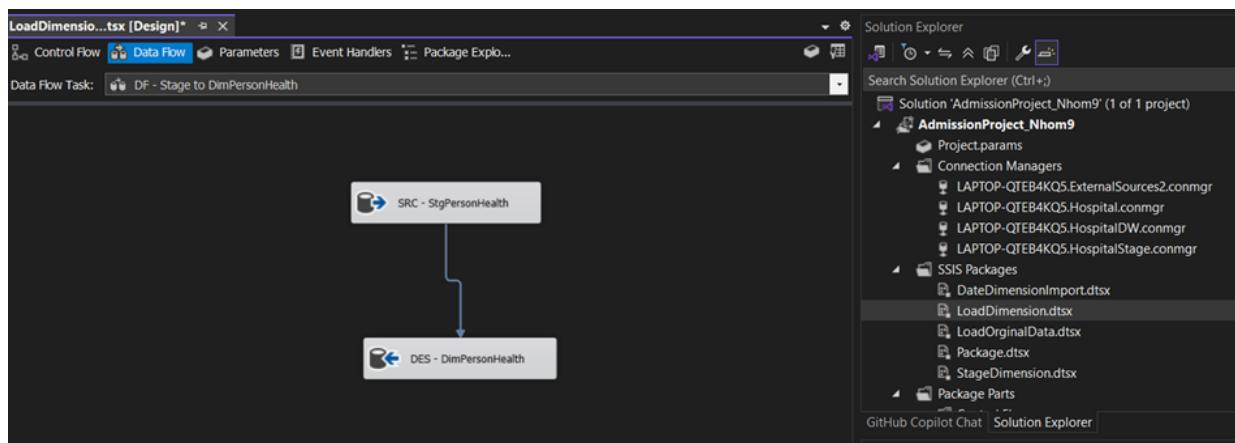
Tiếp tục thực hiện các bước tương tự với DimComorbidity, DimPersonHealth, DimDiagnosis, DimHabit như đã thực hiện với DimPerson, ta có được thiết lập cuối cùng:
DimComorbidity:

Data Flow Task của ta sẽ trông như sau:



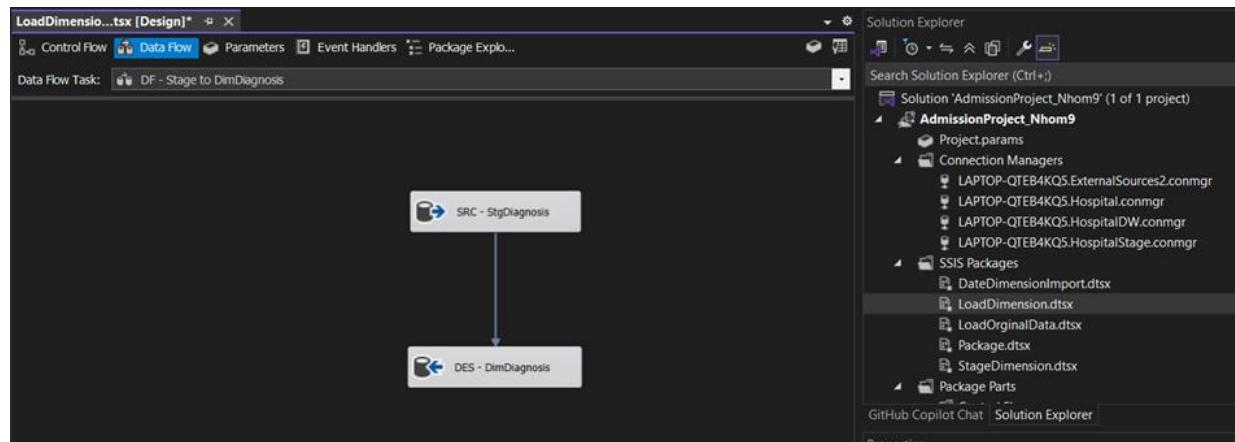
DimPersonHealth:

Data Flow Task của ta sẽ trông như sau:



DimDiagnosis:

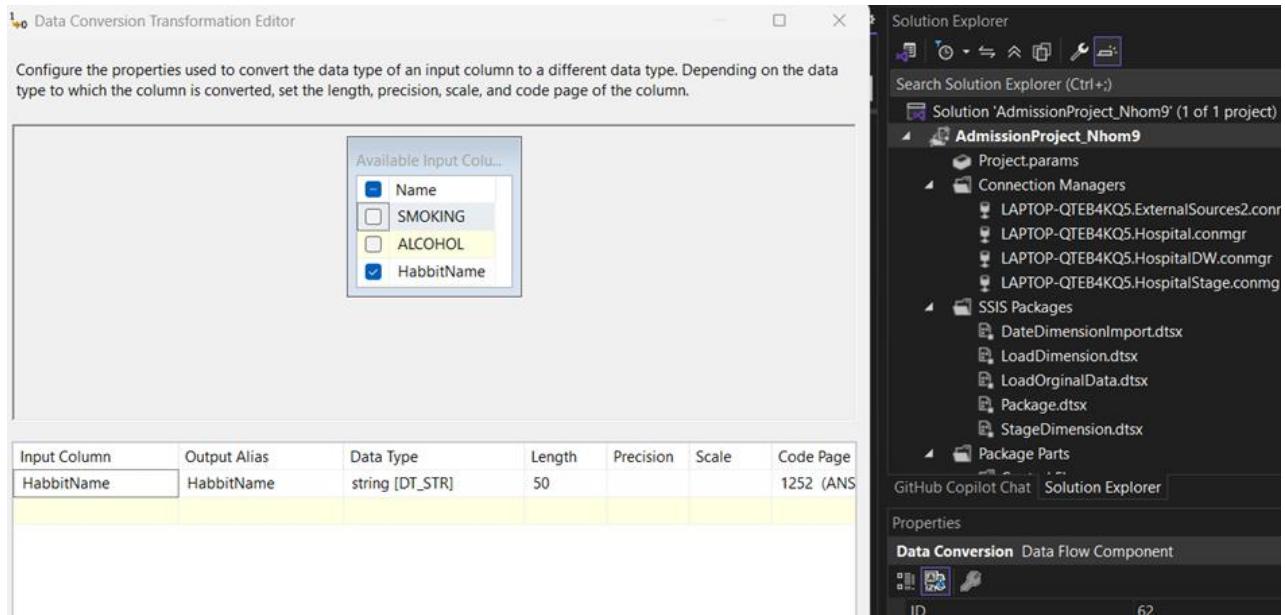
Data Flow Task của ta sẽ trông như sau:



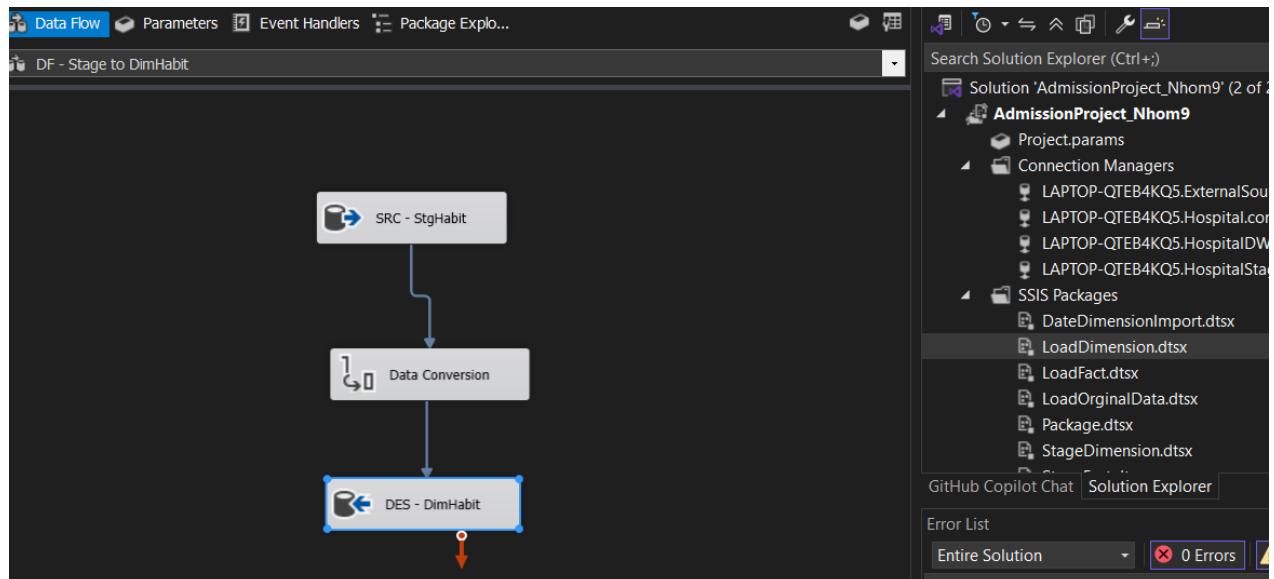
DimHabit: Bổ sung thêm Task Data Conversion để chuyển đổi kiểu dữ liệu HabbitName

cho khớp với kiểu trong bảng Dimension trong Data Warehouse khi load lên không gặp lỗi.

- Cấu hình Data Conversion:



Data Flow Task của ta sẽ trông như sau:

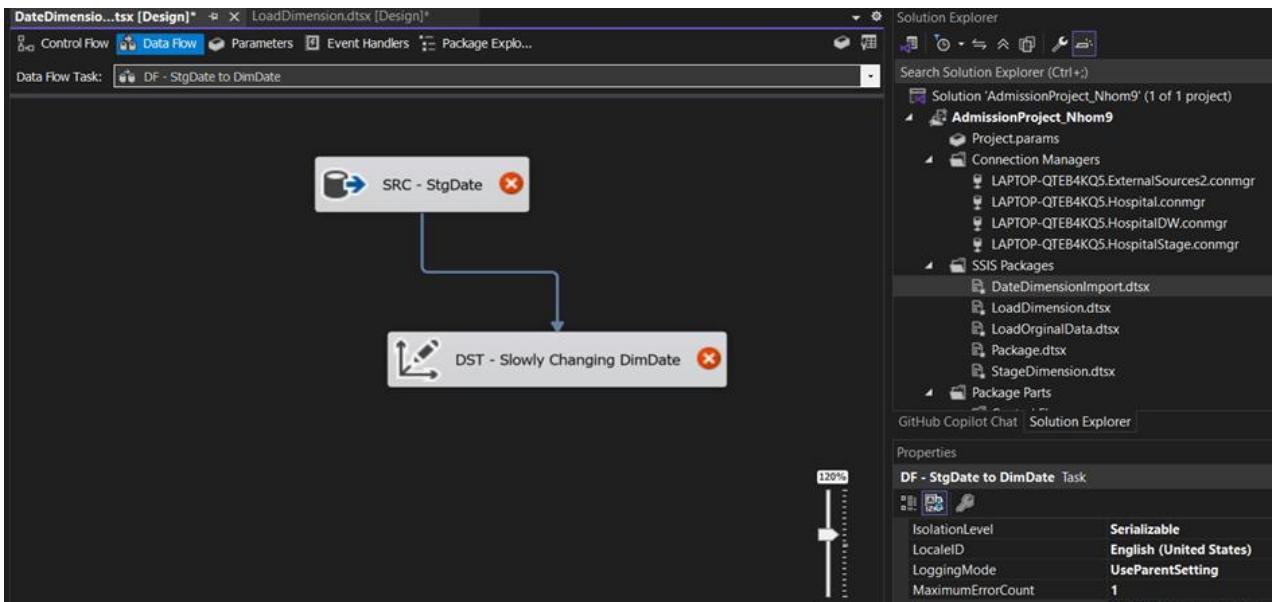


DimDate:

1. Click vào DF - StgDate to DimDate và thêm Source Assistant và Slowly

Changing Dimension.

2. Tiến hành đổi tên và ta có được hình như sau:



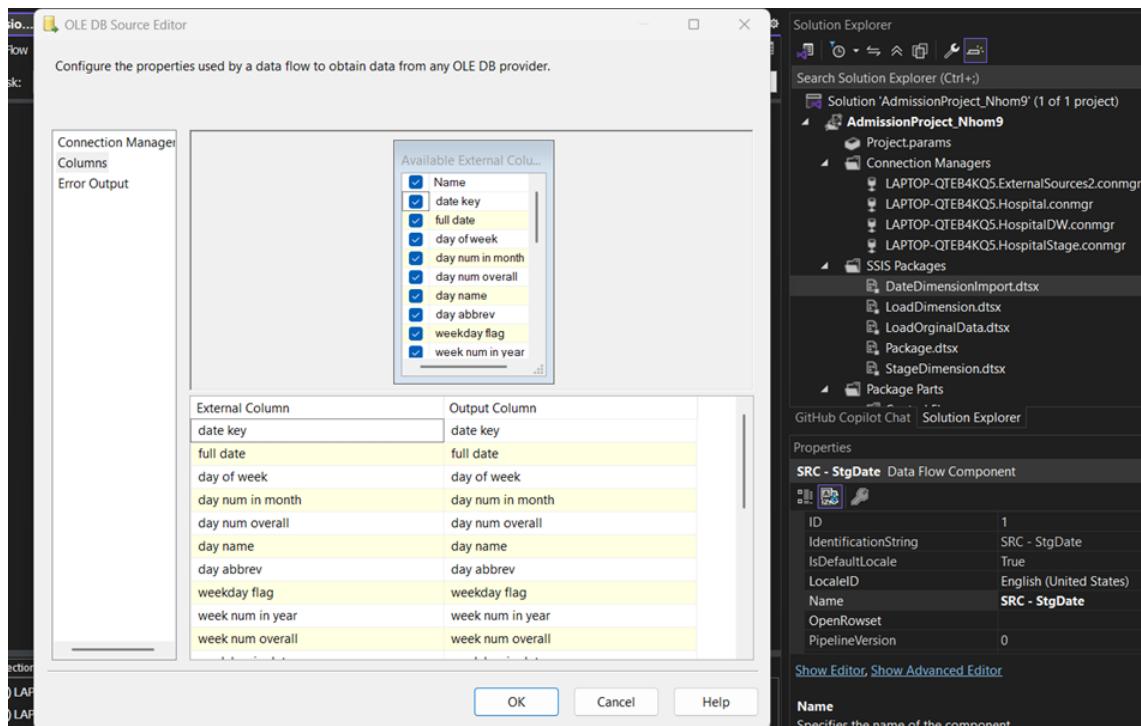
3. Tiến hành cấu hình các task:

- Cấu hình Source Assistant:

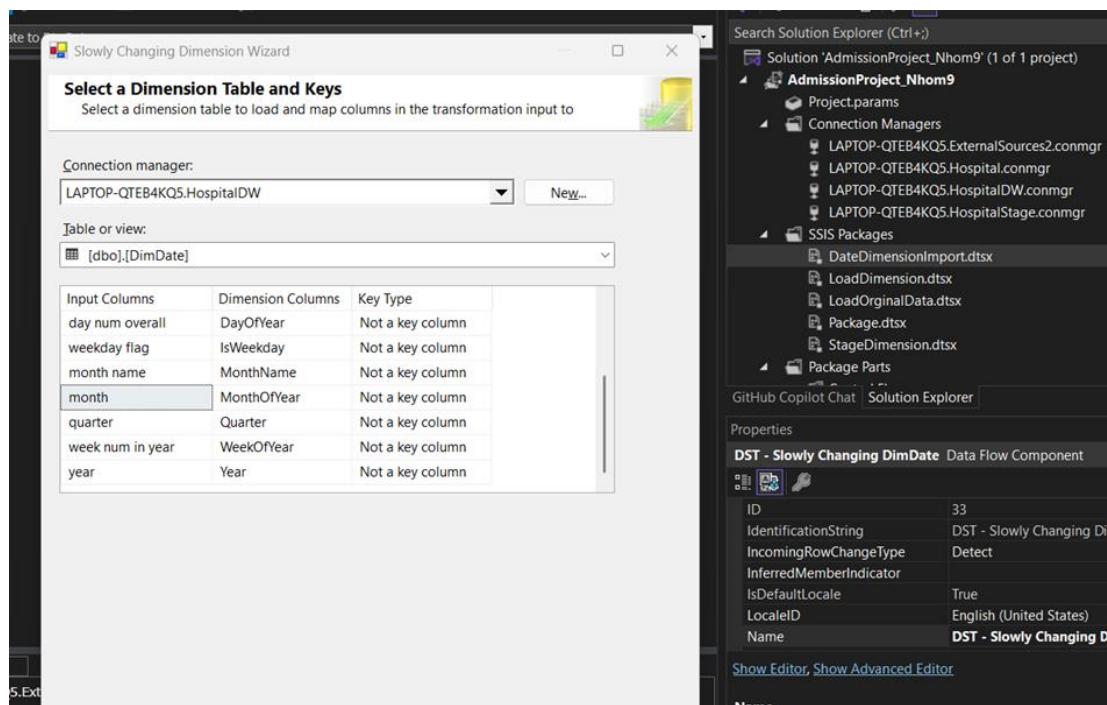
The screenshot shows the OLE DB Source Editor dialog. It displays a preview of the query results, which are the first 200 rows of a table named [dbo].[StgDate]. The configuration settings include:

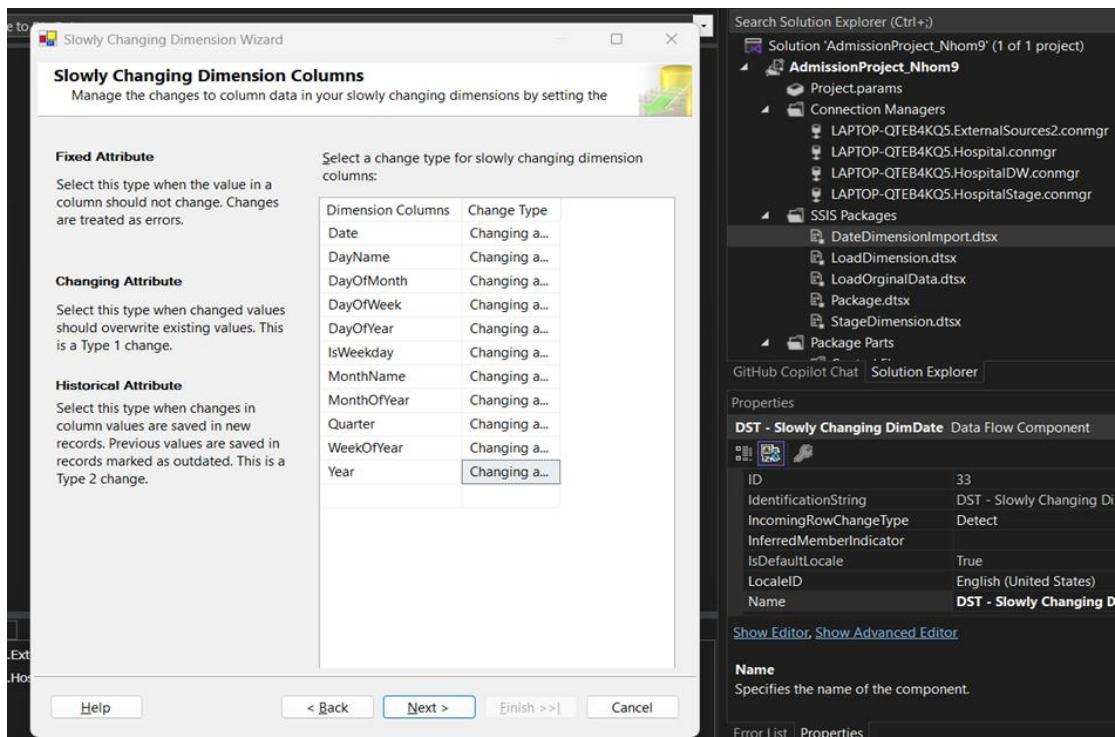
- OLE DB connection manager:** LAPTOP-QTEB4KQ5.HospitalStage
- Data access mode:** Table or view
- Name of the table or the view:** [dbo].[StgDate]

The properties pane on the right shows the component is named 'SRC - StgDate' with ID 1.

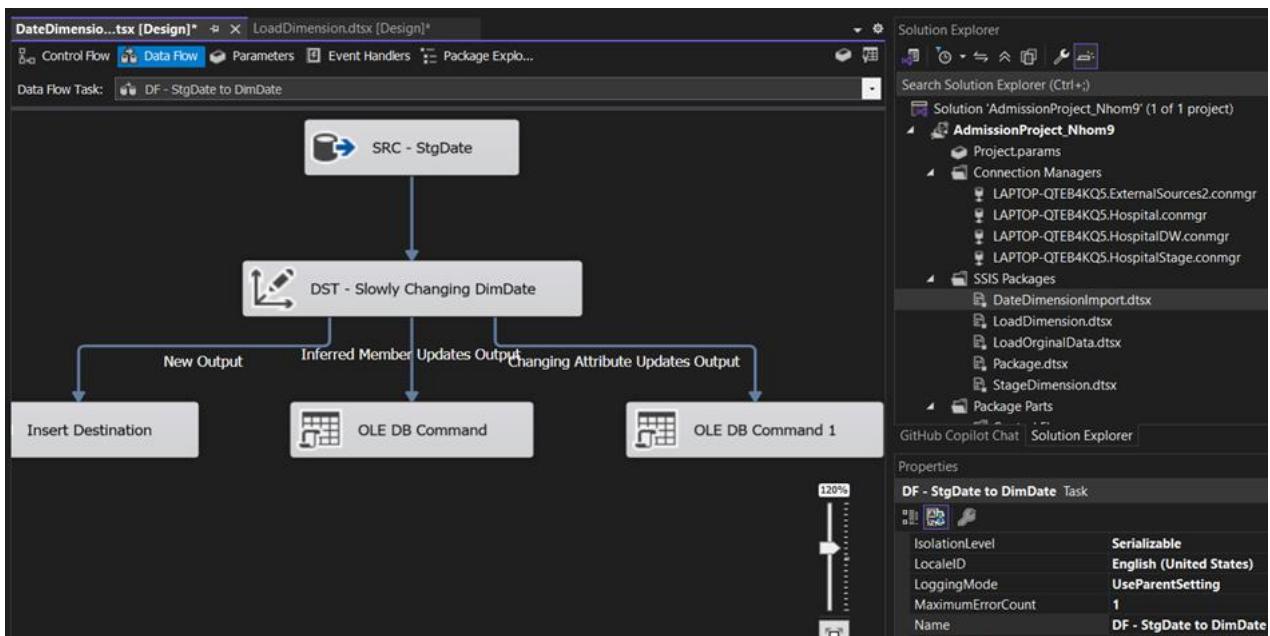


- Cấu hình Slowly Changing Dimension:

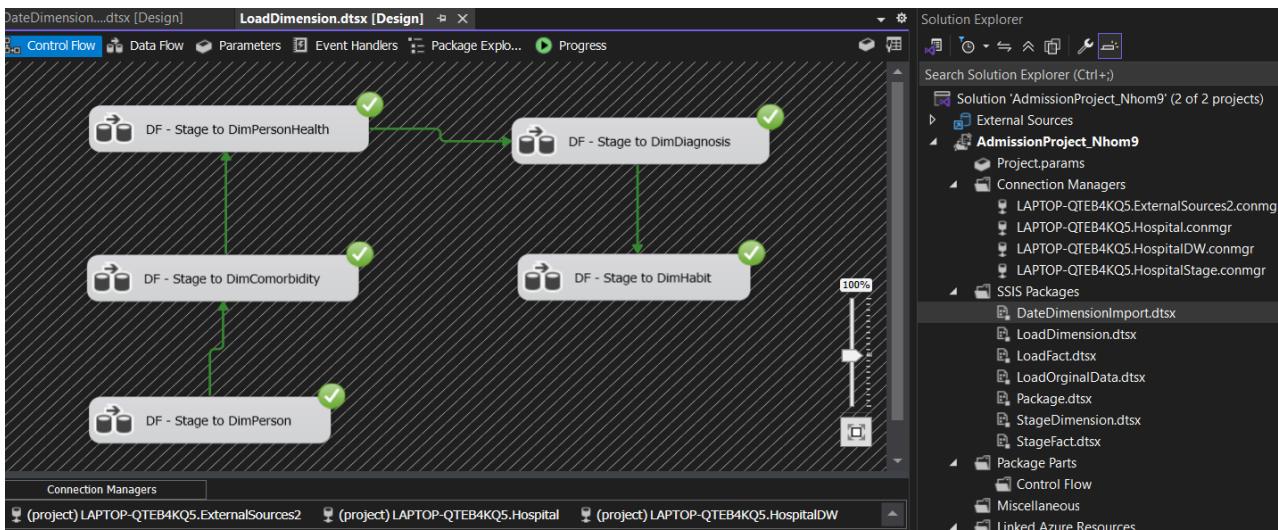




4. Data Flow Task của ta sẽ trông như sau:

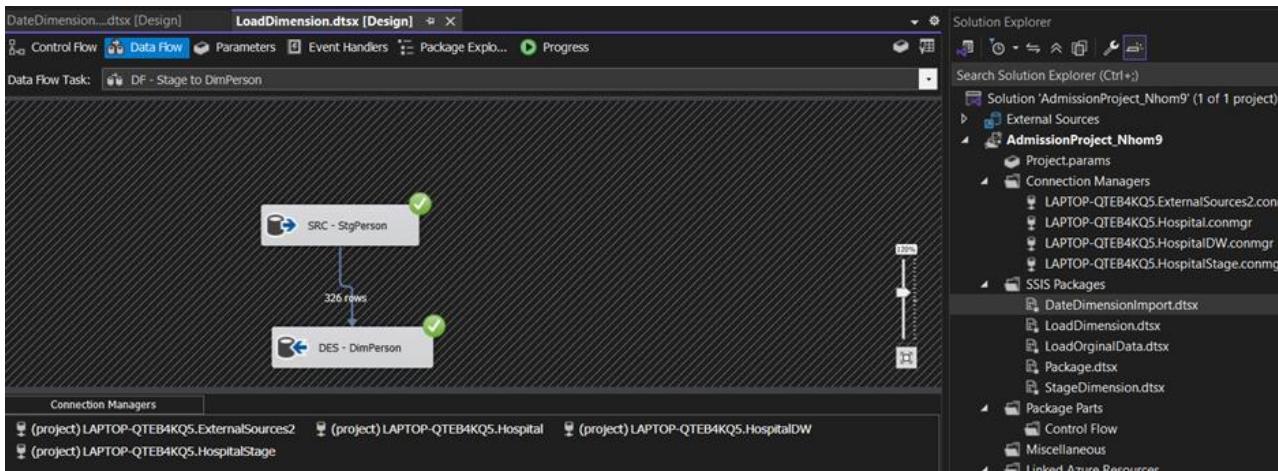


Tiến hành load dữ liệu vào Data Warehouse :

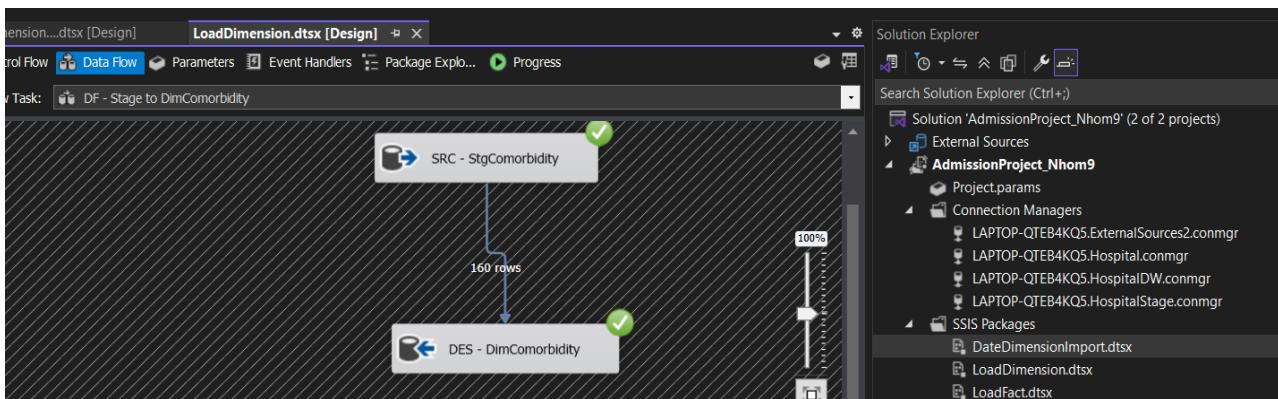


Chi tiết từng Dim:

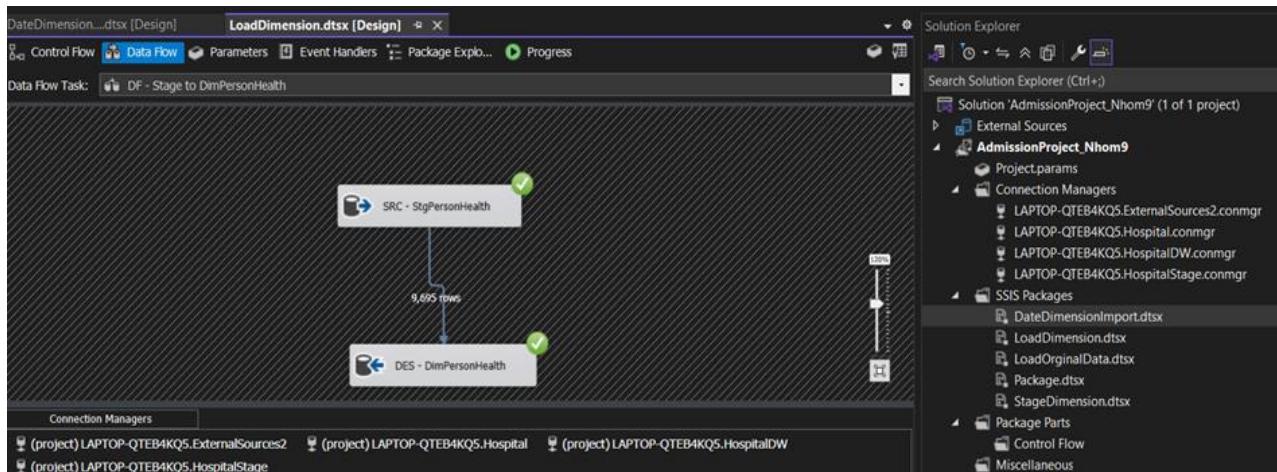
- DimPerson:



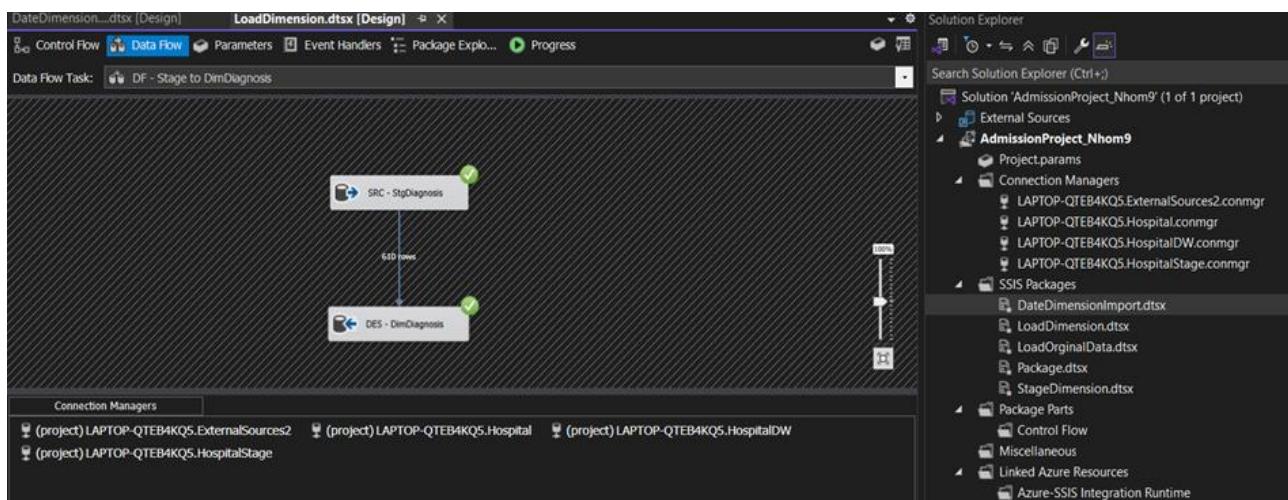
- DimComorbidity:



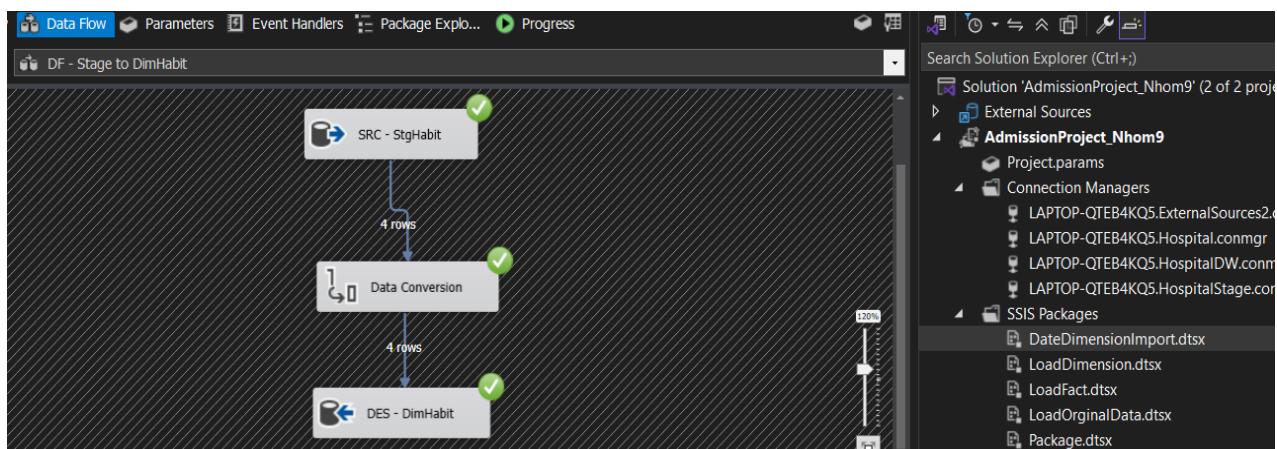
- DimPersonHealth:



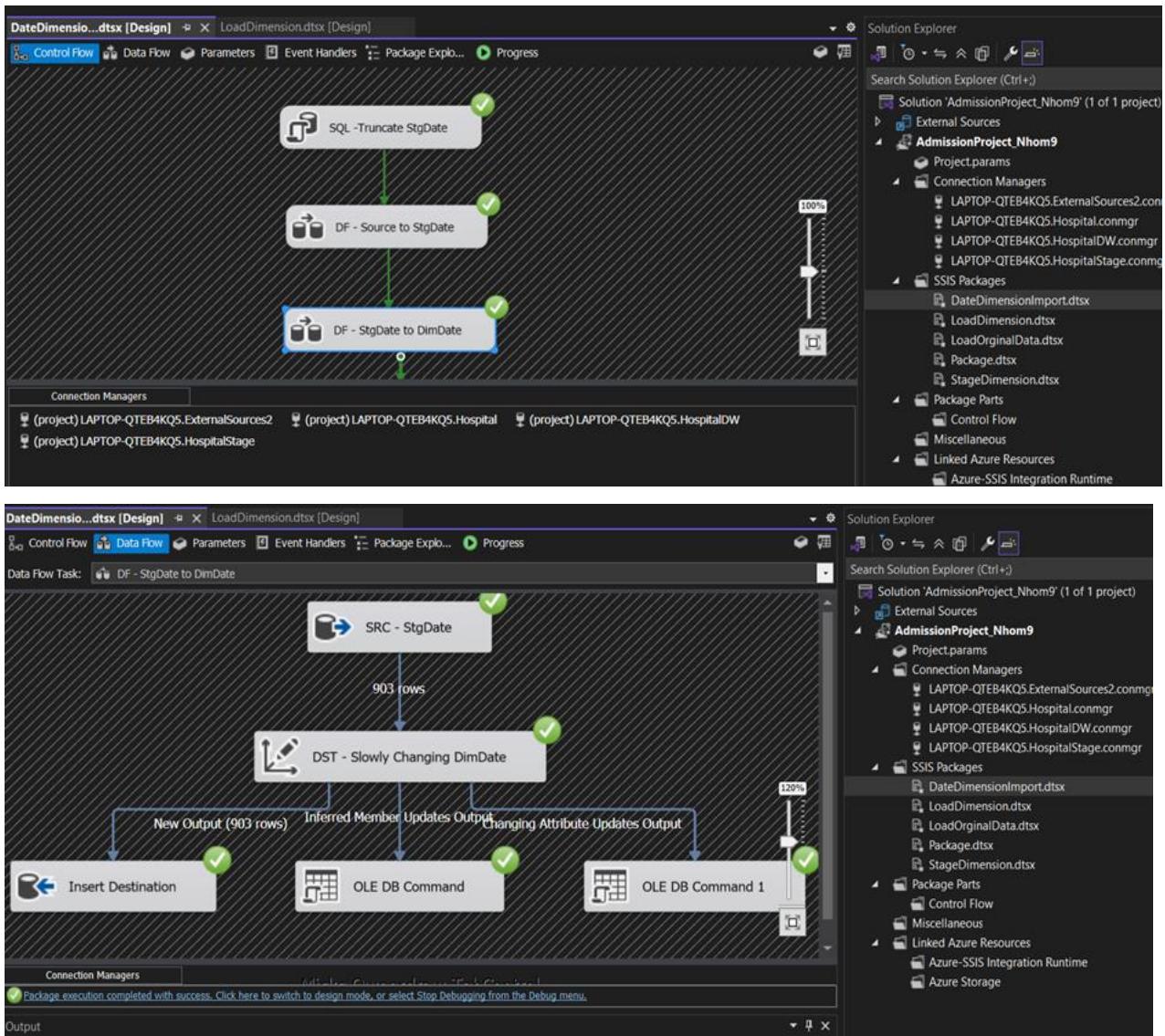
- DimDiagnosis:



- DimHabit:



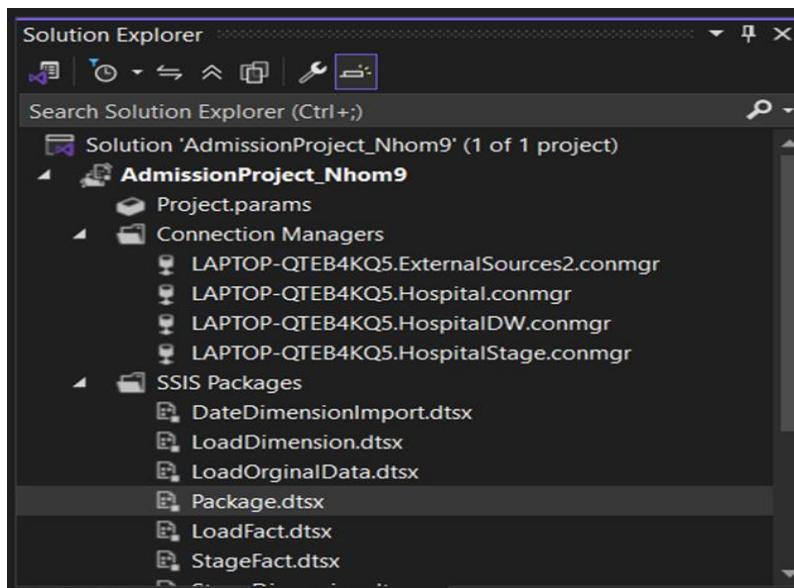
- DimDate:



3.5. Tích hợp Dữ liệu vào các bảng Fact

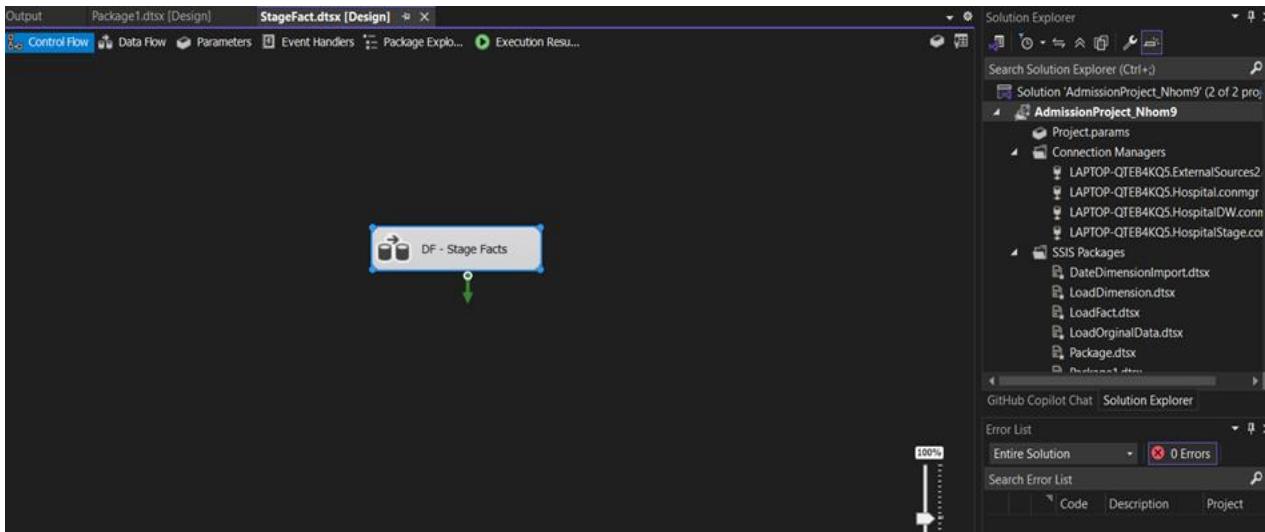
3.5.1. Tạo Package xử lý Fact

Tạo thêm 2 package mới để thực hiện các bước Stage và Load dữ liệu vào các Fact.



3.5.2. Stage các Fact

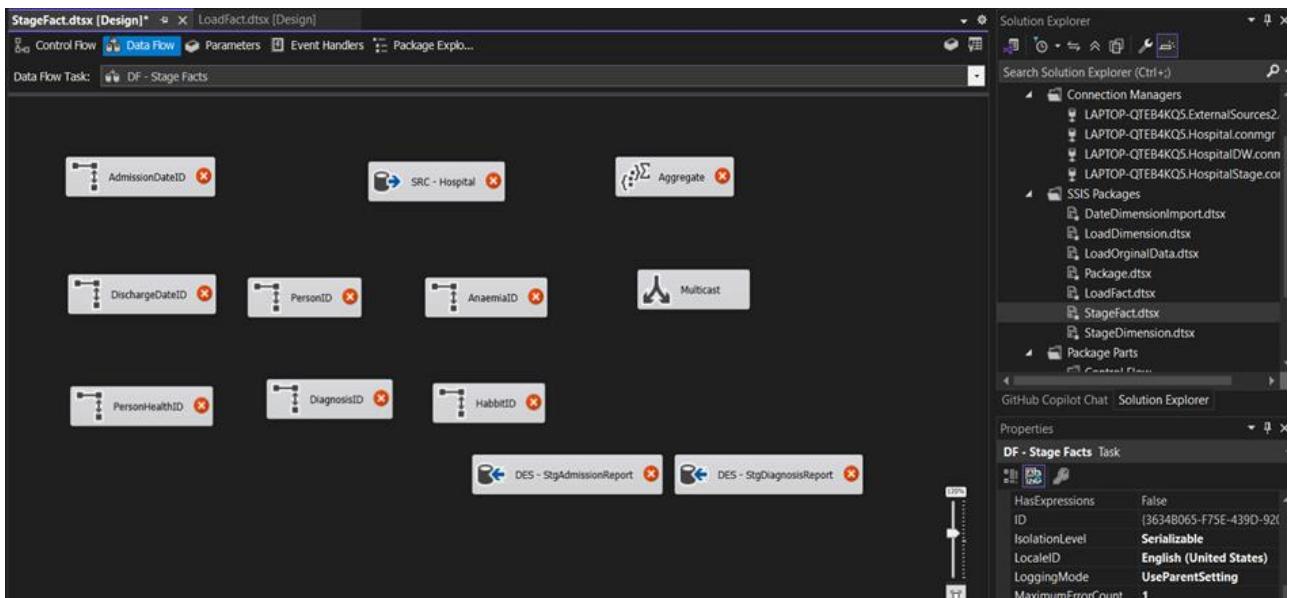
1. Click vào StageFact.dtsx và tiến hành thêm vào 1 Data Flow Task như hình sau:



2. Tiến hành stage hai Fact (StgAdmissionReport và StgDiagnosisReport) sau:

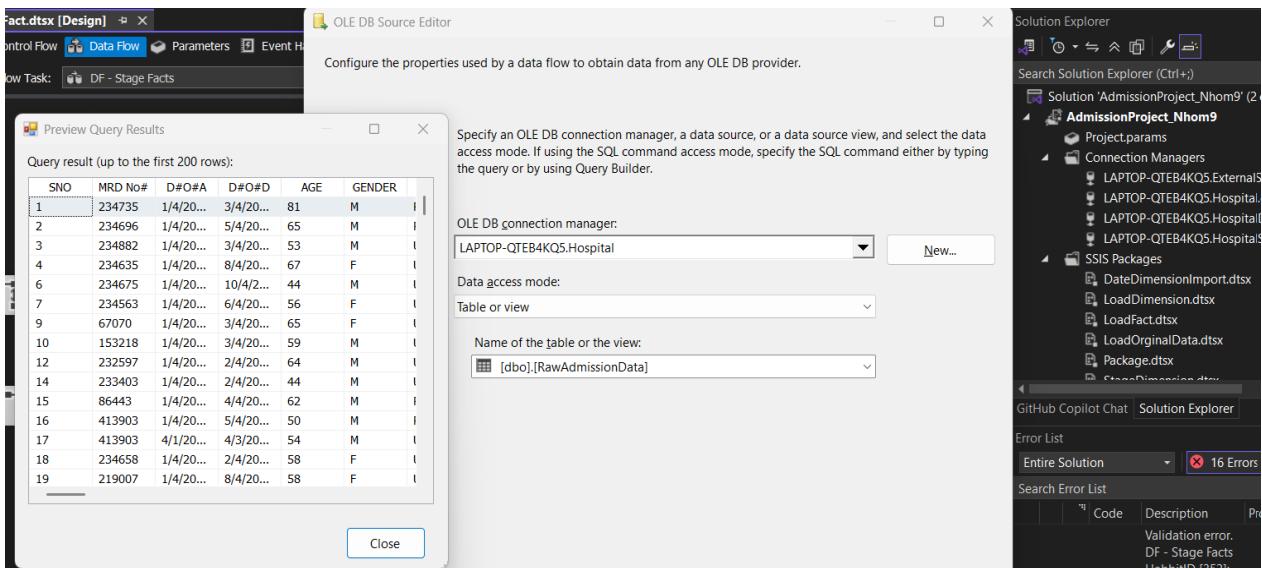
Click vào DF - Stage FactAdmissionReport và thêm Source Assistant, Aggregate, Sort, MultiCast,5 LookUp và 2 Destination Assistant.

Tiến hành đổi tên và ta có được hình như sau:



3. Tiến hành cấu hình các task:

- Cấu hình Source Assistant:



- Cấu hình LookUp Transformation:

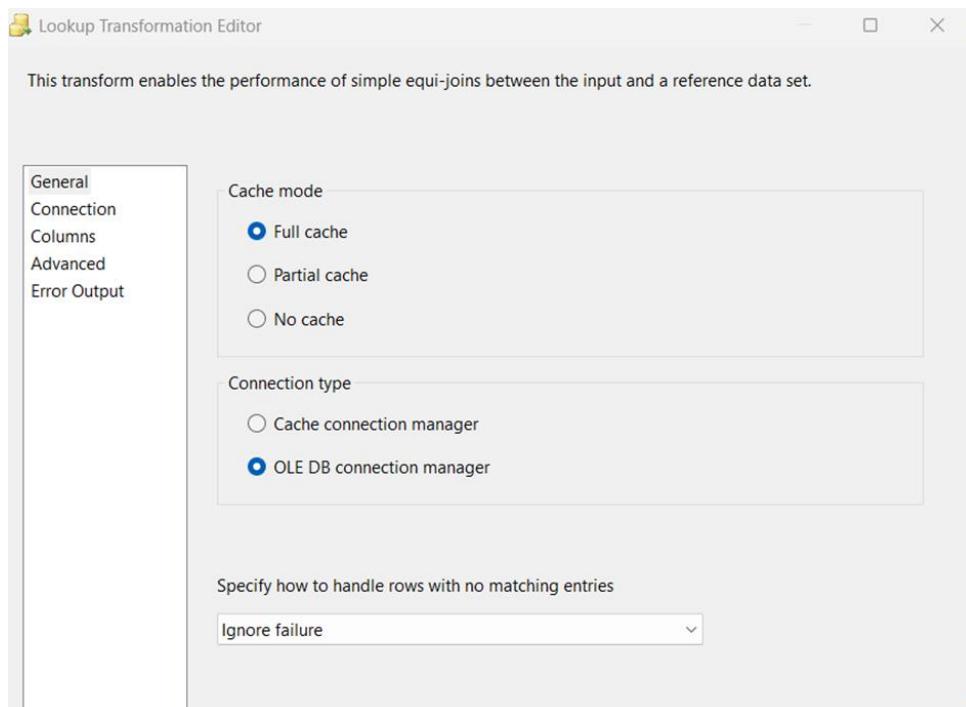
AdmissionDateID:

1. Tab General

Cache mode: Full cache

Connection type: OLE DB connection manager

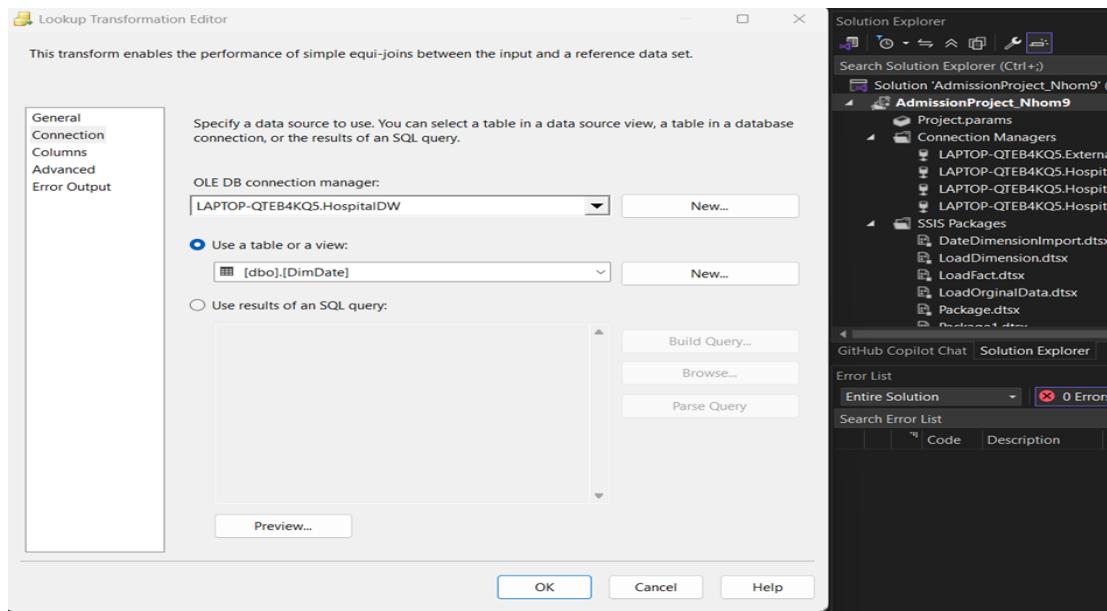
Specify how to handle rows with no matching entries: Ignore failure



2. Tab Connection

OLE DB connection manager: LAPTOP-QTEB4KQ5.HospitalDW

Use a table or a view: [dbo].[DimDate]

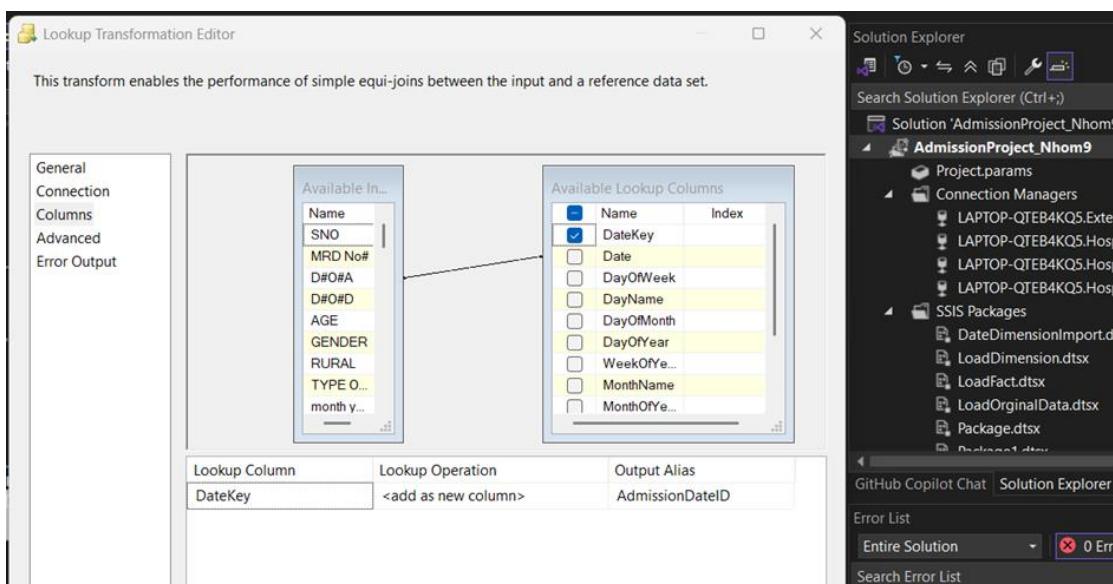


3. Tab Columns

Thiết lập phép nối (Join): Nối cột D.O.A (Input Column) với cột Date (Lookup Column).

Cấu hình Output Column:

- + Lookup Column: DateKey
- + Lookup Operation: <add as new column>
- + Output Alias: AdmissionDateID



DischargeDateID:

1. Tab General

Cache mode: Full cache

Connection type: OLE DB connection manager

Specify how to handle rows with no matching entries: Ignore failure

2. Tab Connection

OLE DB connection manager: LAPTOP-QTEB4KQ5.HospitalDW

Use a table or a view: [dbo].[DimDate]

3. Tab Columns

Thiết lập phép nối (Join): Nối cột D.O.D (Input Column) với cột Date (Lookup Column).

Cấu hình Output Column:

- + Lookup Column: DateKey
- + Lookup Operation: <add as new column>
- + Output Alias: DischargeDateID

PersonID:

1. Tab General

Cache mode: Full cache

Connection type: OLE DB connection manager

Specify how to handle rows with no matching entries: Fail component

2. Tab Connection

OLE DB connection manager: LAPTOP-QTEB4KQ5.HospitalDW

Use a table or a view: [dbo].[DimPerson]

3. Tab Columns

Thiết lập phép nối (Join): Nối cột AGE, GENDER,RURAL (Input Column) với cột AGE, GENDER,RURAL (Lookup Column).

Cấu hình Output Column:

- + Lookup Column: PersonID
- + Lookup Operation: <add as new column>
- + Output Alias: PersonID

ComorbidityID:

1. Tab General

Cache mode: Full cache

Connection type: OLE DB connection manager

Specify how to handle rows with no matching entries: Fail component

2. Tab Connection

OLE DB connection manager: LAPTOP-QTEB4KQ5.HospitalDW

Use a table or a view: [dbo].[DimComorbidity]

3. Tab Columns

Thiết lập phép nối (Join): Nối cột (Input Column) với cột (Lookup Column) các mục có tên giống nhau .

Cấu hình Output Column:

- + Lookup Column: ComorbidityID
- + Lookup Operation: <add as new column>
- + Output Alias: ComorbidityID

PersonHealthID:

1. Tab General

Cache mode: Full cache

Connection type: OLE DB connection manager

Specify how to handle rows with no matching entries: Fail component

2. Tab Connection

OLE DB connection manager: LAPTOP-QTEB4KQ5.HospitalDW

Use a table or a view: [dbo].[DimPersonHealth]

3. Tab Columns

Thiết lập phép nối (Join): Nối cột (Input Column) với cột (Lookup Column).: HB, TLC, PLATELETS, GLUCOSE, UREA, CREATININE, BNP

Cấu hình Output Column:

- + Lookup Column: PersonHealthID
- + Lookup Operation: <add as new column>
- + Output Alias: PersonHealthID

DiagnosisID:

1. Tab General

Cache mode: Full cache

Connection type: OLE DB connection manager

Specify how to handle rows with no matching entries: Fail component

2. Tab Connection

OLE DB connection manager: LAPTOP-QTEB4KQ5.HospitalDW

Use a table or a view: [dbo].[DimDiagnosis]

3. Tab Columns

Thiết lập phép nối (Join): Nối cột (Input Column) với cột (Lookup Column để xác định ID)

Cấu hình Output Column:

- + Lookup Column: DiagnosisID
- + Lookup Operation: <add as new column>
- + Output Alias: DiagnosisID

HabitID:

1. Tab General

Cache mode: Full cache

Connection type: OLE DB connection manager

Specify how to handle rows with no matching entries: Fail component

2. Tab Connection

OLE DB connection manager: LAPTOP-QTEB4KQ5.HospitalDW

Use a table or a view: [dbo].[DimHabit]

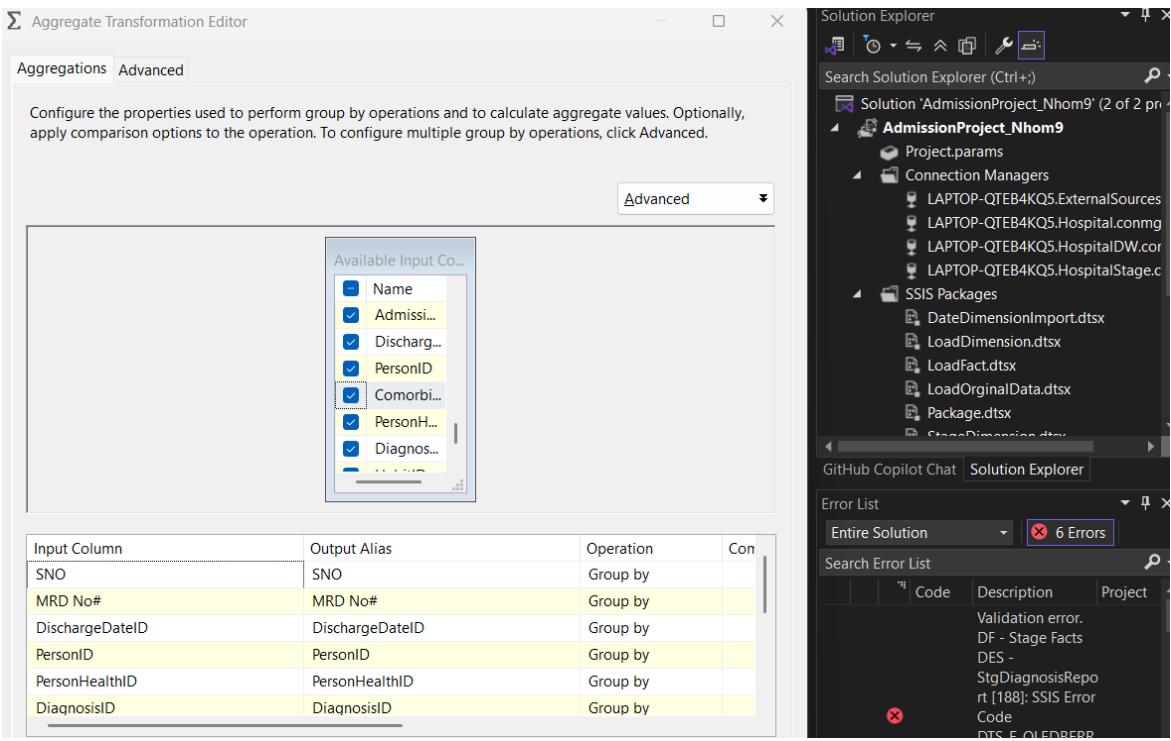
3. Tab Columns

Thiết lập phép nối (Join): Nối cột SMOKING, ALCOHOL (Input Column) với cột SMOKING, ALCOHOL (Lookup Column).

Cấu hình Output Column:

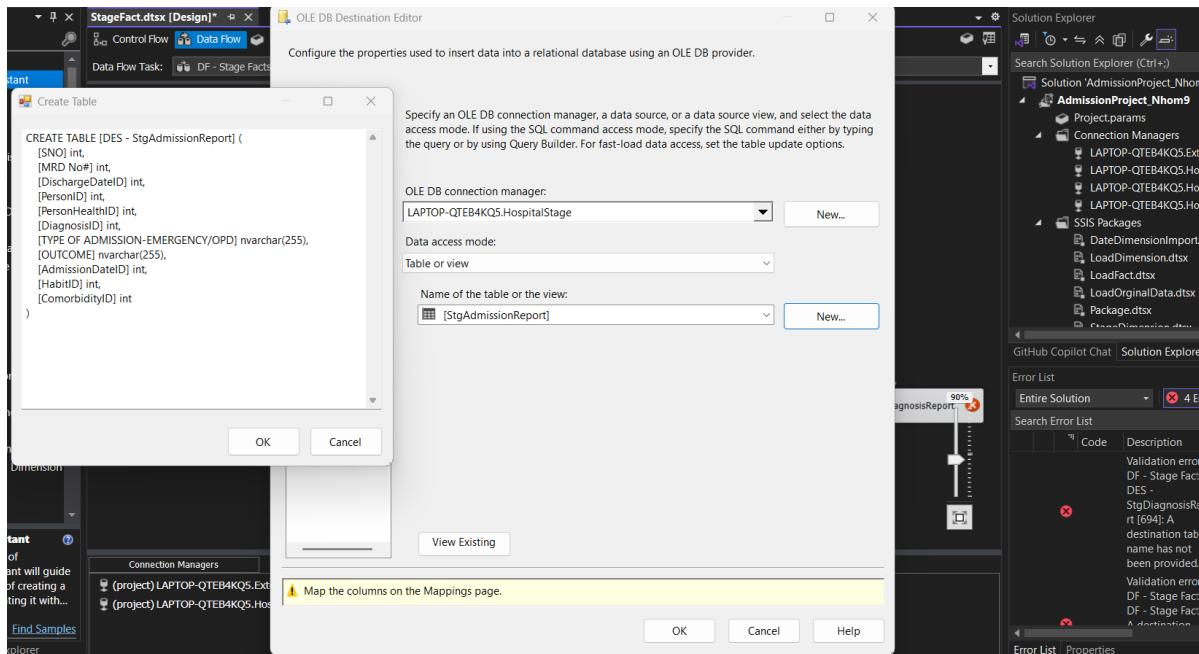
- + Lookup Column: HabitID
- + Lookup Operation: <add as new column>
- + Output Alias: HabitID

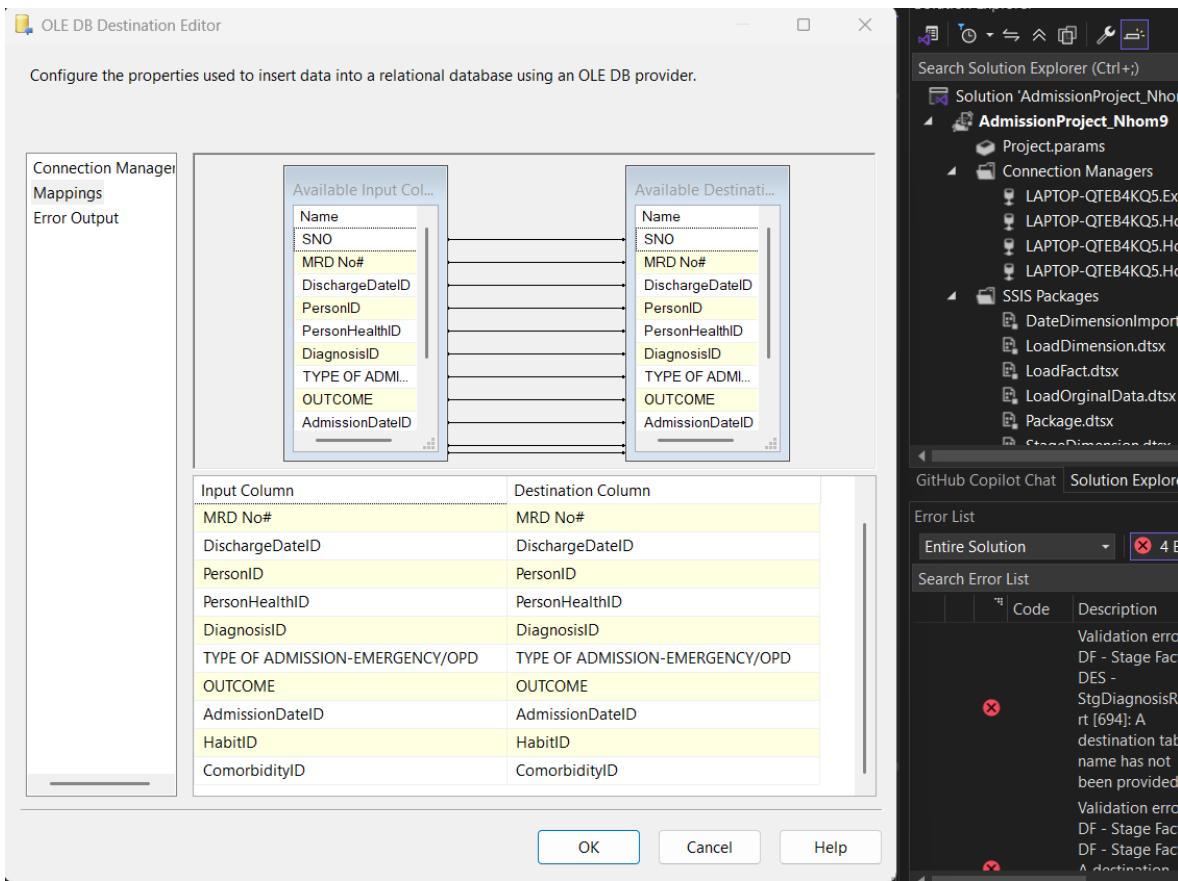
- Cấu hình Aggregate:



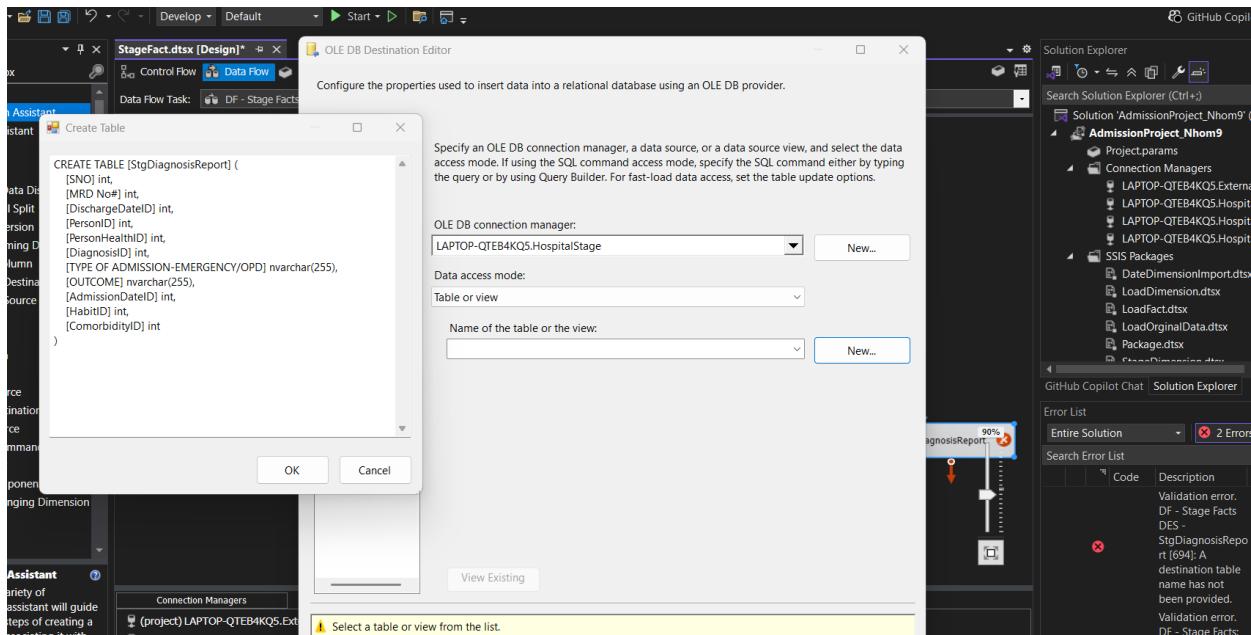
- Cấu hình Destination Assistant:

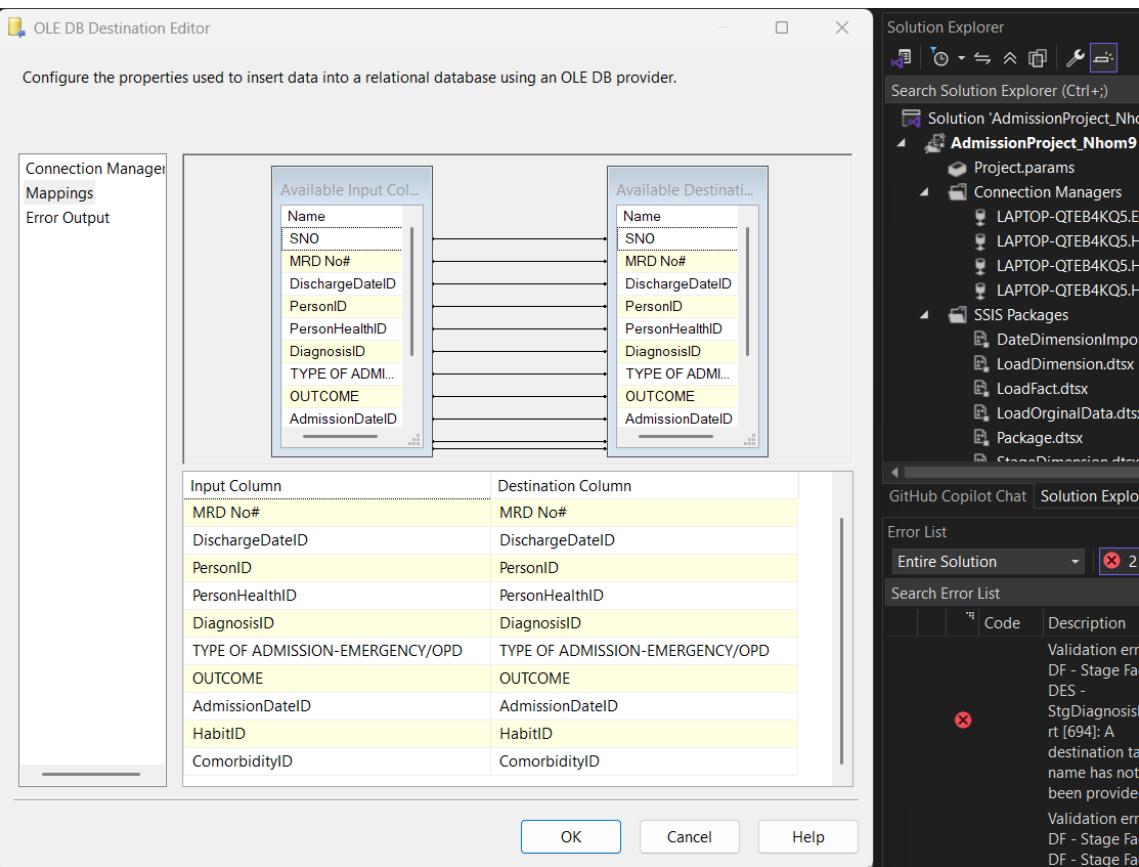
+ AdmissionReport:



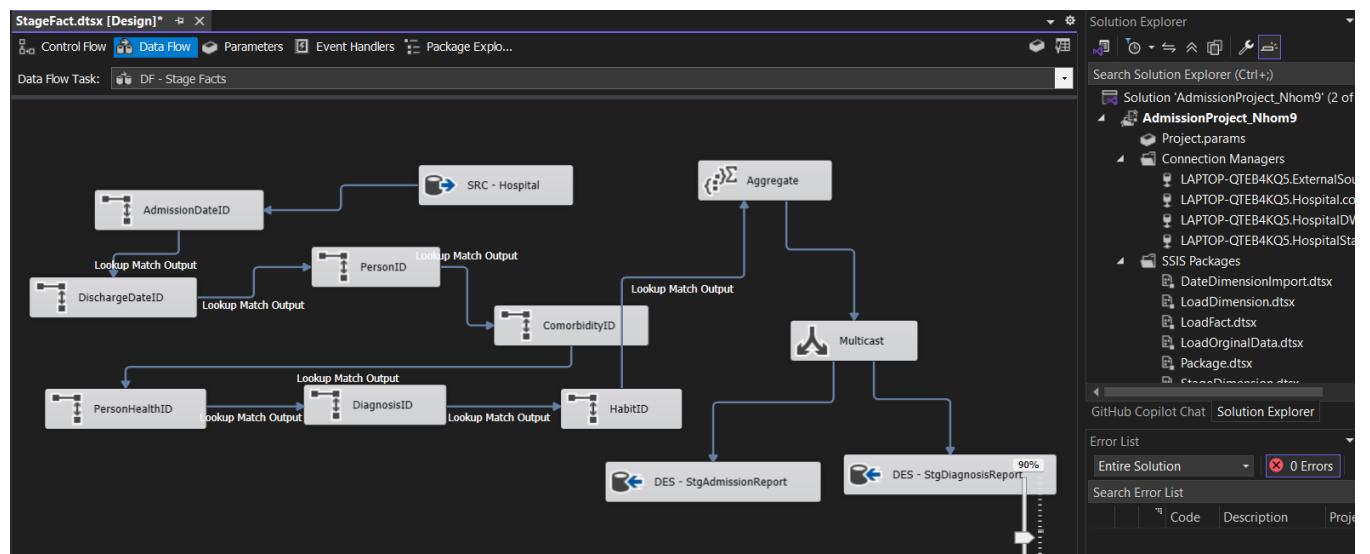


+ DiagnosisReport:

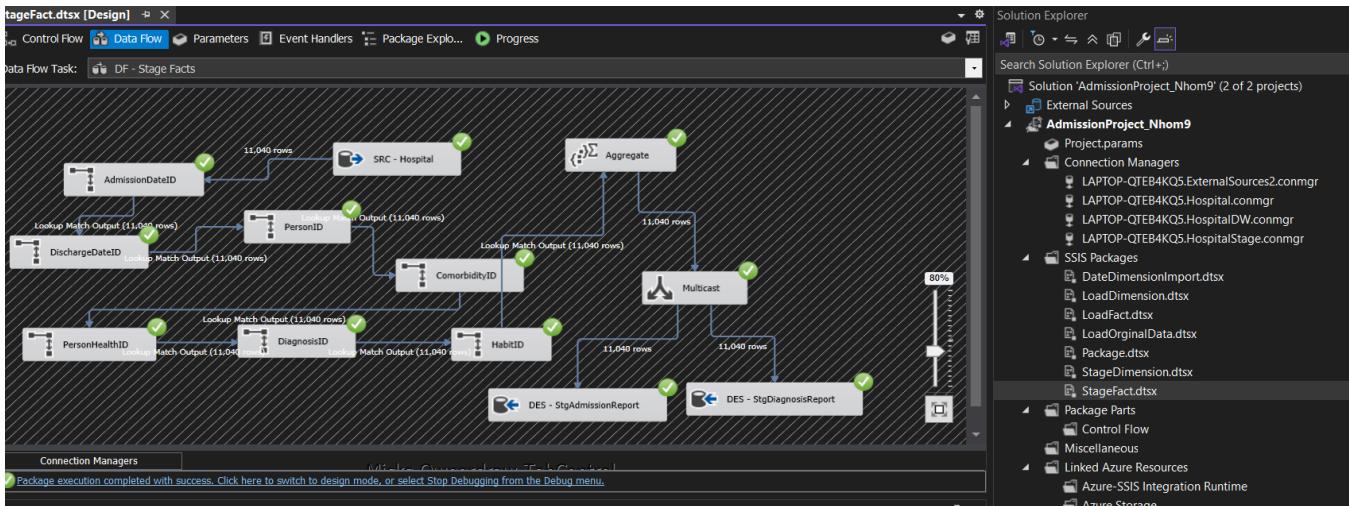




4. Data Flow Task của ta sẽ trông như sau:

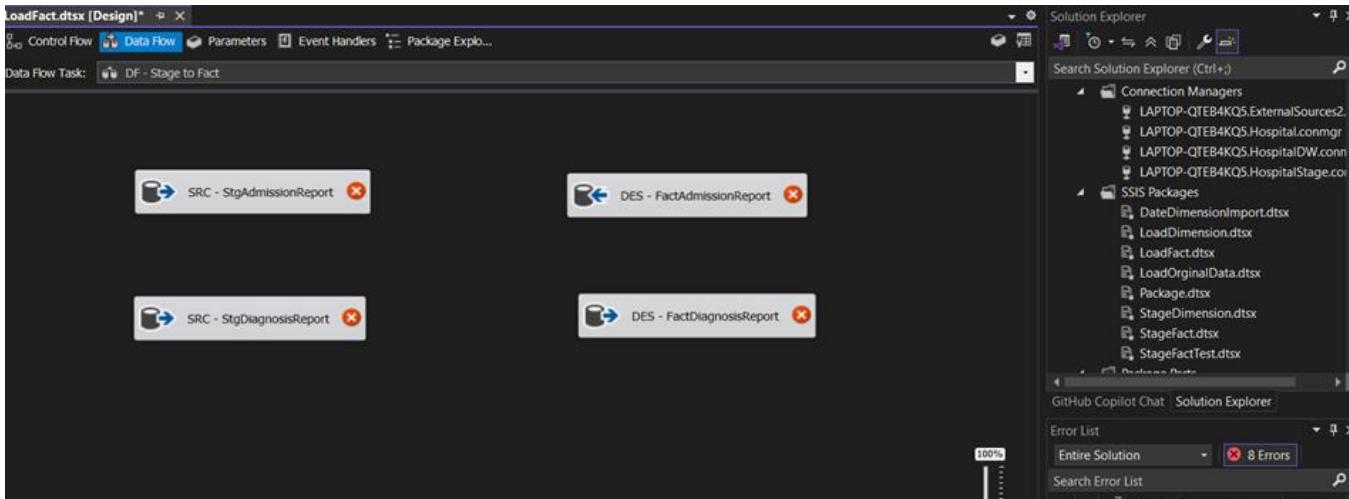


5. Tiến hành chạy Stage các bảng Fact



3.5.3. Load dữ liệu vào các bảng Fact

- Click vào DF - Stage to Fact và thêm 2 Source Assistant và 2 Destination Assistant.
- Tiến hành đổi tên và ta có được hình như sau:

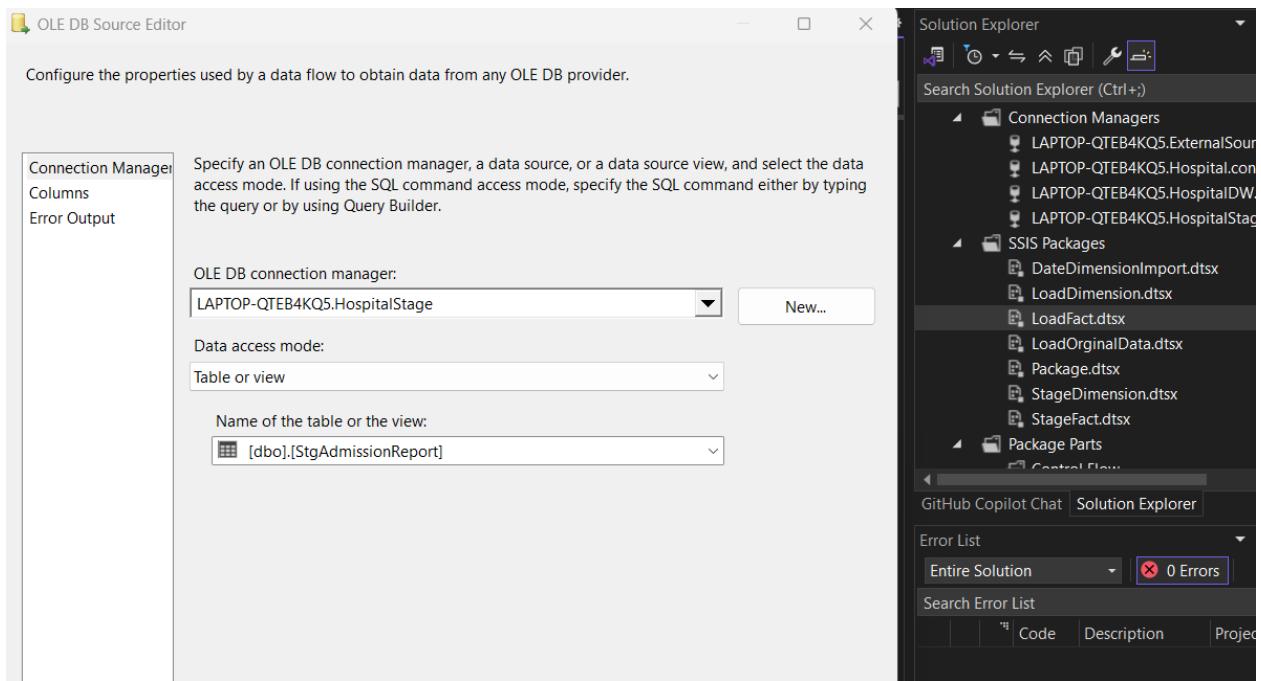


3. Cấu hình OLE DB Source và OLE DB Destination

- Cấu hình OLE DB Source cho StgAdmissionReport

Mục đích: Đọc dữ liệu báo cáo nhập viện từ bảng Stage.

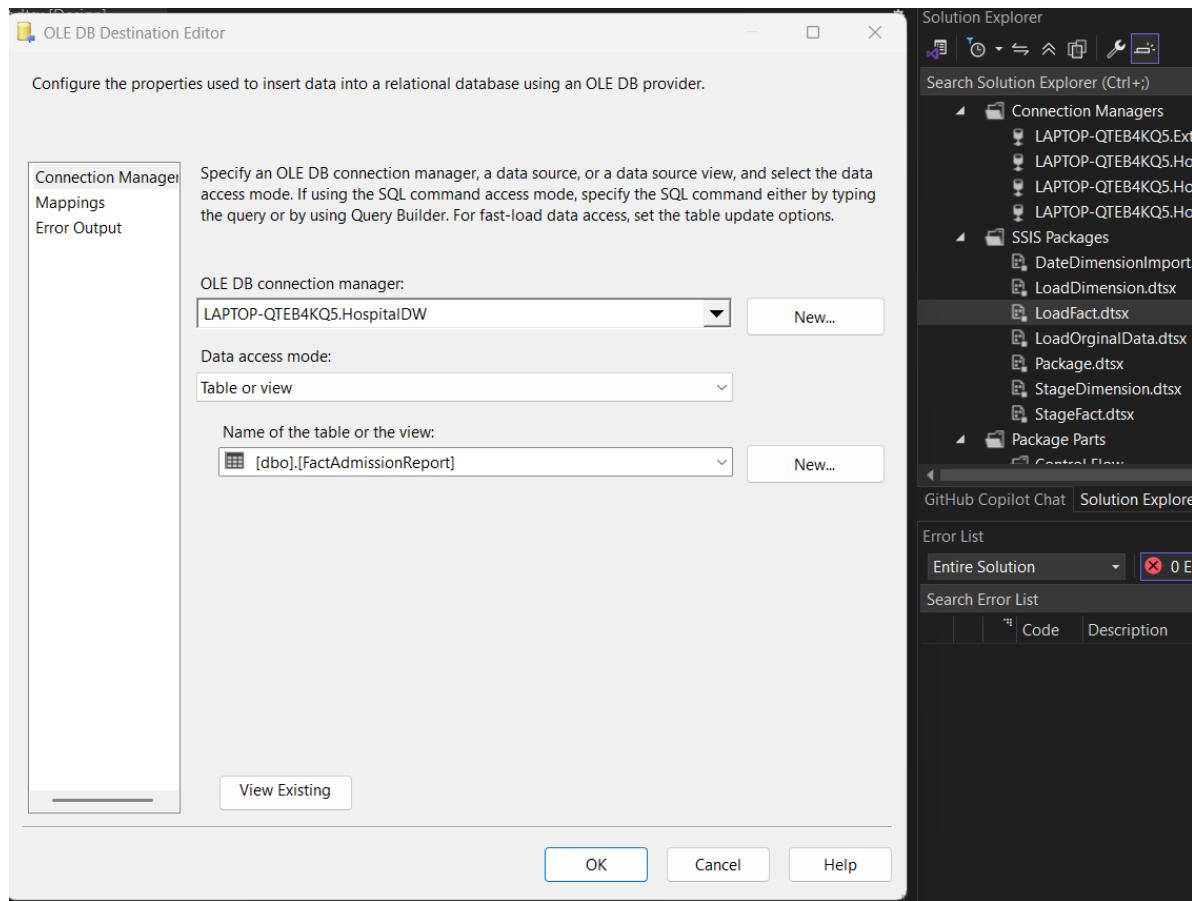
- + Connection Manager: LAPTOP-QTEB4KQ5.HospitalStage (Kết nối tới database Stage).
- + Data access mode: Table or view
- + Name of the table or the view: [dbo].[StgAdmissionReport]



- Cấu hình OLE DB Destination cho FactAdmissionReport

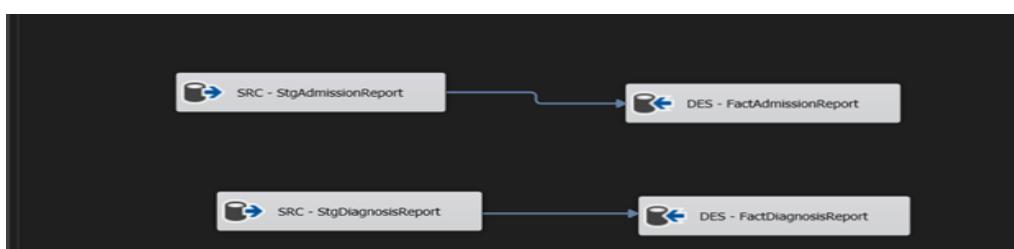
Mục đích: Ghi dữ liệu đã được xử lý vào bảng FactAdmissionReport trong Data Warehouse.

- + Connection Manager: LAPTOP-QTEB4KQ5.HospitalDW (Kết nối tới database Data Warehouse).
- + Data access mode: Table or view
- + Name of the table or the view: [dbo].[FactAdmissionReport]

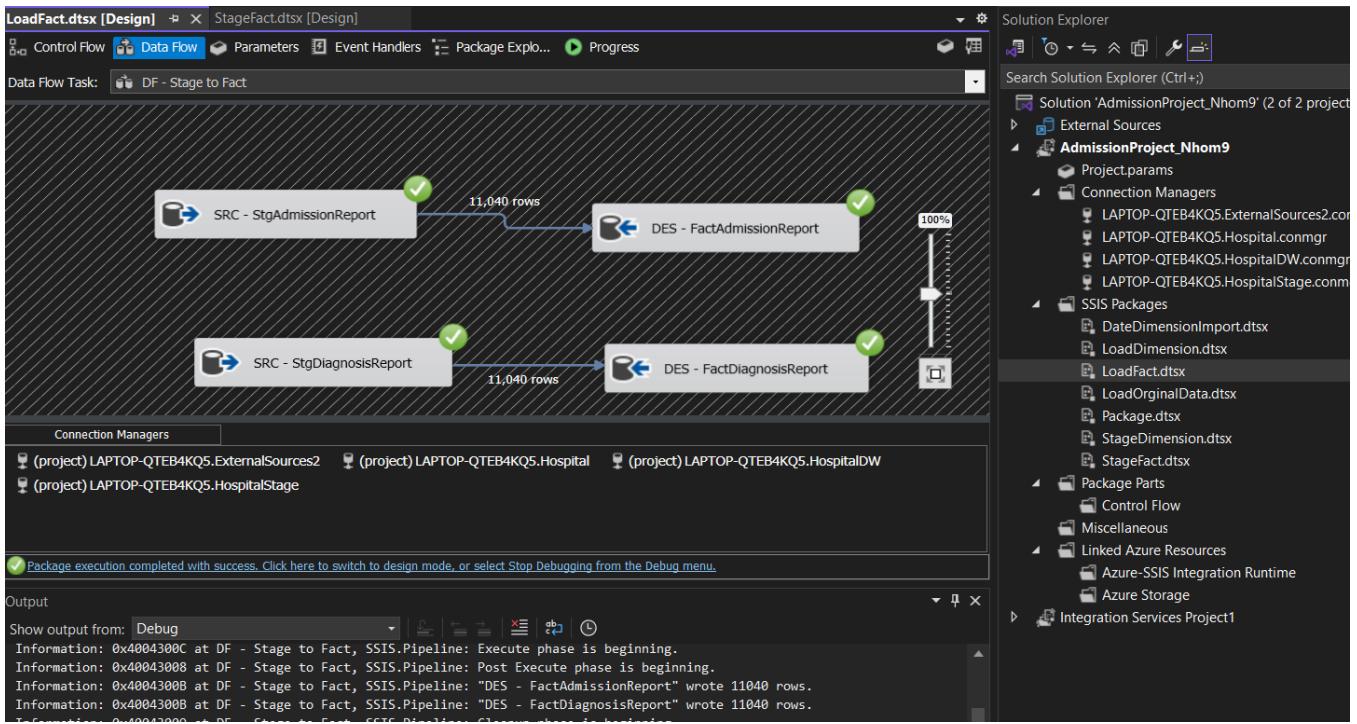


Tương tự, quá trình cấu hình OLE DB Source cho **StgDiagnosisReport** và OLE DB Destination cho **FactDiagnosisReport** cũng sẽ được thực hiện với các bước tương tự, trỏ đến các bảng Stage và Fact tương ứng.

4. Data Flow Task của ta sẽ trông như sau:



5. Tiến hành Load dữ liệu vào Fact



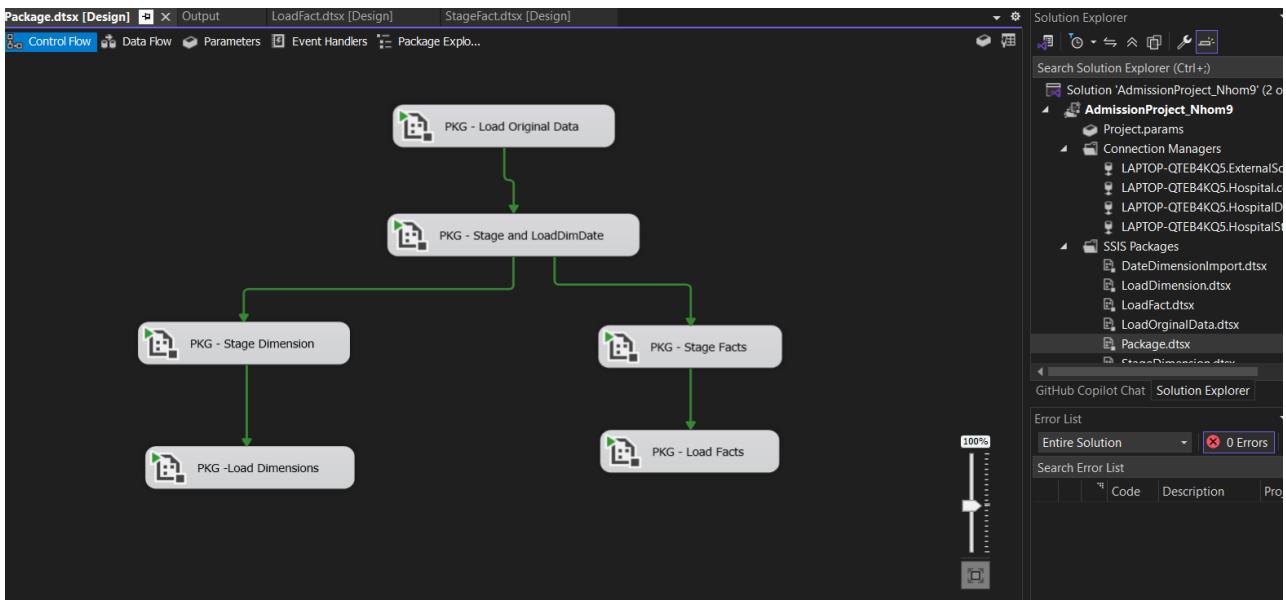
3.6. Tổng hợp quy trình ETL với SSIS

Để quản lý và tự động hóa toàn bộ luồng trích xuất, biến đổi và tải dữ liệu (ETL) vào Data Warehouse, chúng ta sẽ tổng hợp các gói SSIS con đã được xây dựng vào một gói điều khiển chính (**Package.dtsx**). Gói chính này sẽ chịu trách nhiệm thực thi tuần tự các gói con, đảm bảo quy trình ETL diễn ra một cách có tổ chức.

Ta tiến hành mở file **Package.dtsx** và kéo vào các **Execute Package Task** để gọi và thực thi các gói SSIS con.

1. **Stage ETL:** Gói tải dữ liệu vào khu vực Stage (**StageDimension.dtsx**, **StageFact.dtsx**).
2. **Dimension ETL:** Gói tải dữ liệu vào các bảng Dimension (**LoadDimension.dtsx**, **DateDimensionImport.dtsx**).
3. **Fact ETL:** Gói tải dữ liệu vào các bảng Fact (**LoadFact.dtsx**).

Các **Execute Package Task** này sẽ được nối với nhau bằng các đường nối ưu tiên (Precedence Constraints) để đảm bảo rằng một gói chỉ chạy sau khi gói trước đó đã hoàn thành thành công.

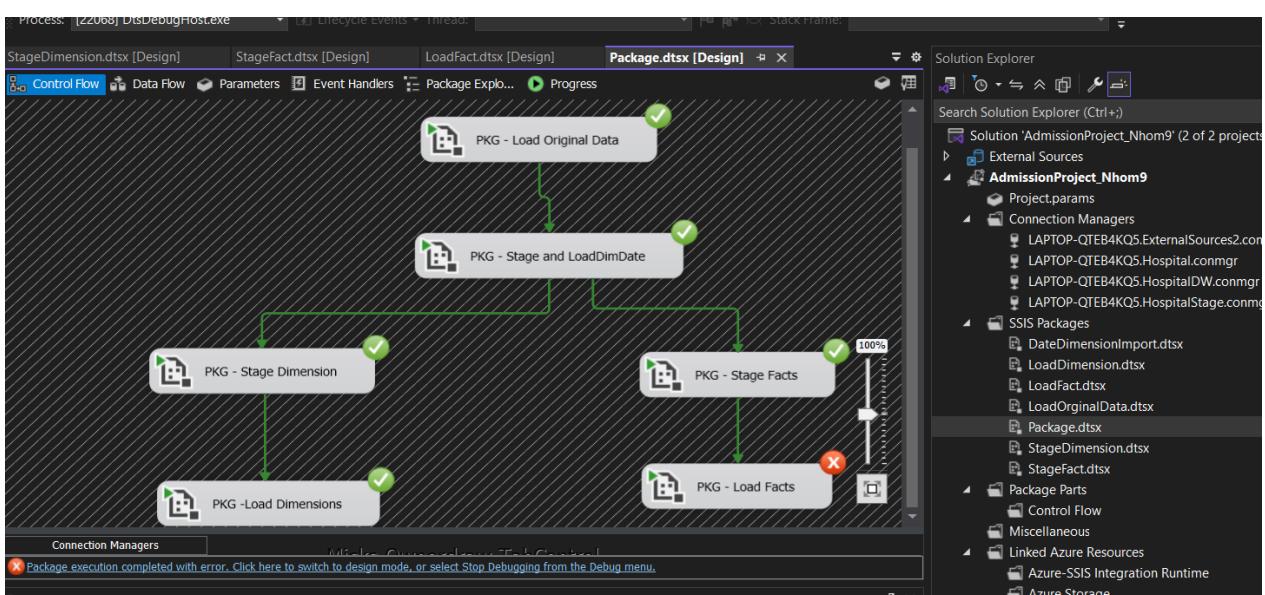


Dánh giá hiệu quả luồng ETL

Sau khi tổng hợp quy trình, việc đánh giá hiệu quả là rất quan trọng:

- **Thời gian thực thi:** Đo lường tổng thời gian cần thiết để hoàn thành toàn bộ luồng ETL.
- **Hiệu suất:** Kiểm tra hiệu suất của từng tác vụ và thành phần để xác định các điểm nghẽn.
- **Chất lượng dữ liệu:** Đảm bảo dữ liệu được tải vào Data Warehouse là chính xác, đầy đủ và nhất quán.

Lưu đồ quy trình tổng thể sẽ minh họa mối quan hệ và thứ tự thực thi của tất cả các gói SSIS trong giải pháp ETL.



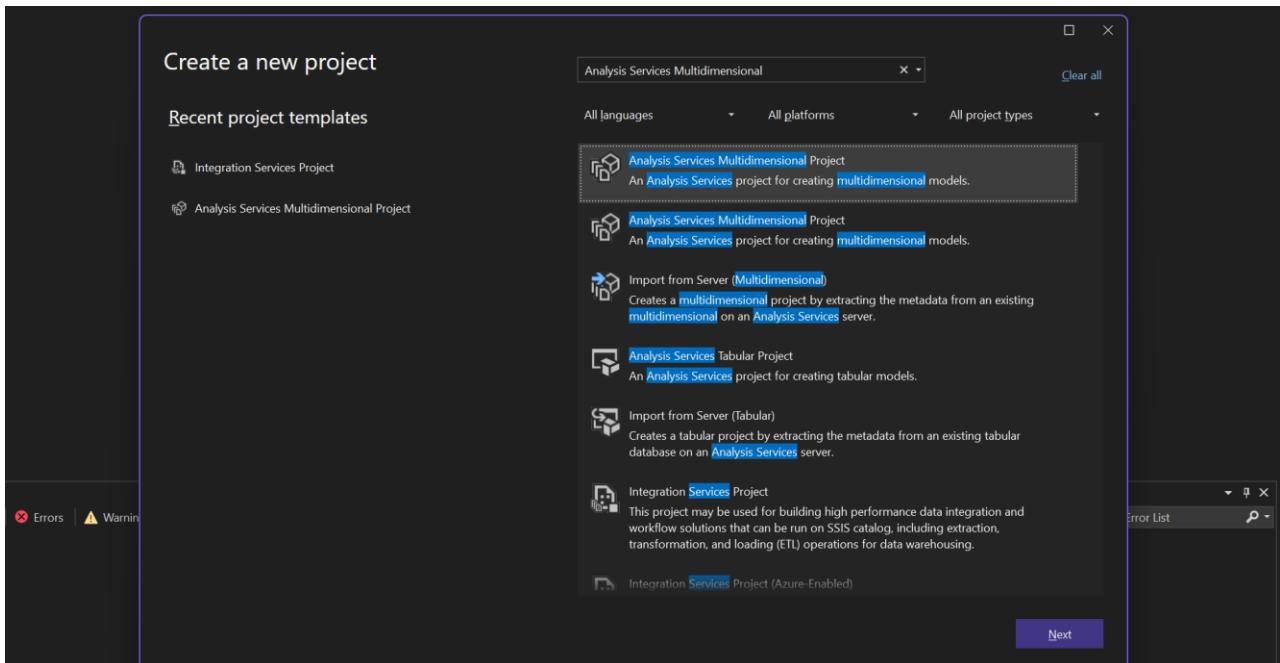
CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU VỚI SSAS

Trong chương này, chúng ta sử dụng *SQL Server Analysis Services (SSAS)* để xây dựng mô hình phân tích dữ liệu dạng đa chiều. Quy trình bao gồm các bước chính như: tạo mô hình OLAP, xây dựng Cube, tạo Hierarchy, truy vấn dữ liệu bằng MDX, và phân tích trực quan qua Excel Pivot Table. Kết quả được đối chiếu với phân tích thủ công nhằm đảm bảo độ chính xác.

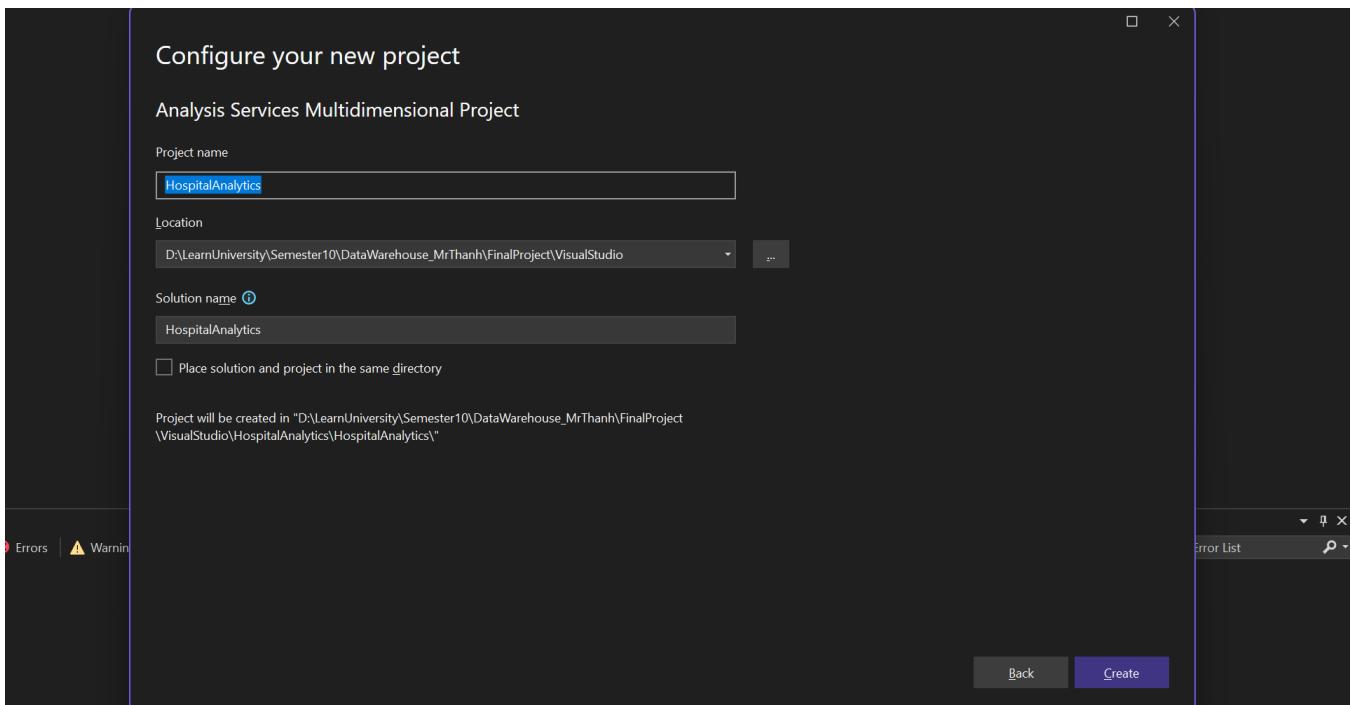
4.1. Xây dựng mô hình OLAP trong SSAS

4.1.1. Tạo Project SSAS

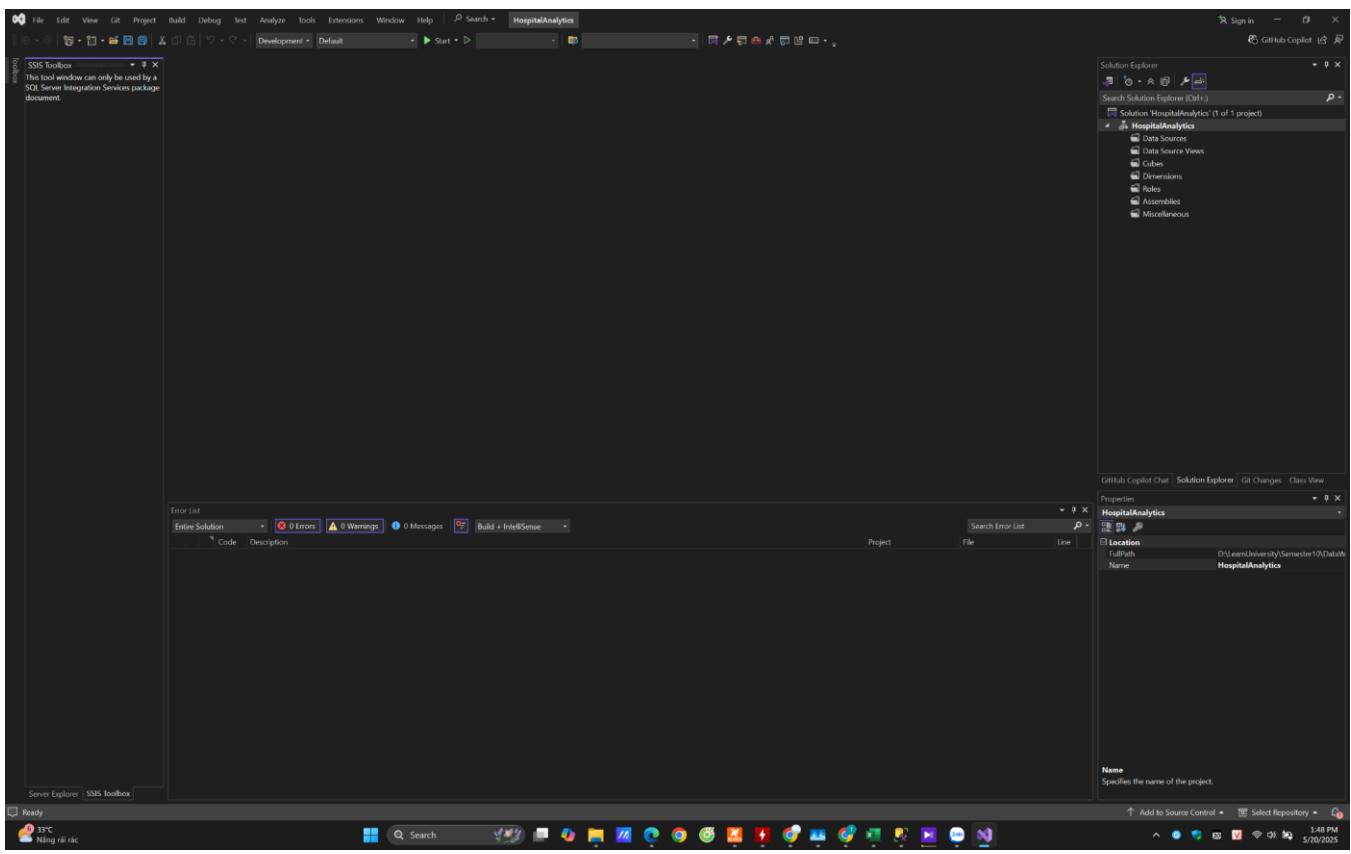
- Mở **Visual Studio** → File → New → Project.
- Chọn template **Analysis Services Multidimensional Project** → Next



- Đặt tên Project, nhấn **Create** để tạo.

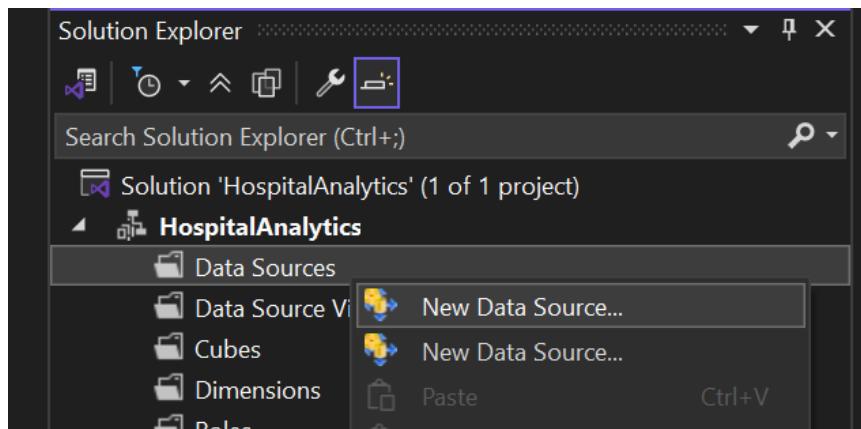


Giao diện của SSAS sau khi tạo một project mới như hình bên dưới:

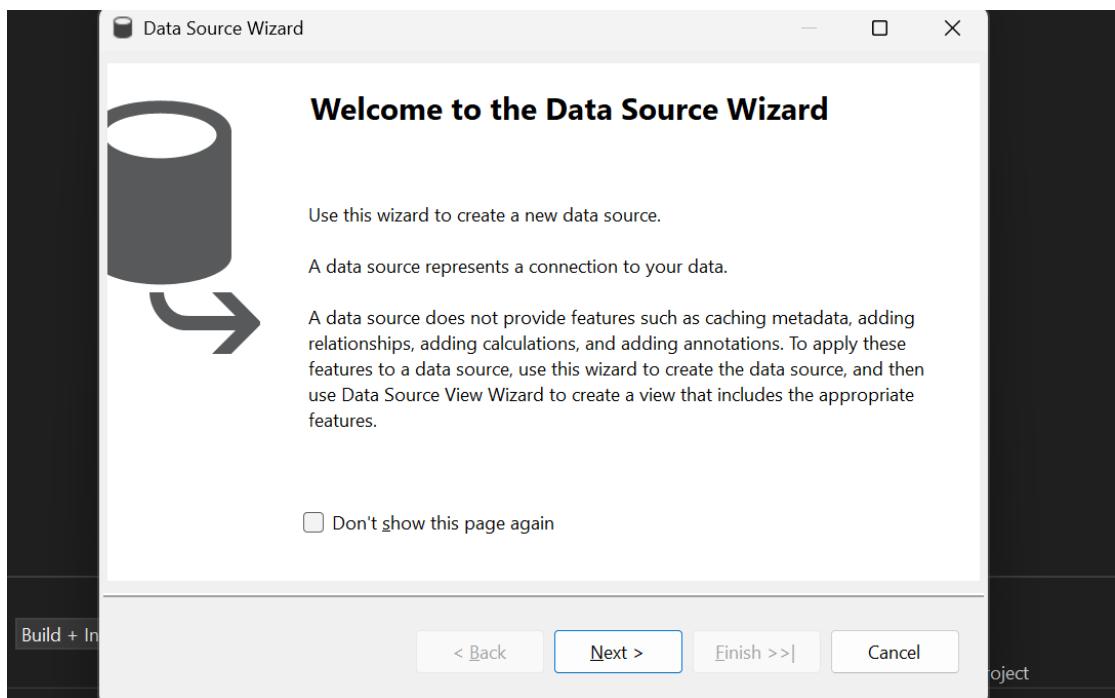


4.1.2. Tạo Data Source

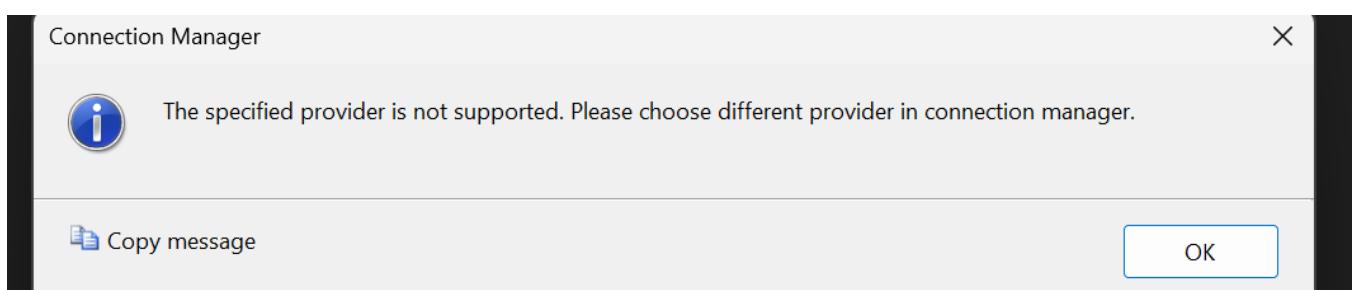
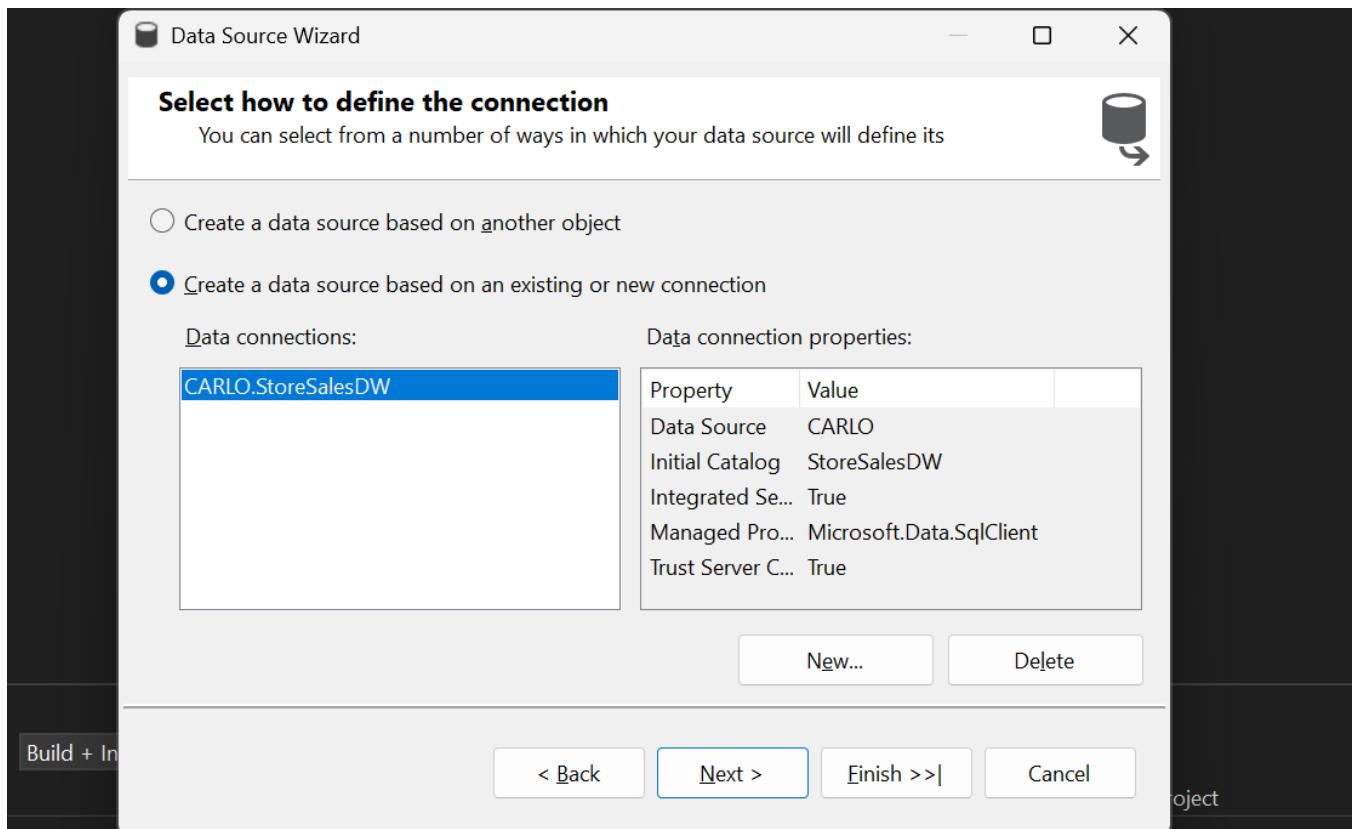
- Trong **Solution Explorer**: Chuột phải Data Sources → New Data Source.



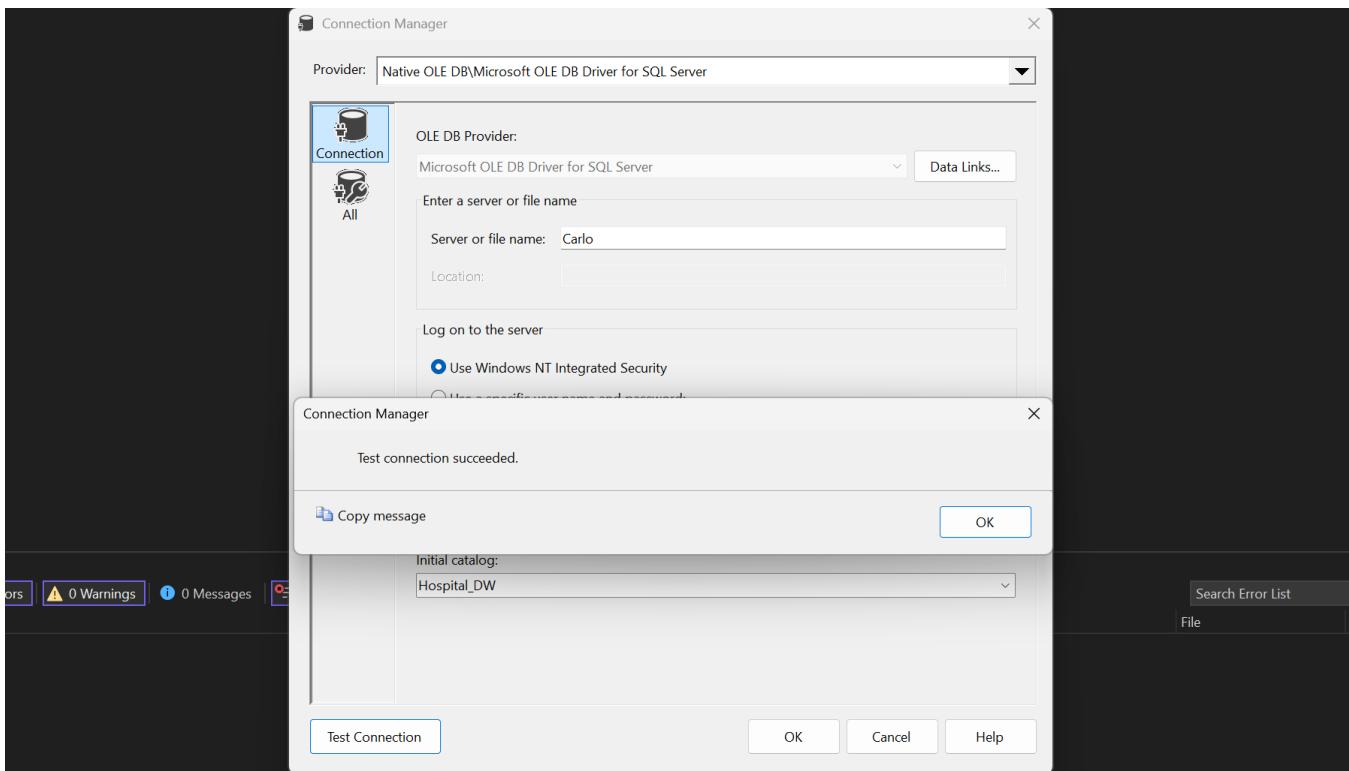
- Kết nối tới cơ sở dữ liệu Hospital_DW: dialog Data Source Wizard hiện lên → Next



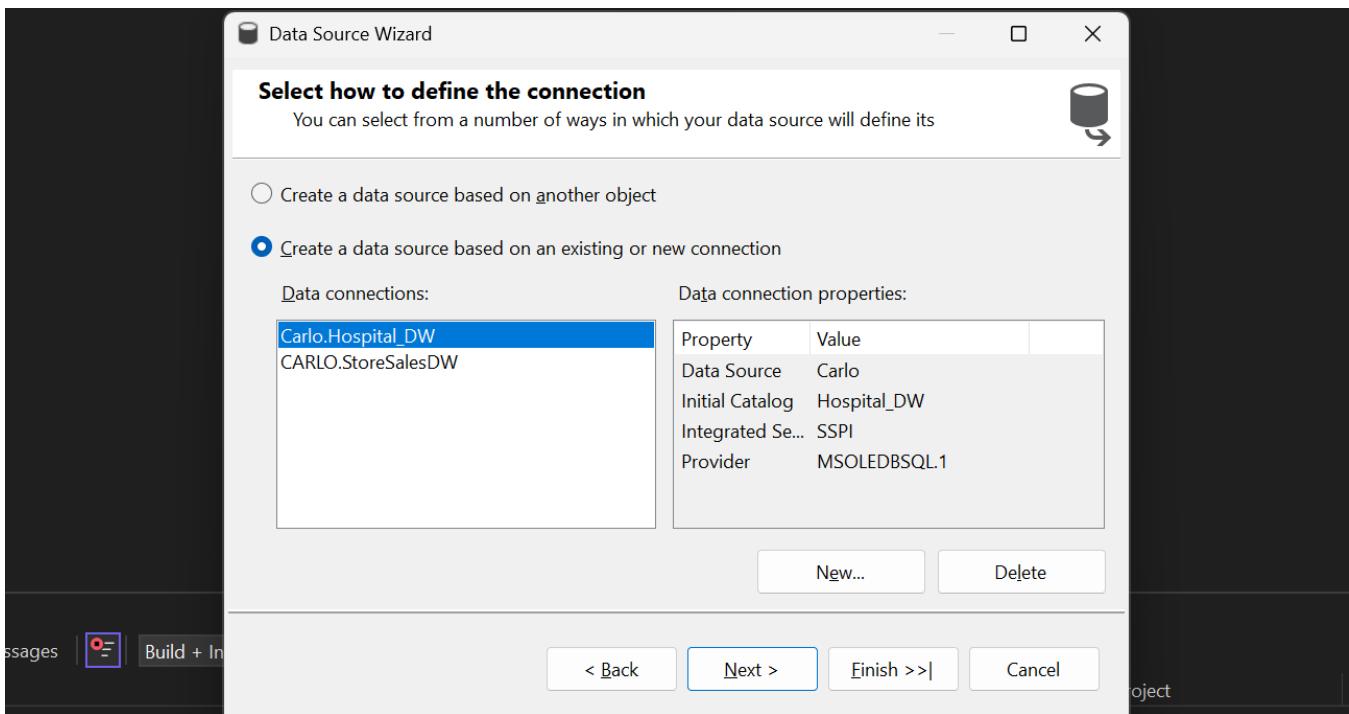
Chọn Create a data source based on an existing or new connection → New...



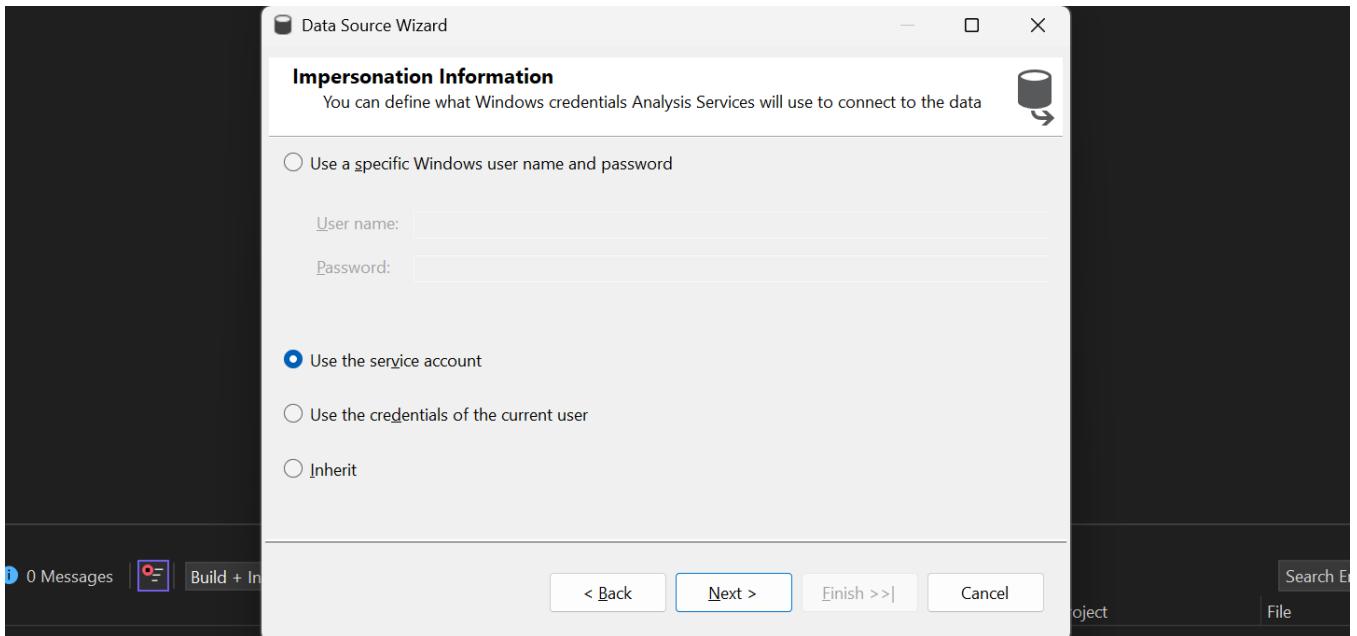
Chọn Provider như hình → điền Server name → chọn DB cần kết nối → Test Connection và hiện ra được như bên dưới là kết nối thành công.



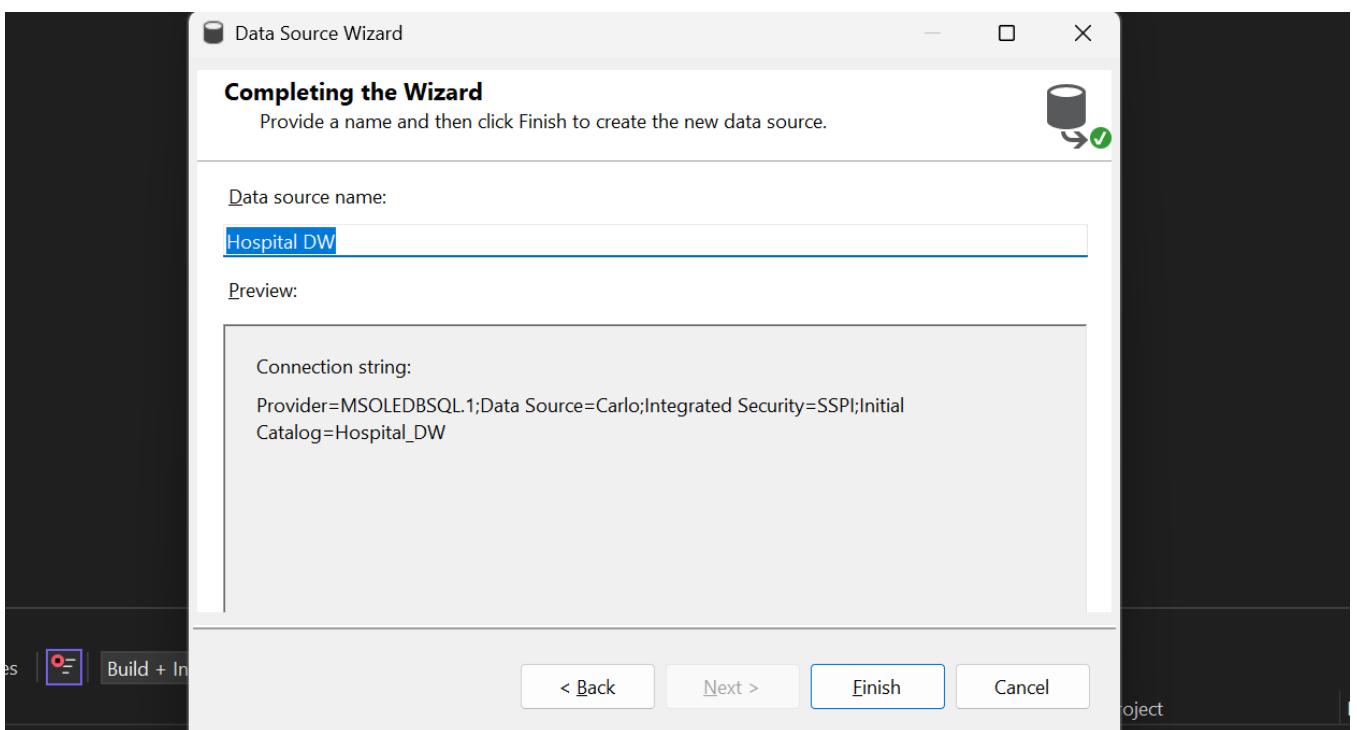
Chọn data vừa tạo và bấm Next



- Chọn chế độ Use the service account → Next

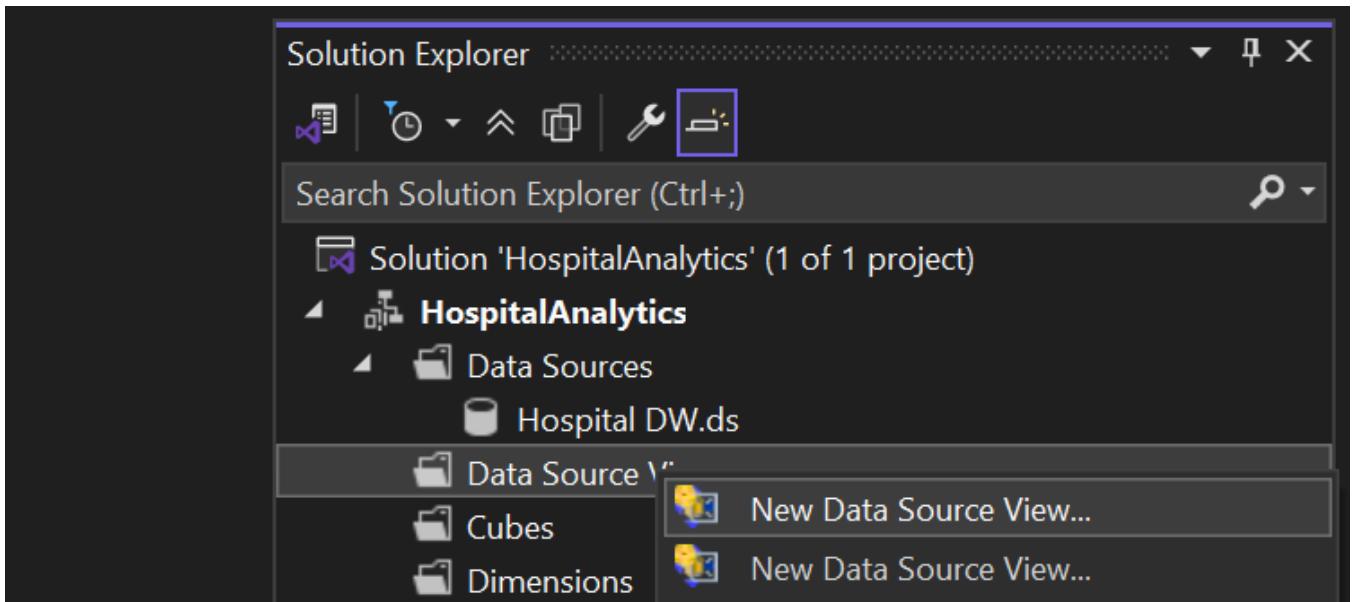


- Nhấn Finish để hoàn tất

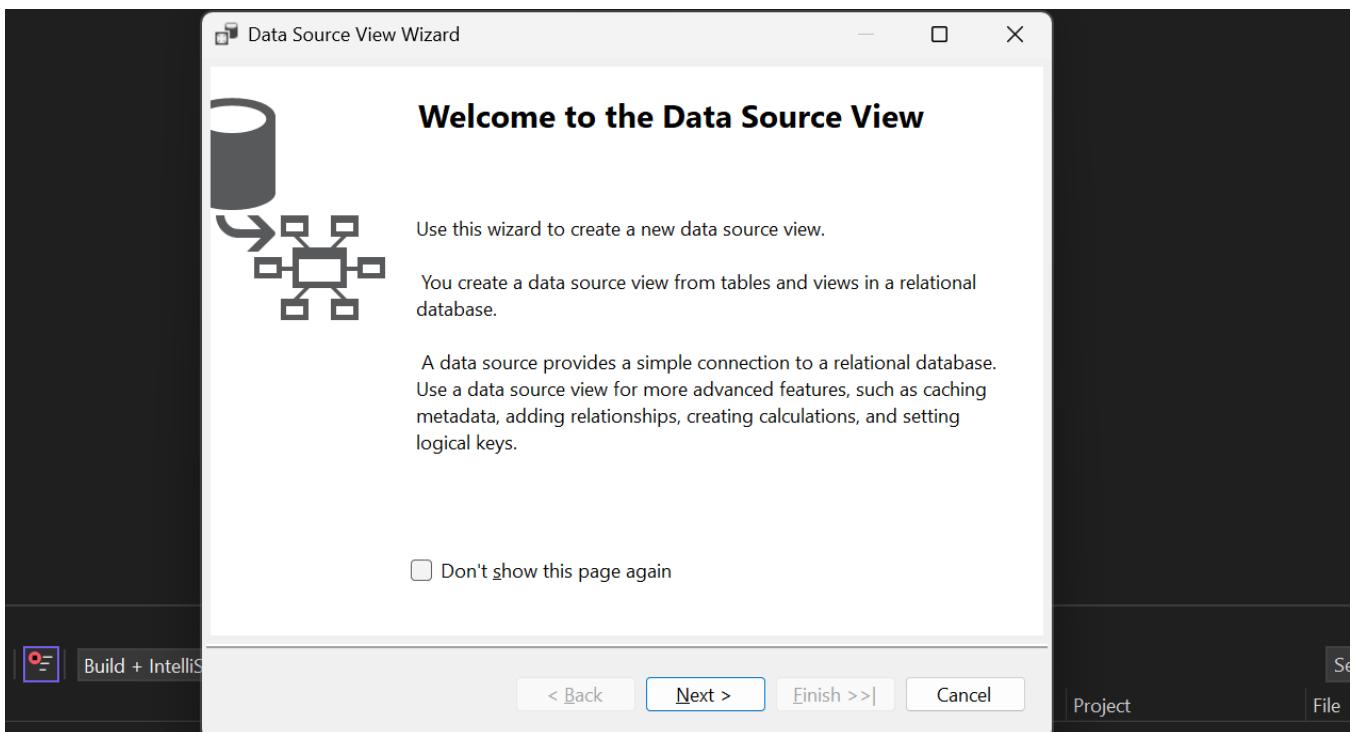


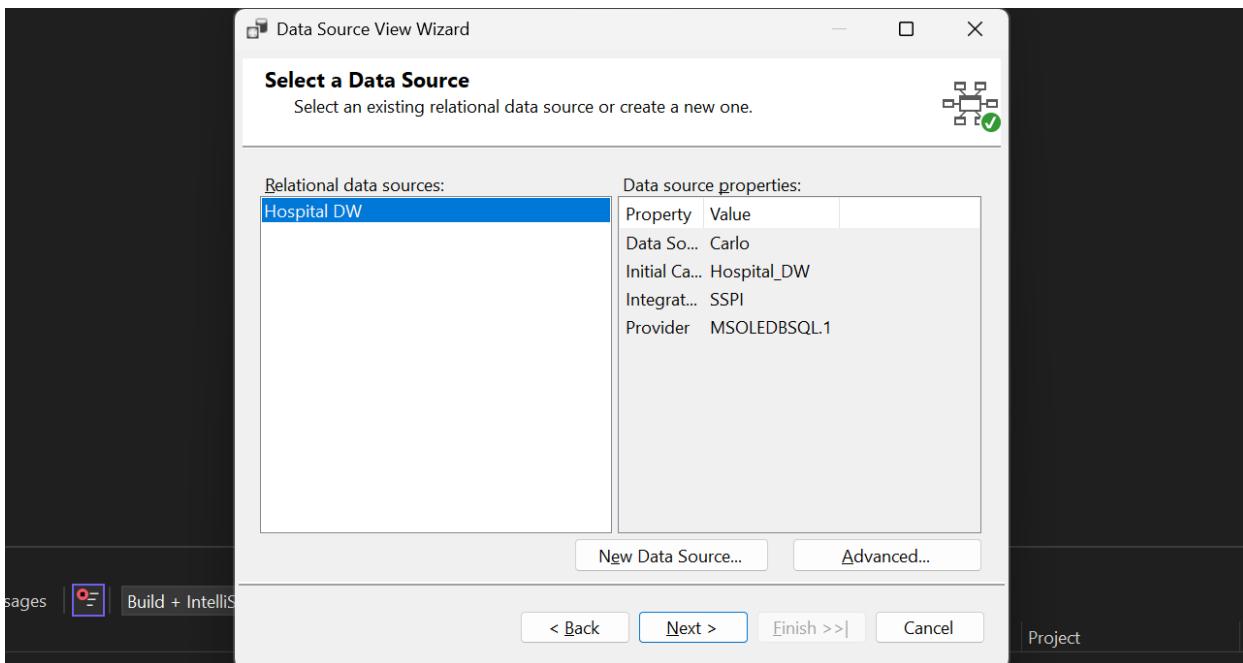
4.1.3. Tạo Data Source View

- Chuột phải Data Source Views → New Data Source View.

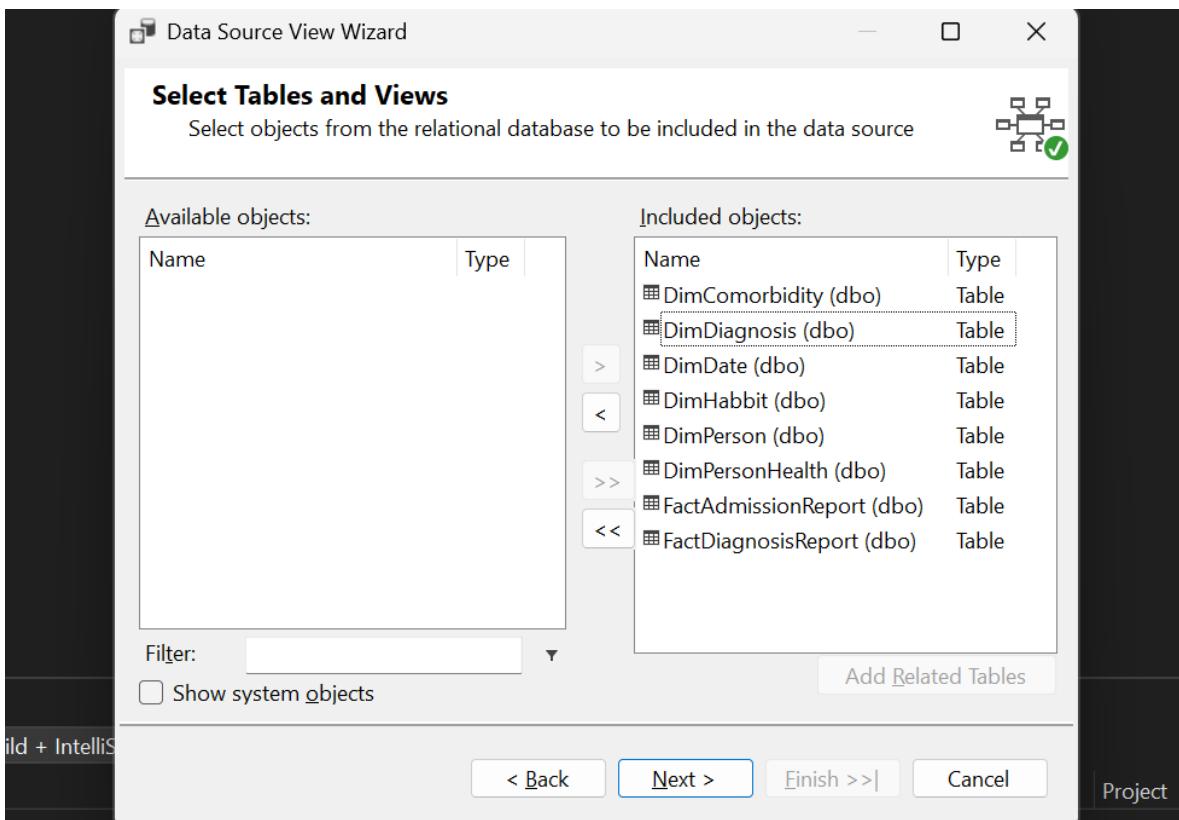


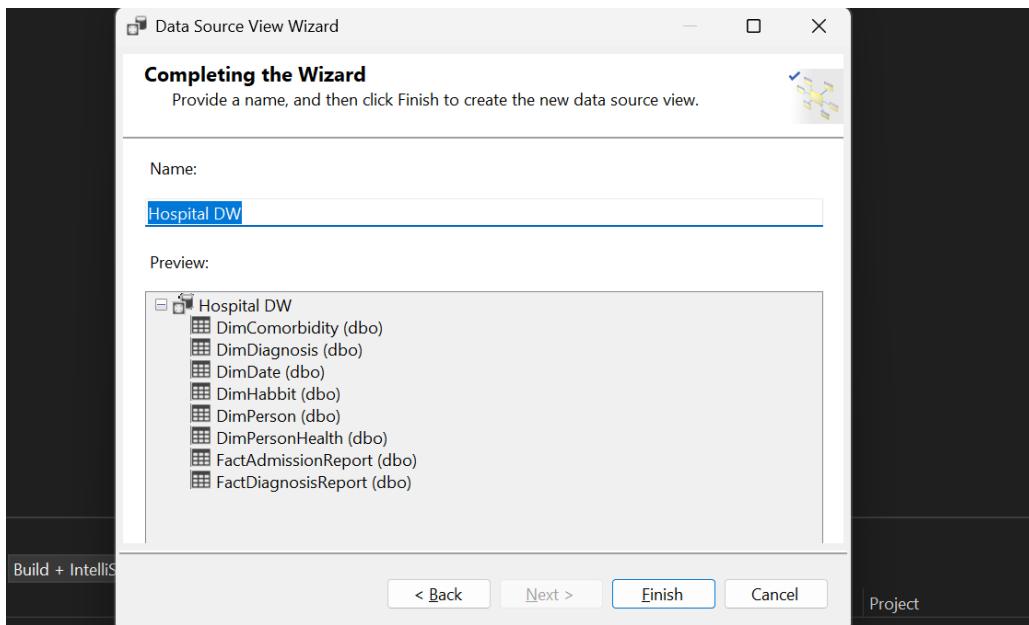
- Dialog Data Source View Wizard hiện lên → Next → chọn Hospital_DW làm nguồn dữ liệu
→ Next



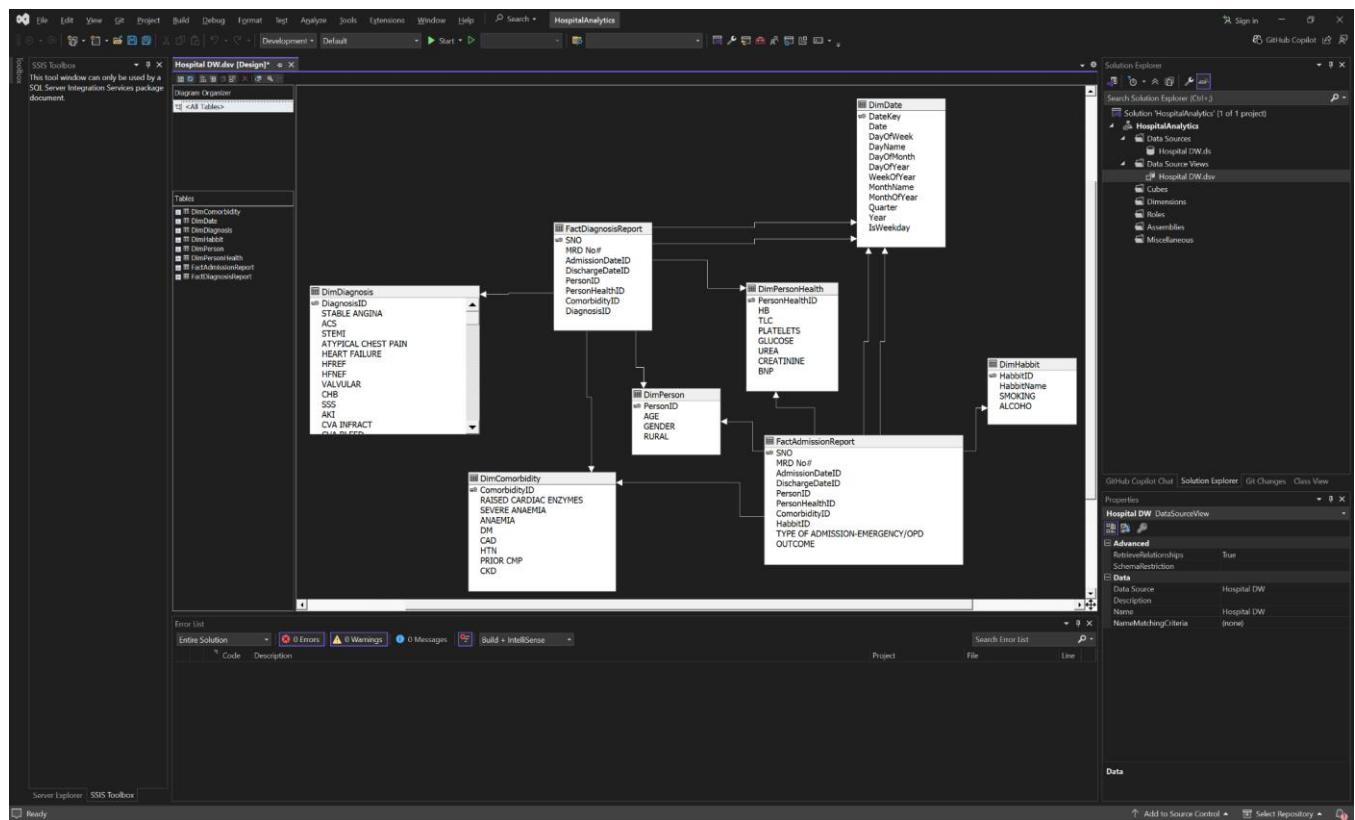


- Chọn hết, nhấn Next → Finish.





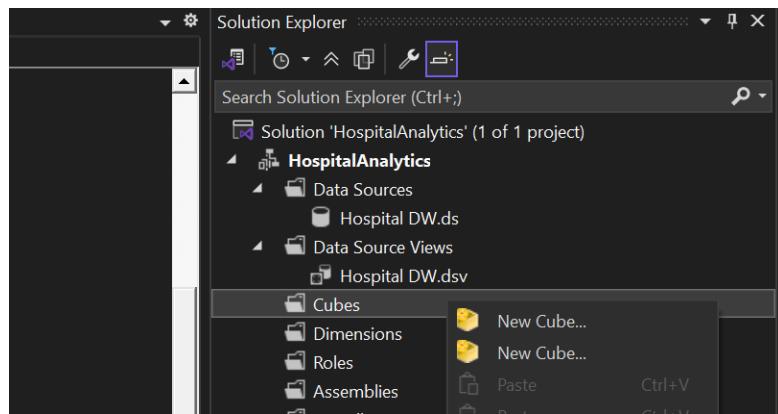
Nhấn 2 lần vào Hospital_DW bên dưới Data Source Views, kết quả hiện ra một **Star Schema** hiển thị tại Design Surface như sau



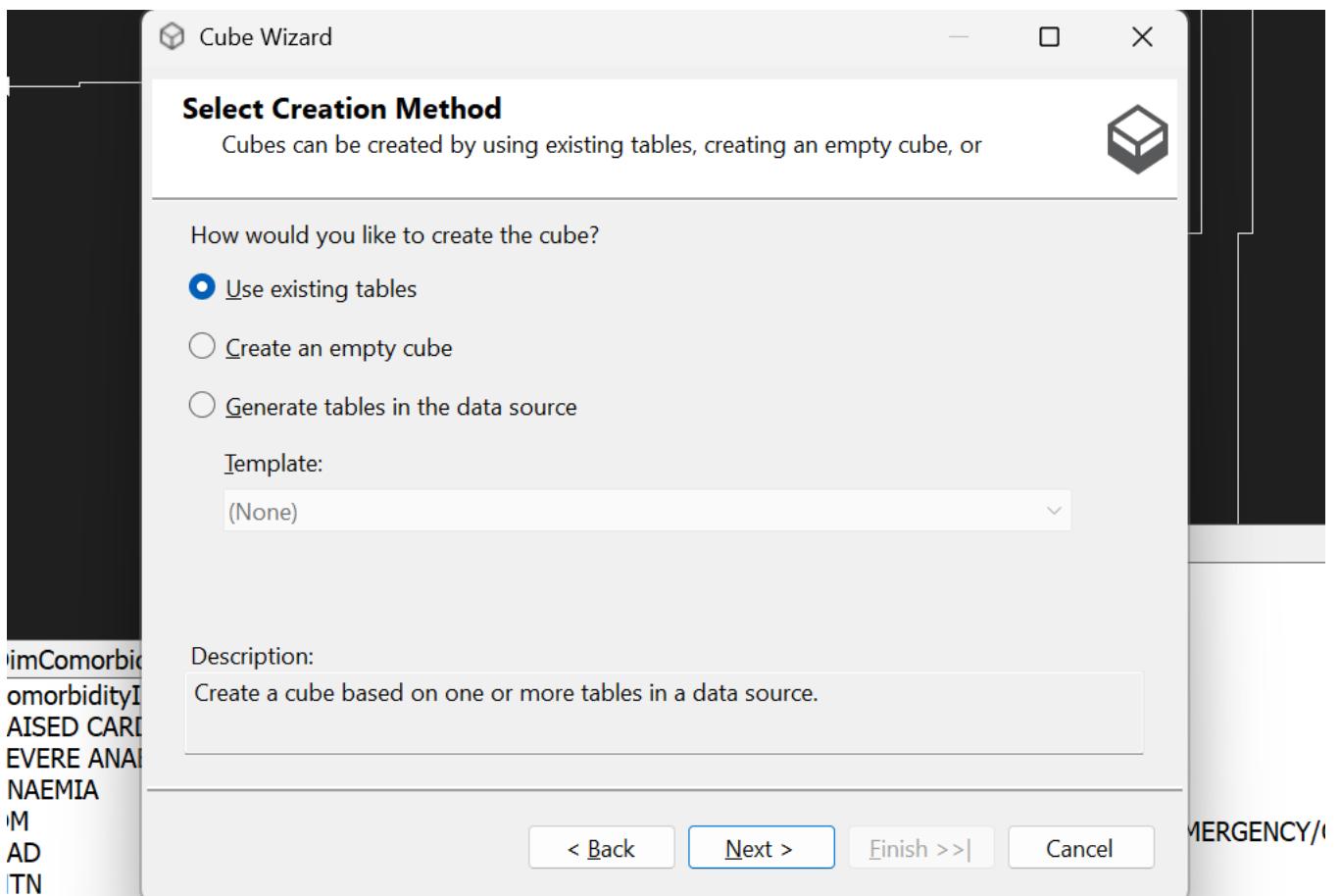
4.2. Tạo và cấu hình Cube

4.2.1. Khởi tạo Cube

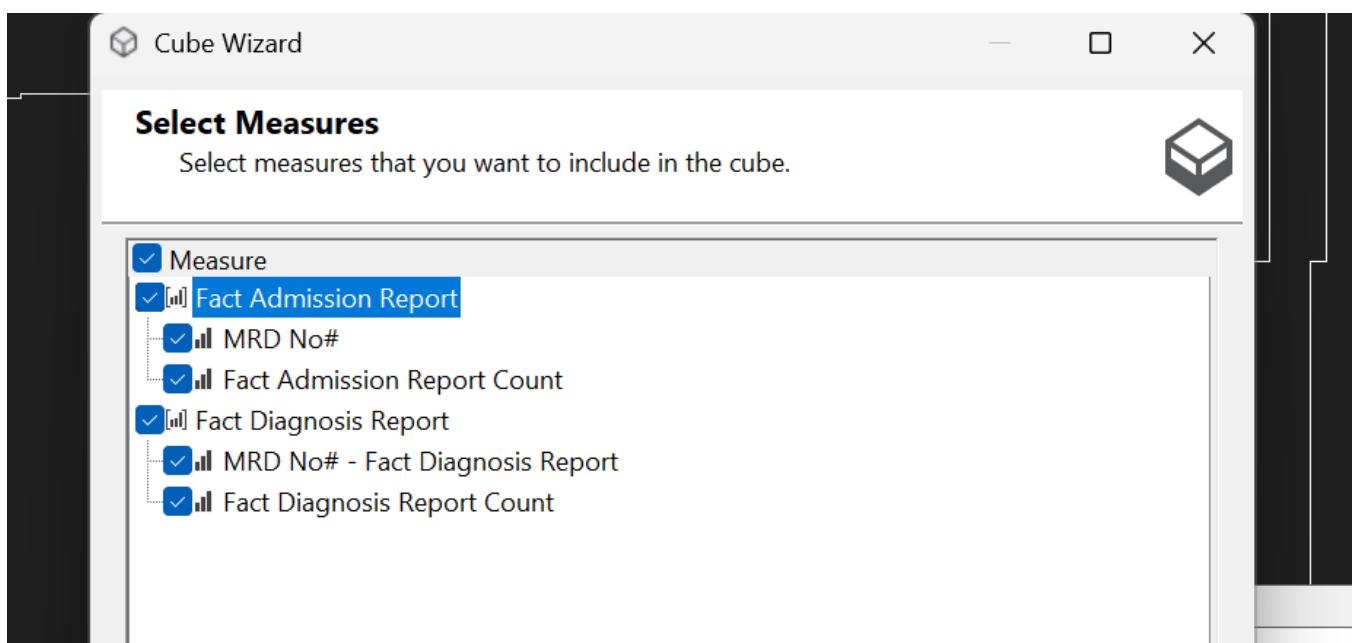
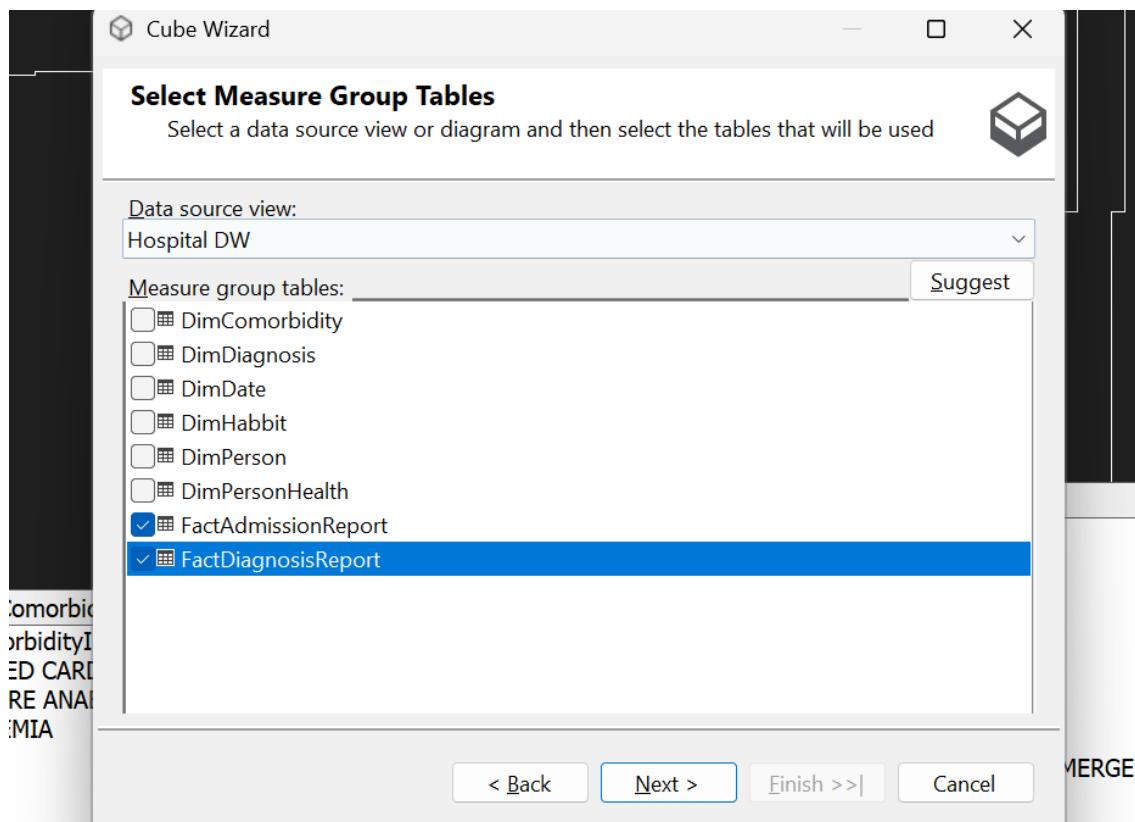
Trong Solution Explorer, chuột phải vào Cubes → New Cube → Next

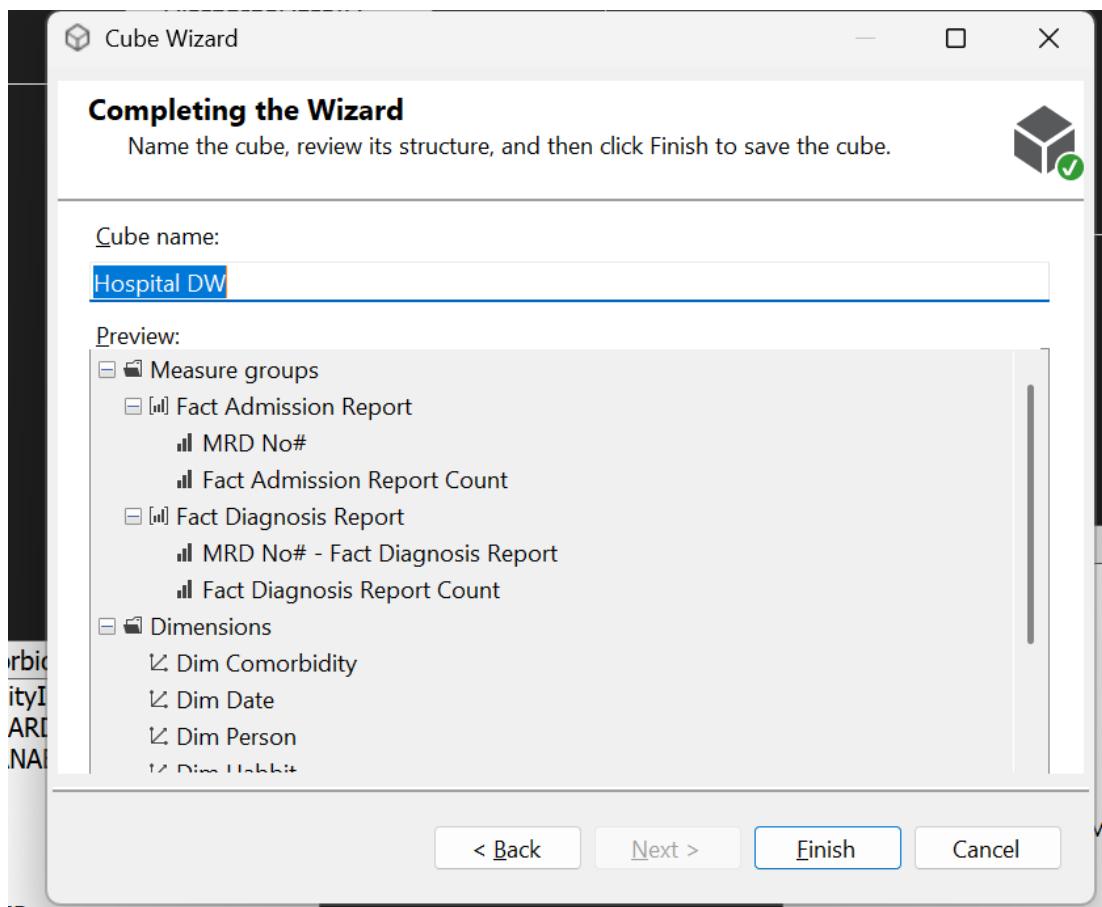


- Chọn Use existing tables → Next, sử dụng Data Source View vừa tạo.



Chọn các bảng Fact như FactAdmissionReport, FactDiagnosisReport → Next → Next → Next



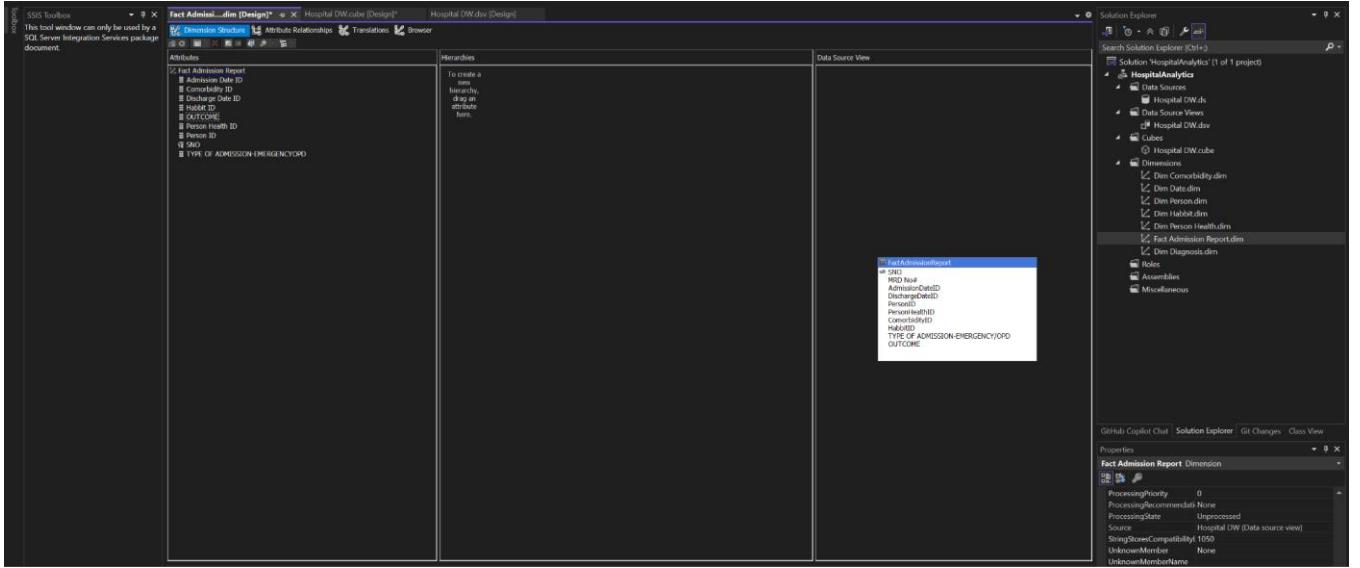


- SSAS tự động tạo các Dimension, người dùng có thể điều chỉnh sau.

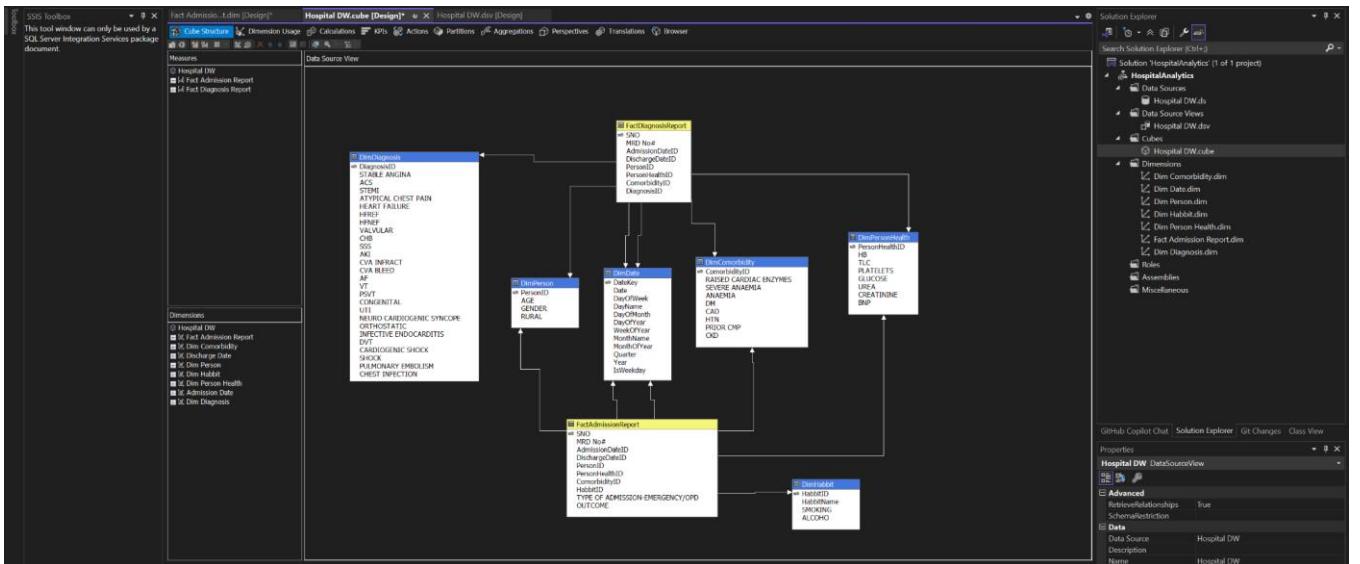
4.2.2. Cập nhật thuộc tính cho Dimension

Sau khi Cube được tạo tự động bởi Wizard, một số thuộc tính quan trọng nằm trực tiếp trong bảng Fact (chẳng hạn như **Loại Hình Nhập Viện** và **Kết Quả Điều Trị** trong **FactAdmissionReport**) cần được định nghĩa rõ ràng để người dùng có thể sử dụng chúng cho việc phân tích. Để làm điều này, nhóm tiến hành chỉnh sửa Dimension được tạo tự động từ bảng Fact (thường gọi là 'Fact Dimension' hoặc 'Degenerate Dimension'). Cụ thể, các cột **TYPE OF ADMISSION-EMERGENCY/OPO** và **OUTCOME** được kéo từ Data Source View vào danh sách thuộc tính (Attributes) của Dimension này. Điều này cho phép các giá trị của **Loại Hình Nhập Viện** và **Kết Quả Điều Trị** được sử dụng như các chiều phân tích trong Cube, giúp người dùng có thể lọc và nhóm dữ liệu theo các tiêu chí này.

Click vào Fact Admission Report.dim trong Solution Explorer, kéo **TYPE OF ADMISSION-EMERGENCY/OPO** và **OUTCOME** từ Data Source View vào Attributes:



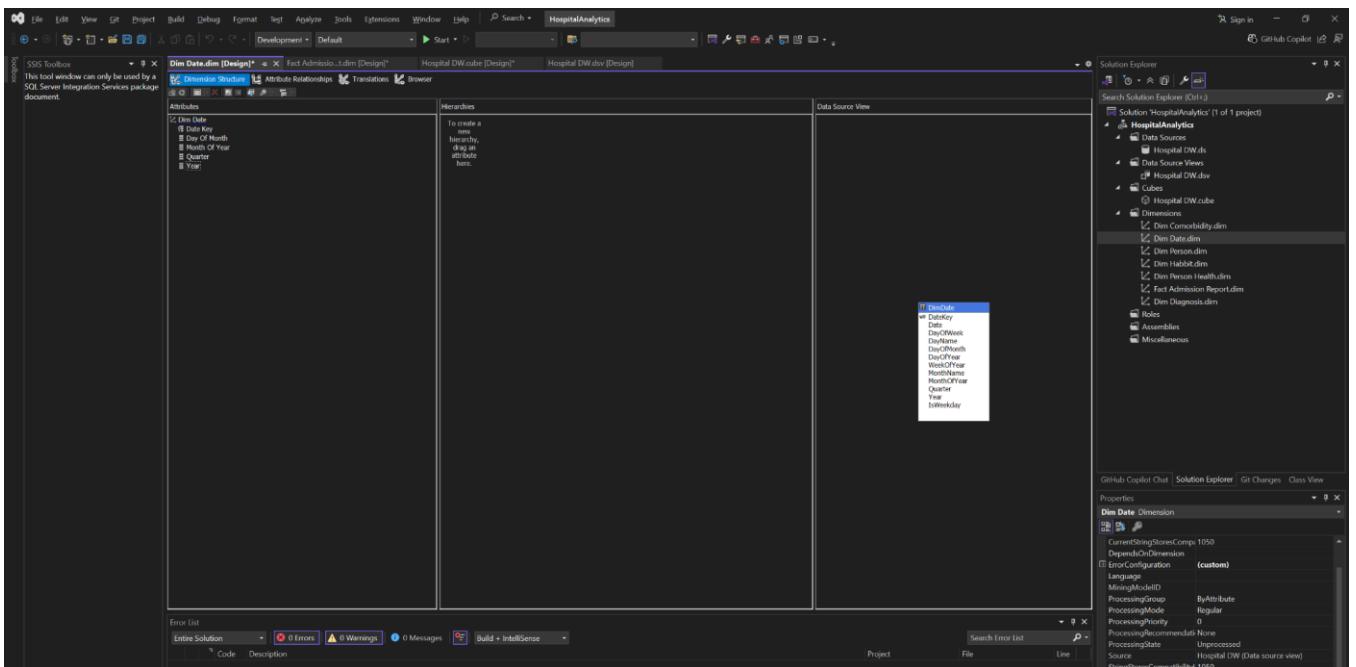
Ta sẽ có được kết quả như sau:



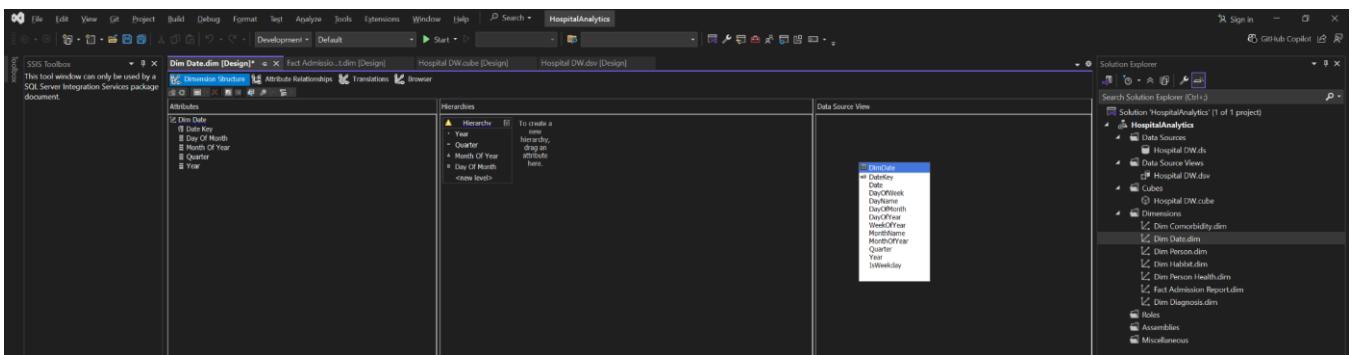
4.3. Tạo Hierarchy (Phân cấp thuộc tính)

4.3.1. Phân cấp thời gian – DimDate

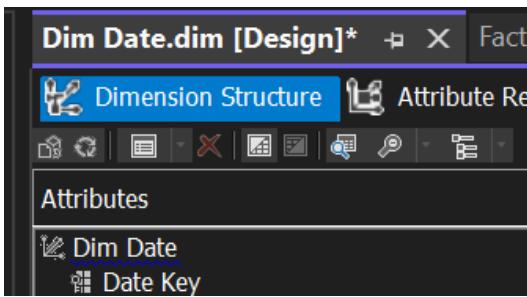
- Trong Solution Explorer, double-click vào Dim Date.dim. Kéo các trường Day Of Month, Month Of Year, Quarter, Year từ mục Data Source View vào Attributes.

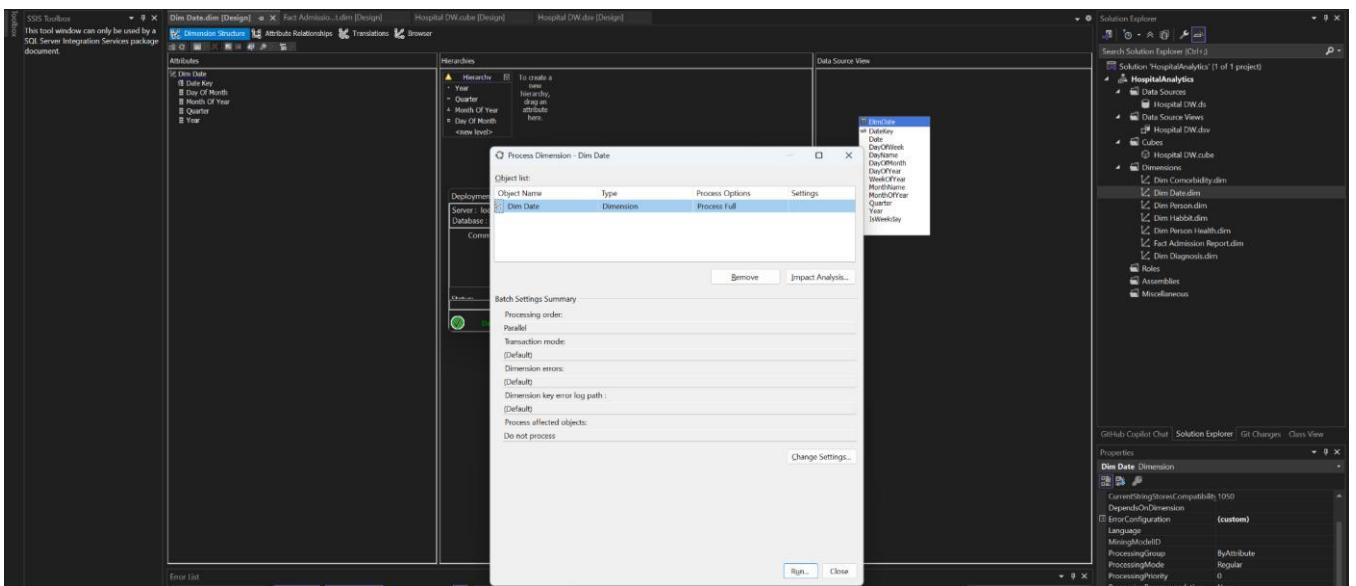
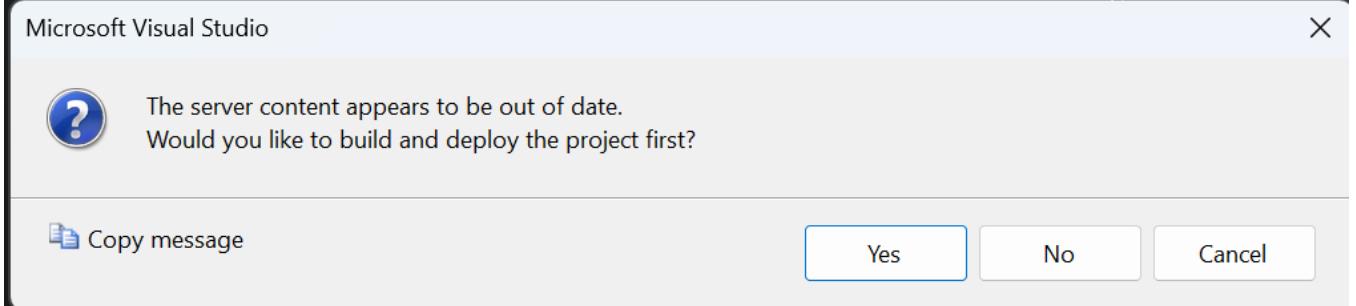


- Tạo Hierarchy theo thứ tự từ năm → quý → tháng → ngày. Kéo từng biến Year, Quarter, Month Of Year, Day Of Month từ Attributes qua Hierarchies.

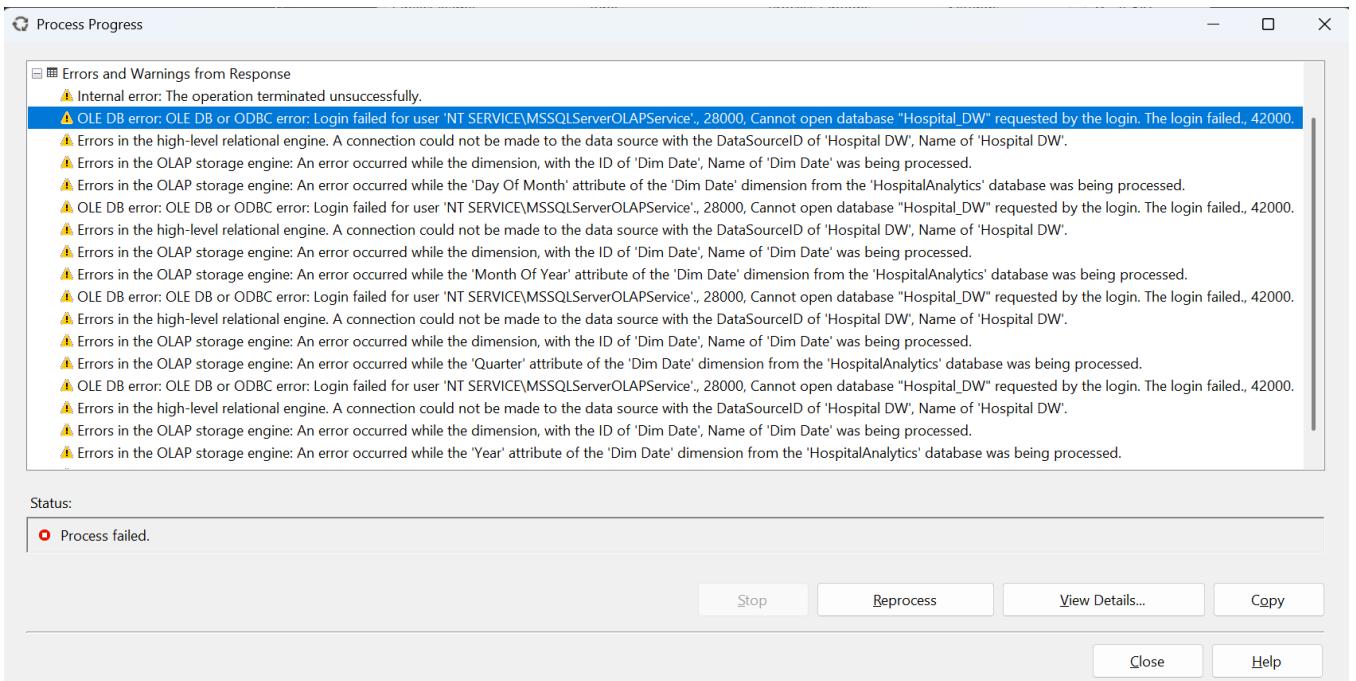


Click **Process** icon (⚙️ - hình dưới) ở toolbar click **Yes** khi được hỏi để build và deploy project.

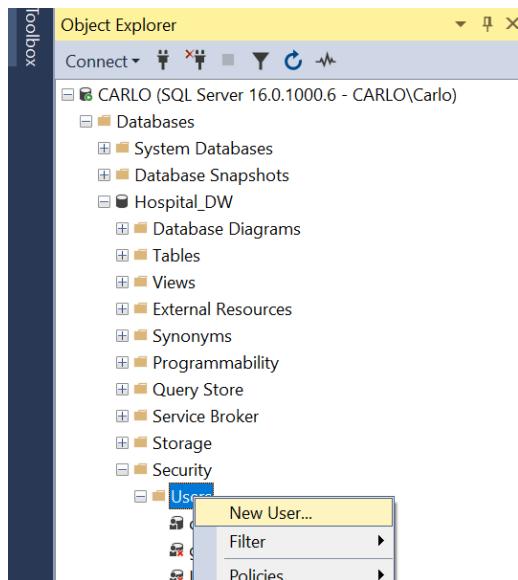




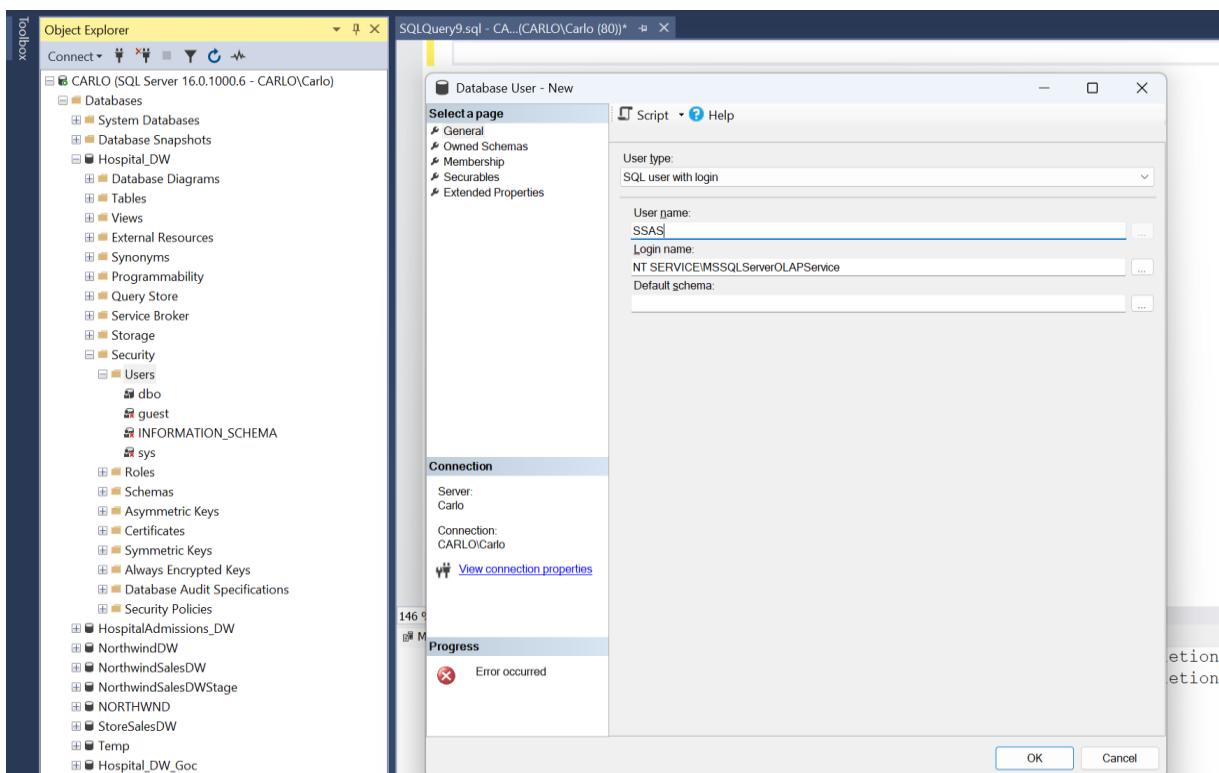
Click Run để thực thi



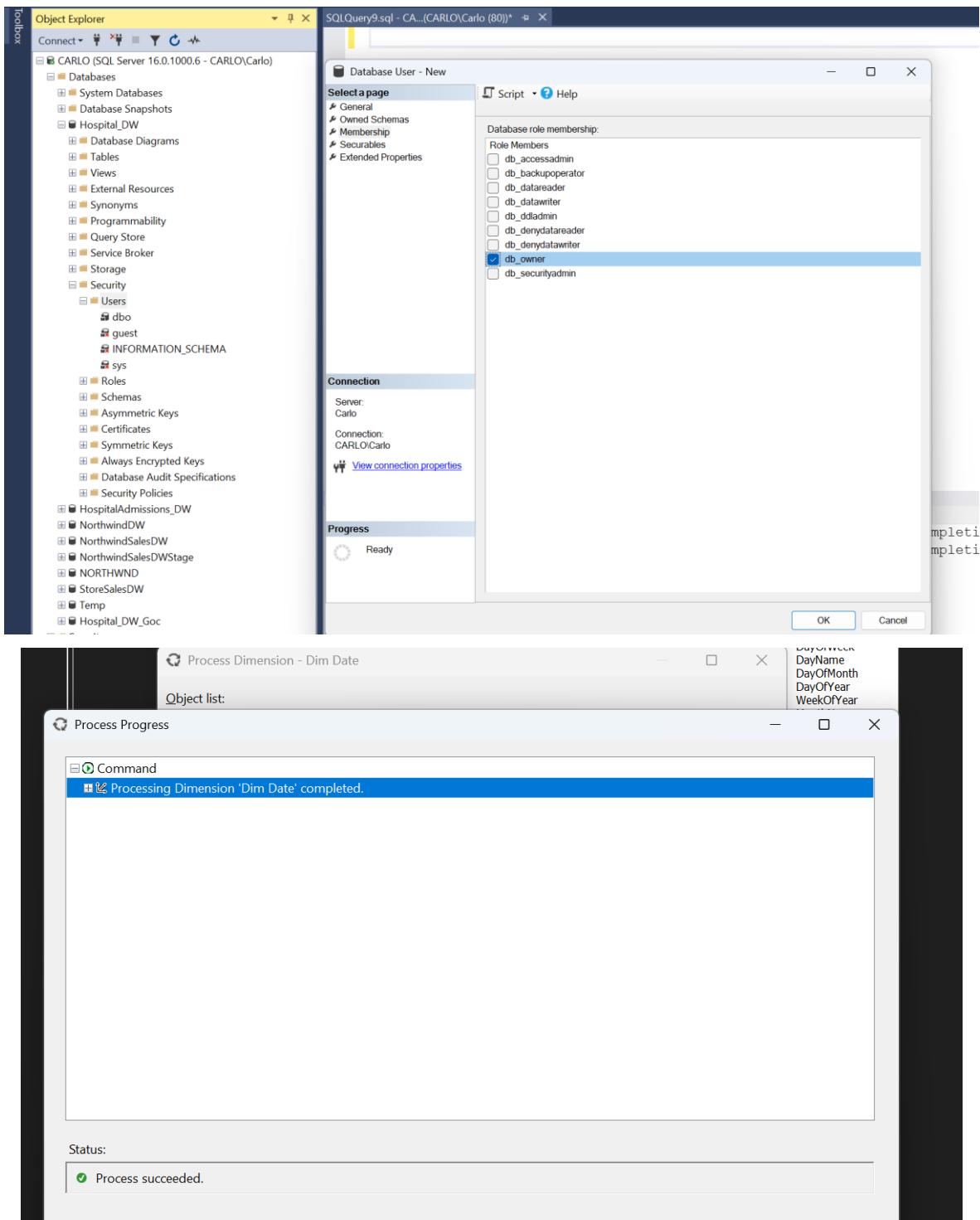
Lỗi login thất bại vì chưa thêm user. Vào SQL Server kết nối đến instance SQL Server. Mở rộng DB và thêm user → Security → User → New User...



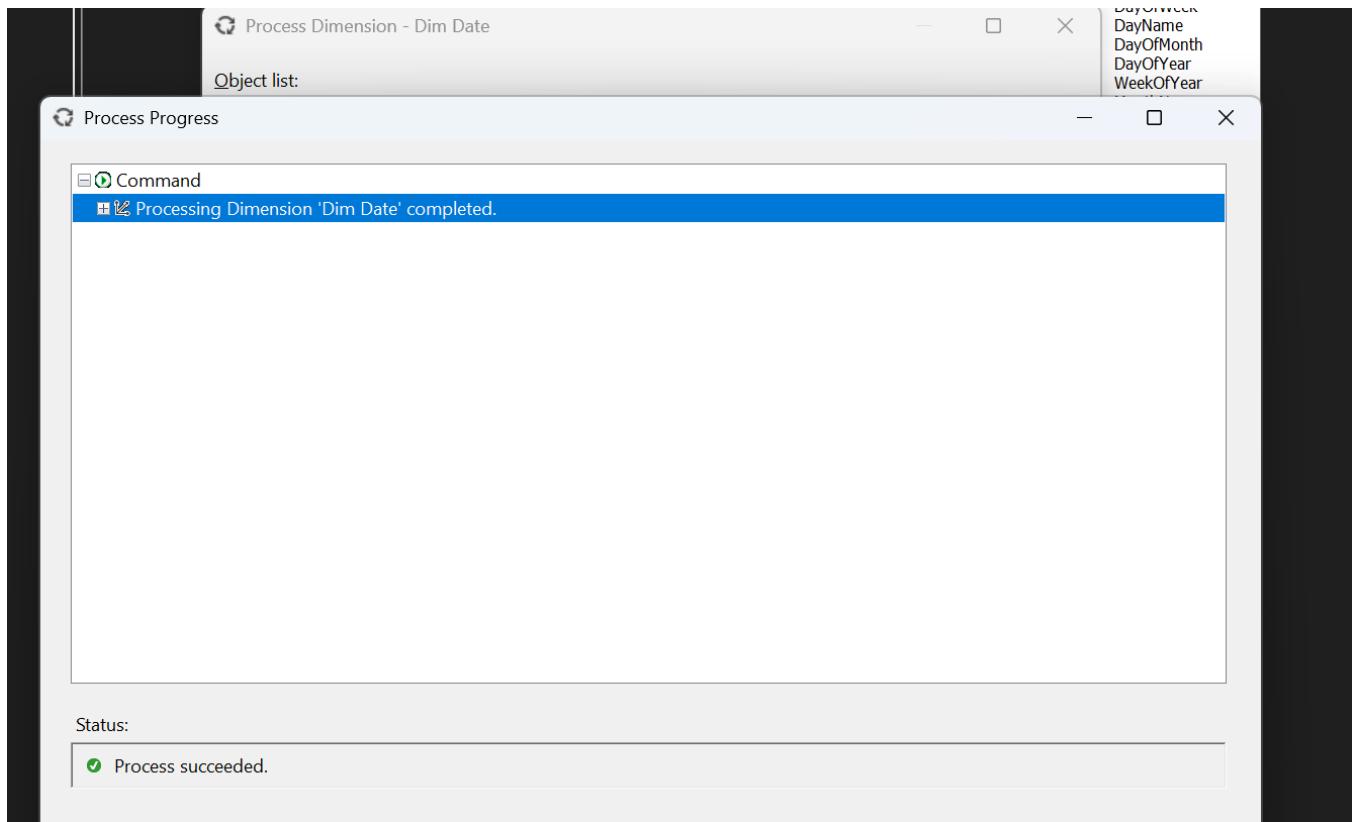
Nhập User name bất kỳ, Login name nhập đúng tên dịch vụ trong lỗi:



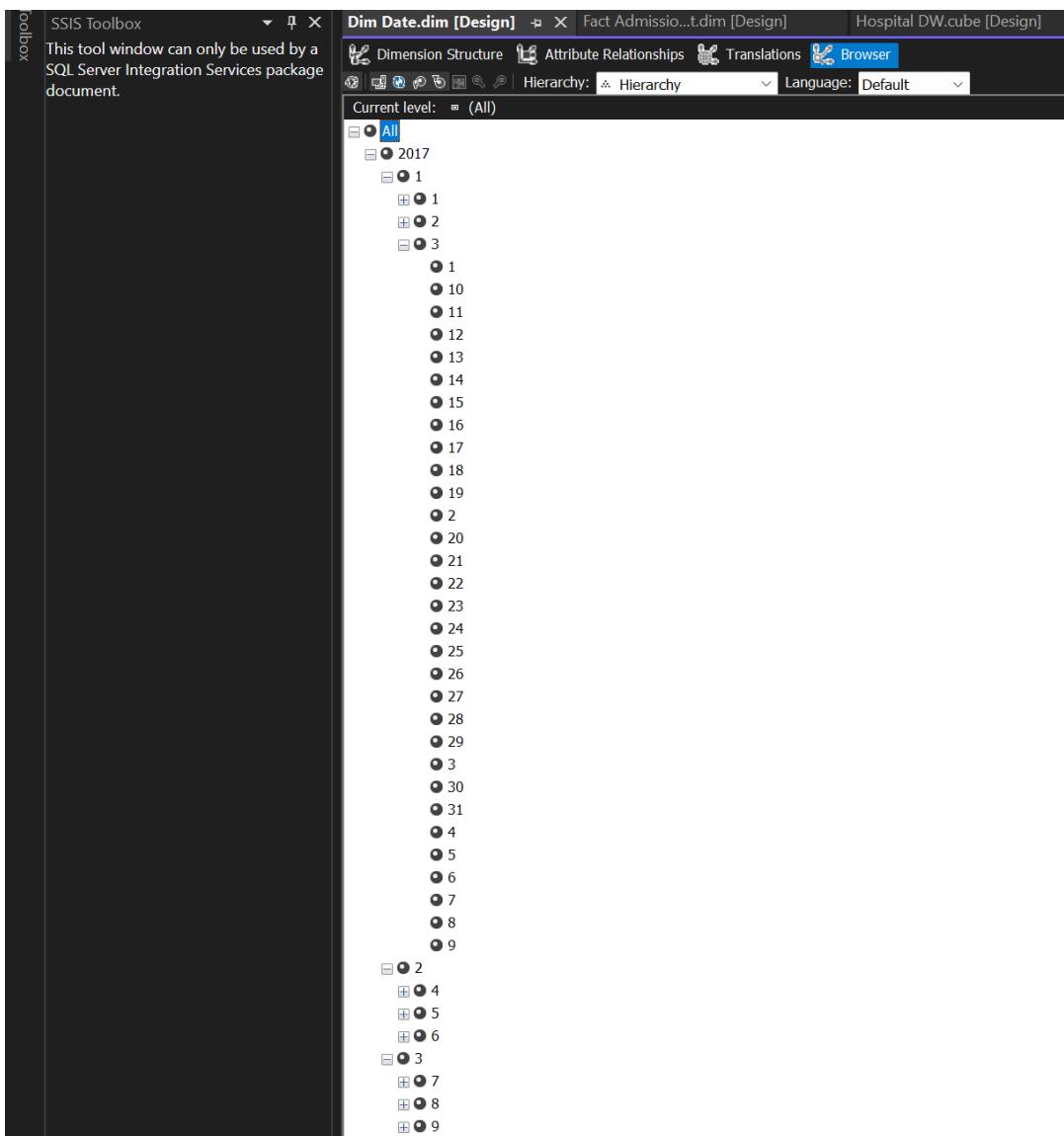
Vào mục Membership → tích chọn db_owner → OK



Click Run để thực thi. Nếu thành công, **Process Progress** dialog sẽ xuất hiện như sau:



Chuyển đến tab Browser ta sẽ thấy thời gian được phân cấp theo năm → quý → tháng → ngày



4.3.2. Phân cấp bệnh lý – DimDiagnosis

HeartFailureByType (Phân Loại Suy Tim): Phân loại suy tim dựa trên phân suất tổng máu (EF - Ejection Fraction), gồm: suy tim với phân suất tổng máu giảm (HFREF) và suy tim với phân suất tổng máu bảo tồn (HFNEF).

HeartShock (Tình Trạng Sốc Tim): Phân loại tình trạng sốc, tập trung vào sốc tim (Cardiogenic Shock) và các dạng sốc chung khác (Shock). Sốc tim là tình trạng nguy hiểm cần can thiệp y tế khẩn cấp.

HeartRate (Các Vấn Đề Nhịp Tim): Nhóm các chẩn đoán liên quan đến rối loạn nhịp tim, bao gồm bệnh tim bẩm sinh (Congenital), rung nhĩ (AF), nhịp nhanh thất (VT), và nhịp nhanh trên thất kịch phát (PSVT).

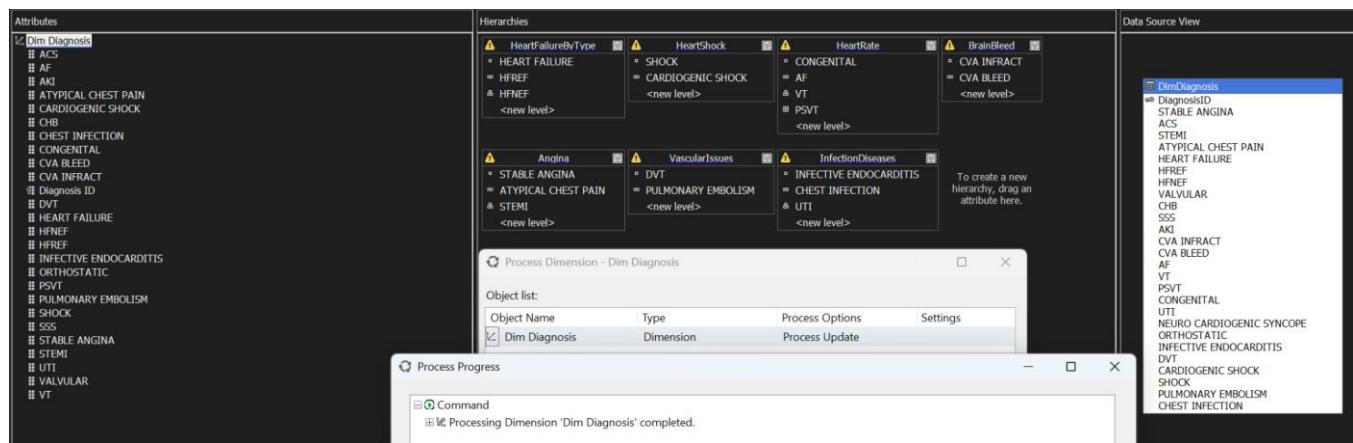
BrainBleed (Tai Biến Mạch Máu Não): Phân loại tai biến mạch máu não, gồm nhồi máu não (CVA Infarct) và xuất huyết não (CVA Bleed).

Angina (Đau Thắt Ngực): Phân loại tình trạng đau thắt ngực, như đau thắt ngực ổn định (Stable Angina), đau ngực không điển hình (Atypical Chest Pain), hội chứng mạch vành cấp khác có thể liên quan như nhồi máu cơ tim có ST chênh lên (STEMI)

VascularIssues (Các Vấn Đề Mạch Máu): Nhóm các bệnh lý liên quan đến mạch máu như huyết khối tĩnh mạch sâu (DVT) và thuyên tắc phổi (Pulmonary Embolism).

InfectionDiseases (Bệnh Lý Nhiễm Trùng): Nhóm các bệnh nhiễm trùng thường gặp như nhiễm trùng đường tiết niệu (UTI), viêm nội tâm mạc nhiễm khuẩn (Infective Endocarditis) và nhiễm trùng ngực/phổi (Chest Infection).

Cách thực hiện tương tự nhiên phân cấp thời gian. Ta sẽ được bảng sau:



4.3.3. Phân cấp liên quan tới bệnh đồng mắc – DimComorbidity

HasAnaemia: Mô tả tình trạng bệnh nhân có hoặc không có thiếu máu. Nhóm này bao gồm các mức độ thiếu máu khác nhau được ghi nhận, từ thiếu máu tổng quát (**ANAEMIA**) đến thiếu máu nghiêm trọng (**SEVERE ANAEMIA**), giúp phân loại bệnh nhân theo mức độ thiếu máu hiện tại.

HasDM (Tình trạng Đái tháo đường): Tập hợp các thuộc tính liên quan đến bệnh đái tháo đường (**DM**) và các biến chứng hoặc bệnh lý thường đi kèm như bệnh thận mãn tính (**CKD**), tăng huyết áp (**HTN**), và bệnh mạch vành (**CAD**).

HeartDisease (Bệnh lý Tim Mạch): Nhóm các thuộc tính liên quan đến các vấn đề về tim mạch, bao gồm bệnh mạch vành (**CAD**), chỉ số men tim tăng (**RAISED CARDIAC ENZYMES**) – một dấu hiệu của tổn thương cơ tim, và tiền sử bệnh cơ tim (**PRIOR CMP**)

Cách thực hiện tương tự nhiên phân cấp thời gian. Ta sẽ được bảng sau:

Đối với những Dimension khác chưa được cấu hình chi tiết (ví dụ, chưa tạo các phân cấp phức tạp), hoặc đối với các thuộc tính còn lại trong các Dimension đã có mà không cần phân cấp, chúng ta sẽ thực hiện thao tác cơ bản là kéo các cột (fields) tương ứng từ Data Source View vào mục Attributes của Dimension đó trong SSAS.

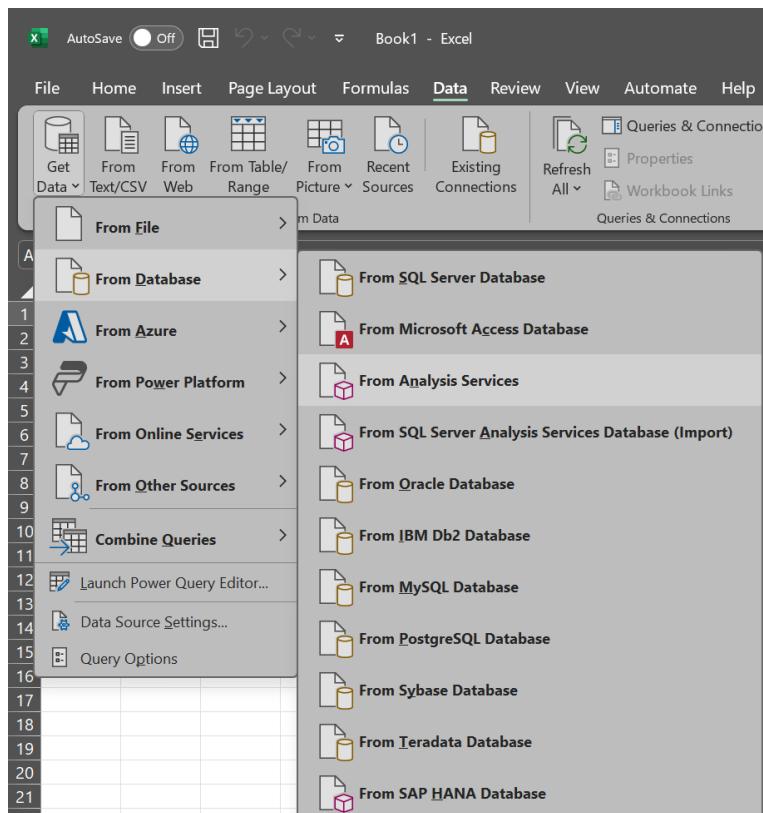
Sau khi hoàn thành các bước trên thì vào Hospital DW.cube bấm nút Progress để tạo Cube

4.4. Xây dựng & phân tích qua Excel Pivot Table

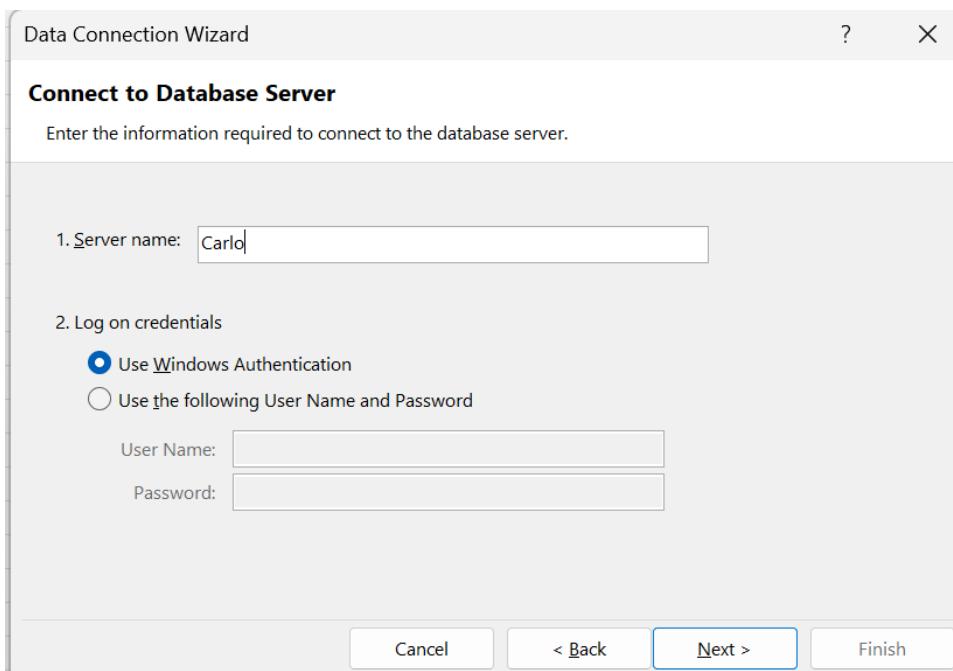
4.4.1. Kết nối SSAS với Excel

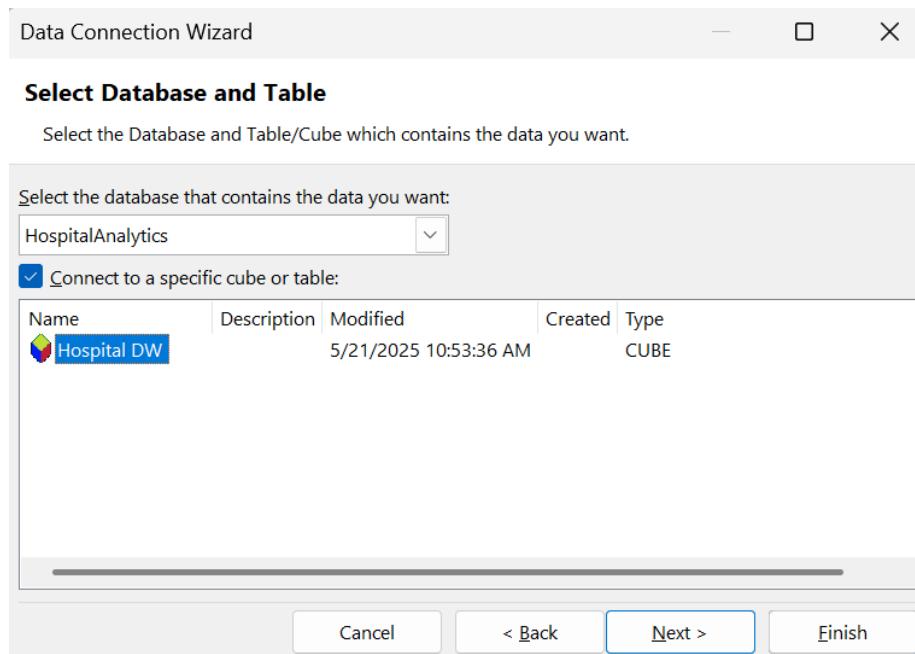
Mở Excel, tạo một new blank workbook.

Trên thanh Ribbon, chọn Data → Get Data → From Database → From Analysis Services

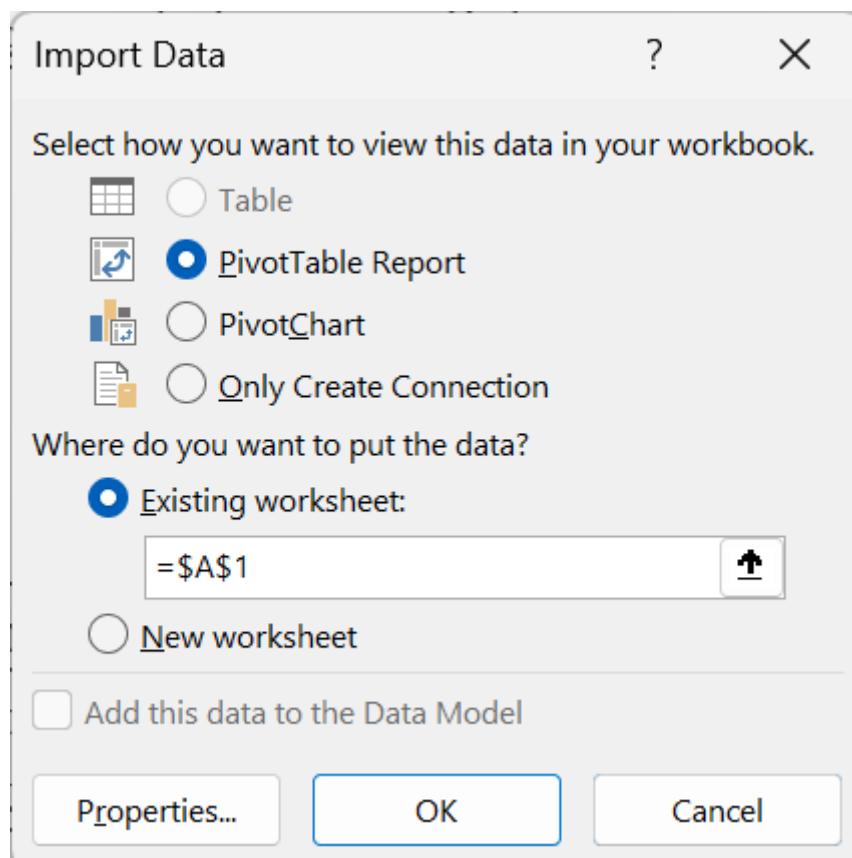


Chọn Server name phù hợp → Next → chọn database và cube → click Finish chứ không Next



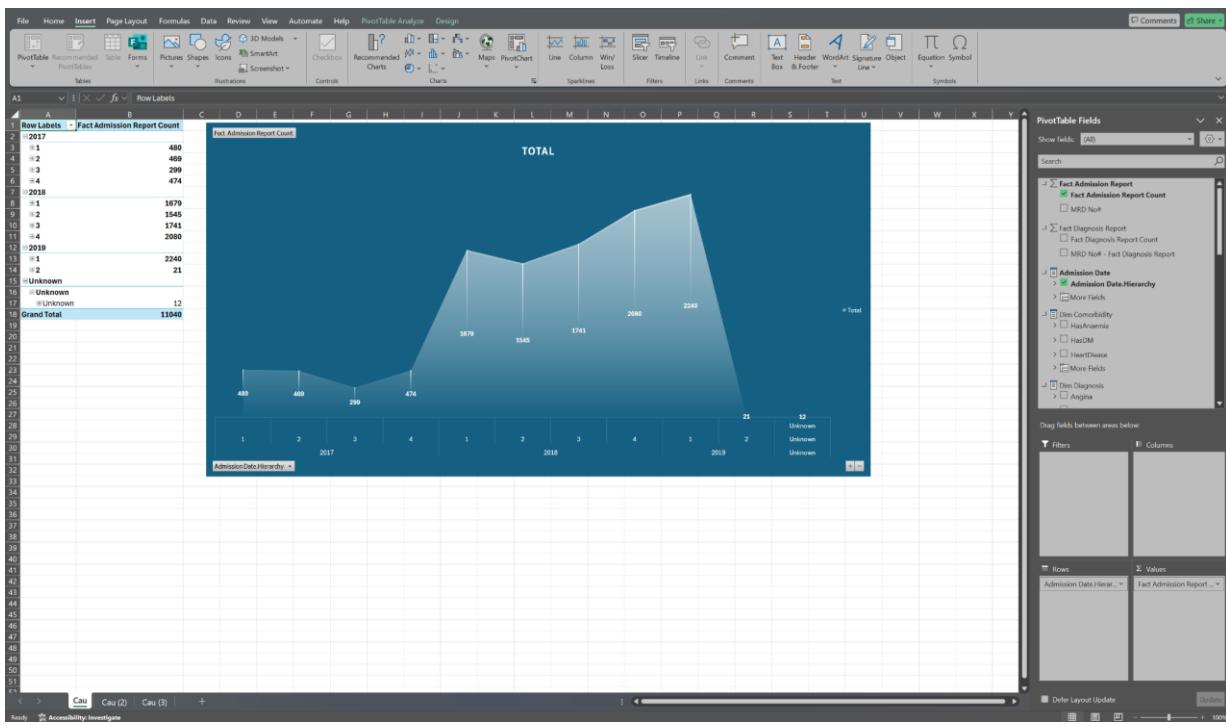


Import Data dialog xuất hiện, chọn **PivotTable Report** và **Existing Worksheet** và click **OK**.



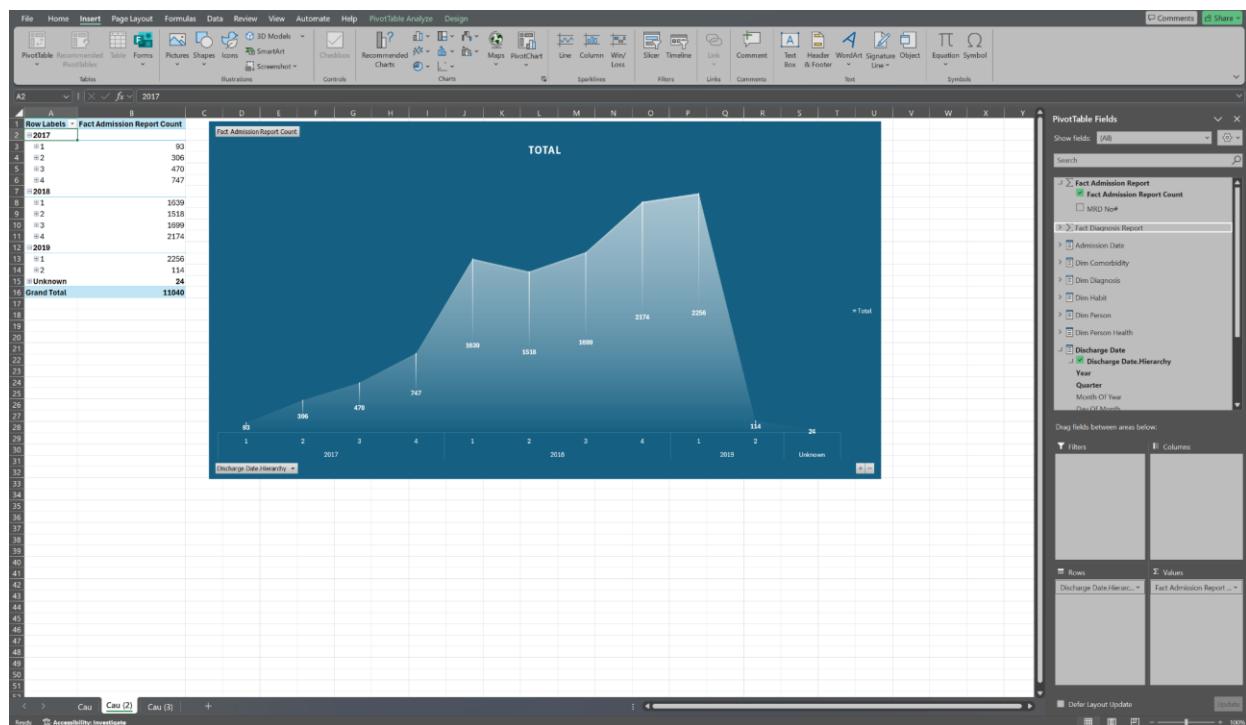
4.4.2. Trả lời các câu hỏi nghiệp vụ từ các bảng Pivot

Câu 1: Số lượng bệnh nhân nhập viện



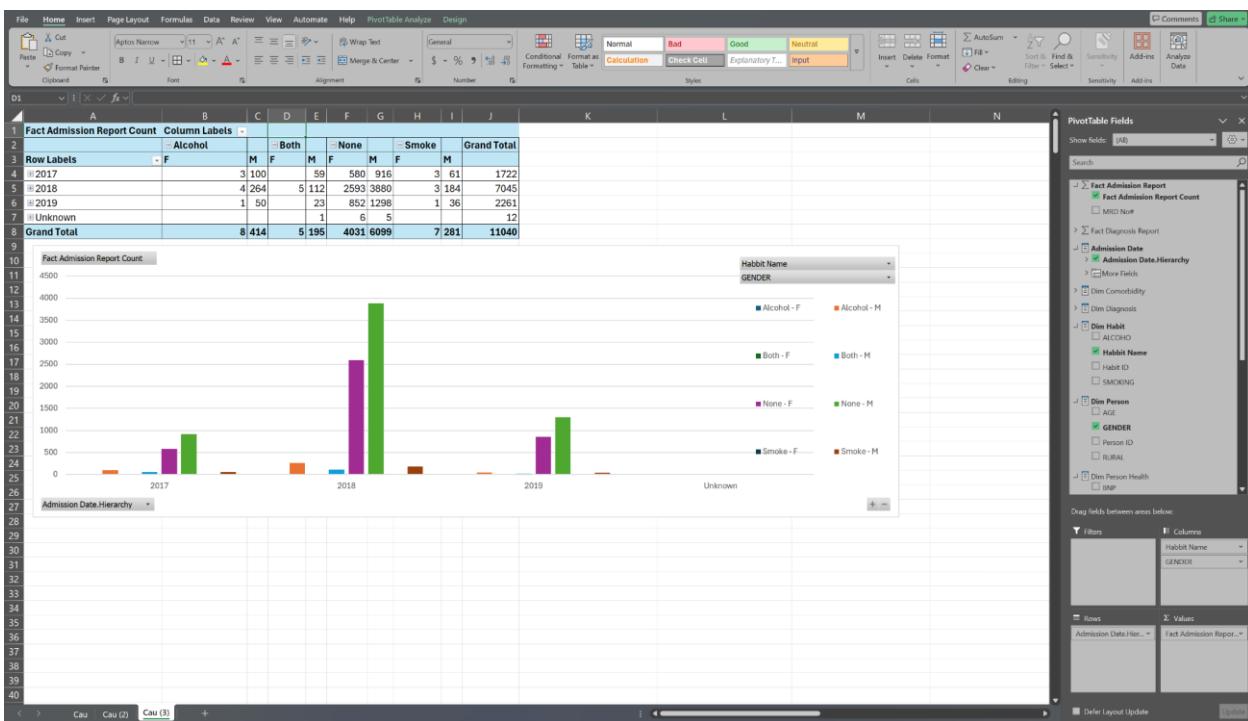
Ta thấy số lượng bệnh nhân nhập viện nhiều nhất là vào quý 1 năm 2019

Câu 2: Số lượng bệnh nhân xuất viện



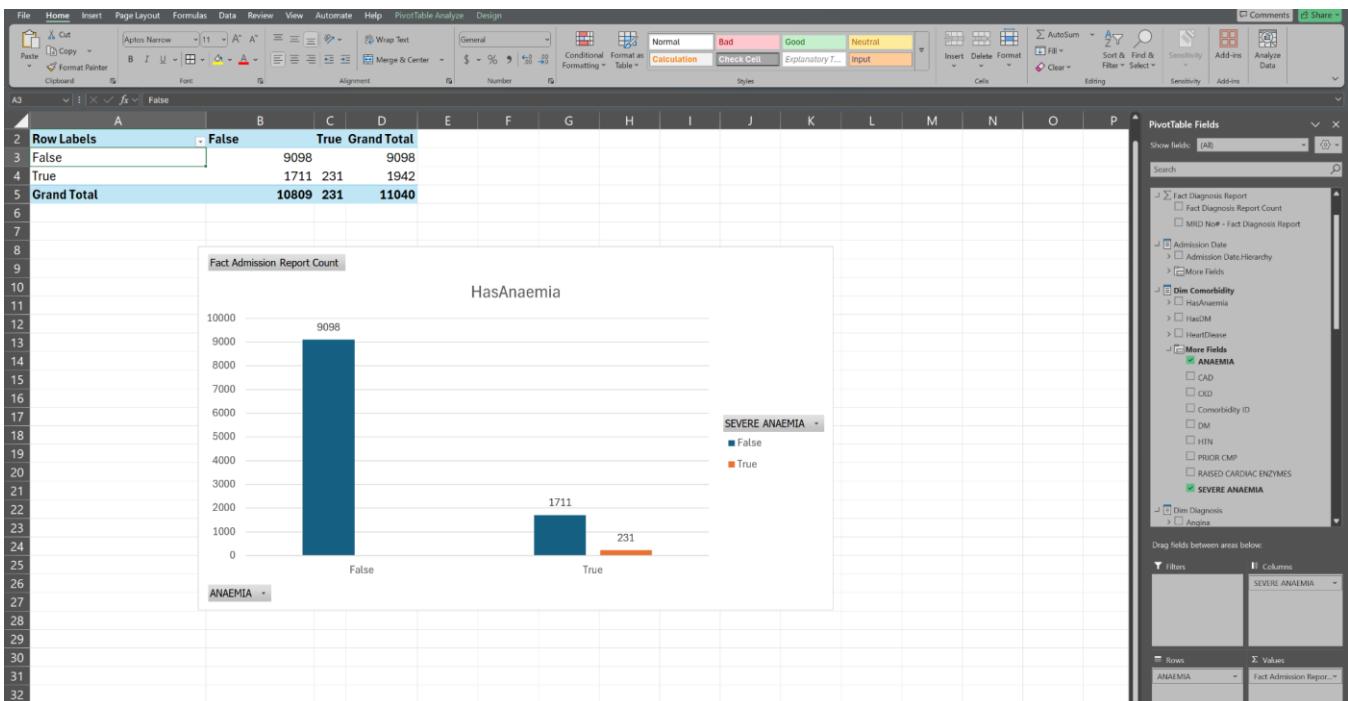
Số lượng bệnh nhân xuất viện cao nhất cũng là quý 1 năm 2019

Câu 3: Số lượng bệnh nhân với mỗi loại thói quen



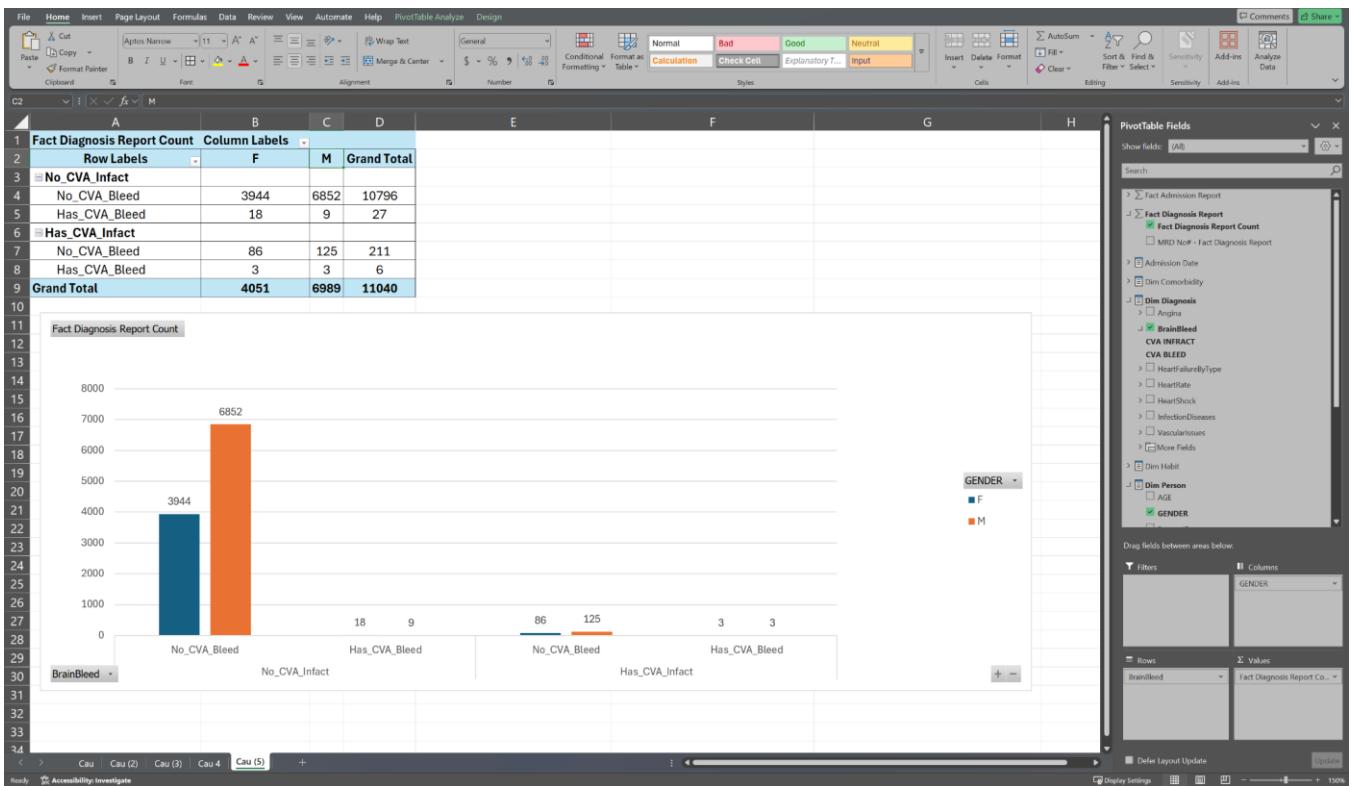
Ta thấy đa số bệnh nhân không hút thuốc và không rượu bia nhiều hơn hẳn. Trong đó số lượng phụ nữ hút thuốc và uống rượu ít hơn hẳn so với đàn ông.

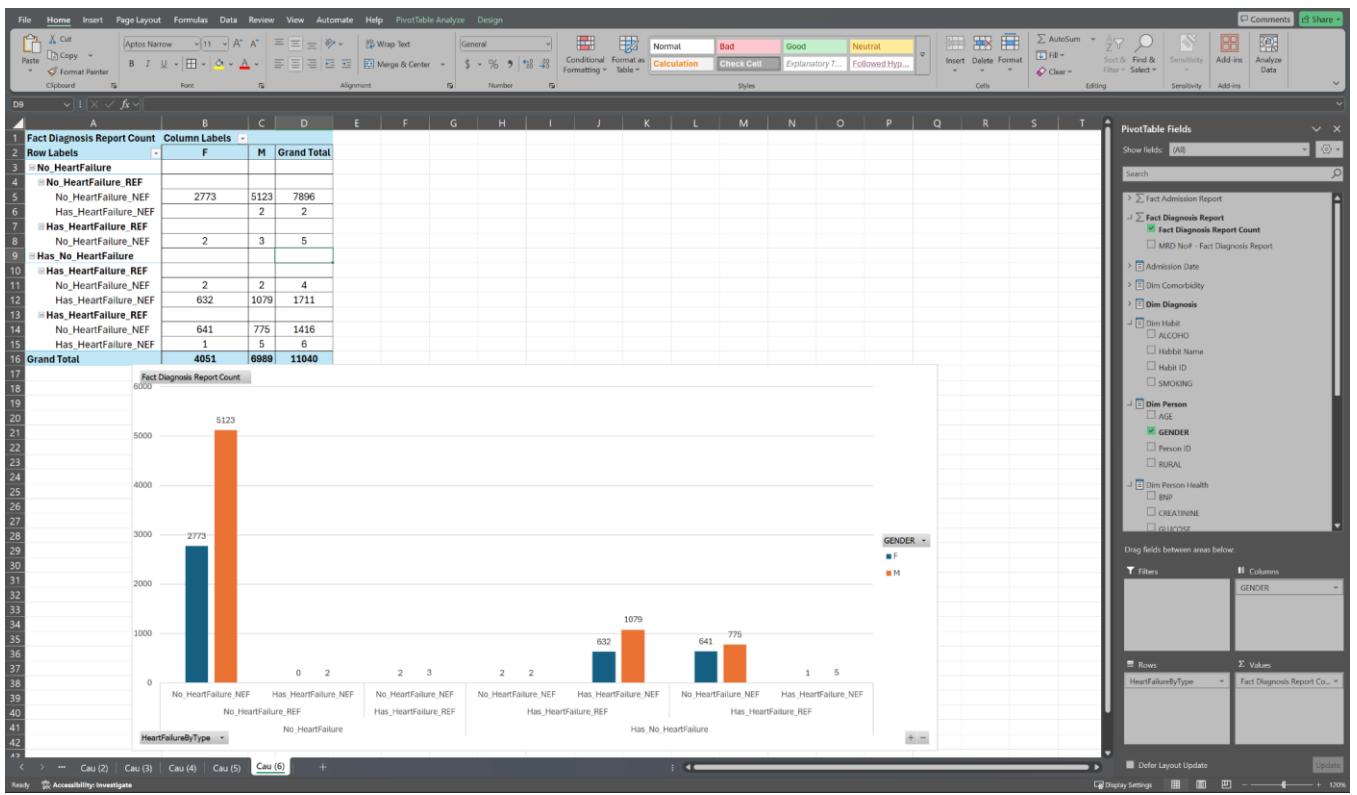
Câu 4: Số lượng bệnh nhân bị bệnh thiếu máu ở các tình trạng khác nhau



Đa phần bệnh nhân không bị thiếu máu (khoảng 85%) và chỉ có khoảng 10% bệnh nhân là bị thiếu máu nghiêm trọng.

Câu 5: Nguyên nhân nào khiến bệnh nhân bị đột quỵ nhiều hơn?

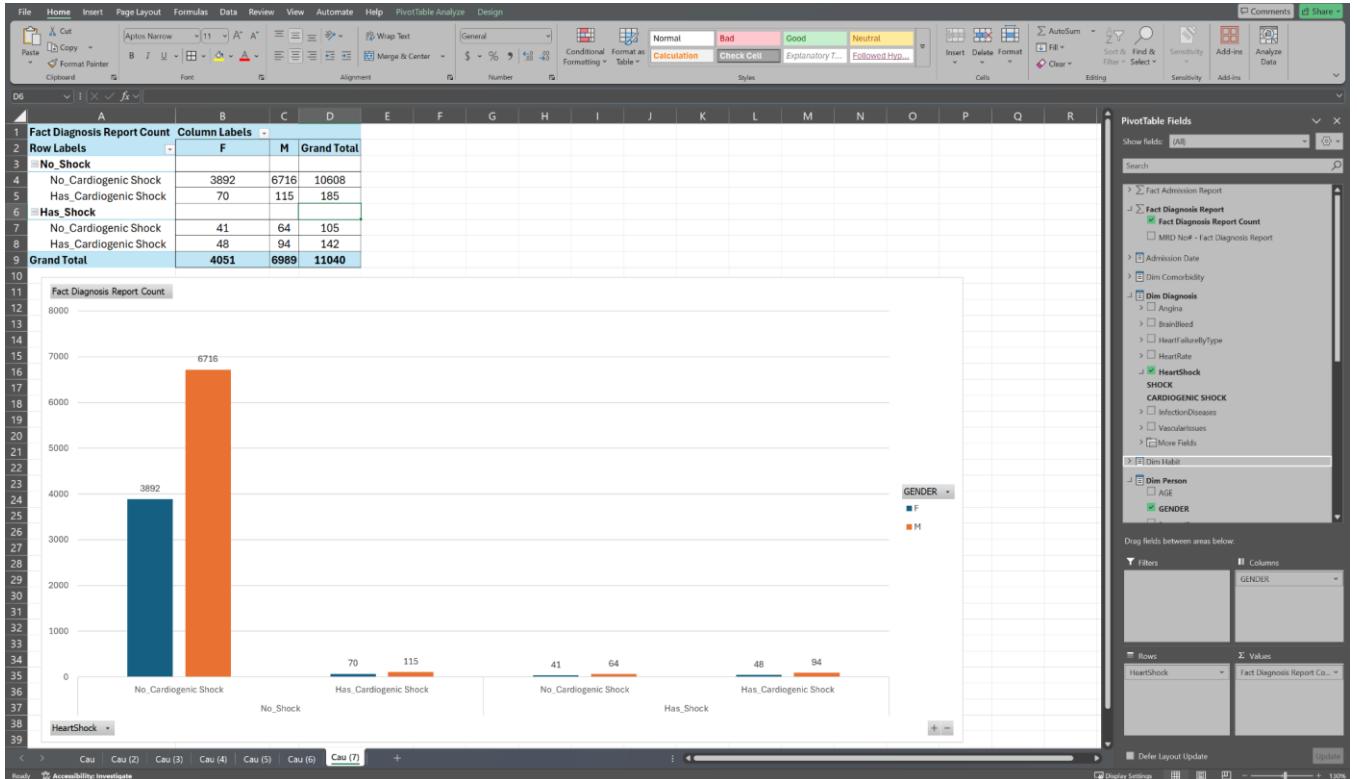




Phần lớn bệnh nhân không mắc hội chứng suy tim, với tổng cộng 7.896 ca không bị (No_HeartFailure_REF/NEF), so với 3.137 ca có bị (Has_HeartFailure_REF/NEF).

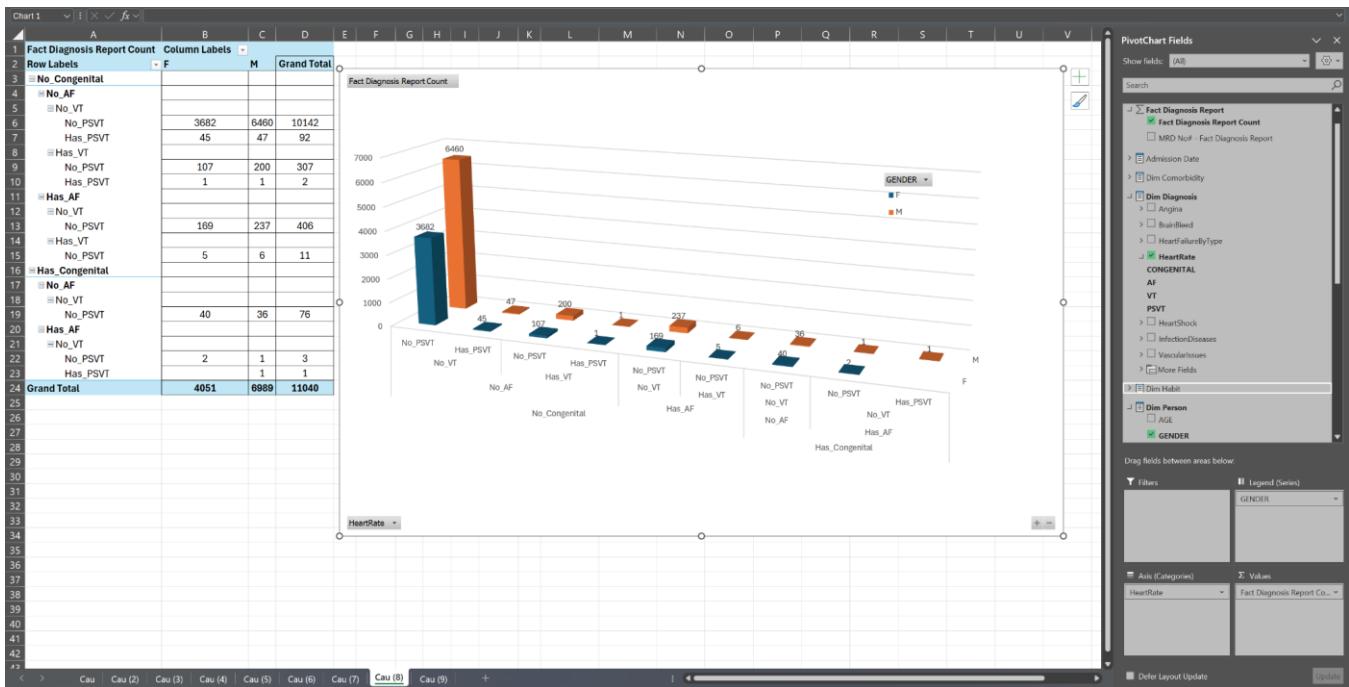
Trong số những người bị suy tim, **nam giới chiếm đa số**. Điều này cho thấy **giới tính có thể là một yếu tố nguy cơ đáng chú ý trong hội chứng suy tim**.

Câu 7: Số lượng bệnh nhân bị sốc tim



Ta thấy khoảng 90% bệnh nhân không bị sốc tim. Có **247** **bệnh nhân bị sốc tim**, trong đó **142** trường hợp là **sốc tim do nguyên nhân tim mạch (Cardiogenic Shock)**. Nam giới chiếm tỷ lệ cao hơn rõ rệt trong cả hai nhóm.

Câu 8: Số lượng bệnh nhân mắc các vấn đề về nhịp tim bất thường



Phần lớn bệnh nhân **không mắc bệnh tim bẩm sinh (No_Congenital)** với tổng số **10.142** ca, chiếm ~92% tổng bệnh nhân. Trong khi đó, chỉ có ~8% **bệnh nhân có tiền sử tim bẩm sinh (Has_Congenital)**.

Đáng chú ý, trong nhóm **Has_Congenital**, gần như không có bệnh nhân nào được ghi nhận mắc đồng thời các rối loạn nhịp tim khác như AF, VT hoặc PSVT. Điều này gợi ý rằng:

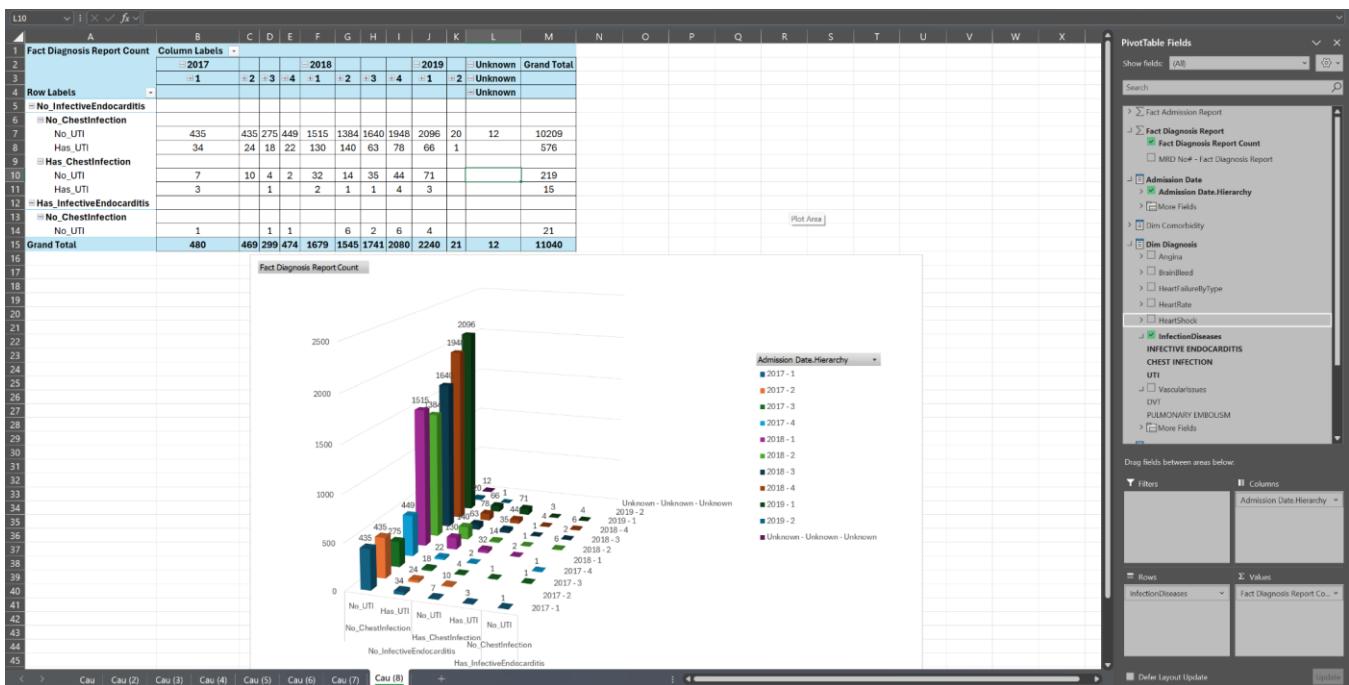
- + **Tim bẩm sinh có thể là một nhóm riêng biệt**, không đi kèm các rối loạn nhịp tim phổ biến.
- + Có thể do **bệnh nhân tim bẩm sinh đã được tầm soát và điều trị sớm** nên ít gặp biến chứng rối loạn nhịp.

Ngược lại, các rối loạn nhịp tim như **AF (Rung nhĩ)**, **VT** và **PSVT** chủ yếu xảy ra ở nhóm không có bệnh tim bẩm sinh, đặc biệt là nam giới — được thể hiện rõ qua các cột màu cam trên biểu đồ.

→ Kết quả này giúp bác sĩ xác định nhóm nguy cơ rối loạn nhịp tim rõ ràng hơn, hỗ trợ xây dựng phác đồ điều trị riêng biệt cho:

- Nhóm **bệnh nhân tim bẩm sinh**, tập trung theo dõi biến chứng phát triển khác hơn là các rối loạn nhịp tim.
- Nhóm **bệnh nhân không có tim bẩm sinh**, cần sàng lọc và giám sát chặt nhịp tim (AF, VT, PSVT).

Câu 9: Số lượng bệnh nhân mắc các bệnh nhiễm trùng theo thời gian?



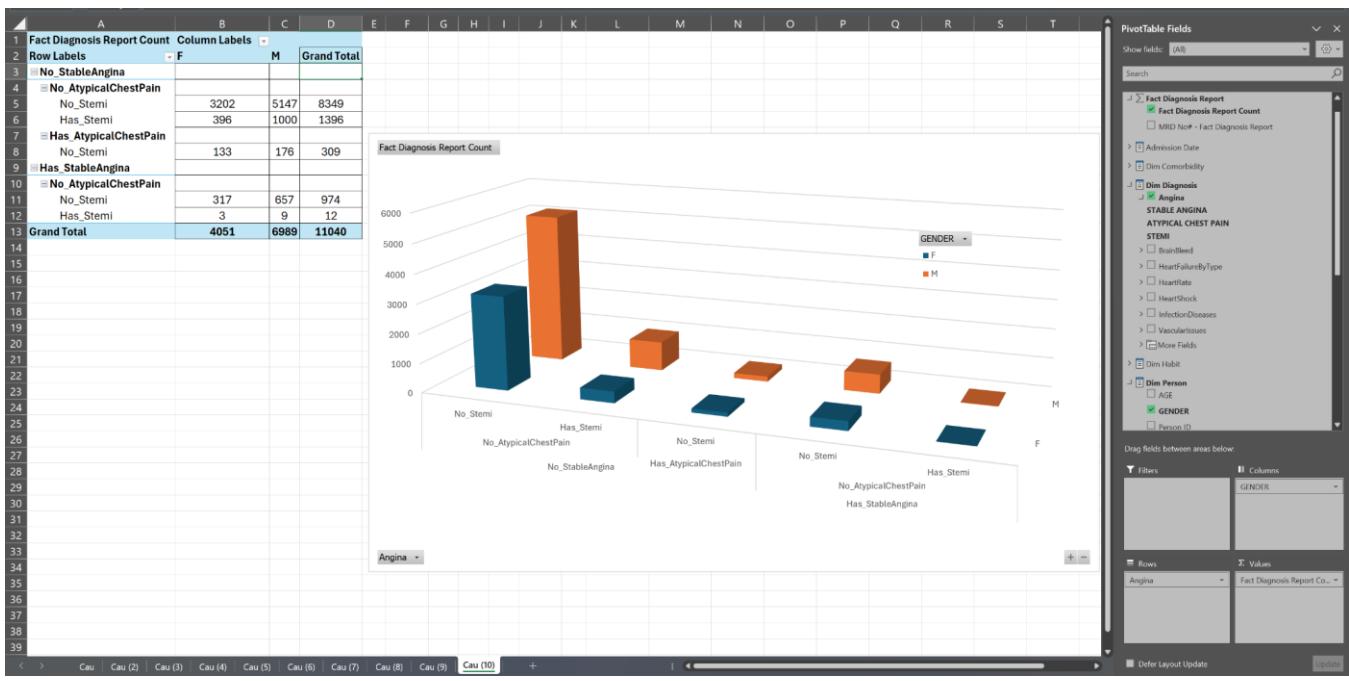
Từ năm 2017 đến 2019, số lượng bệnh nhân mắc **các bệnh nhiễm trùng (Infectious Diseases)** như Nhiễm trùng đường tiết niệu (UTI), nhiễm trùng phổi (Chest Infection), và viêm nội tâm mạc (Infective Endocarditis) **tăng dần theo thời gian**.

Đặc biệt, **quý 1 năm 2019** ghi nhận **số ca cao nhất**. Trong đó, **UTI là loại nhiễm trùng phổ biến nhất** với **576 trường hợp**.

Đáng chú ý, **hầu hết bệnh nhân không mắc viêm nội tâm mạc (No_InfectiveEndocarditis)**. Trong nhóm nhỏ mắc bệnh này (**Has_InfectiveEndocarditis**), **không có trường hợp nào bị đồng thời UTI hoặc Chest Infection** — cho thấy khả năng các bệnh nhiễm trùng này có xu hướng loại trừ nhau trong chẩn đoán. Đồng thời, chỉ khoảng **4% trong nhóm không bị viêm nội tâm mạc** có nhiễm các bệnh khác như UTI hoặc Chest Infection.

→ **Điều này có thể giúp tiết kiệm chi phí điều trị cho nhóm bệnh nhân này**, vì nguy cơ lây nhiễm chéo thấp hơn.

Câu 10: Số lượng bệnh nhân bị đau thắt ngực và nhồi máu cơ tim cấp tính



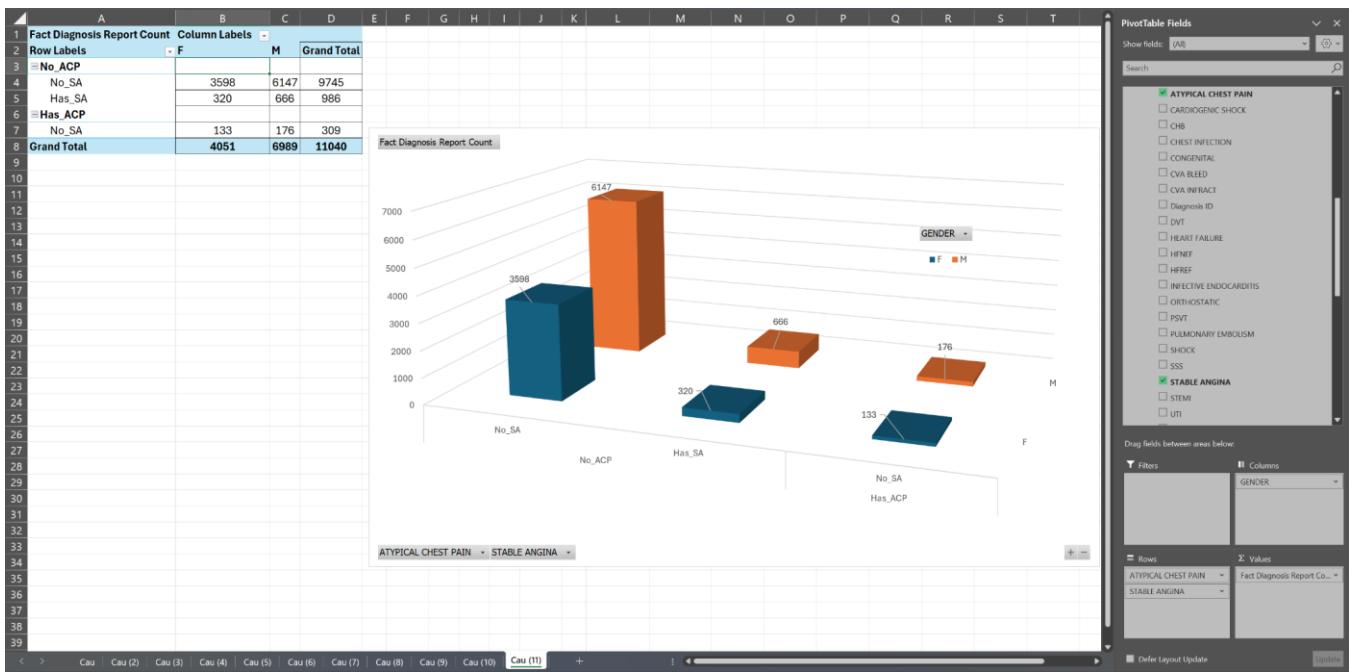
Có 1.408 bệnh nhân bị nhồi máu cơ tim cấp tính (STEMI) và 986 bệnh nhân bị đau thắt ngực ổn định (Stable Angina).

Chỉ có 12 bệnh nhân mắc đồng thời cả hai, cho thấy **hai bệnh này hiếm khi xảy ra cùng lúc.**

Ngoài ra, có 309 bệnh nhân bị đau ngực không điển hình (Atypical Chest Pain) nhưng không mắc STEMI hay Stable Angina, đây là nhóm cần theo dõi vì triệu chứng có thể không rõ ràng nhưng vẫn có nguy cơ tim mạch tiềm ẩn.

→ Sự phân tách rõ giữa STEMI và Stable Angina hỗ trợ bác sĩ phân loại lâm sàng rõ ràng, còn nhóm Atypical Chest Pain lại yêu cầu theo dõi kỹ hơn vì **dễ bị bỏ sót hoặc chẩn đoán nhầm.**

Câu 11: Số lượng bệnh nhân mắc các bệnh đau ngực theo giới tính



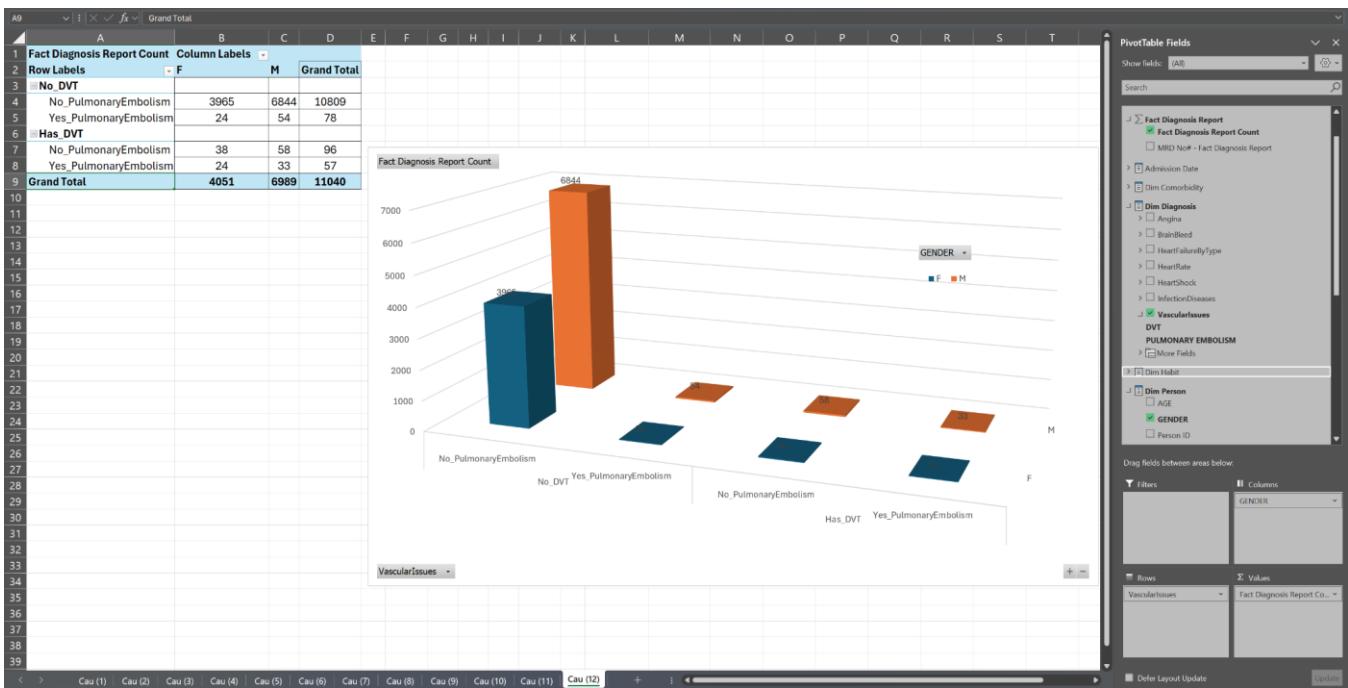
Có **986** bệnh nhân bị **đau thắt ngực ổn định (Stable Angina)** mà không kèm theo đau ngực không điển hình, trong đó **nam giới chiếm gần gấp đôi nữ giới**.

Đồng thời, có **309** **bệnh nhân bị đau ngực không điển hình (Atypical Chest Pain)** mà không bị đau thắt ngực ổn định, cho thấy đây là một nhóm riêng biệt đáng chú ý.

Nhìn chung, **nam giới có tỷ lệ mắc các triệu chứng đau ngực cao hơn nữ giới rõ rệt**, đặc biệt là ở dạng **đau thắt ngực ổn định**. Điều này có thể liên quan đến **các yếu tố nguy cơ phổ biến ở nam giới** như hút thuốc, rượu, stress hoặc ít kiểm tra y tế định kỳ.

→ Sự chênh lệch giới tính trong các triệu chứng đau ngực là rõ rệt và có ý nghĩa lâm sàng. Việc **tầm soát và tuyên truyền thay đổi lối sống cho nam giới** có thể giúp giảm tỷ lệ bệnh tim mạch và cải thiện sức khỏe cộng đồng.

Câu 12: Số lượng bệnh nhân bị các vấn đề về mạch máu như huyết khối tĩnh mạch sâu hay thuyên tắc phổi



Có 153 bệnh nhân bị huyết khối tĩnh mạch sâu (DVT) và 135 bệnh nhân bị thuỷt tắc phổi (Pulmonary Embolism). Trong đó, 57 người mắc cả hai bệnh lý, chiếm 42% trong nhóm PE.

Điều này khẳng định mối liên hệ chặt chẽ giữa DVT và PE, và cho thấy DVT là yếu tố nguy cơ quan trọng cần theo dõi sớm để phòng ngừa thuỷt tắc phổi nguy hiểm.

→ Các bệnh lý mạch máu như DVT và PE thường đi kèm nhau. Việc phát hiện và điều trị DVT sớm có thể giúp ngăn ngừa PE, từ đó giảm tỷ lệ biến chứng nặng và tử vong ở bệnh nhân tim mạch

Kết luận từ các câu hỏi:

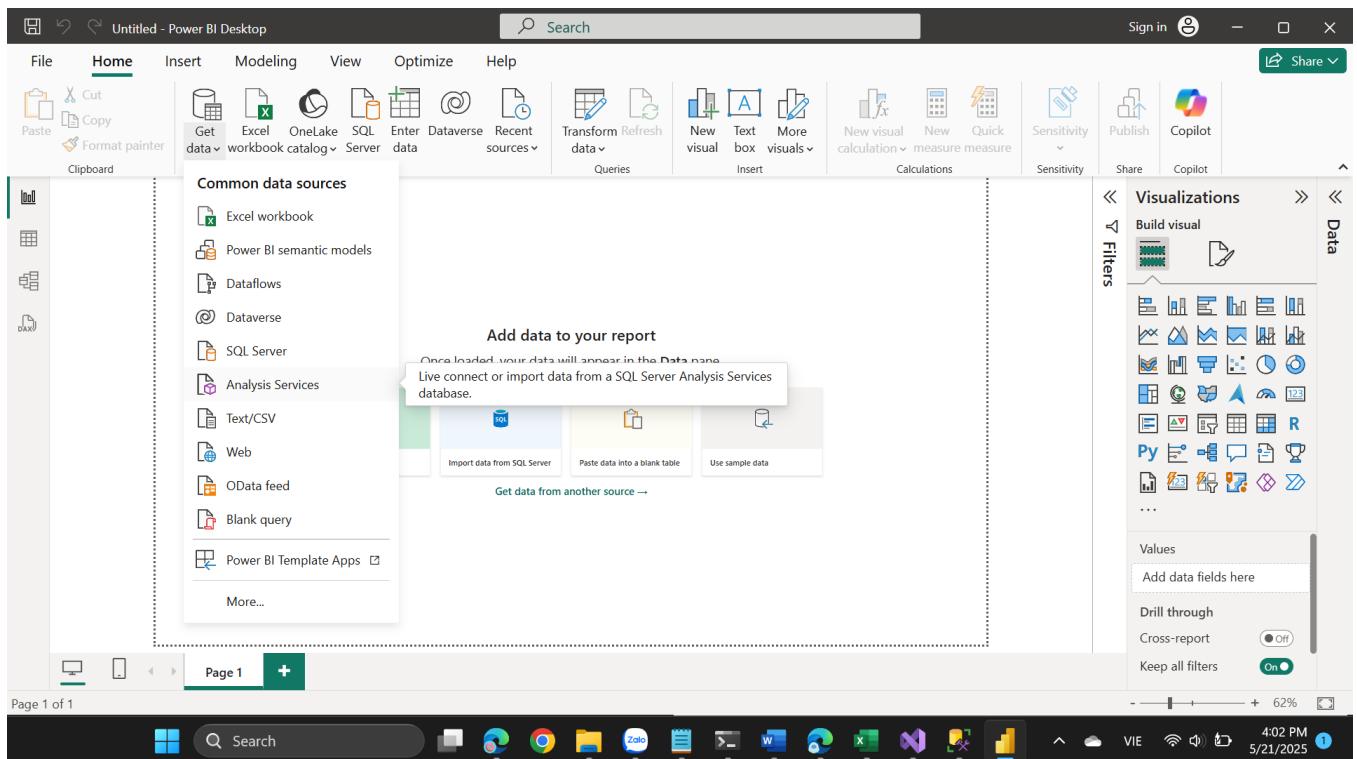
STT	Câu hỏi phân tích	Kết quả phân tích	Kết luận hỗ trợ ra quyết định
1	Bệnh nhân nhập viện mỗi ngày?	Nhiều nhất vào quý 1 năm 2019	Tăng cường nguồn lực bệnh viện trong quý có số ca tăng
2	Bệnh nhân xuất viện mỗi ngày?	Cũng cao nhất vào quý 1 năm 2019	Cân đối thời gian lưu viện
3	Thói quen bệnh nhân (hút thuốc, uống rượu)?	Đa số không hút thuốc, uống rượu	Tập trung tư vấn thói quen sống cho nhóm nguy cơ
4	Bệnh thiếu máu?	~85% không thiếu máu	Chỉ định xét nghiệm thêm với nhóm có nguy cơ
5	Nguyên nhân đột quỵ?	Đa số do đông máu	Kê thuốc chống đông máu sớm

STT	Câu hỏi phân tích	Kết quả phân tích	Kết luận hỗ trợ ra quyết định
6	Bệnh nhân suy tim?	~30% có hội chứng suy tim	Theo dõi sát triệu chứng suy tim nhẹ
7	Sốc tim theo giới tính?	Nam bị sốc tim nhiều hơn nữ	Ưu tiên điều trị nhóm nam giới nguy cơ cao
8	Nhip tim bất thường?	Chủ yếu không bị	Xây dựng phác đồ điều trị phù hợp
9	Bệnh nhiễm trùng theo thời gian?	Tỷ lệ rất thấp nếu không viêm nội tâm mạc	Giảm xét nghiệm tốn kém cho nhóm không nguy cơ
10	Đau thắt ngực & STEMI?	Không quá liên quan nhau	Cần phân biệt rõ chẩn đoán
11	Bệnh đau ngực theo giới?	Nam giới bị nhiều hơn nữ	Tăng cường tầm soát đau ngực cho nam
12	Vấn đề về mạch máu	DVT là yếu tố nguy cơ dẫn đến PE	Phát hiện và điều trị DVT sớm có thể giúp ngăn ngừa PE

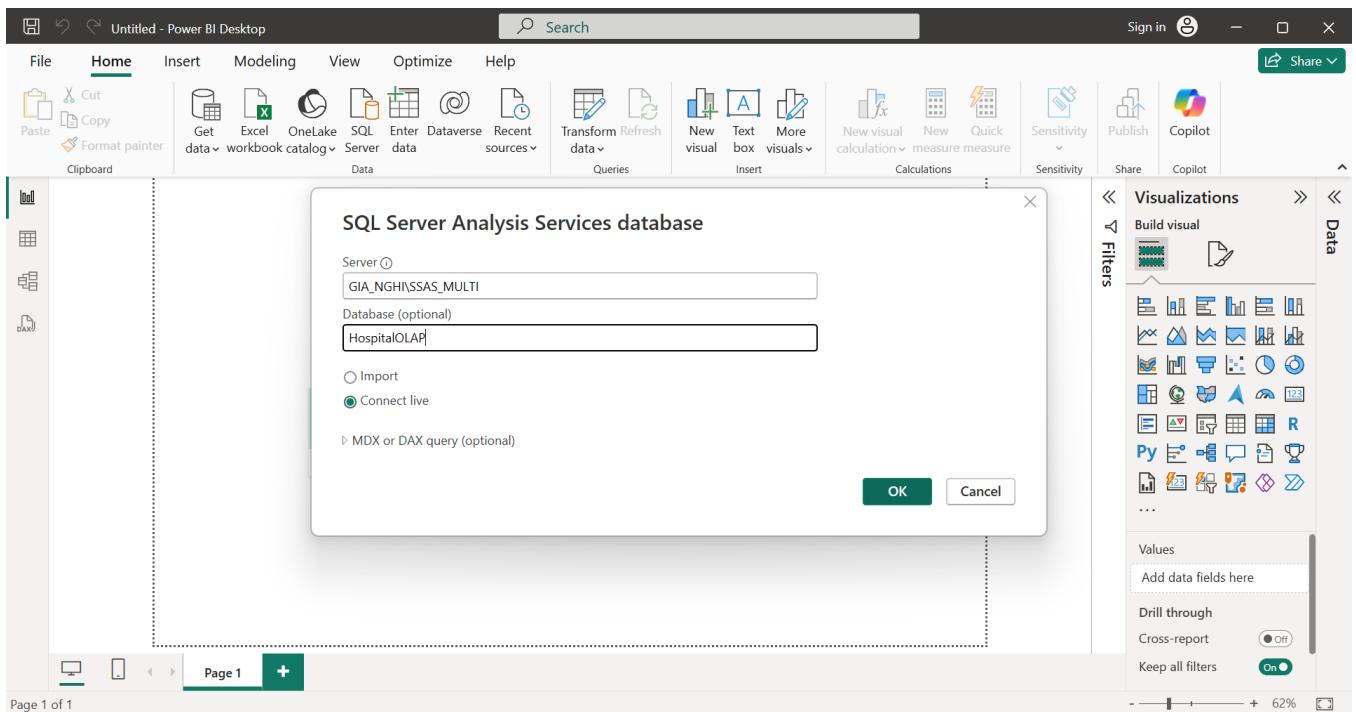
4.5. Xây dựng báo cáo qua Power BI

Để có thể tiến hành các câu lệnh, ta sẽ tải tập dữ liệu vào Power BI bằng cách thực hiện các bước như sau:

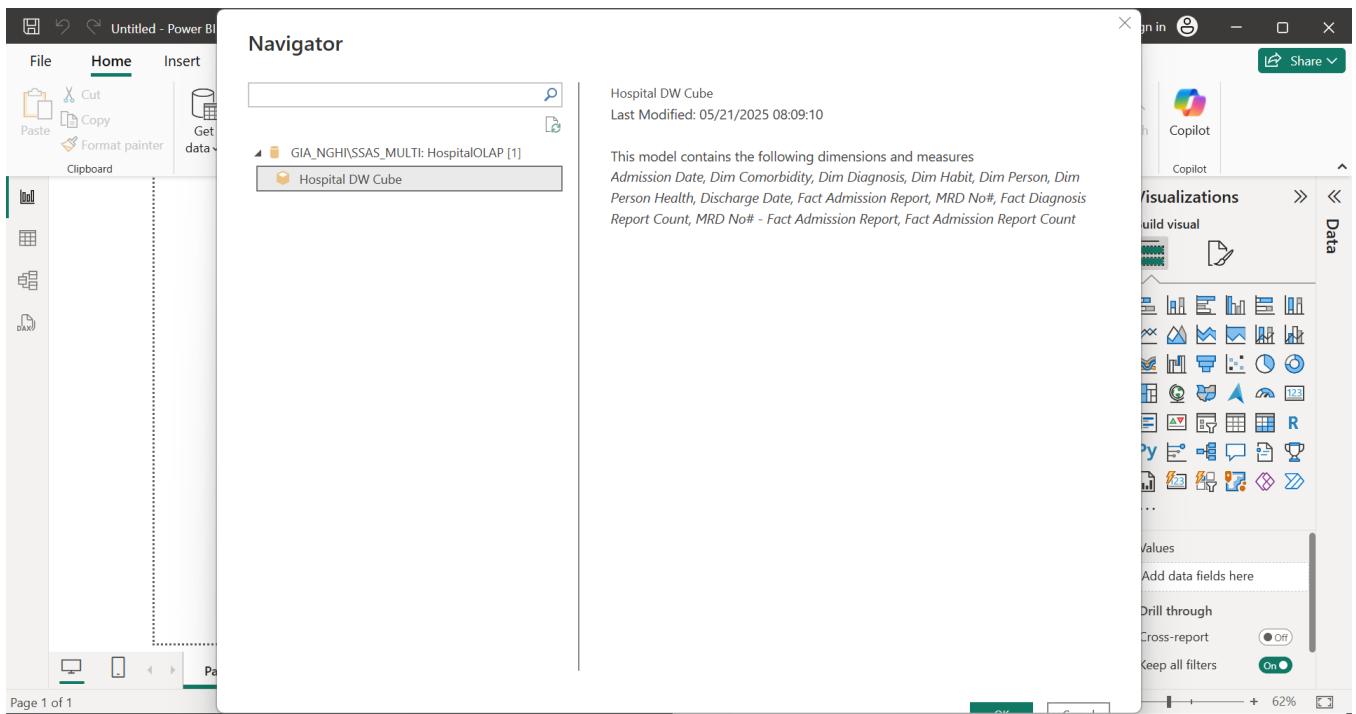
1. Chọn kết nối tới SQL Server Analysis Services database



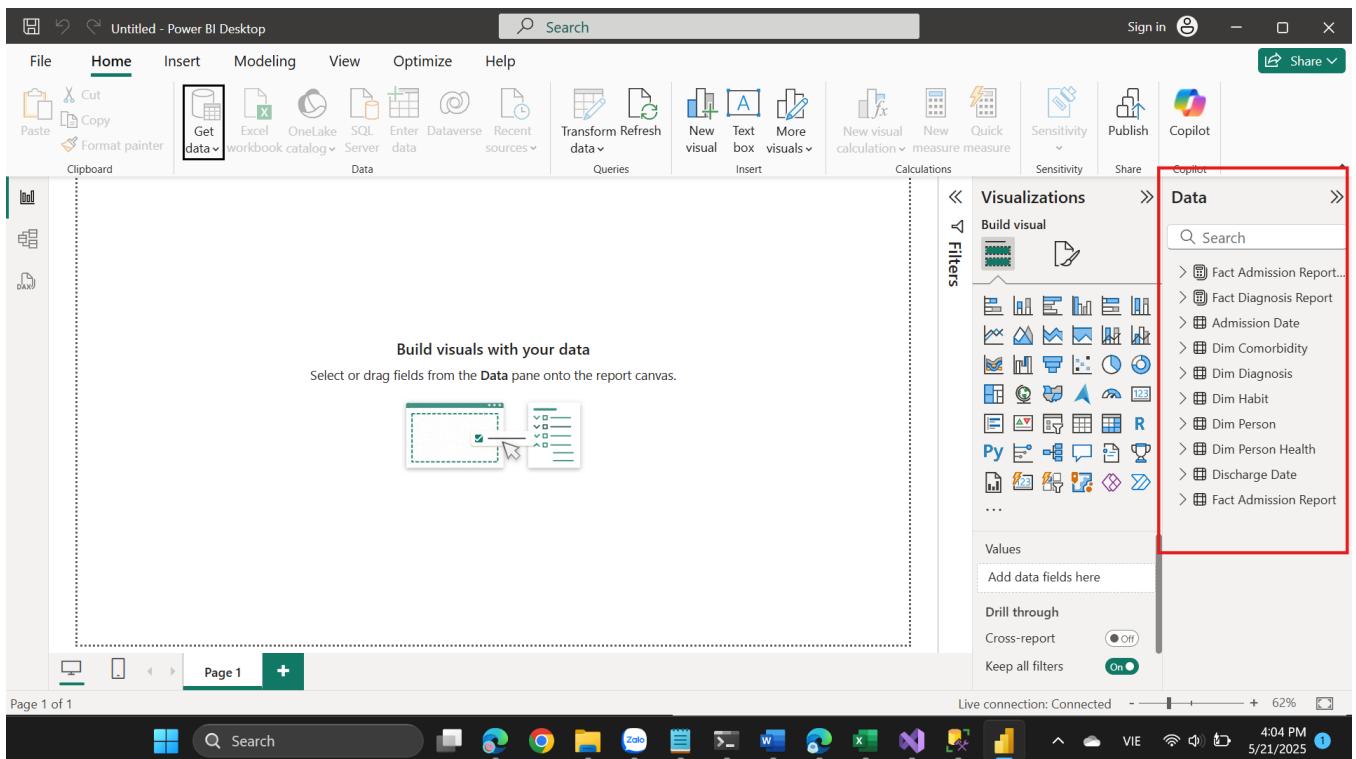
2. Chọn tên Server và Database



3. Nhấn OK để kết nối



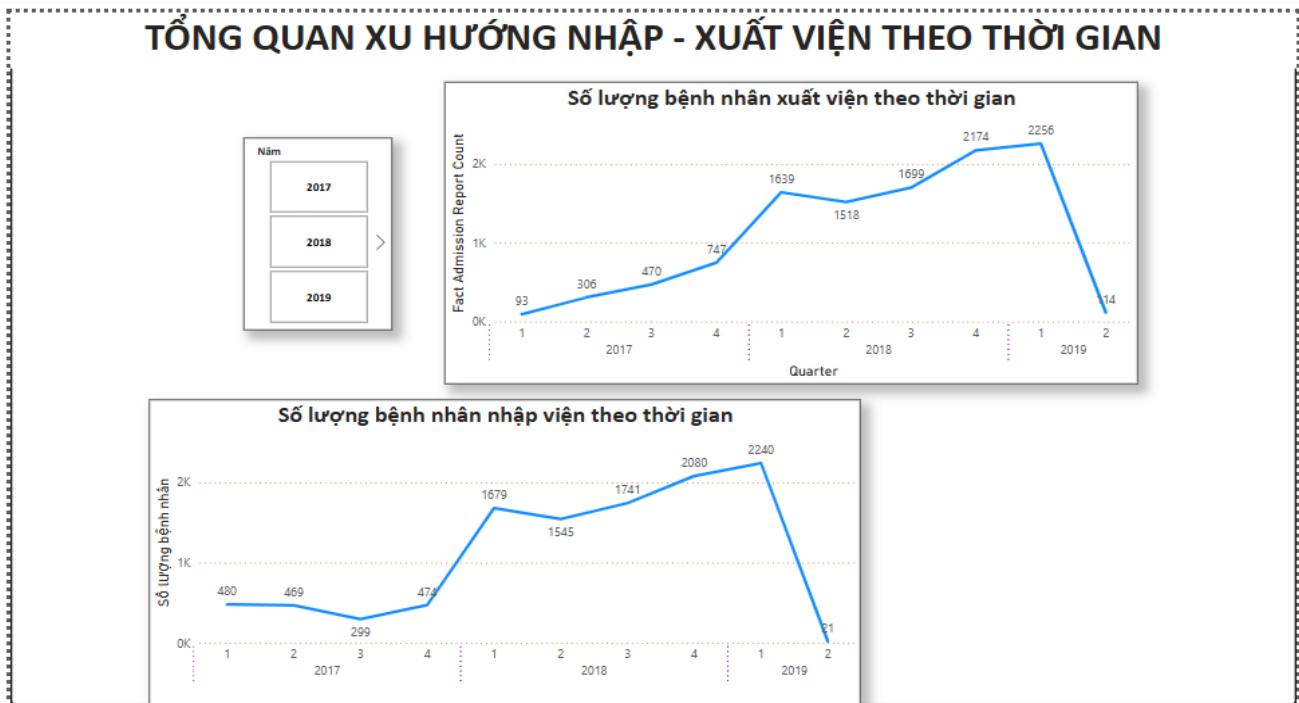
4. Cửa sổ Data hiện ra, kết nối thành công.



5. Tiến hành làm báo cáo

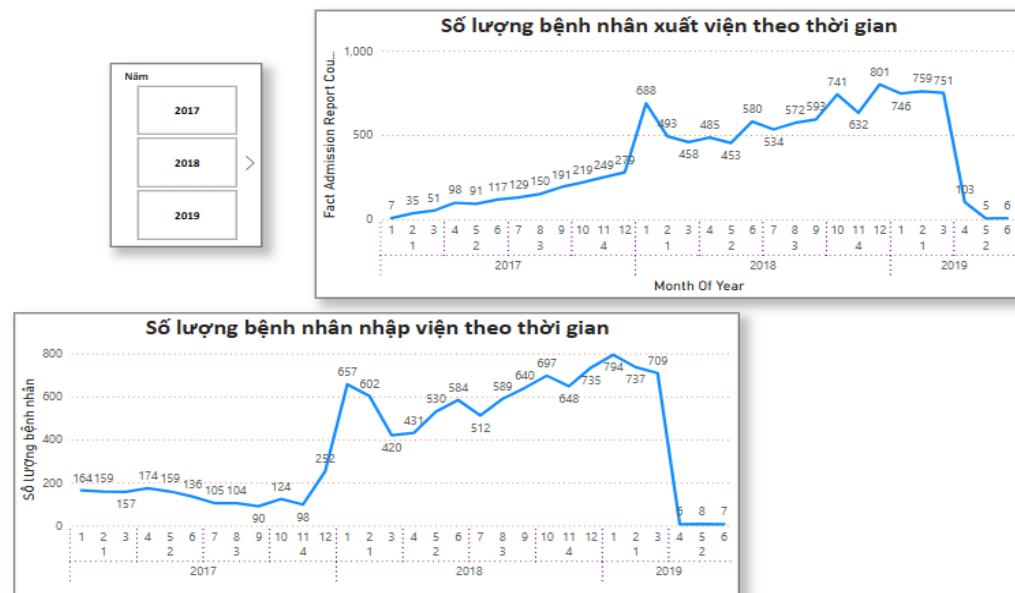
Câu 1: Tổng quan xu hướng nhập - xuất viện theo thời gian từ 2017-2019

- Theo quý:



- Theo tháng:

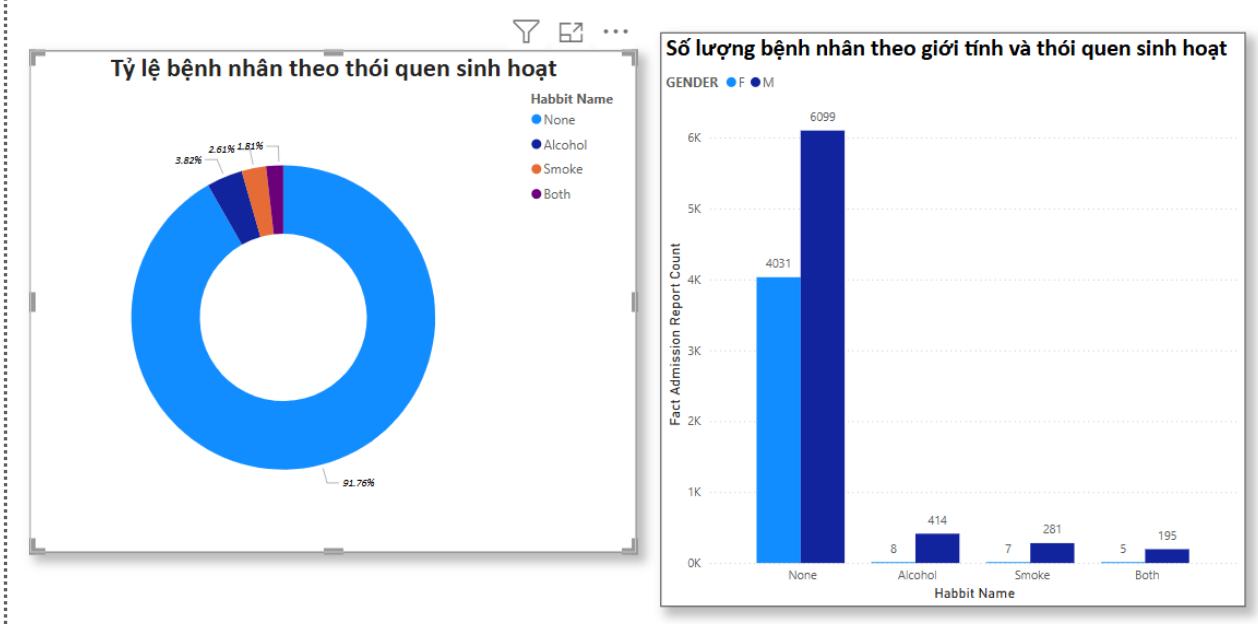
TỔNG QUAN XU HƯỚNG NHẬP - XUẤT VIỆN THEO THỜI GIAN



- Nhận xét:
- + Số lượng bệnh nhân xuất viện có xu hướng tăng rõ rệt từ quý 1/2017 đến quý 1/2018.
- + Số lượng bệnh nhân nhập viện ở mức thấp trong năm 2017.
- + Từ năm 2017 đến đầu 2019, số lượng nhập viện và xuất viện có xu hướng tăng mạnh. Đặc biệt, giai đoạn 2018–Q1/2019 là thời kỳ hoạt động y tế diễn ra sôi nổi nhất.

Câu 2: Thói quen của bệnh nhân

THÓI QUEN CỦA BỆNH NHÂN NHẬP VIỆN



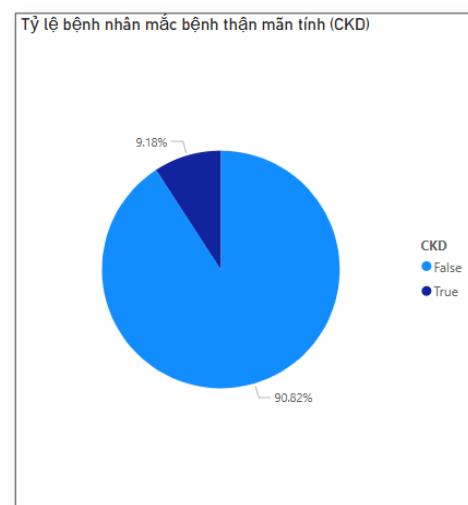
- Nhận xét:

- + Đa số bệnh nhân nhập viện không có thói quen rượu/bia hoặc thuốc lá: chiếm tới 91.76%.
- + Nhóm có thói quen chỉ uống rượu chiếm 3.82%, chỉ hút thuốc 2.61%, và cả hai thói quen là 1.81%.
- + Khi phân tích theo giới tính:
 - Nam giới (M) có số lượng cao hơn rõ rệt trong cả 4 nhóm thói quen.
 - Nhóm “None” vẫn chiếm tỷ lệ cao nhất ở cả nam và nữ, tuy nhiên sự khác biệt giới tính rõ rệt nhất ở nhóm “Alcohol” và “Both”, cho thấy số lượng nam giới có thói quen xấu cao hơn.

Câu 3: Tổng quan bệnh thận mạn tính (CKD) và mối liên hệ với thiếu máu

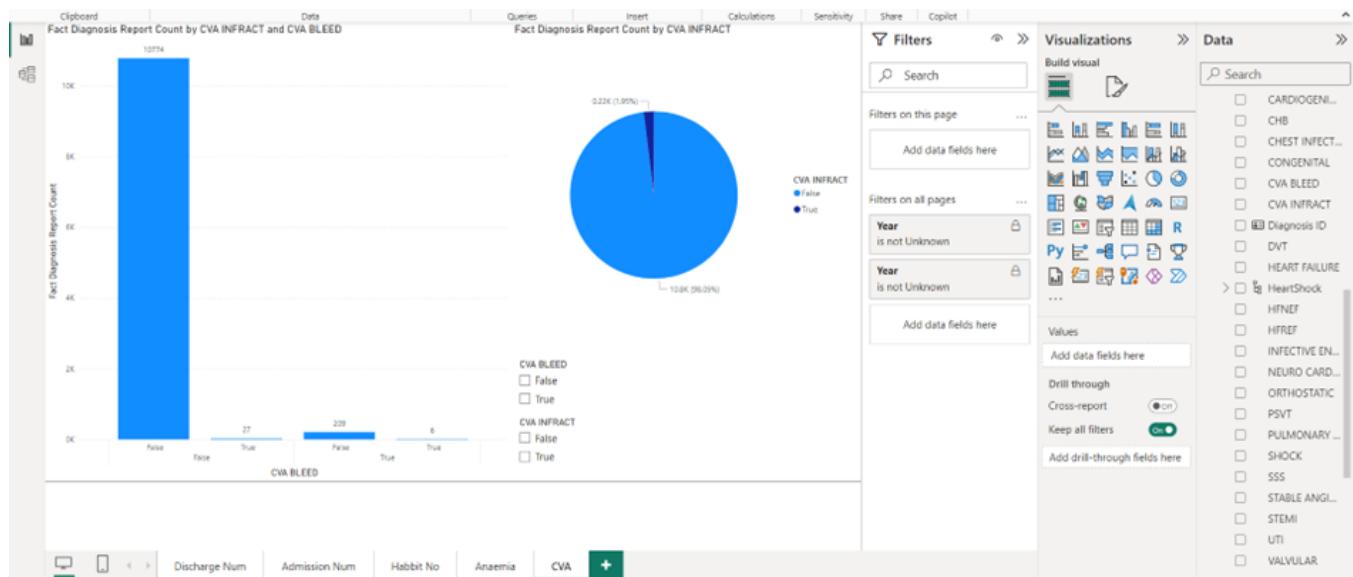
Tổng quan bệnh thận mạn tính (CKD) và mối liên hệ với thiếu máu

Số lượng bệnh nhân theo tổ hợp bệnh thận mạn tính, thiếu máu và thiếu máu nghiêm trọng		
False	False	False
8591		
	1263	True False False
	507	True True False
	448	False True True
	173	Tr...
	58	



- Nhận xét:
 - + Phần lớn bệnh nhân (8.591 người) không có CKD, không thiếu máu thường hay nghiêm trọng (False False False)
 - + Trong nhóm có CKD:
 - 1.263 bệnh nhân có CKD nhưng không thiếu máu nghiêm trọng
 - 448 người có cả CKD và thiếu máu
 - 58 người có cả CKD và thiếu máu nghiêm trọng → nhóm này rất nhỏ nhưng là tổ hợp 3 bệnh lý nặng cùng lúc, cần ưu tiên theo dõi và can thiệp y tế.

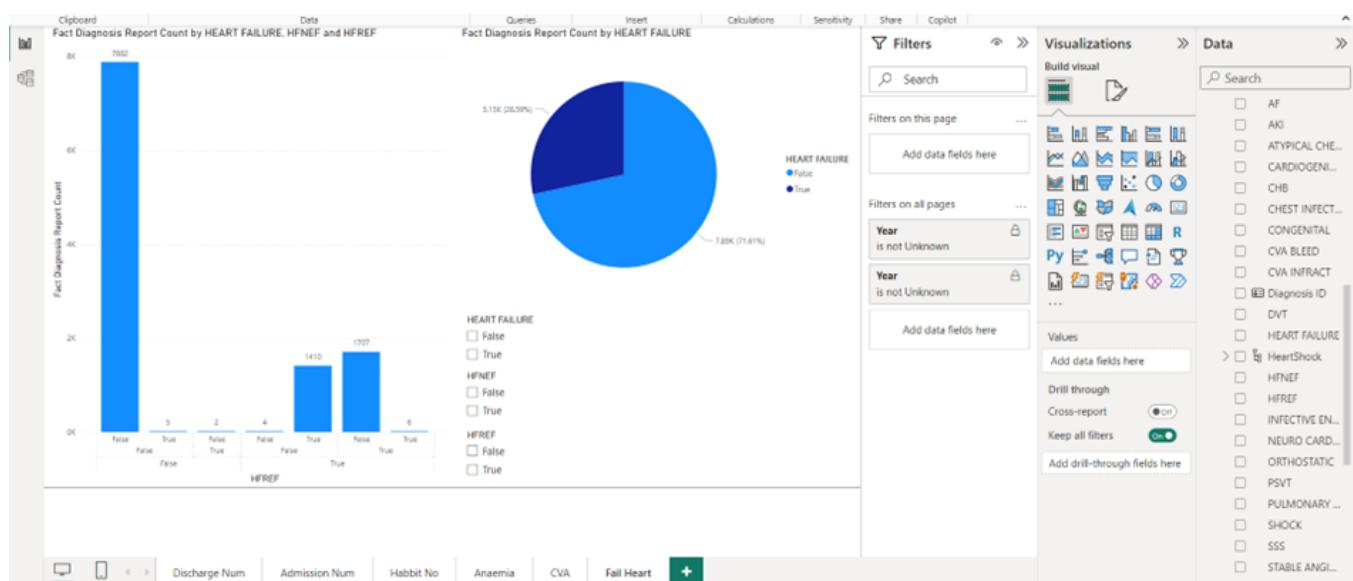
Câu 4: Nguyên nhân dẫn đến đột quy



33 ca đột quy do xuất huyết não, 217 ca đột quy do đông máu, 6 ca đột quy do vừa xuất huyết não vừa đông máu. Ta có thể thấy đa số pha đột quy là do tình trạng đông máu.

-> Kê thuốc chống đông máu cho những bệnh nhân có khả năng đột quy.

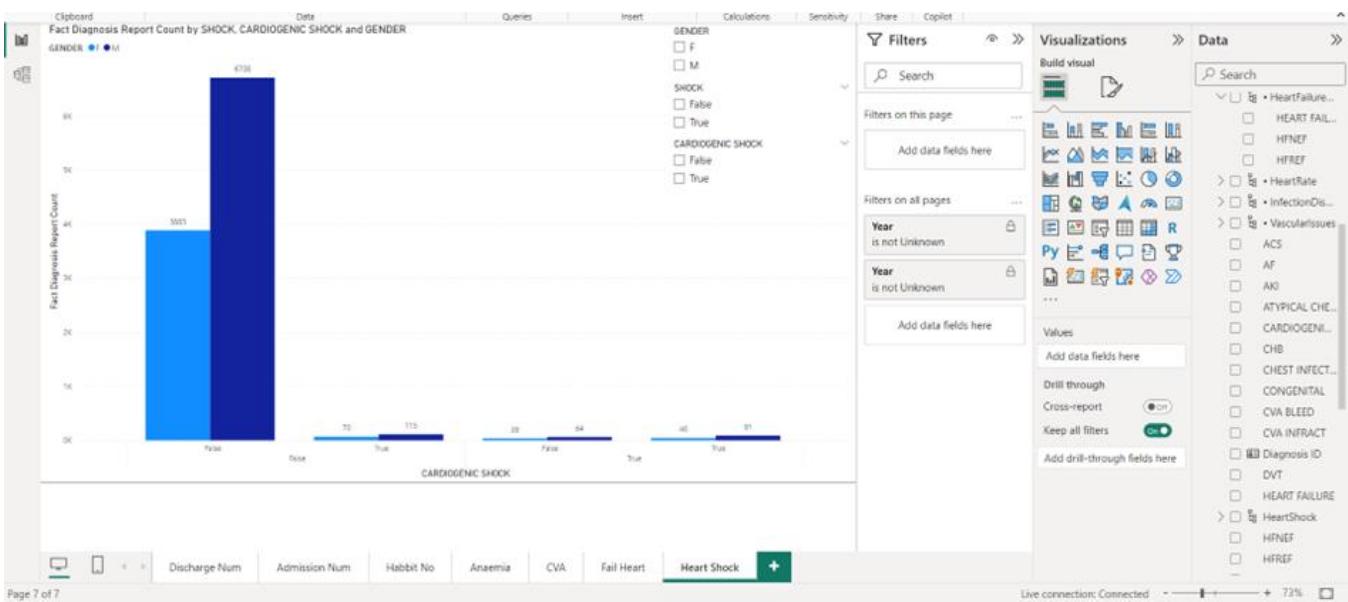
Câu 5: Tỷ lệ bệnh nhân bị suy tim



Hầu hết bệnh nhân đều không bị hội chứng suy tim (~70%) và nếu các bệnh nhân có hội chứng suy tim thì khả năng họ bị suy tim nặng hoặc nhẹ là 50%.

-> Hỗ trợ bệnh nhân trong việc khám bệnh.

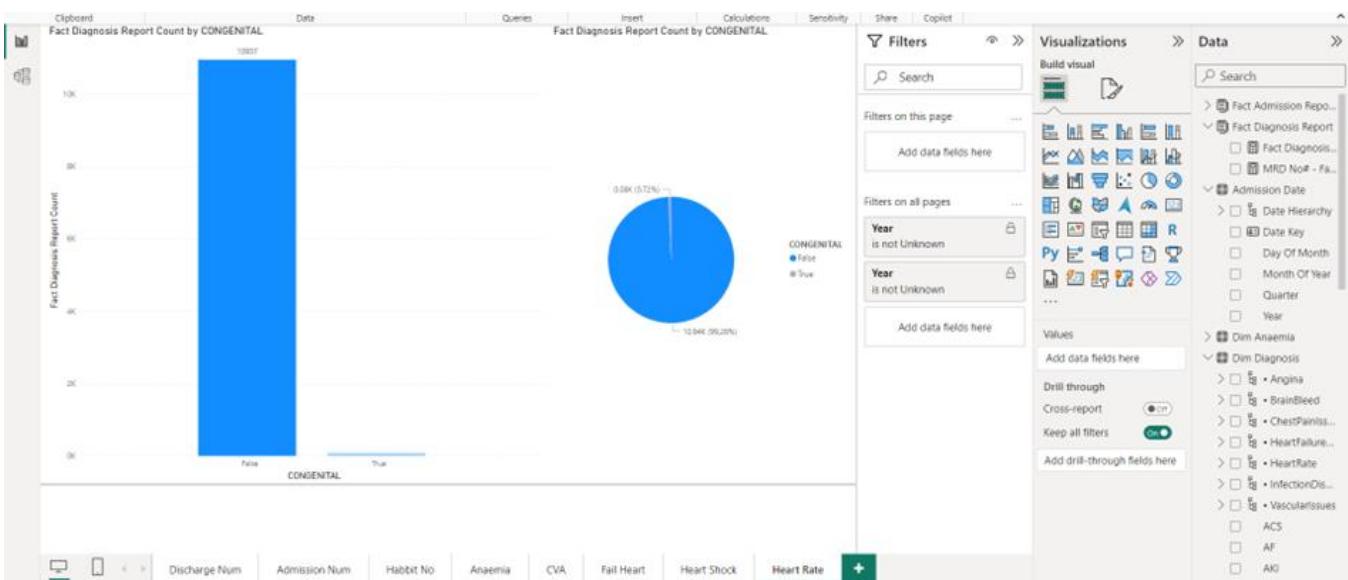
Câu 6: Tỷ lệ bệnh nhân bị sốc tim theo giới tính



Ta thấy 90% bệnh nhân không bị sốc tim, và trong số những người bị sốc tim, khả năng bệnh họ trở nặng là ~50%. Ta cũng thấy số người bị sốc tim là nam cao gần gấp đôi so với nữ.

-> Ta có thể tập trung điều trị cho các bệnh nhân nam khi khả năng họ trở nặng là cao hơn so với nữ.

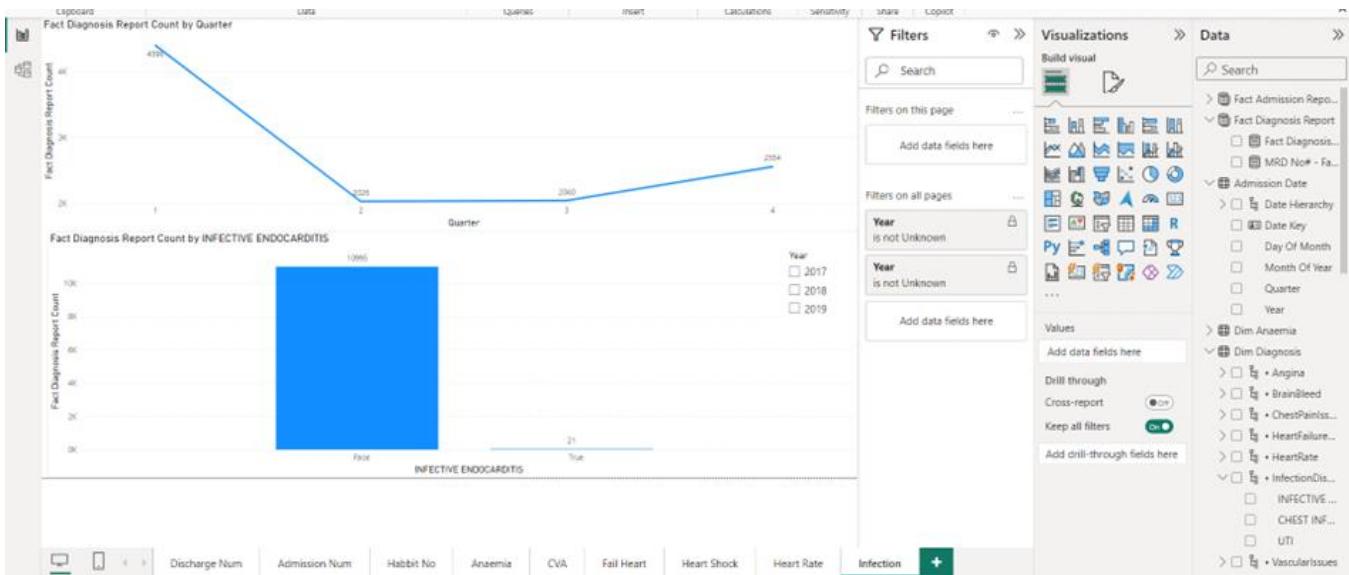
Câu 7: Bệnh nhân mắc các vấn đề về nhịp tim bất thường



Ta thấy phần đông bệnh nhân không bị tim bẩm sinh và những người bị bệnh tim bẩm sinh sẽ không có các vấn đề về rối loạn nhịp tim.

-> Giúp bác sĩ trong việc xây dựng phác đồ điều trị.

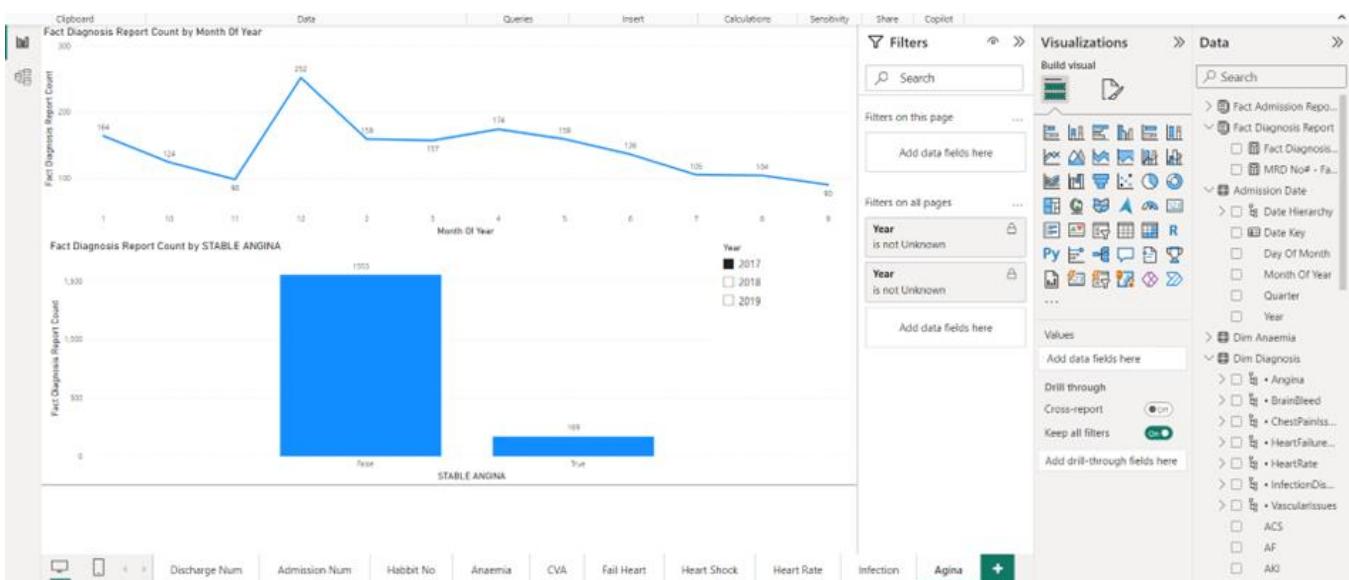
Câu 8: Có bao nhiêu bệnh nhân mắc các bệnh nhiễm trùng theo thời gian?



Ta thấy hầu hết bệnh nhân đều không bị viêm nội tâm mạc. Tuy nhiên những người bị viêm nội tâm mạc sẽ không bị các bệnh nhiễm trùng khác. Đồng thời những người không bị viêm nội tâm mạc có khả năng rất nhỏ bị nhiễm trùng (~4%).

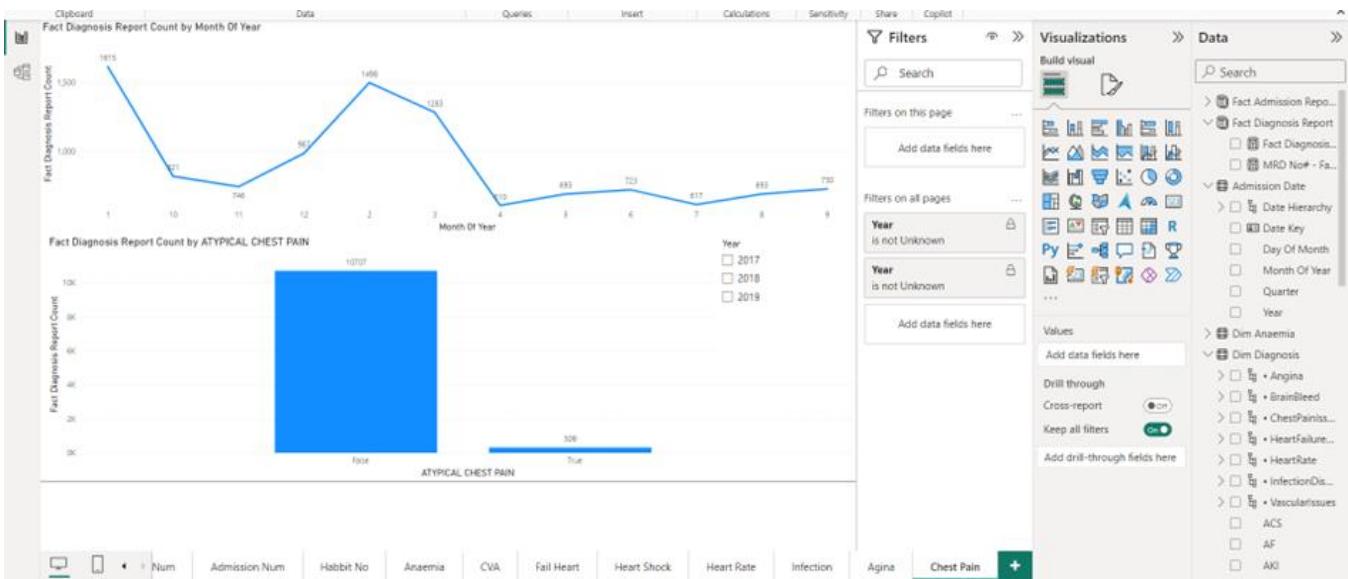
-> Tiết kiệm chi phí cho bệnh nhân nếu họ không bị viêm nội tâm mạc.

Câu 9: Có bao nhiêu bệnh nhân bị đau thắt ngực liên tục và nhồi máu cơ tim cấp tính?



Có 1396 ca bị nhồi máu cơ tim cấp tính (STEMI). 974 ca bị đau thắt ngực liên tục. 12 ca bị cả 2 đau thắt ngực liên tục và nhồi máu cơ tim. Ta thấy đau thắt ngực liên tục (Stable Angina) và nhồi máu cơ tim (STEMI) có vẻ không liên quan nhau lắm.

Câu 10: Có bao nhiêu bệnh nhân mắc các bệnh đau ngực



Ta thấy có 986 ca co thắt ngực liên tục nhưng không bị đau ngực không điếm hình (ACP).
309 ca đau ngực không điếm hình (ACP) nhưng không bị co thắt ngực liên tục.

Câu 11: Có bao nhiêu bệnh nhân bị các vấn đề về mạch máu như thuyên tắc tĩnh mạch sâu hay thuyên tắc tuyến phổi?



321 ca bị sốc tim (CARDIOGENIC SHOCK). 72 ca thuyên tắc phổi (PULMONARY EMBOLISM). 6 ca vừa thuyên tắc phổi vừa sốc tim. 96 ca thuyên tắc tĩnh mạch sâu (DVT). 57 ca vừa thuyên tắc tĩnh mạch sâu vừa thuyên tắc phổi.

-> Ta thấy các bệnh về thuyên tắc mạch máu khá liên quan với nhau

CHƯƠNG 5: KẾT LUẬN

5.1. Kết quả đạt được

Trong suốt quá trình triển khai đồ án xây dựng hệ thống "KHO DỮ LIỆU CHO QUẢN LÝ NHẬP VÀ XUẤT VIỆN BỆNH NHÂN", chúng em đã thu được những kết quả học tập và ứng dụng thực tế đáng ghi nhận. Dự án này đã củng cố và làm sâu sắc hơn kiến thức nền tảng về môn học Kho Dữ liệu. Đồng thời, chúng em đã trở nên thành thạo hơn trong việc sử dụng các công cụ hỗ trợ quá trình ETL, đặc biệt là SQL Server Integration Services (SSIS), để chuyển đổi và tích hợp dữ liệu vào môi trường kho dữ liệu. Bên cạnh đó, việc khám phá và làm việc với SQL Server Analysis Services (SSAS) đã trang bị cho chúng em khả năng xây dựng các cấu trúc phân tích đa chiều, mở ra nhiều góc nhìn về dữ liệu. Cuối cùng, việc sử dụng Excel PivotTable và Power BI Desktop đã giúp chúng em hiện thực hóa dữ liệu thông qua những biểu đồ và báo cáo trực quan, dễ dàng nắm bắt và diễn giải.

Không chỉ dừng lại ở kiến thức chuyên môn, dự án còn tạo cơ hội để chúng em rèn luyện kỹ năng thu thập, phân tích và xử lý thông tin đầu vào một cách có hệ thống. Quá trình đặt câu hỏi, thảo luận và cùng nhau tìm kiếm giải pháp cho các thách thức đã nâng cao khả năng làm việc nhóm và tư duy giải quyết vấn đề của mỗi thành viên trong chúng em.

5.2. Hạn chế

Mặc dù nhóm đã nỗ lực hoàn thành các yêu cầu trong phạm vi đề tài, chúng em nhận thấy dự án xây dựng kho dữ liệu vẫn còn tồn tại một số hạn chế nhất định. Trước hết, do giới hạn về thời gian thực hiện cũng như kinh nghiệm còn đang trong quá trình tích lũy, nhóm mới chỉ triển khai được các bước cơ bản trong quy trình xây dựng kho dữ liệu, từ thiết kế mô hình dữ liệu đến quá trình ETL và trực quan hóa ban đầu. Việc khai thác sâu các khía cạnh nâng cao như tối ưu hóa hiệu suất truy vấn, tích hợp dữ liệu từ nhiều nguồn phức tạp, xây dựng các chỉ số phân tích chuyên sâu (advanced KPIs) hay triển khai các tính năng bảo mật vẫn chưa được thực hiện.

Bên cạnh đó, việc phân tích và khai thác triệt để tiềm năng của tập dữ liệu gốc vẫn còn hạn chế, dẫn đến các dashboard trực quan và báo cáo phân tích chưa đạt được chiều sâu mong

muốn. Những hạn chế này là điều khó tránh khỏi đối với một đề tài thực hành trong khuôn khổ học phần, tuy nhiên, chúng em xem đây là những bài học quý giá cho việc phát triển năng lực cá nhân cũng như định hướng nâng cao chất lượng các dự án trong tương lai.

Tài liệu tham khảo

- [1] Các tài liệu định dạng PDF hướng dẫn về Tích hợp Dữ liệu sử dụng SQL Server, SSIS, SSAS từ giảng viên hướng dẫn, Thạc sĩ Nguyễn Văn Thành.
- [2] Minewiskan (năm không xác định), *Bài học 3: Chính sửa các Measures, Attributes và Hierarchies*. [trực tuyến] learn.microsoft.com. Truy cập tại:
<https://learn.microsoft.com/en-us/analysis-services/multidimensional-tutorial/lesson-3-modifying-measures-attributes-and-hierarchies?view=asallproducts-allversions>
- [3] Minewiskan (năm không xác định), *Bài học 4: Định nghĩa các thuộc tính và thuộc tính Dimension nâng cao*. [trực tuyến] learn.microsoft.com. Truy cập tại:
<https://learn.microsoft.com/en-us/analysis-services/multidimensional-tutorial/lesson-4-defining-advanced-attribute-and-dimension-properties?view=asallproducts-allversions>
- [4] Nguyễn Văn Chúc (năm không xác định). *Dự án ETL từ nguồn dữ liệu Excel đến Star Schema với SSIS*. [trực tuyến] youtube.com. Truy cập tại:
https://www.youtube.com/watch?v=Yp8fXLnVCp8&ab_channel=ChucNguyenVan
- [5] Bollepalli, S.C.; Sahani, A.K.; Aslam, N.; Mohan, B.; Kulkarni, K.; Goyal, A.; Singh, B.; Singh, G.; Mittal, A.; Tandon, R.; Chhabra, S.T.; Wander, G.S.; Armoundas, A.A. (2022). An Optimized Machine Learning Model Accurately Predicts In-Hospital Outcomes at Admission to a Cardiac Unit. *Diagnostics*, 12(2), 241. <https://doi.org/10.3390/diagnostics12020241>