

TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH

KHOA CÔNG NGHỆ THÔNG TIN



ĐỀ TÀI CUỐI KỲ MÔN KHO DỮ LIỆU

TÊN ĐỀ TÀI: Xây dựng kho dữ liệu về việc nhập và xuất viện của bệnh nhân

GVHD: ThS. Nguyễn Văn Thành

Lớp HP: DAWH430784_23_2_01

Nhóm thực hiện: Nhóm 13

Học kỳ: 2

Năm học: 2023 - 2024

Thành phố Hồ Chí Minh, tháng 5, năm 2024

DANH SÁCH SINH VIÊN NHÓM THỰC HIỆN

HỌC KỲ II NĂM HỌC 2023-2024

Nhóm 13

Đề tài: Xây dựng kho dữ liệu về việc nhập và xuất viện của bệnh nhân.

STT	Họ và tên	MSSV	Tỷ lệ đóng góp
1	Trần Quốc Bảo	21133005	100%
2	Lê Đức Thắng	21133082	100%
3	Nguyễn Tấn Phát	21133107	100%
4	Huỳnh Nguyễn Tuấn Kiệt	20132213	100%

Nhận xét của giảng viên:

Ngày ... tháng 05 năm 2024

Giảng viên chấm điểm
Ths. Nguyễn Văn Thành

LỜI CẢM ƠN

Lời đầu tiên, nhóm em xin được gửi lời cảm ơn đặc biệt đến **Thầy – Ths. Nguyễn Văn Thành** - Giảng viên phụ trách môn **Kho Dữ Liệu – Trường Đại học Sư Phạm Kỹ Thuật TP. Hồ Chí Minh.**

Trong thời gian nhóm em làm đồ án, chúng em đã nhận được nhiều sự giúp đỡ từ thầy. Thầy đã cung cấp đầy đủ kiến thức, chỉ bảo và đóng góp những ý kiến quý báu giúp chúng em có thể hoàn thành được đồ án môn học của mình một cách tốt nhất.

Xuất phát từ mục đích học tập, tìm hiểu sâu hơn các kiến thức về kho dữ liệu, cũng như tìm hiểu kỹ về quy trình nghiệp vụ của lên ý tưởng, xây dựng một kho dữ liệu. Nhóm chúng em đã thực hiện đồ án “**Xây dựng kho dữ liệu về việc nhập và xuất viện của bệnh nhân**”. Trong quá trình thực hiện đồ án, dựa trên kiến thức được Thầy cung cấp qua các buổi học lý thuyết cũng như thực hành trên lớp, kết hợp với việc tự tìm hiểu những công cụ và kiến thức mới, nhóm đã cố gắng thực hiện đồ án một cách tốt nhất. Tuy nhiên, đồ án còn chưa được hoàn thiện và có nhiều sai sót.

Nhóm rất mong nhận được sự góp ý từ thầy nhằm rút ra những kinh nghiệm quý báu và hoàn thiện vốn kiến thức để nhóm có thể hoàn thành những đồ án, dự án khác trong tương lai .

Nhóm chúng em xin chân thành cảm ơn thầy!

PHÂN CÔNG NHIỆM VỤ

Nhiệm vụ	Trần Quốc Bảo	Huỳnh Nguyễn Tuấn Kiệt	Lê Đức Thắng	Nguyễn Tấn Phát
Tìm kiếm tập dữ liệu.		X		
Mô hình hóa dữ liệu vào workbook.				
+ DimDate. + DimPerson.			X	
+ DimAnaemia. + DimPersonHealth.				X
+ DimDiagnosis. + FactAdmissionReport.	X			
+ DimHabbit. + FactDiagnosisReport.		X		
Thực hiện ETL đưa dữ liệu vào các bảng Dimension và Fact dùng SSIS.				
- Tiến hành đưa dữ liệu gốc vào database để phục vụ quá trình ETL.	X			
- Tiến hành bước Stage và đưa dữ liệu vào các Dimension.				
+ DimDate.			X	
+ DimPerson. + DimAnaemia. + DimPersonHealth.	X			
+ DimDiagnosis. + DimHabbit.				X
- Tiến hành bước Stage và đưa dữ liệu vào các Fact.		X		

Nhập dữ liệu vào SSAS, tạo Data Source, Data Source View và Cube				X
Tạo các phân cấp hierarchy cho các bảng Dimension và các Attribute Relationships.				
+ Hierarchy cho DimDate.		X		
+ Hierarchy cho DimDiagnosis.	X			X
+ Hierarchy cho DimAnaemia.				X
Thực hiện trả lời các câu hỏi đã đặt ra cho Business Process: AdmissionReport và DiagnosisReport				
- Cho biết số lượng bệnh nhân nhập/xuất viện mỗi ngày?	X			
- Cho biết số lượng bệnh nhân với các loại thói quen?				X
- Cho biết số lượng bệnh nhân với tiền sử các bệnh lý?			X	
- Có bao nhiêu bệnh nhân bị bệnh thiếu máu ở các tình trạng khác nhau?		X		
- Cho biết có bao nhiêu bệnh nhân bị sốc tim theo giới tính?				X
- Nguyên nhân do đâu mà bệnh nhân bị đột quy?	X			
- Có bao nhiêu bệnh nhân mắc các vấn đề về nhịp tim bất thường?		X		
- Có bao nhiêu bệnh nhân mắc các bệnh nhiễm trùng theo thời gian?			X	

- Có bao nhiêu bệnh nhân bị các vấn đề về mạch máu như thuyên tắc tĩnh mạch sâu hay thuyên tắc phổi?				X
- Có bao nhiêu bệnh nhân bị đau thắt ngực liên tục và nhồi máu cơ tim cấp tính?	X			
- Có bao nhiêu bệnh nhân mắc các bệnh đau ngực?		X		
Thực hiện trả lời các câu hỏi đã đặt ra cho Business Process: AdmissionReport và DiagnosisReport.				
+ Trả lời câu hỏi dùng Power BI Desktop			X	
+ Trả lời câu hỏi dùng Excel Pivot		X		
Tạo bảng báo cáo sử dụng SSRS		X		

MỤC LỤC

CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI	1
1.1. Lý do chọn đề tài	1
1.2. Tổng quan tập dữ liệu	1
1.2.1. Nguồn dữ liệu sử dụng	1
1.2.2. Mô tả tập dữ liệu.....	2
1.3. Giới thiệu các công cụ được sử dụng	6
CHƯƠNG 2: THIẾT KẾ XÂY DỰNG CƠ SỞ DỮ LIỆU TÁC NGHIỆP (OLAP)	7
2.1. Xác định các Business Process và bảng Fact.....	7
2.1.1. Business Process: Admission Report	7
2.1.2. Business Process: Diagnosis Report	7
2.2. Xây dựng các bảng Dimension	8
2.2.1. DimPerson	8
2.2.2. DimAnaemia	8
2.2.3. DimPersonHealth	8
2.2.4. DimDiagnosis	9
2.2.5. DimHabbit	9
2.3. Galaxy Schema (Lược đồ hình ngân hà)	9
CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS).....	10
3.1. Tạo ETL Project và Solution	10
3.2. Thiết lập các Connection.....	11
3.2. Import Dữ liệu vào Database gốc	14
3.2.1. Tạo Package.....	14
3.2.2. Thiết lập Control Flow	15
3.2.3. Load dữ liệu từ Source File vào Database	16
3.3. Import Dữ liệu vào các bảng Dimension	20

3.3.1. Tạo package.....	20
3.3.2. Stage các Dimension	21
3.3.3. Load dữ liệu vào các Dimension.....	56
3.4. Import Dữ liệu vào các bảng Fact.....	84
3.4.1. Tạo package.....	84
3.3.2. Stage các Fact.....	84
3.3.3. Load các Fact.....	93
3.5. Tổng hợp quá trình Tích hợp dữ liệu (SSIS)	98
CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS).....	99
4.1. Quá trình xây dựng mô hình.....	99
4.1.1. Tạo Project.....	99
4.1.2. Tạo Data Source	100
4.1.3. Tạo Data Source View	103
4.2. Quá trình tạo Cube	105
4.3. Quá trình tạo Hierarchy	107
4.3.1. Tạo Hierarchy cho DimDate	107
4.3.2. Tạo Hierarchy cho DimDiagnosis	110
4.3.3. Tạo Hierarchy cho DimAnaemia	111
4.4. Tiến hành trả lời các câu hỏi Business Process	112
4.4.1. Sử dụng công cụ Excel Pivot Table	112
4.4.2. Sử dụng Power BI	119
CHƯƠNG 5: BÁO CÁO - REPORT.....	130
5.1. Cấu hình SSRS (SQL Server Reporting Services)	130
5.2. Quá trình tạo Project SSRS.....	138
5.3. Báo cáo về số lượng bệnh nhân và các vấn đề về nhịp tim theo từng năm.....	140
5.4. Báo cáo về các chỉ số trung bình của các bệnh nhân.....	144

CHƯƠNG 6: KẾT LUẬN	147
6.1. Kết quả đạt được	147
6.2. Những hạn chế	147
6.3. Hướng phát triển	147
6.4. Tài liệu tham khảo	147

NỘI DUNG

CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI

Giới thiệu tổng quan về tập dữ liệu, lý do hình thành dự án, nguồn dữ liệu thực hiện.

Tính cần thiết của xây dựng cơ sở dữ liệu cũng như tiến hành khảo sát, nghiên cứu và phân tích các báo cáo nghiệp vụ cần phục vụ cho kho dữ liệu.

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, 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 dữ liệu sử dụng

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: [Hospital Admissions Data \(kaggle.com\)](#)

1.2.2. Mô tả tập dữ liệu

1.2.2.1. Dữ liệu của tập dữ liệu

Dữ liệu này được thu thập từ bệnh nhân được nhập viện trong khoảng hai năm (1 tháng 4 năm 2017 đến 31 tháng 3 năm 2019) tại Viện tim mạch Hero DMC, đơn vị thuộc Trường Đại học Y khoa và Bệnh viện Dayanand, Ludhiana, Punjab, Ấn Độ. Đây là một trường y tế cao cấp và bệnh viện. Trong thời gian nghiên cứu, đơn vị tim mạch có 14.845 lần nhập viện tương ứng với 12.238 bệnh nhân, trong đó có 1.921 bệnh nhân có nhiều lần nhập viện.

Cụ thể, dữ liệu liên quan đến bệnh nhân; ngày nhập viện; ngày xuất viện; thông tin nhân khẩu học như tuổi, giới tính, địa phương (đô thị hoặc nông thôn); loại nhập viện (khẩn cấp hoặc ngoại trú); tiền sử bệnh nhân, bao gồm hút thuốc, uống rượu, tiểu đường (DM), tăng huyết áp (HTN), bệnh động mạch vành (CAD), bệnh cơ tim trước đó (CMP) và bệnh thận mãn tính (CKD); và các chỉ số thí nghiệm tương ứng với huyết globin (HB), tổng số bạch cầu (TLC), tiểu cầu, đường huyết, ure, creatinine, peptide natriuret (BNP), men tim yếu (RCE) và tỷ lệ nhảy (EF). Những bệnh lý đồng bệnh khác và các đặc điểm (28 đặc điểm), bao gồm suy tim, STEMI và tràn khí phổi, đã được ghi lại và phân tích.

Các kết quả chỉ ra liệu bệnh nhân đã xuất viện hay tử vong trong bệnh viện cũng được ghi lại.

1.2.2.2. Thông số tập dữ liệu

Tập dữ liệu [Hospital Admissions Data](#) là một nguồn tài nguyên quý giá để xây dựng kho dữ liệu cho dữ liệu nhập viện và xuất viện bệnh viện. Tập dữ liệu chứa thông tin về hơn 15.000 lượt nhập viện bệnh viện và 56 cột thông tin, bao gồm các thông tin về đặc điểm của bệnh nhân, ngày nhập viện và xuất viện, kết quả xét nghiệm và lịch sử bệnh nhân.

1.2.2.3. Mô tả chi tiết tập 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 2 tệp với mô tả như sau:

Tên tập dữ liệu	Mô tả	Số dòng
HDHI Admission data.csv	Lưu trữ thông tin nhập xuất viện của các bệnh nhân	15800 dòng

Bảng HDHI Admission data

Tên hiển thị	Tên đầy đủ	Kiểu dữ liệu	Giải thích
SNO	Serial Number	Int	
MRD No.	Admission Number	Int	
D.O.A	Date of Admission	Date	Ngày nhập viện
D.O.D	Date of Discharge	Date	Ngày xuất viện
AGE	AGE	Int	Tuổi
GENDER	GENDER	Varchar(1)	Giới tính
RURAL	RURAL(R) /Urban(U)	Varchar(1)	Nơi ở
TYPE OF ADMISSION EMERGENCY/OPD	TYPE OF ADMISSION EMERGENCY/OPD	Varchar(1)	Loại tiếp nhận
month year	month year	Varchar(10)	
DURATION OF STAY	DURATION OF STAY	Int	Thời gian trong viện
duration of intensive unit stay	duration of intensive unit stay	Int	Thời gian trong phòng hồi sức tích cực
OUTCOME	OUTCOME	Varchar(50)	
SMOKING	SMOKING	Bit	Có hút thuốc
ALCOHOL	ALCOHOL	Bit	Có sử dụng rượu bia
DM	Diabetes Mellitus	Bit	Có bị tiểu đường
HTN	Hypertension	Bit	Có bị cao huyết áp
CAD	Coronary Artery Disease	Bit	Có bị bệnh động mạch vành

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

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

CARDIOGENIC SYNCOPE	CARDIOGENIC SYNCOPE		kinh phế vị
ORTHOSTATIC	ORTHOSTATIC	Bit	Có bị chấn thương chỉnh hình
INFECTIVE ENDOCARDITIS	INFECTIVE ENDOCARDITIS	Bit	Có bị viêm nội tâm mạc
DVT	Deep venous thrombosis	Bit	Có bị huyết khối tĩnh mạch sâu
CARDIOGENIC SHOCK	CARDIOGENIC SHOCK	Bit	Có bị sốc tim
SHOCK	SHOCK	Bit	Có bị sốc
PULMONARY EMBOLISM	PULMONARY EMBOLISM	bit	Có bị thuyên tắc phổi
CHEST INFECTION	CHEST INFECTION	Bit	Có bị nhiễm trùng đường phổi

1.3. Giới thiệu các công cụ được sử dụng

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

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

Chương này sẽ đi vào tiền xử lý dữ liệu. Phân tích các yêu cầu khai thác từ kho dữ liệu từ đó xây dựng lược đồ phù hợp với kho dữ liệu trước khi tiến hành tích hợp dữ liệu vào kho (Quá trình SSIS)

2.1. Xác định các Business Process và bảng Fact

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

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		X	X	X		X	X	X

2.1.1. Business Process: Admission Report

Các câu hỏi được đặt ra:

- Cho biết số lượng bệnh nhân nhập/ xuất viện mỗi ngày?
- Cho biết số lượng bệnh nhân với các loại thói quen?
- Cho biết số lượng bệnh nhân với tiền sử các bệnh lý?
- Có bao nhiêu bệnh nhân bị bệnh thiếu máu ở các tình trạng khác nhau?

Table Name	Admission_fact	Column Name	Display Name
		AdmissionID	AdmissionID
Table Type	Fact	AnaemiaKey	AnaemiaKey
Display Name	Admission_fact	PersonKey	PersonKey
Database Schema	dbo	PersonalHealthKey	PersonalHealthKey
Table Description	Admission_fact	HabbitKey	HabbitKey
Comment	Sourced from Covid19	TimeKey	TimeKey
Biz Filter Logic	One row per patient admission report details	MRD No	MRD No
Size	Y	TypeOfAdmission	TypeOfAdmission
Generate Script?		Outcome	Outcome

2.1.2. Business Process: Diagnosis Report

Các câu hỏi được đặt ra:

- Cho biết có bao nhiêu bệnh nhân bị sốc tim theo giới tính?
- Nguyên nhân do đâu mà bệnh nhân bị đột quy?
- Có bao nhiêu bệnh nhân mắc các vấn đề về nhịp tim bất thường?
- Có bao nhiêu bệnh nhân mắc các bệnh nhiễm trùng theo thời gian?
- Có bao nhiêu bệnh nhân bị các vấn đề về mạch máu như thuyên tắc tĩnh mạch sâu hay thuyên tắc phổi?
- Có bao nhiêu bệnh nhân bị đau thắt ngực liên tục và nhồi máu cơ tim cấp tính?
- Có bao nhiêu bệnh nhân mắc các bệnh đau ngực?

Table Name	Diagnosis_fact	Column Name	Display Name
Table Type	Fact		
Display Name	Diagnosis_fact	DiagnosisReportID	DiagnosisReportID
Database Schema	dbo	AnaemiaKey	AnaemiaKey
Table Description	Diagnosis_fact		
Comment	Sourced from Covid19	PersonKey	PersonKey
Biz Filter Logic		PersonalHealthKey	PersonalHealthKey
Size	One row per patient diagnosis details	DiagnosisKey	DiagnosisKey
Generate Script?	Y	TimeKey	TimeKey
		MRD No	MRD No

2.2. Xây dựng các bảng Dimension

2.2.1. DimPerson

Instructions!	Attribute / Fact Name	Description	Alternate Names	Sample Values
Dimension / Fact Table				
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)

2.2.2. DimAnaemia

Instructions!	Attribute / Fact Name	Description	Alternate Names	Sample Values
DimAnaemia	DM	Diabetes Mellitus (Tiểu đường)		1/0
	HTN	Hypertension (Cao huyết áp)		1/0
	CAD	Coronary Artery Disease (Bệnh động mạch vành)		1/0
	PRIOR CMP	CARDIOMYOPATHY (Bệnh cơ tim phì đại)		1/0
	CKD	CHRONIC KIDNEY DISEASE (Bệnh suy thận mạn)		1/0
	RAISED CARDIAC ENZYMES	RAISED CARDIAC ENZYMES (Tăng enzyme tim)		1/0
	SEVERE ANAEMIA	SEVERE ANAEMIA (Có tiền sử bệnh nặng)		1/0
	ANAEAMIA	ANAEAMIA (Có tiền sử bệnh)		1/0

2.2.3. DimPersonHealth

Instructions!	Attribute / Fact Name	Description	Alternate Names	Sample Values
DimPersonHealth	HB	Haemoglobin (Chỉ số haemoglobin)		9.5
	TLC	TOTAL LEUKOCYTES COUNT (Chỉ số bạch cầu)		16.1
	PLATELETS	PLATELETS (Chỉ số tiểu cầu)		337
	GLUCOSE	GLUCOSE (Chỉ số đường huyết)		80
	UREA	UREA (Chỉ số u-rê)		34
	EF	Ejection Fraction (Phản suất tổng máu)		35
	CREATININE	CREATININE (Chỉ số creatinine)		0.9
	BNP	B-TYPE NATRIURETIC PEPTIDE (peptide natriuretic loại B)		1880

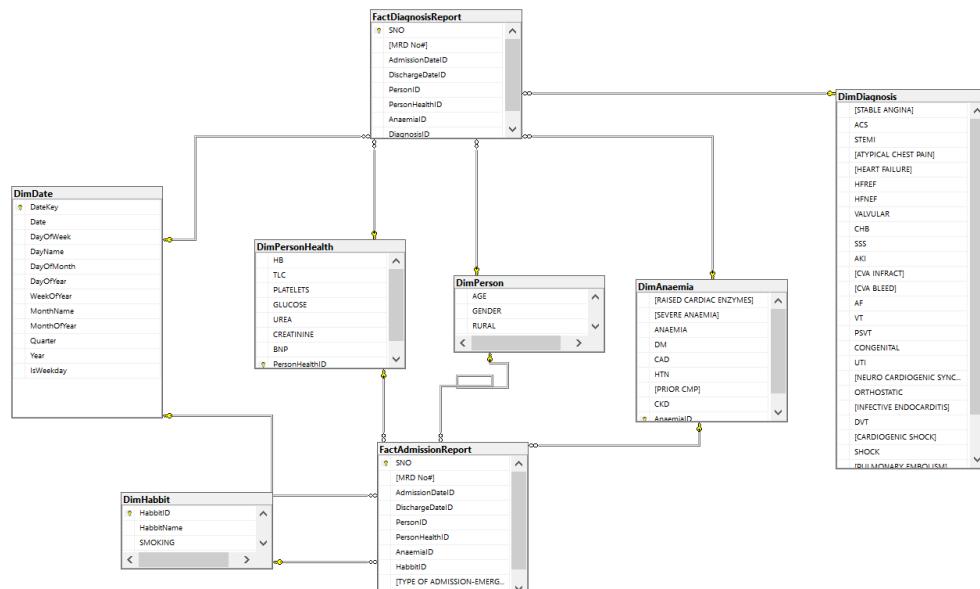
2.2.4. DimDiagnosis

Instructions! Dimension / Fact Table	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
	HFRF	HEART FAILURE WITH REDUCED EJECTION FRACTION (Suy tim giảm phân suất tổng máu binh thường)		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

2.2.5. DimHabbit

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

2.3. Galaxy Schema (Lược đồ hình ngân hà)

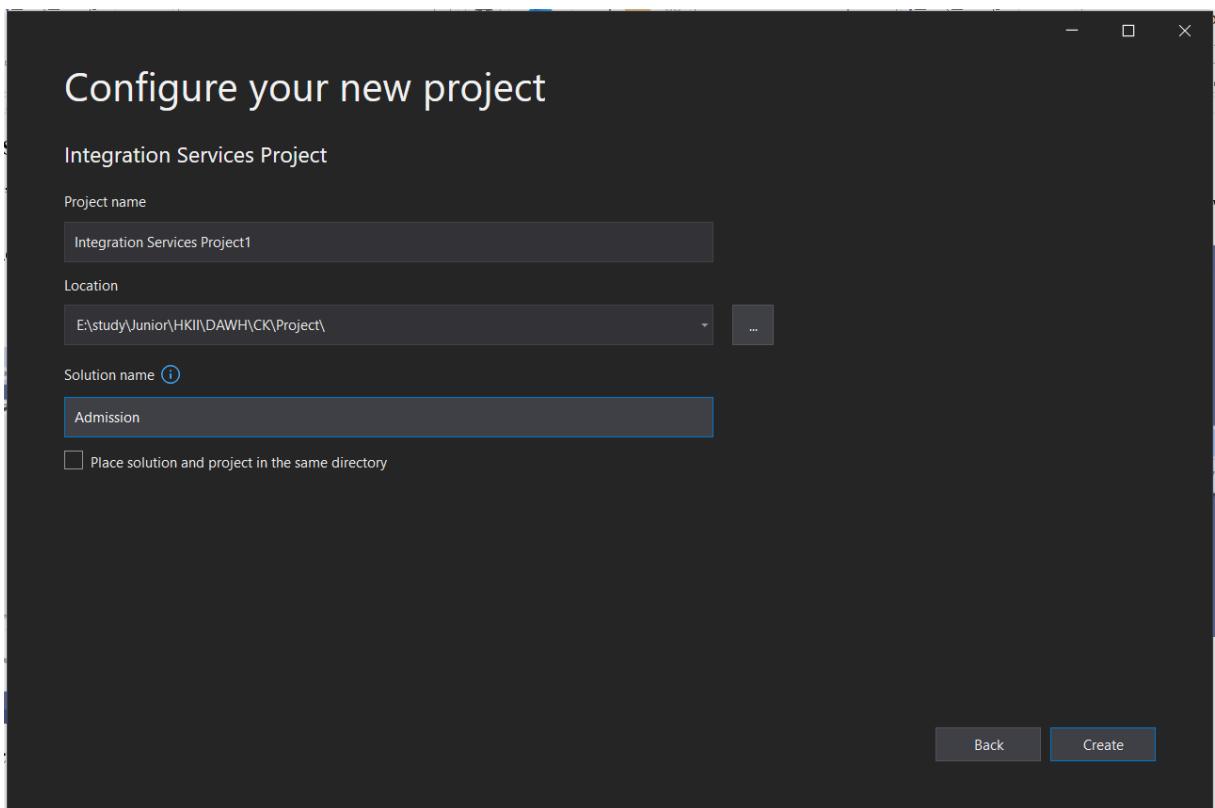


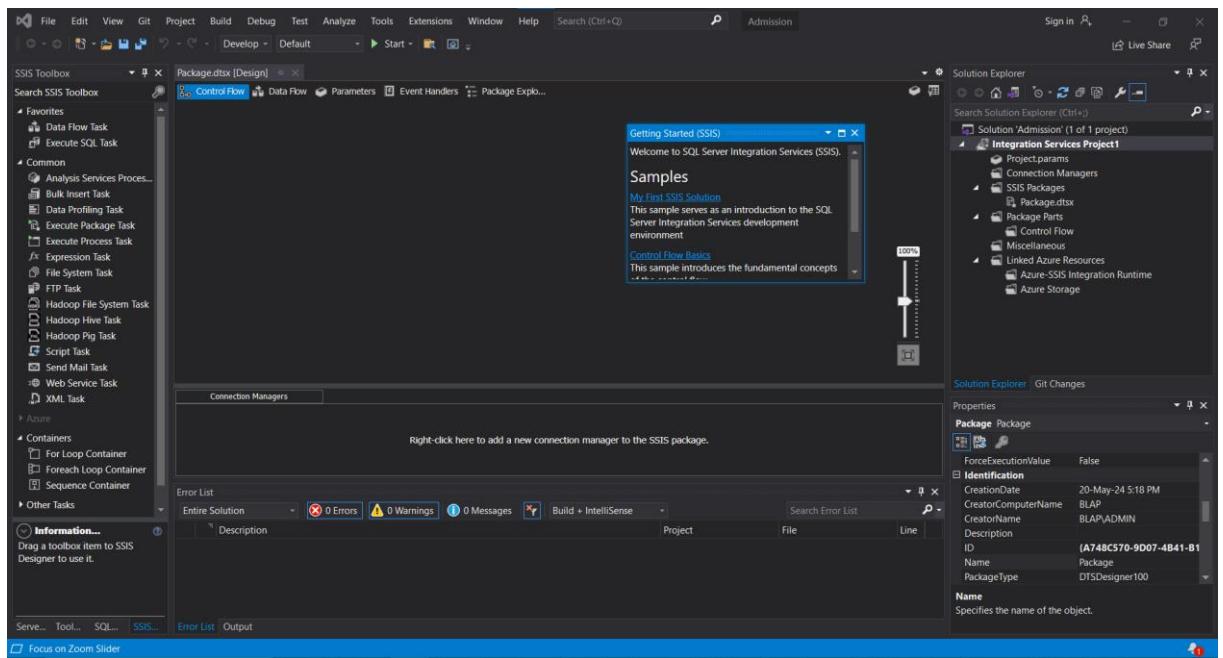
CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS)

Ở chương này, chúng em sử dụng công cụ SSIS để xây dựng các gói tích hợp dữ liệu, cài đặt các gói chạy tự động. Viết truy vấn đưa dữ liệu từ OLTP đến kho dữ liệu.

3.1. Tạo ETL Project và Solution

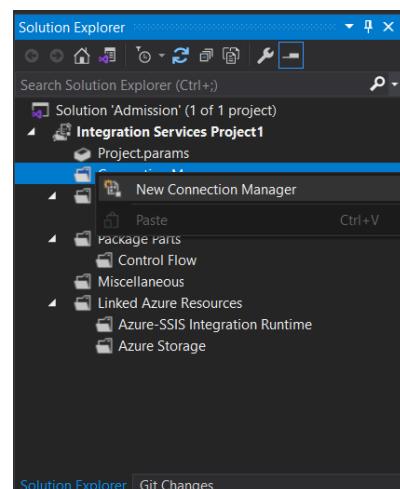
1. Từ Visual Studio menu, chọn **File → New → Project**.
2. Từ **dialog**:
 - a. Chọn **Integration Services Project**
 - b. Đặt tên: **Admission**
3. Click **OK**



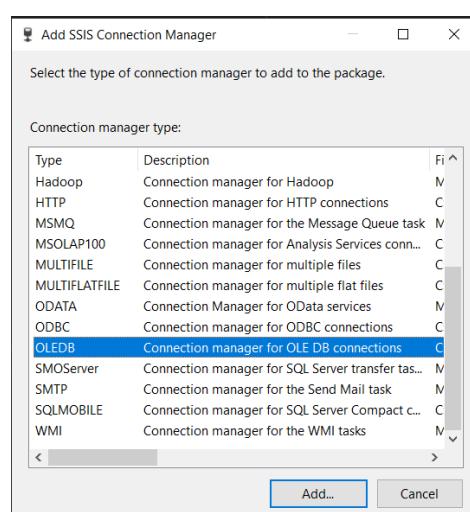


3.2. Thiết lập các Connection

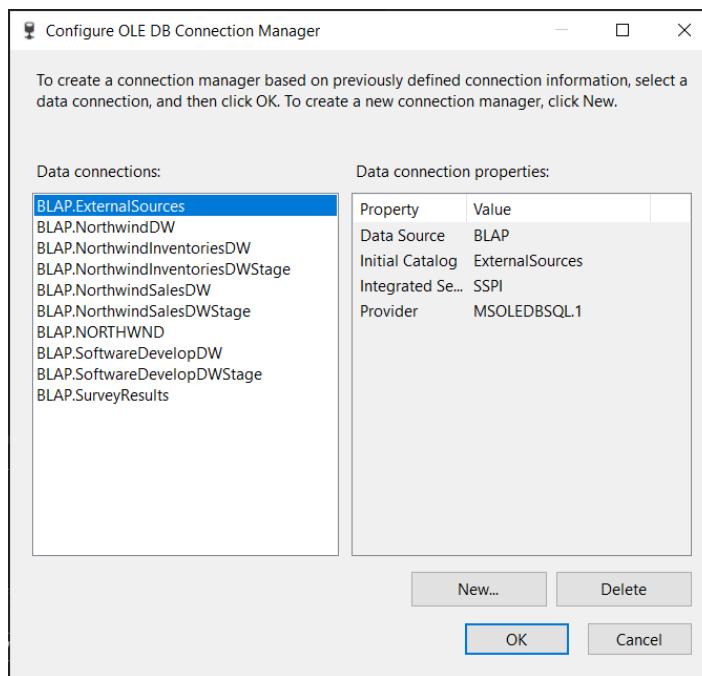
1. Tại Solution Explorer, right-click vào Connection Managers, chọn New Connection Manager.



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



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



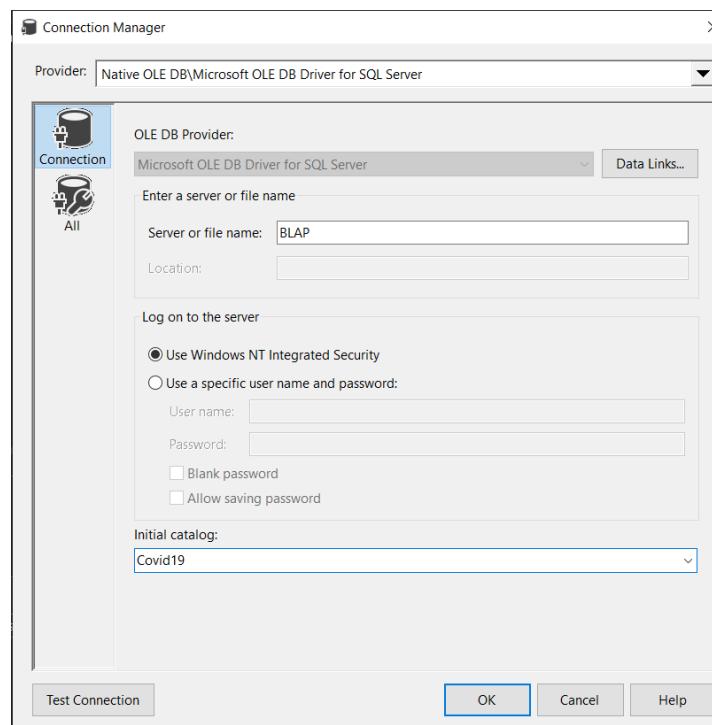
Đặt tên cho các thông tin như bên dưới:

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

Server Name: <Tên Server của bạn>

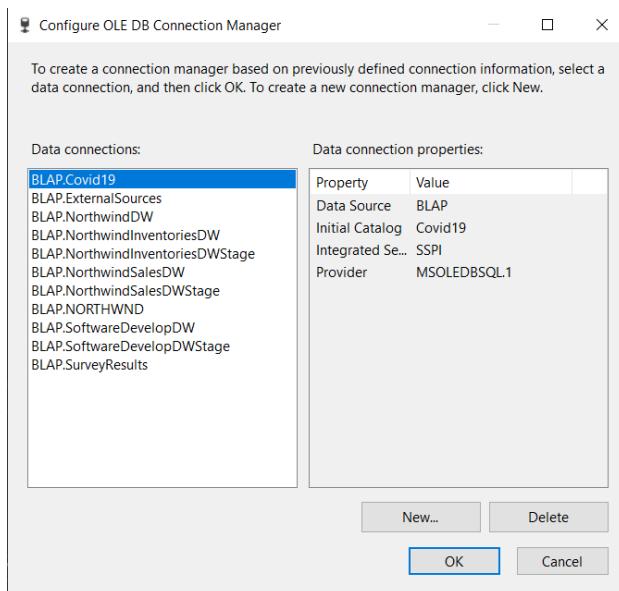
Log on to the Server: **Use Windows Authentication**

Connect to a database: <Tên database của bạn>



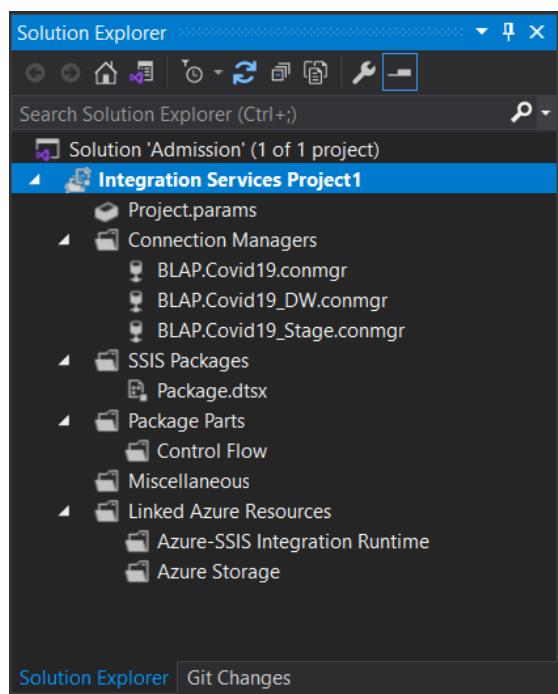
Click **Test Connection** để kiểm tra connection, sau đó click **OK** để lưu connection.

4. Trở lại màn hình **Configure OLE DB Connection Manager**, ta được <Tên Server của bạn>.<Tên database của bạn> vừa được tạo:



5. Chọn <Tên database của bạn> connection và click **OK** để thêm connection này vào project.

6. Lặp lại các bước trên (bước 1 → 5) ba lần cho lần lượt **Covid19_DW** và **Covid19_Stage** source database. Sau khi tạo xong **Connection Managers** folder sẽ như bên dưới.

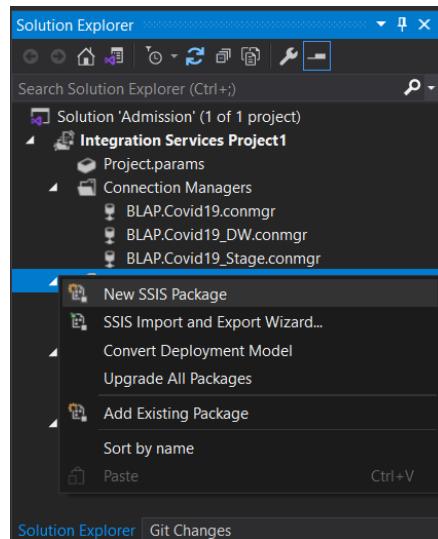


3.2. Import Dữ liệu vào Database gốc

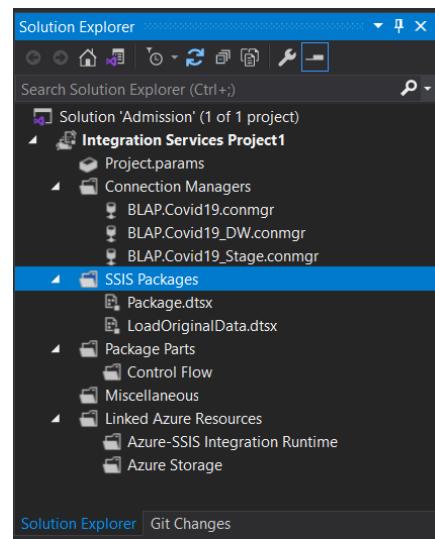
3.2.1. Tạo Package

Tạo một package mới, LoadOrginalData.

- Ở Solution Explorer, right-click vào SSIS Packages và chọn New SSIS Package.



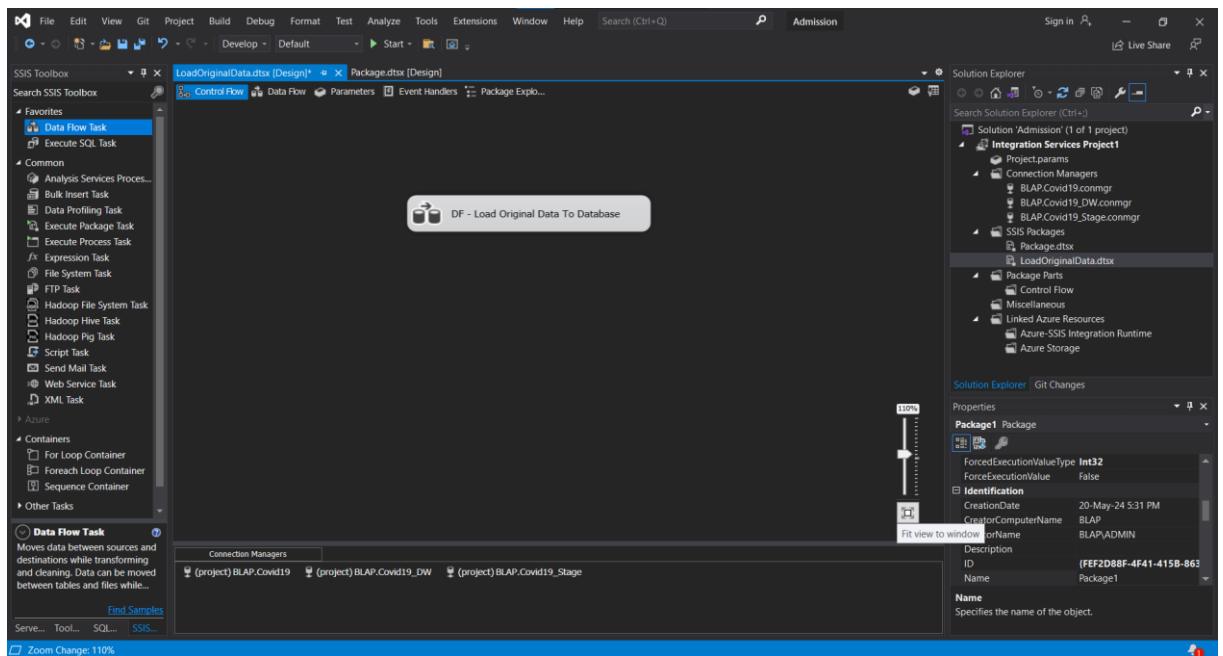
- Right-Click vào package có tên **Package1.dtsx** và chọn **Rename**.
- Đặt tên package là **LoadOrginalData**.
- Sau khi thực hiện xong ta được **Package.dtsx** và **LoadOrginalData.dtsx** như hình sau:



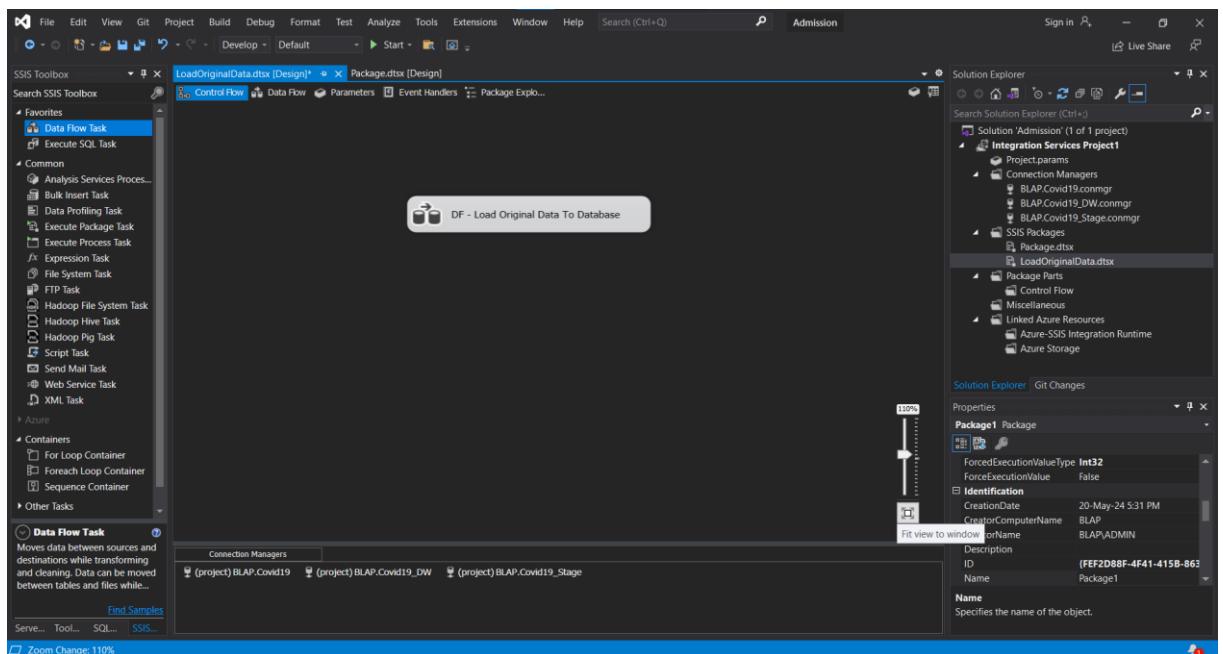
- Double-click vào **LoadOrginalData.dtsx** package để mở nó.

3.2.2. Thiết lập Control Flow

1. Từ SSIS Toolbox tạo 1 Data flow tasks vào surface design như hình sau:

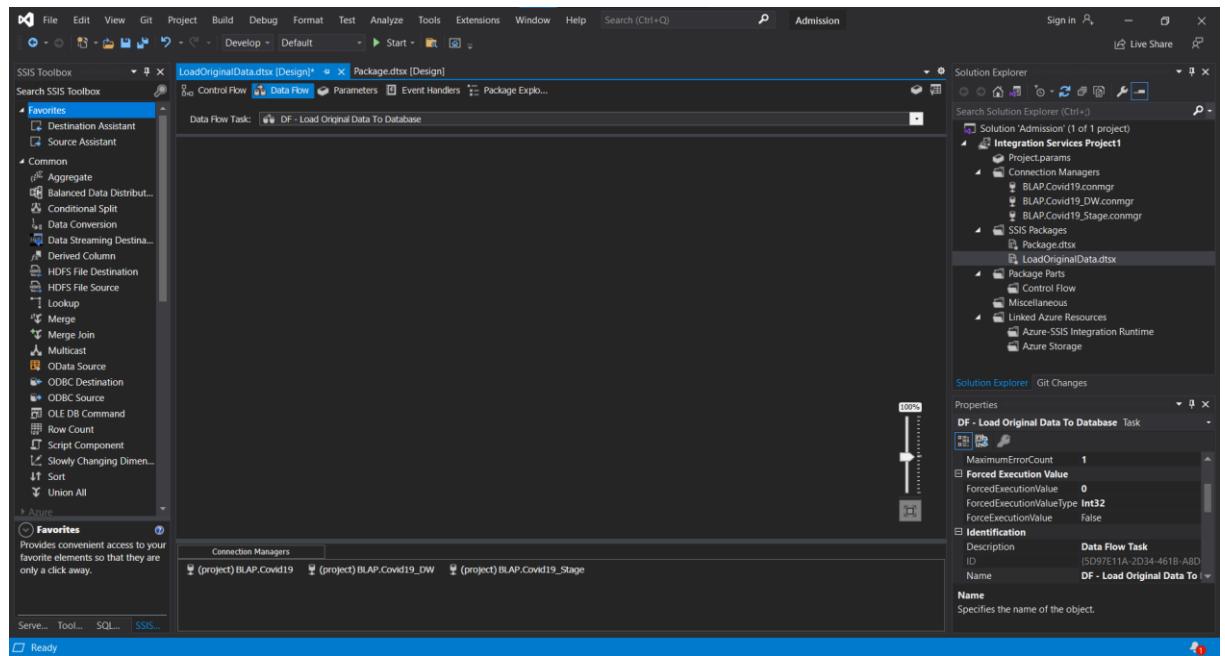


2. Đặt tên cho task Data Flow Task là DF - Load Original Data To Database.

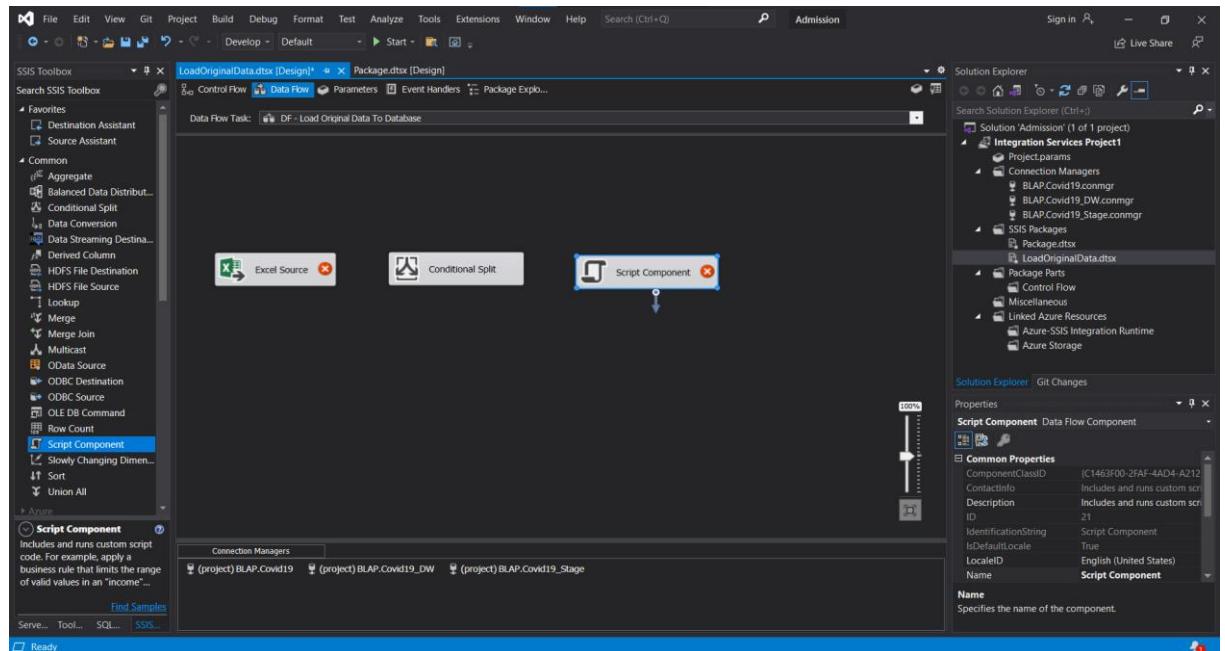


3.2.3. Load dữ liệu từ Source File vào Database

1. Double-click DF - Load Original Data To Database task để mở nó trong Data Flow design surface:

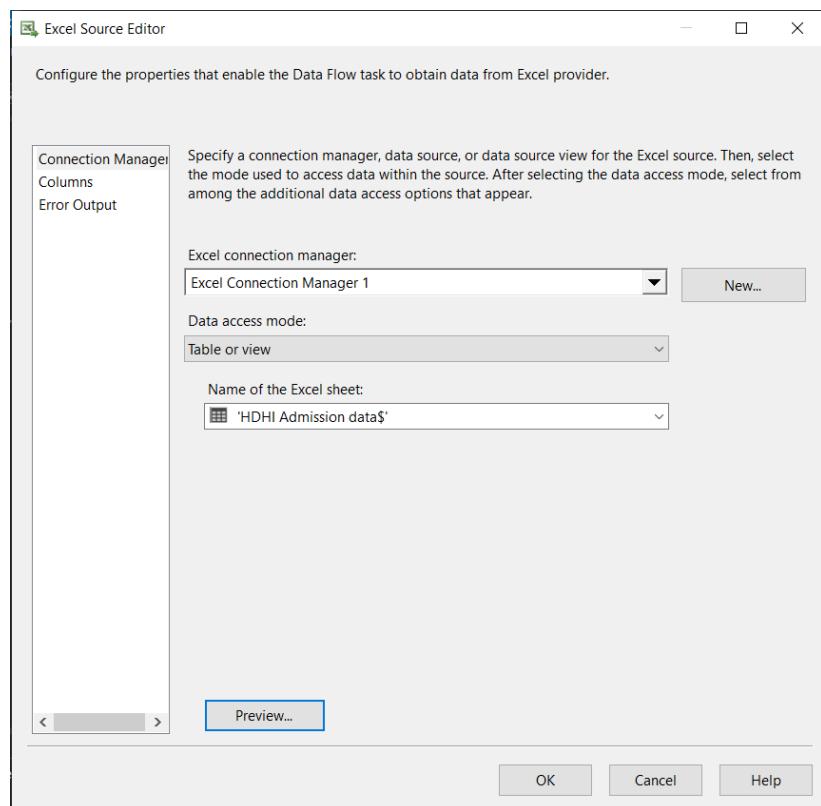


2. Từ SSIS Toolbox, kéo rê Excel Source, Conditional Split, Script Component và Destination Assistant vào design surface. Khi ta kéo Script Component sẽ xuất hiện 1 hộp thoại, vui lòng chọn Transformation. Design Surface sẽ trông giống sau:

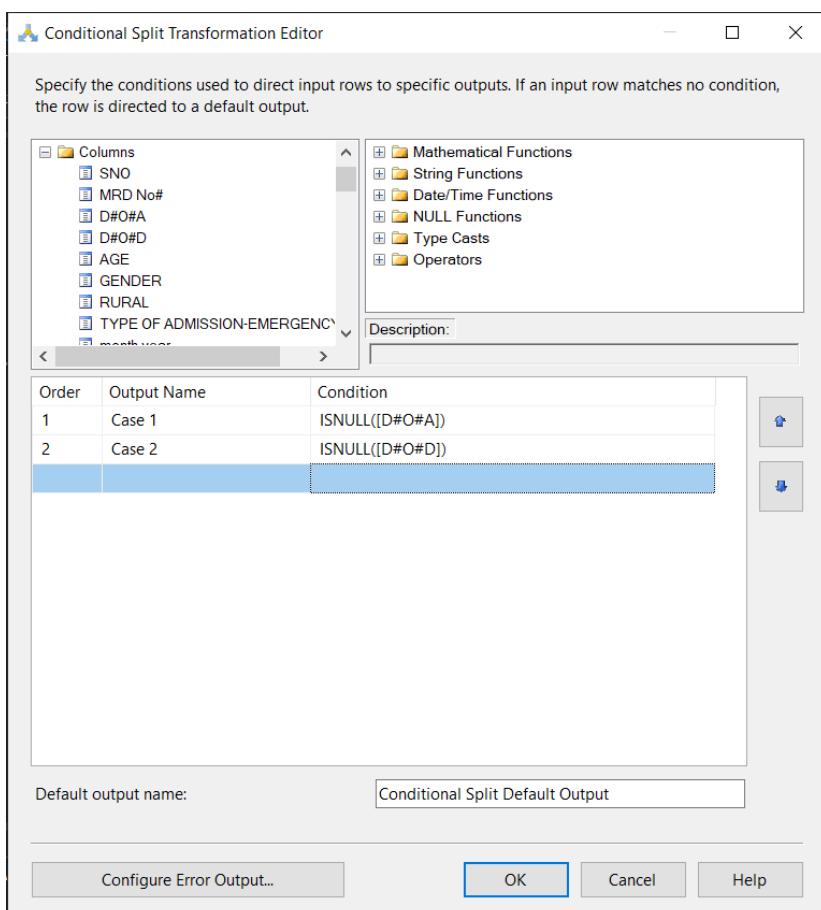


3. Tiến hành nối các task lại với nhau và tiến hành cấu hình:

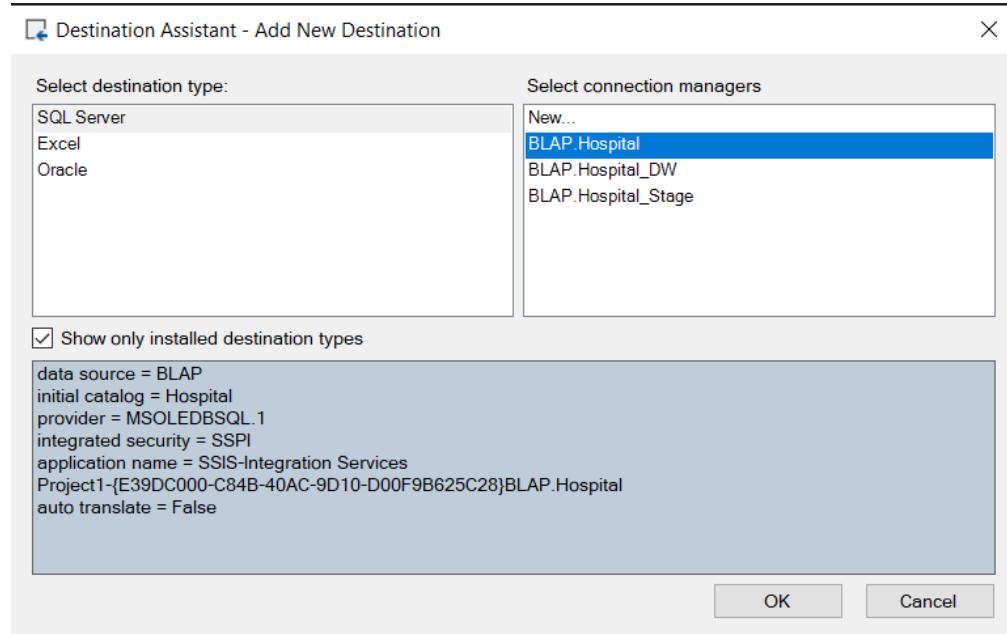
- Cấu hình Excel Source:



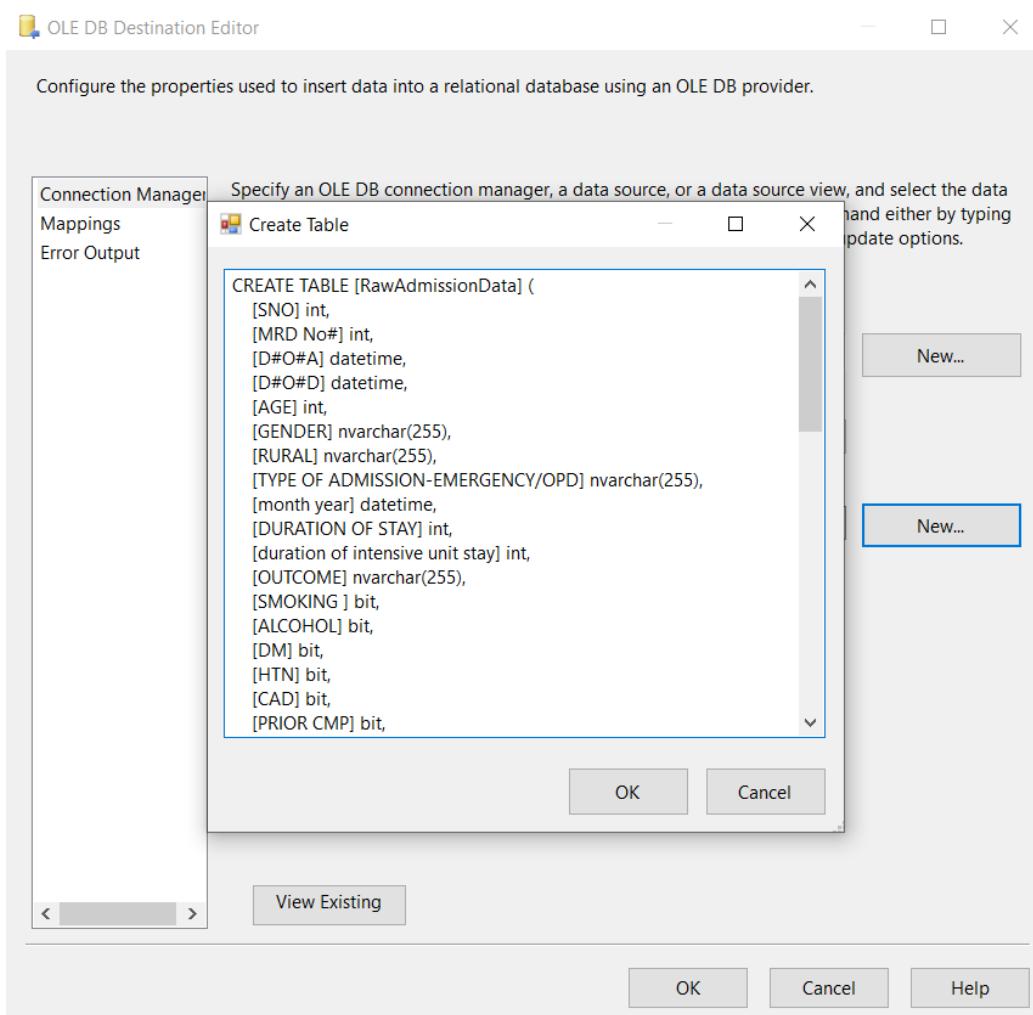
- Cấu hình Conditional Split:

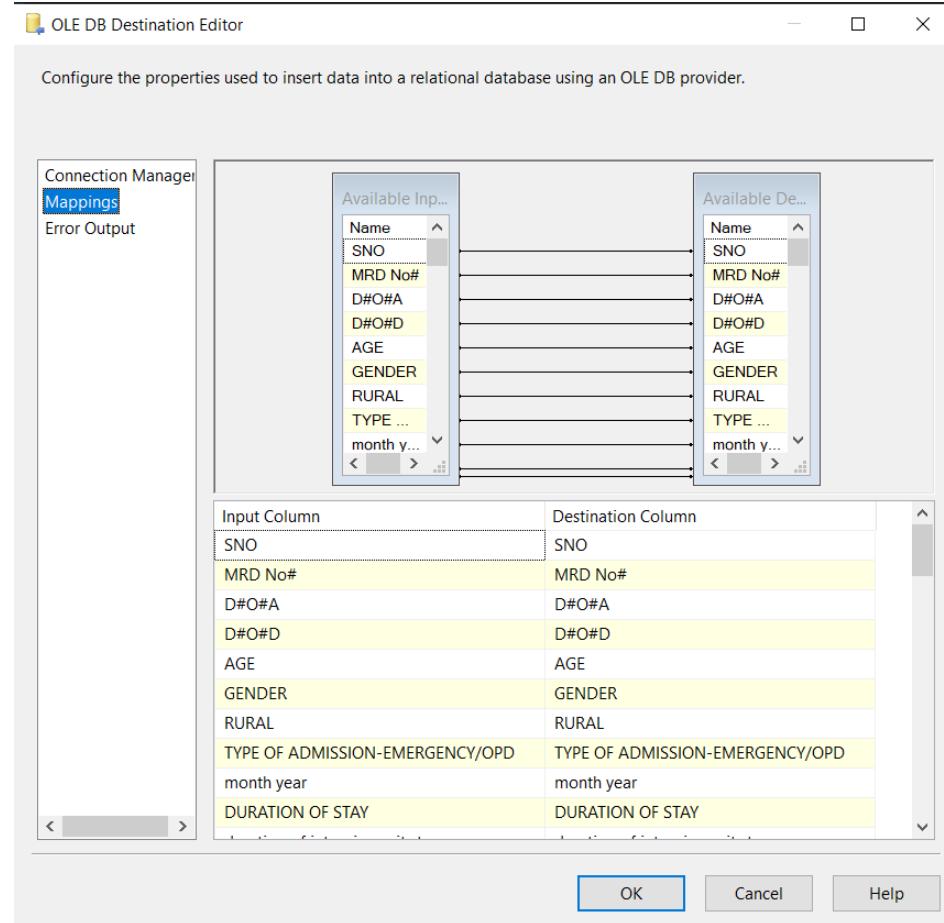
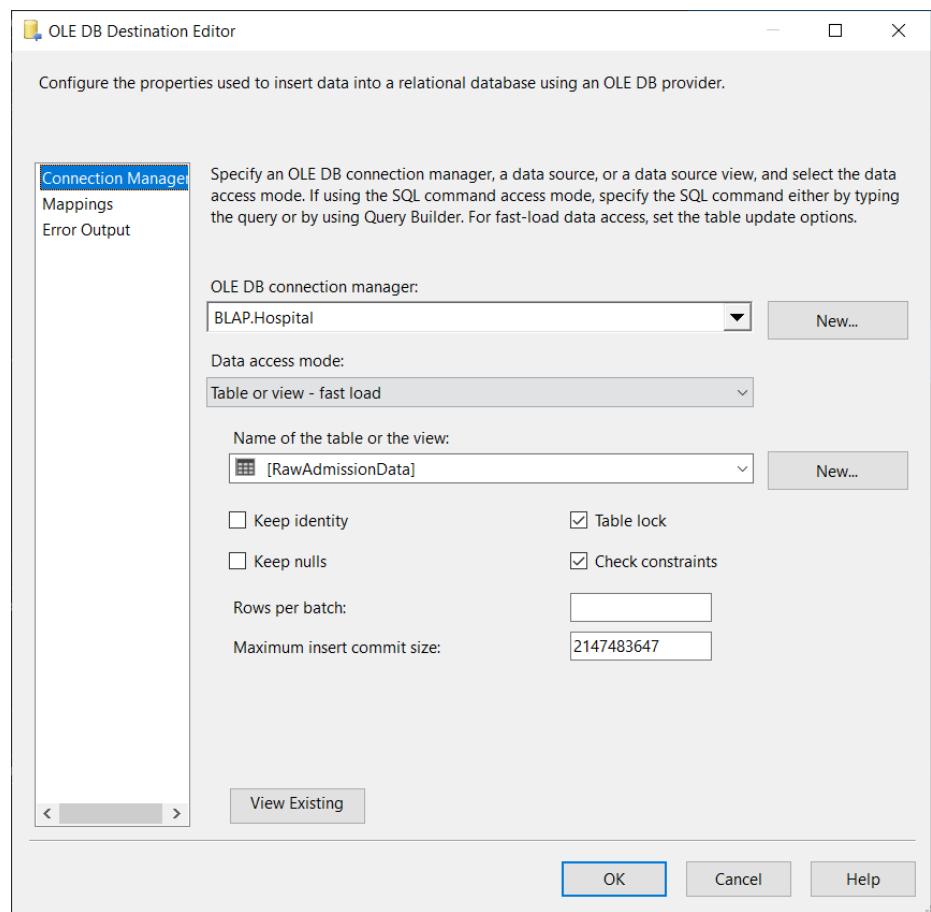


4. Kéo rê **Destination Assistant** vào Design Surface sẽ xuất hiện bảng sau:

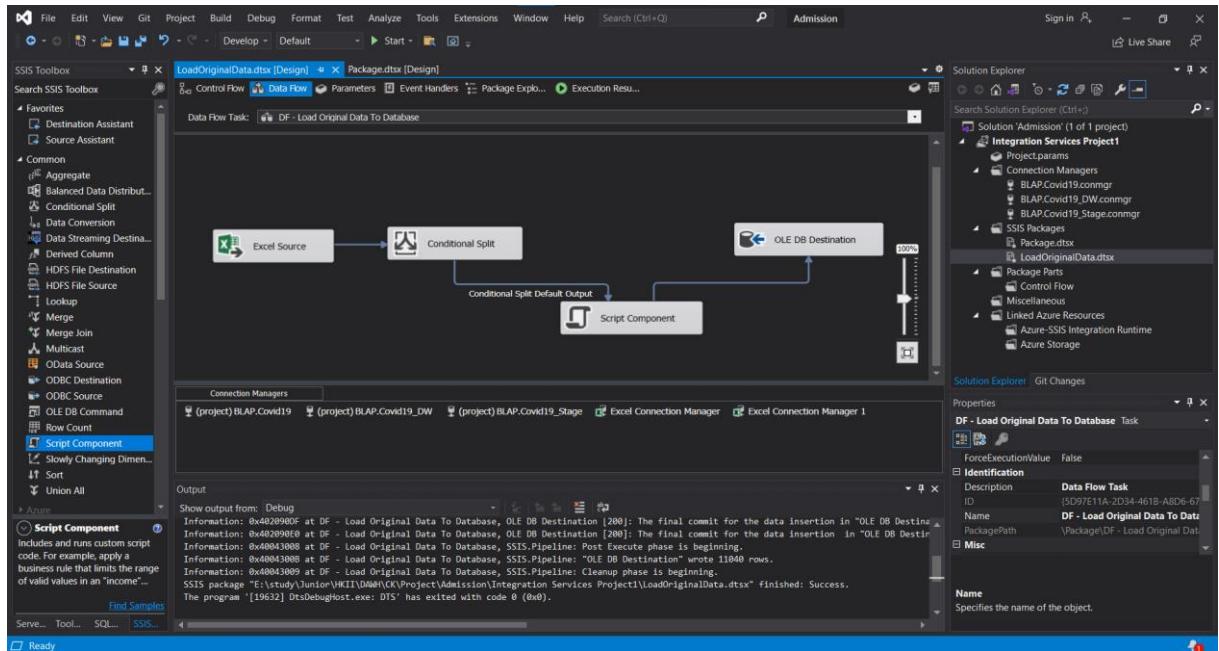


5. Cấu hình **Destination Assistant**. Click nút **New...** để tạo bảng trong CSDL stage, ta sẽ thấy window **Create Table** như sau: (Đổi tên bảng lại thành **RawAdmissionData**).





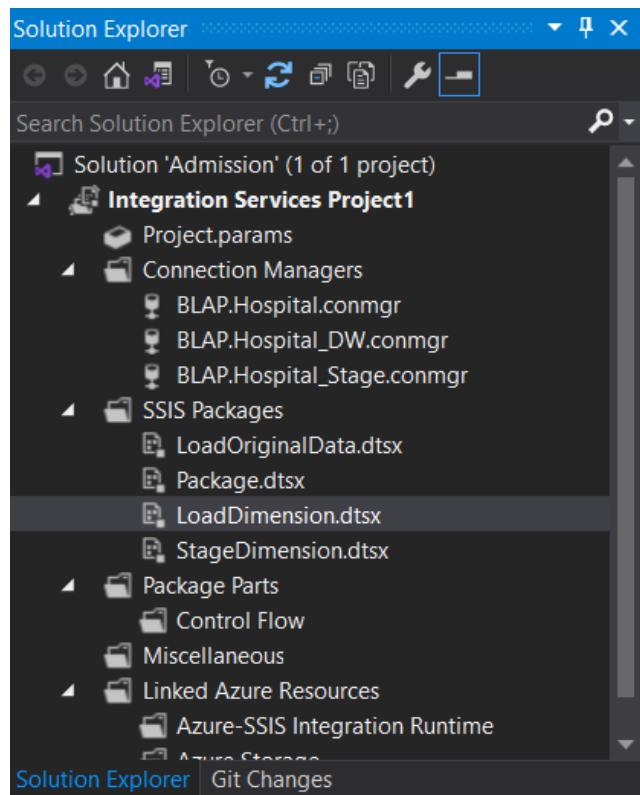
Click **OK** để hoàn tất việc cấu hình. Màn hình data flow design surface sẽ như bên dưới.



3.3. Import Dữ liệu vào các bảng Dimension

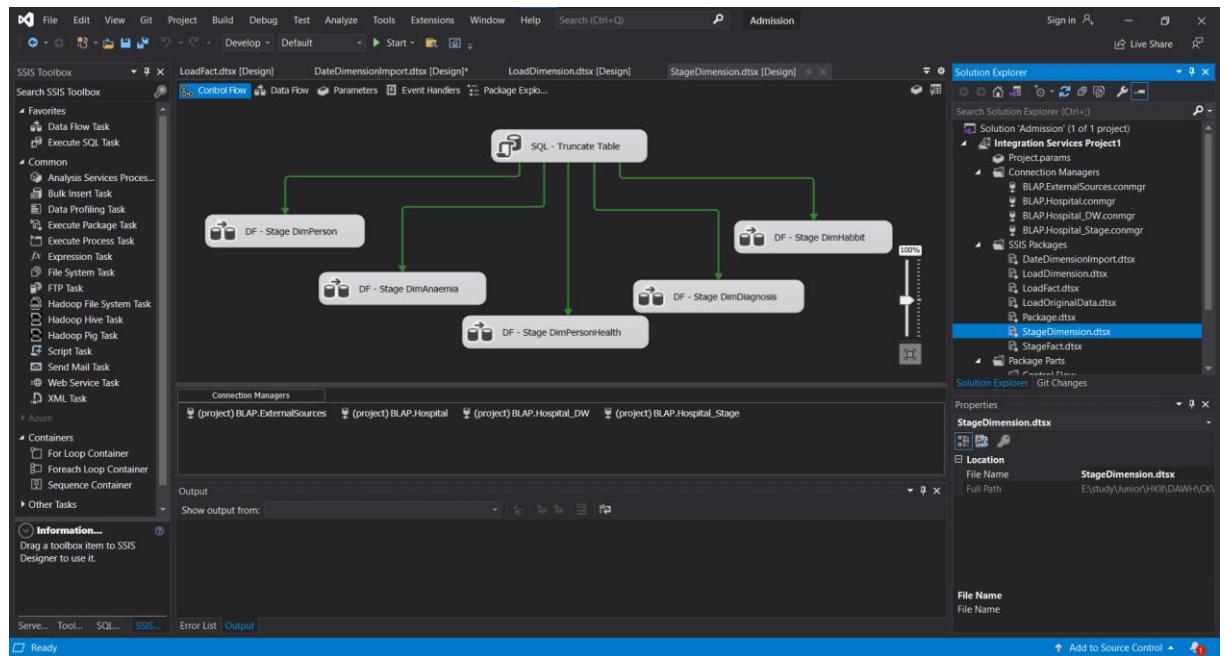
3.3.1. Tạo package

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

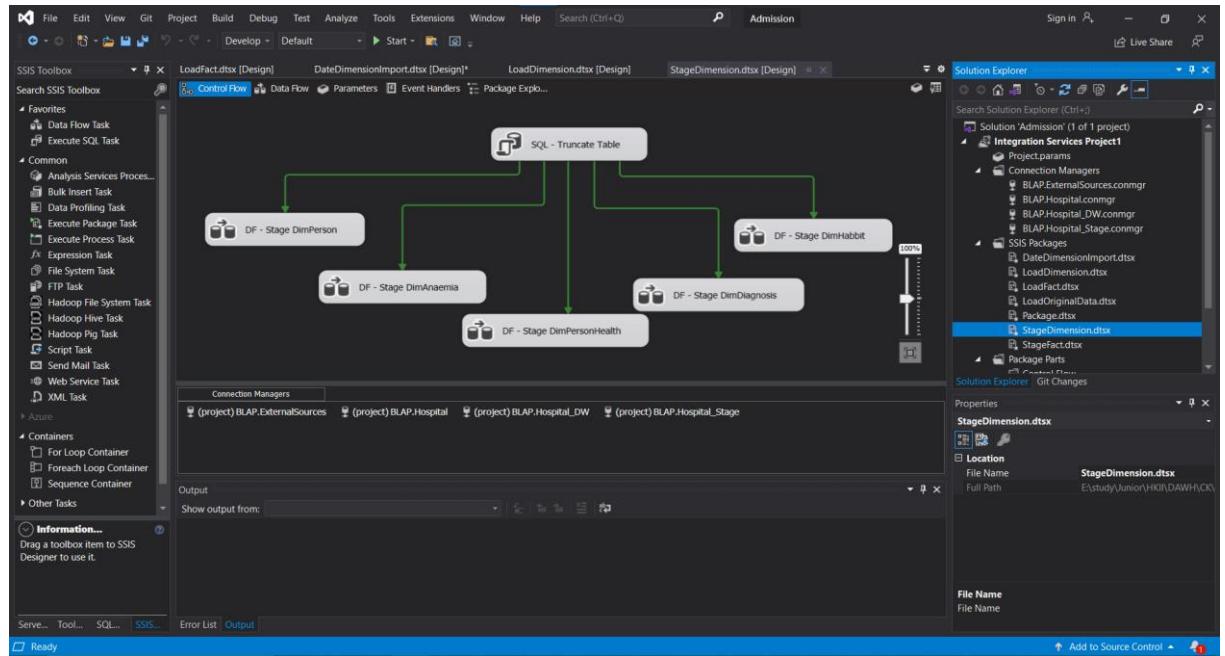


3.3.2. Stage các Dimension

1. Click vào StageDimension.dtsx và tiến hành thêm vào 1 Execution SQL Task và 6 Data Flow Task như hình sau:



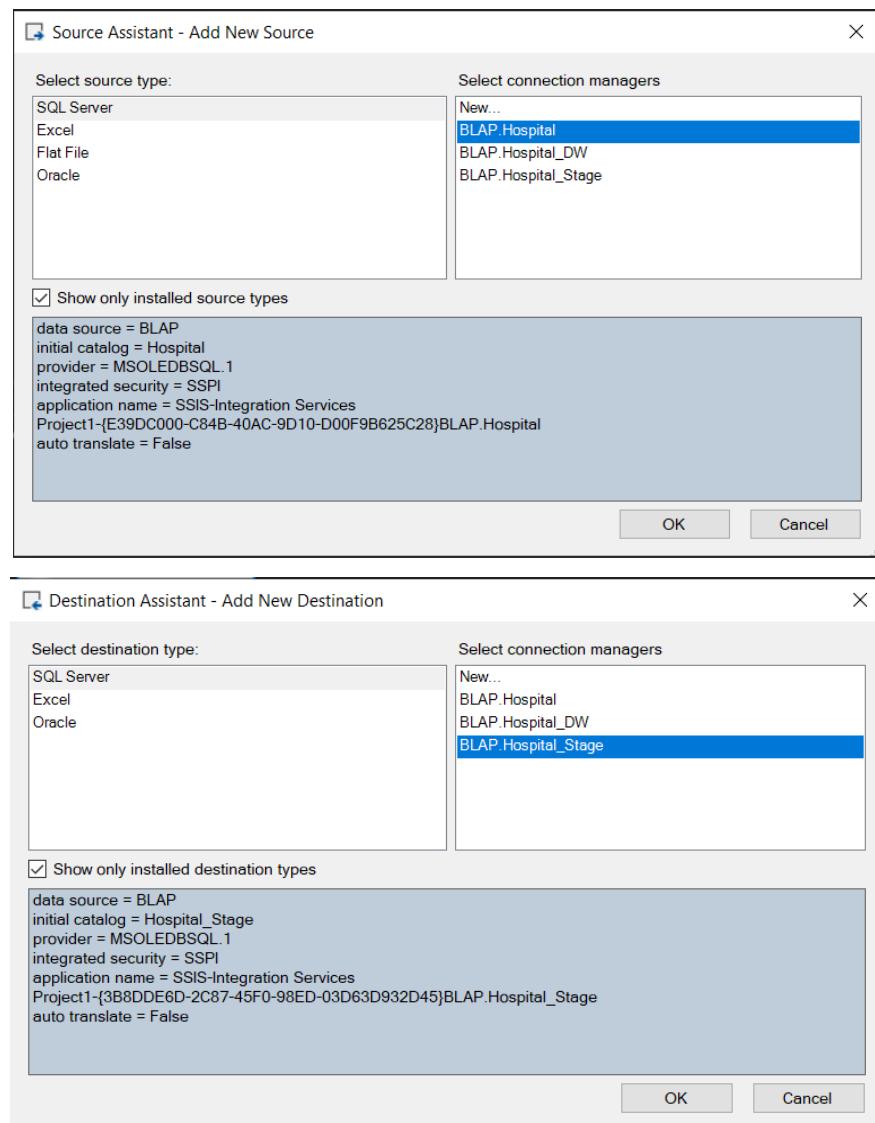
2. Đổi tên các task như sau:



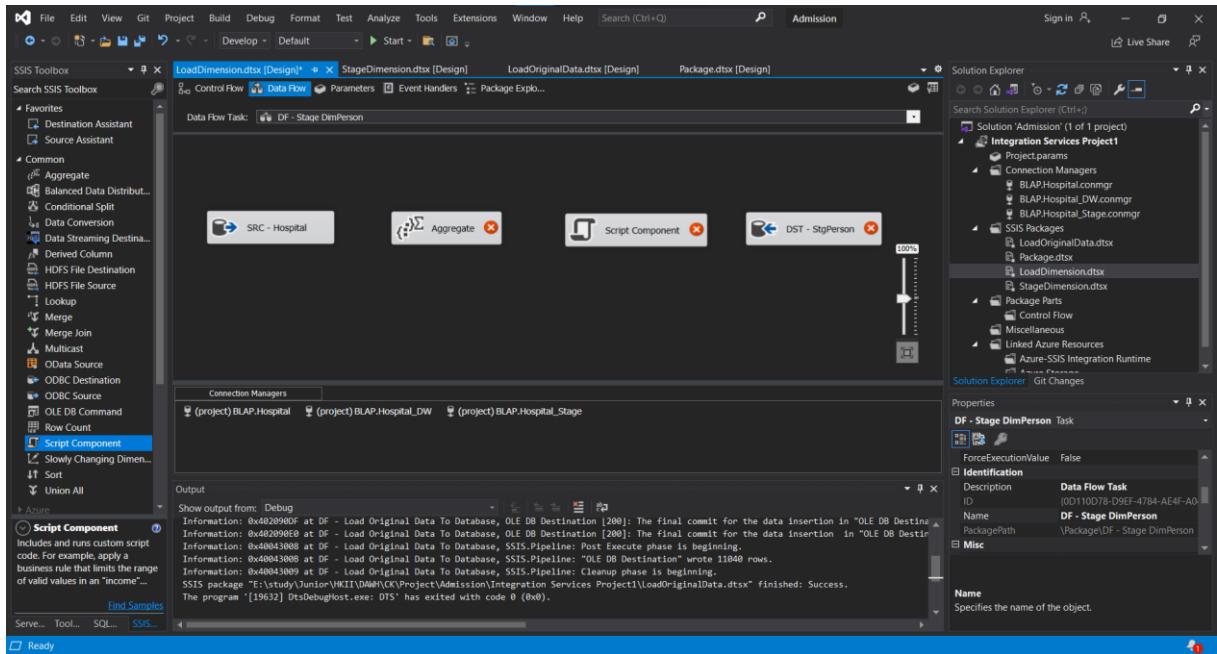
3. Tiến hành stage các **Dimension** sau:

3.3.2.1. StgPerson

- Click vào **DF - Stage DimPerson** và thêm **Source Assistant, Aggregate, Script Component** và **Destination Assistant**

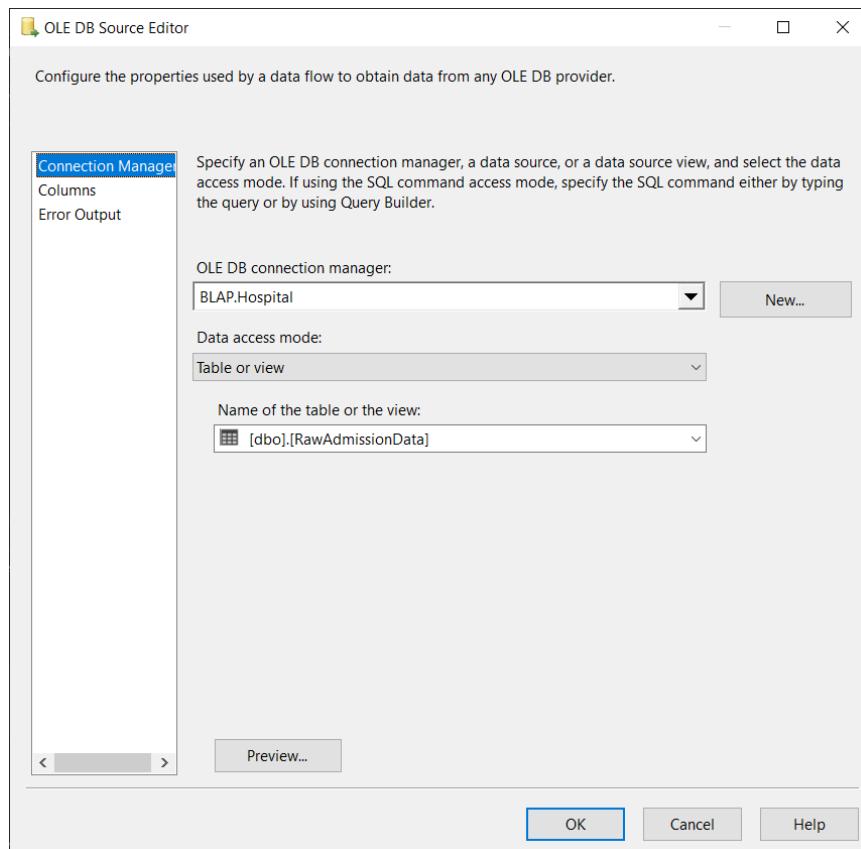


2. Tiến hành đổi tên và ta có được hình như sau:

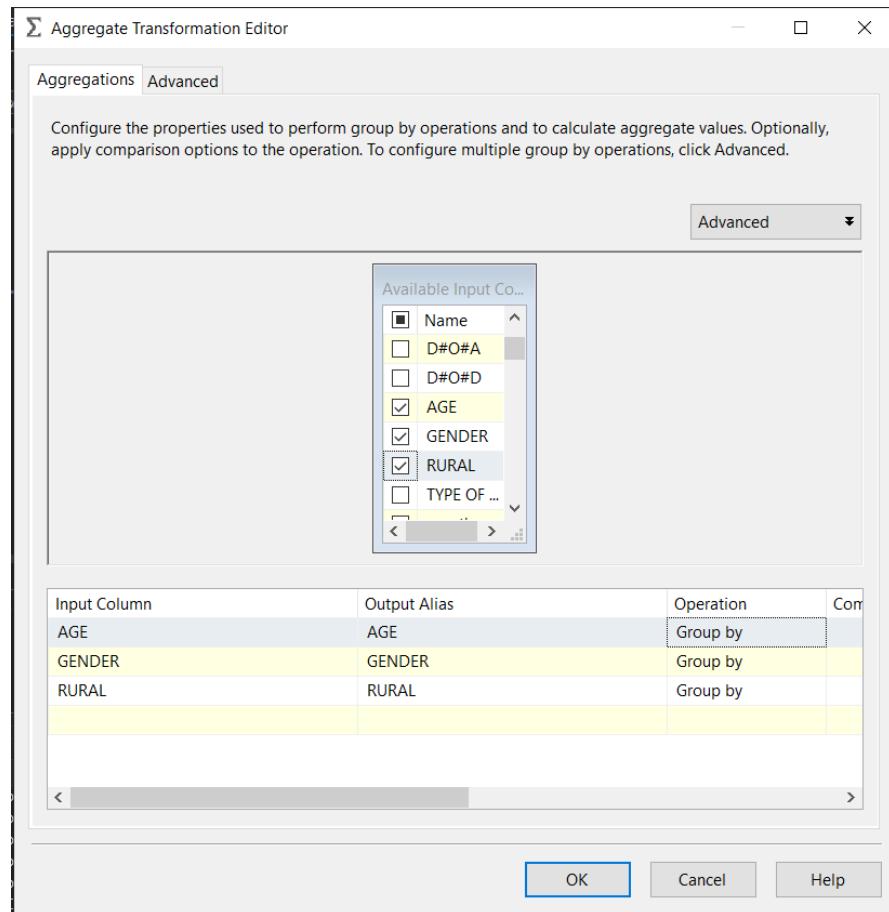


3. Tiến hành cấu hình các task:

- Cấu hình Source Assistant:



- Cấu hình Aggregate:

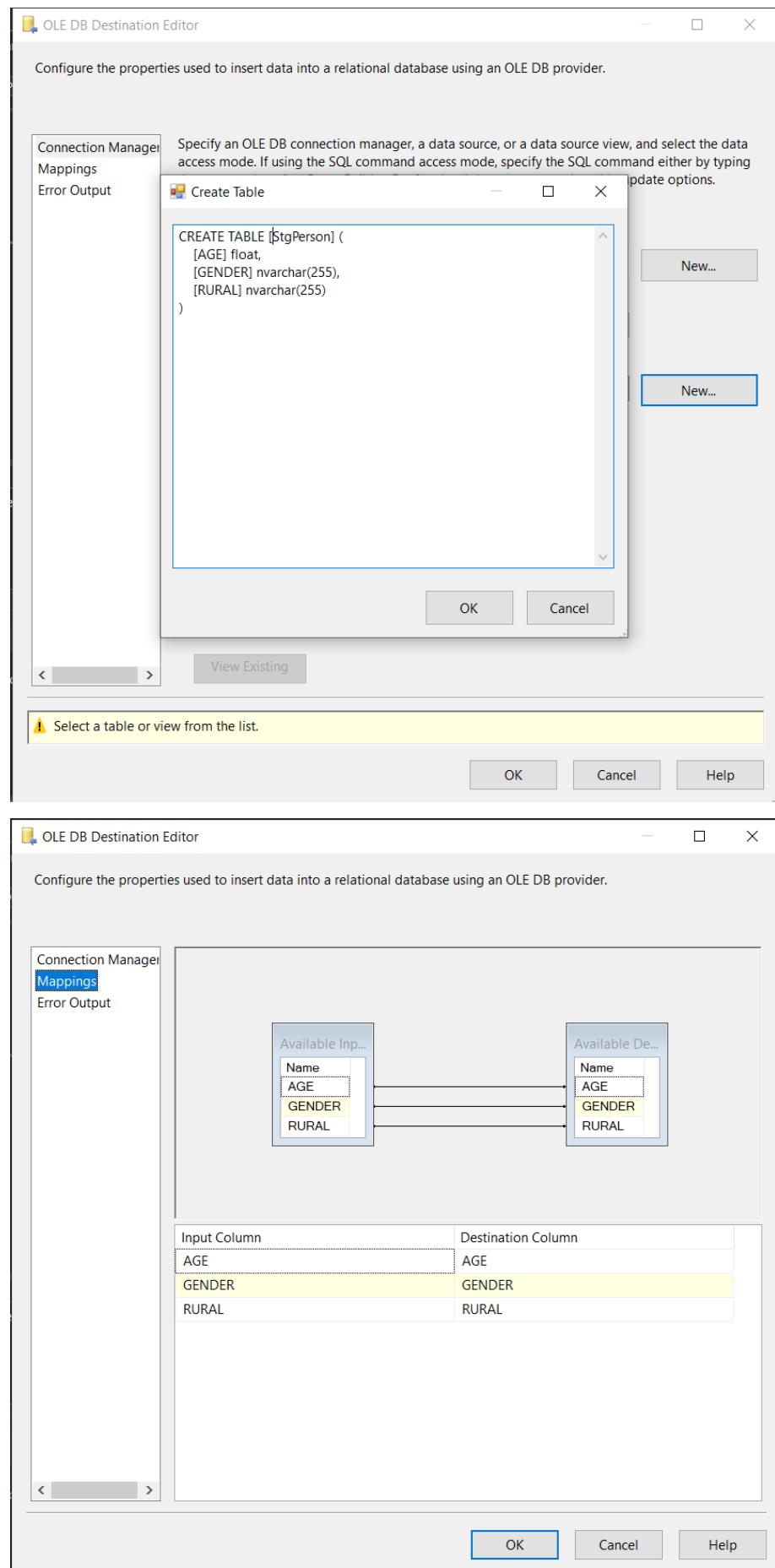


- Cấu hình Script Component:

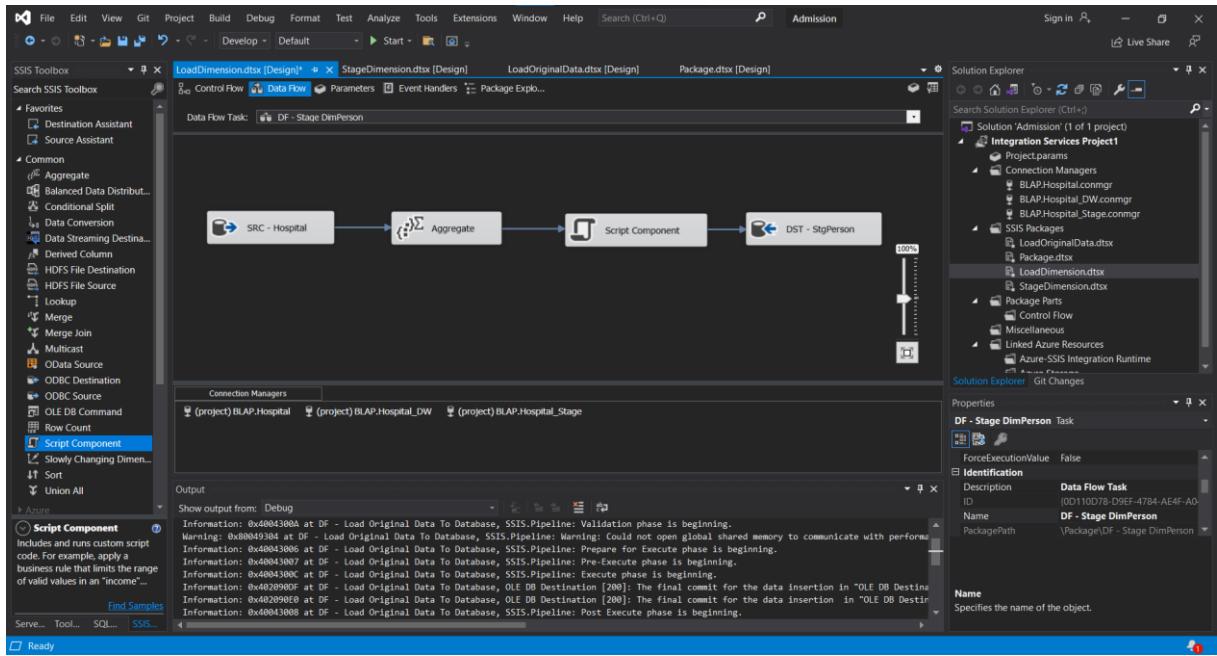
```
1 reference
public Int32 PersonID
{
    set
    {
        this[0] = value;
    }
}

public override void Input0_ProcessInputRow(Input0Buffer Row)
{
    /*
     * Add your code here
     */
    count++;
    Row.PersonID = count;
}
```

- Cấu hình Destination Assistant:

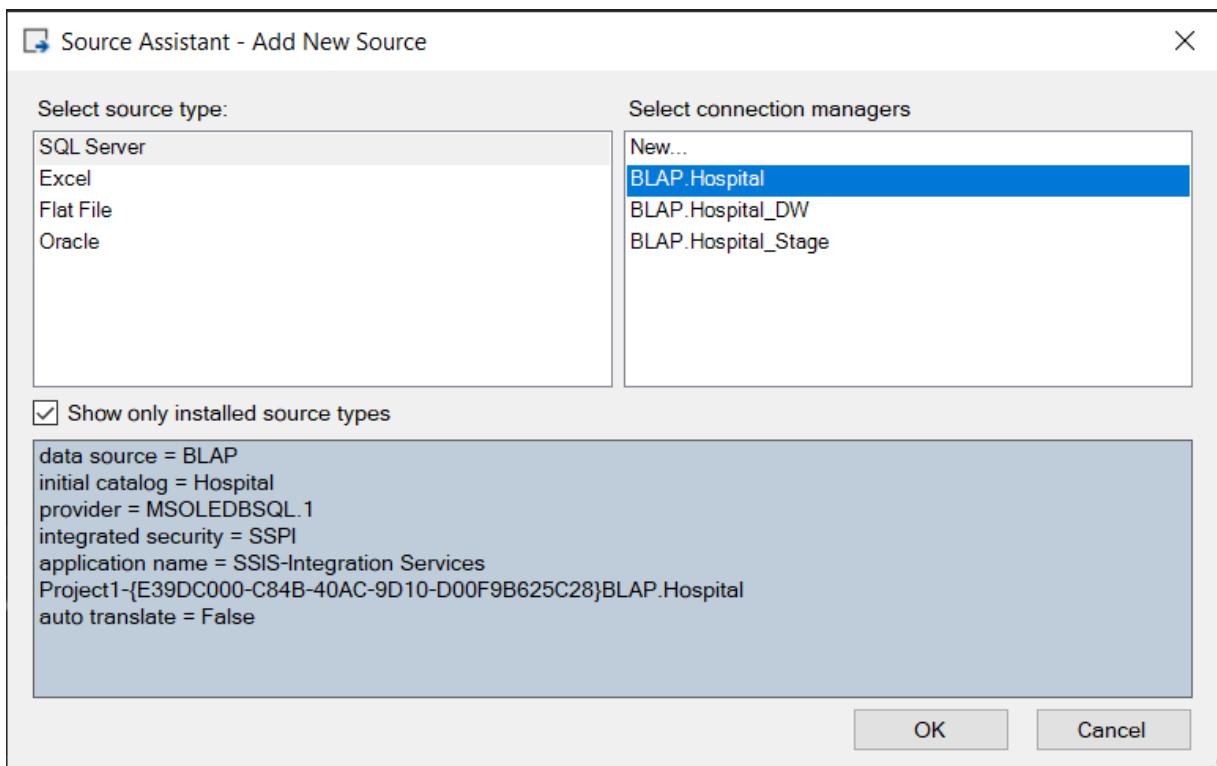


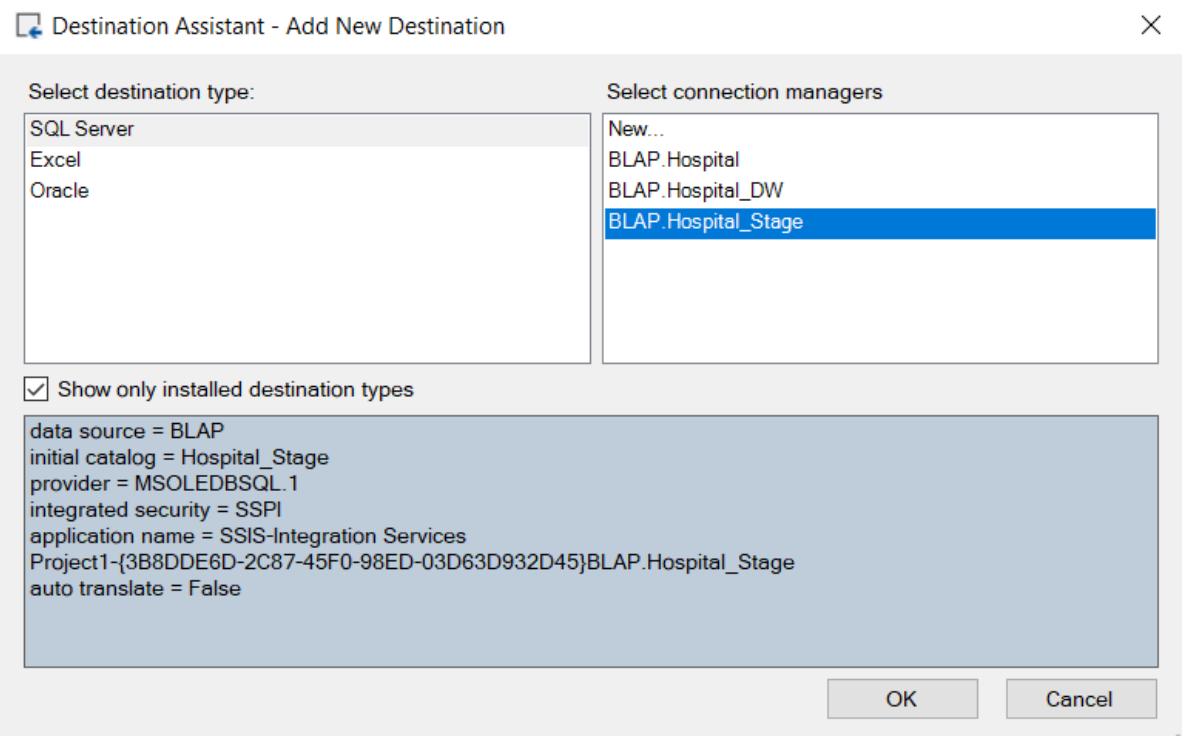
4. Data Flow Task của ta sẽ trông như sau:



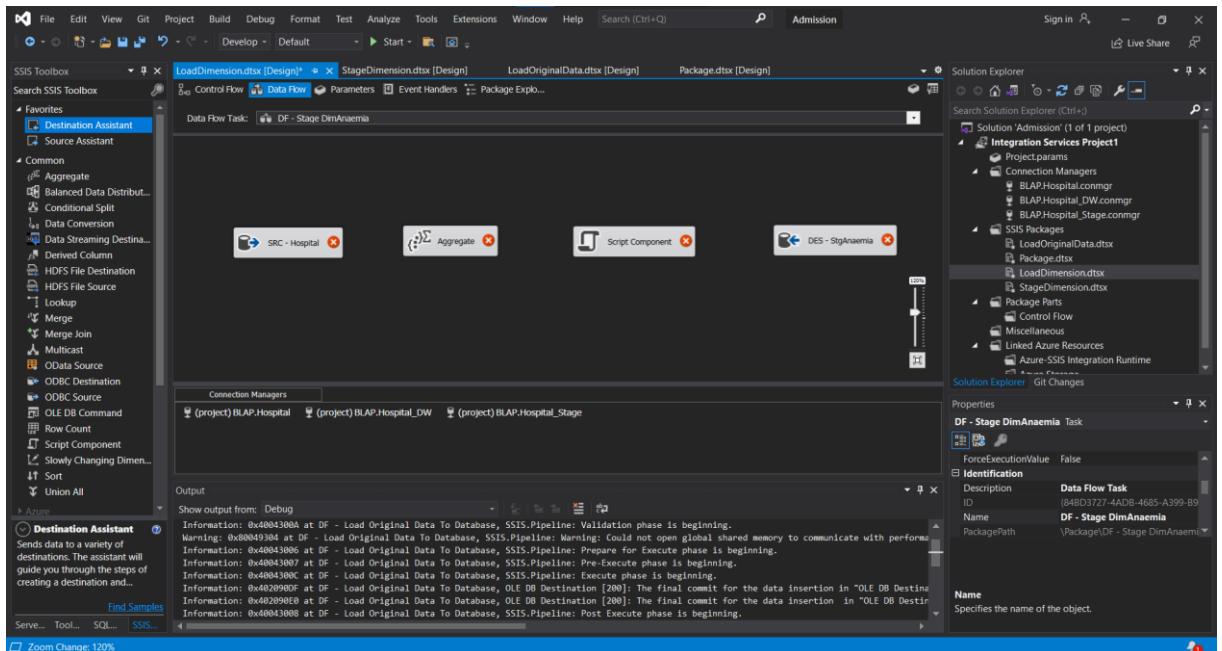
3.3.2.2. StgAnaemia

1. Click vào DF - Stage DimAnaemia và thêm Source Assistant, Aggregate, Script Component và Destination Assistant



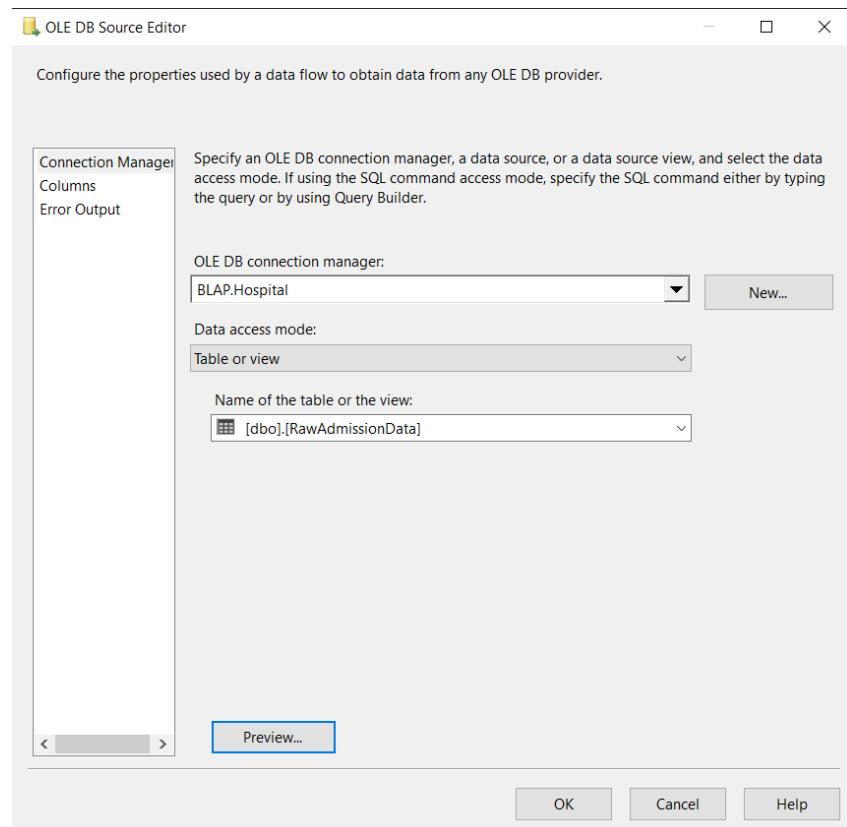


2. Tiến hành đổi tên và ta có được hình như sau:

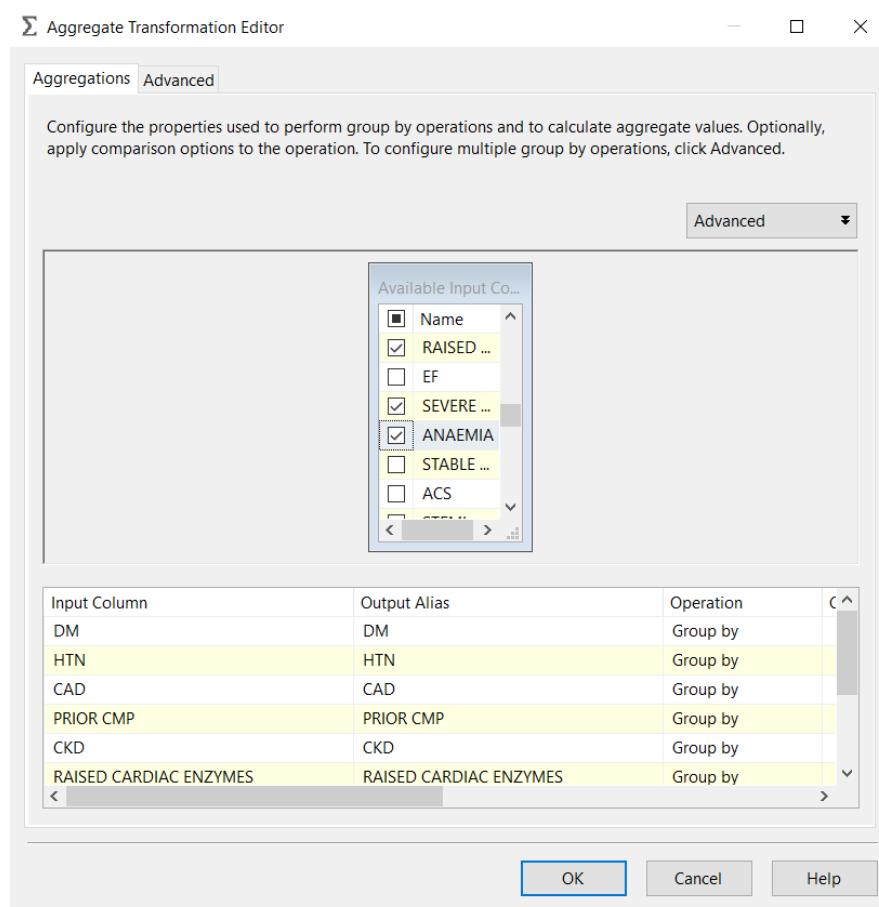


3. Tiến hành cấu hình các task:

- Cấu hình Source Assistant:

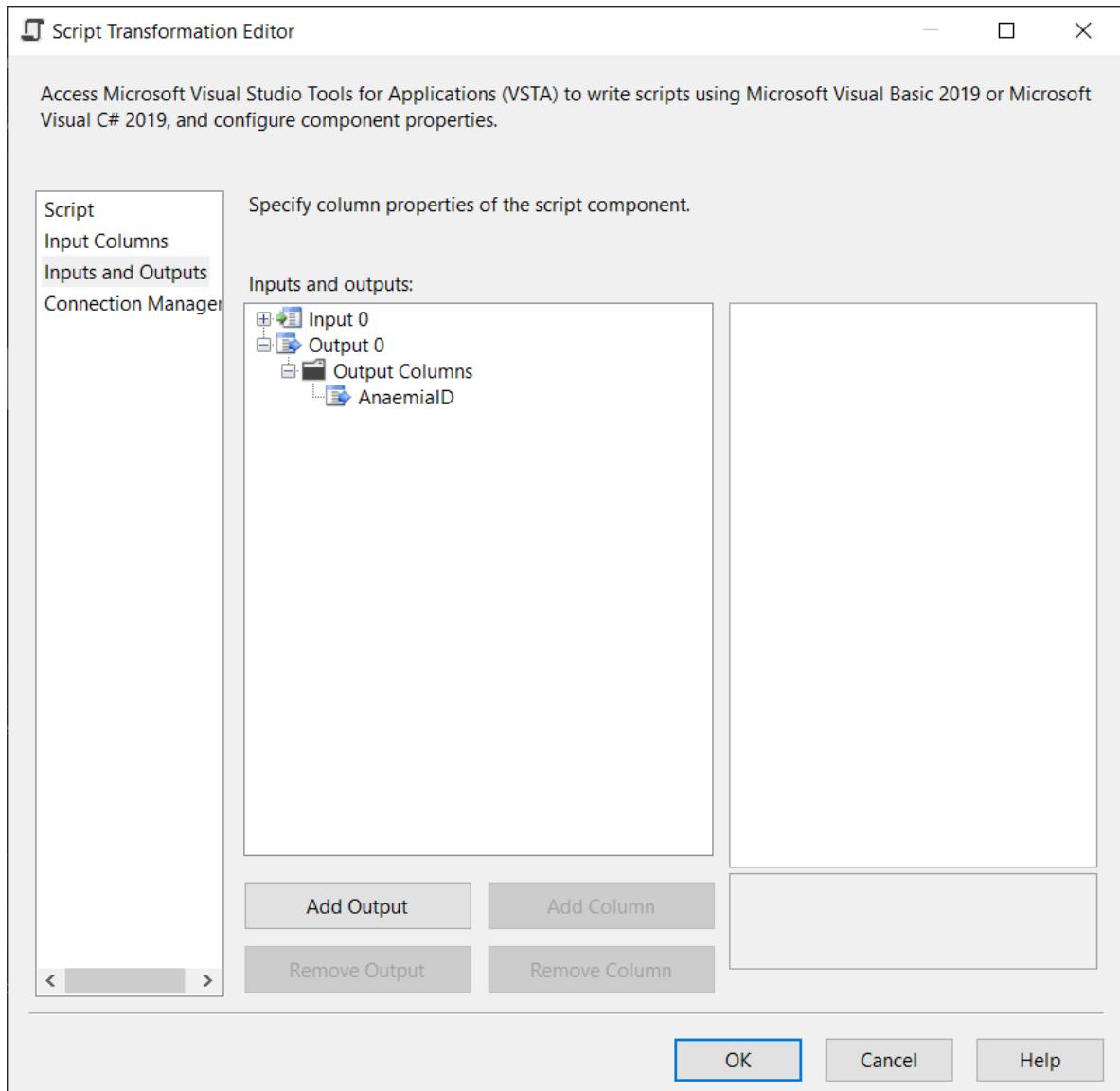


- Cấu hình Aggregate:

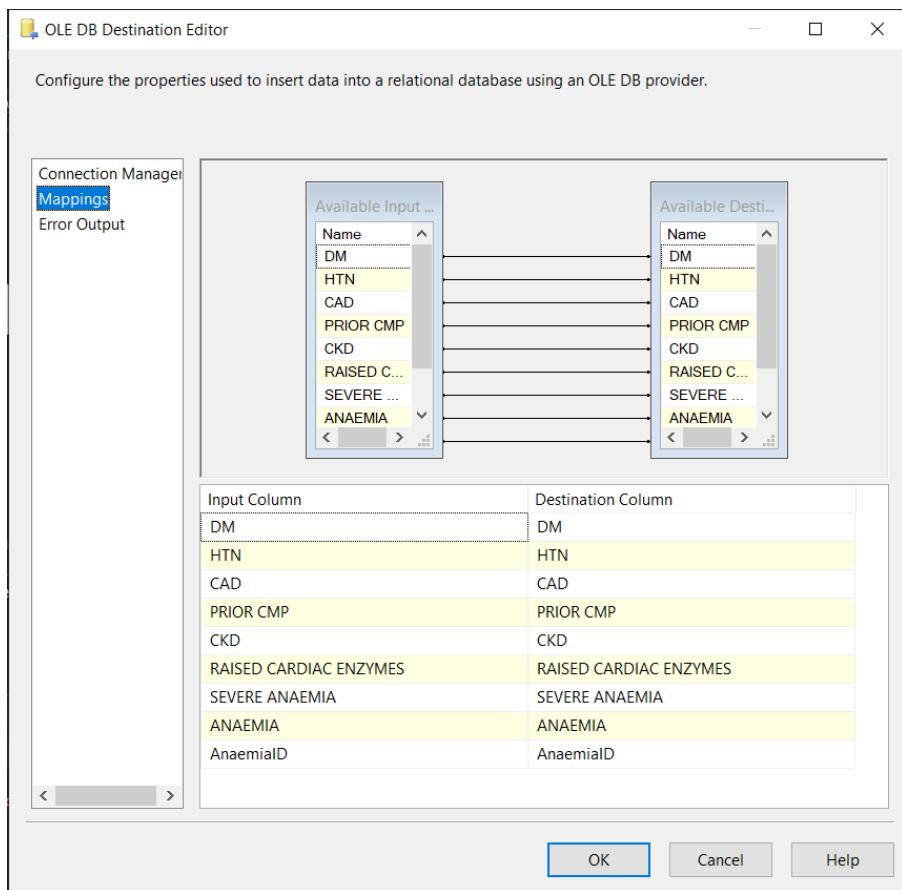
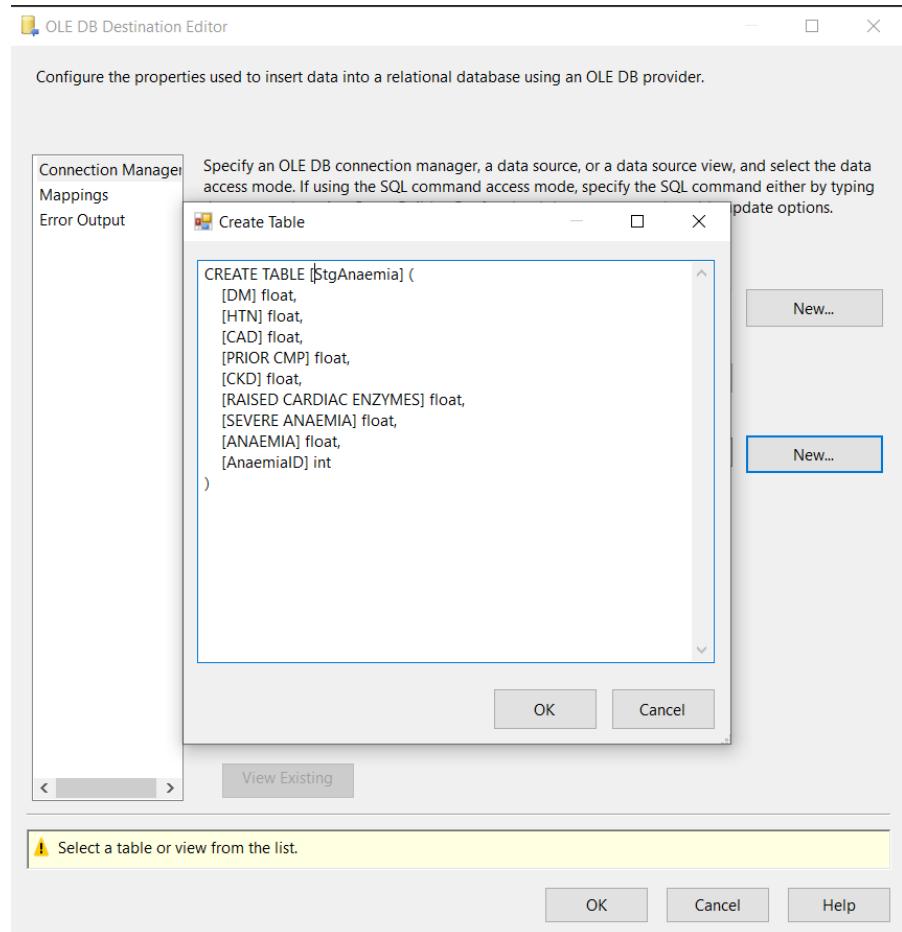


- Cấu hình Script Component:

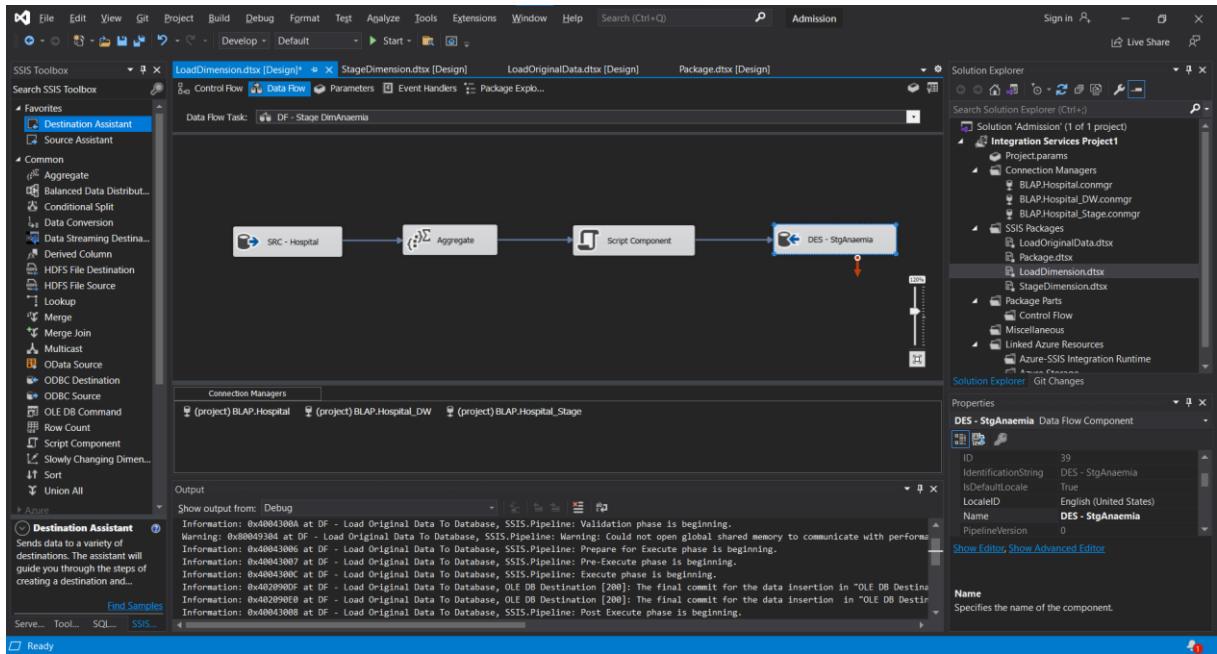
```
public override void Input0_ProcessInputRow(Input0Buffer Row)
{
    /*
     * Add your code here
     */
    count++;
    Row.AnemiaID = count;
}
```



- Cấu hình Destination Assistant:

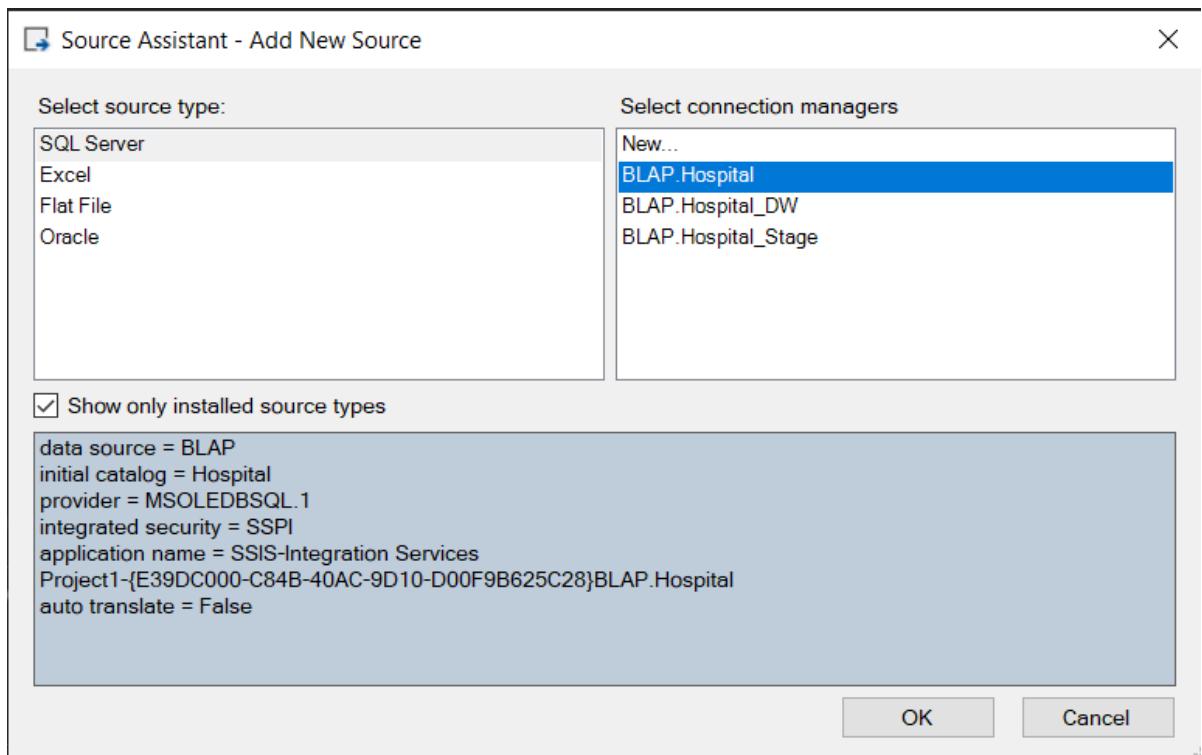


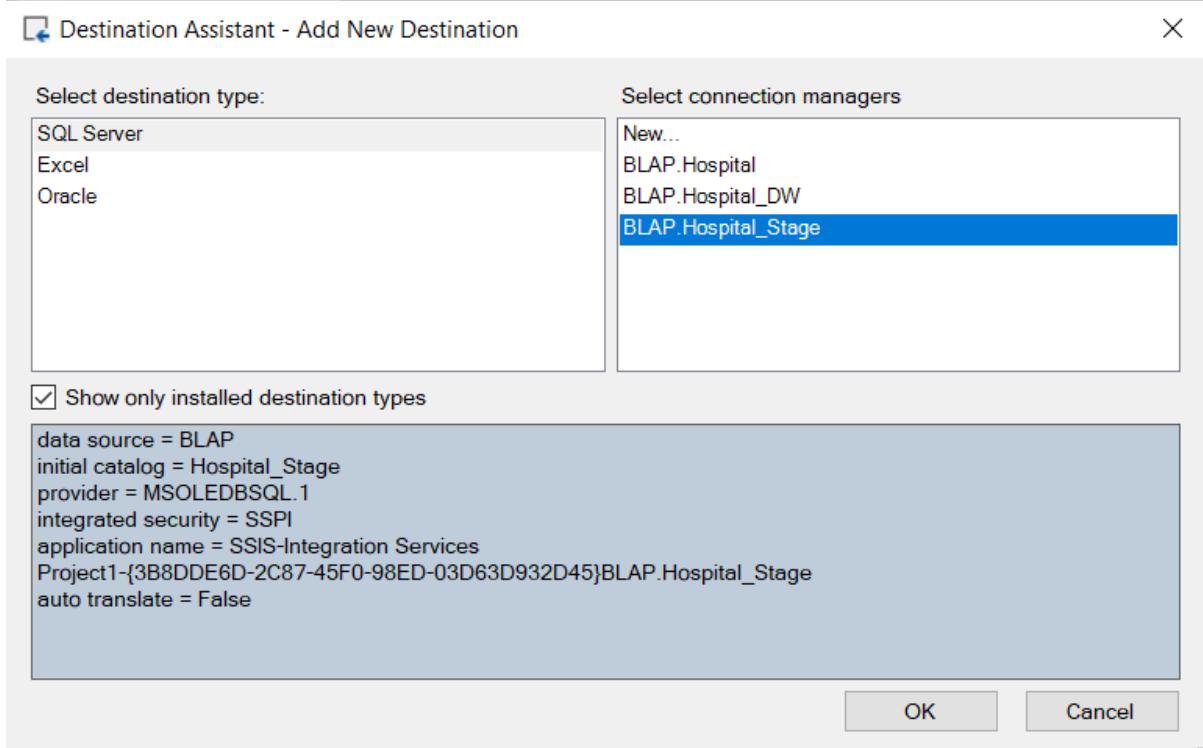
4. Data Flow Task của ta sẽ trông như sau:



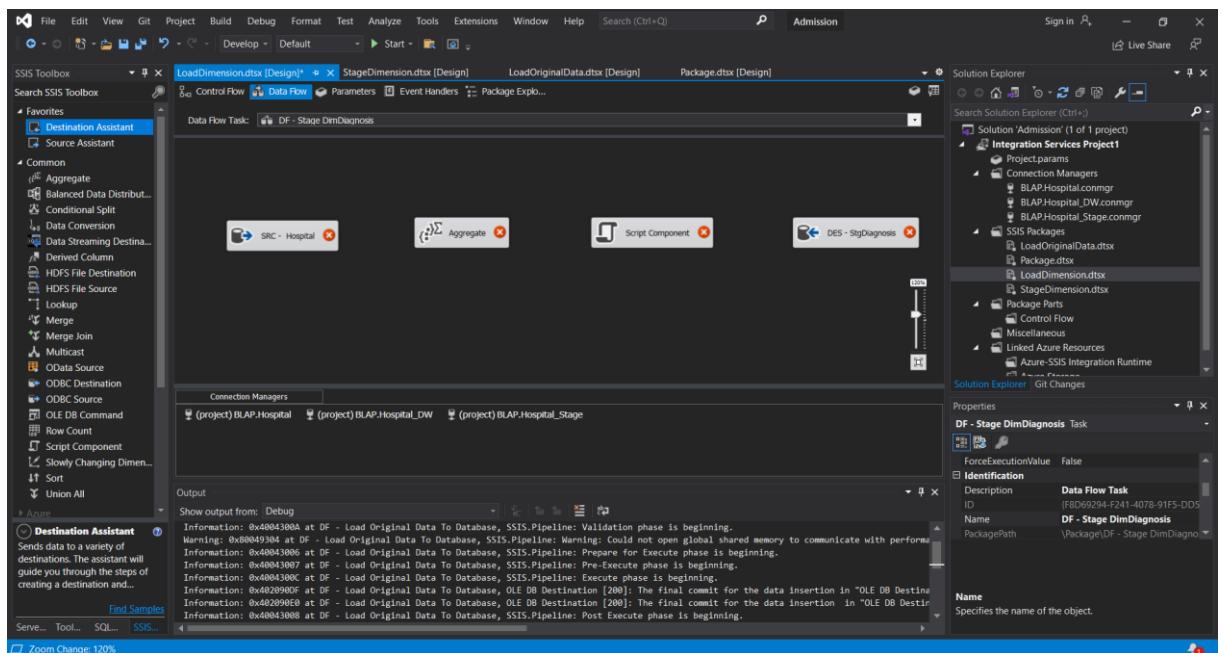
3.3.2.3. StgDiagnosis

1. Click vào DF - Stage DimDiagnosis và thêm Source Assistant, Aggregate, Script Component và Destination Assistant



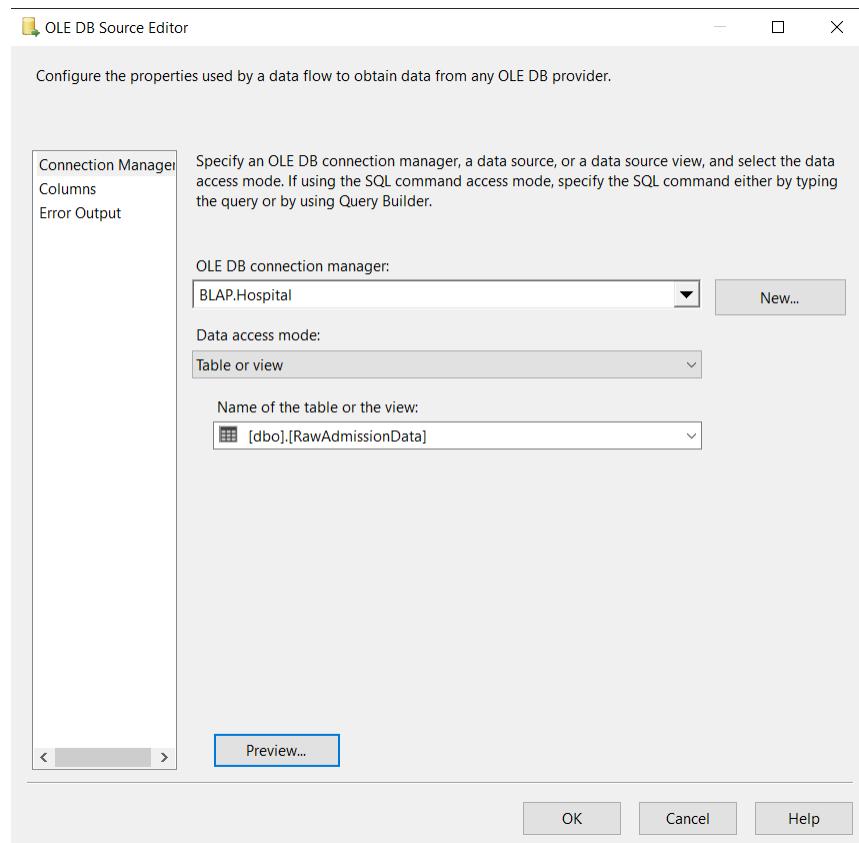


2. Tiến hành đổi tên và ta có được hình như sau:

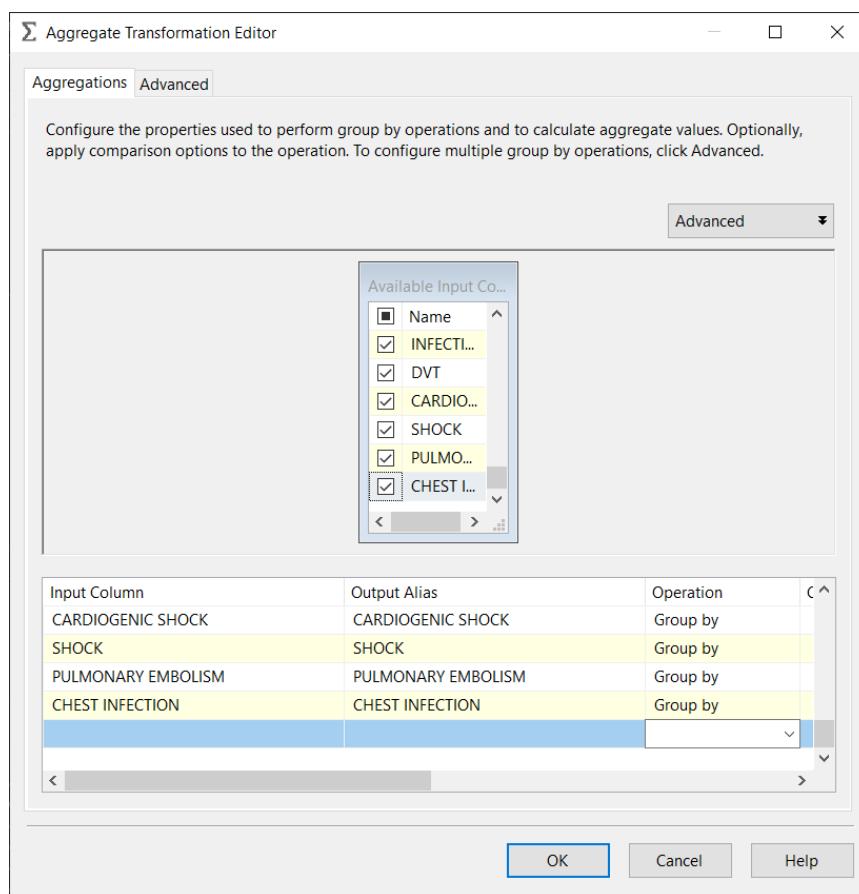


3. Tiến hành cấu hình các task:

- Cấu hình Source Assistant:

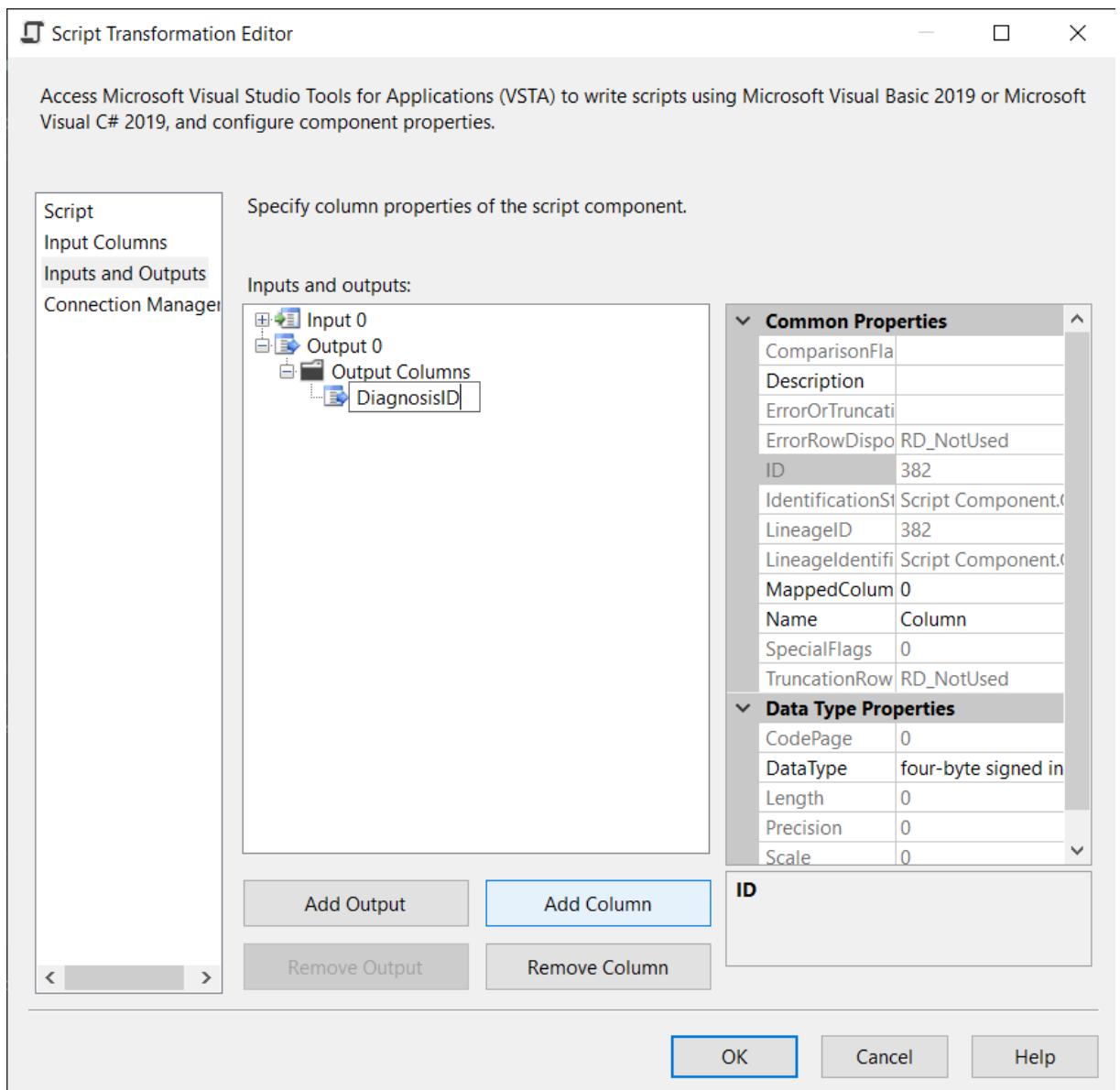


- Cấu hình Aggregate:

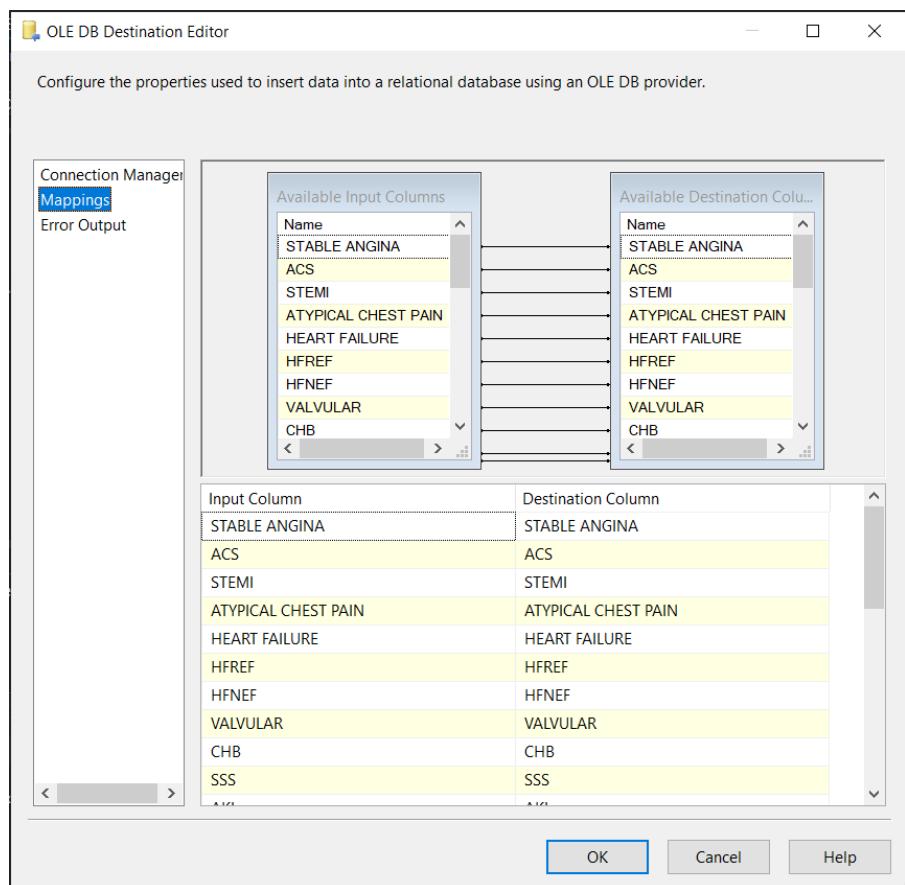
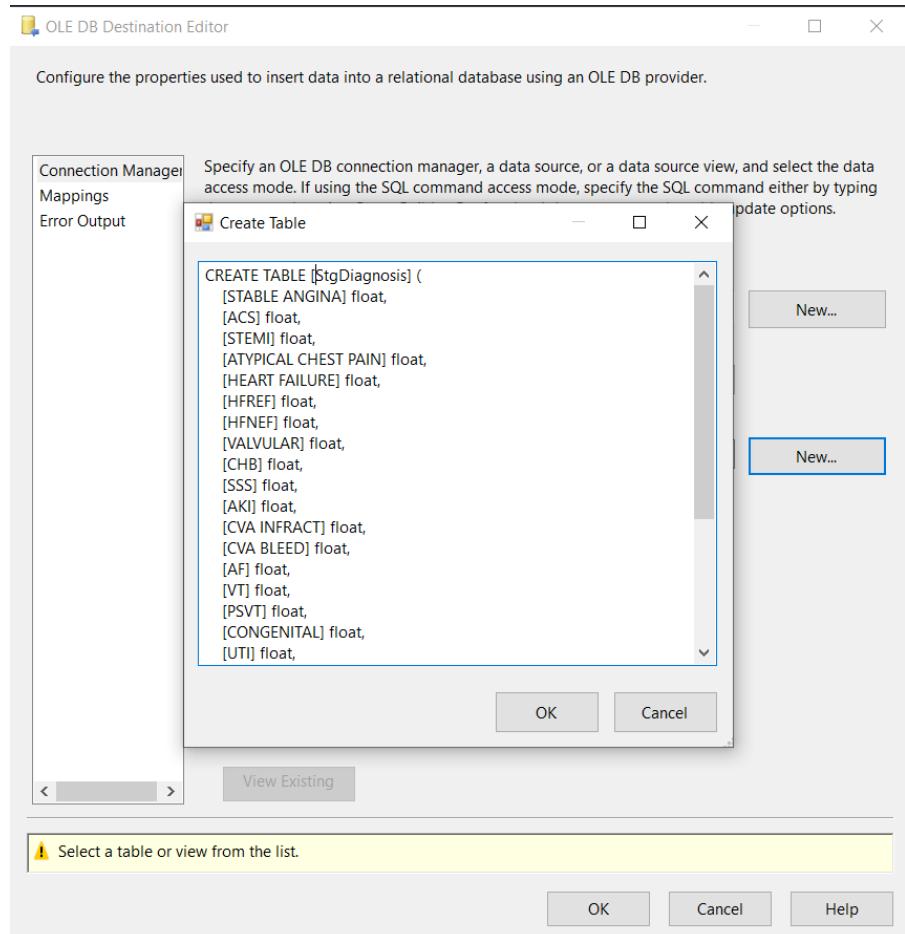


- Cấu hình Script Component:

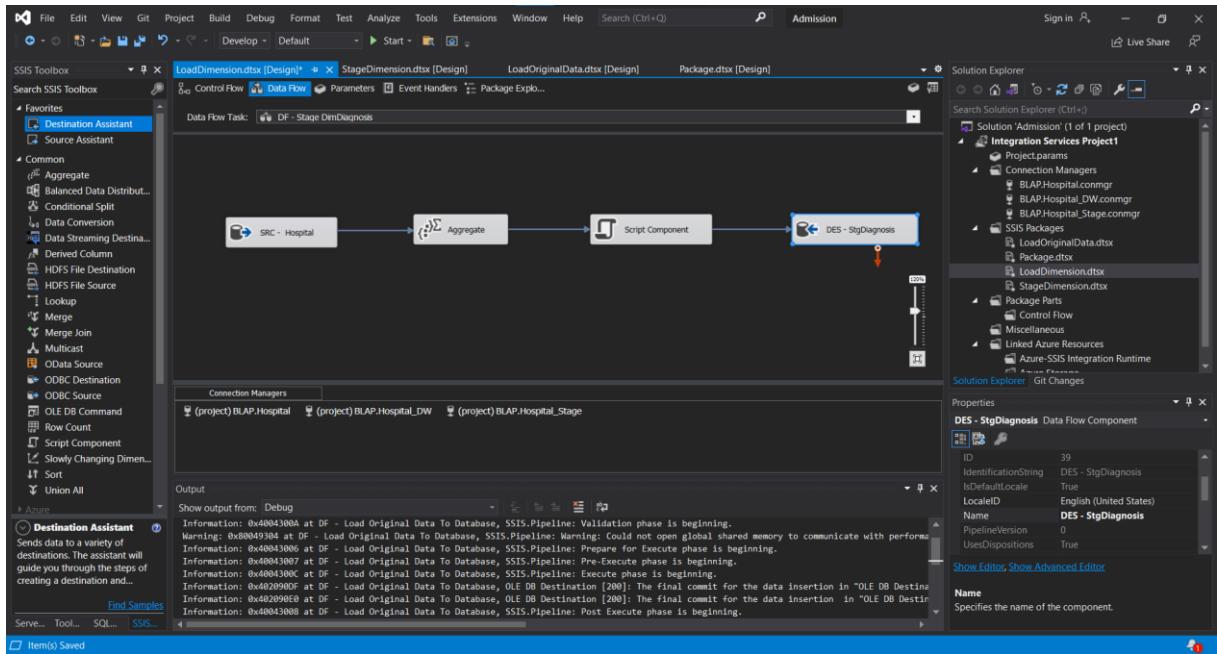
```
int count = 0;
2 references
public override void Input0_ProcessInputRow(Input0Buffer Row)
{
    /*
     * Add your code here
     */
    count++;
    Row.DiagnosisID = count;
}
```



- Cấu hình Destination Assistant:

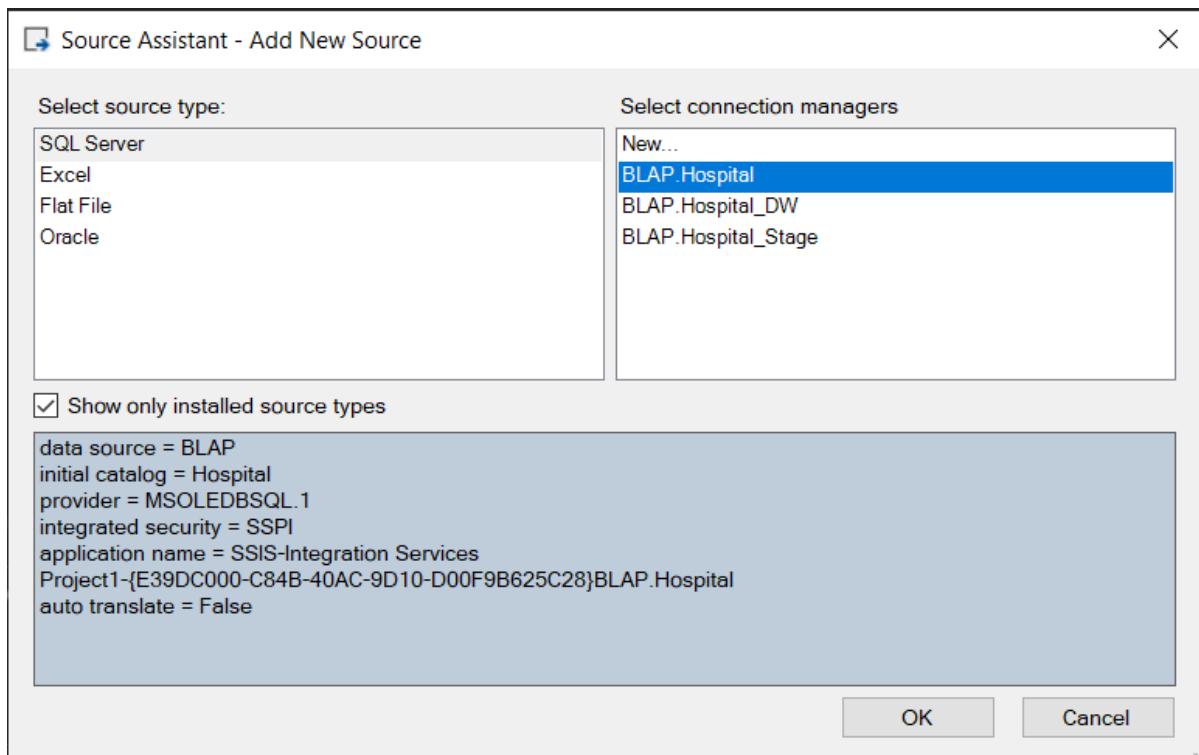


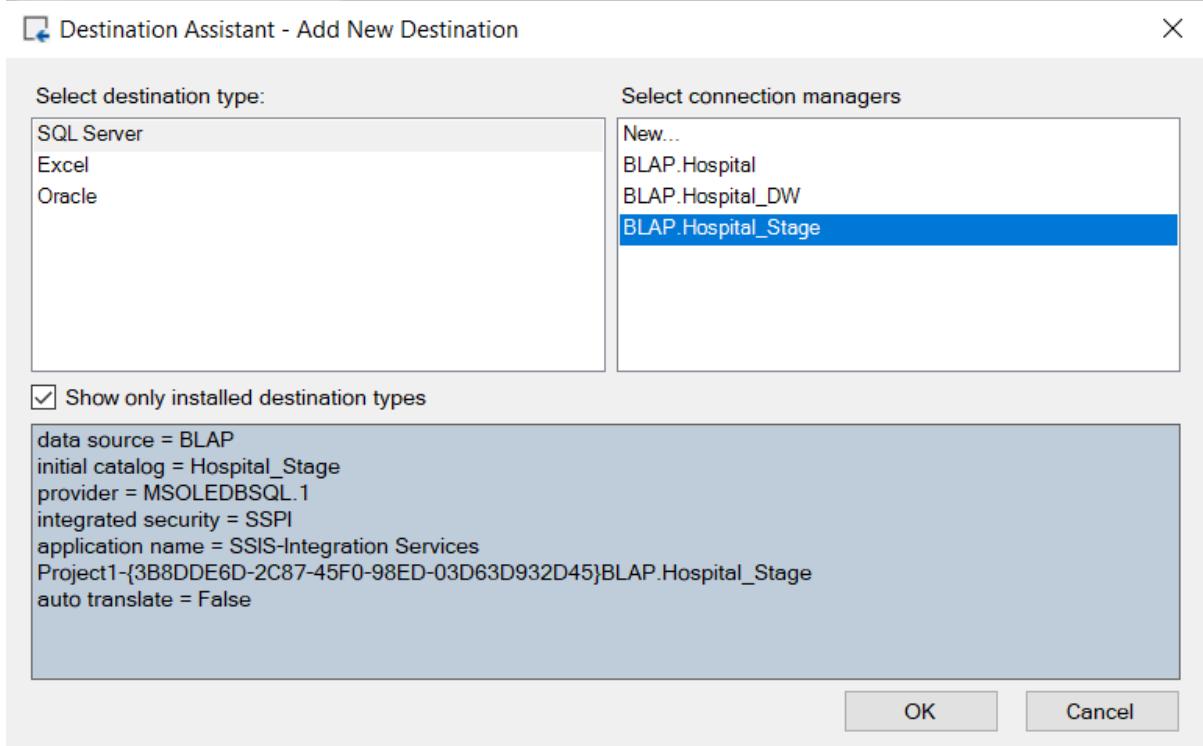
4. Data Flow Task của ta sẽ trông như sau:



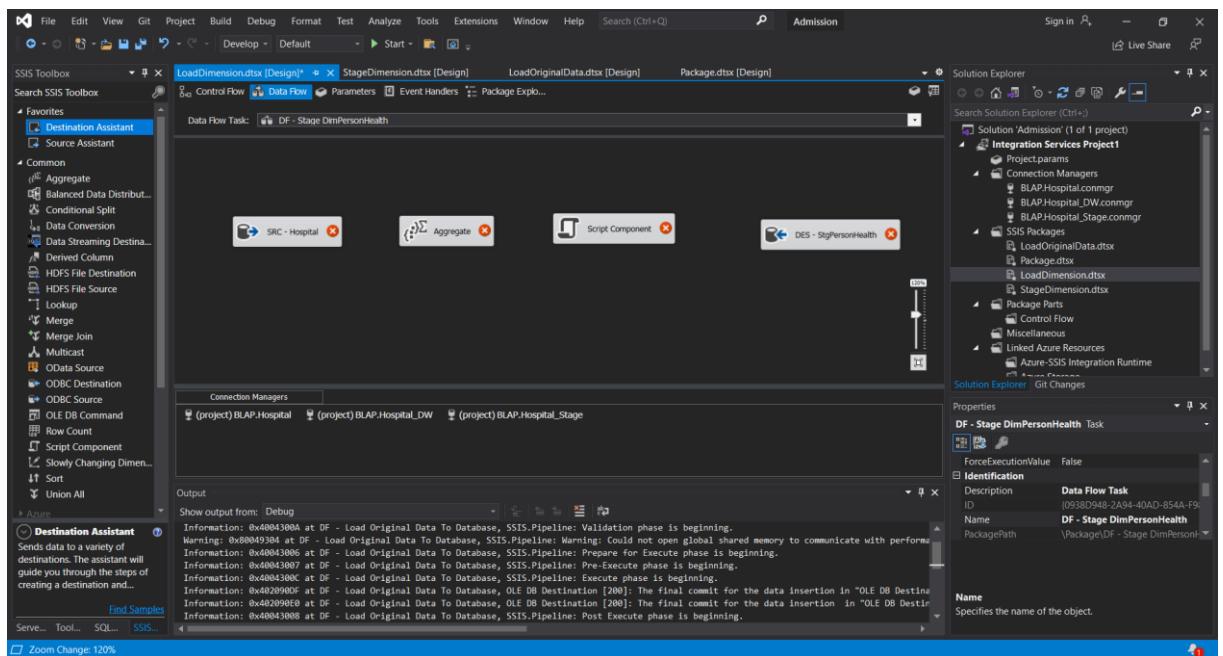
3.3.2.4. StgPersonHealth

1. Click vào DF - Stage DimPersonHealth và thêm Source Assistant, Aggregate, Script Component và Destination Assistant



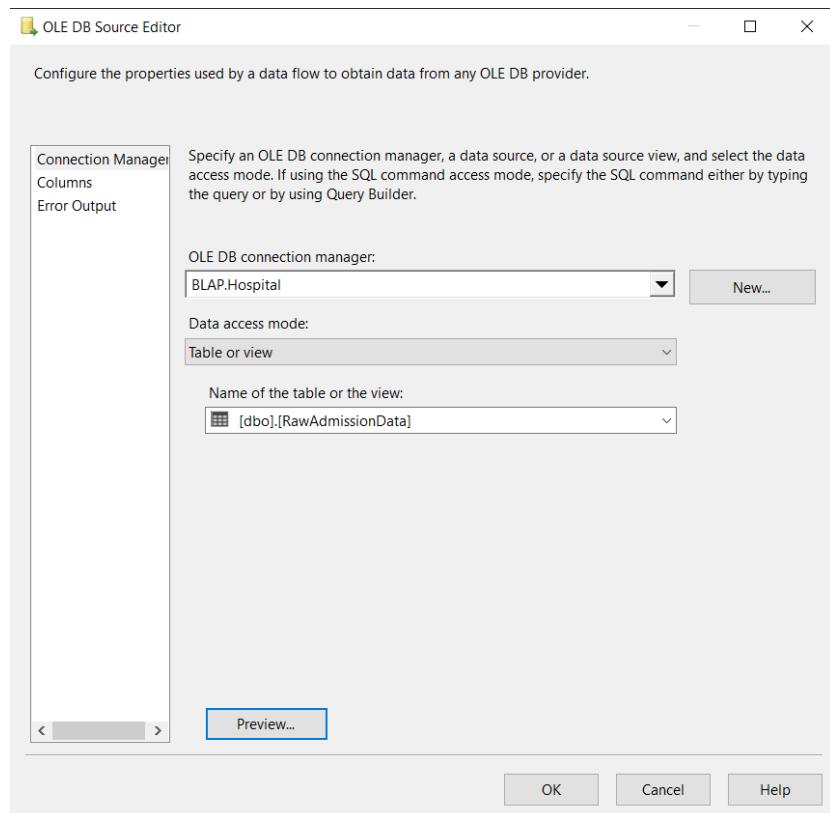


2. Tiến hành đổi tên và ta có được hình như sau:

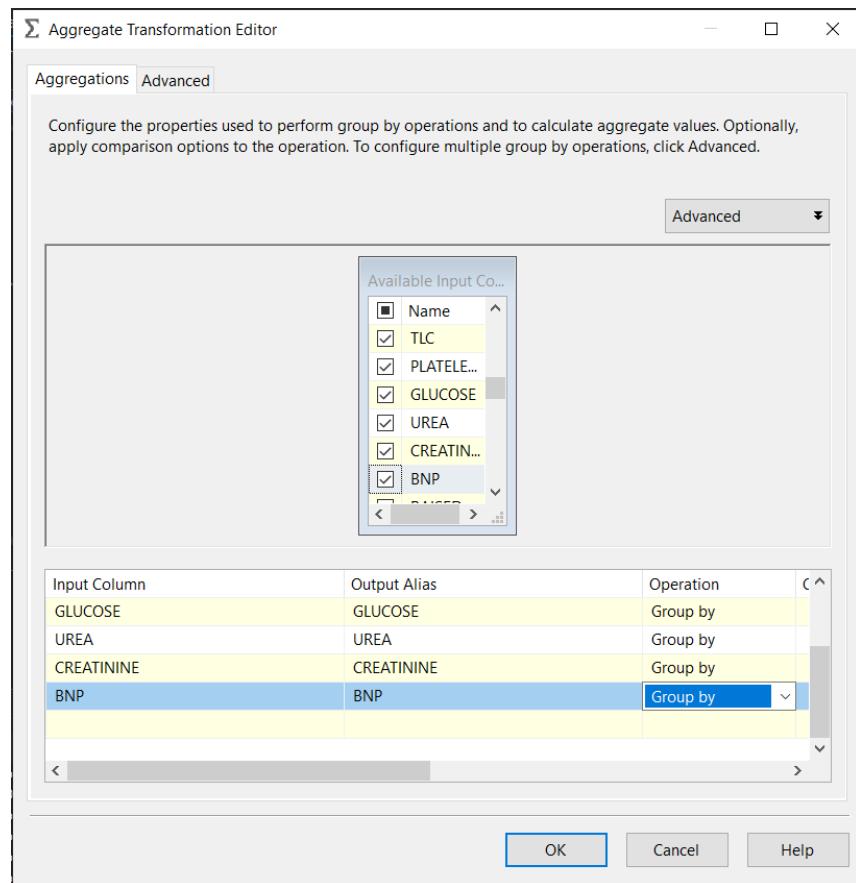


3. Tiến hành cấu hình các task:

- Cấu hình Source Assistant:

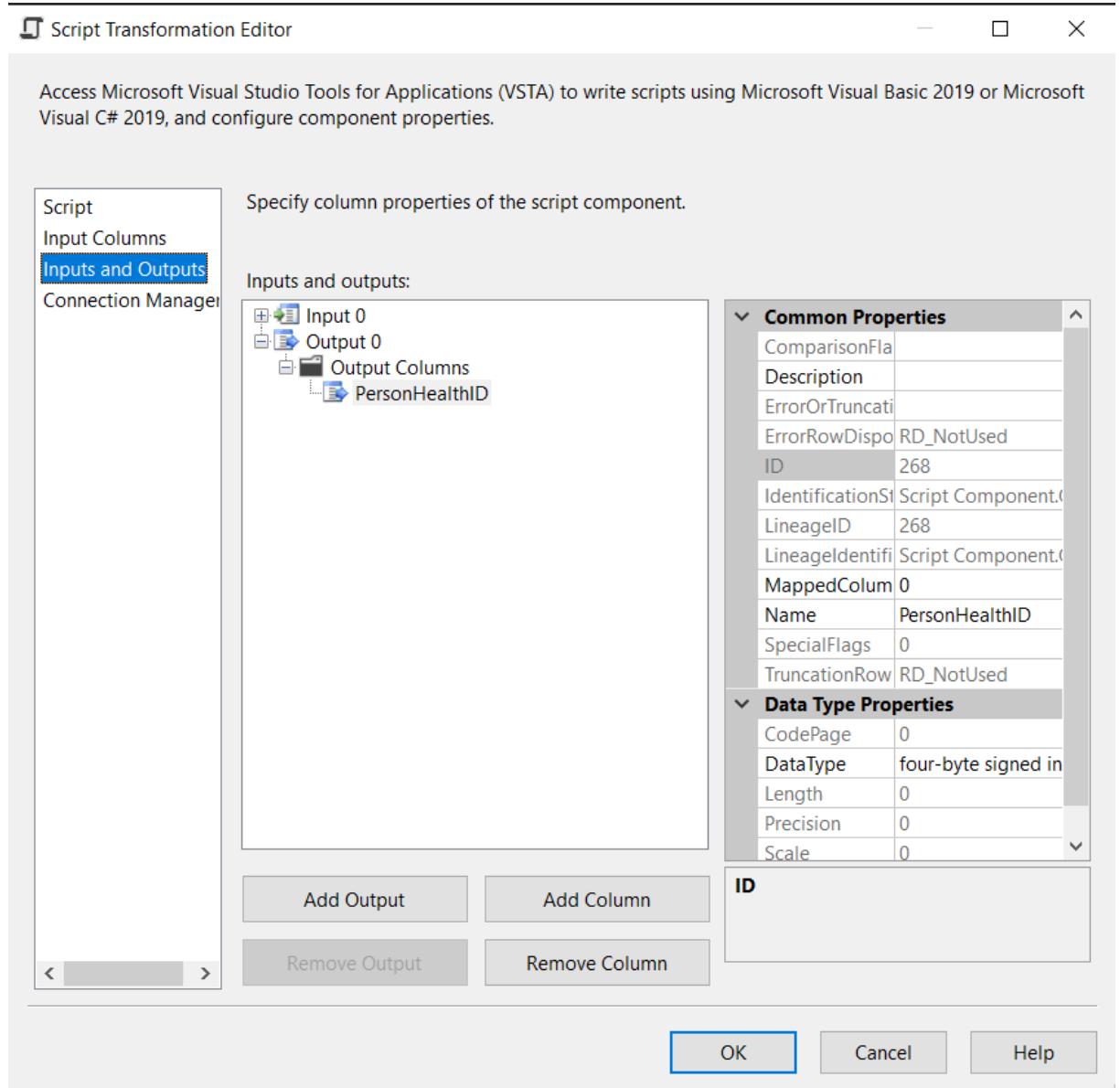


- Cấu hình Aggregate:

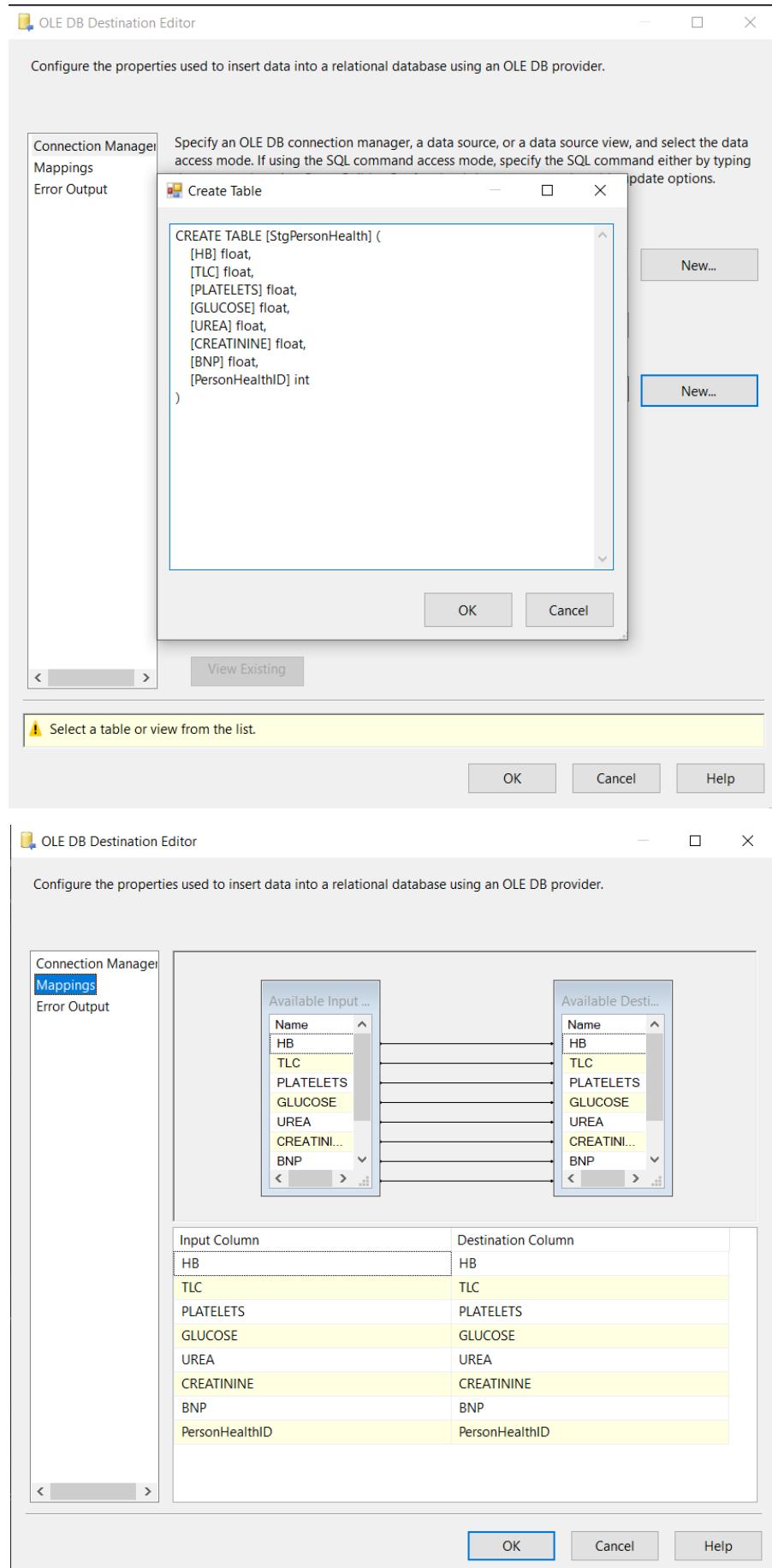


- Cấu hình Script Component:

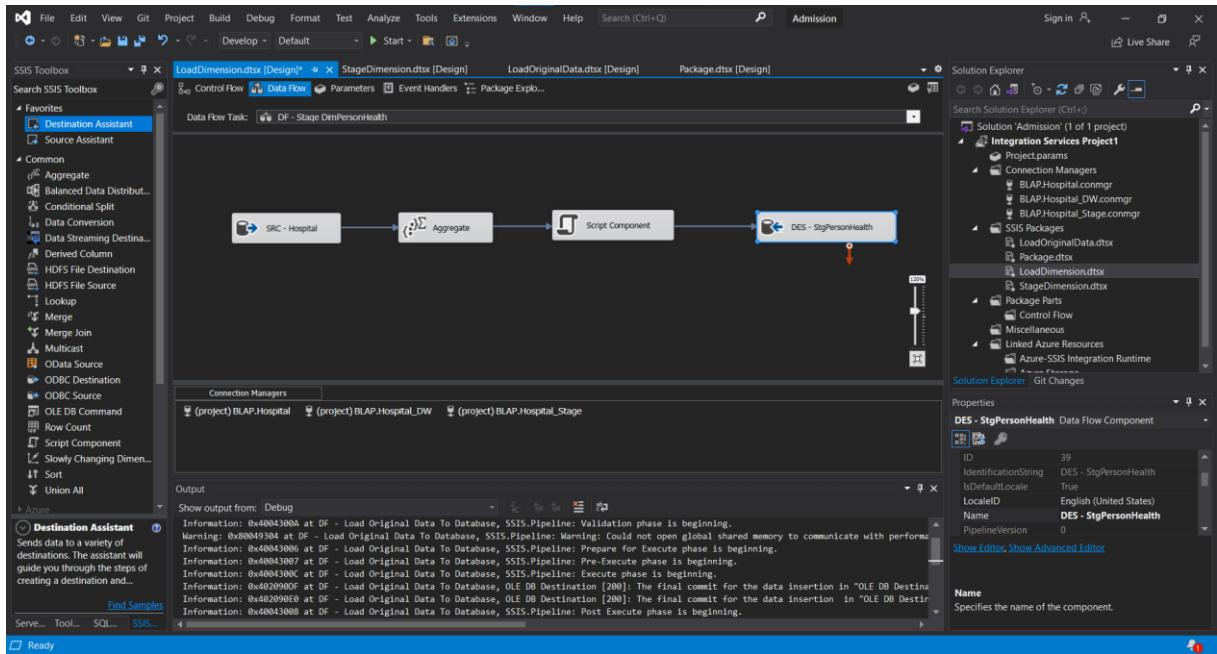
```
int count = 0;
2 references
public override void Input0_ProcessInputRow(Input0Buffer Row)
{
    /*
     * Add your code here
     */
    count++;
    Row.PersonHealthID = count;
}
```



- Cấu hình Destination Assistant:

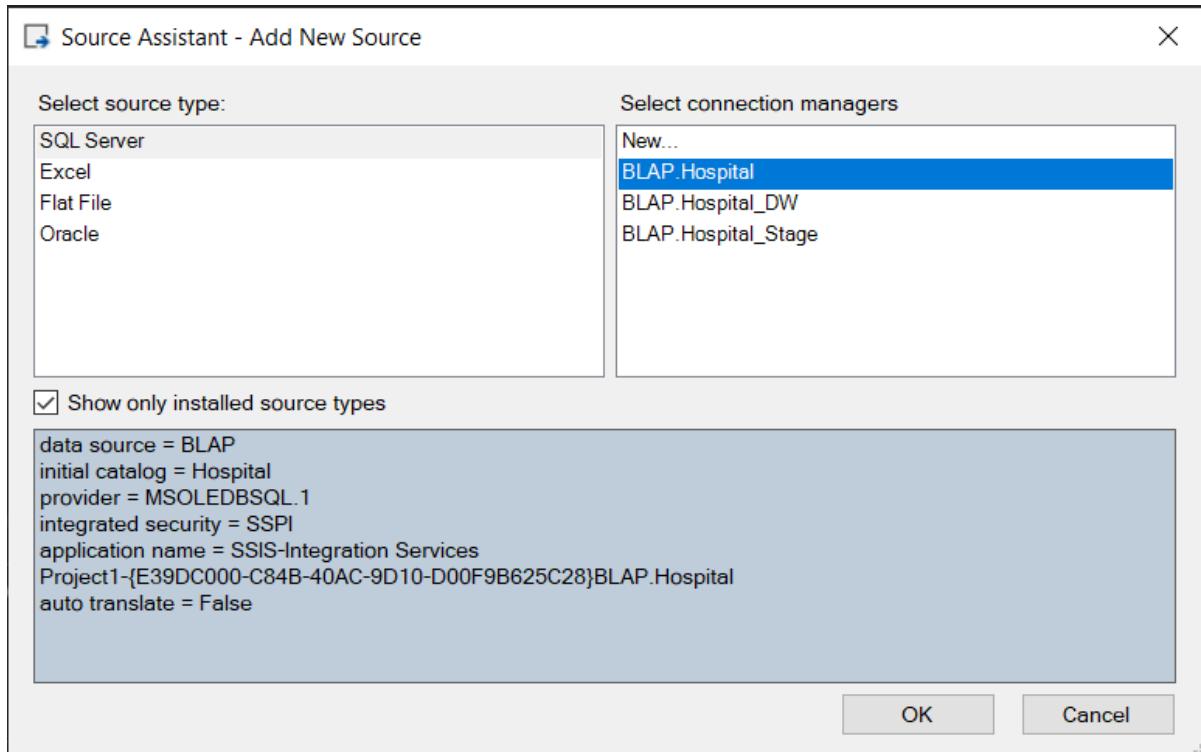


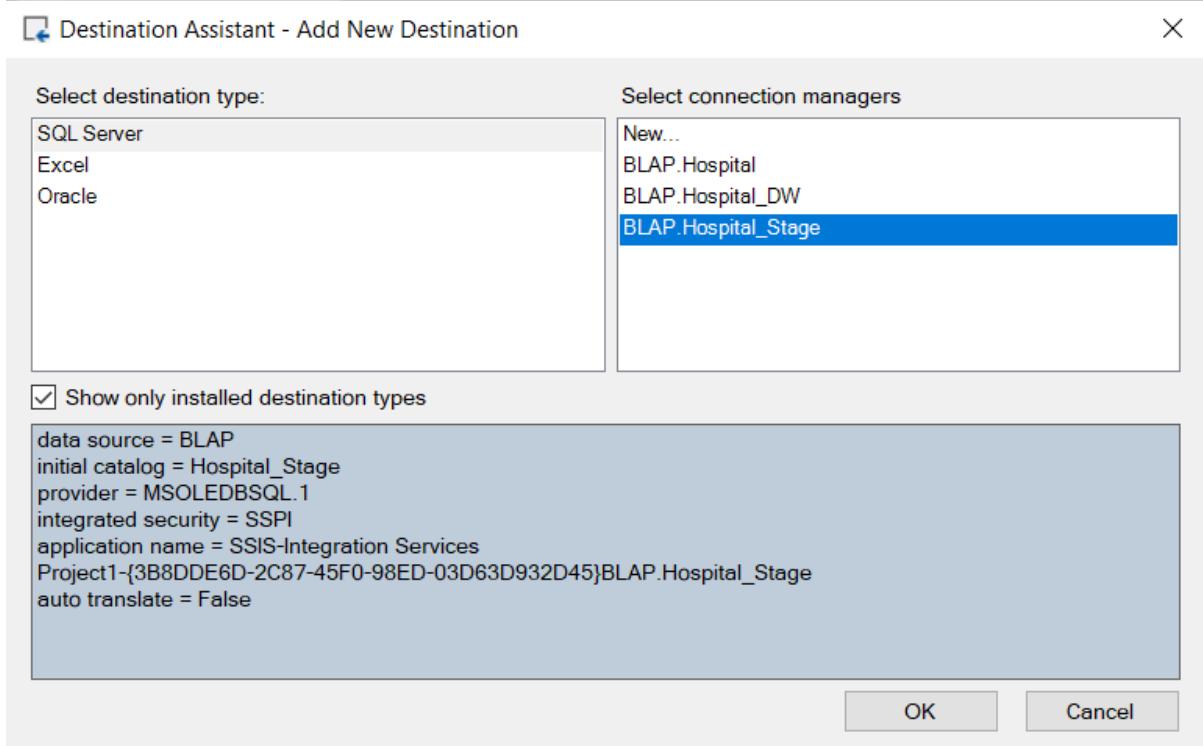
4. Data Flow Task của ta sẽ trông như sau:



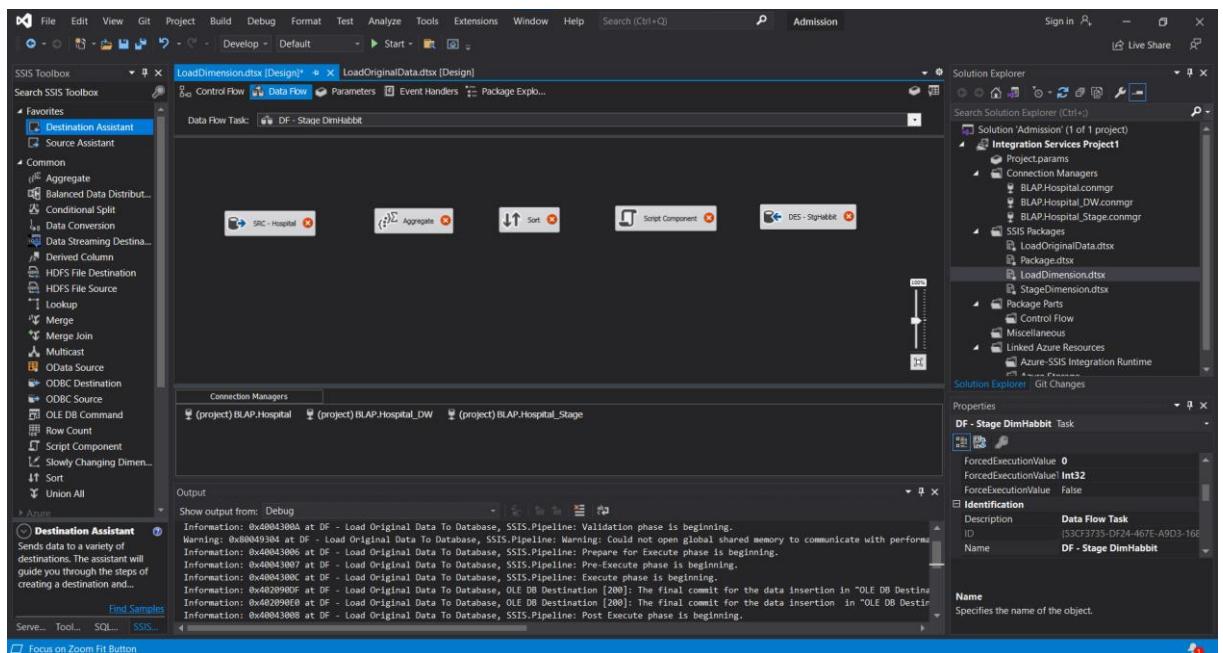
3.3.2.5. StgHabbit

1. Click vào DF - Stage DimHabbit và thêm Source Assistant, Aggregate, Sort, Script Component và Destination Assistant



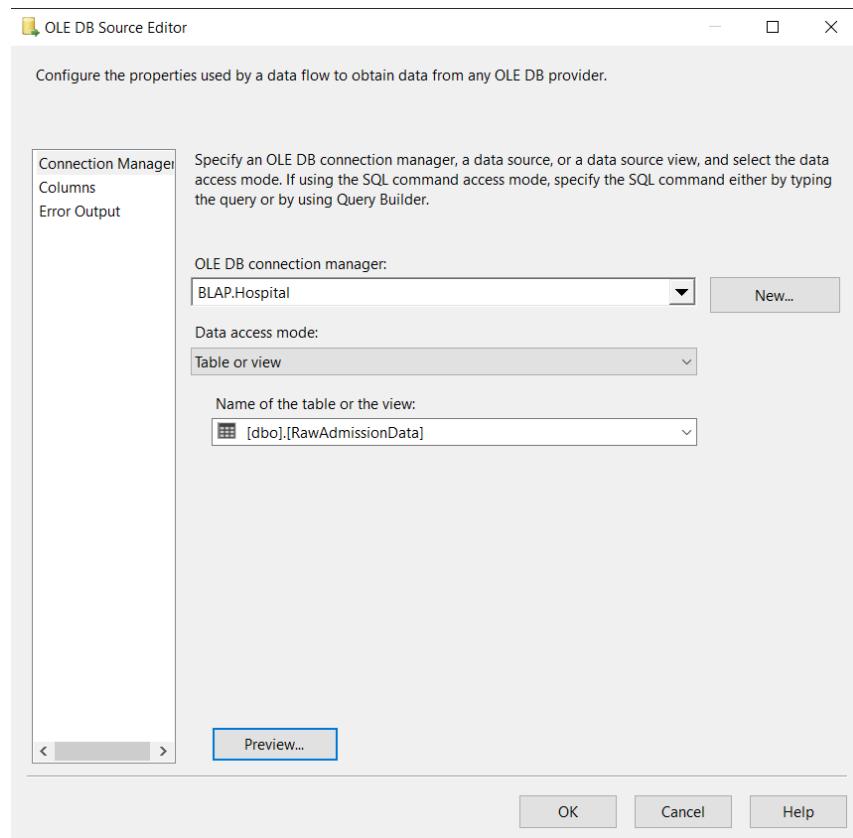


2. Tiến hành đổi tên và ta có được hình như sau:

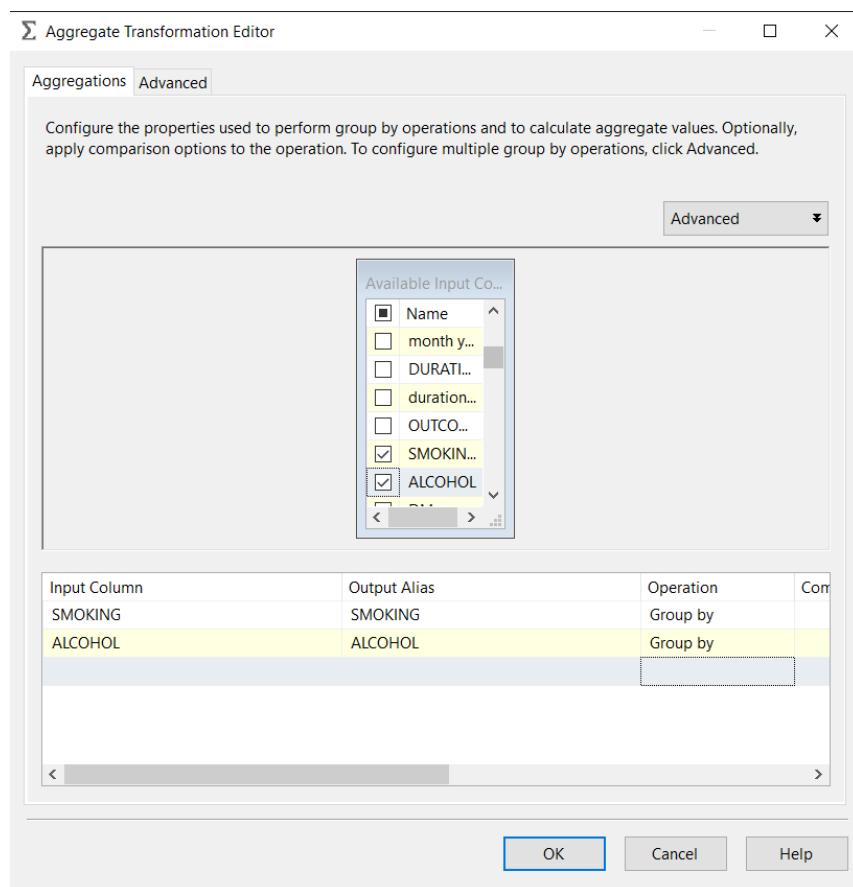


3. Tiến hành cấu hình các task:

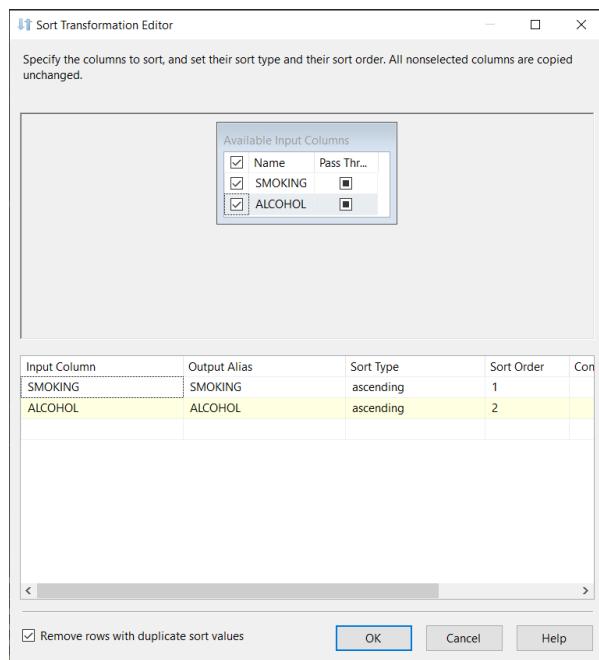
- Cấu hình Source Assistant:



- Cấu hình Aggregate:



- Cấu hình Sort:

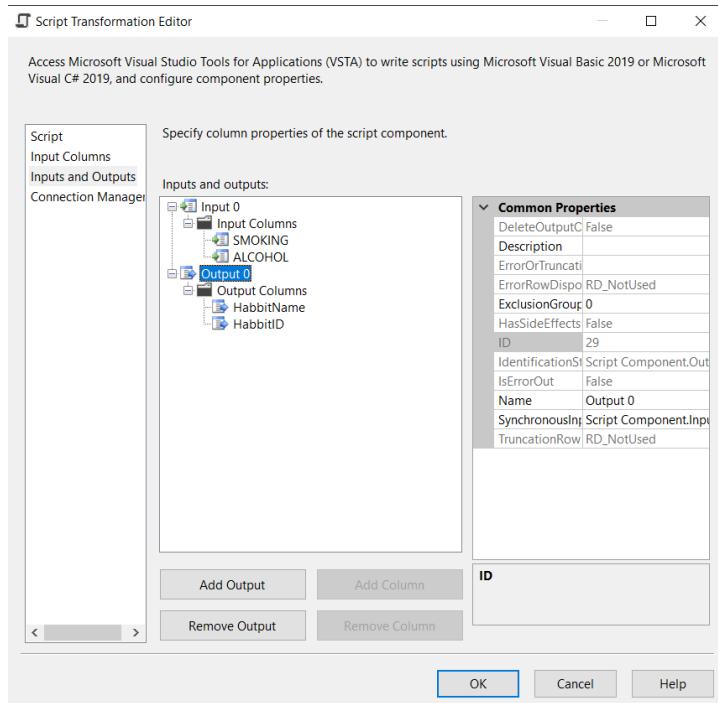


- Cấu hình Script Component:

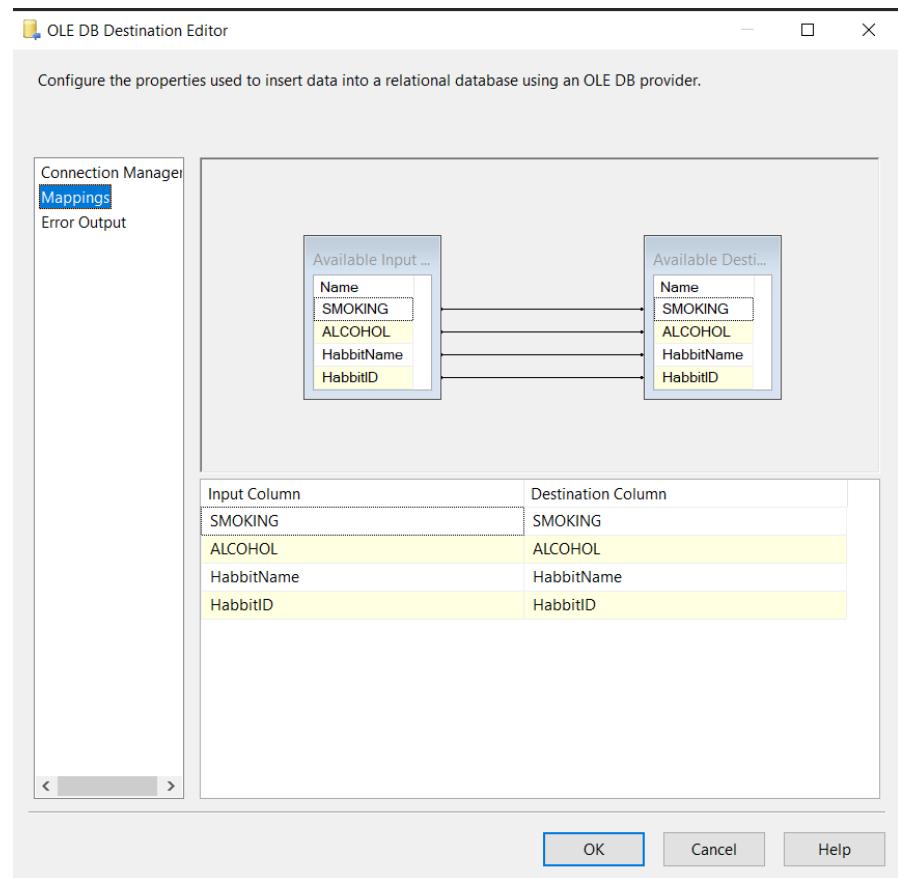
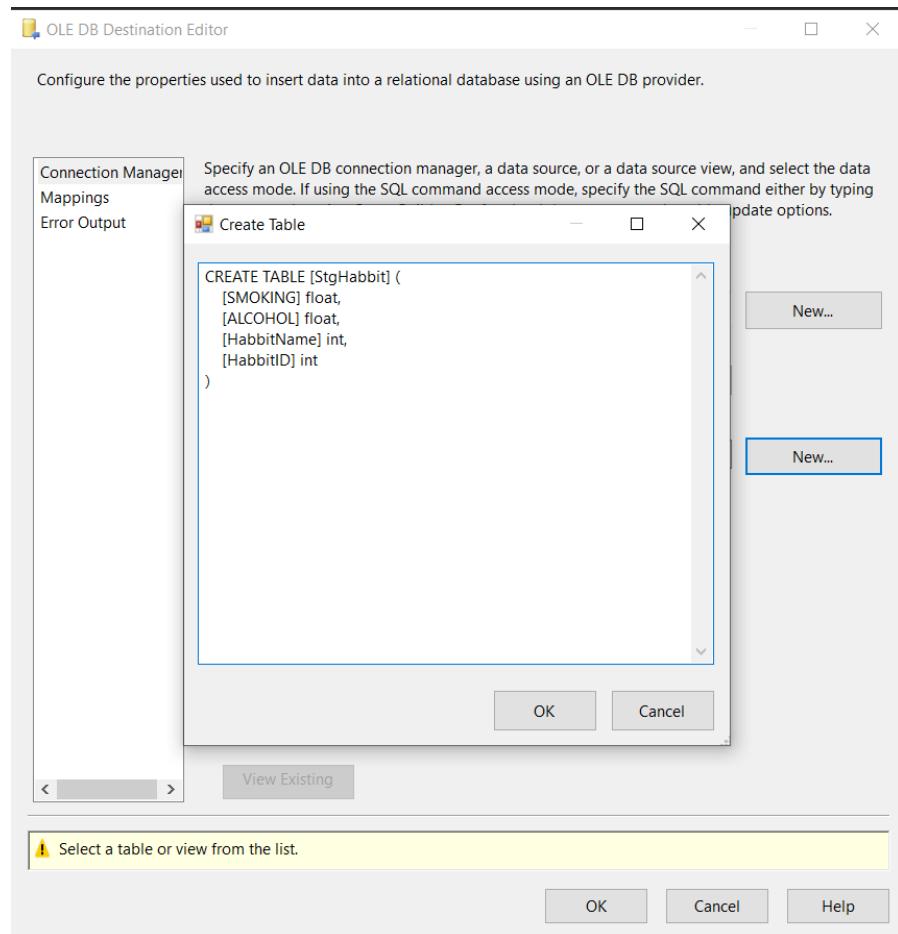
```

int count = 0;
2 references
public override void Input0_ProcessInputRow(Input0Buffer Row)
{
    /*
     * Add your code here
     */
    count++;
    Row.HabbitID = count;
    if (checkSmoke(Row.ALCOHOL) && checkSmoke(Row.SMOKING))
    {
        Row.HabbitName = "Both";
    }
    else if (checkSmoke(Row.SMOKING) && !checkAlcoho(Row.ALCOHOL))
    {
        Row.HabbitName = "Smoke";
    }
    else if (!checkSmoke(Row.SMOKING) && checkAlcoho(Row.ALCOHOL))
    {
        Row.HabbitName = "Alcohol";
    }
    else if (!checkSmoke(Row.SMOKING) && !checkAlcoho(Row.ALCOHOL))
    {
        Row.HabbitName = "None";
    }
}

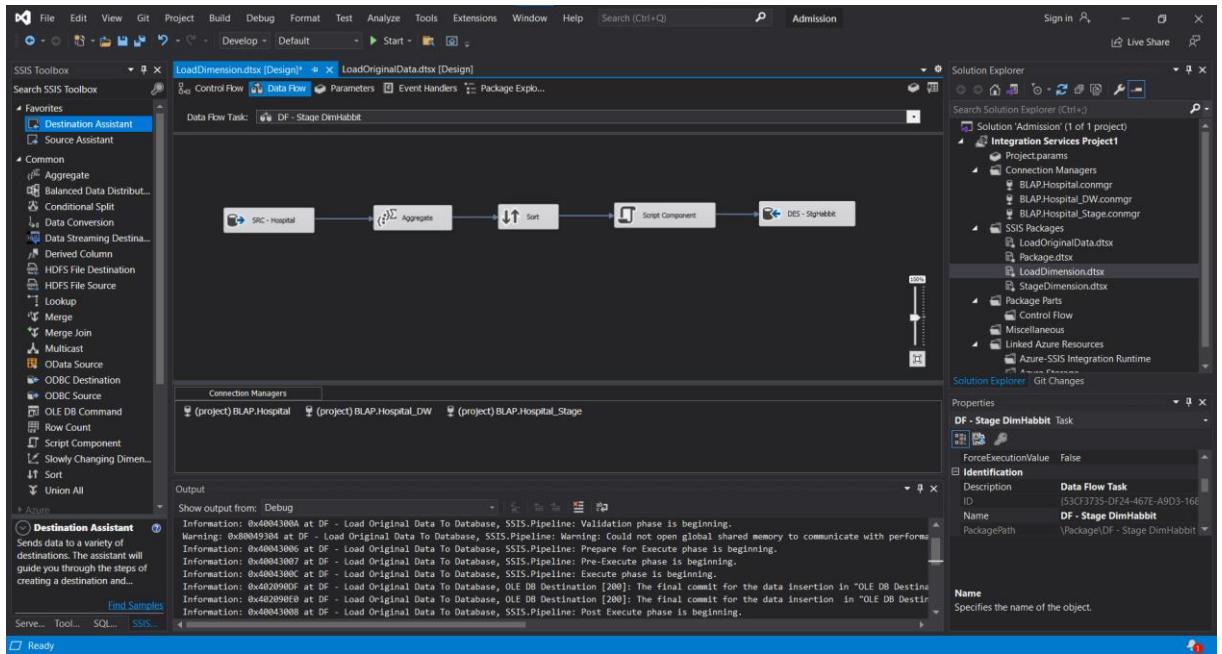
```



- Cấu hình Destination Assistant:

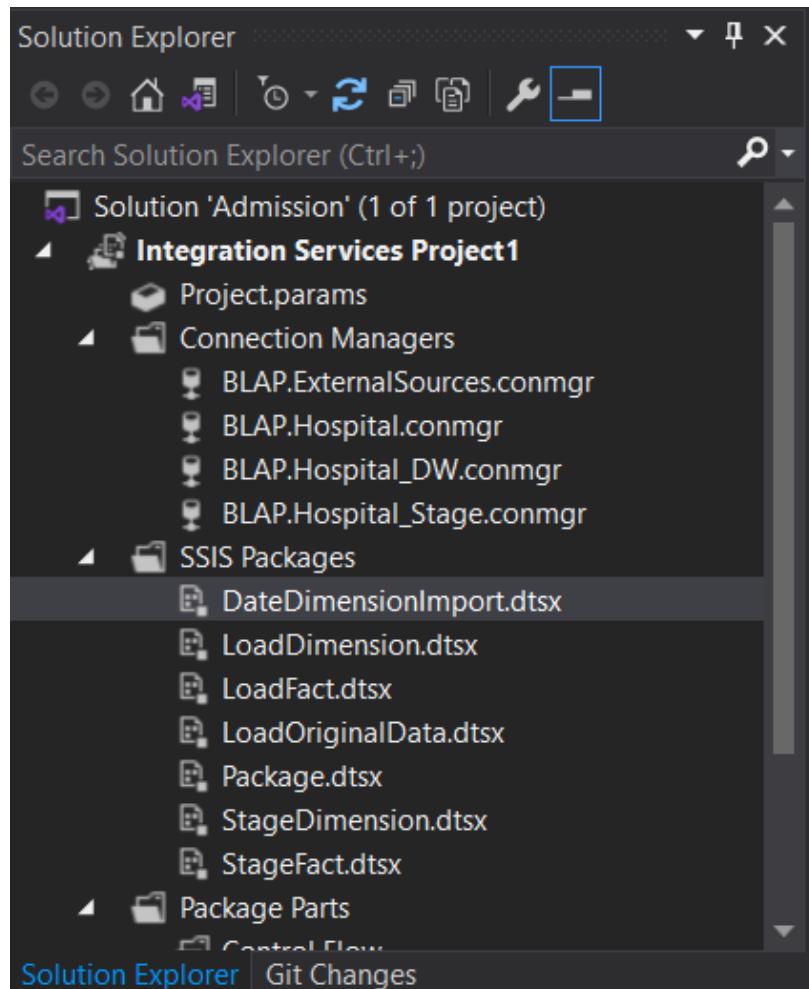


4. Data Flow Task của ta sẽ trông như sau:

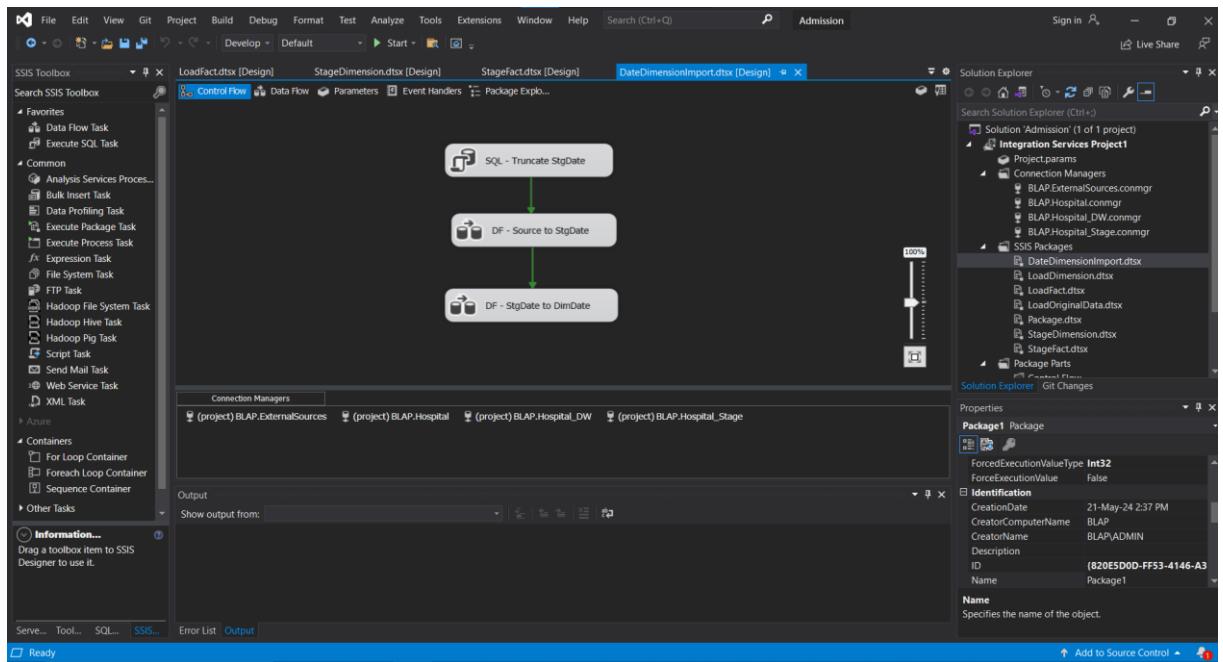


3.3.2.6. StgDate

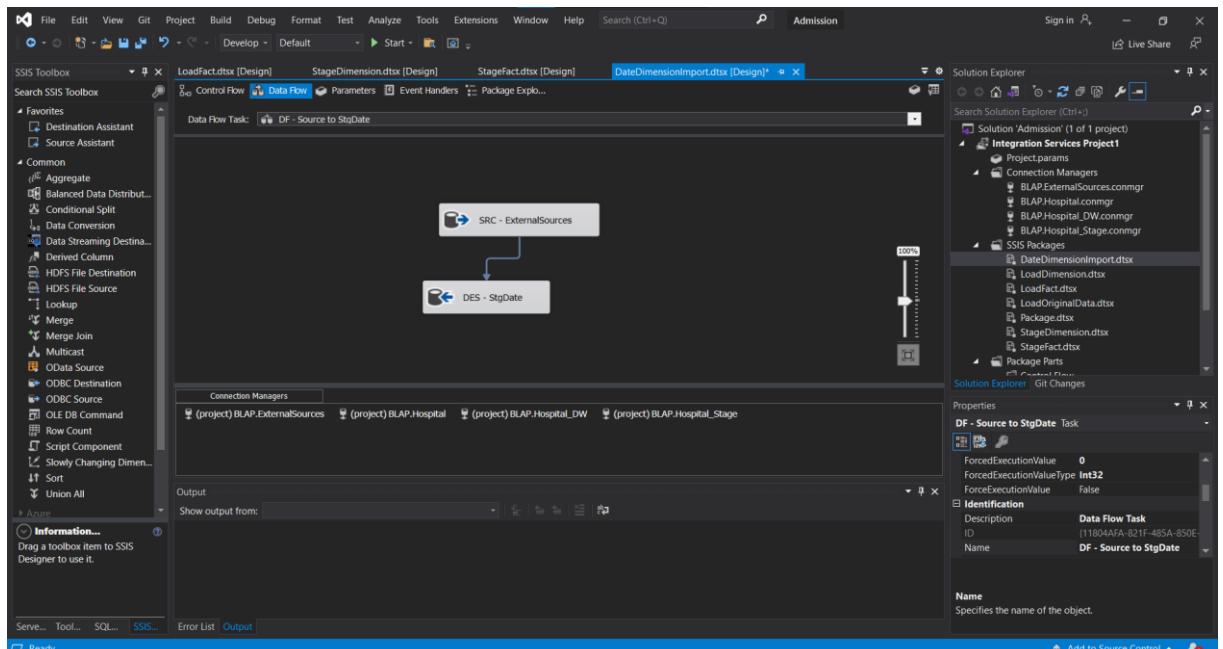
1. Ta tạo một package mới mang tên **DateDimensionImport** đồng thời ta tạo thêm một kết nối mới tới **External Sources** để lấy dữ liệu thời gian.



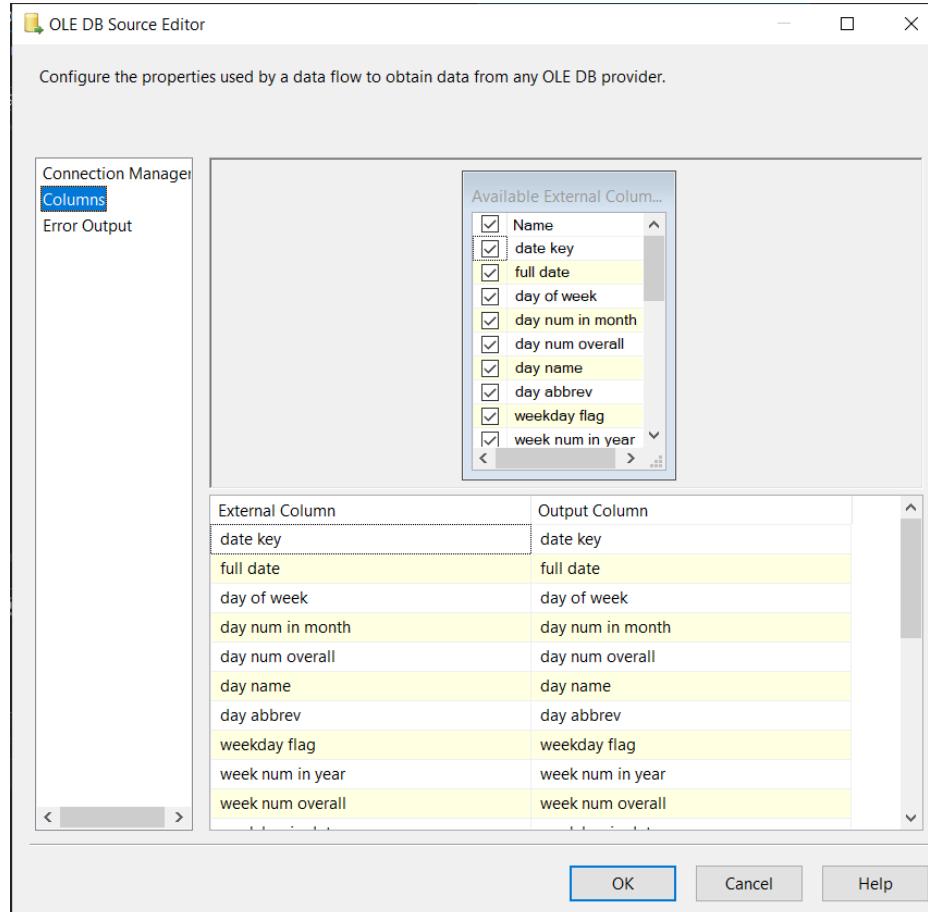
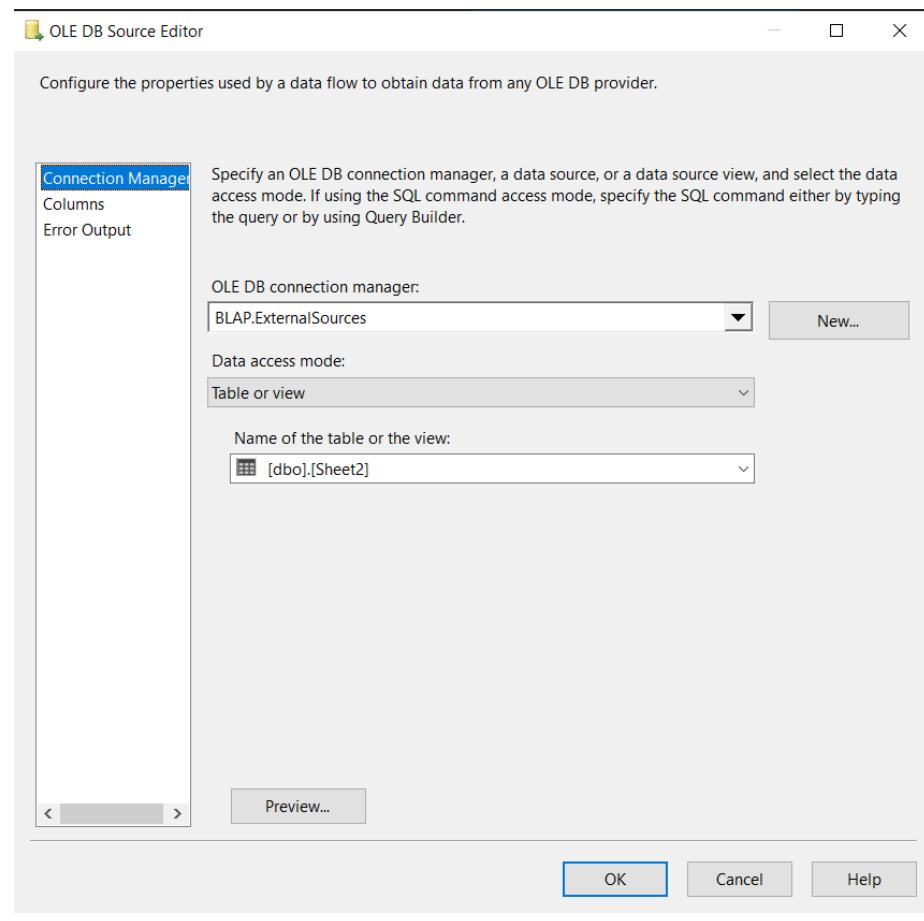
2. Nhập chọn **DateDimensionImport** và kéo thả 1 **Execute SQL task** và 2 **Data flow tasks** vào **surface design** và tiến hành đặt tên cho các task. Sau khi hoàn thành ta sẽ có kết quả như hình sau:



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:



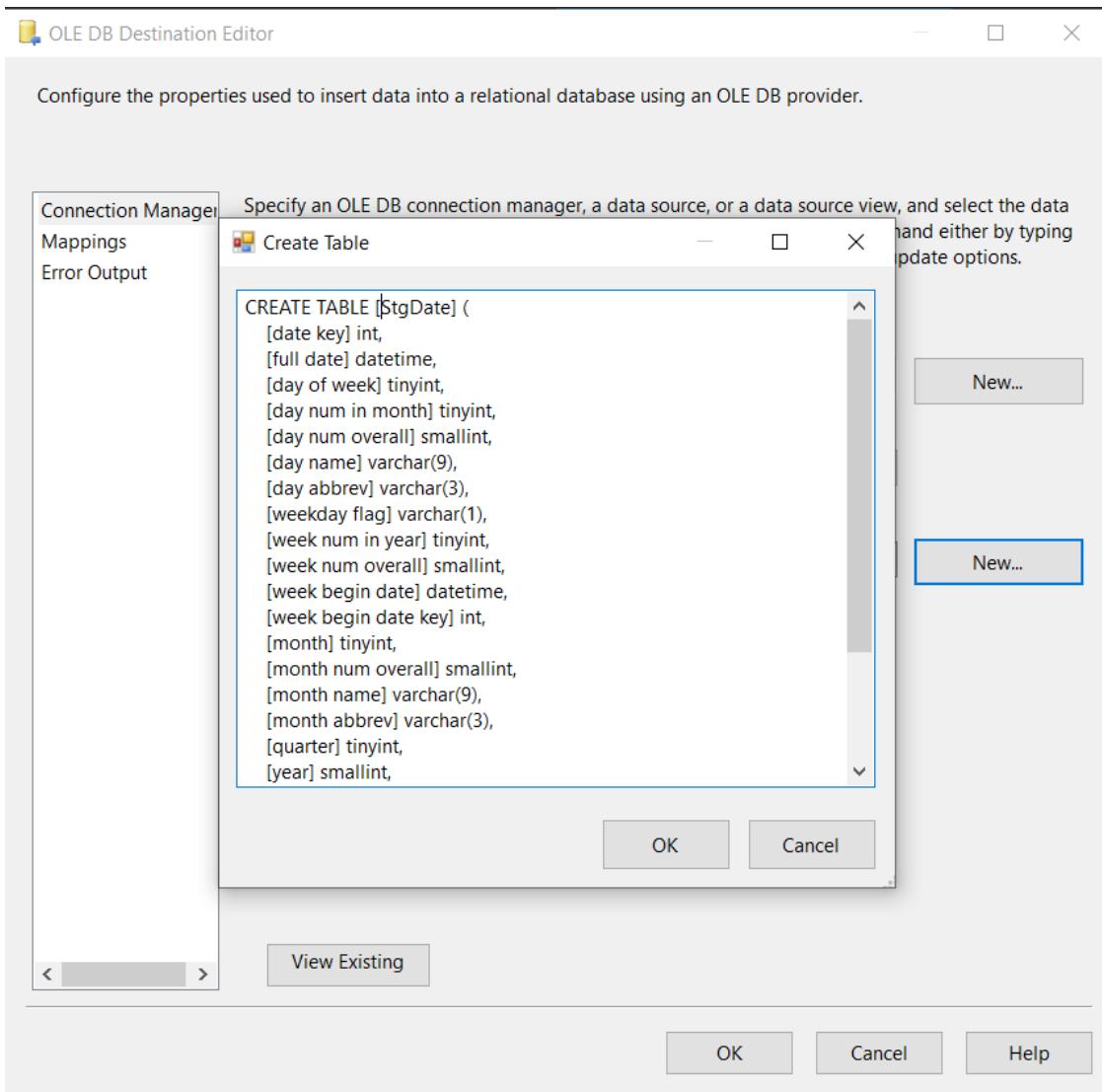
- Cấu hình Source Assistant:

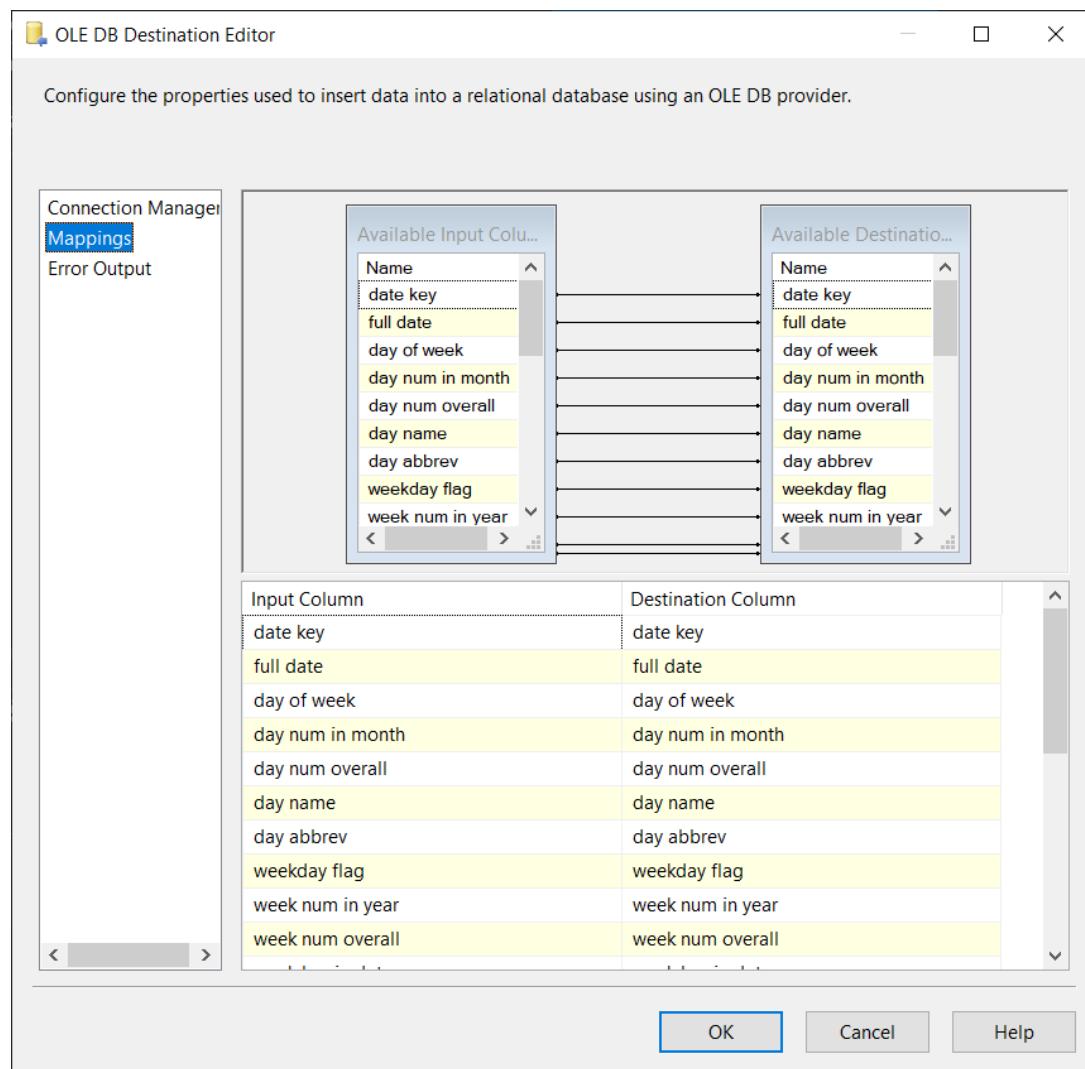


Dữ liệu của bảng được lấy từ file Excel DateStage.xls với nội dung dữ liệu như sau

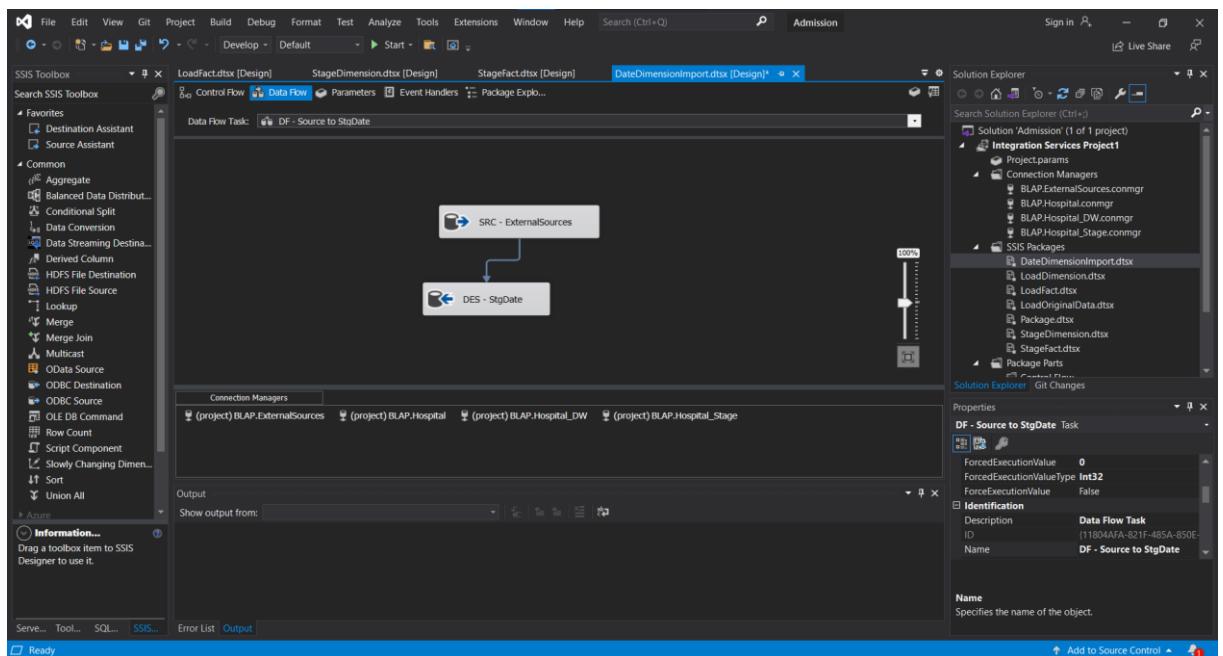
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W				
1	date key	full date	day of week	num in day	day name	day abbrev	weekday	week num	week begin date	month	year	month	num in month	month overall	month name	month abbrev	quarter	year	yearmo	7	3	fiscal year	mon	fiscal year	mon	quarter	year
2	20170101	01-01-17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

- Cấu hình Destination Assistant:

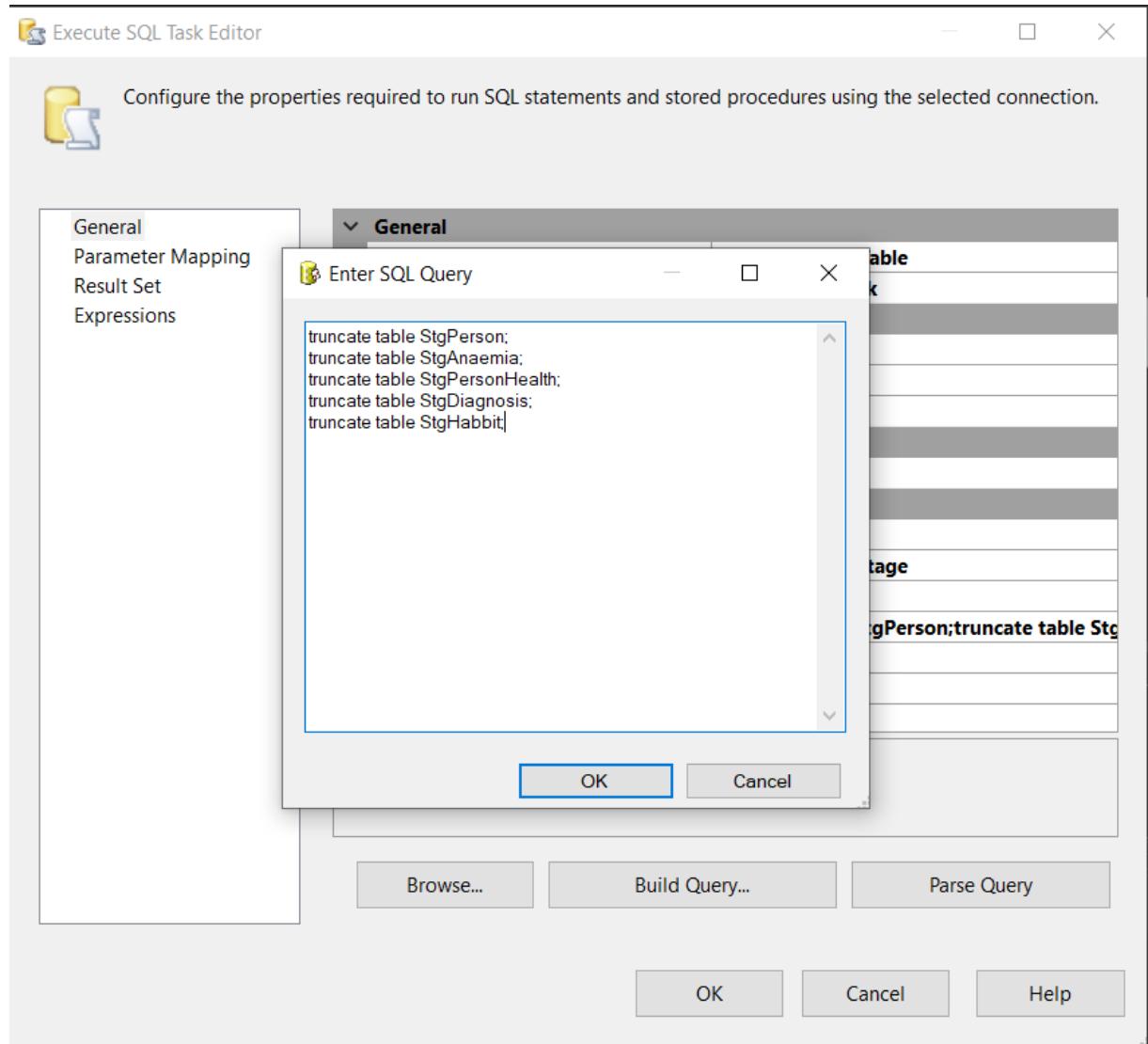


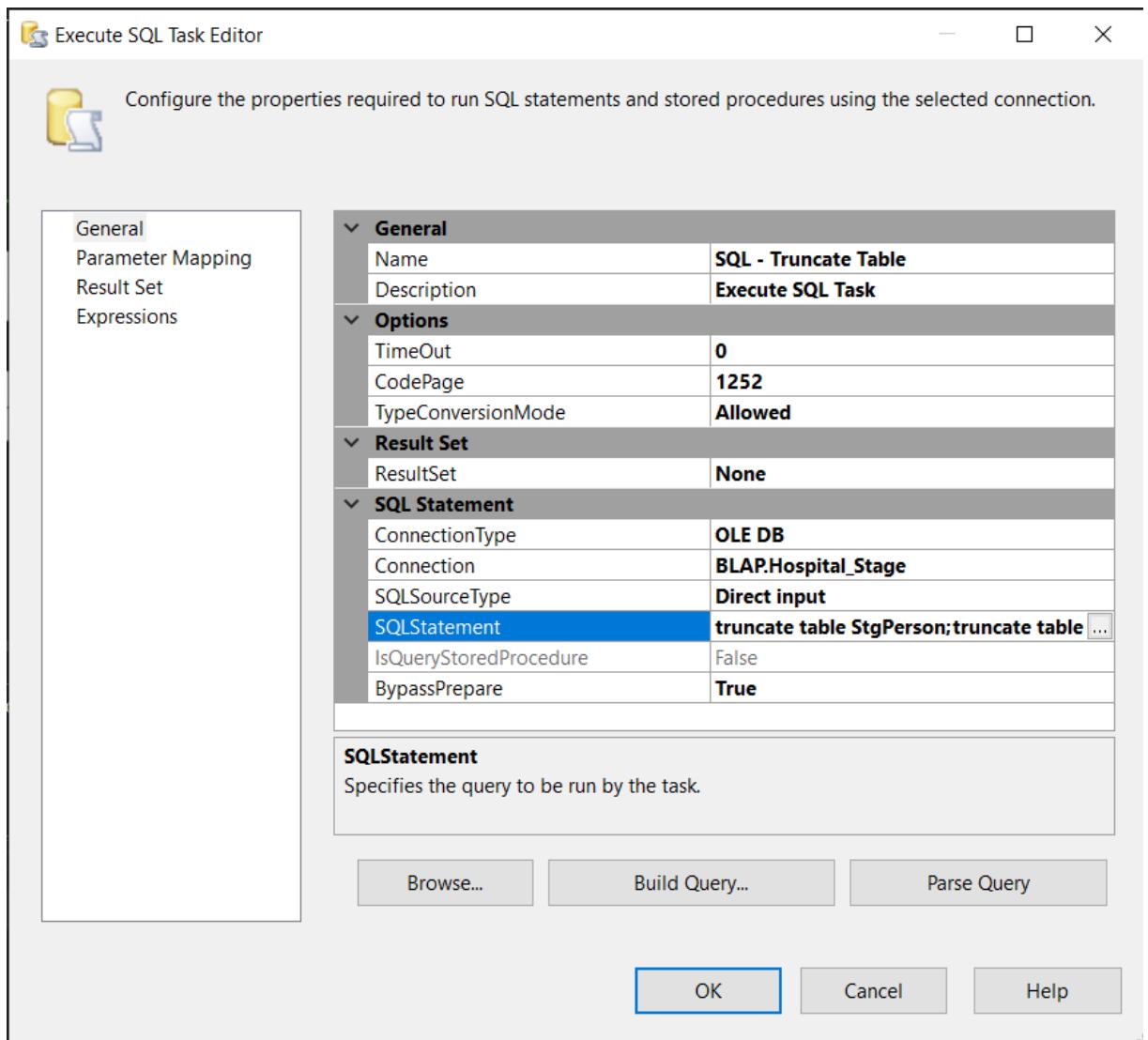


4. Data Flow Task của ta sẽ trông như sau:

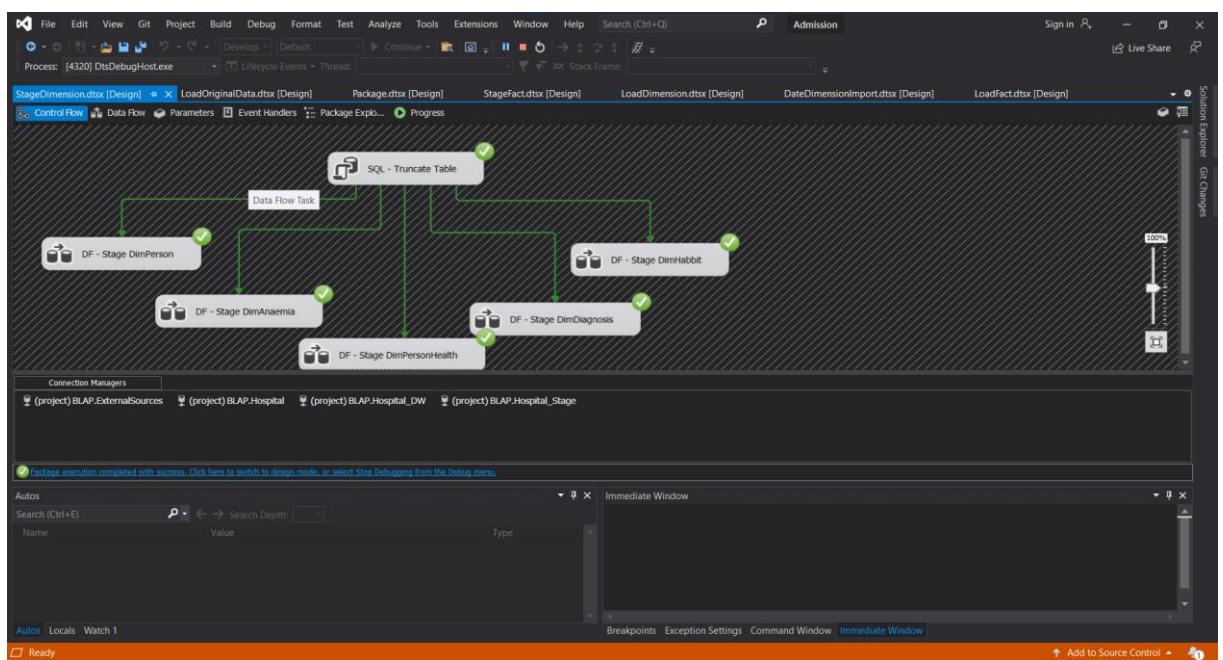


3.3.2.7. Cấu hình Execute SQL Task

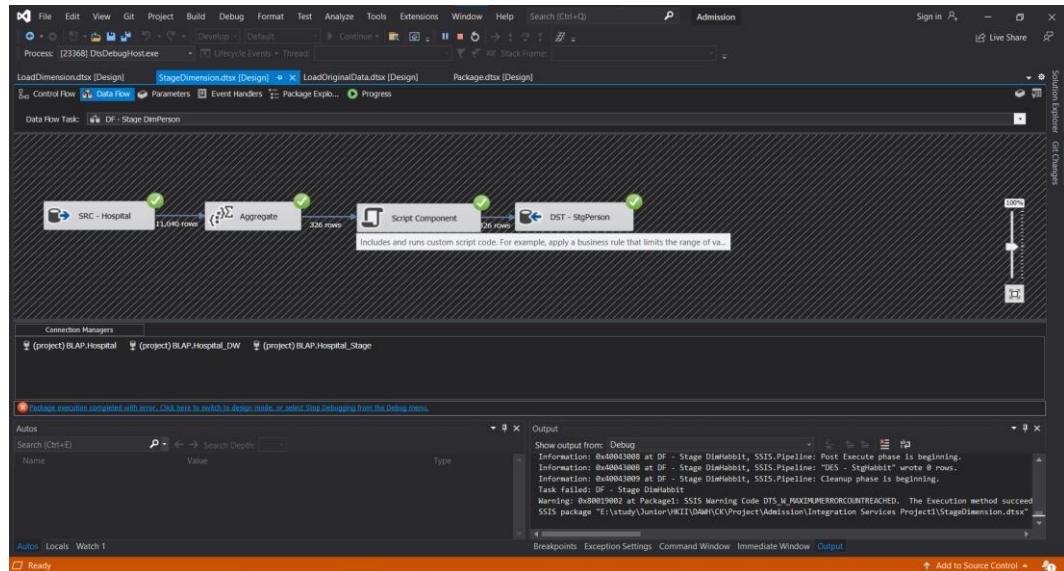




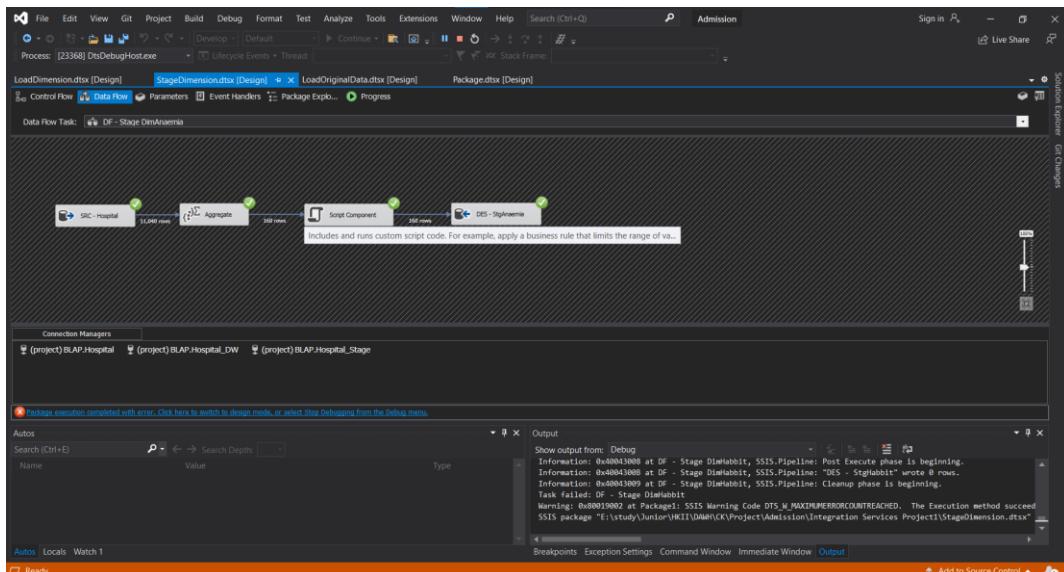
3.3.2.8. Tiến hành chạy Stage dữ liệu



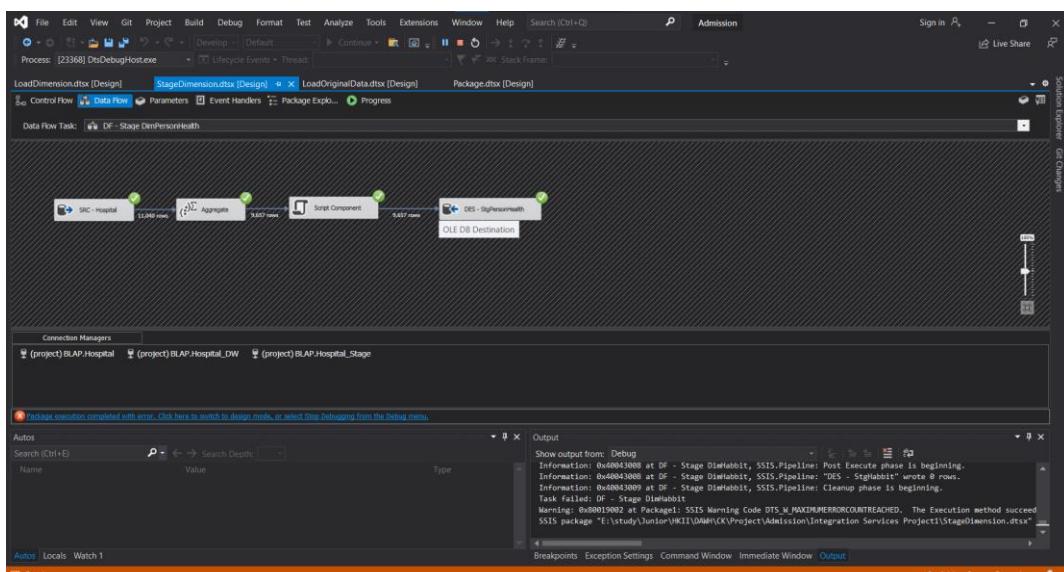
- StgPerson



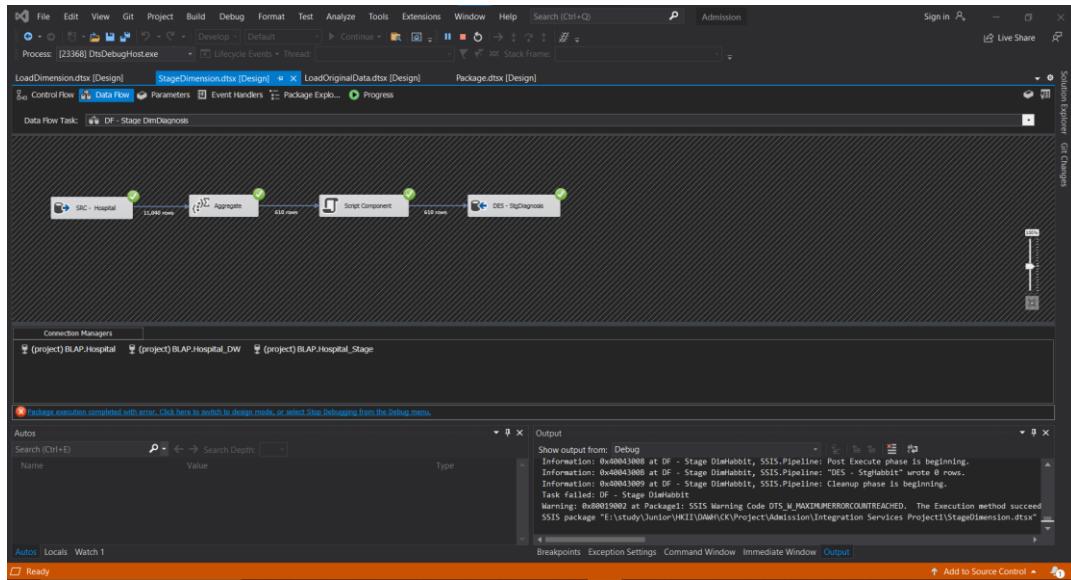
- StgAnaemia



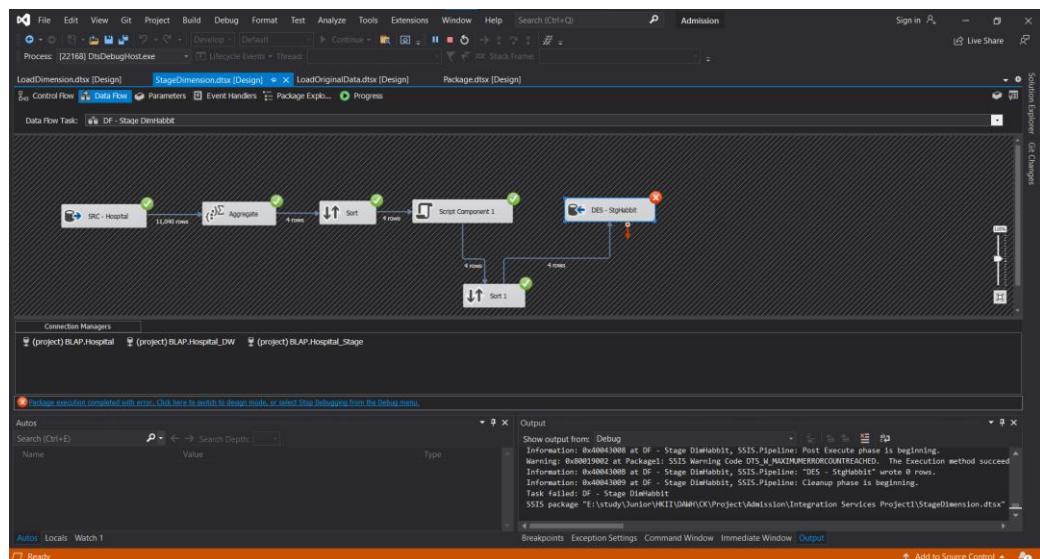
- StgPersonHealth



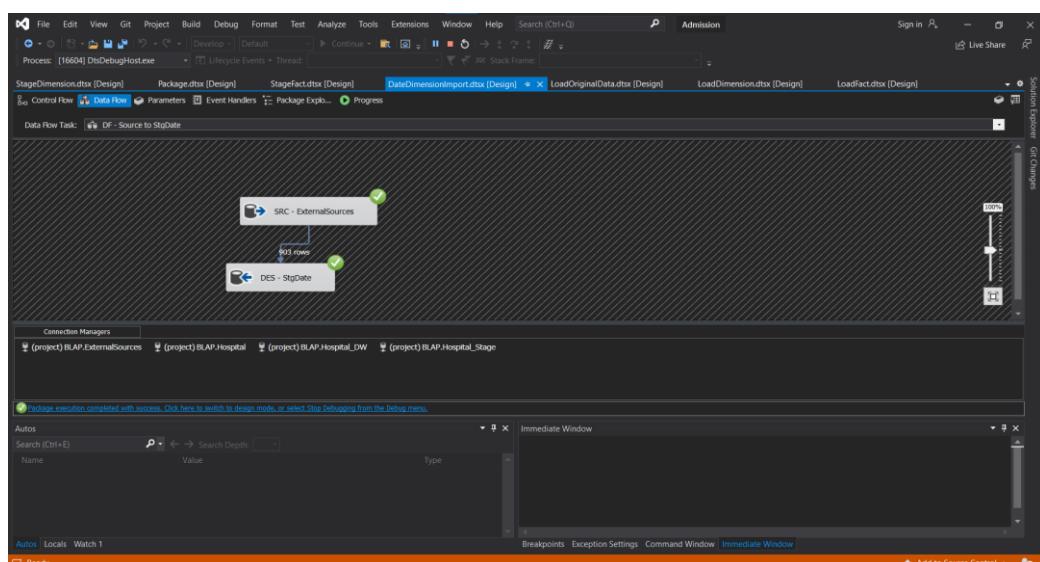
- StgDiagnosis



- StgHabbit

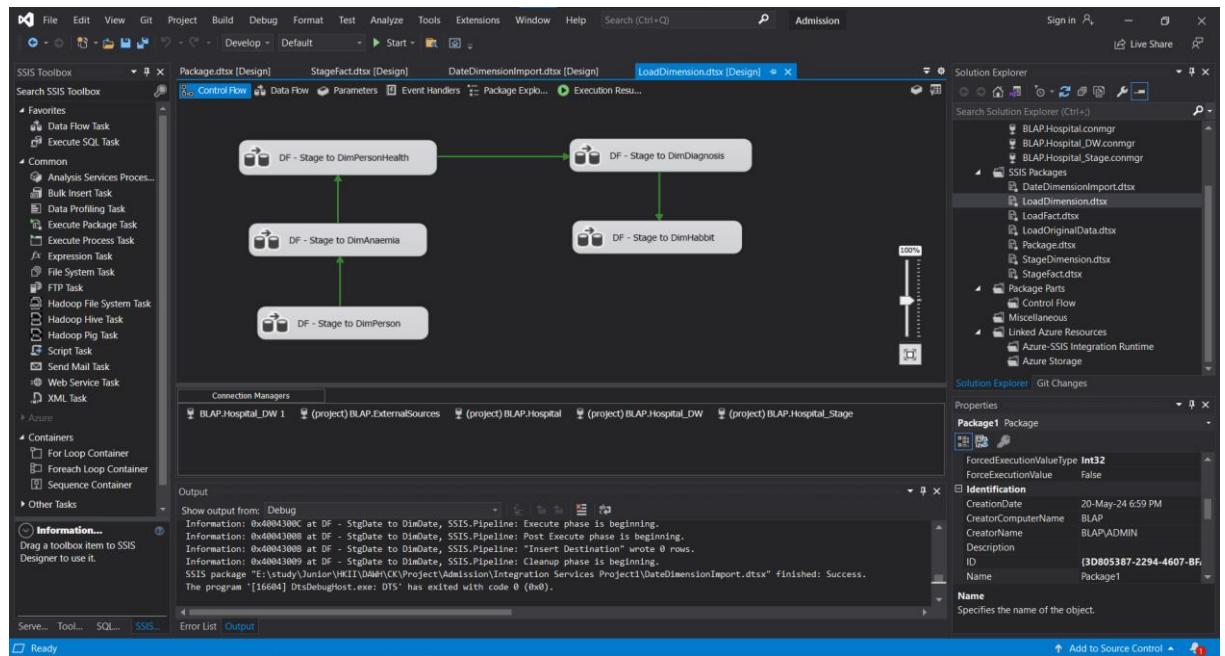


- StgDate

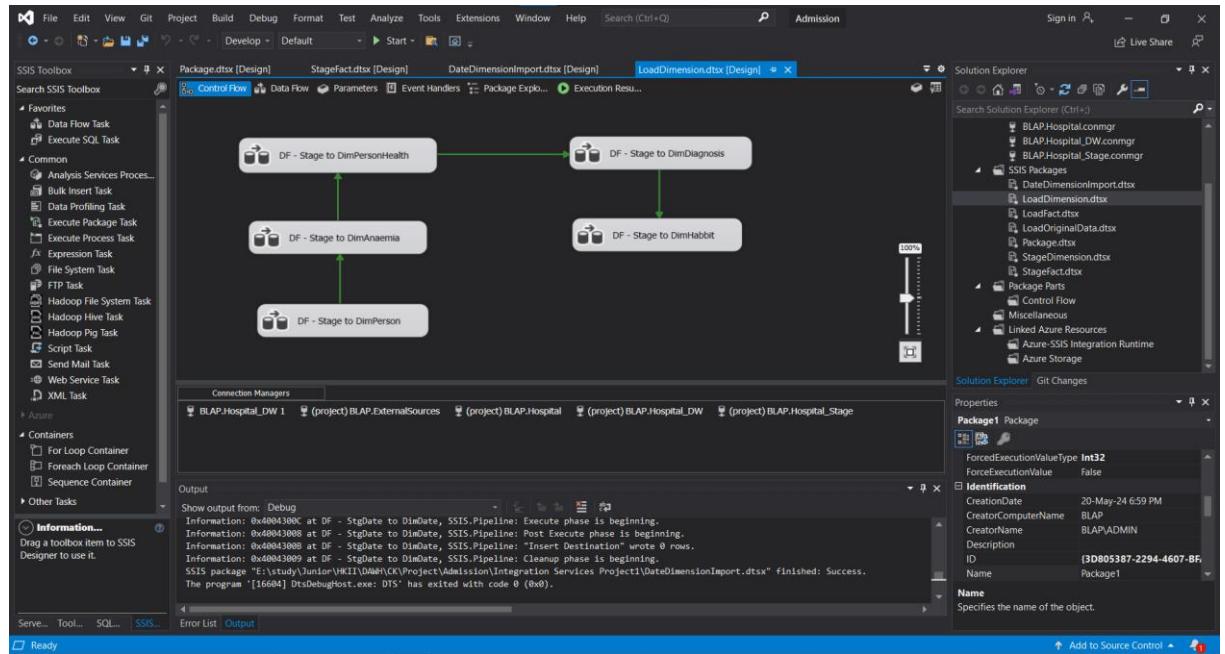


3.3.3. Load dữ liệu vào các Dimension

1. Click vào **LoadDimension.dtsx** và tiến hành thêm vào 6 Data Flow Task như hình sau:



2. Đổi tên các task như sau:



3. Tiến hành load các **Dimension** sau:

3.3.3.1. DimPerson

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

The image contains two side-by-side screenshots of the Microsoft SQL Server Integration Services (SSIS) Source and Destination Assistants.

Top Screenshot (Source Assistant - Add New Source):

- Select source type:** SQL Server, Excel, Flat File, Oracle.
- Select connection managers:** New..., BLAP.Hospital, BLAP.Hospital_DW, **BLAP.Hospital_Stage** (highlighted in blue).
- Show only installed source types:**
- Data source configuration:**

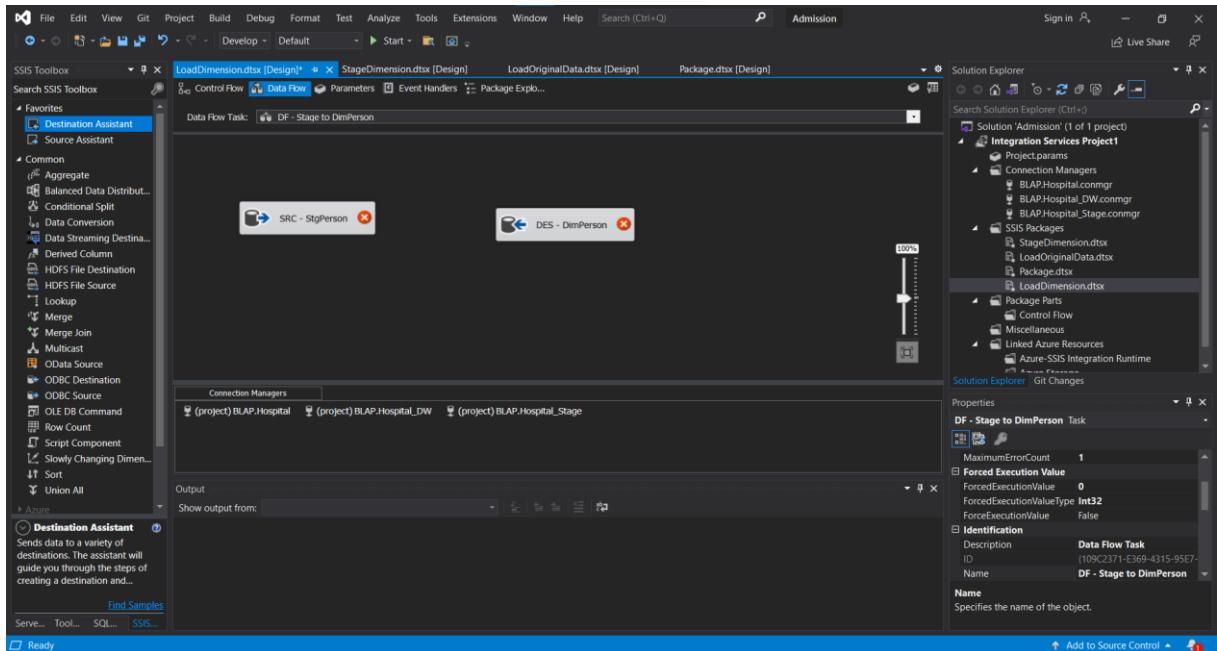
```
data source = BLAP
initial catalog = Hospital_Stage
provider = MSOLEDBSQL.1
integrated security = SSPI
application name = SSIS-Integration Services
Project1-{3B8DDE6D-2C87-45F0-98ED-03D63D932D45}BLAP.Hospital_Stage
auto translate = False
```
- Buttons:** OK, Cancel.

Bottom Screenshot (Destination Assistant - Add New Destination):

- Select destination type:** SQL Server, Excel, Oracle.
- Select connection managers:** New..., BLAP.Hospital, **BLAP.Hospital_DW** (highlighted in blue), BLAP.Hospital_Stage.
- Show only installed destination types:**
- Data source configuration:**

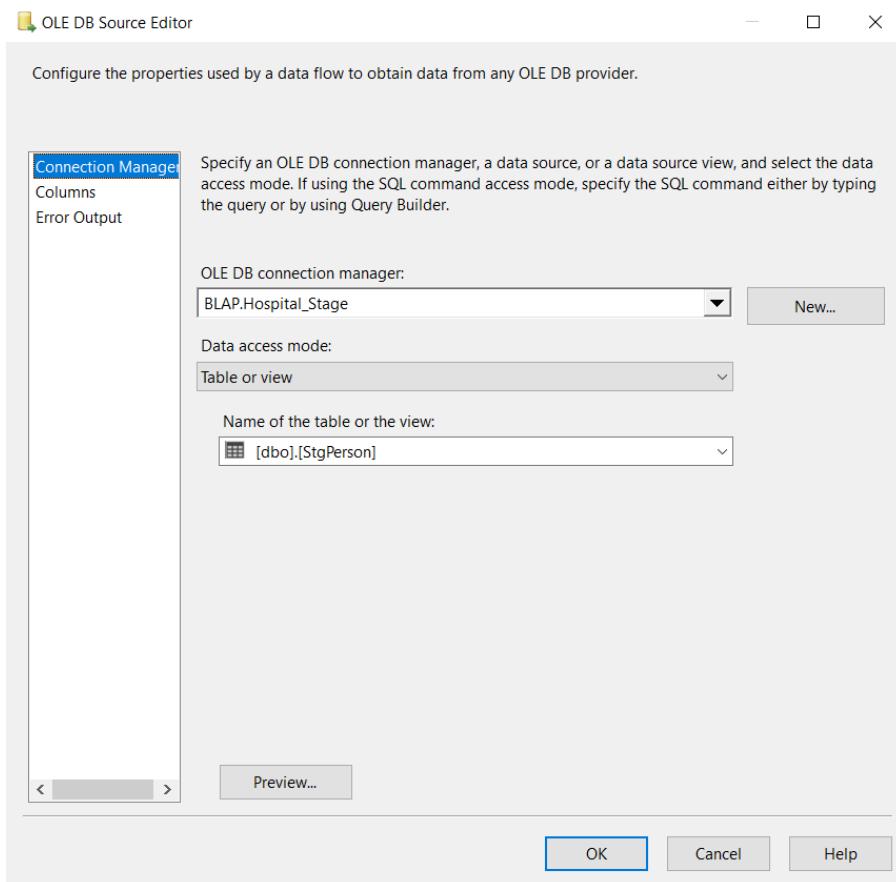
```
data source = BLAP
initial catalog = Hospital_DW
provider = MSOLEDBSQL.1
integrated security = SSPI
application name = SSIS-Integration Services
Project1-{D8259DD8-8870-4375-894E-D0CB615BC29B}BLAP.Hospital_DW
auto translate = False
```
- Buttons:** OK, Cancel.

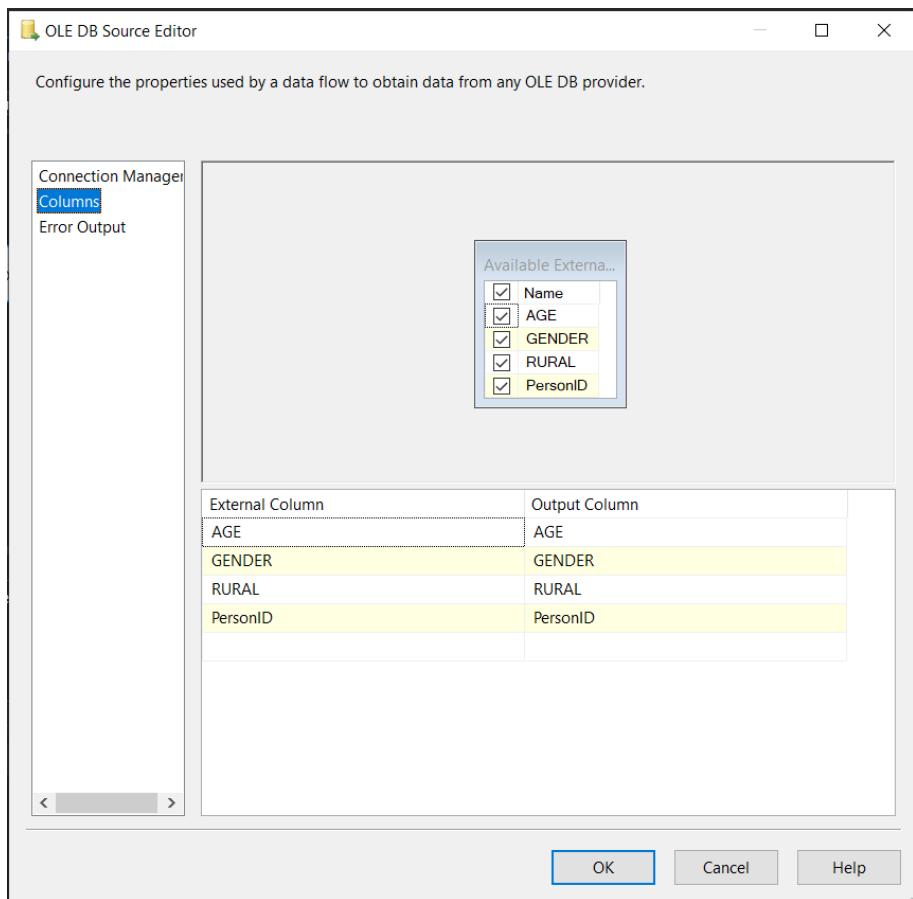
2. Tiến hành đổi tên và ta có được hình như sau:



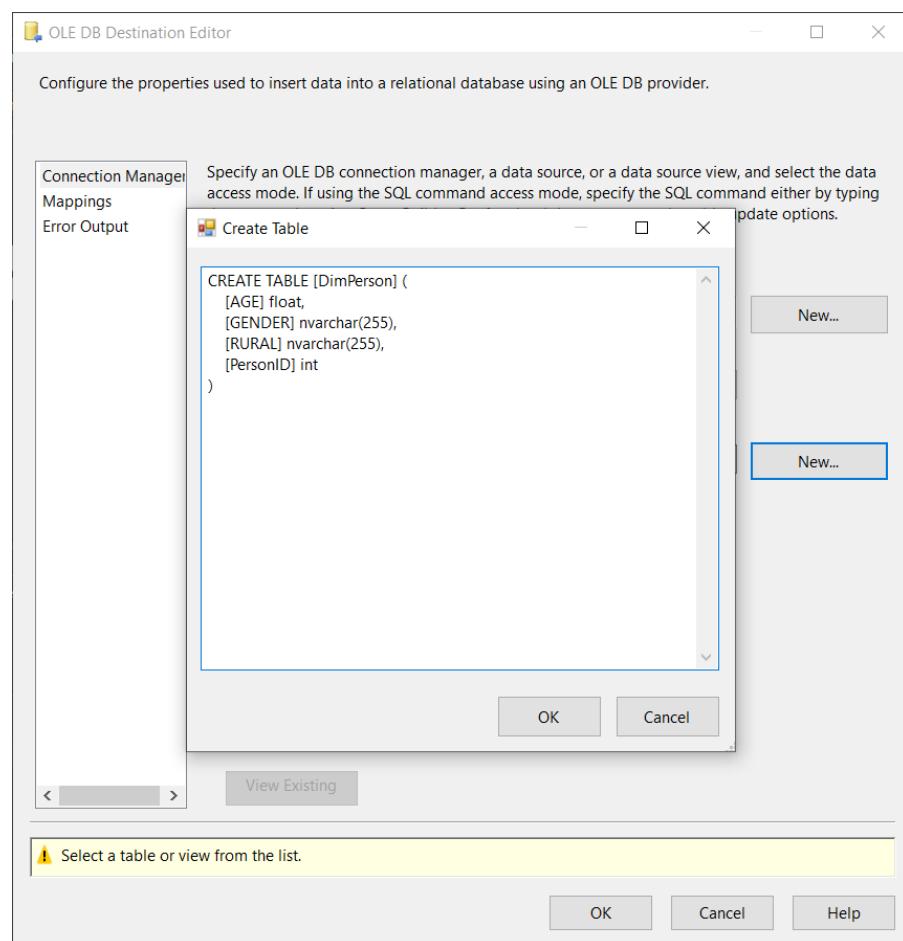
3. Tiến hành cấu hình các task:

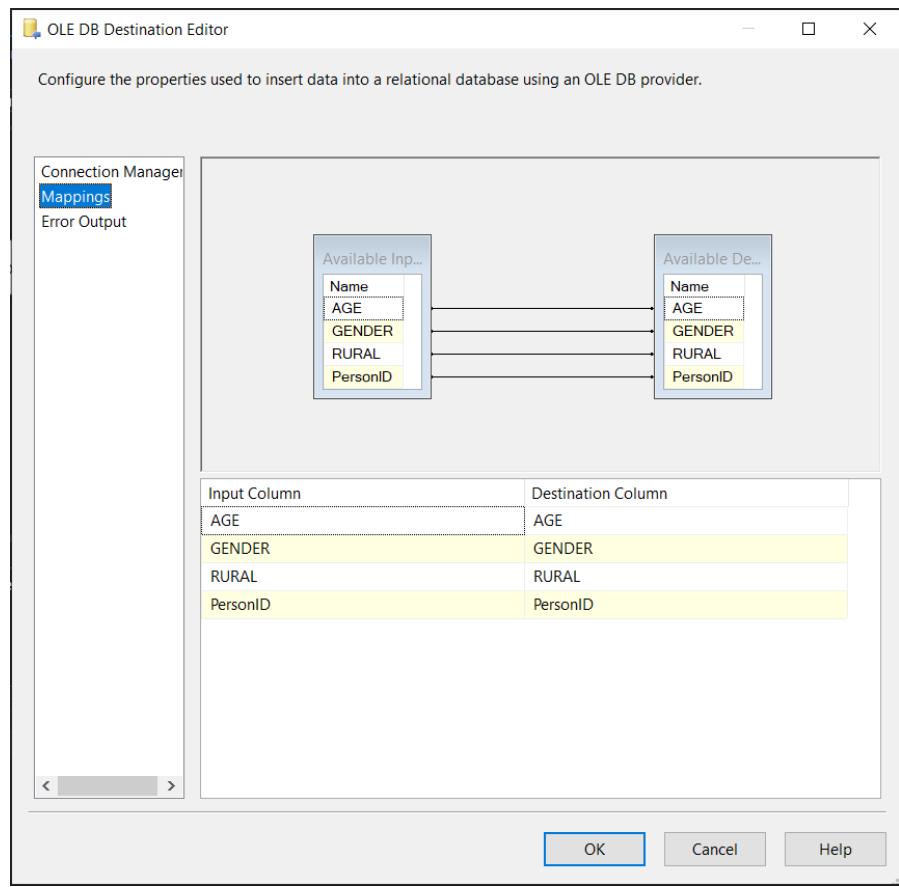
- Cấu hình Source Assistant:



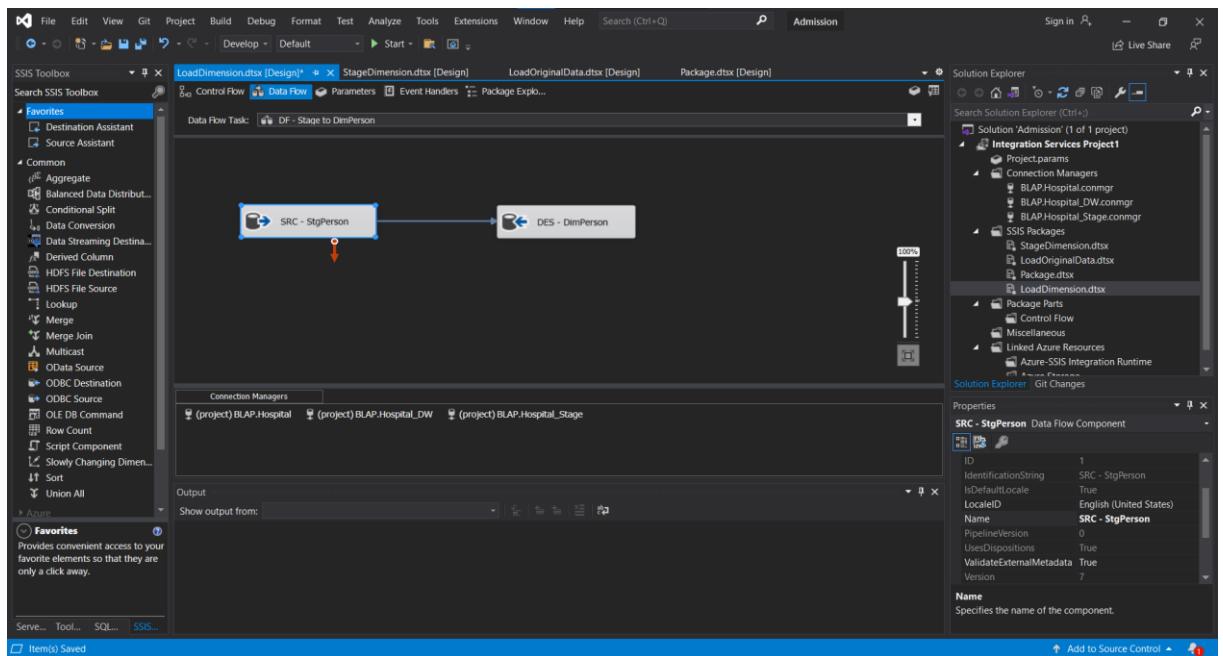


- Cấu hình Destination Assistant:



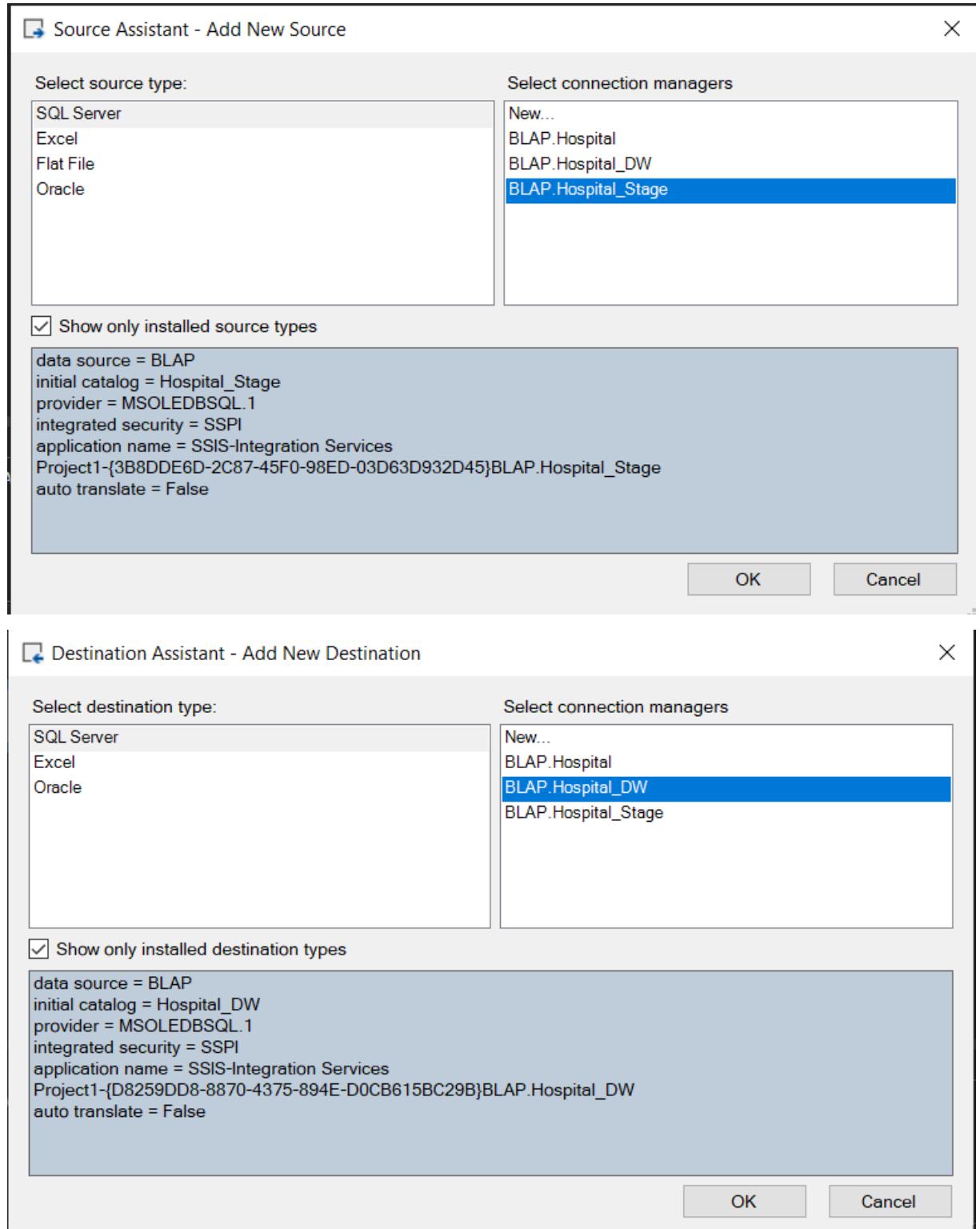


4. Data Flow Task của ta sẽ trông như sau:

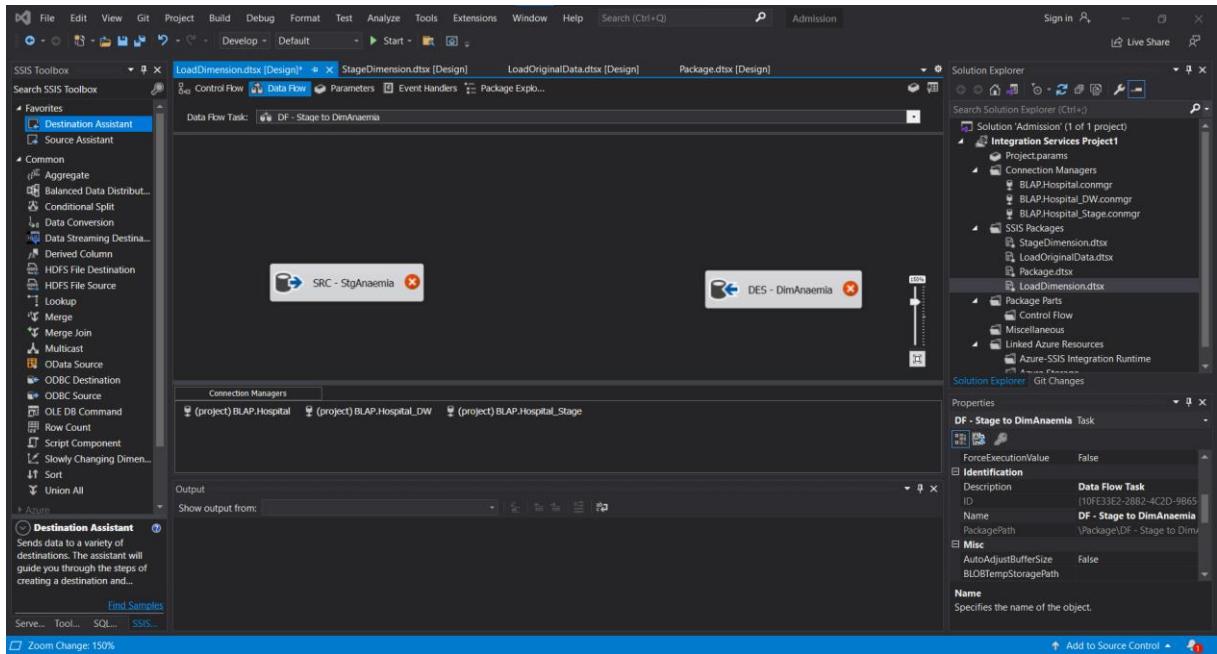


3.3.3.2. DimAnaemia

- Click vào DF - Stage to DimAnaemia và thêm Source Assistant và Destination Assistant.

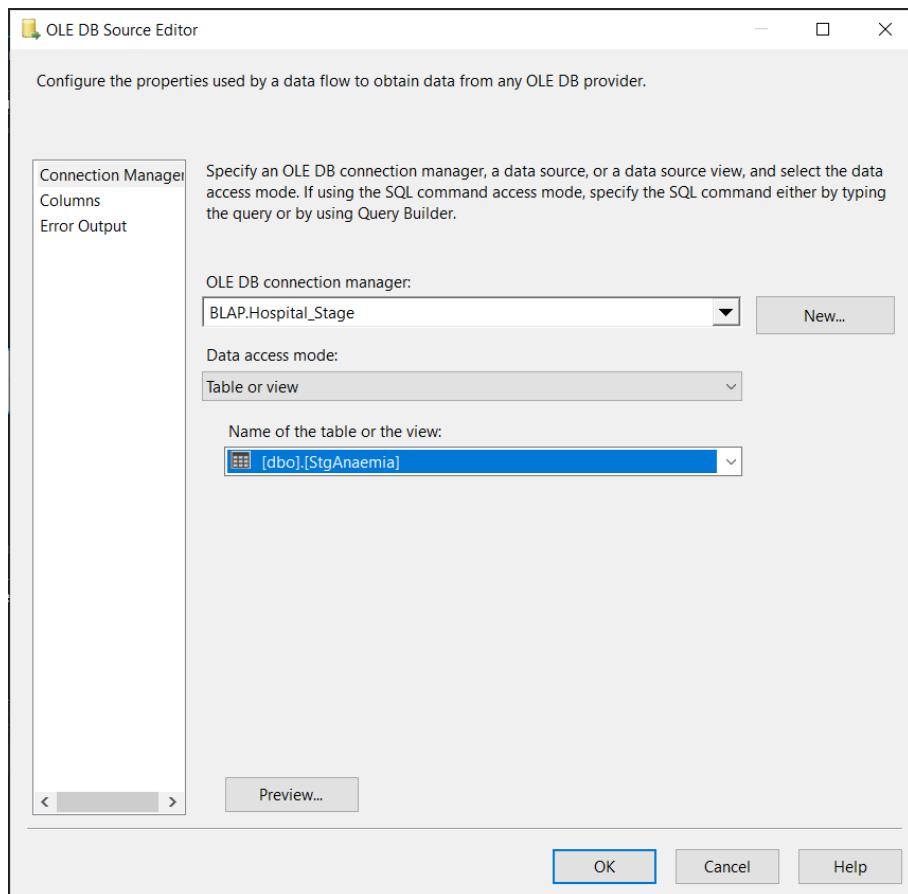


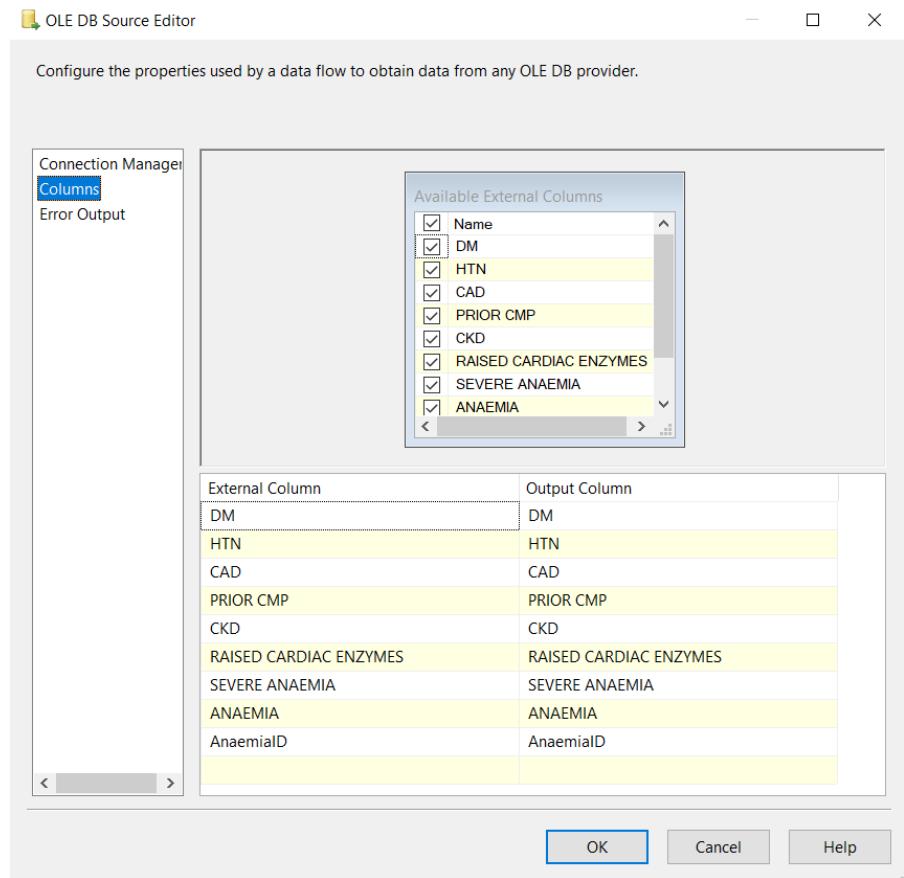
2. Tiến hành đổi tên và ta có được hình như sau:



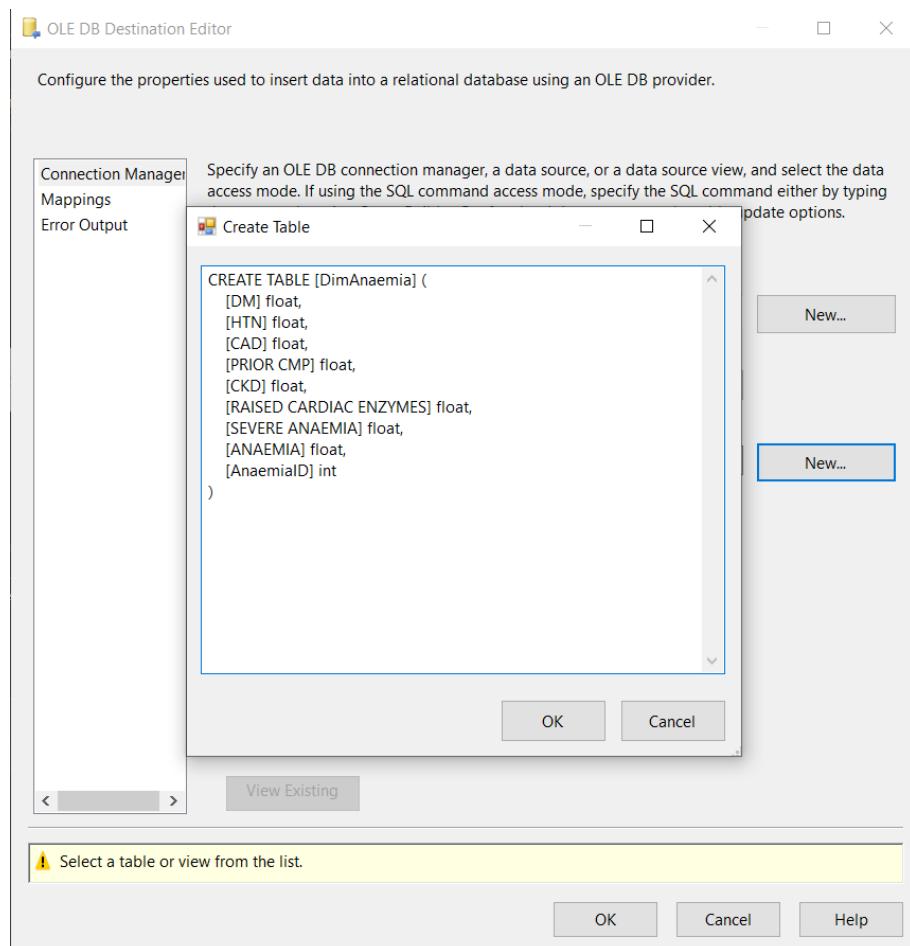
3. Tiến hành cấu hình các task:

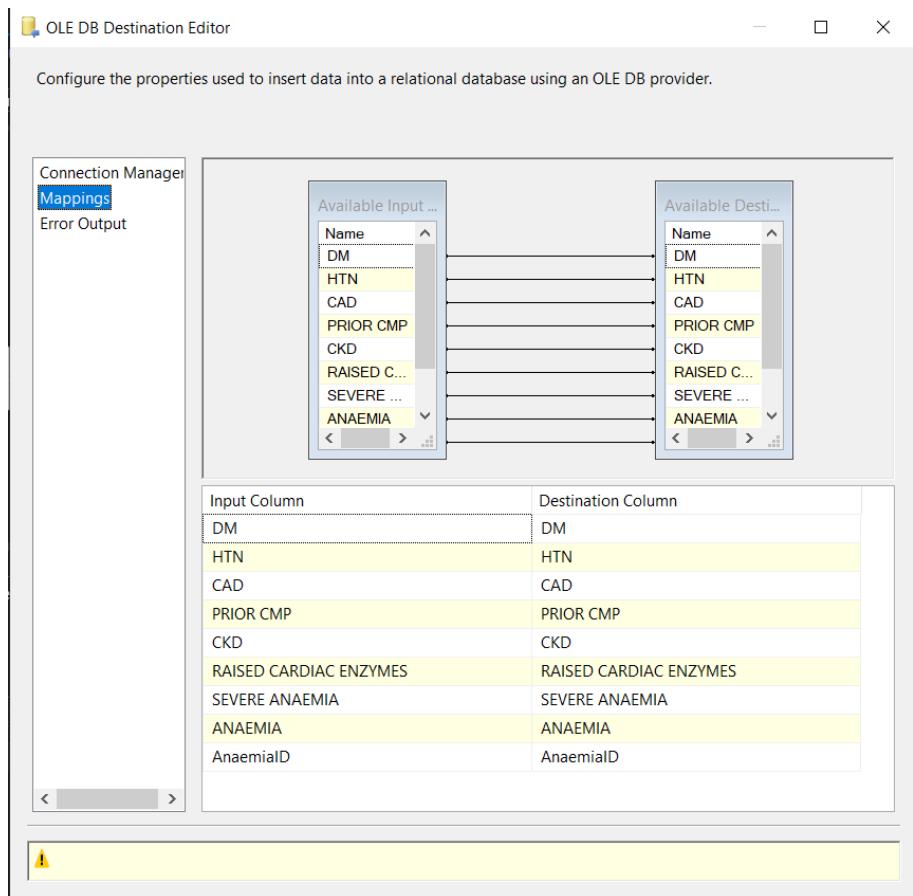
- Cấu hình Source Assistant:



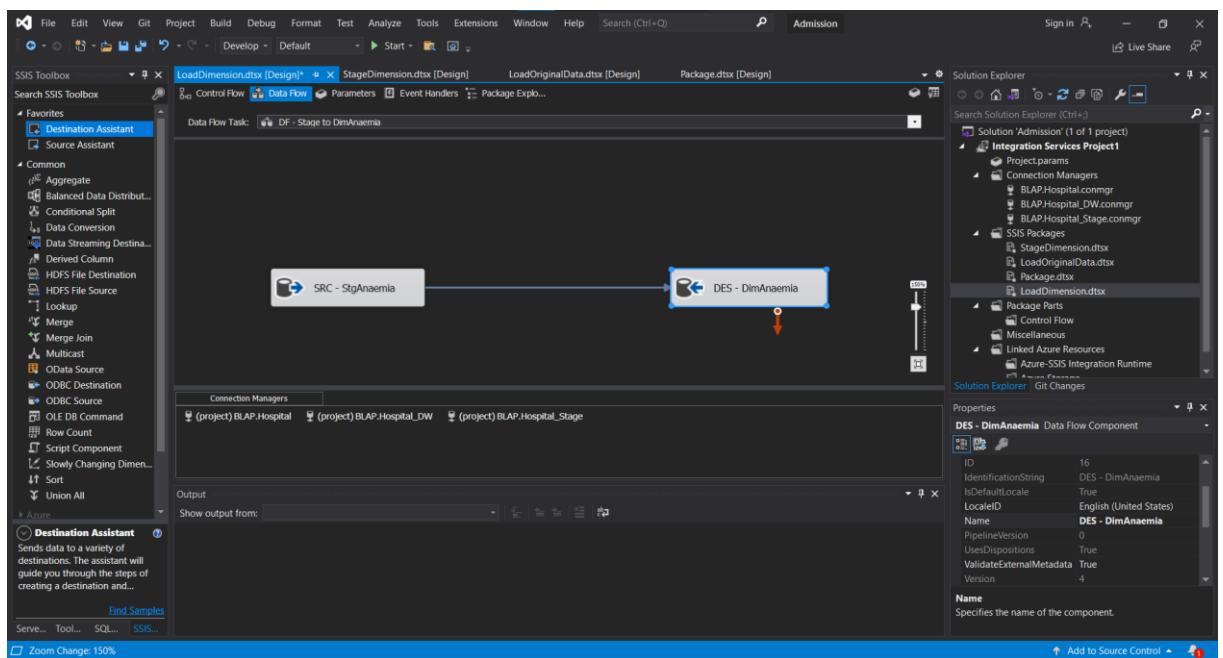


- Cấu hình Destination Assistant:



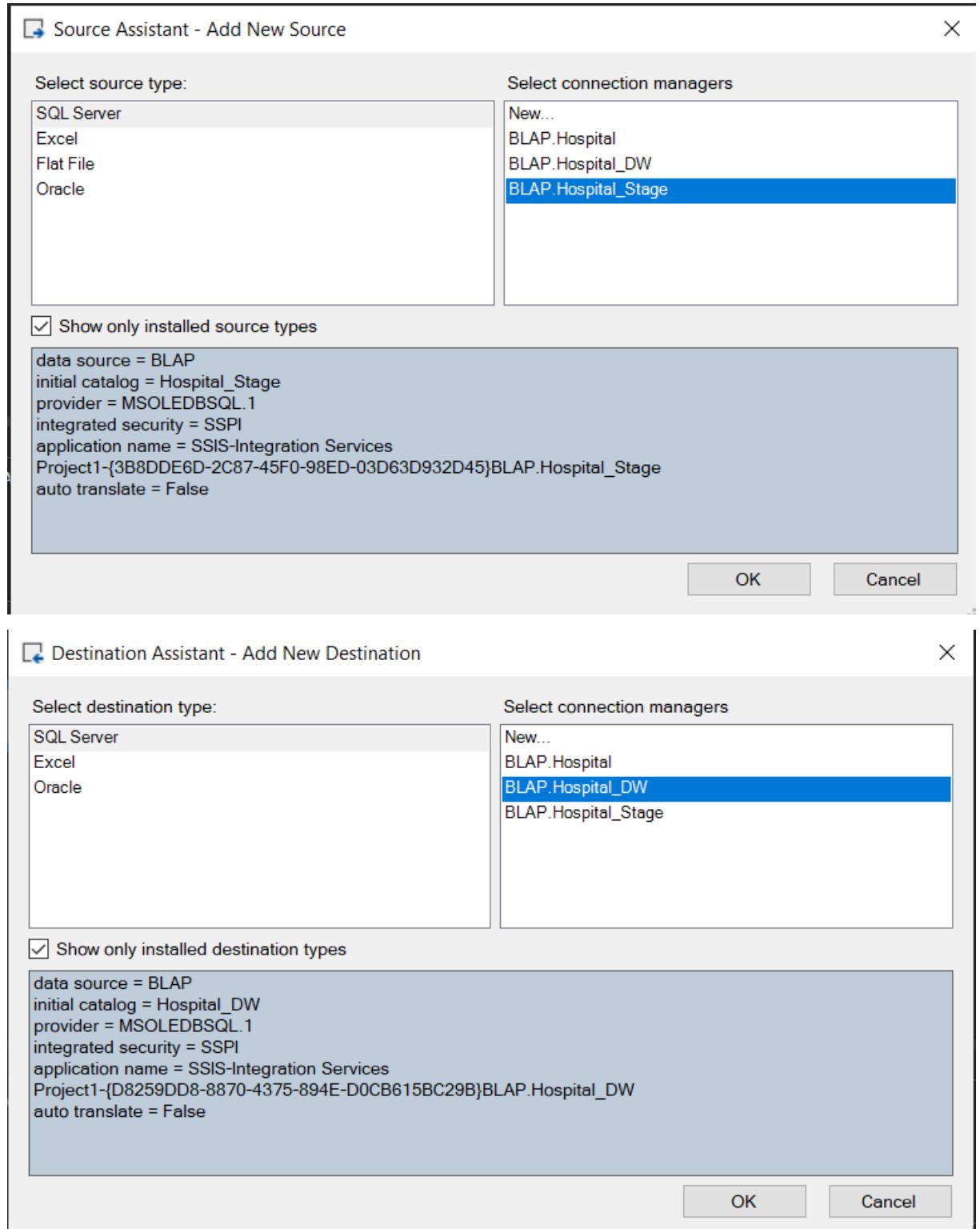


4. Data Flow Task của ta sẽ trông như sau:

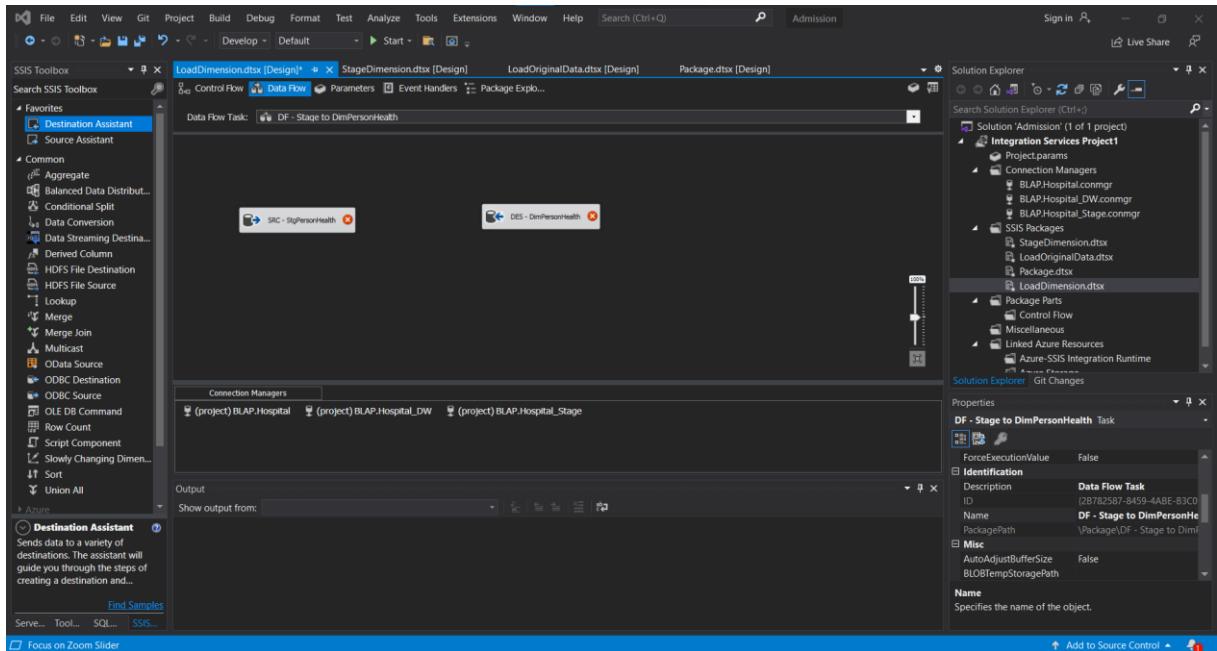


3.3.3.3. DimPersonHealth

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

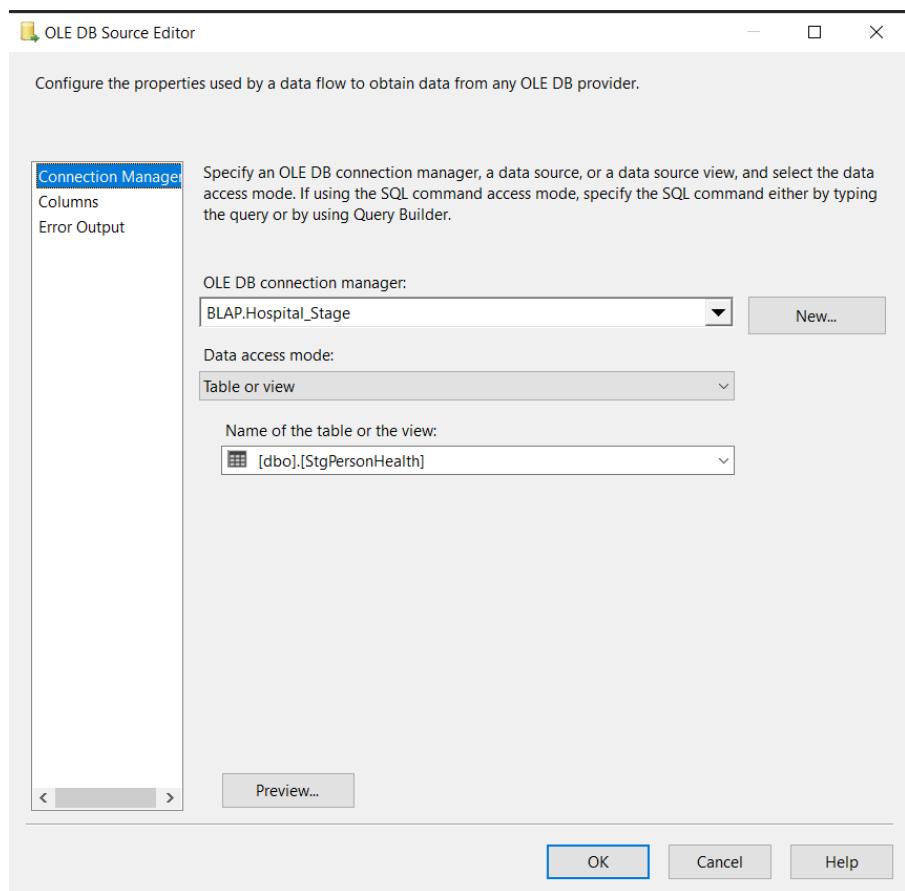


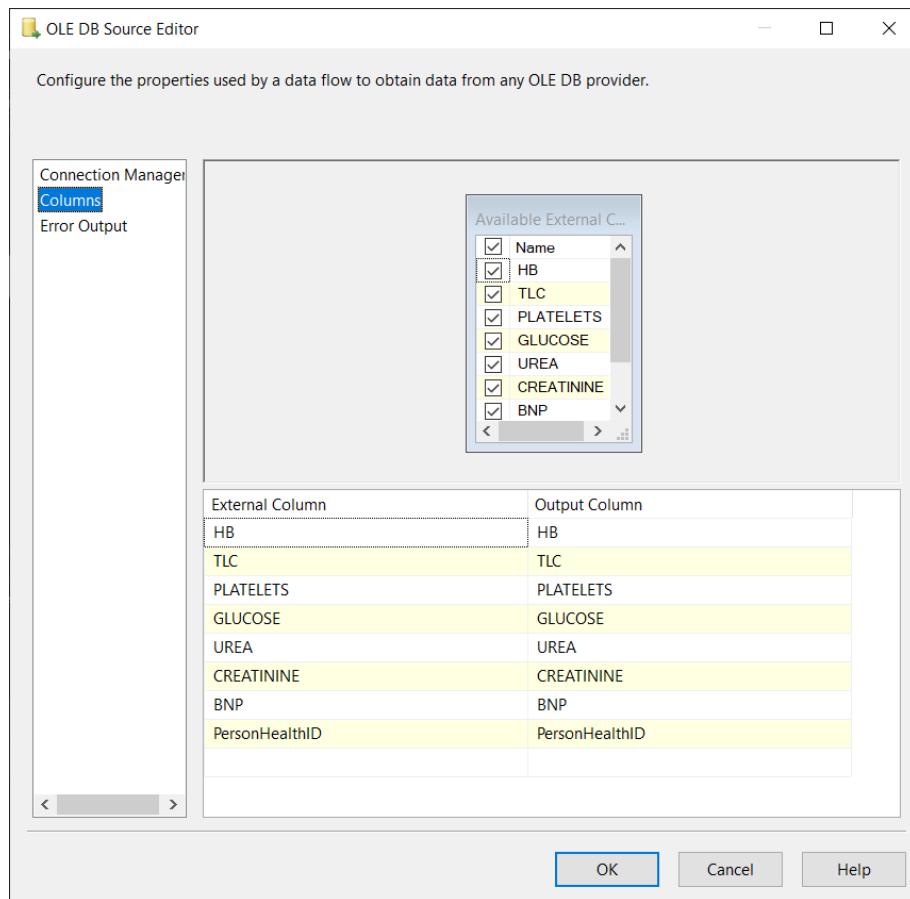
2. Tiến hành đổi tên và ta có được hình như sau:



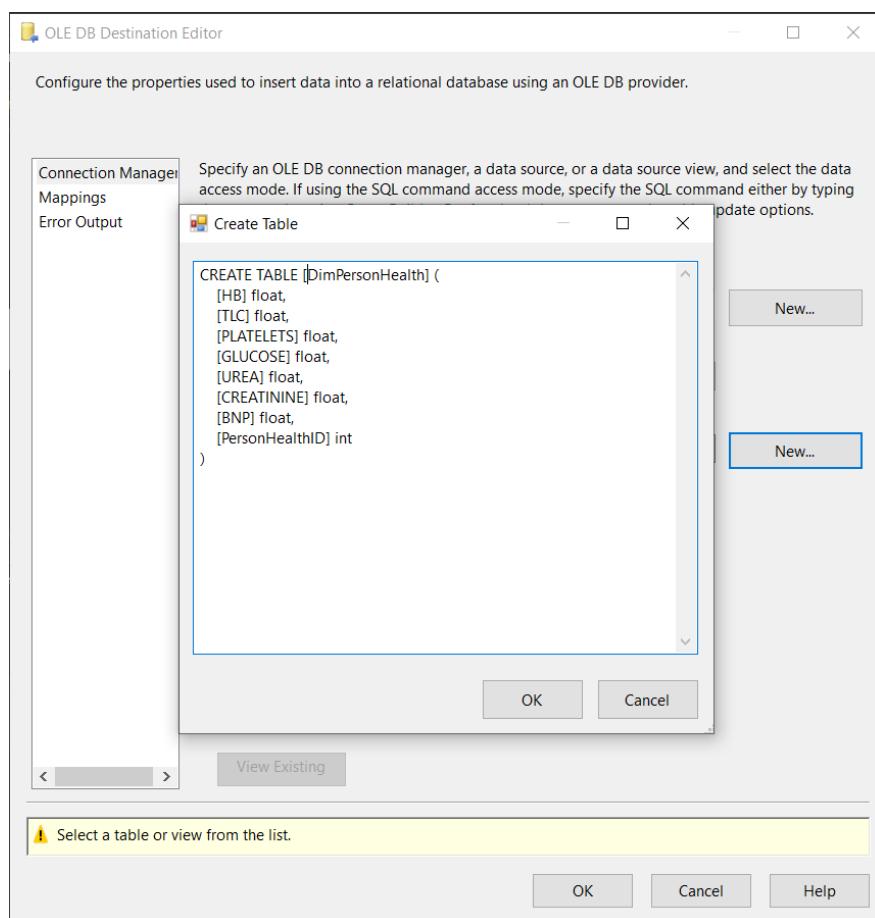
3. Tiến hành cấu hình các task:

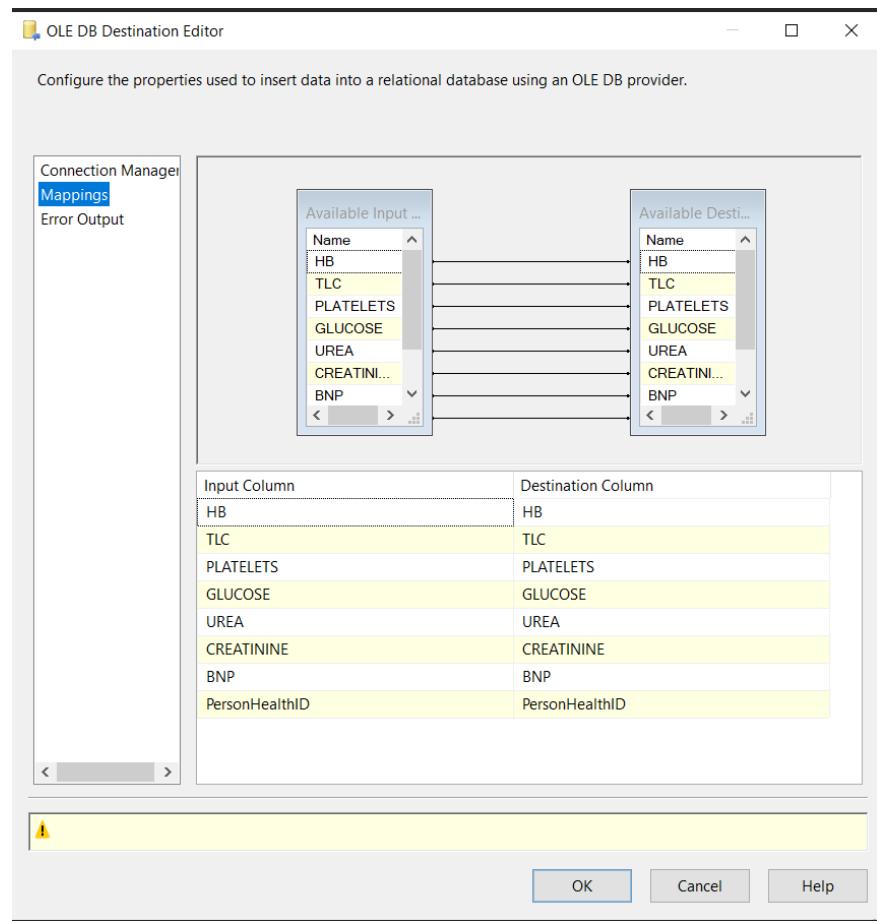
- Cấu hình Source Assistant:



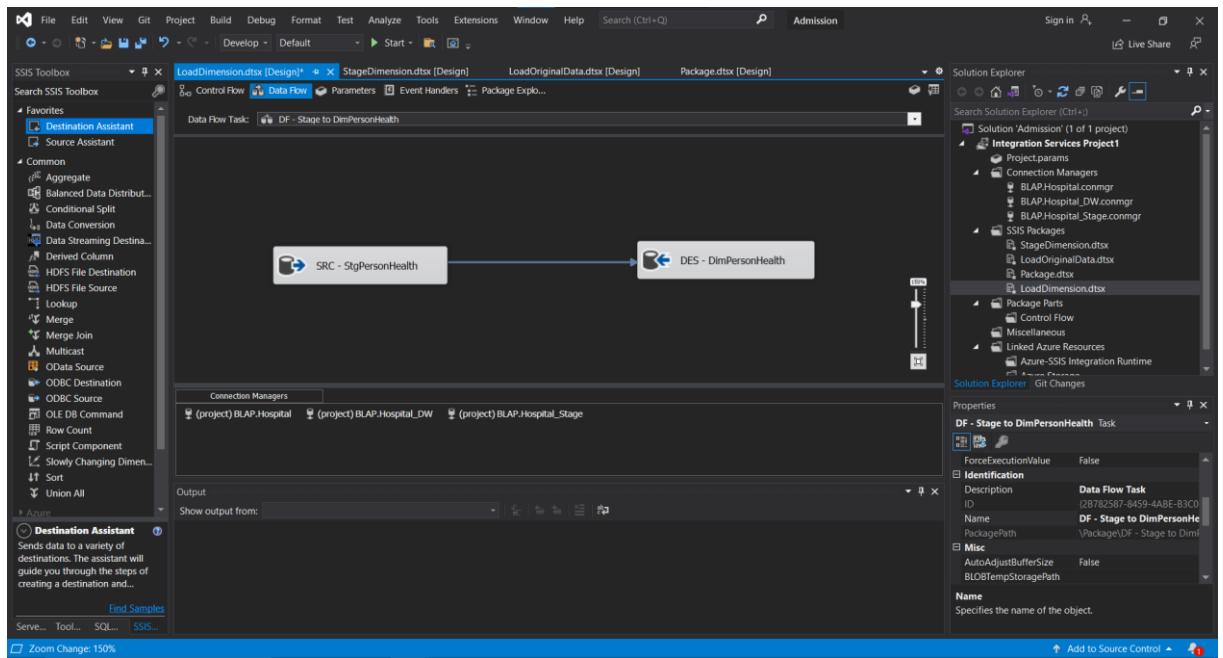


- Cấu hình Destination Assistant:



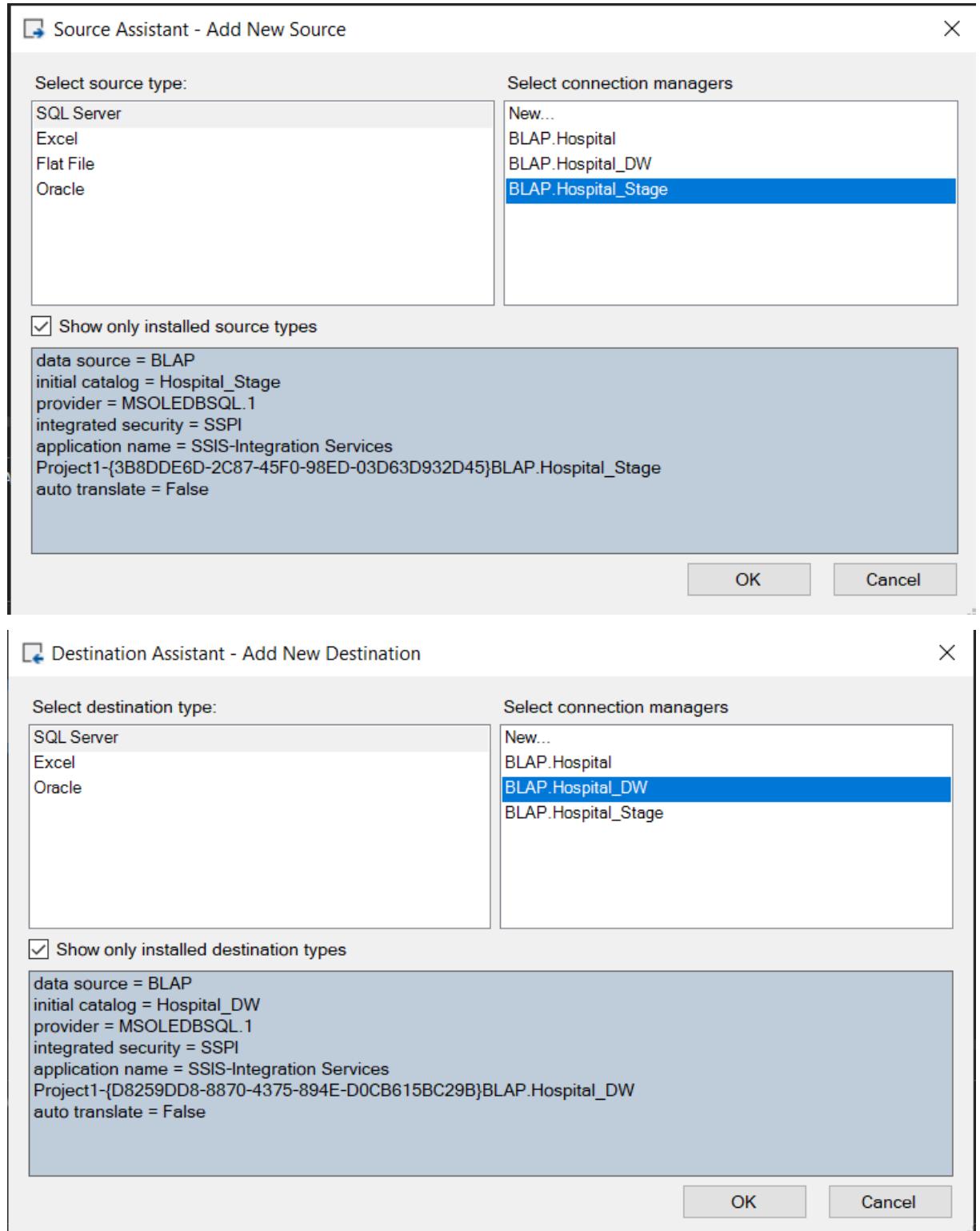


4. Data Flow Task của ta sẽ trông như sau:

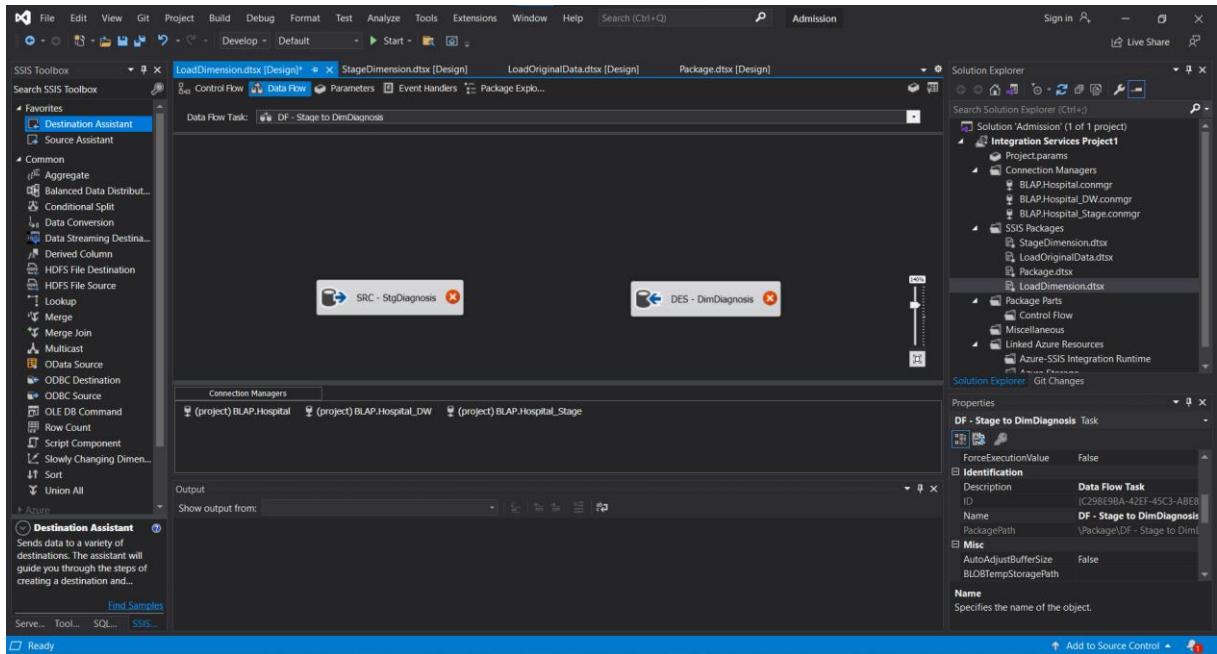


3.3.3.4. DimDiagnosis

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

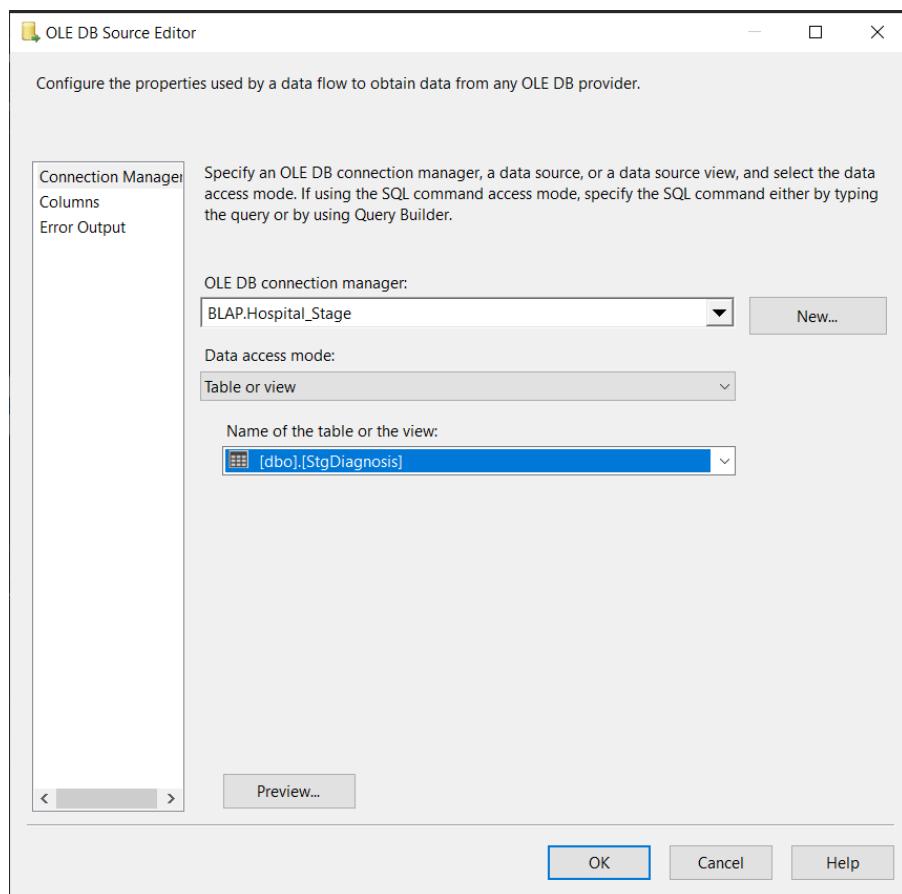


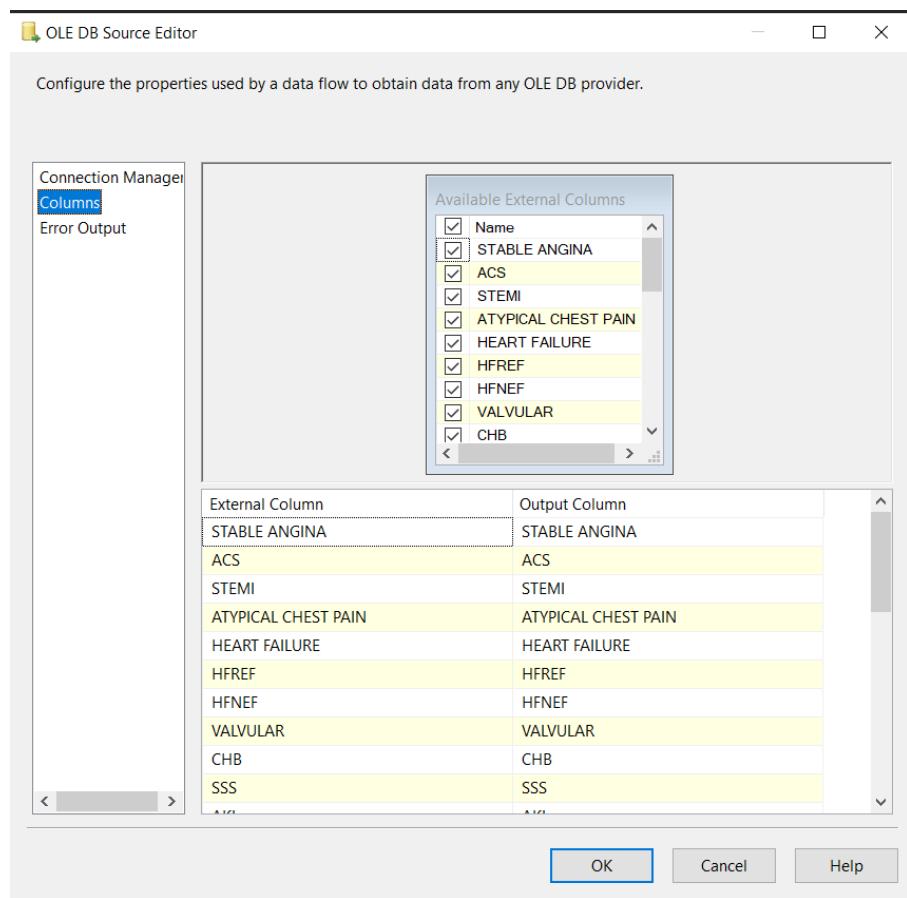
2. Tiến hành đổi tên và ta có được hình như sau:



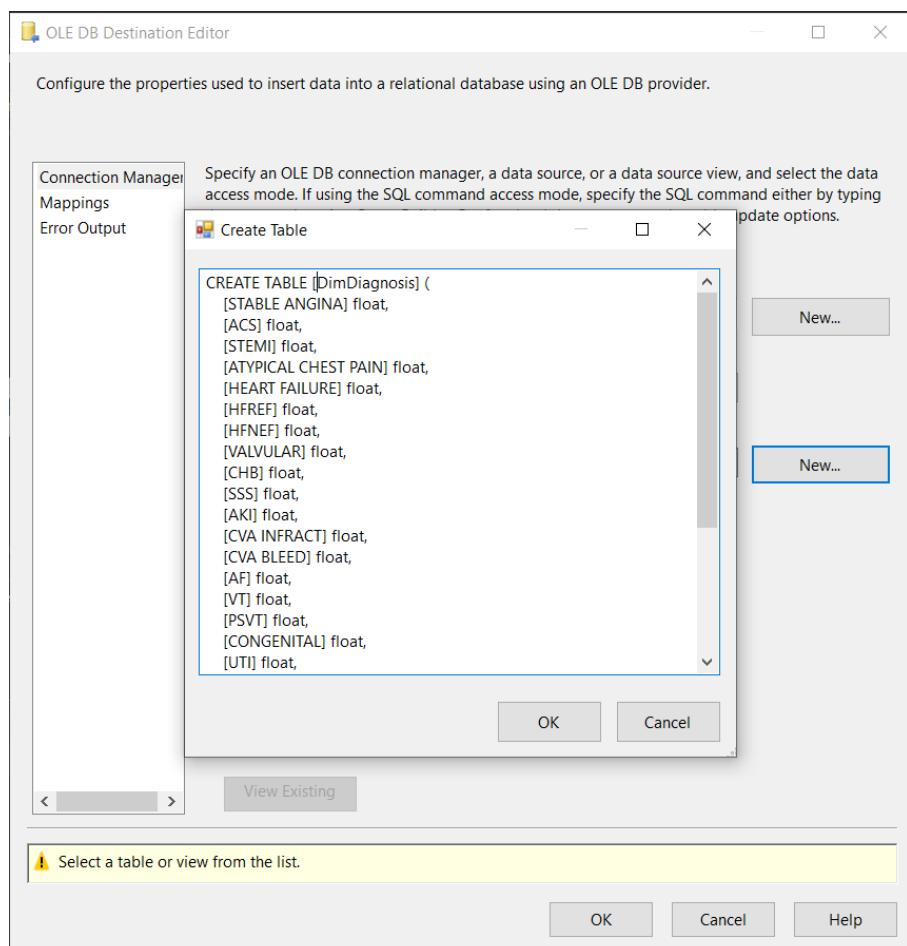
3. Tiến hành cấu hình các task:

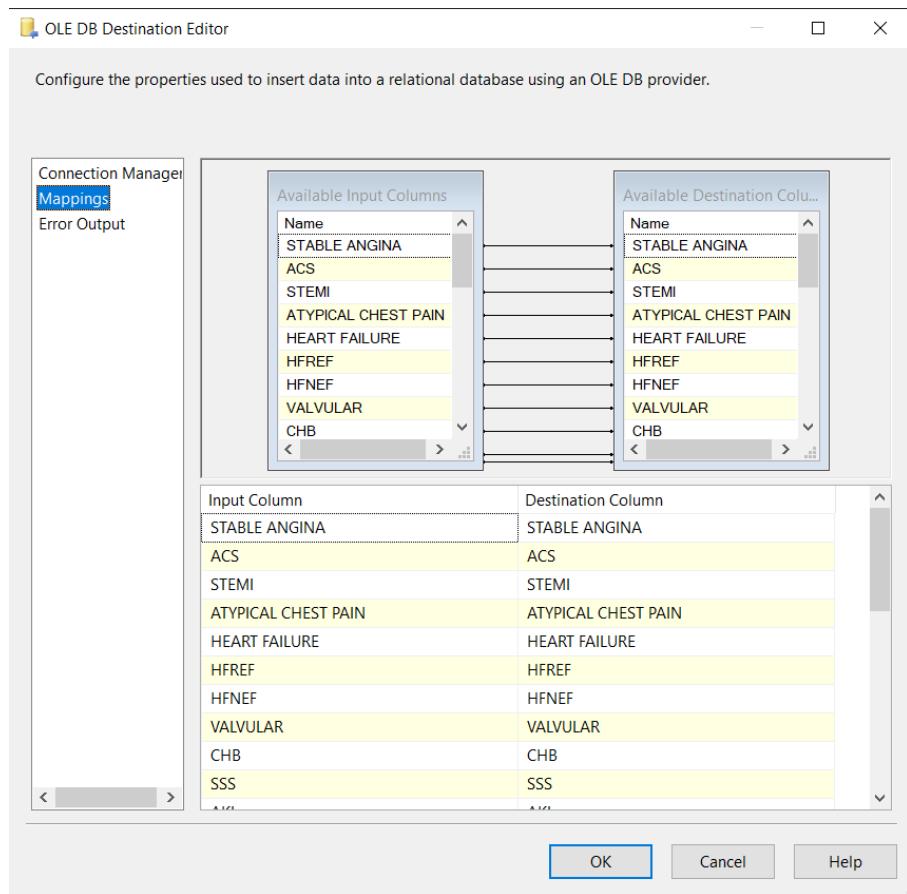
- Cấu hình Source Assistant:



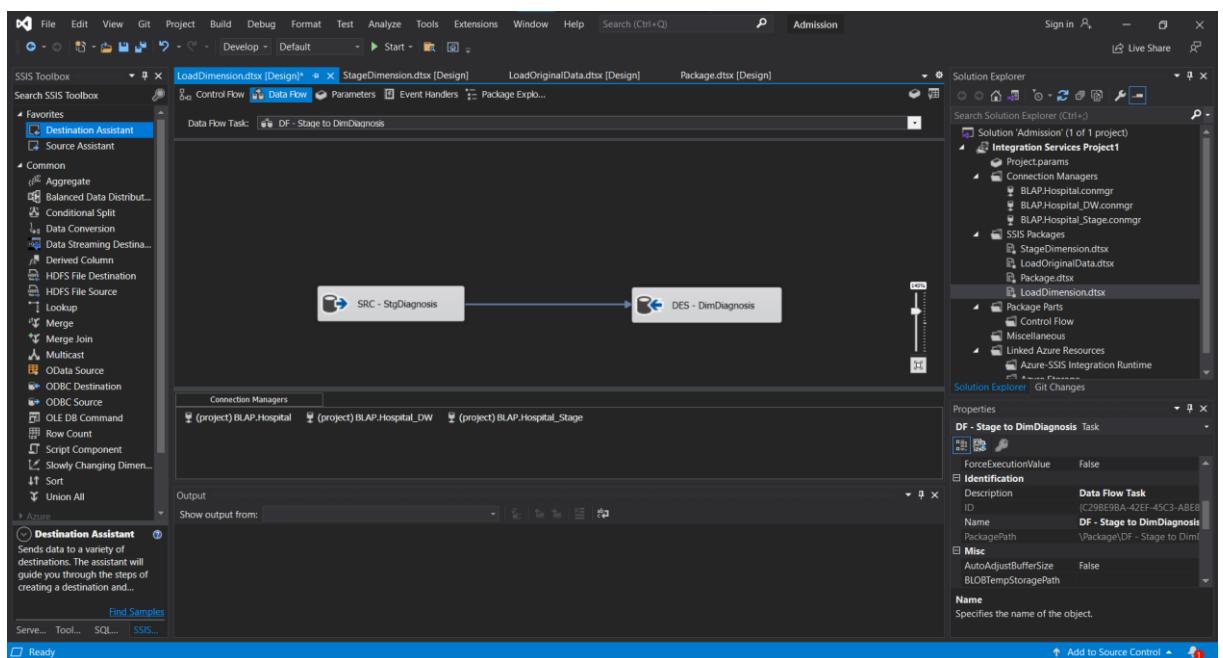


- Cấu hình Destination Assistant:



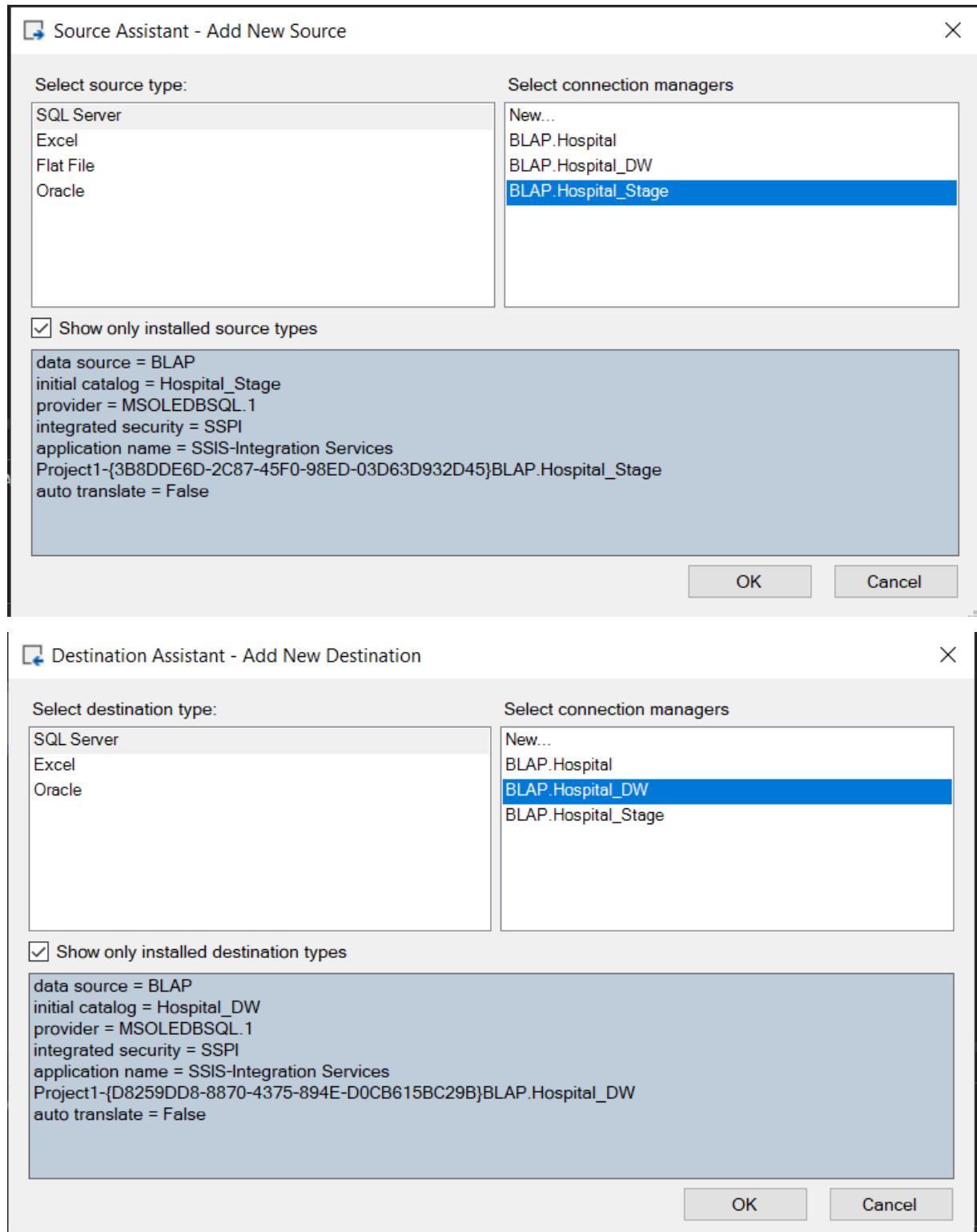


4. Data Flow Task của ta sẽ trông như sau:

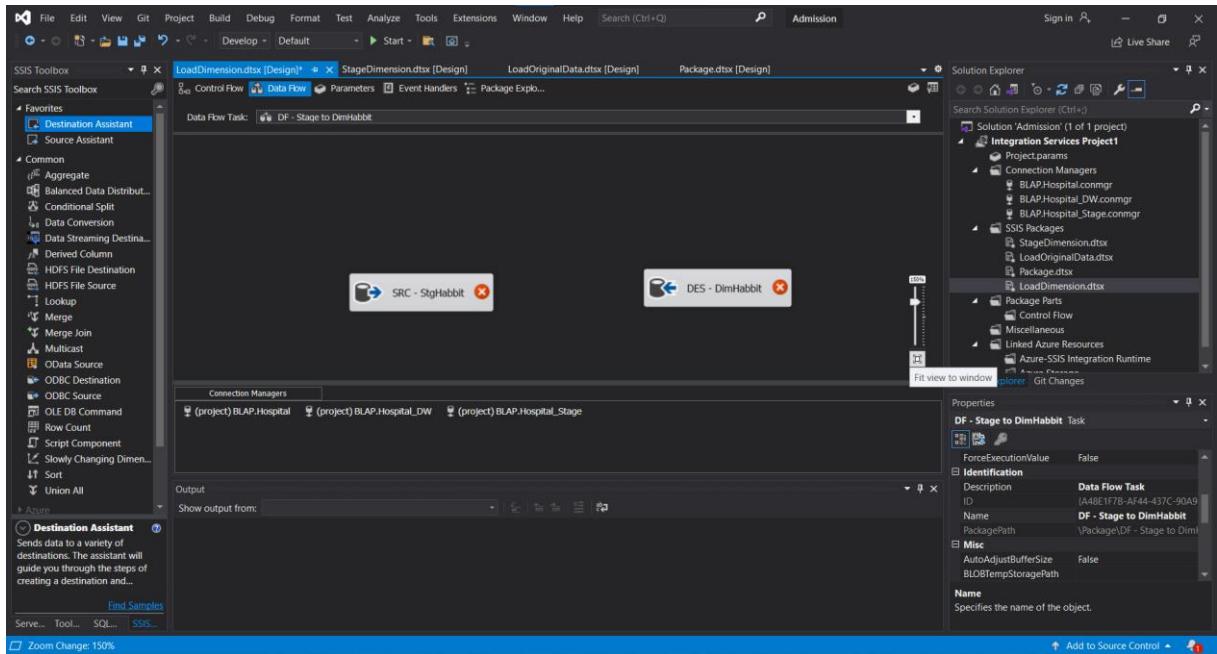


3.3.3.5. DimHabbit

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

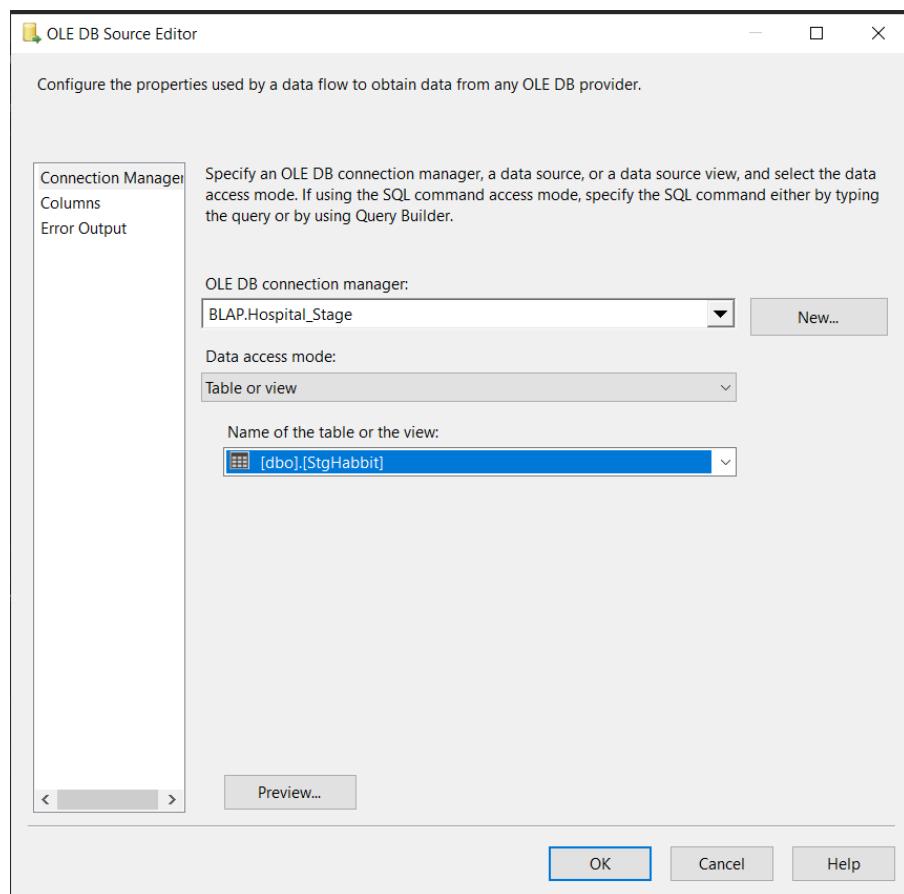


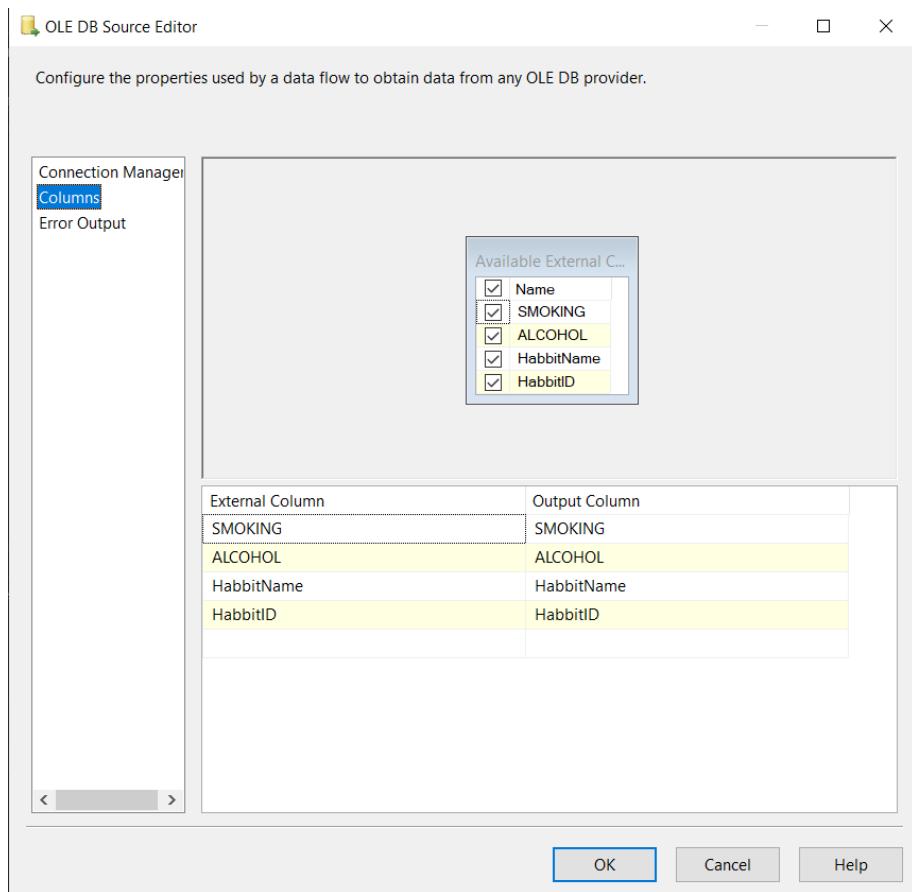
2. Tiến hành đổi tên và ta có được hình như sau:



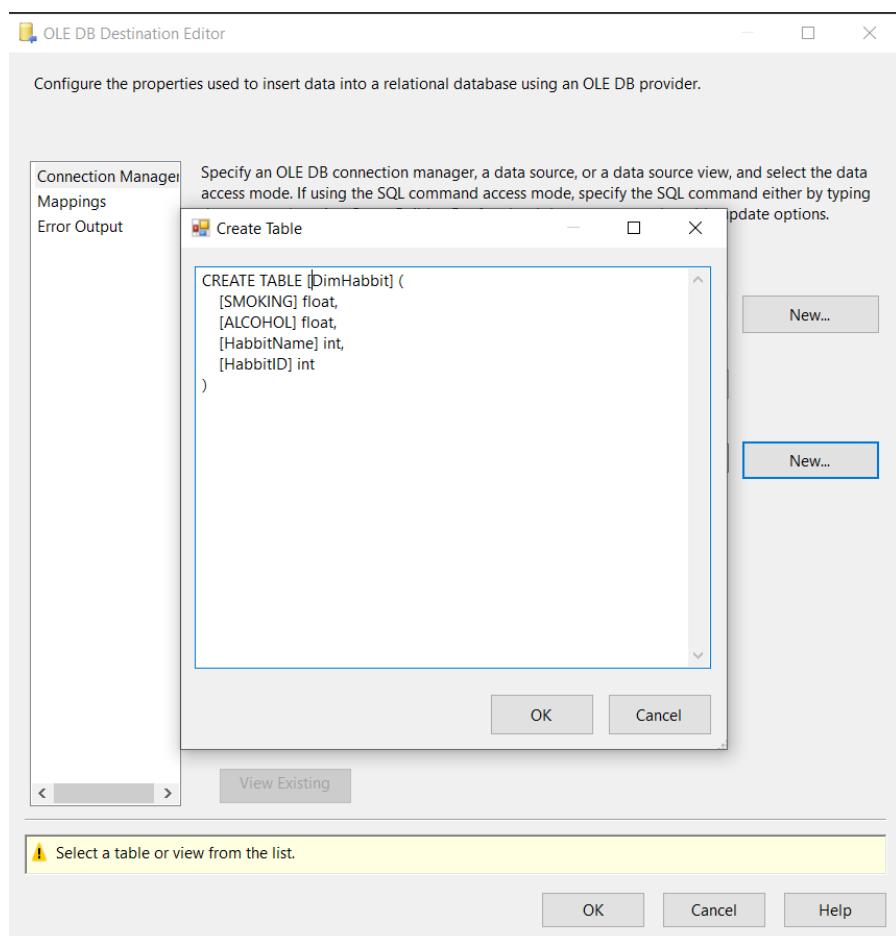
3. Tiến hành cấu hình các task:

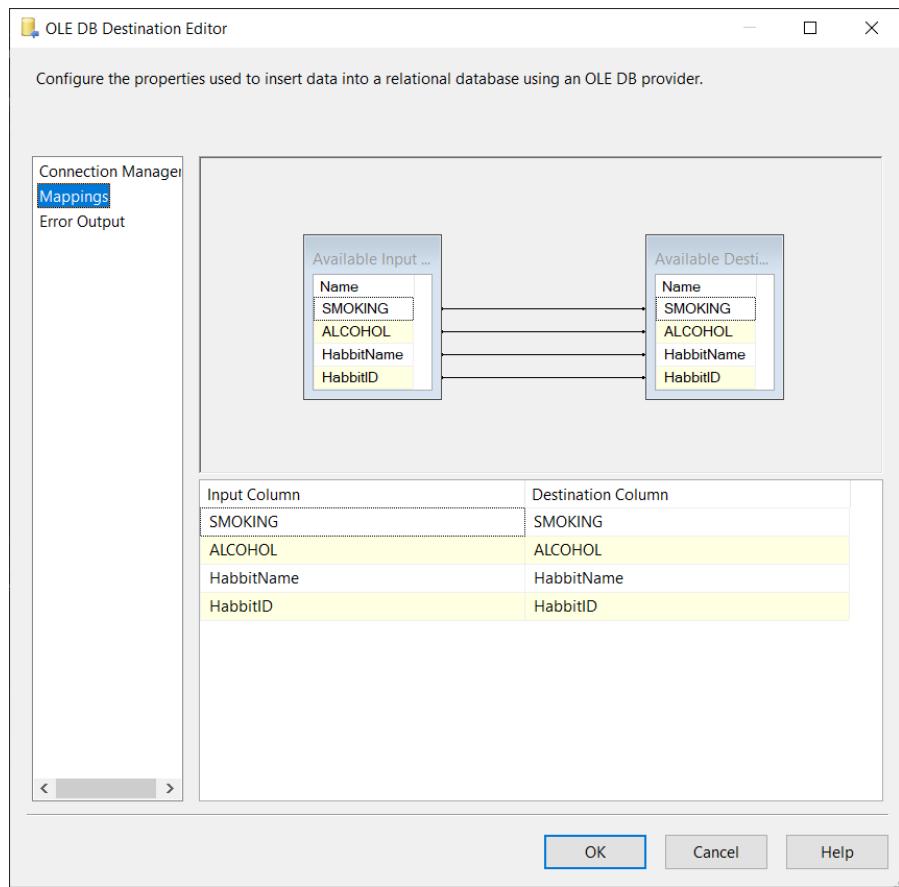
- Cấu hình Source Assistant:



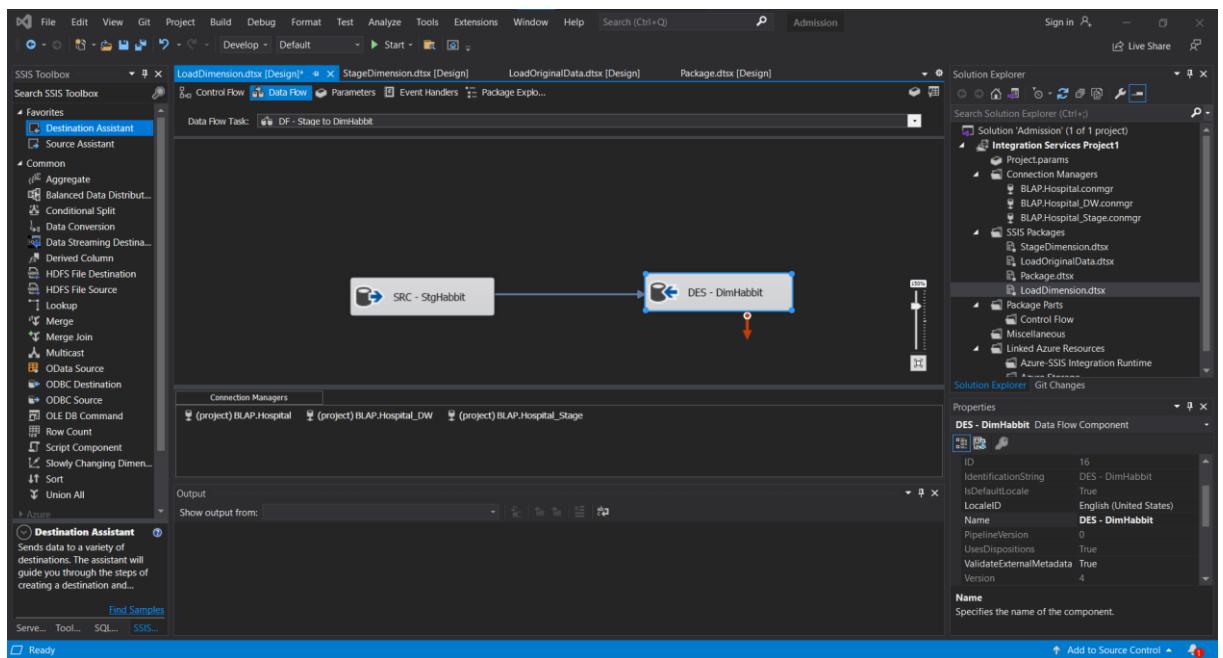


- Cấu hình Destination Assistant:



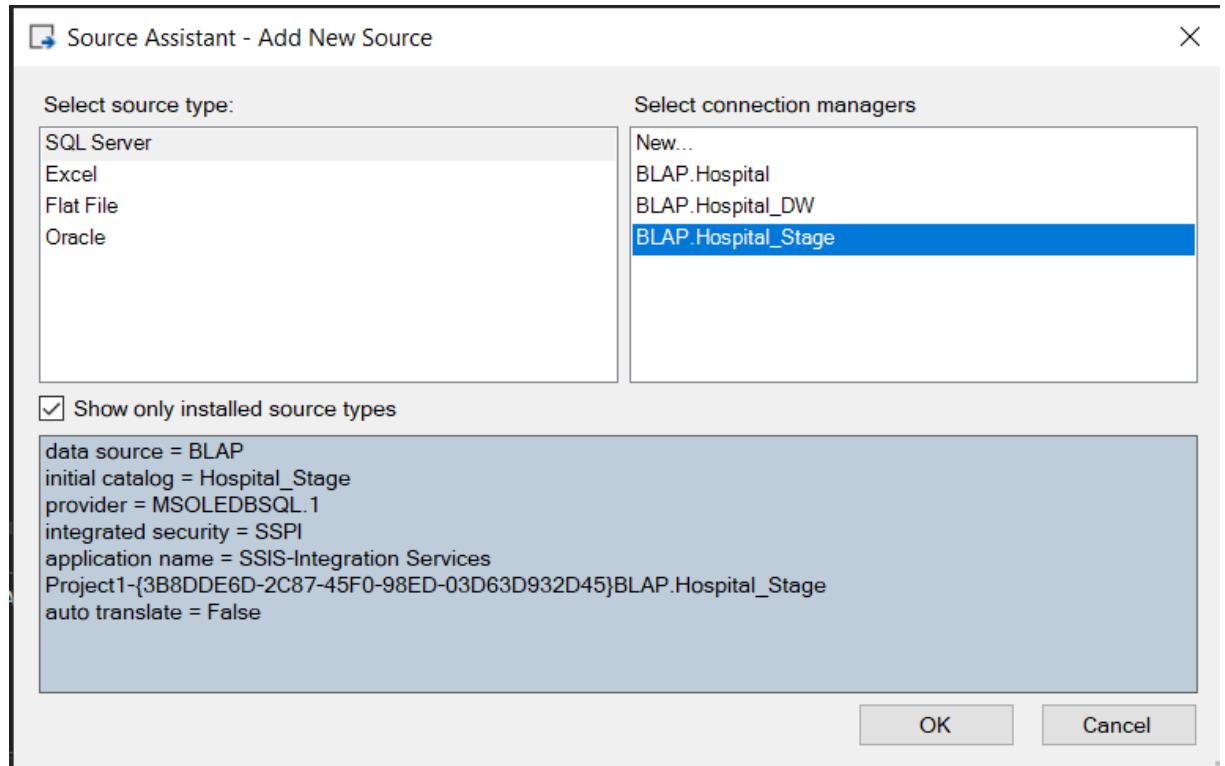


4. Data Flow Task của ta sẽ trông như sau:

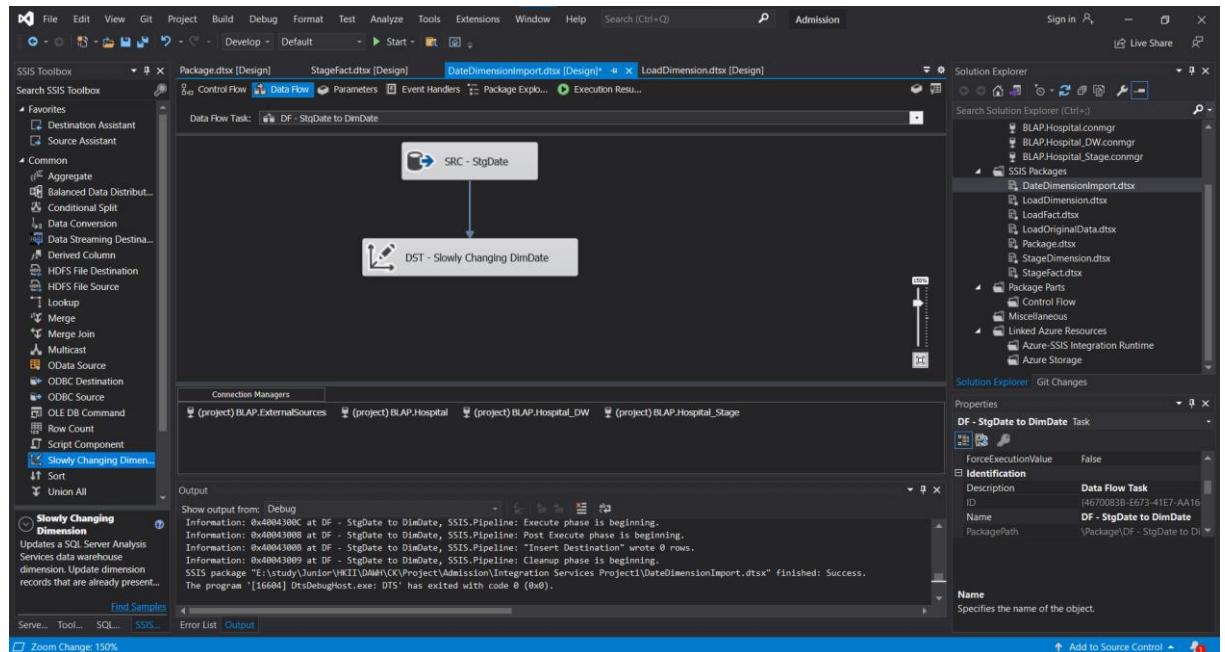


3.3.3.6. DimDate

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

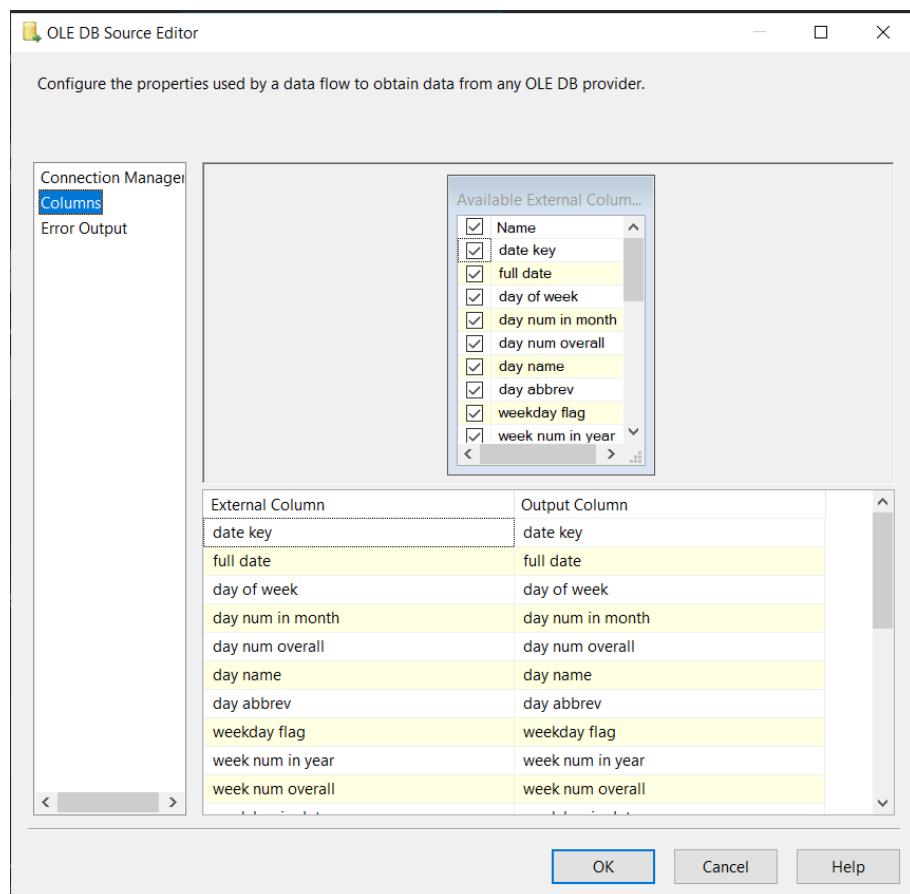
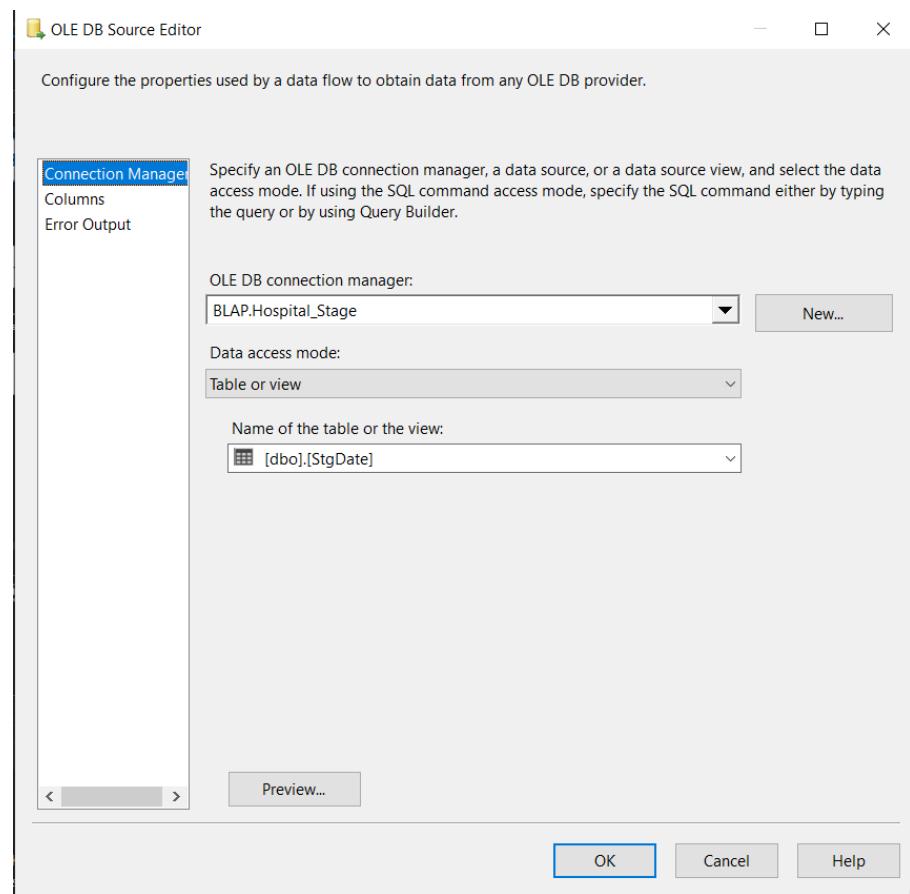


- Tiến hành đổi tên và ta có được hình như sau:

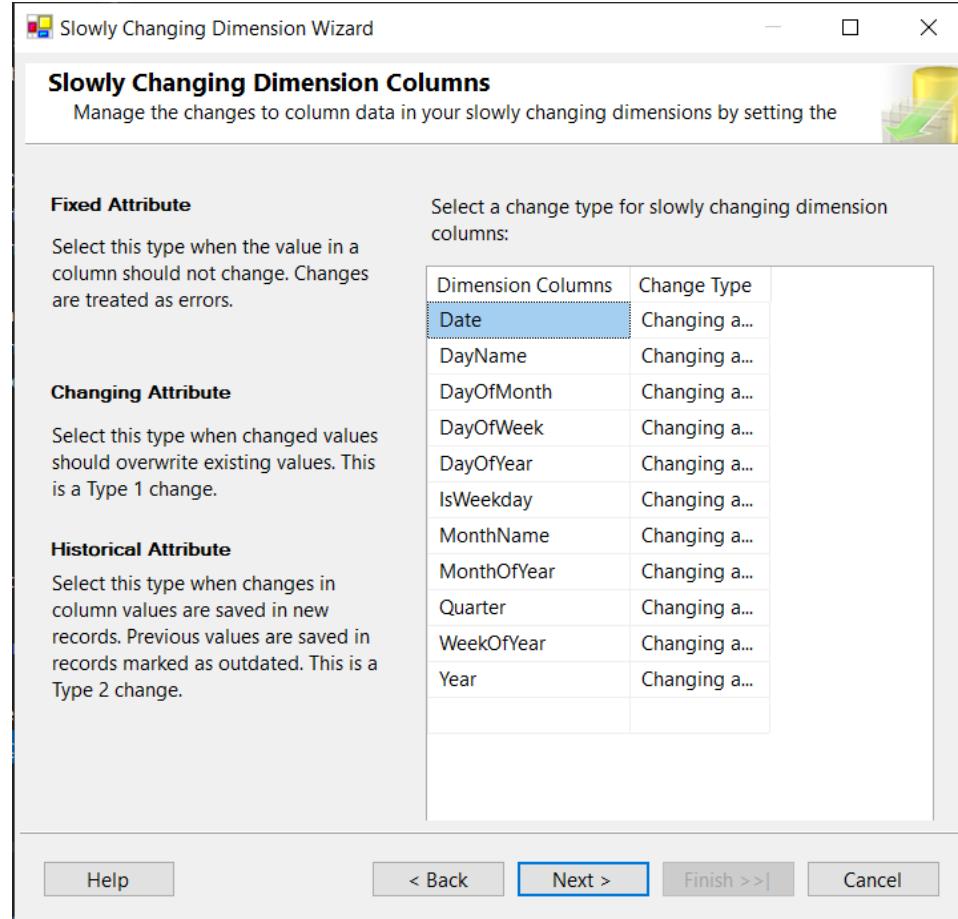
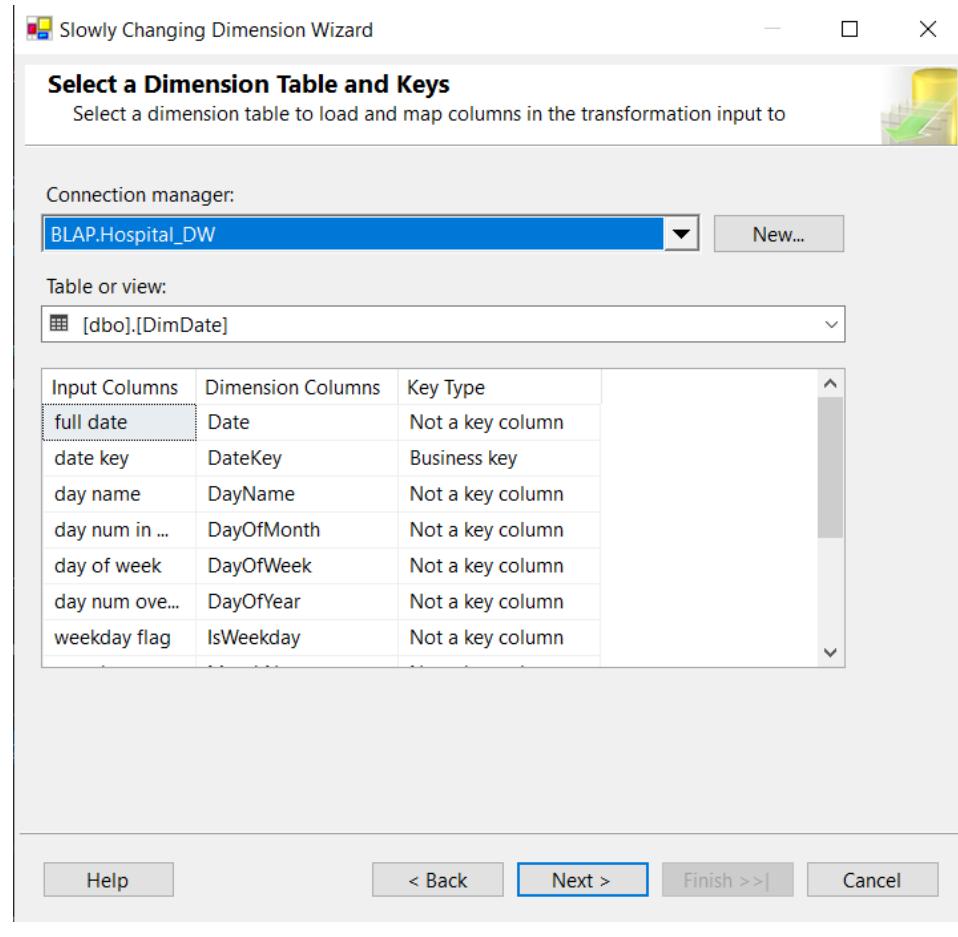


3. Tiến hành cấu hình các task:

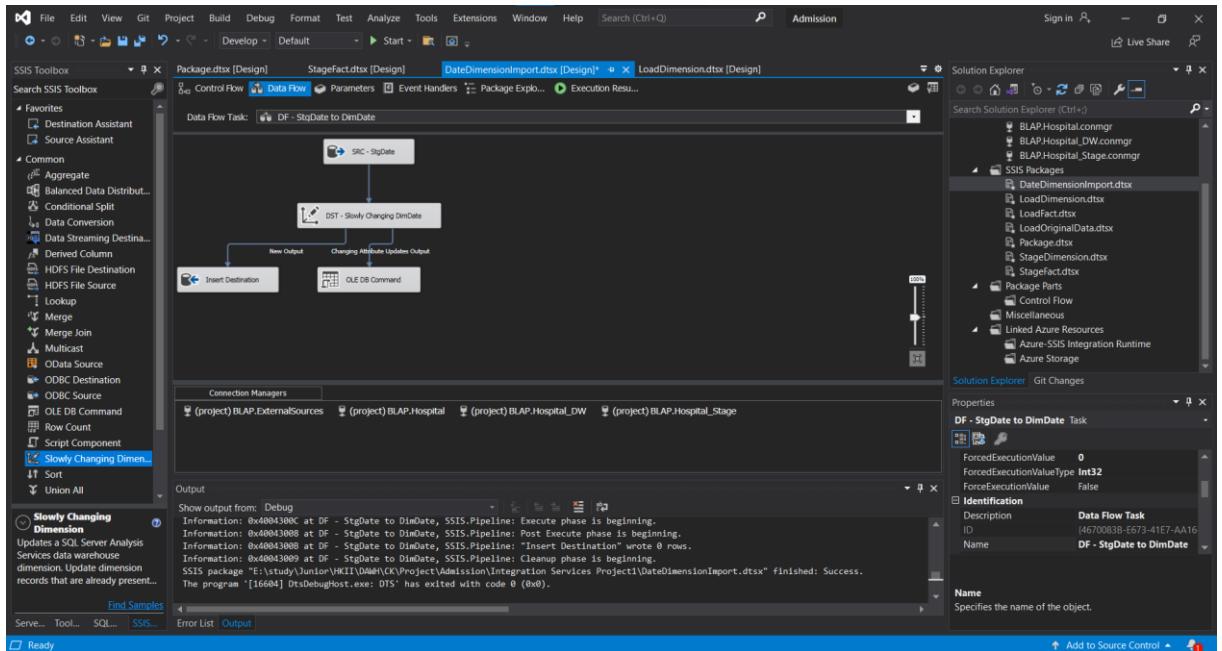
- Cấu hình Source Assistant:



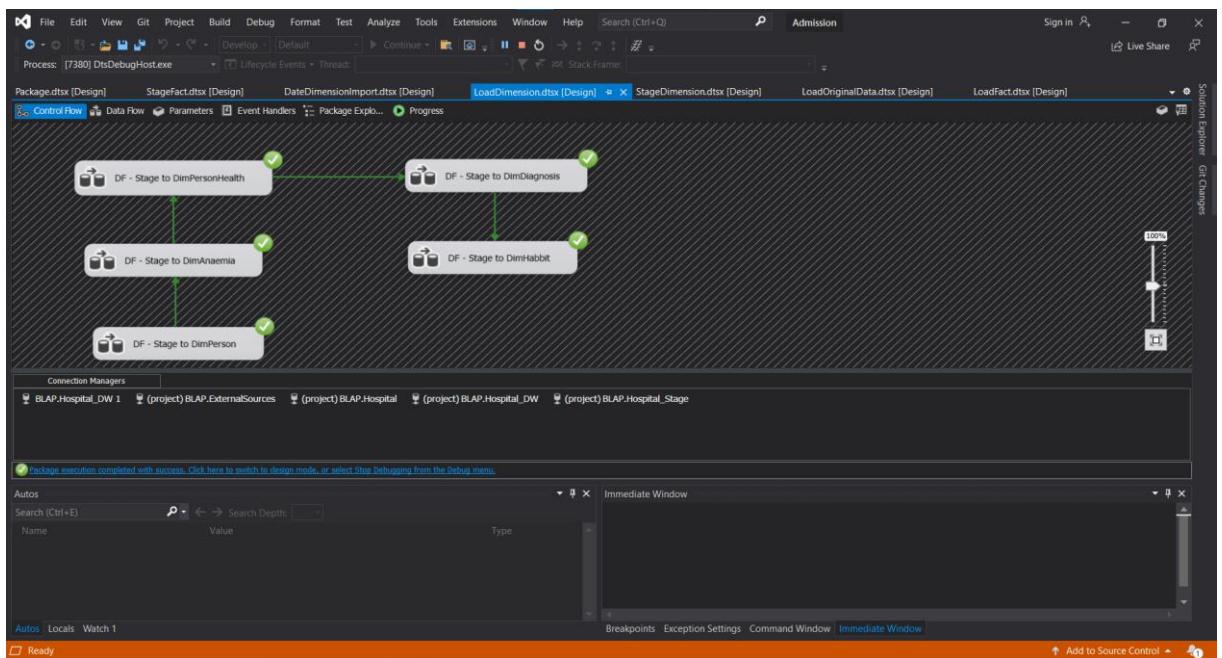
- Cấu hình Slowly Changing Dimension:



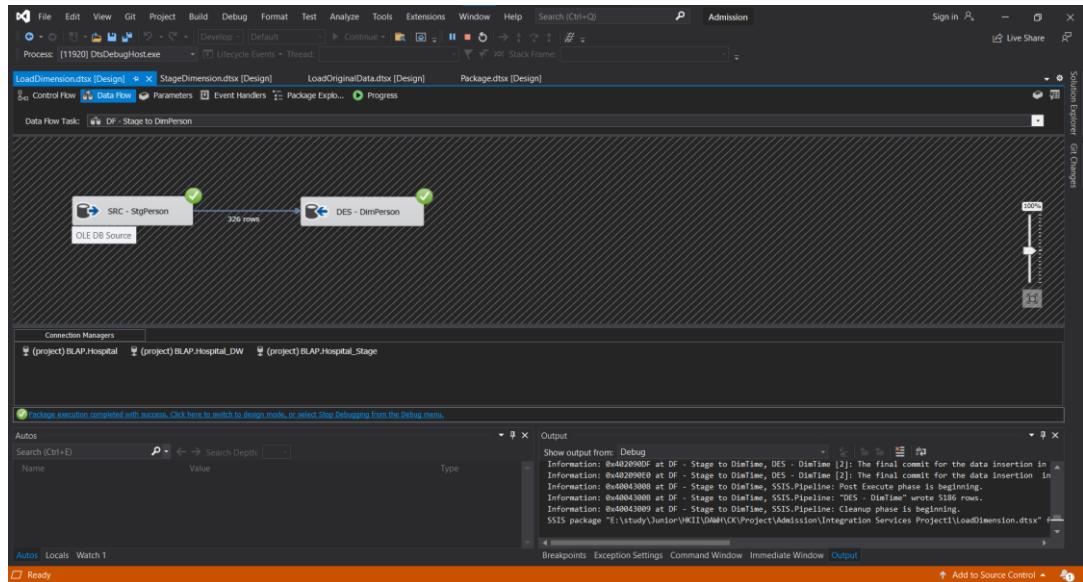
4. Data Flow Task của ta sẽ trông như sau:



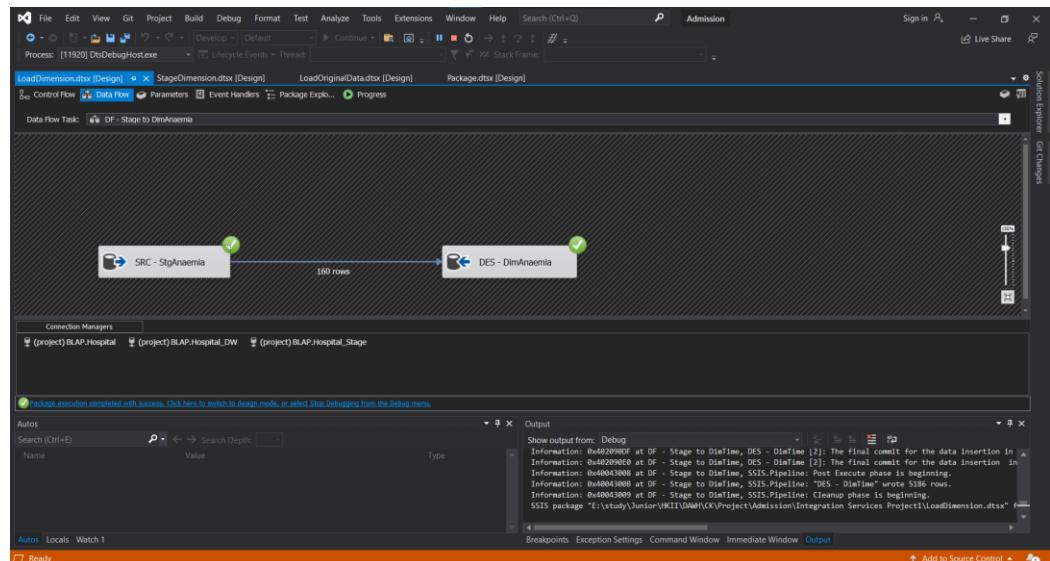
3.3.3.7. Tiến hành Load dữ liệu vào Data Warehouse



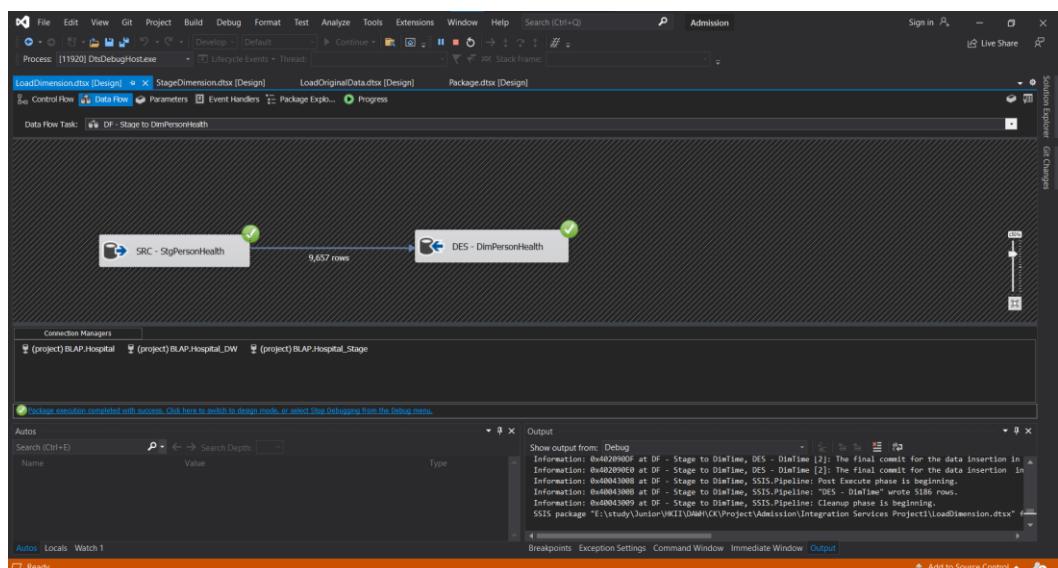
- DimPerson



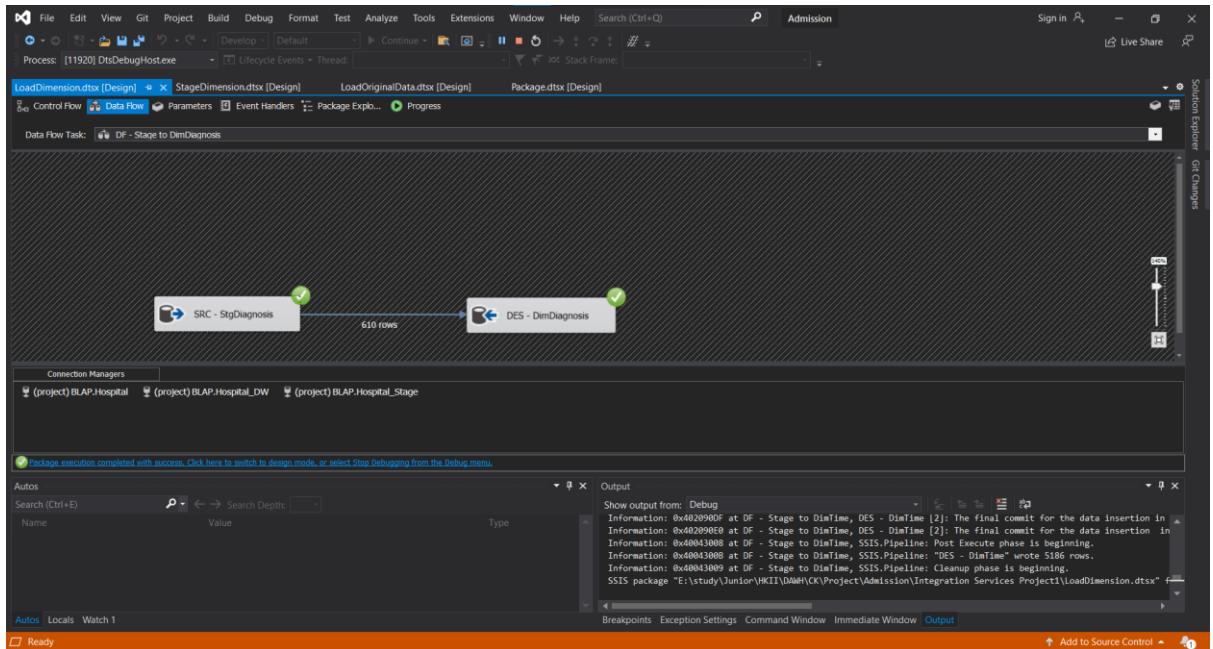
- DimAnaemia



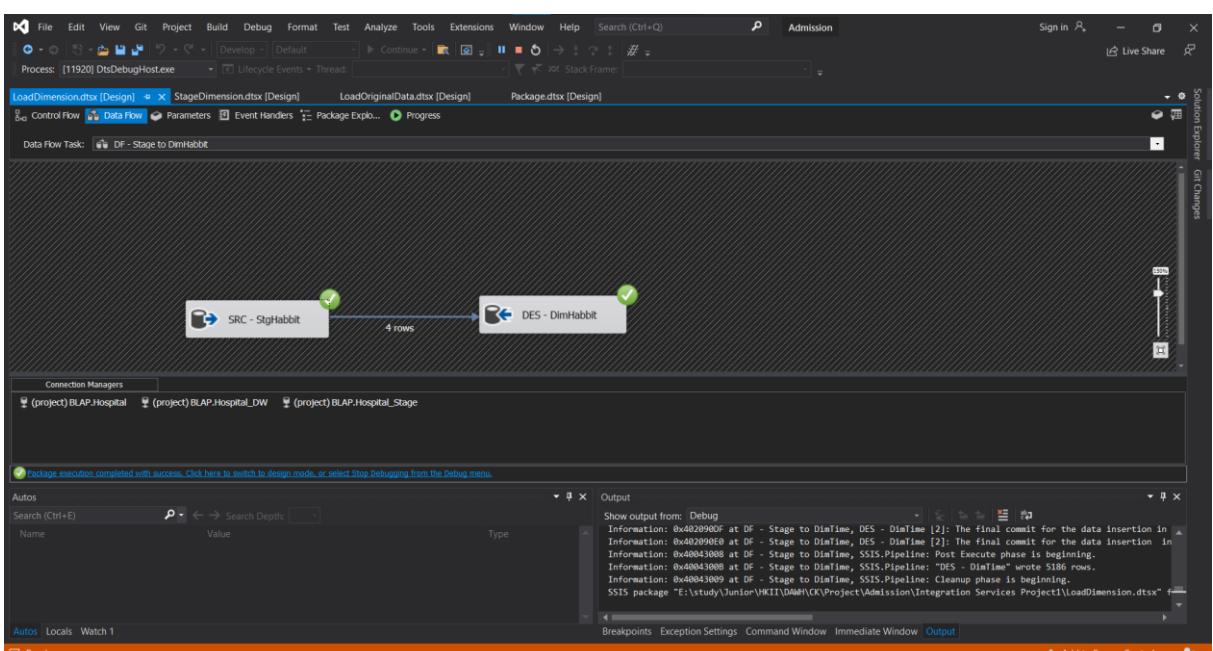
- DimPersonHealth



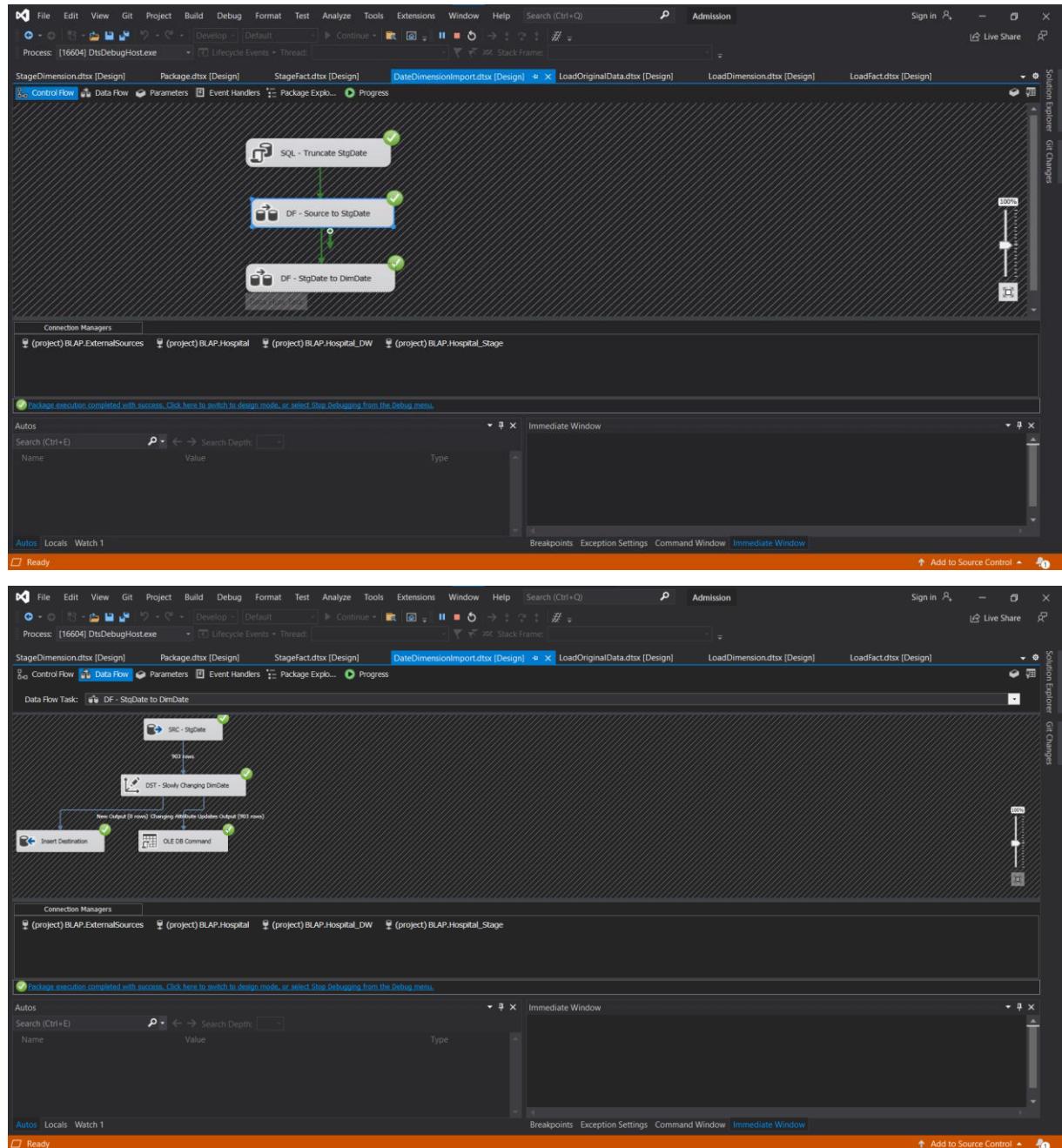
- DimDiagnosis



- DimHabbit



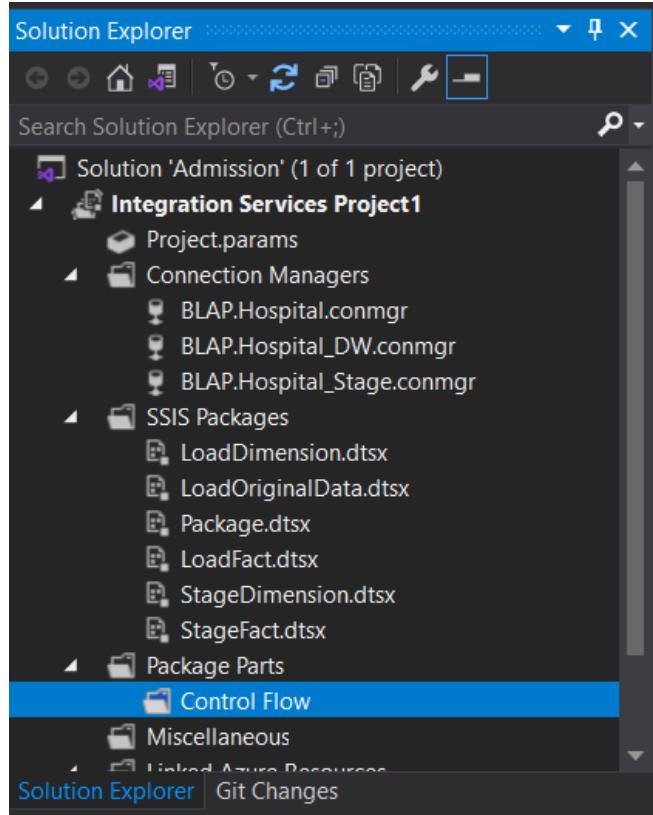
- DimDate



3.4. Import Dữ liệu vào các bảng Fact

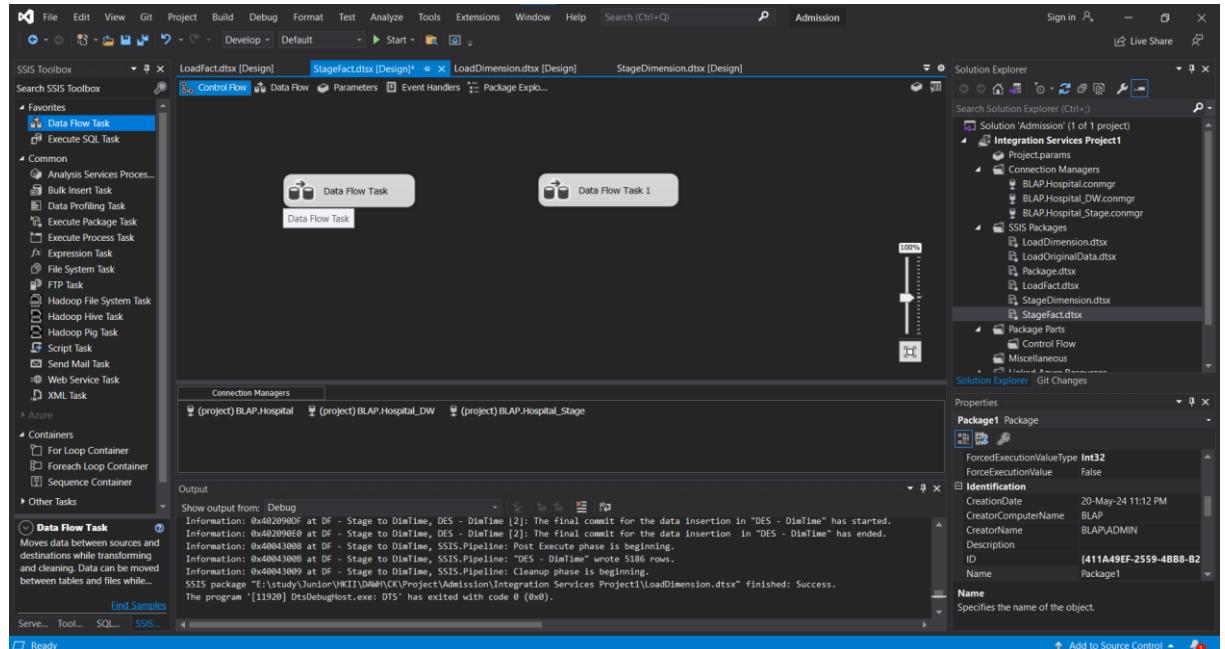
3.4.1. Tạo package

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

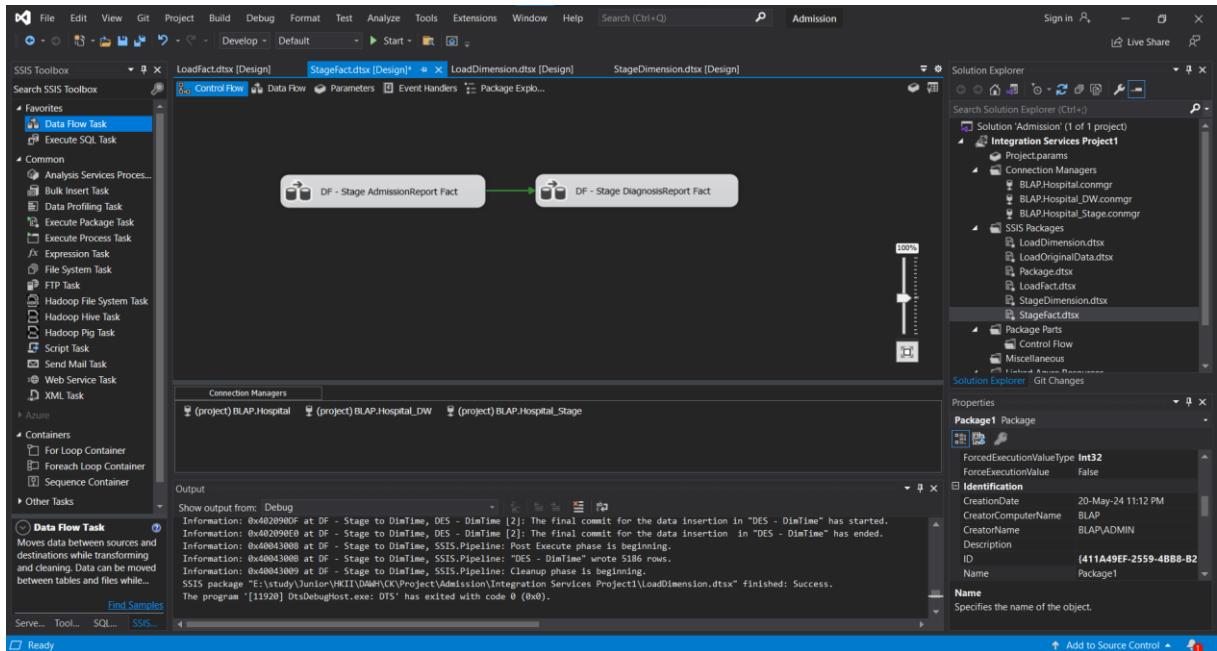


3.3.2. Stage các Fact

1. Click vào **StageFact.dtsx** và tiến hành thêm vào 2 **Data Flow Task** như hình sau:

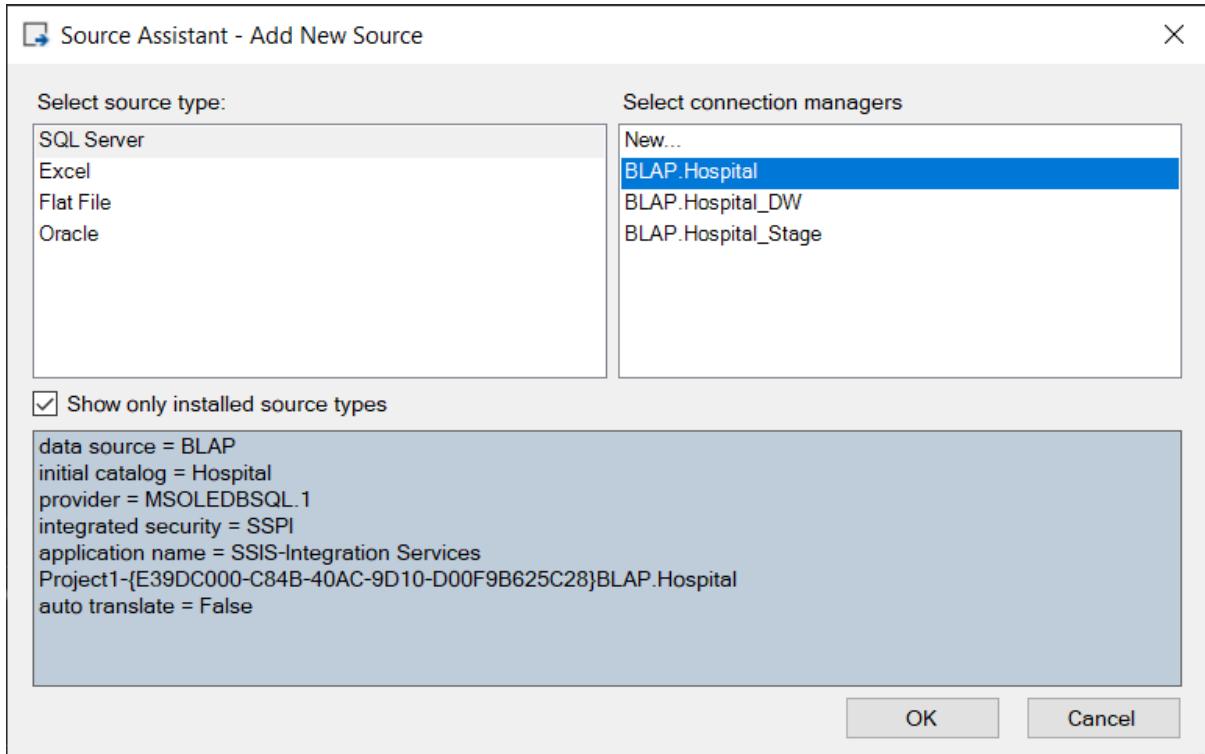


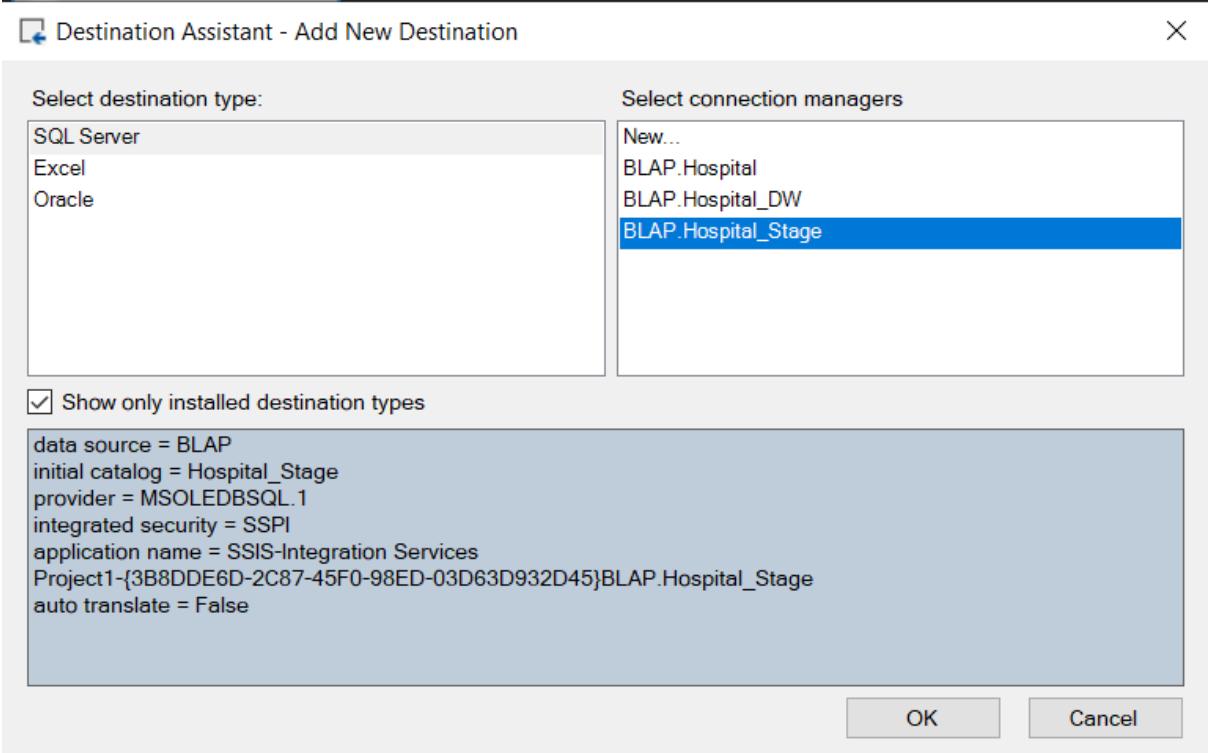
2. Đổi tên các task như sau:



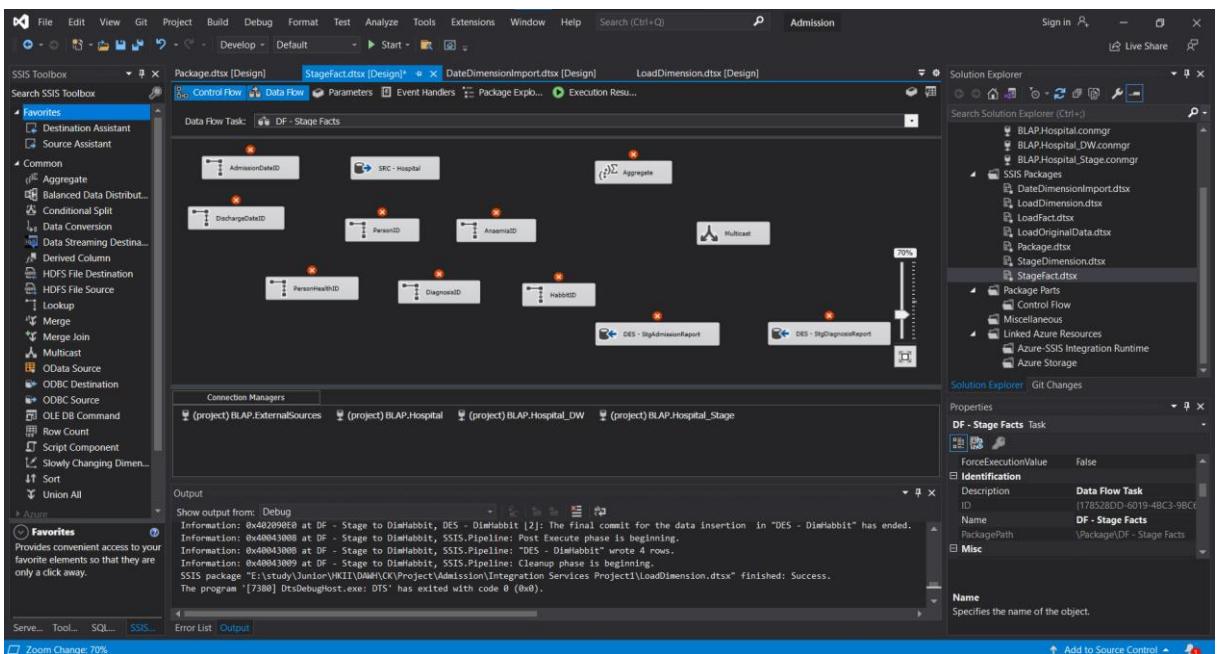
3. Tiến hành stage hai Fact (StgAdmissionReport và StgDiagnosisReport) sau:

- Click vào DF - Stage FactAdmissionReport và thêm Source Assistant, Aggregate, Sort, MultiCast, 6 LookUp và 2 Destination Assistant.



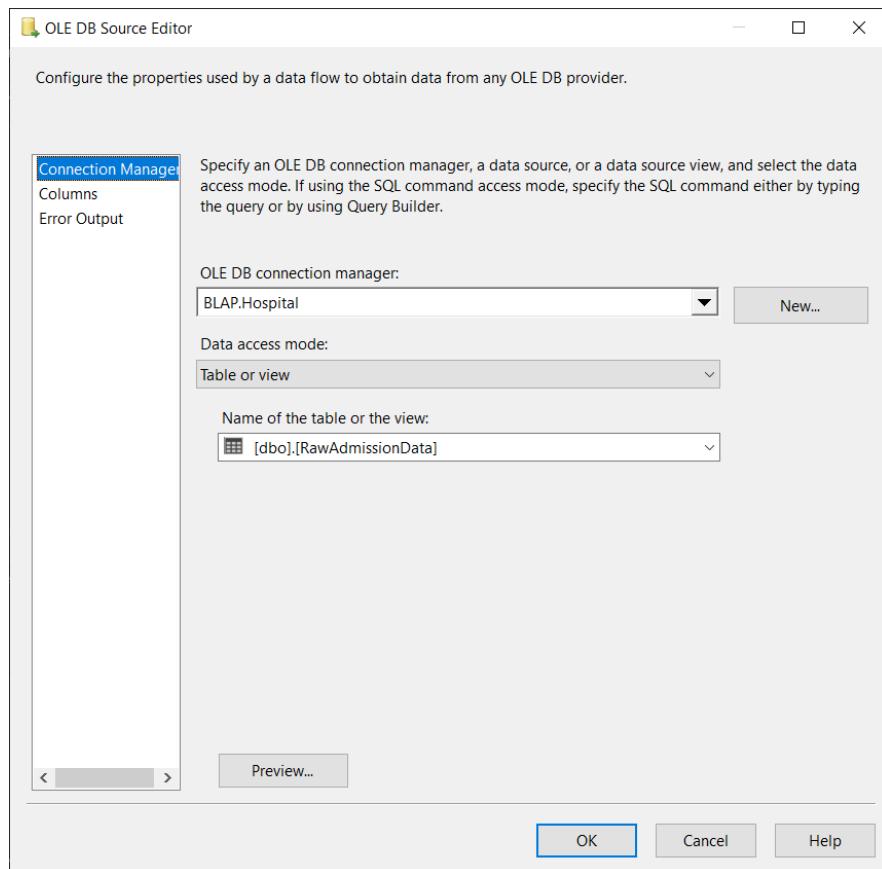


2. Tiến hành đổi tên và ta có được hình như sau:

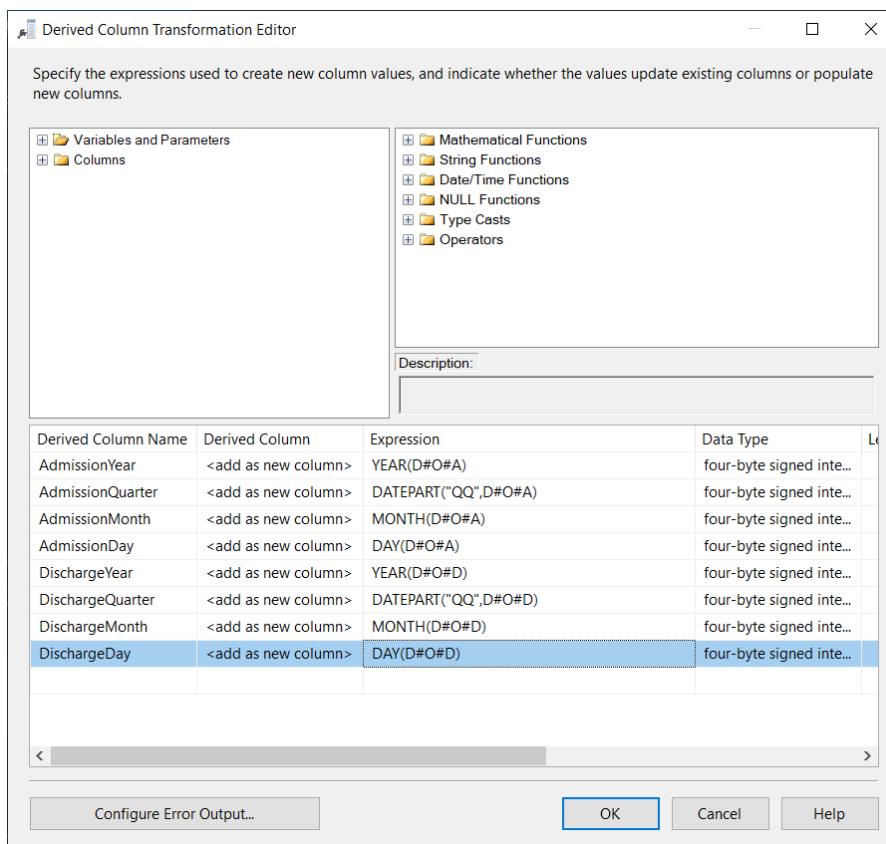


3. Tiến hành cấu hình các task:

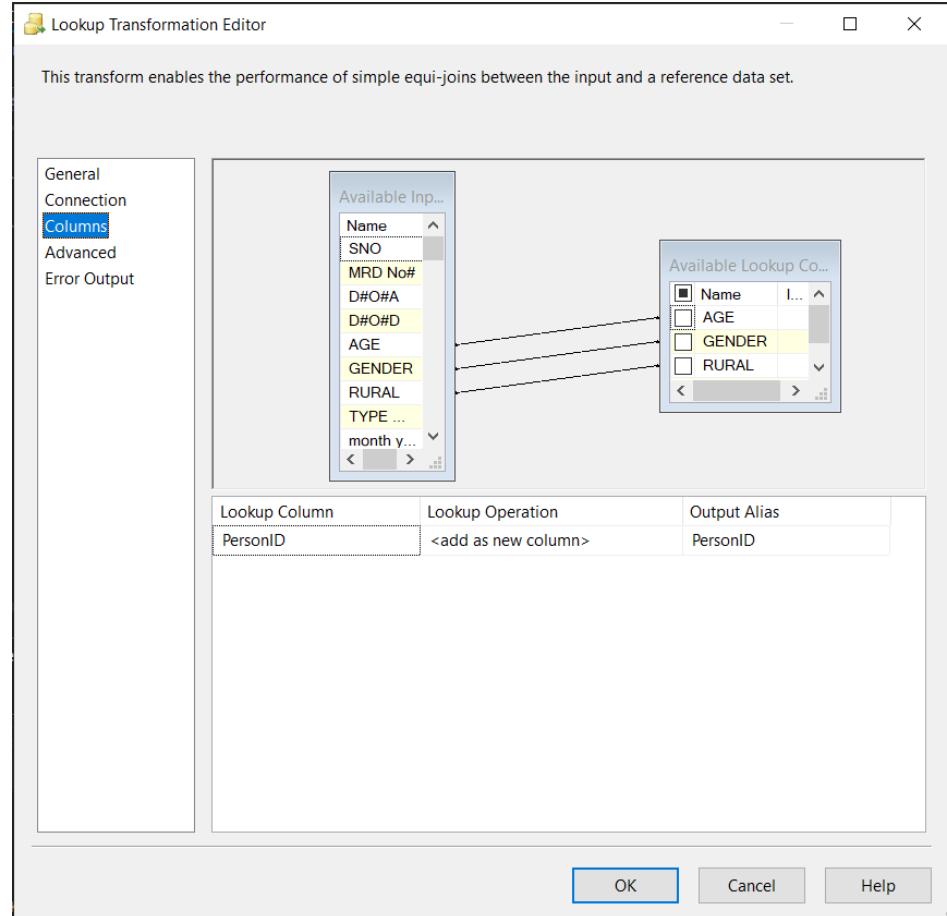
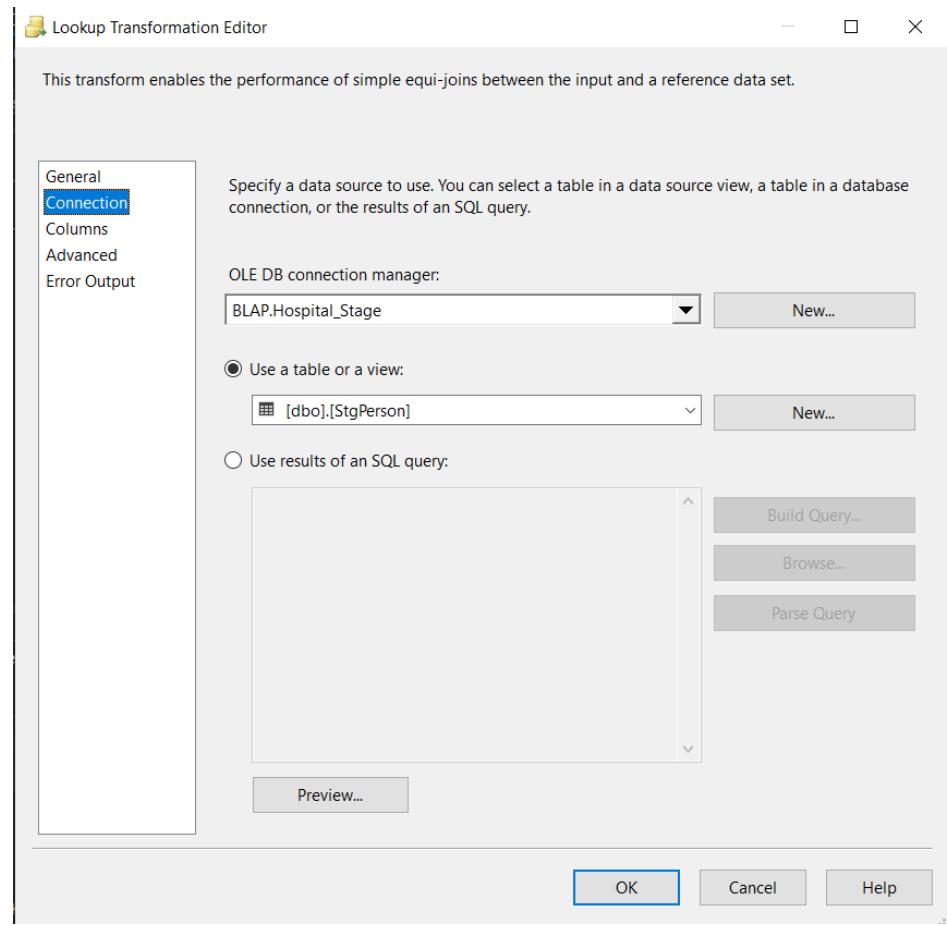
- Cấu hình Source Assistant:



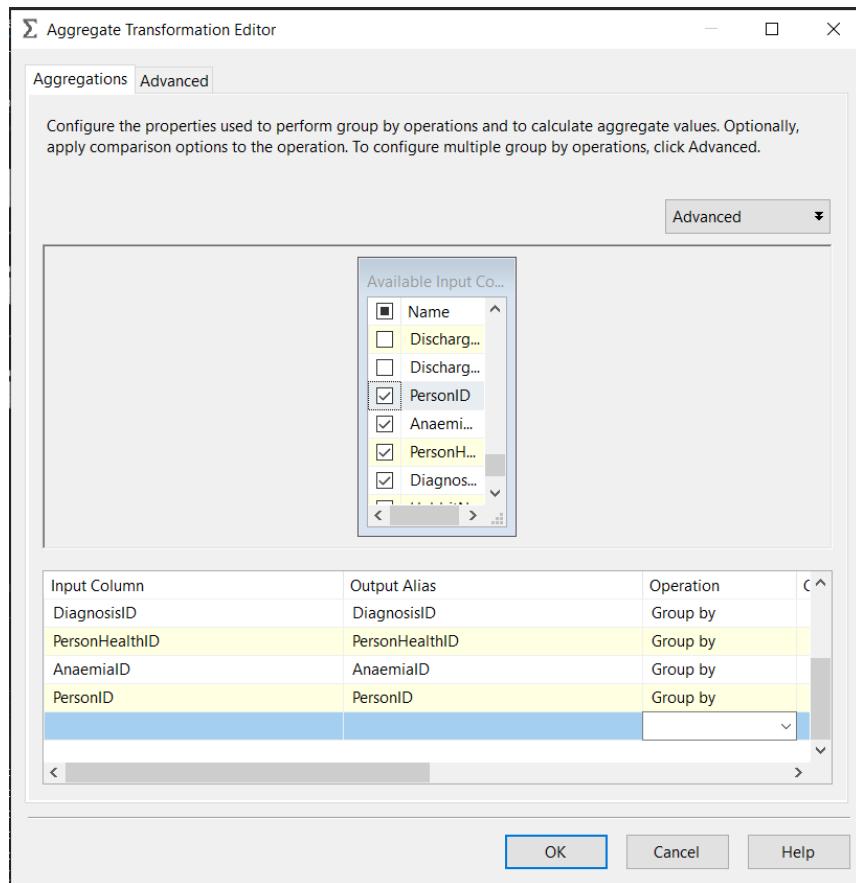
- Cấu hình Derived Column:



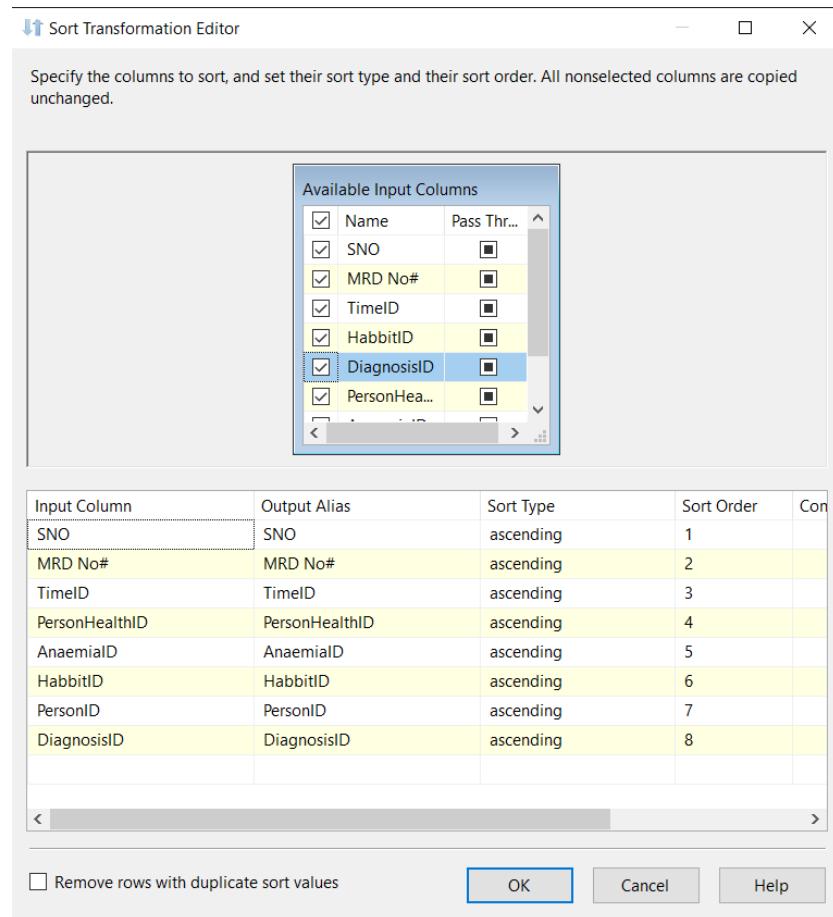
- Cấu hình LookUp Transformation:



- Cấu hình Aggregate:

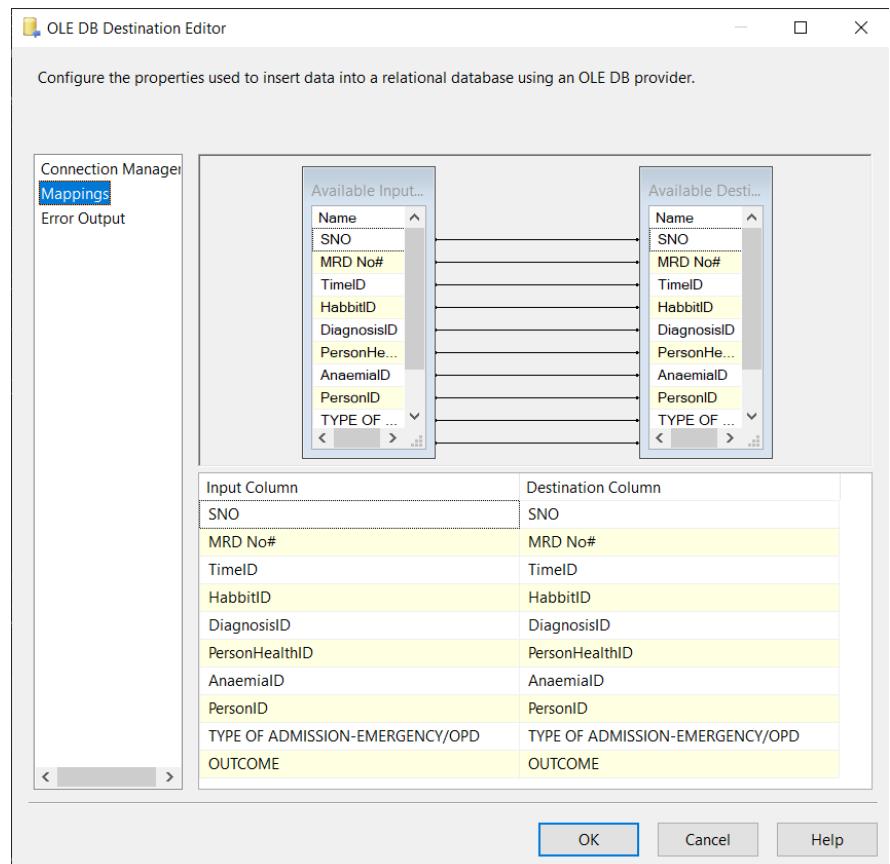
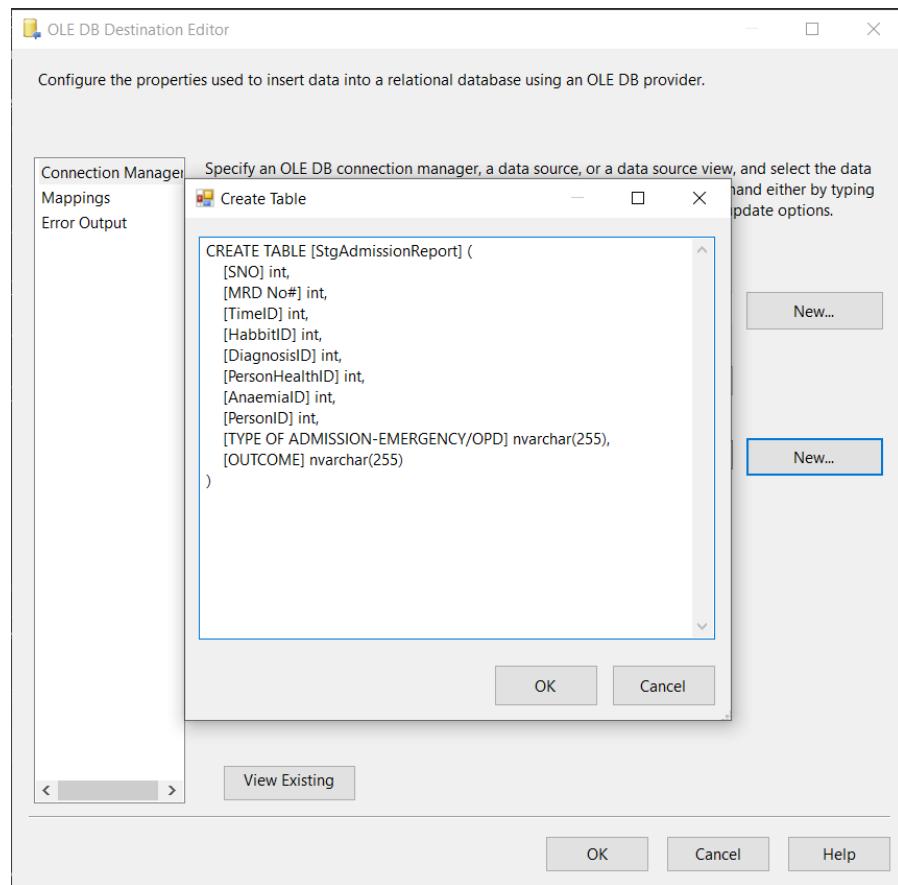


- Cấu hình Sort:

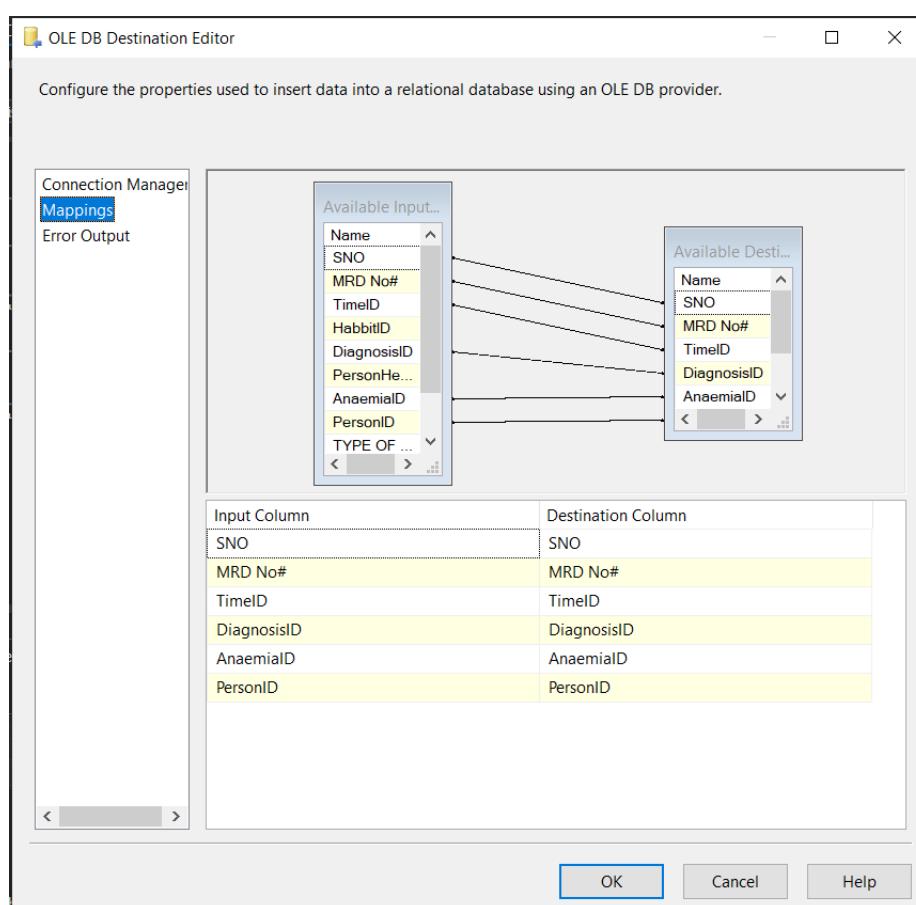
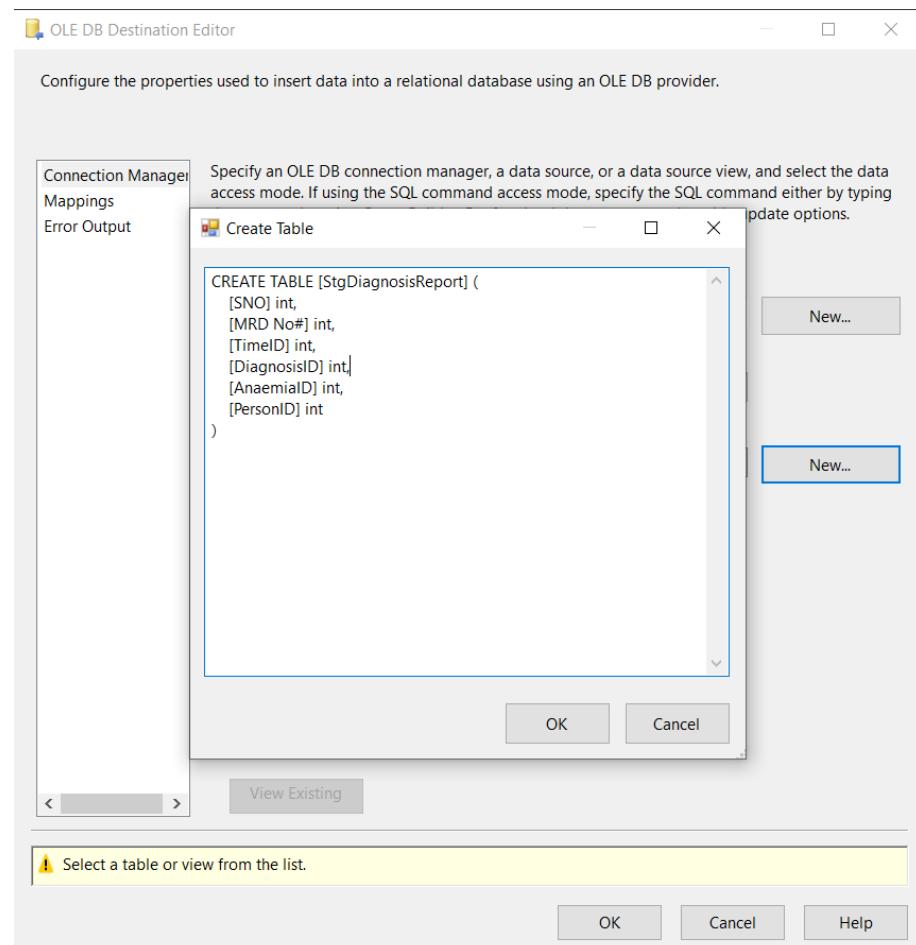


- Cấu hình Destination Assistant:

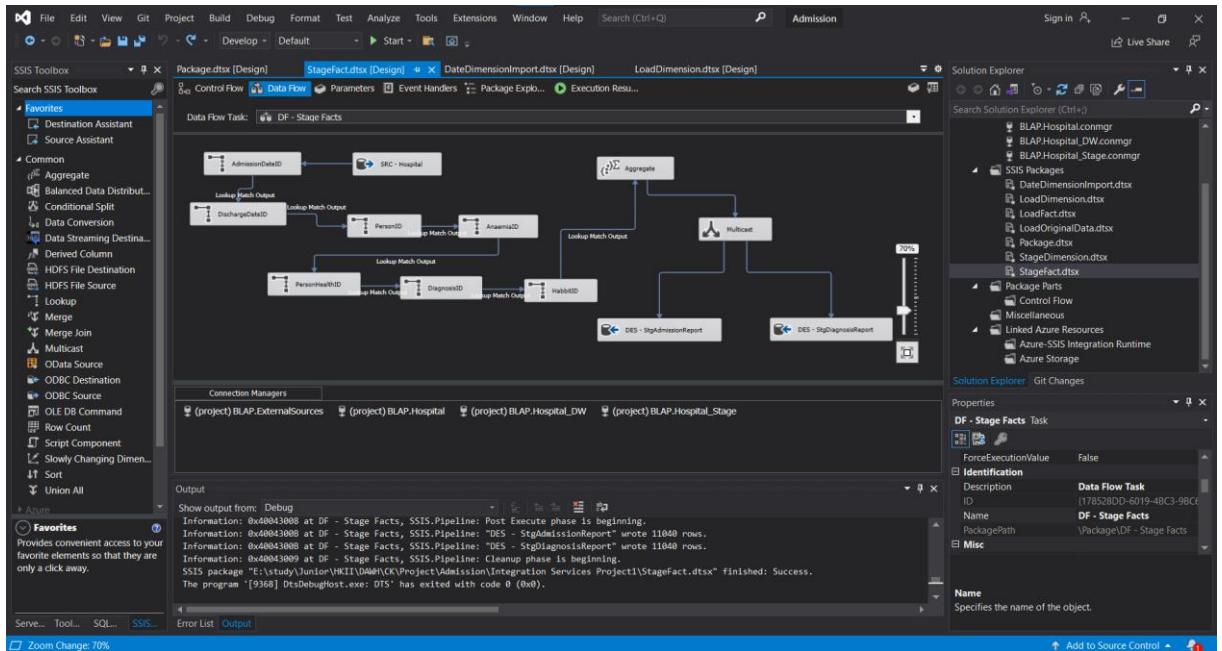
+ AdmissionReport:



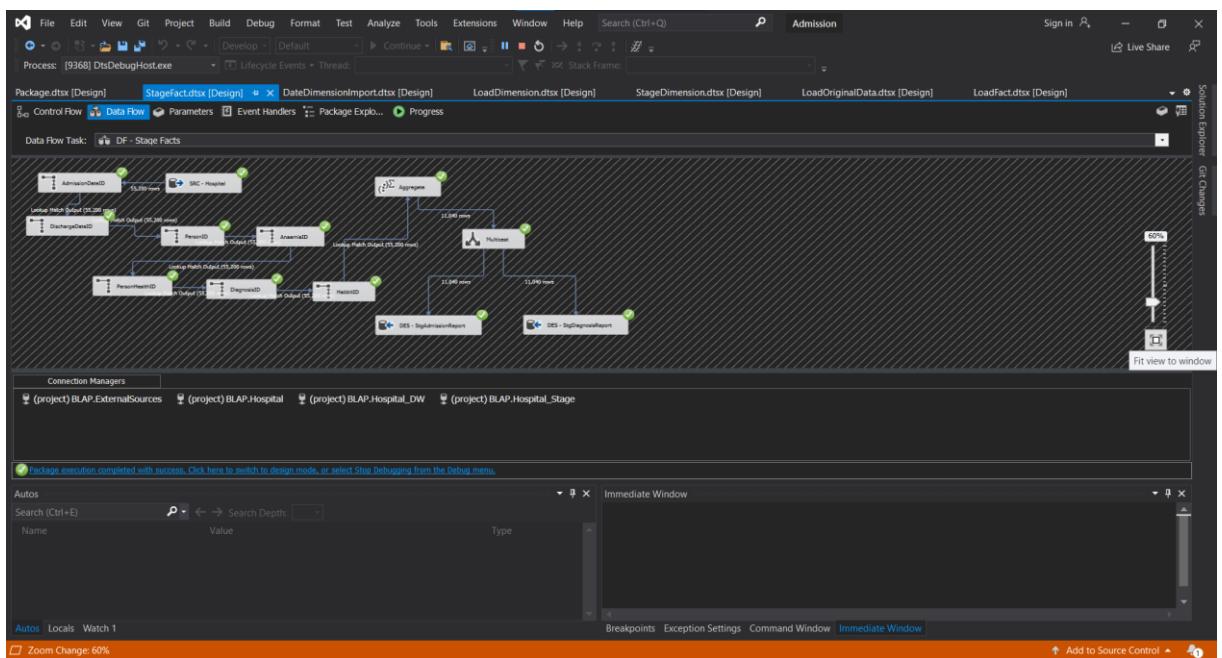
+ DiagnosisReport:



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

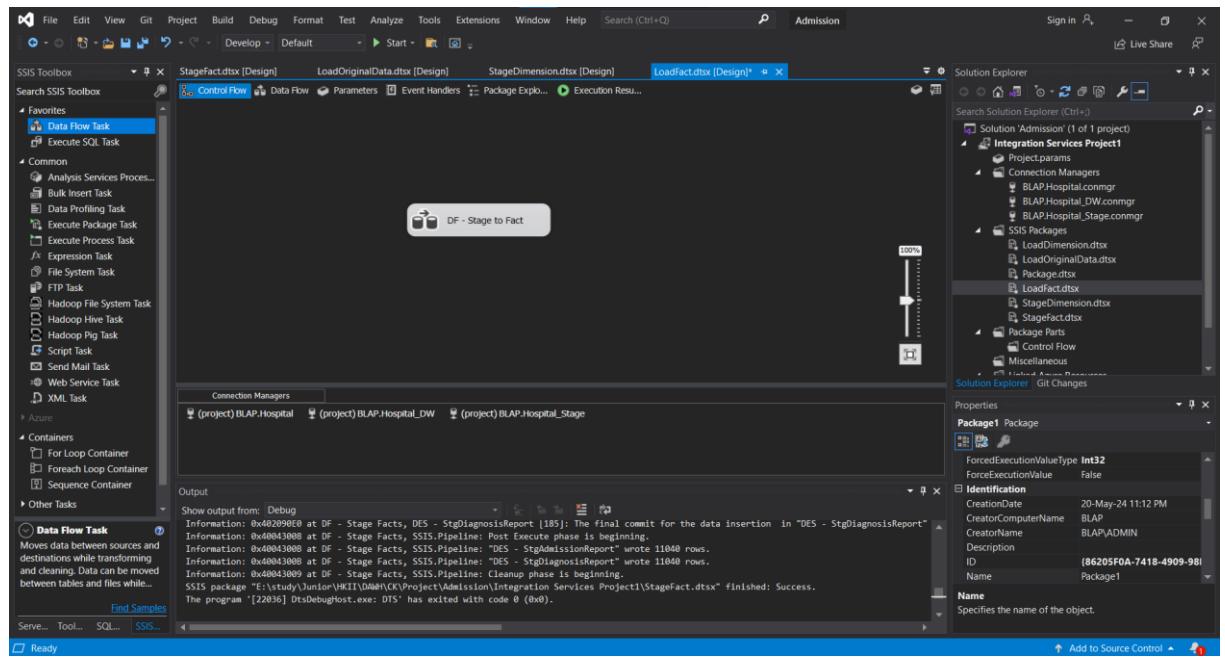


5. Tiến hành chạy Stage các bảng Fact

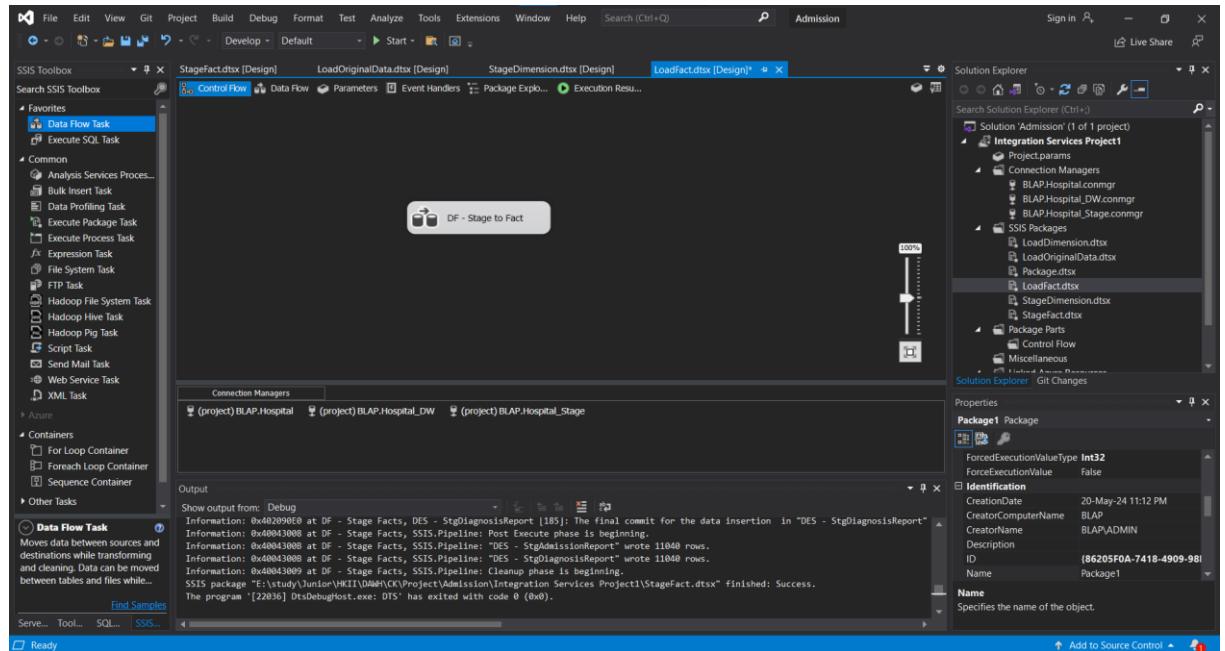


3.3.3. Load các Fact

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

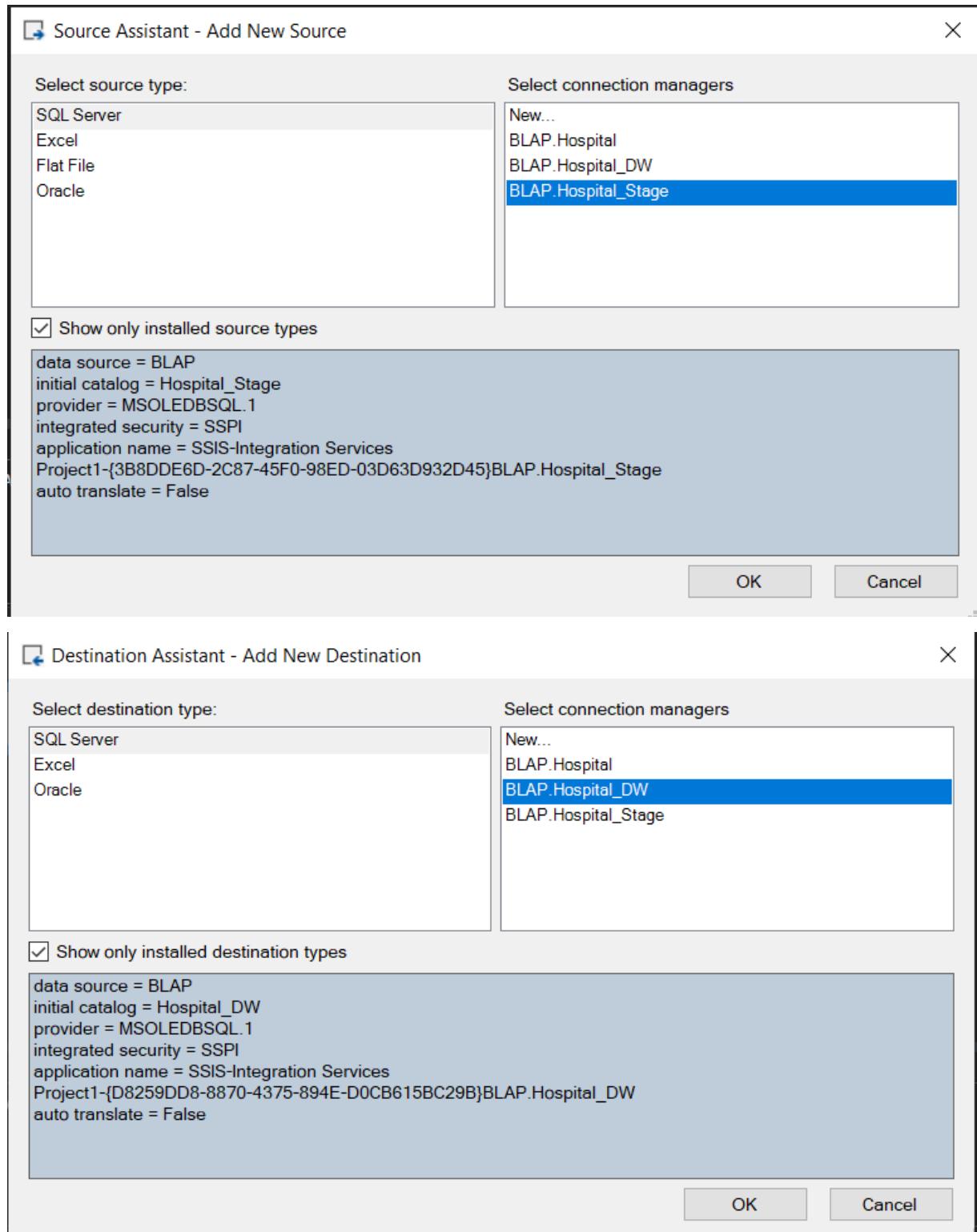


2. Đổi tên các task như sau:

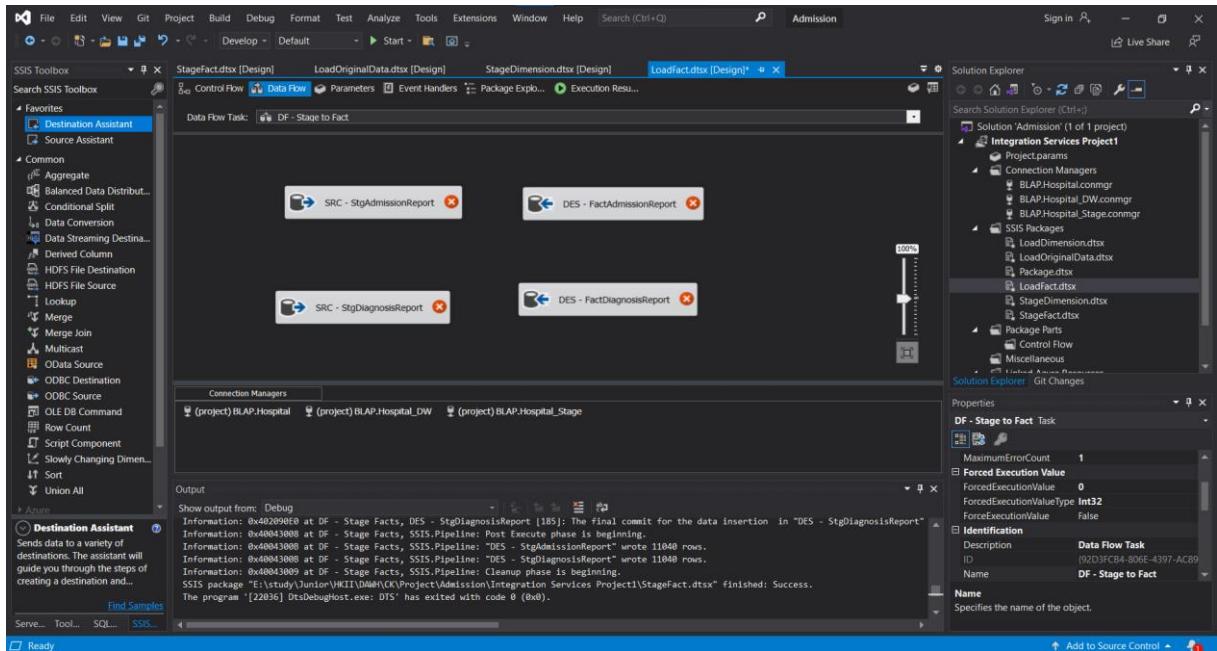


3. Tiến hành load các Fact sau:

1. Click vào **DF - Stage to Fact** và thêm 2 **Source Assistant** và 2 **Destination Assistant**.

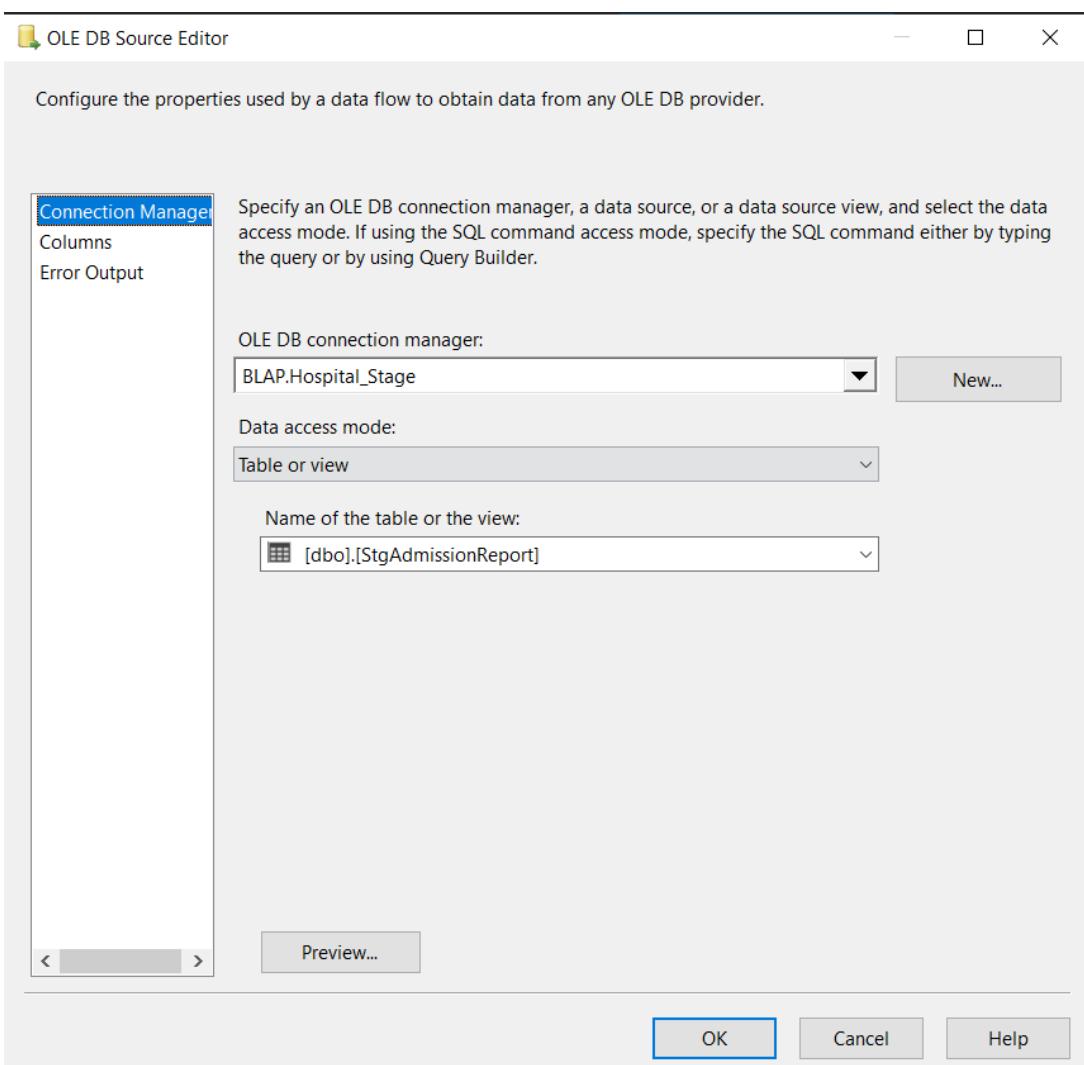


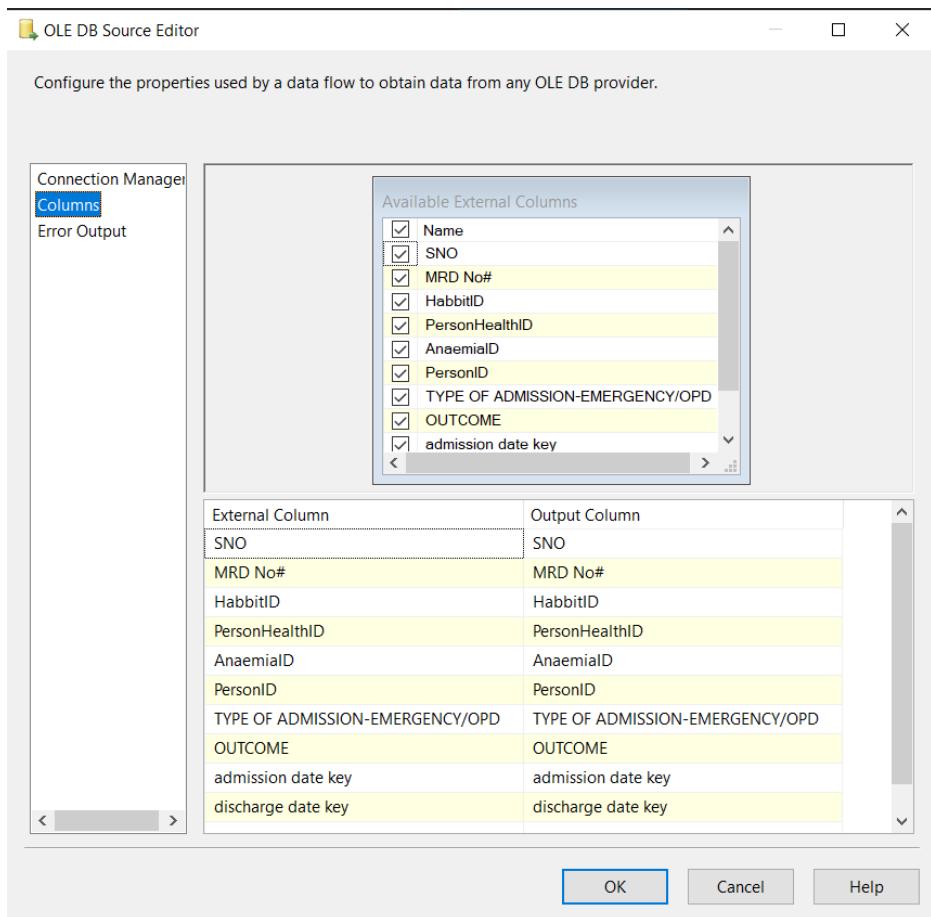
2. Tiến hành đổi tên và ta có được hình như sau:



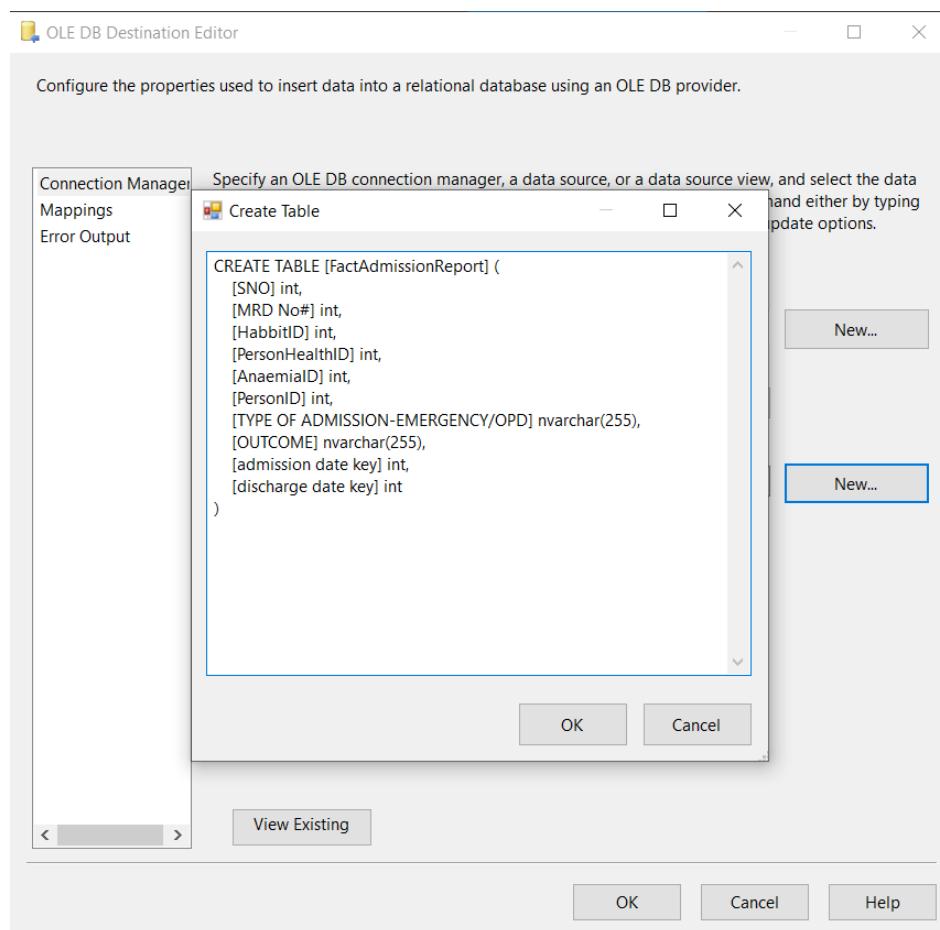
3. Tiến hành cấu hình các task:

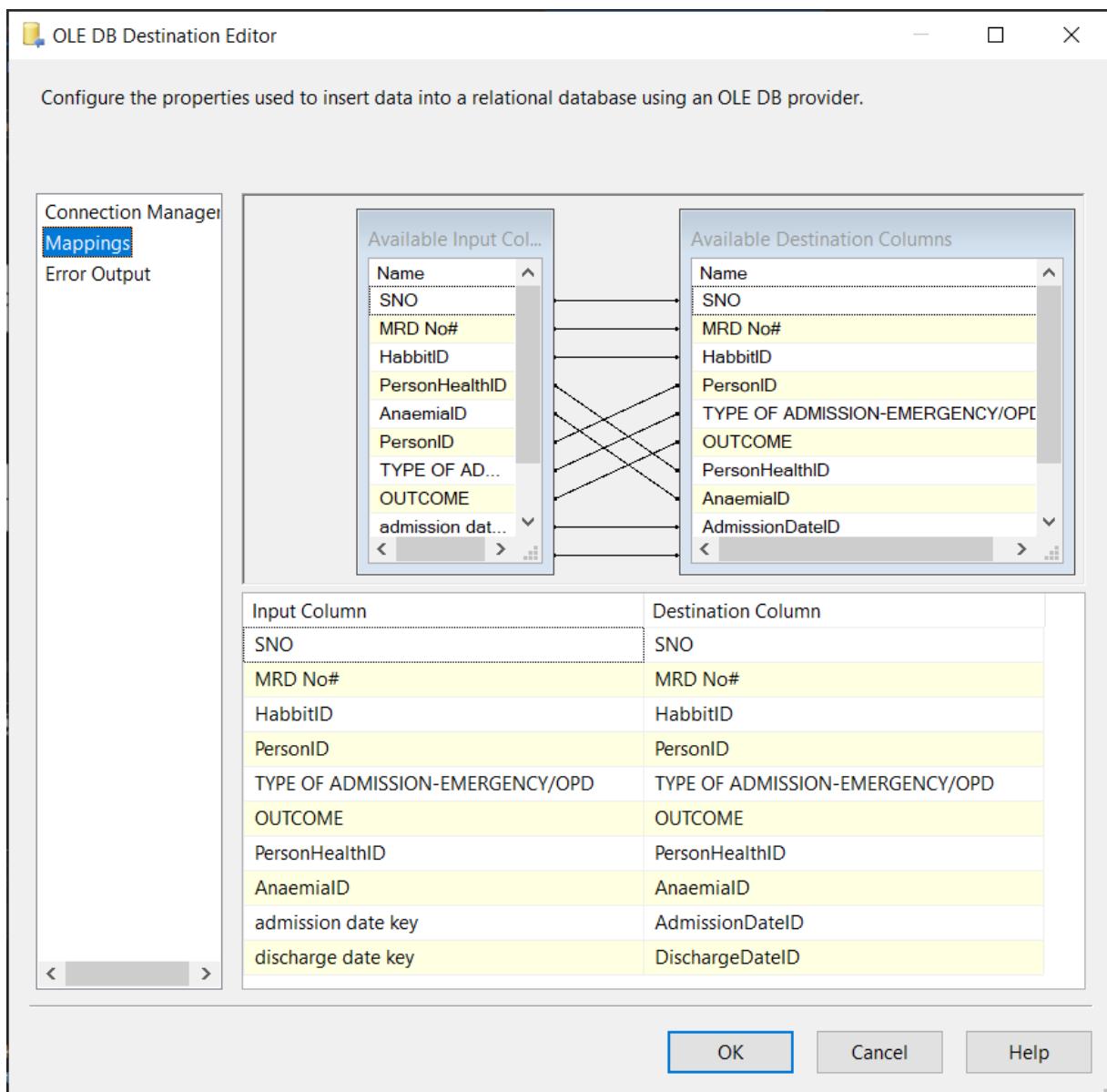
- Cấu hình Source Assistant:



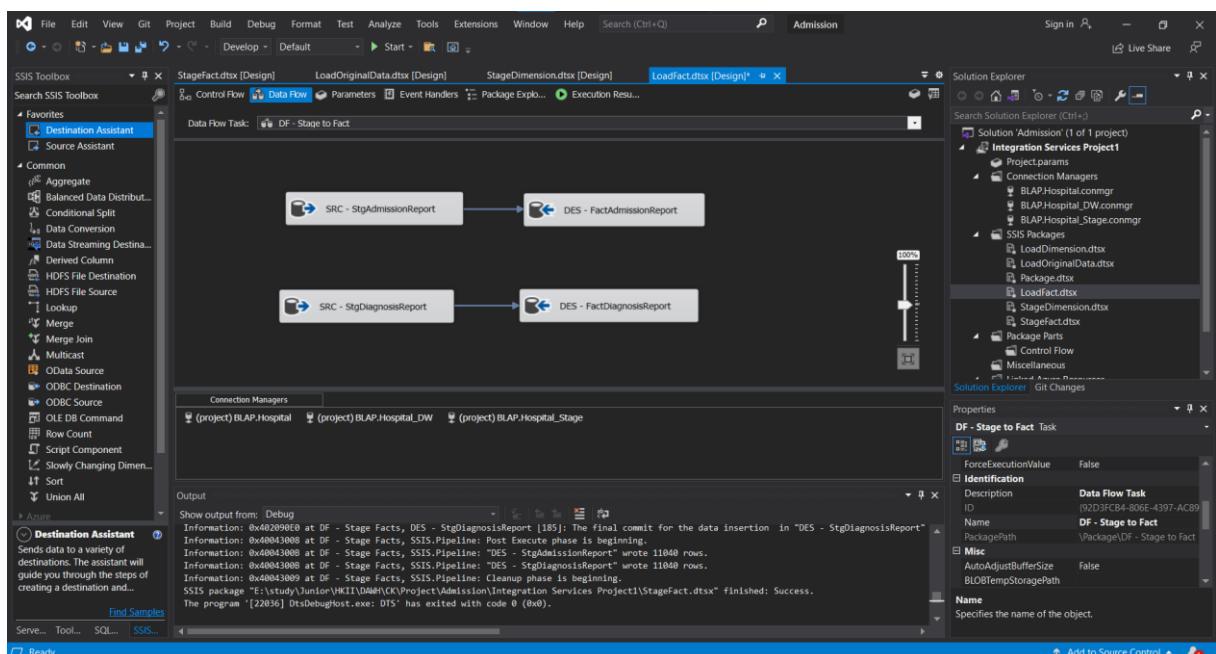


- Cấu hình Destination Assistant:

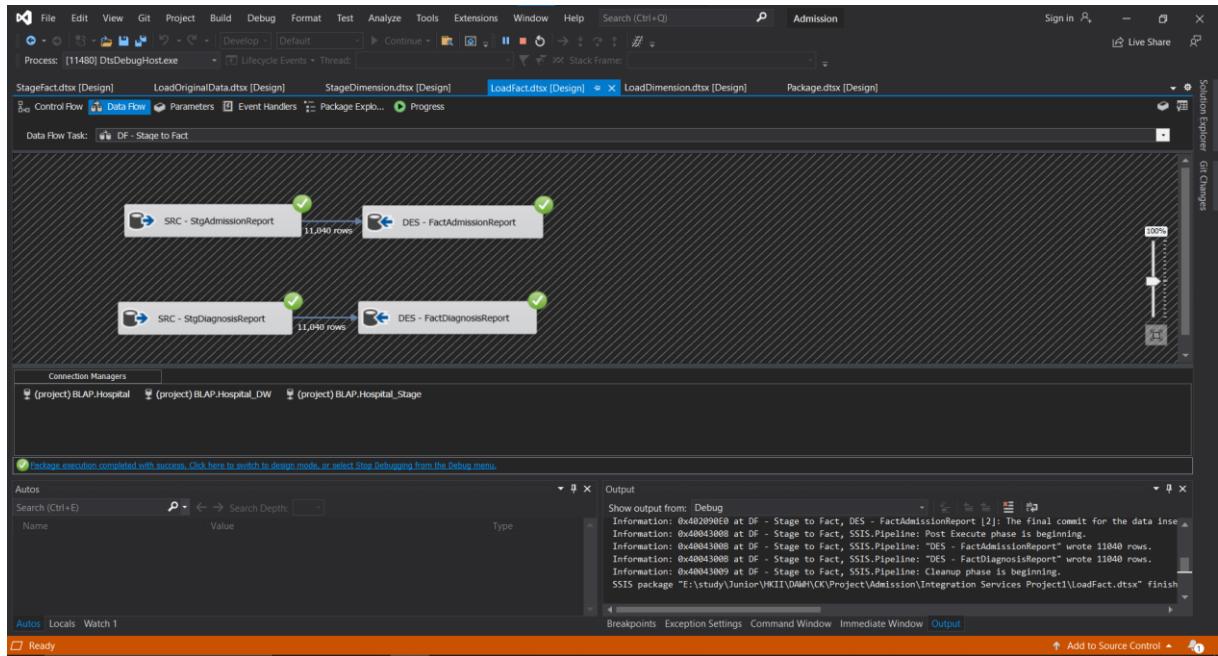




4. Data Flow Task của ta sẽ trông như sau:

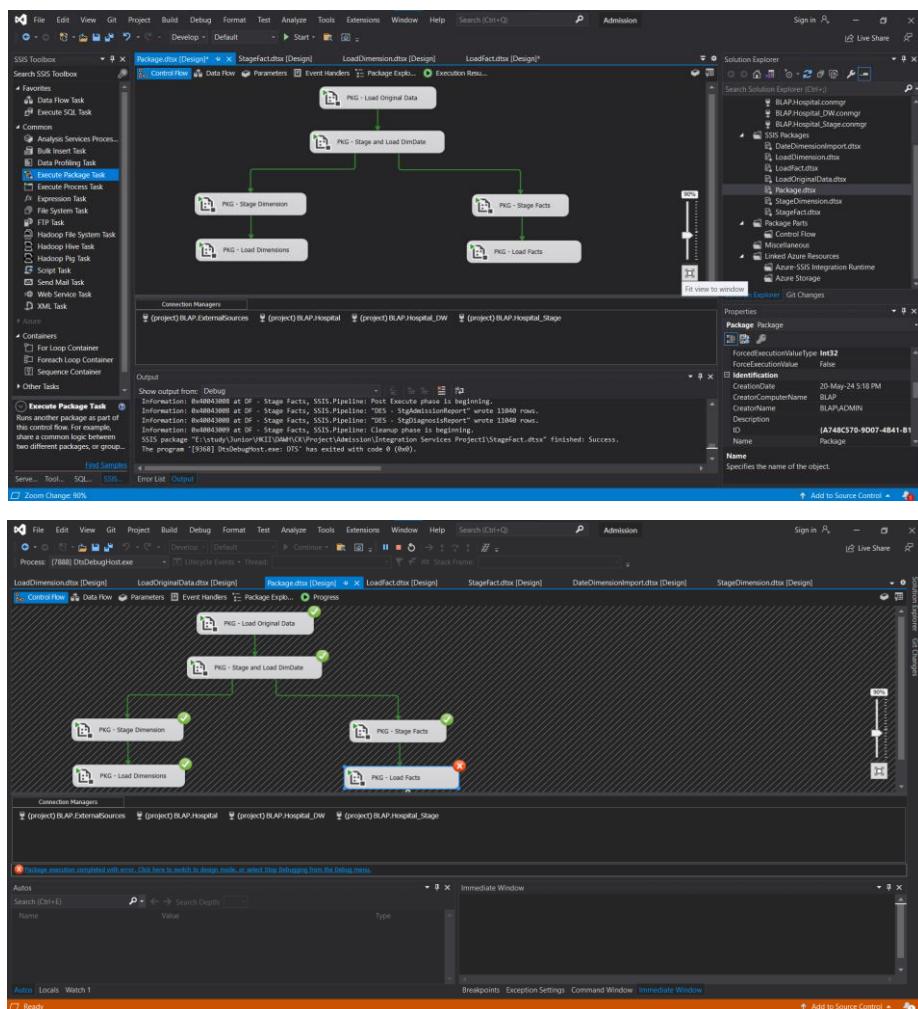


5. Tiến hành Load dữ liệu vào Fact



3.5. Tổng hợp quá trình Tích hợp dữ liệu (SSIS)

Ta tiến hành mở file Package.dtsx và kéo vào 6 Execution Package Task như hình sau:



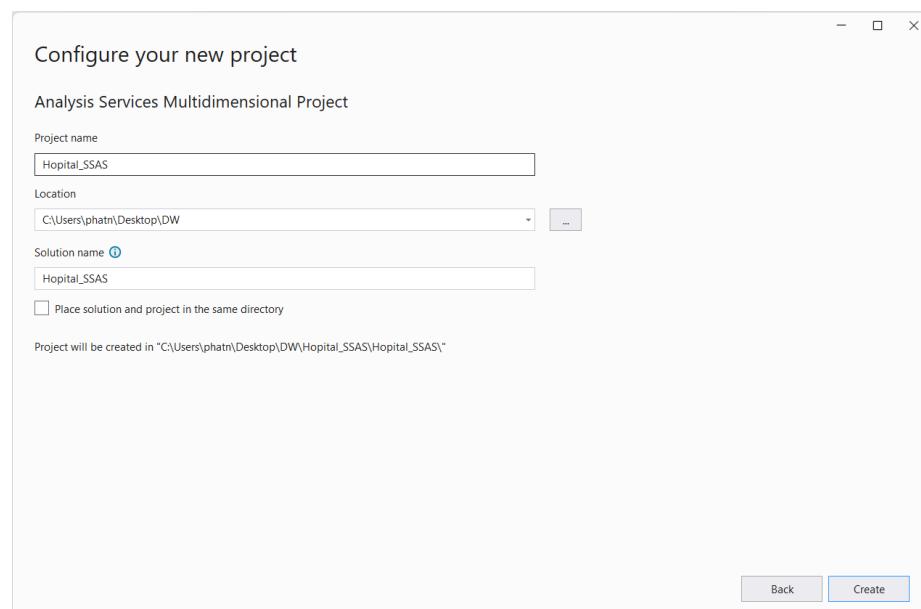
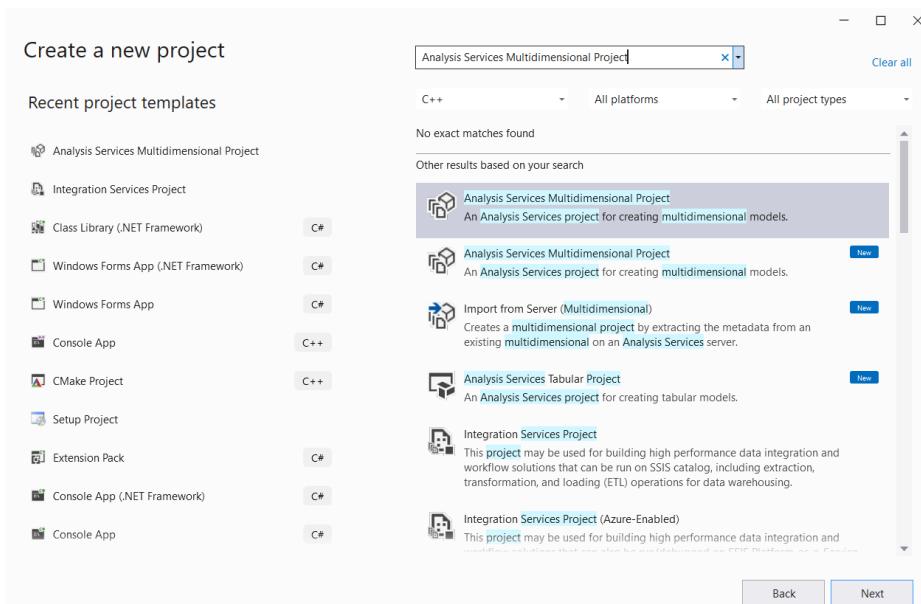
CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS)

Chương này, chúng ta dùng công cụ SSAS xây dựng mô hình và xây dựng khối – cube, xây dựng các API, phân tích dữ liệu trên Cube. Sử dụng chức năng Pivot table trong Excel thực hiện tương tự như quá trình SSAS và tiến hành đổi chiều kết quả với phân tích bằng tay. Sử dụng ngôn ngữ truy vấn MDX thực hiện truy vấn 4 loại: roll up, drill down, slice and dice, pivot table, các hàm hỗ trợ và đổi chiều kết quả với thực hiện bằng tay và Excel.

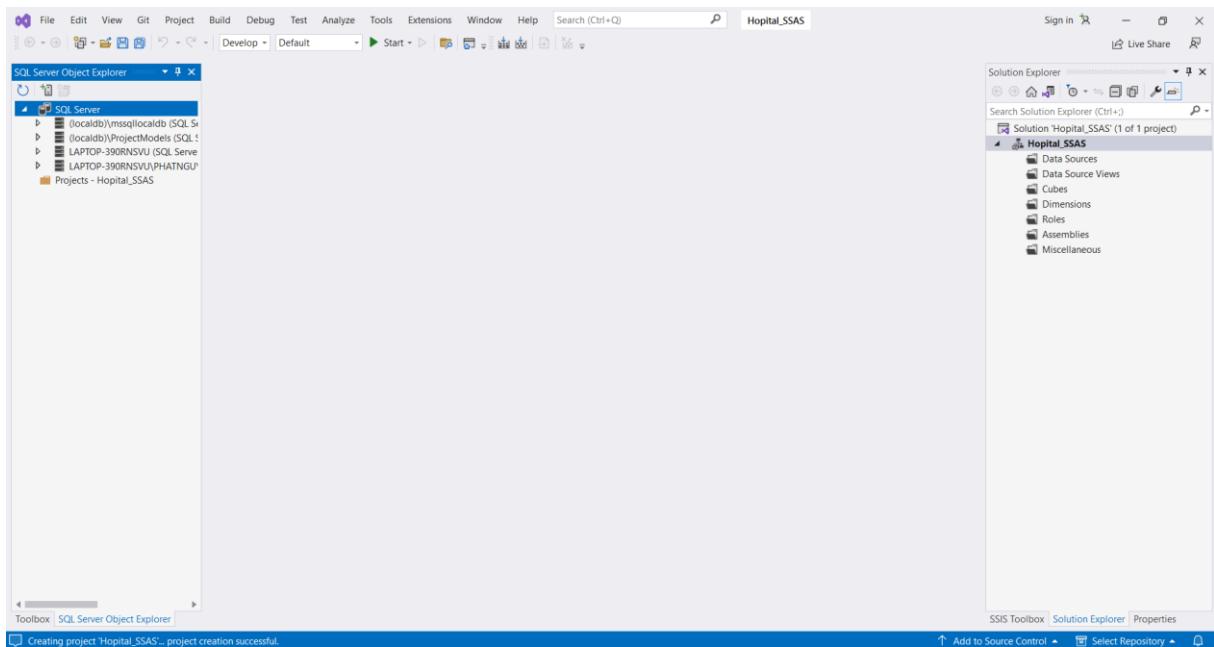
4.1. Quá trình xây dựng mô hình

4.1.1. Tạo Project

Mở Visual Studio → File → New → Project → Analysis Services Multidimensional Project, đặt tên cho project và click OK.

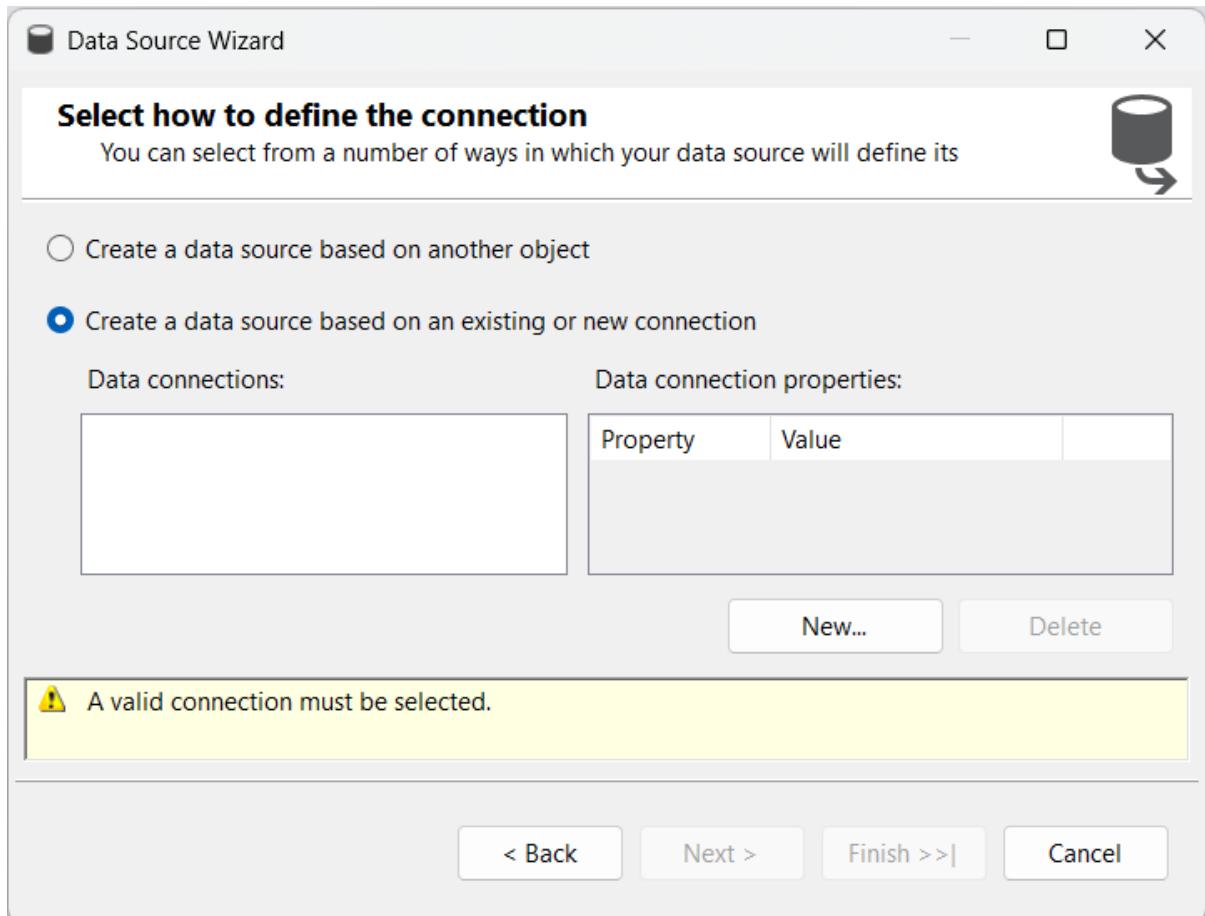


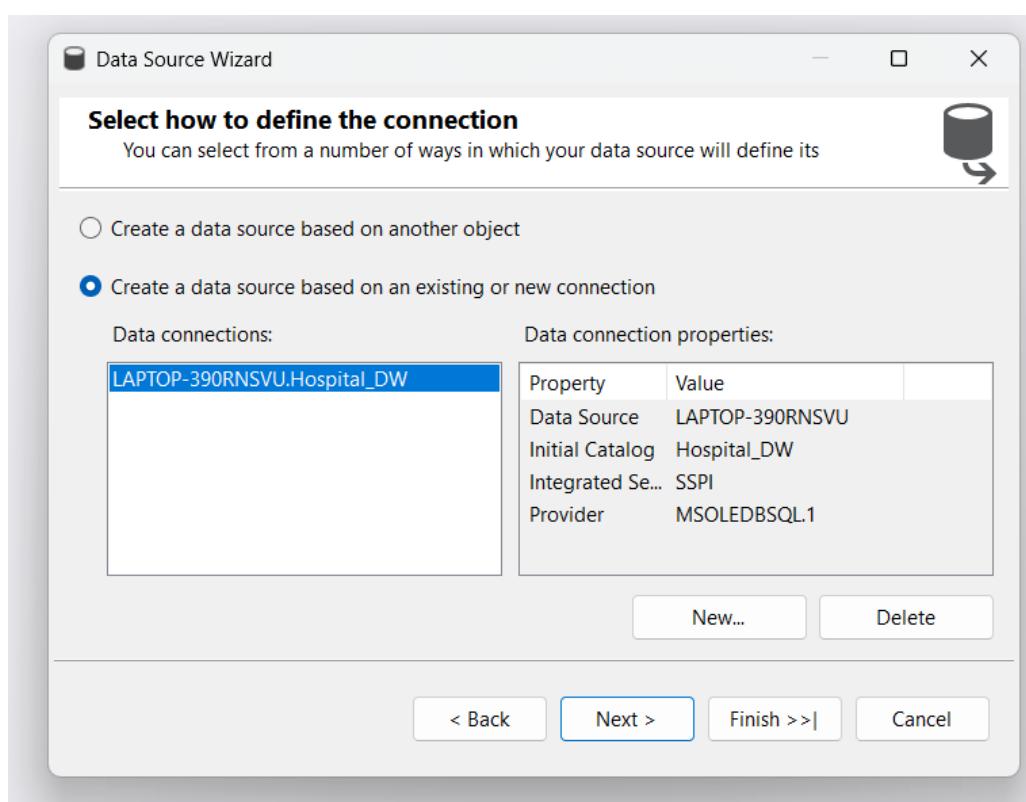
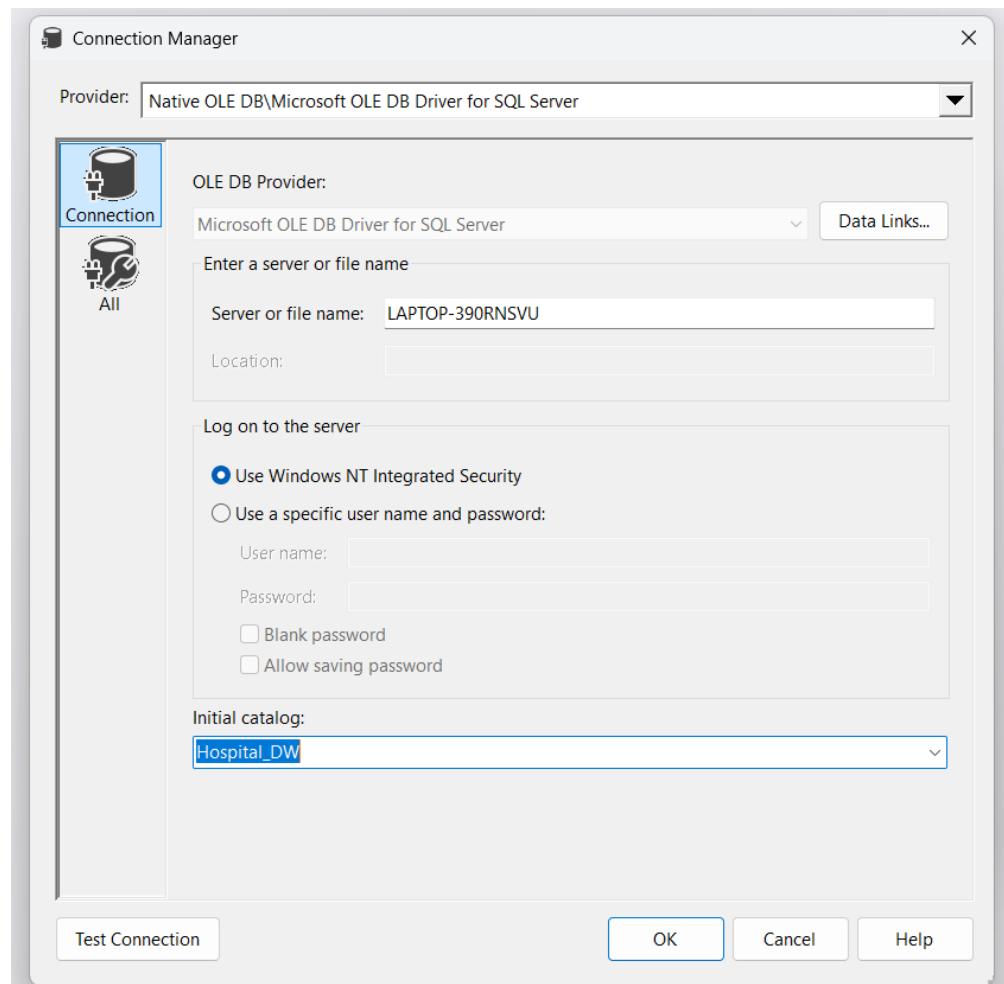
Giao diện của SSAS sau khi tạo một project mới như hình bên dưới:

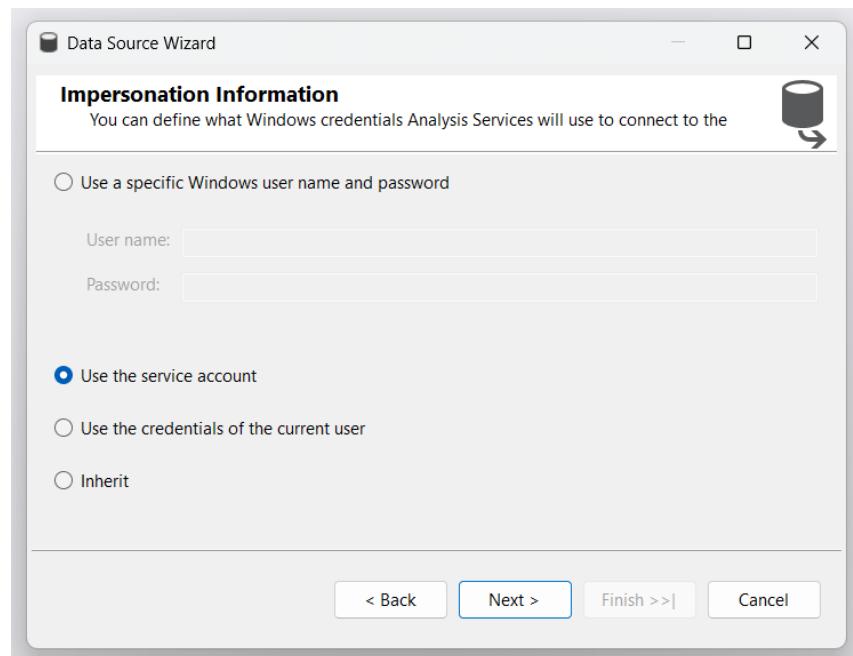


4.1.2. Tạo Data Source

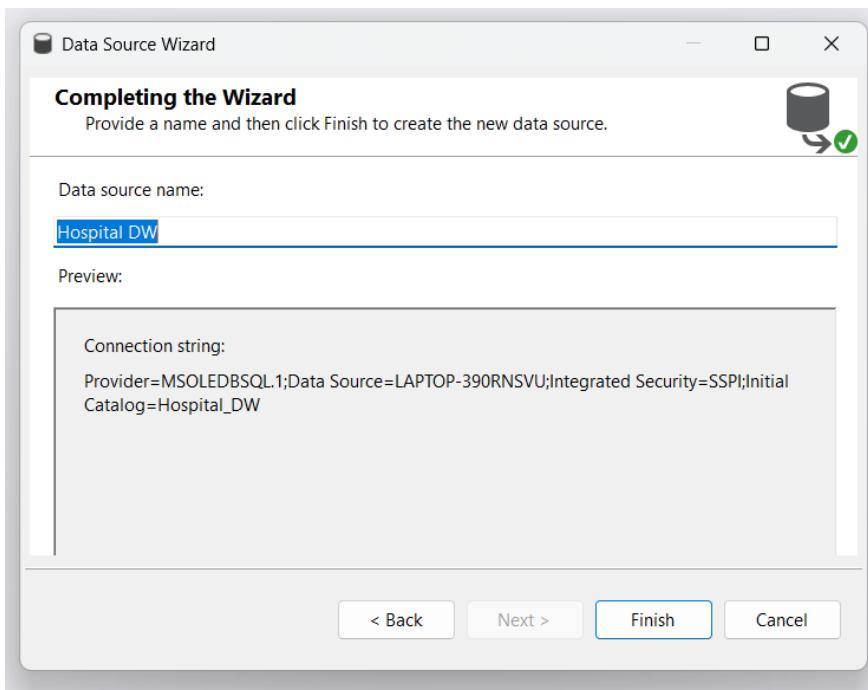
Right-click vào **Data Sources** ở khung **Solution Explorer**. Chon **New Data Source** để tạo Data Source mới. Kết nối đến Database **Hospital_DW**.



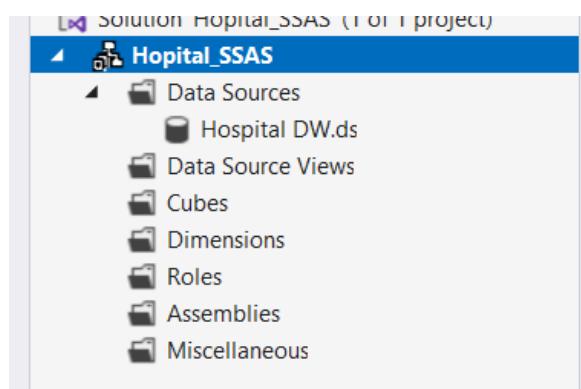




Chọn “Use the service account”, Nhấn next:



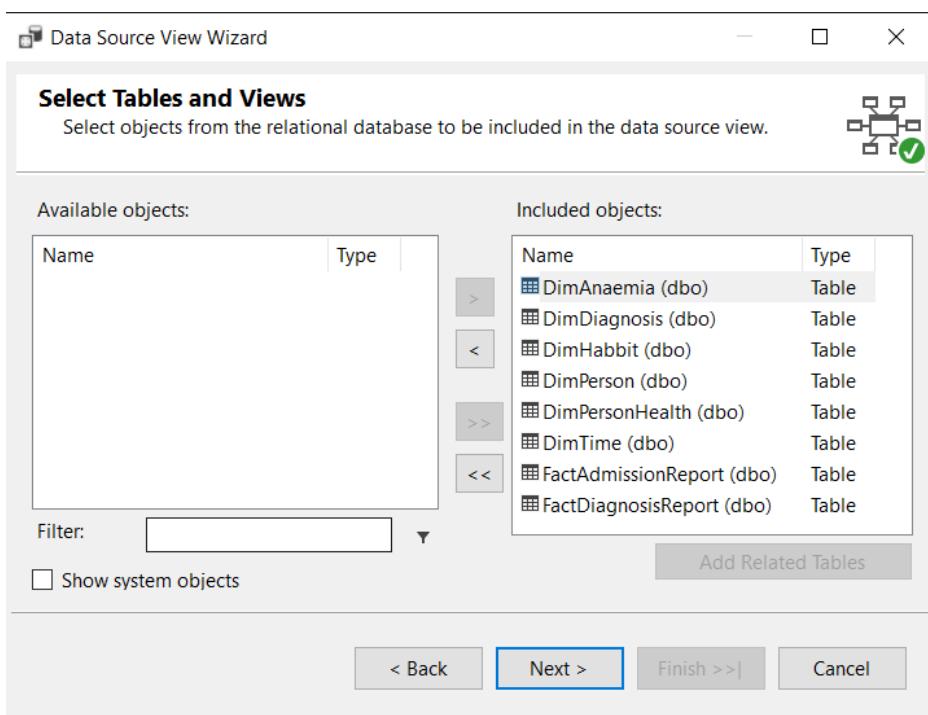
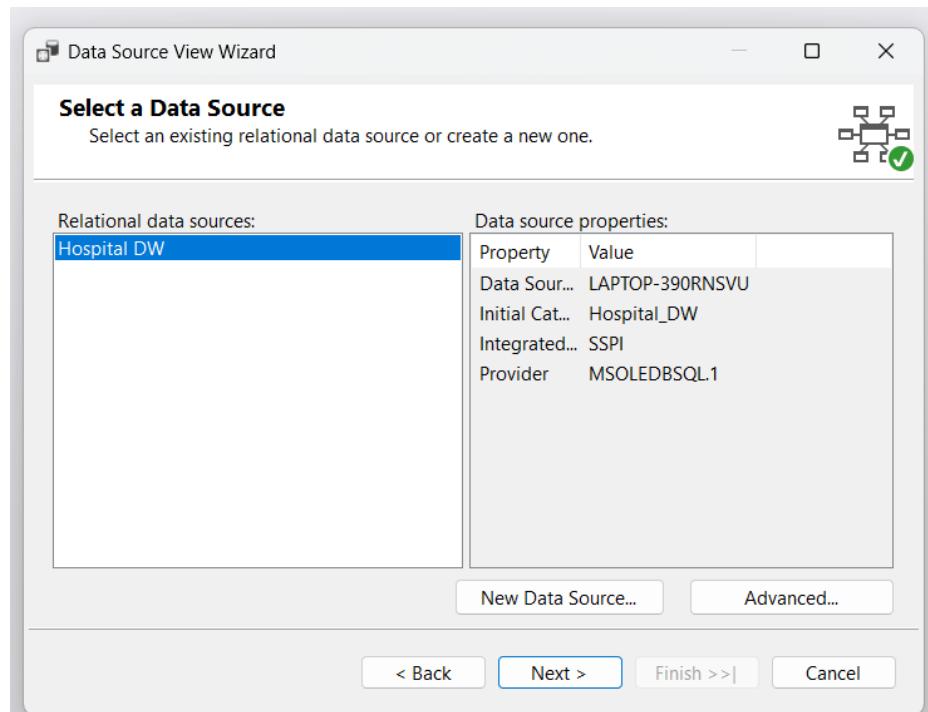
Click Finish để hoàn thành việc kết nối. Sau khi hoàn thành bạn sẽ thấy thông tin về Data Source được hiển thị ở mục Solution Explorer như hình bên dưới.

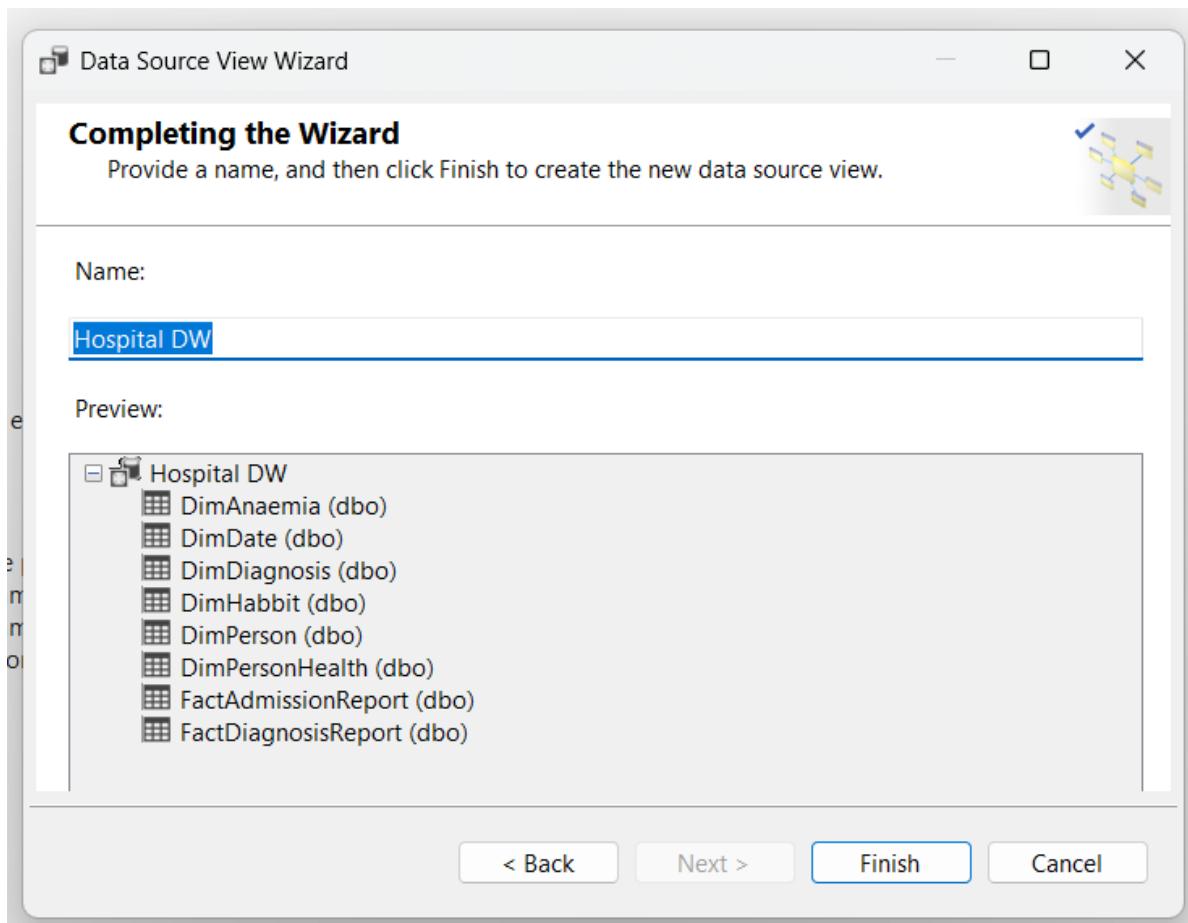


4.1.3. Tạo Data Source View

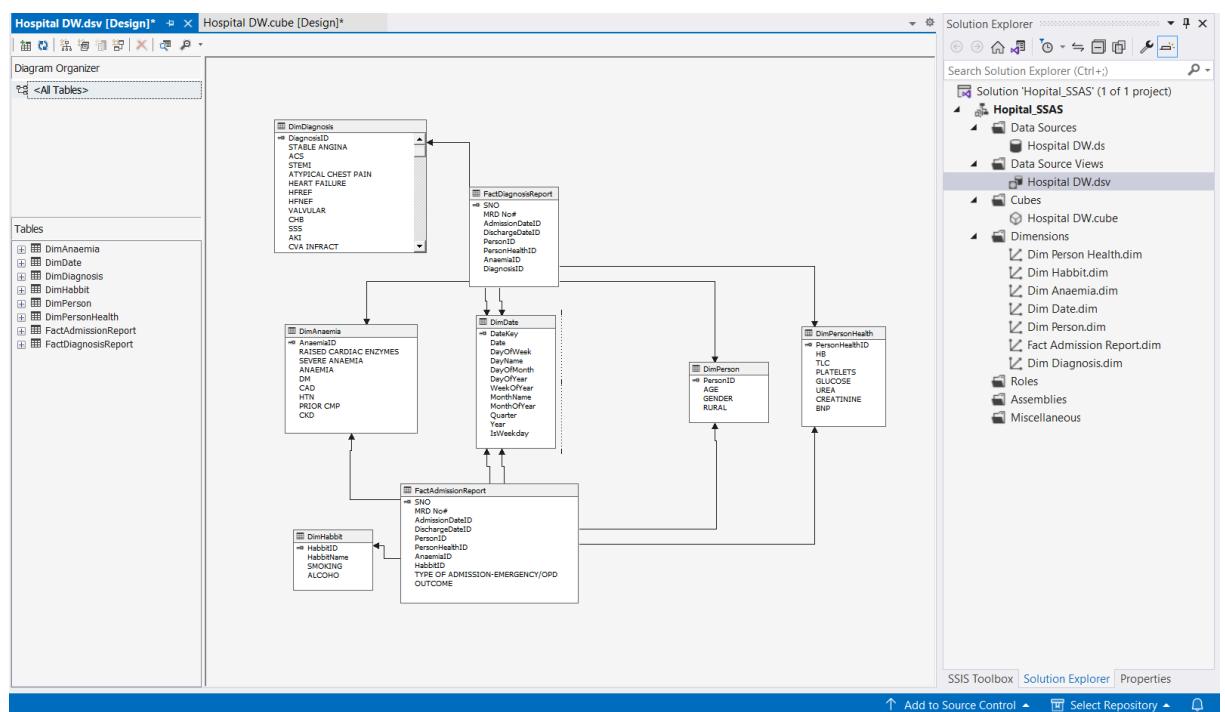
Data Source View chứa logical model của CSDL (tables, keys, columns, và các constraints) sẽ được sử dụng bởi OLAP database để tạo các data cube.

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



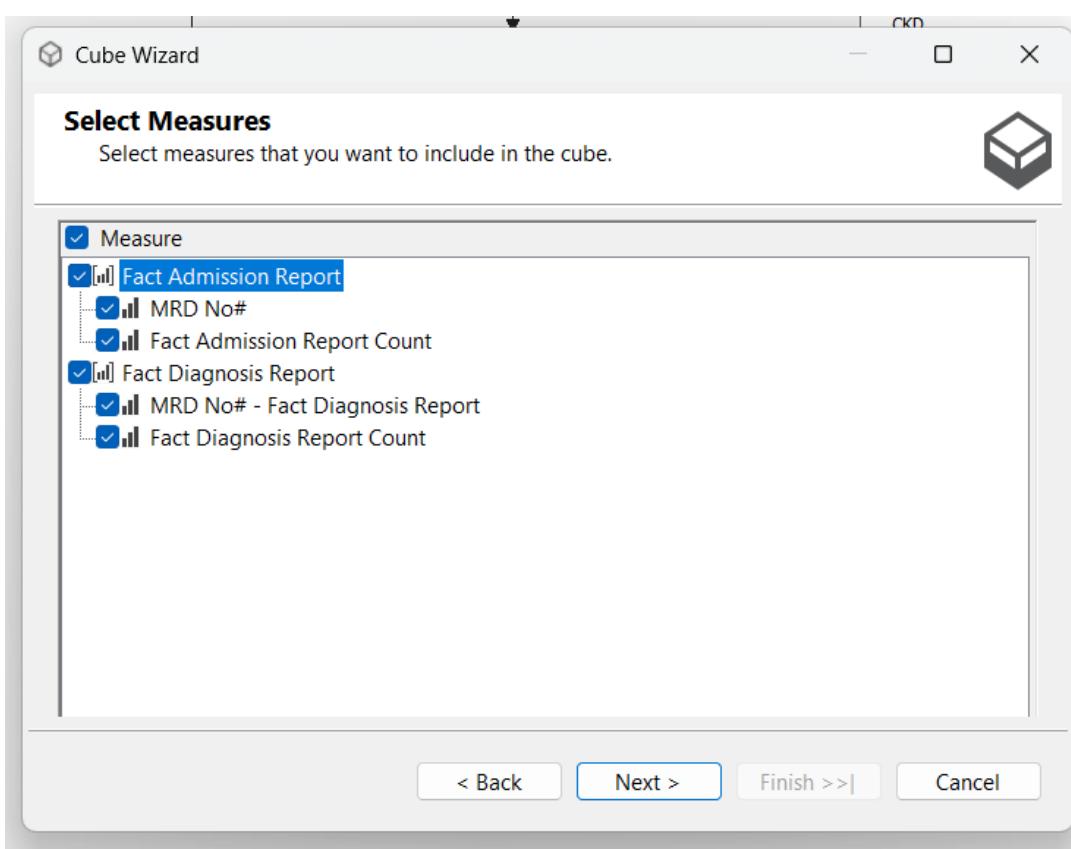
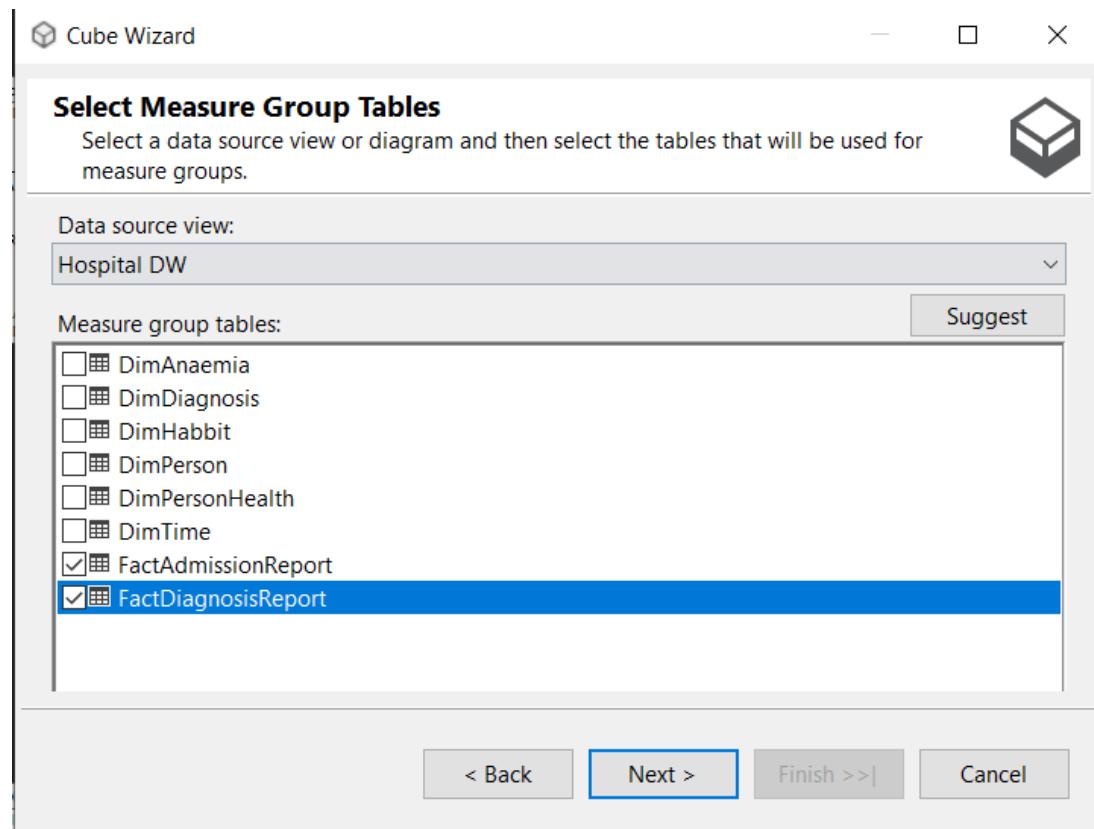


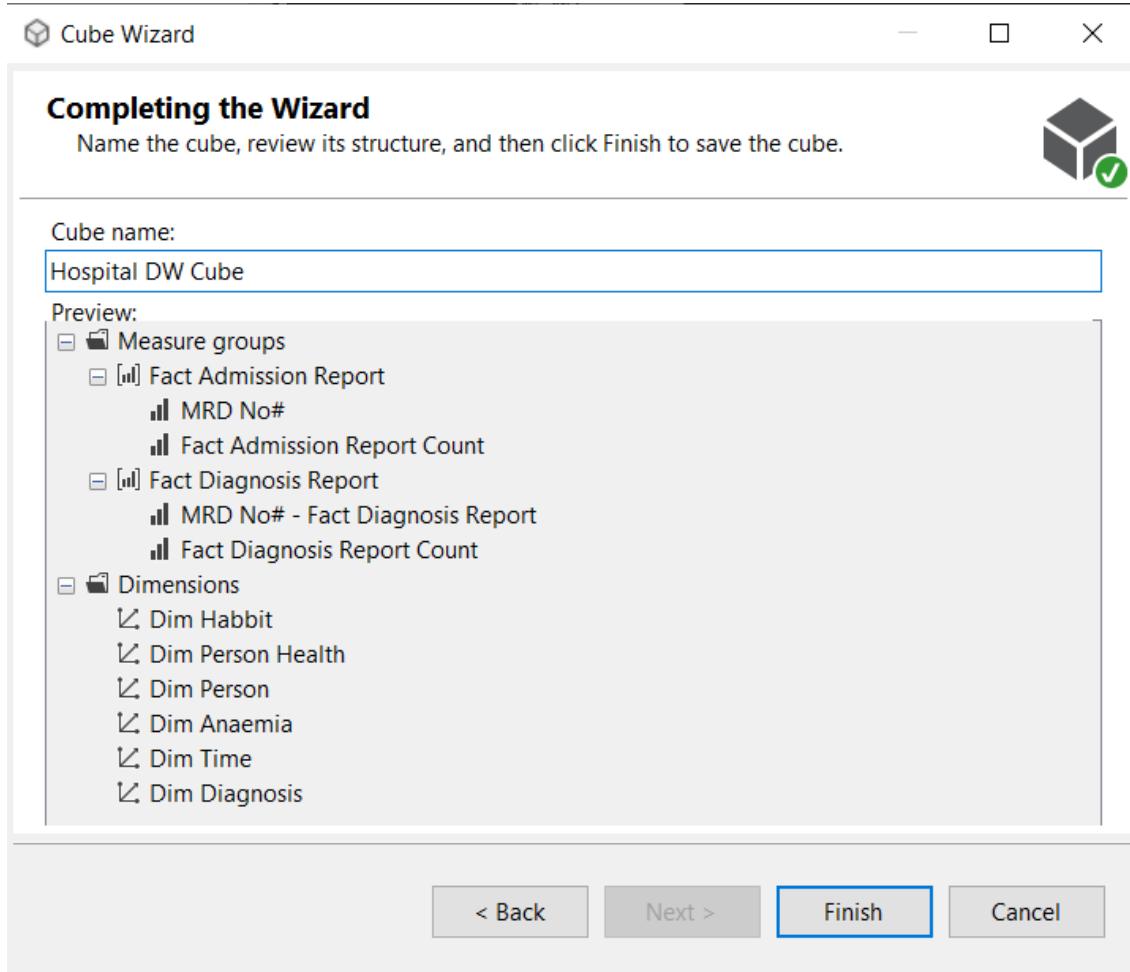
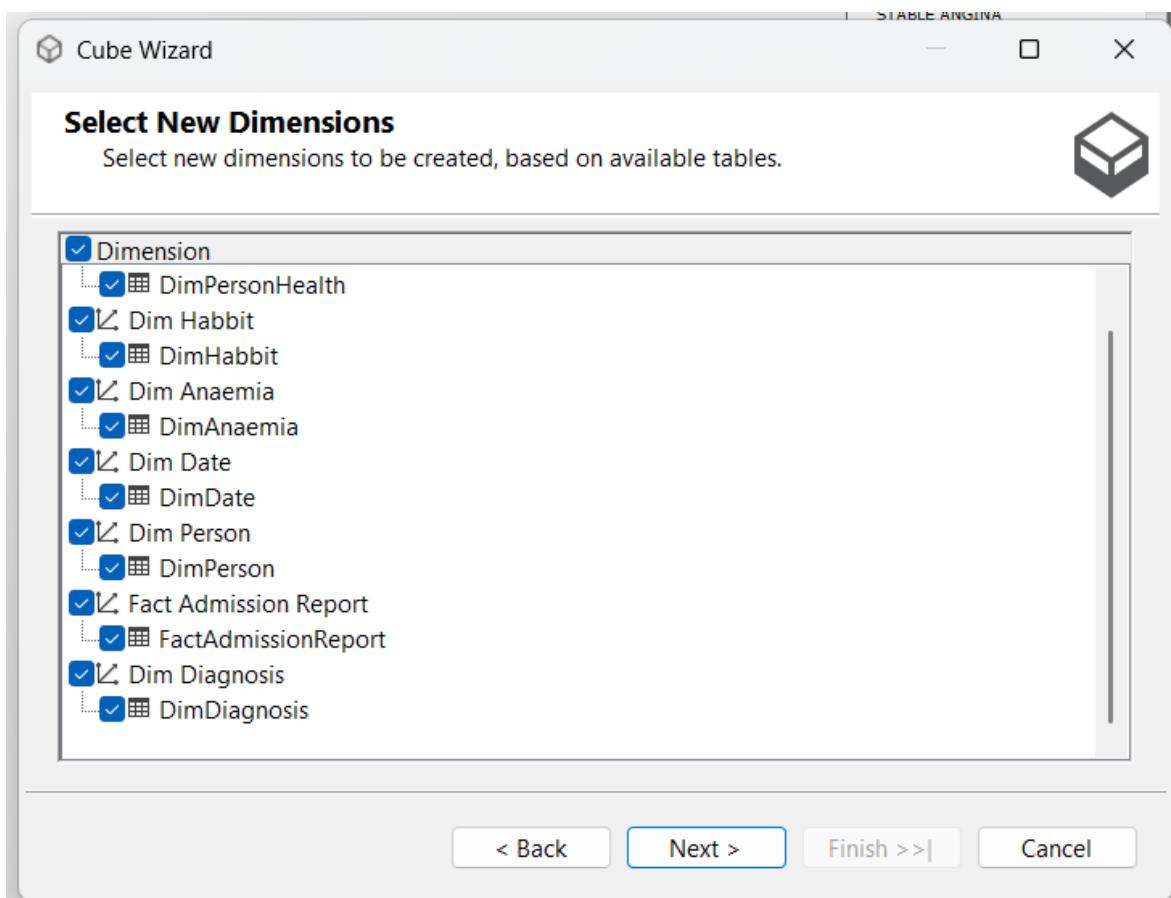
Double-click vào Store Sales Data Source View vừa mới tạo ta sẽ thấy star schema ở Design surface như sau:



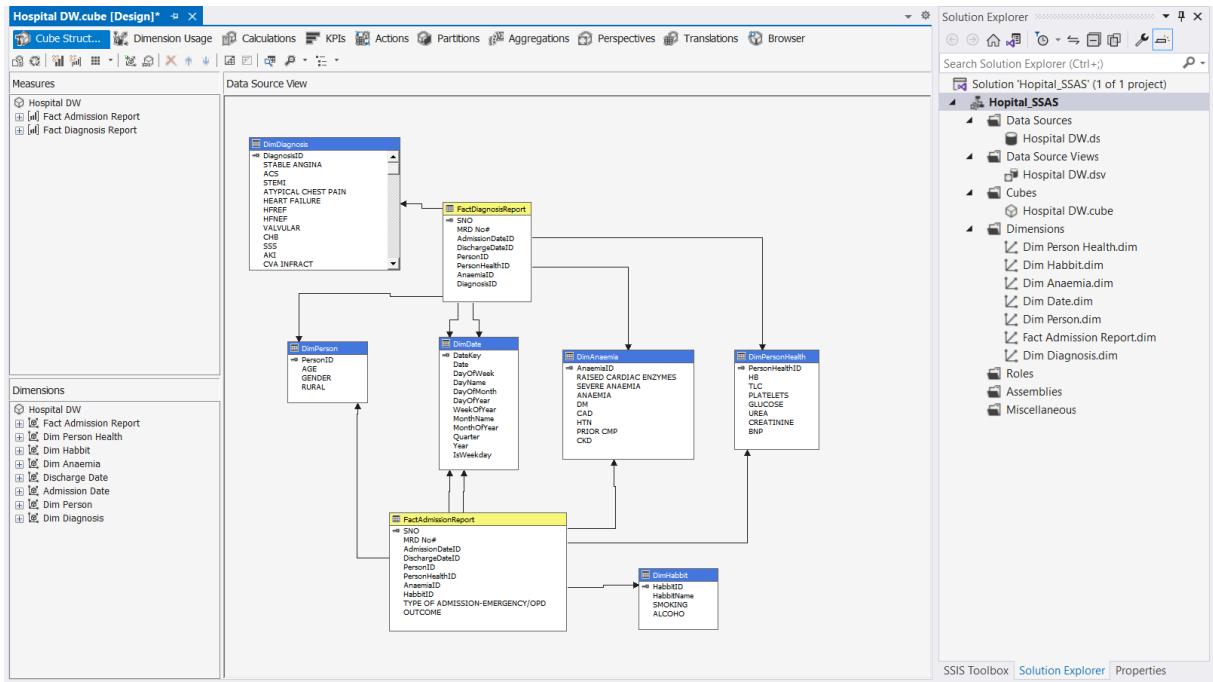
4.2. Quá trình tạo Cube

Right-click vào **Cubes** ở khung **Solution Explorer**. Chon New Cube để tạo Cube mới. Chọn **Data Source View** là Hospital DW View. Tích chọn 2 bảng **FactAdmissionReport** và **FactDiagnosisReport**, nhán Next.



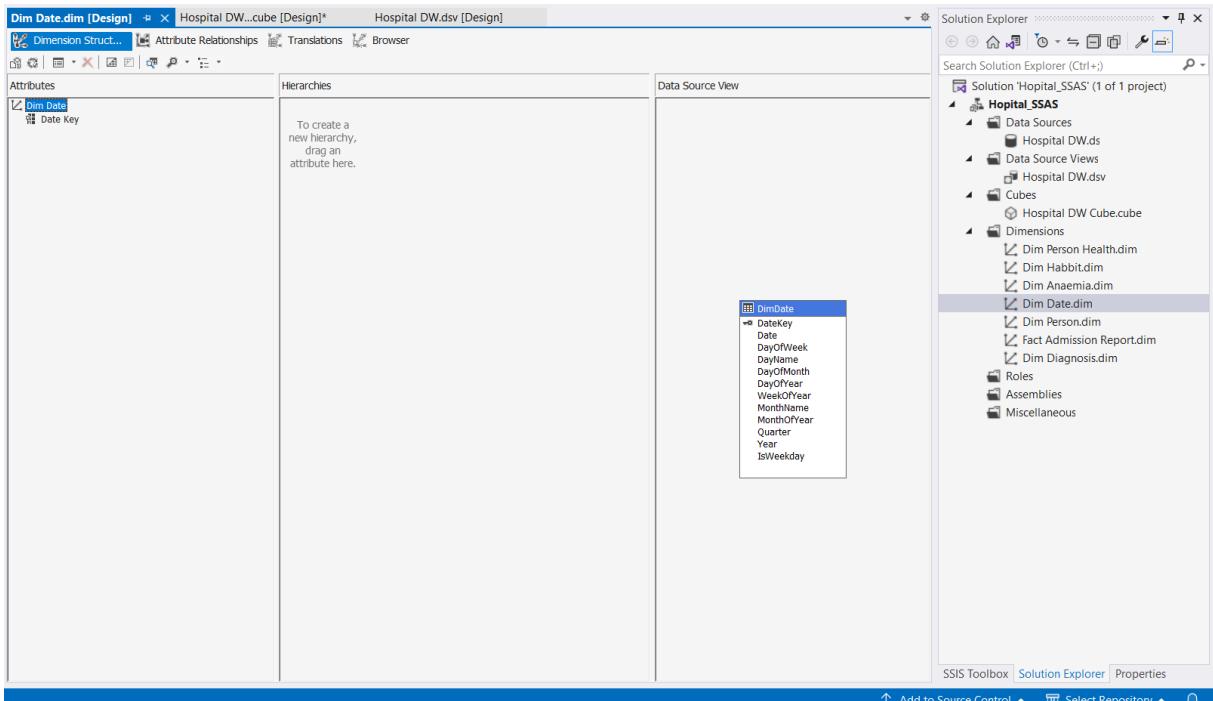


Mặc dù Wizard ở bước trước đã tạo các dimension, ta cần chỉnh sửa lại chúng để có thể thay đổi các thuộc tính (attributes). Click vào Fact Admission Report.dim kéo TYPE OF ADMISSION-EMERGENCY/OPO, OUTCOME từ Data Source View vào Attributes. Ta sẽ có được kết quả như sau:

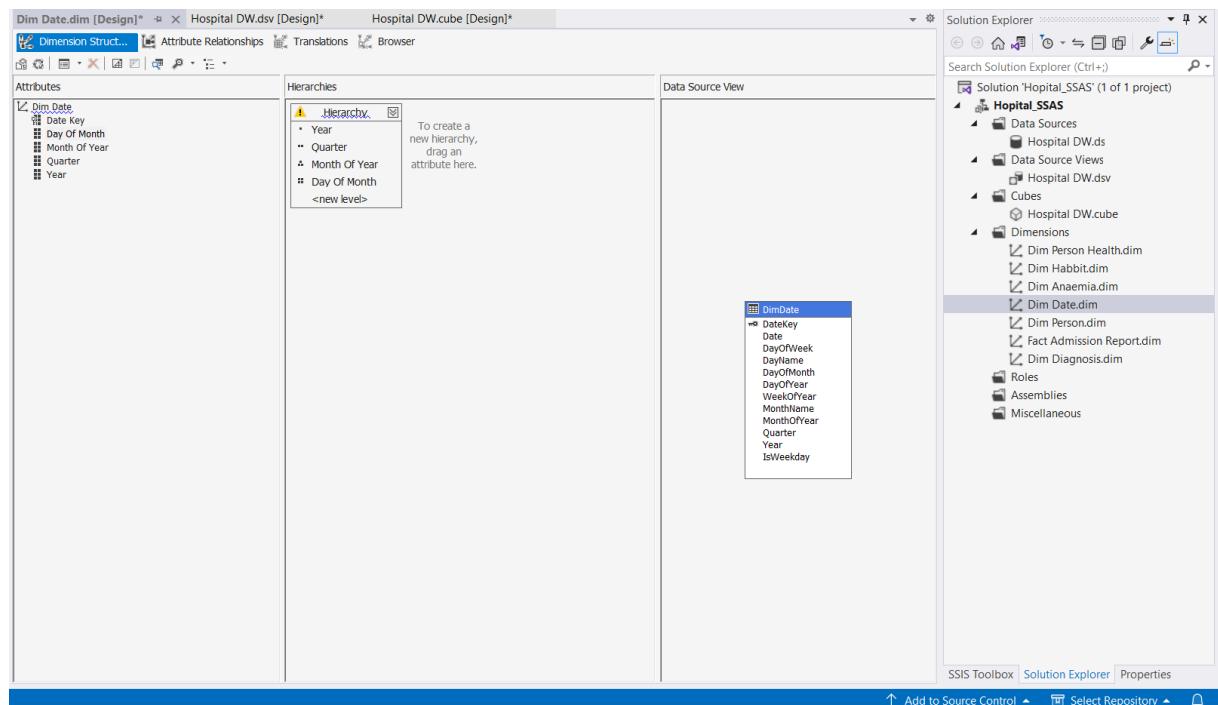


4.3. Quá trình tạo Hierarchy

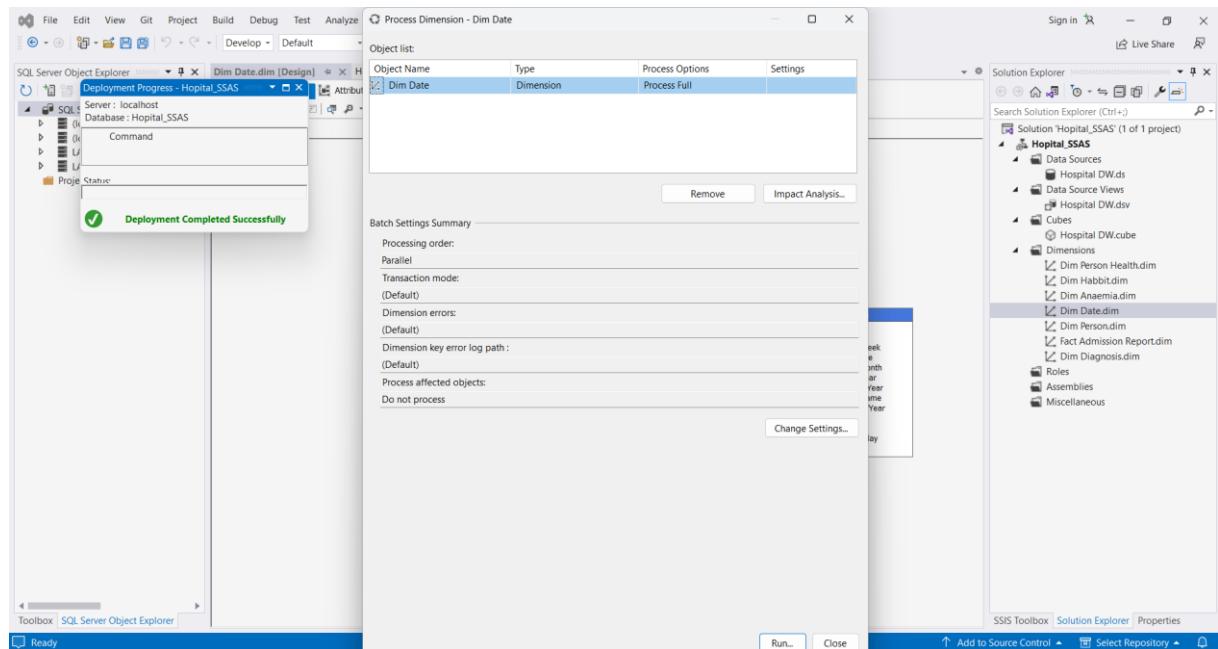
4.3.1. Tạo Hierarchy cho DimDate



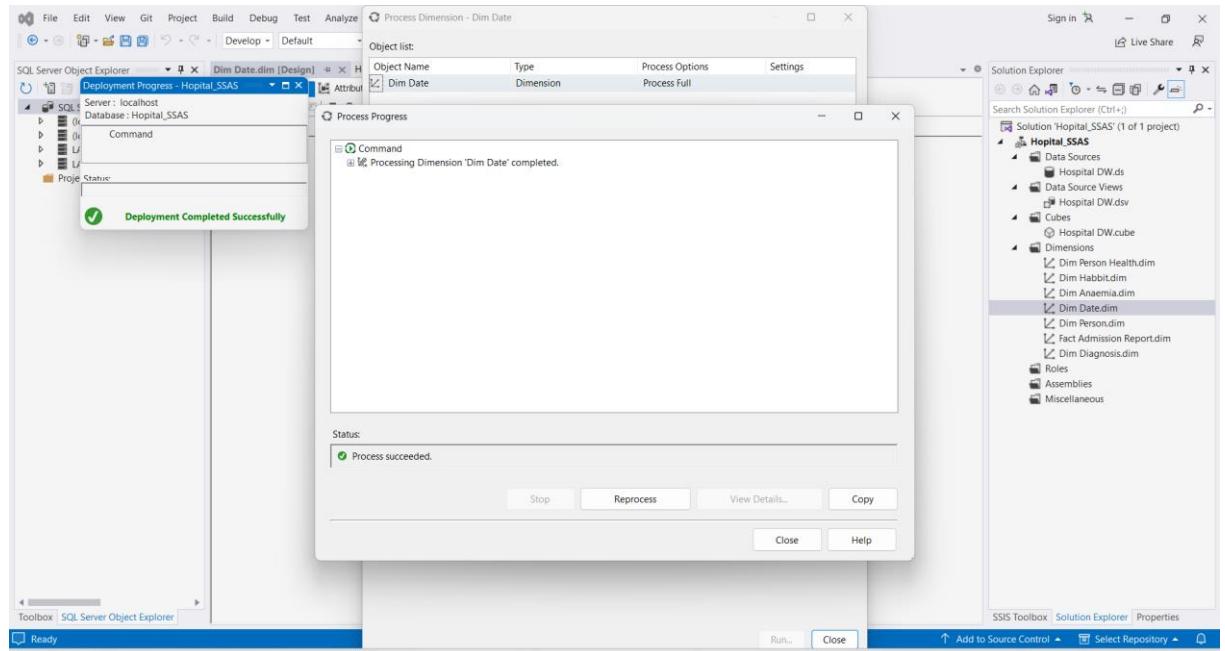
Kéo rê (drag and drop) các cột Day Of Month, Month Of Year, Quarter và Year từ mục Data Source View vào mục Attributes. Ta sẽ được kết quả như hình sau:



Click Process icon () ở toolbar click Yes khi được hỏi để build và deploy project. Nếu build thành công, dialog sau sẽ xuất hiện



Click Run

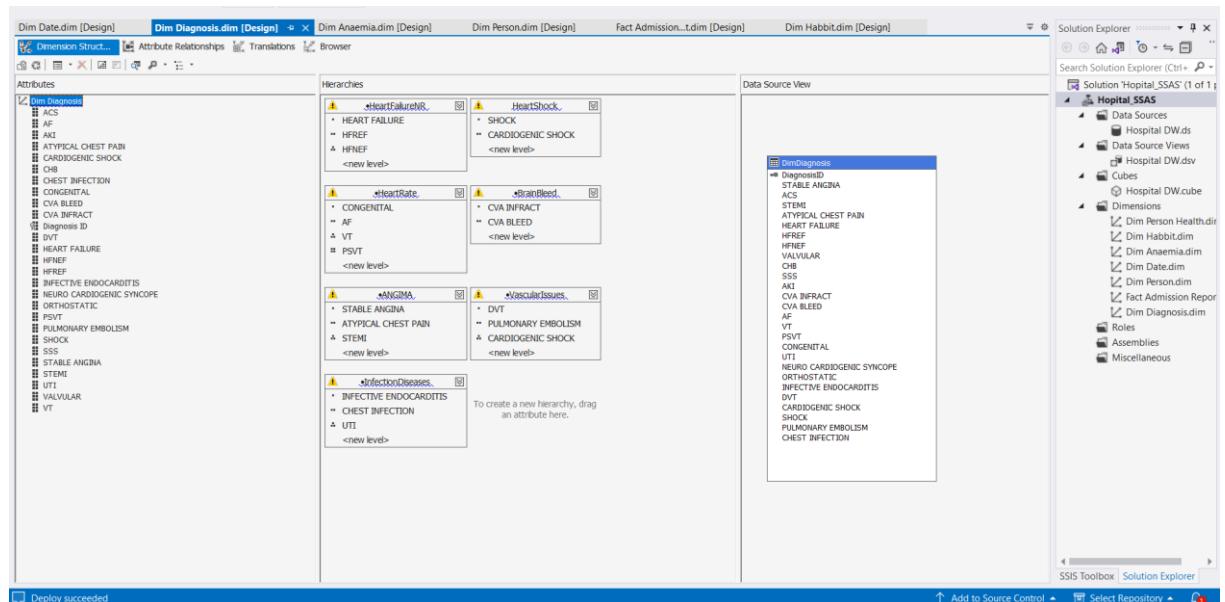


Chuyển đến tab Browser. Ta thấy thời gian được phân cấp theo năm → quý → tháng → ngày.

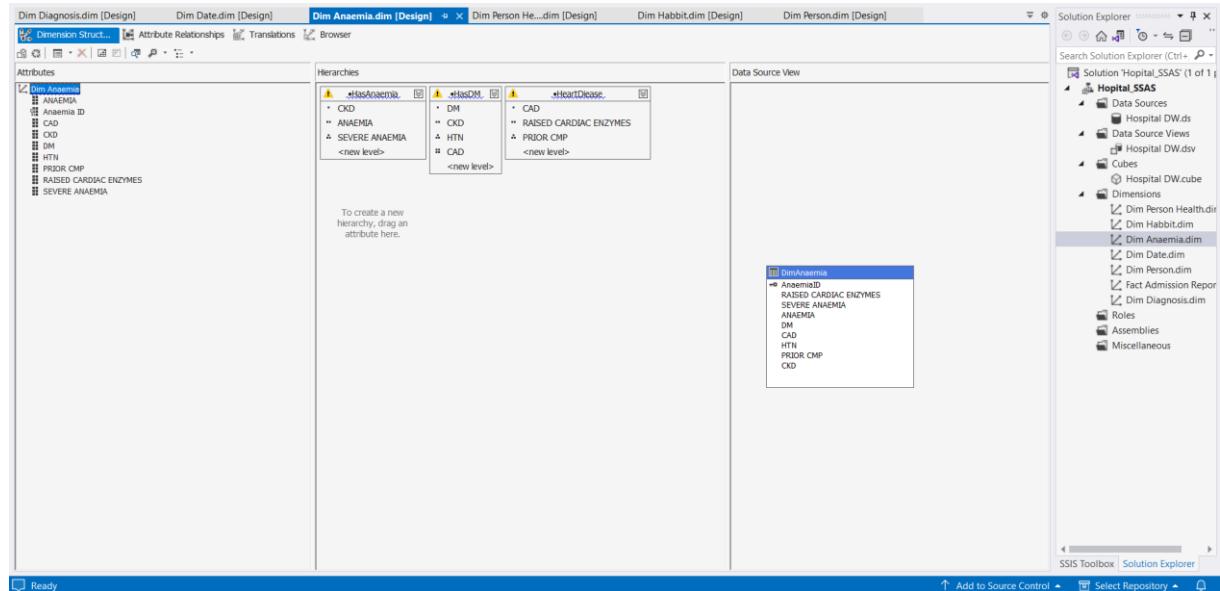
The screenshot shows the 'Browser' tab in the SSAS interface. It displays a hierarchical tree under the 'Current level: (All)' heading. The hierarchy starts with '2017', which branches into '1', '2', and '3'. '1' further branches into '1', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19', '2', '20', '21', '22', '23', '24', '25', '26', '27', '28', '3', '4', '5', '6', '7', '8', and '9'. This visualizes the time dimension's hierarchy from year down to day.

4.3.2. Tạo Hierarchy cho DimDiagnosis

- HeartFailureNR: Phân cấp dựa trên suy tim. Kết quả đánh giá nguyên nhân suy tim thường thông qua nhiều giai đoạn trong đó có sử dụng thước đo phân suất tổng máu (EF - Ejection Fraction). Sau khi sàng lọc và chẩn đoán thường đưa ra 1 trong 2 kết luận về suy tim là bệnh suy tim với phân suất tổng máu giảm và suy tim với phân suất tổng máu bảo tồn.
- HeartShock: Phân cấp dựa trên triệu chứng shock. Shock có nhiều nguyên nhân trong đó có shock tim. Nếu không được điều trị nhanh chóng, shock tim có thể gây nguy hiểm đến tính mạng người bệnh. Biến chứng nguy hiểm khác là gây tổn thương không phục hồi tại gan, thận và một số cơ quan khác do bị thiếu cung cấp oxy.
- HeartRate: Phân cấp dựa trên các yếu tố liên quan đến nhịp tim.
- BrainBleed: Phân cấp dựa trên hiện tượng tai biến mạch máu não và tình trạng nghiêm trọng của nó.
- Angina: Phân cấp dựa trên mức độ đau thắt ngực. Đau thắt ngực không ổn định thường đi kèm theo triệu trứng nhồi máu cơ tim.
- VascularIssues: Phân cấp dựa trên các vấn đề về mạch máu.
- InfectionDiseases: Phân cấp dựa trên các vấn đề liên quan đến nhiễm trùng.



4.3.3. Tạo Hierarchy cho DimAnaemia

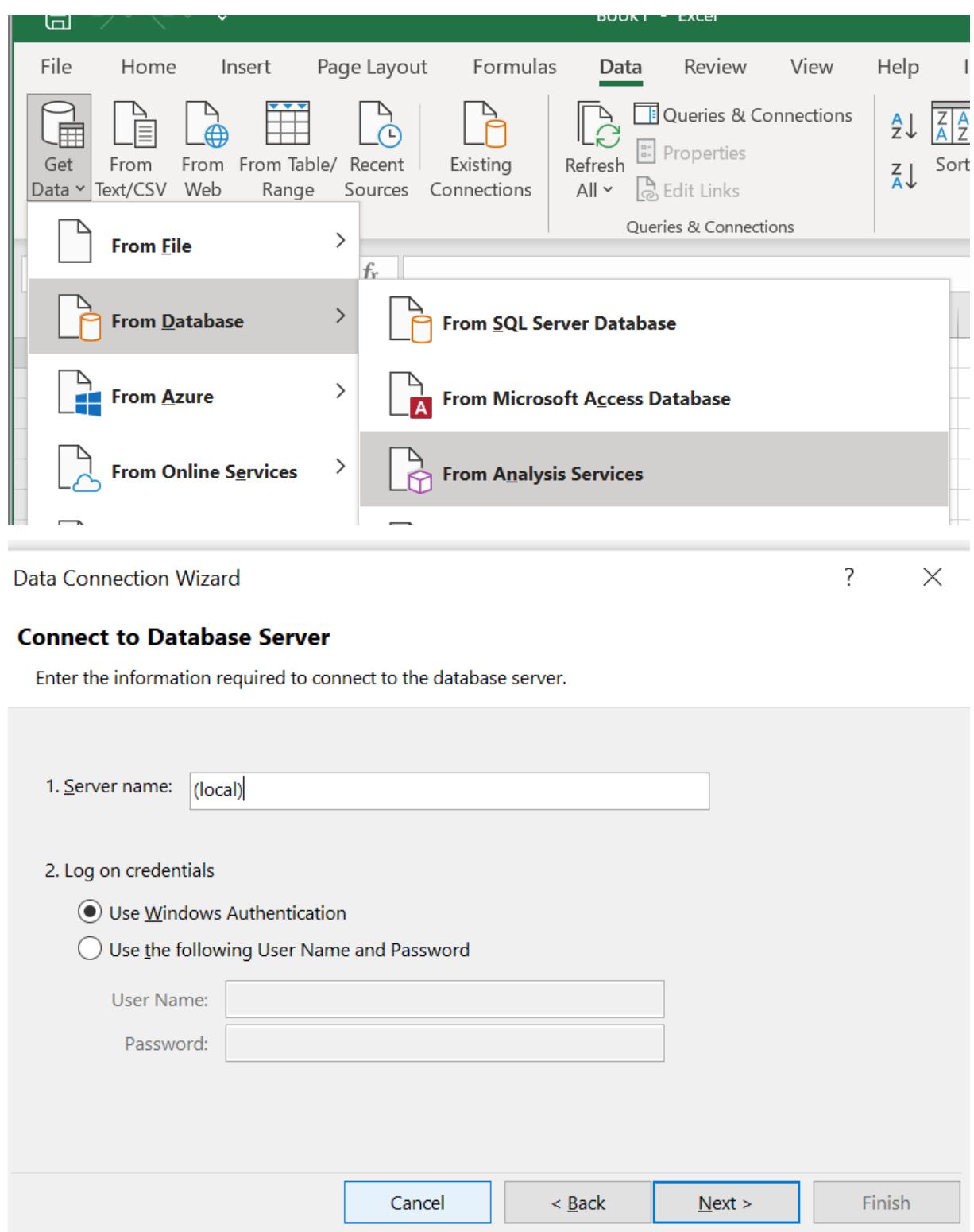


- HasAnaemia: Thiếu máu ở bệnh nhân suy thận mạn là một tình trạng liên quan đến việc suy giảm chức năng thận, với các rối loạn huyết học, rối loạn hormone và dạ dày ruột. Cơ chế thiếu máu ở bệnh nhân suy thận mạn là do thiếu Erythropoietin (EPO) và thiếu sắt. Trong đó, cơ chế chủ yếu là thiếu Erythropoietin (EPO).
- HasDM: Đái tháo đường nếu không được kiểm soát sẽ gây ra các biến chứng cấp và mạn đến vi mạch thận, dẫn đến suy thận. Mà suy thận thường dẫn đến huyết áp cao. Tăng huyết áp lâu ngày sẽ làm hư lớp nội mạc của mạch vành và gây ra những bệnh liên quan đến động mạch vành.
- HeartDisease: bệnh liên quan đến động mạch ngày càng phổ biến. Và thường bác sĩ sẽ chỉ định xét nghiệm men tim khi mắc hội chứng này. Men tim như creatine phosphokinase, lactate dehydrogenase, troponine T, CK-MB, SGOT (AST), SGPT (ALT) có thể tăng khi viêm cơ tim cấp và mạn tính. Tuy nhiên nếu các men này là âm tính cũng không loại trừ được tình trạng viêm cơ tim. Đối với cái dim còn lại chưa chung ta sẽ kéo các cột từ Data Source View vào Attributes.

4.4. Tiến hành trả lời các câu hỏi Business Process

4.4.1. Sử dụng công cụ Excel Pivot Table

Ta tiến hành cấu hình liên kết từ Excel tới Dữ liệu từ Analysis Service.



Data Connection Wizard

?

X

Select Database and Table

Select the Database and Table/Cube which contains the data you want.

Select the database that contains the data you want:

Hopital_SSAS

Connect to a specific cube or table:

Name	Description	Modified	Created	Type
Hospital DW		5/22/2024 12:44:31 PM		CUBE



Cancel

< Back

Next >

Finish

Import Data

?

X

Select how you want to view this data in your workbook.



Table



PivotTable Report



PivotChart



Only Create Connection

Where do you want to put the data?

Existing worksheet:

=A\$1



New worksheet

Add this data to the Data Model

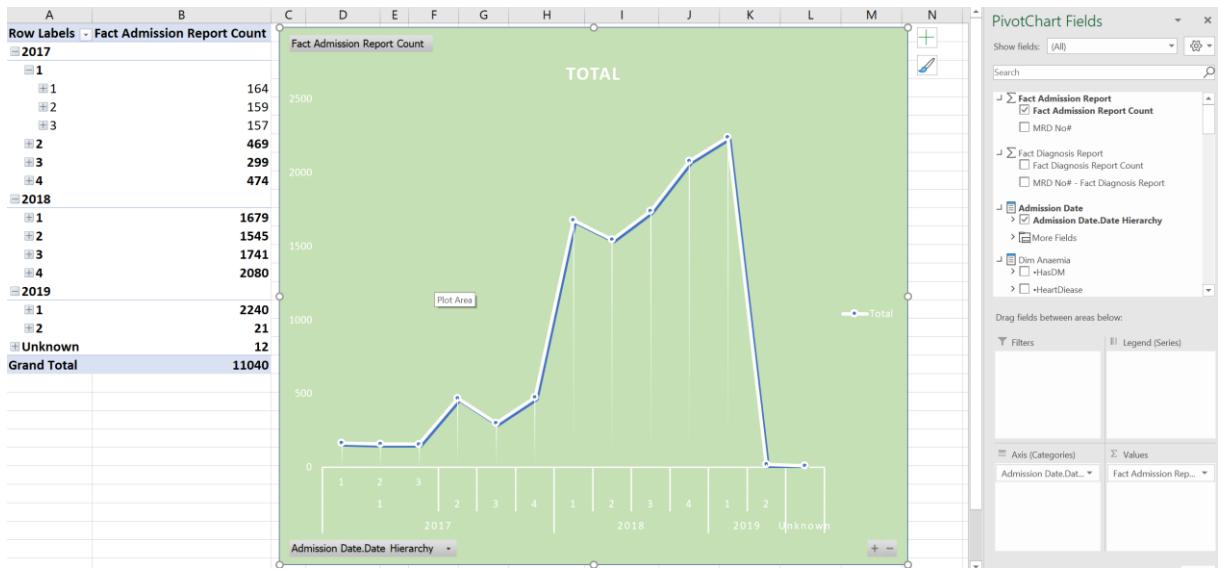
Properties...

OK

Cancel

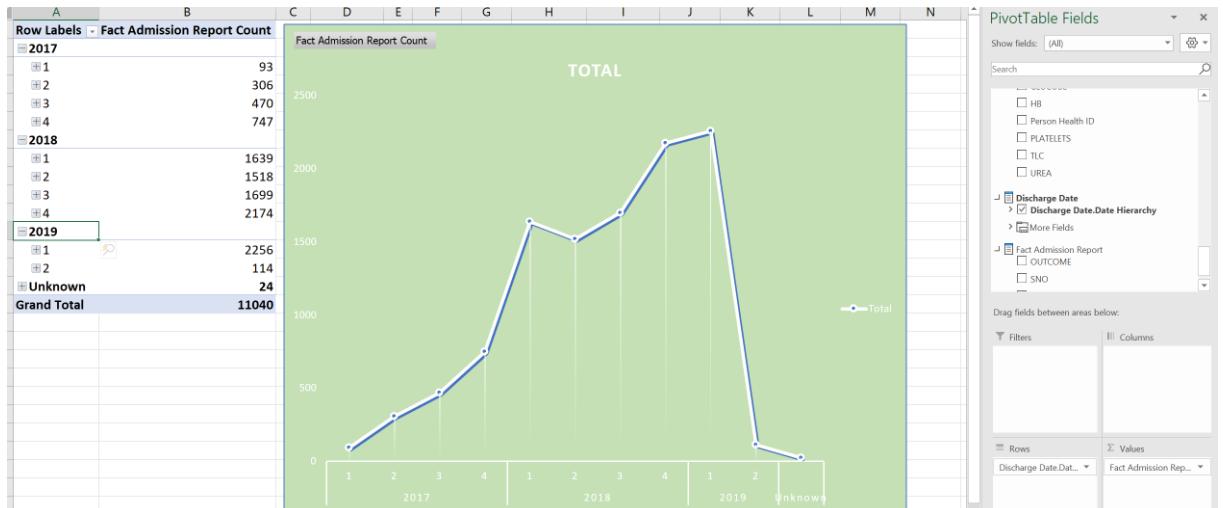
Ta tiến hành tạo các bảng pivot để trả lời các câu hỏi sau:

Câu 1: Số lượng bệnh nhân nhập viện mỗi ngày?



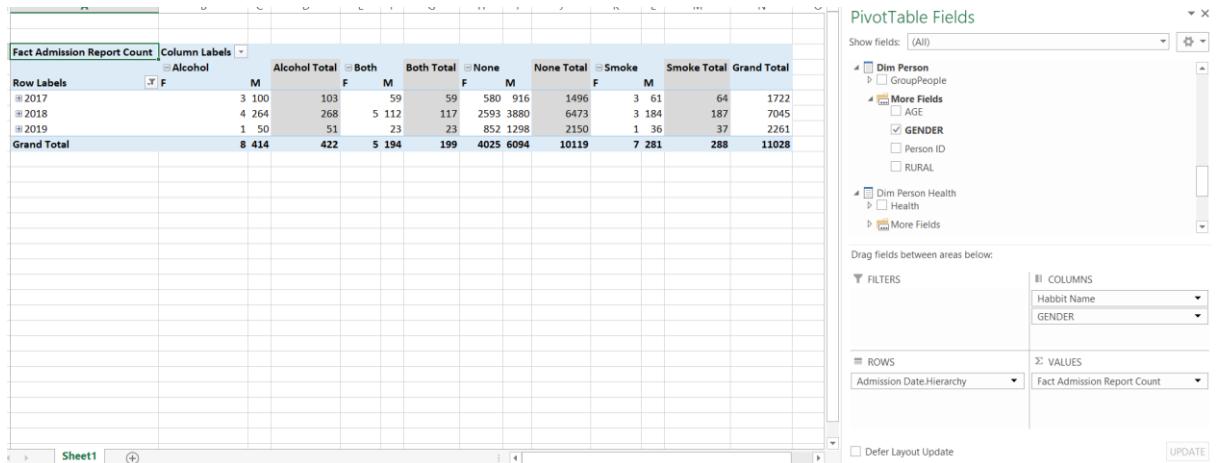
Ta thấy số lượng bệnh nhân nhập viện nhiều nhất là vào năm 2018.

Câu 2: Số lượng bệnh nhân xuất viện mỗi ngày?



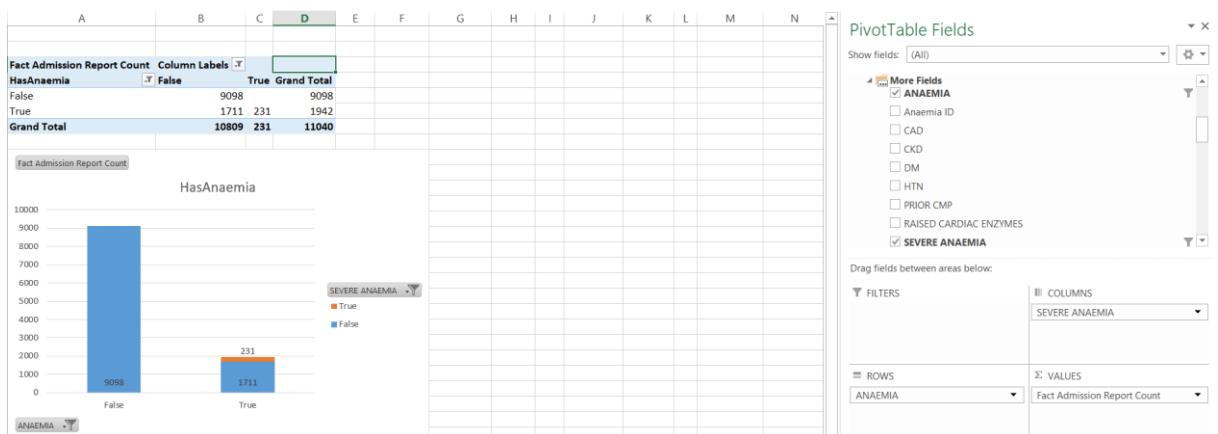
Ta cũng thấy được số lượng bệnh nhân xuất viện nhiều nhất cũng là vào năm 2018.

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



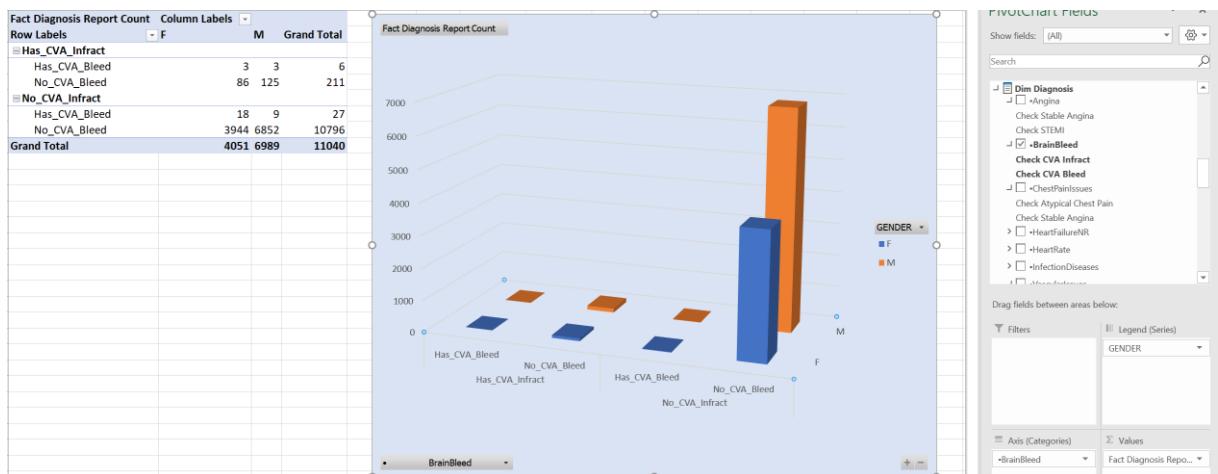
Ta thấy đa số bệnh nhân không hút thuốc và uống rượu (10119 bệnh nhân). Trong đó số lượng phụ nữ có những thói quen như hút thuốc, rượu bia thấp hơn hẳn so với đàn ông.

Câu 4: Có bao nhiêu 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% những người bị thiếu máu bị thiếu máu trầm trọng.

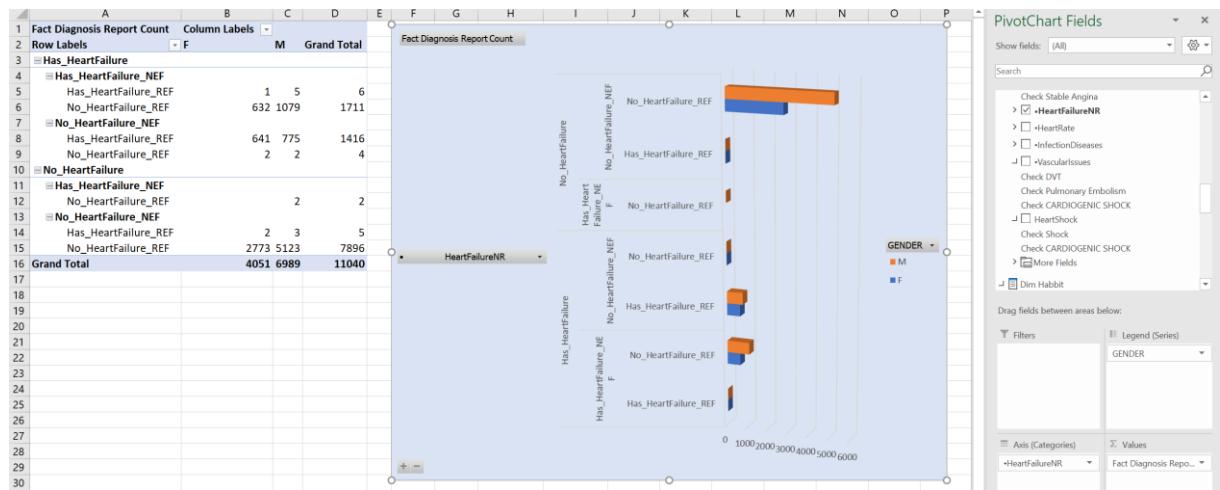
Câu 5: Nguyên nhân do đâu mà bệnh nhân bị đột quỵ?



27 ca đột quy chỉ do xuất huyết não, 211 ca đột quy chỉ 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.

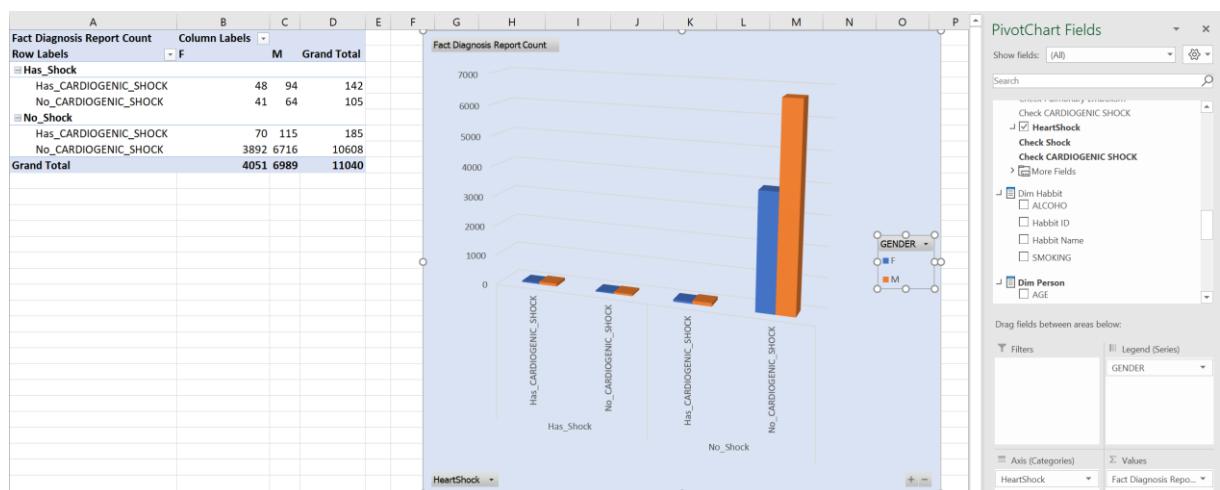
Câu 6: Có bao nhiêu bệnh nhân bị hội chứng suy tim?



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.

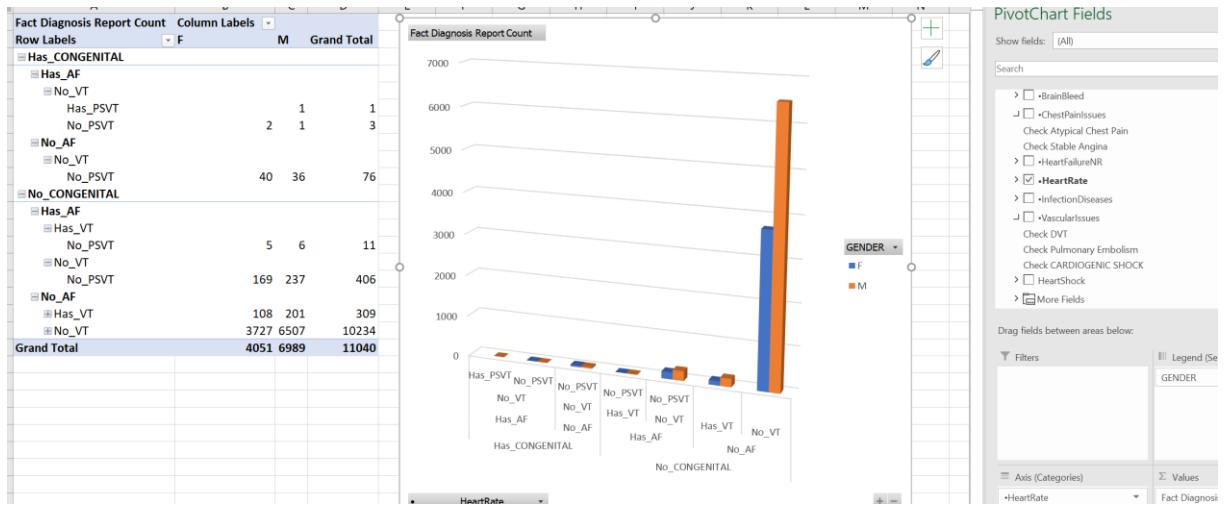
Câu 7: Có bao nhiêu bệnh nhân bị sốc tim theo giới tính?



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.

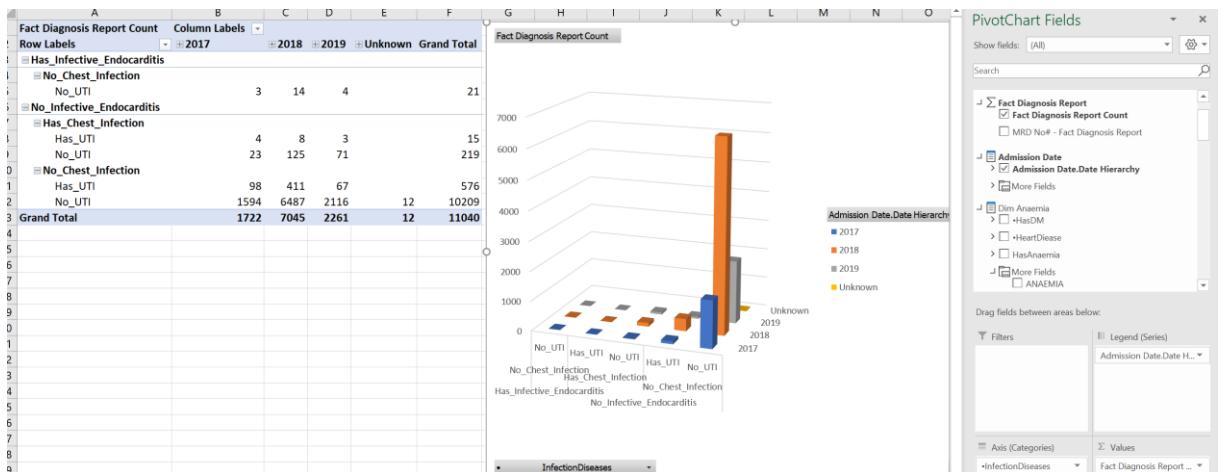
Câu 8: Có bao nhiêu bệnh nhân mắc các vấn đề về nhịp tim bất thường?



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.

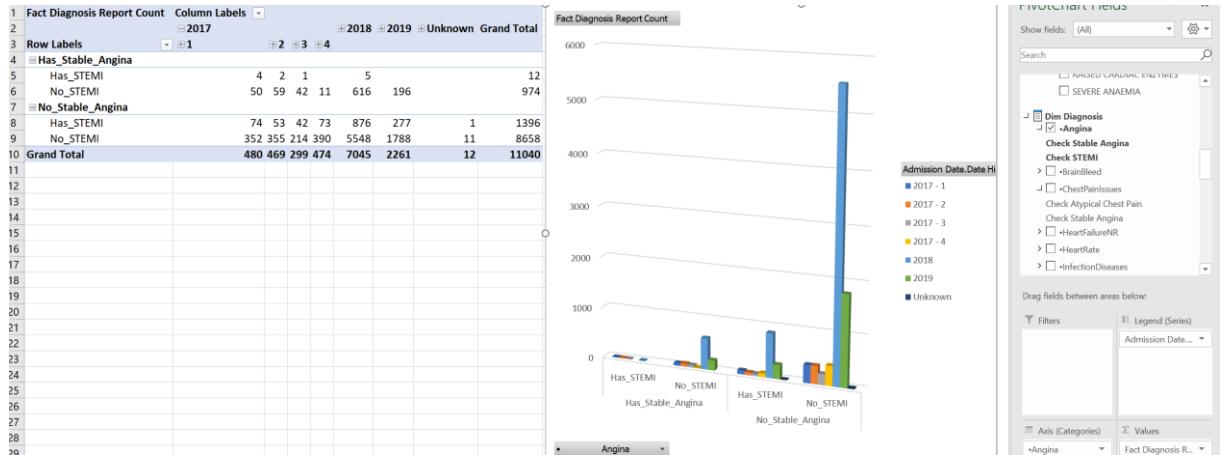
Câu 9: Có bao nhiêu bệnh nhân mắc các bệnh nhiễm trùng theo thời gian?



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



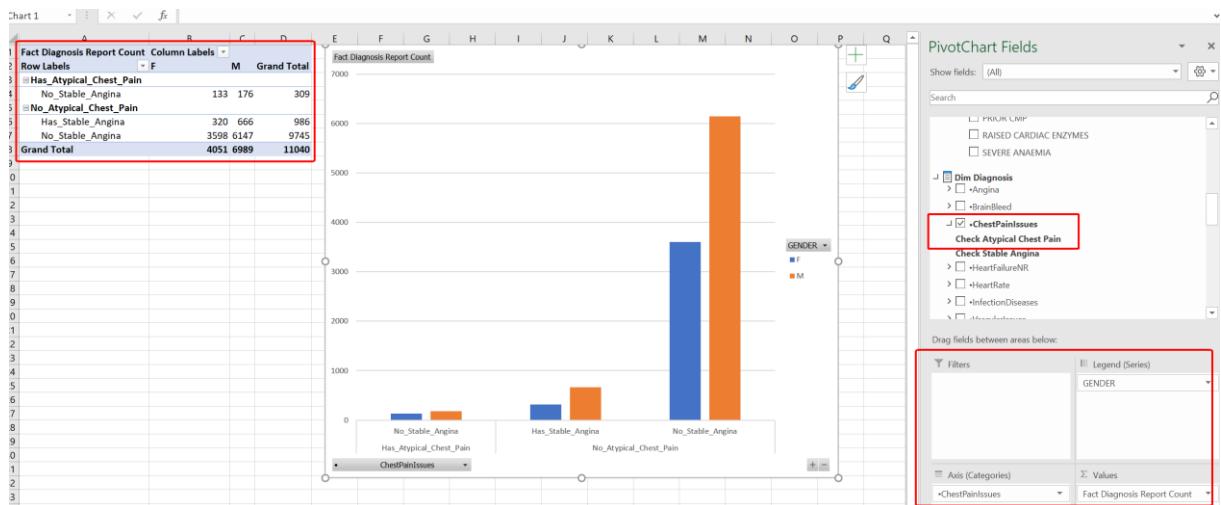
Có 1396 ca bị nhồi máu cơ tim cấp tính (STEMI). 974 ca bị đau thắt ngực liên tục (Stable Angina). 12 ca bị cả 2 đau thắt ngực liên tục và nhồi máu cơ tim. Ta thấy đau thắt ngực liên tục (Stable Angina) và nhồi máu cơ tim (STEMI) có vẻ không liên quan nhau lắm.

Câu 11: Có bao nhiêu bệnh nhân mắc các bệnh đau ngực?

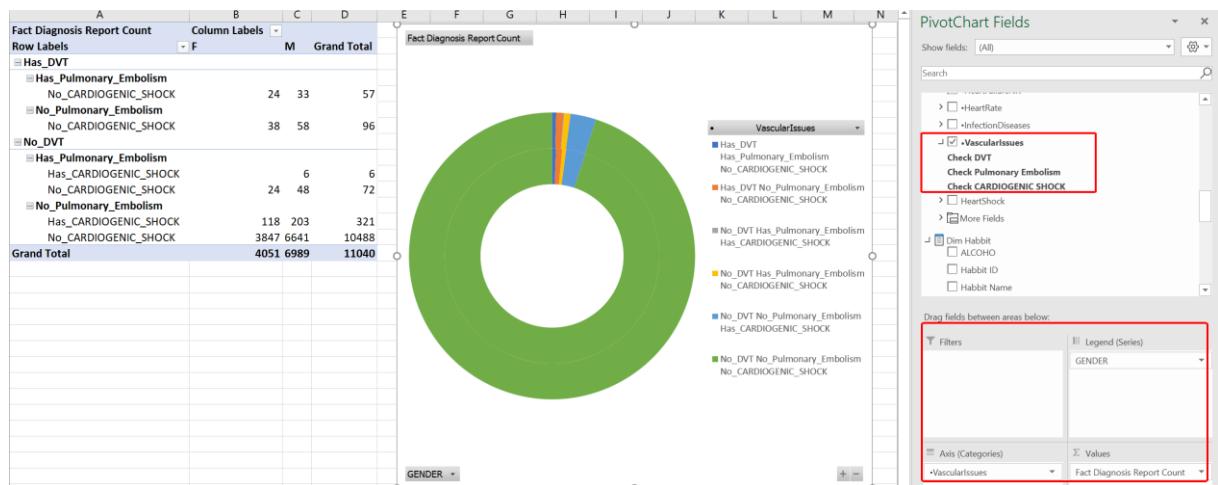
Ta thấy có

320 ca co thắt ngực liên tục nhưng không bị đau ngực không đi kèm hình (ACP) là nữ và 666 ca là nam. 133 ca đau ngực không đi kèm hình (ACP) nhưng không bị co thắt ngực liên tục là nữ và 176 là nam.

→ Ta thấy nam giới gấp nhiều ván đè về đau ở ngực hơn nữ giới (có thể do hút thuốc)



Câu 12: Có bao nhiêu bệnh nhân bị các vấn đề về mạch máu như thuyên tắc tĩnh mạch sâu hay thuyên tắc phổi?



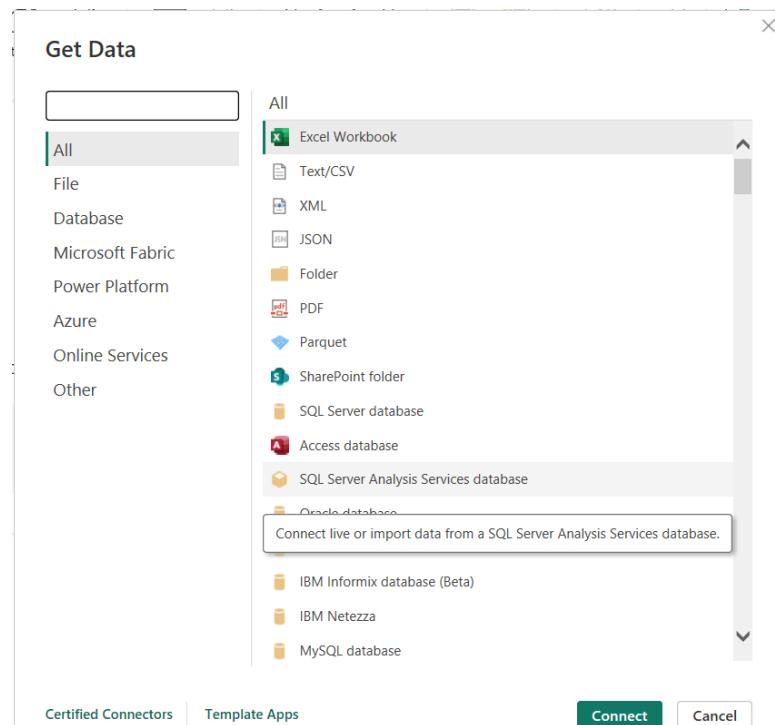
321 ca bị sốc tim (CARDIOGENIC SHOCK). 72 ca thuyên tắc phổi (PULMONARY EMBOLISM). 6 ca vừa thuyên tắc phổi vừa sốc tim. 96 ca thuyên tắc tĩnh mạch sâu (DVT). 57 ca vừa thuyên tắc tĩnh mạch sâu vừa thuyên tắc phổi. Nhưng vẫn có thể thấy rằng nam giới gấp các vấn đề về mạch máu nhiều hơn nữ.

→ Ta thấy các bệnh về thuyên tắc mạch máu khá liên quan với nhau.

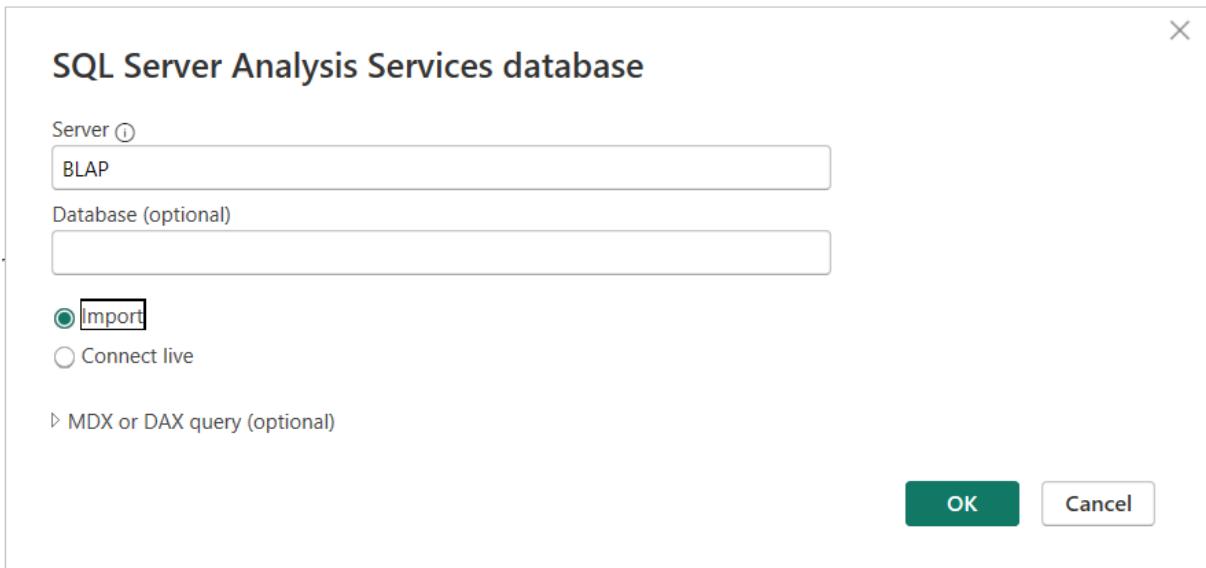
4.4.2. Sử dụng Power BI

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** và click **Connect**.



2. Chọn tên **Server** và đặt chế độ kết nối là **Import**.



3. Chọn các Dim cần dùng tới và tiến hành Load dữ liệu.

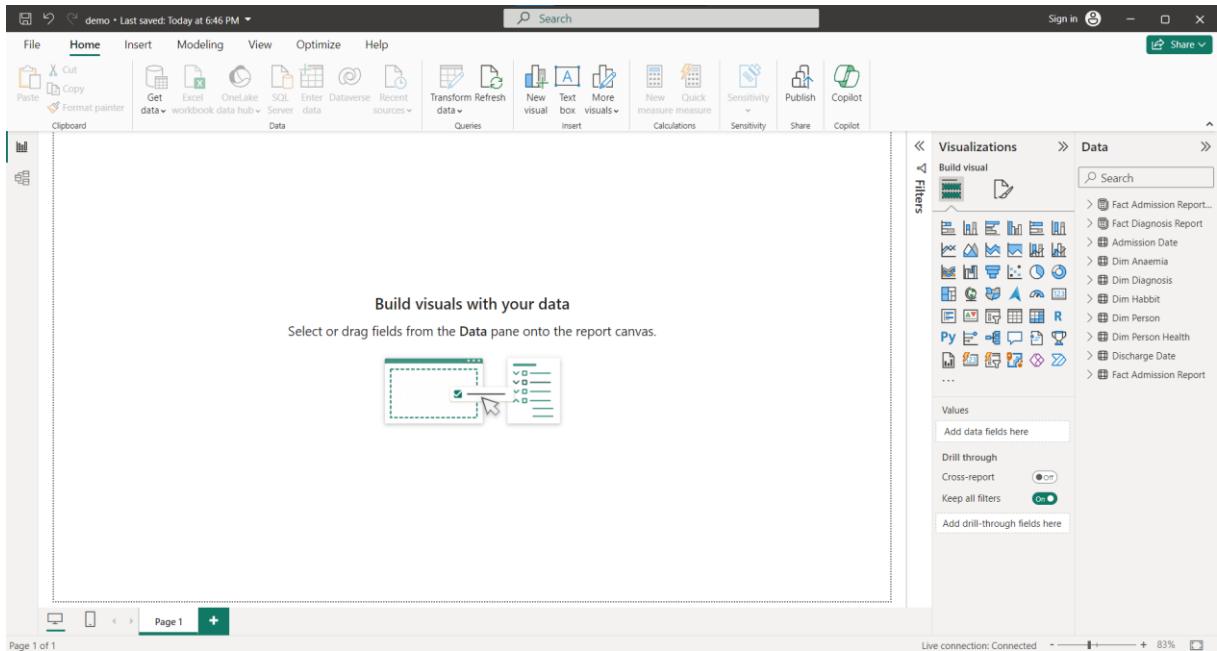
Hospital DW

Admission Date.Date Key	Admission Date.Day Of Month	Admission Date.Month Of Year
20170101	1	1
20170101	1	1
20170101	1	1
20170101	1	1
20170101	1	1

The data in the preview has been truncated due to size limits.

Load Transform Data Cancel

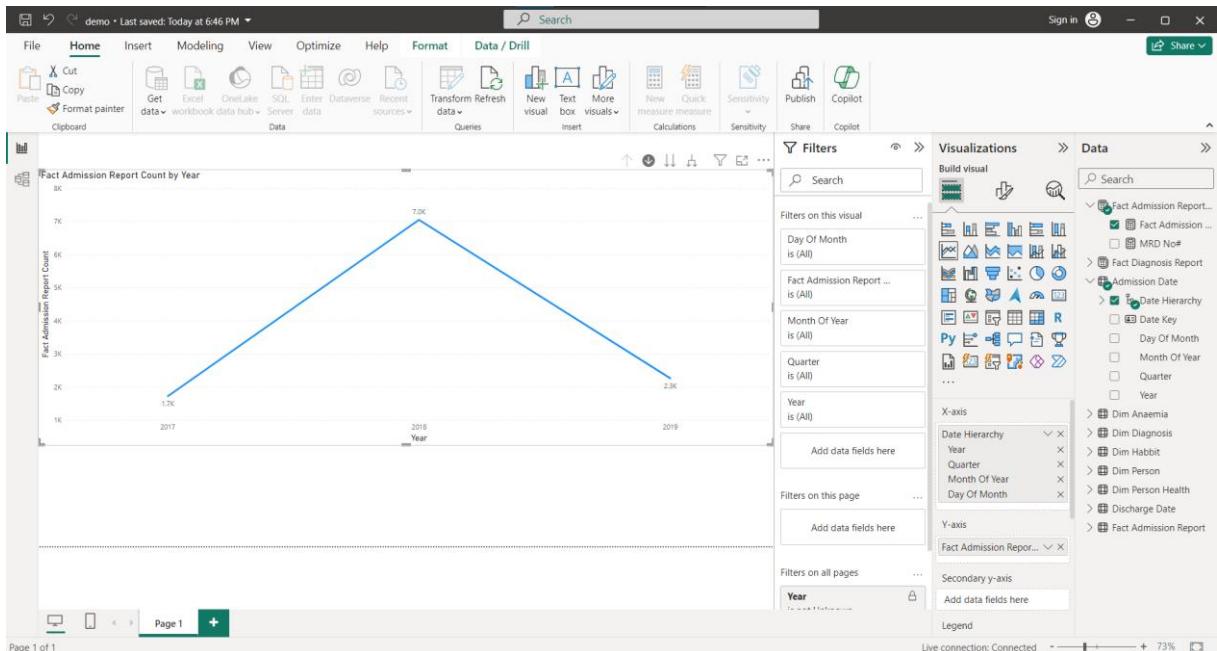
4. Cửa sổ sẽ hiện ra như sau:



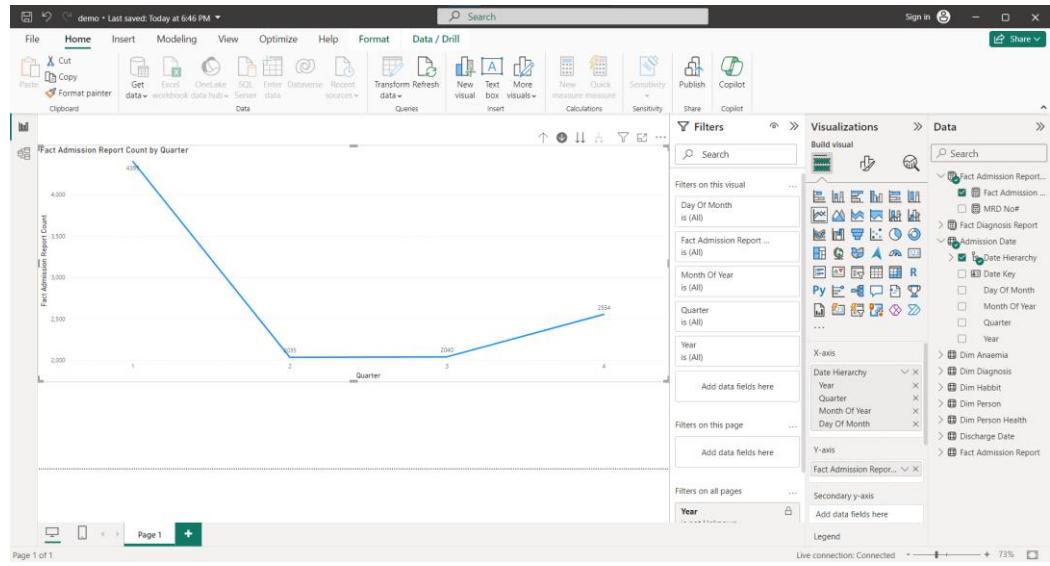
5. Tiến hành trả lời các câu hỏi

Câu 1: Số lượng bệnh nhân nhập viện theo thời gian?

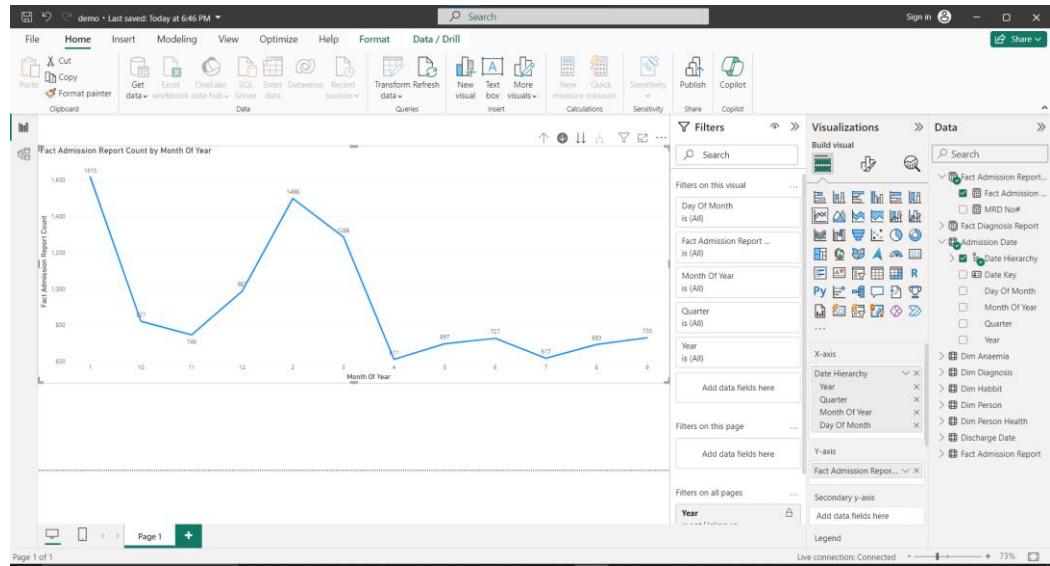
- Theo năm:



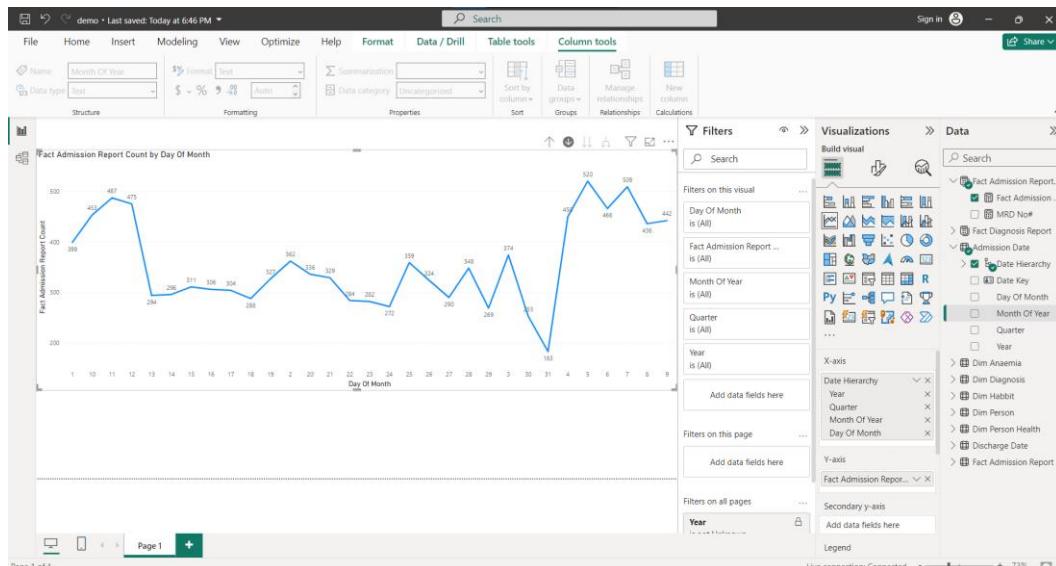
- Theo quý:



- Theo tháng:



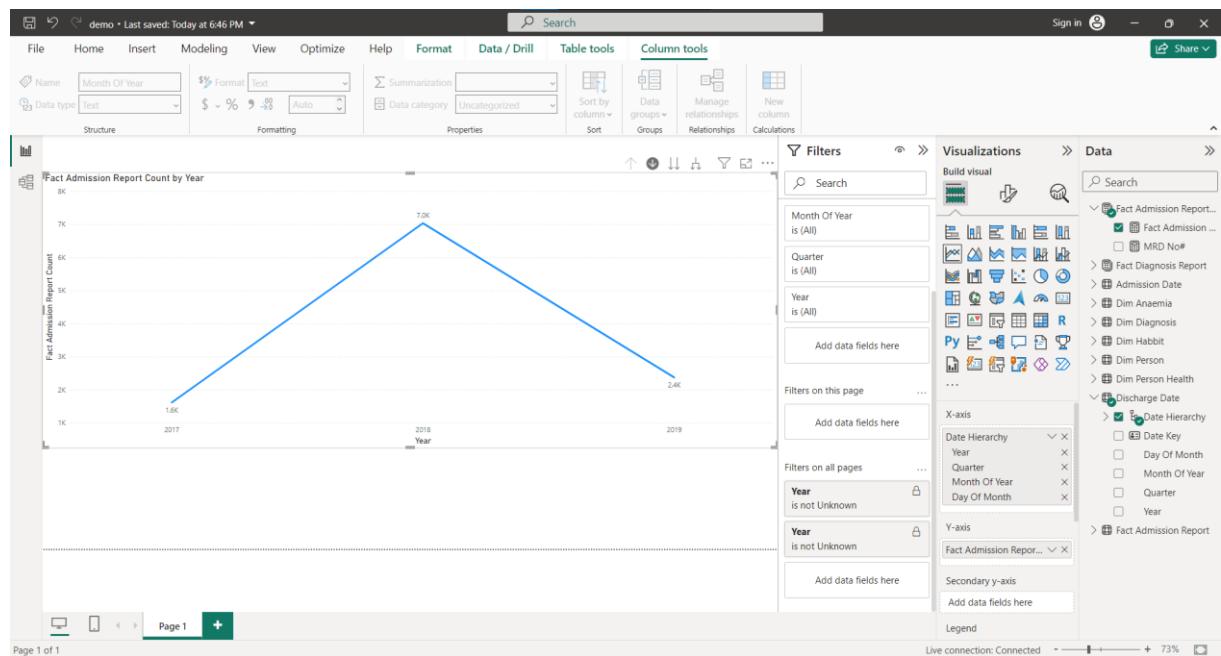
- Theo ngày:



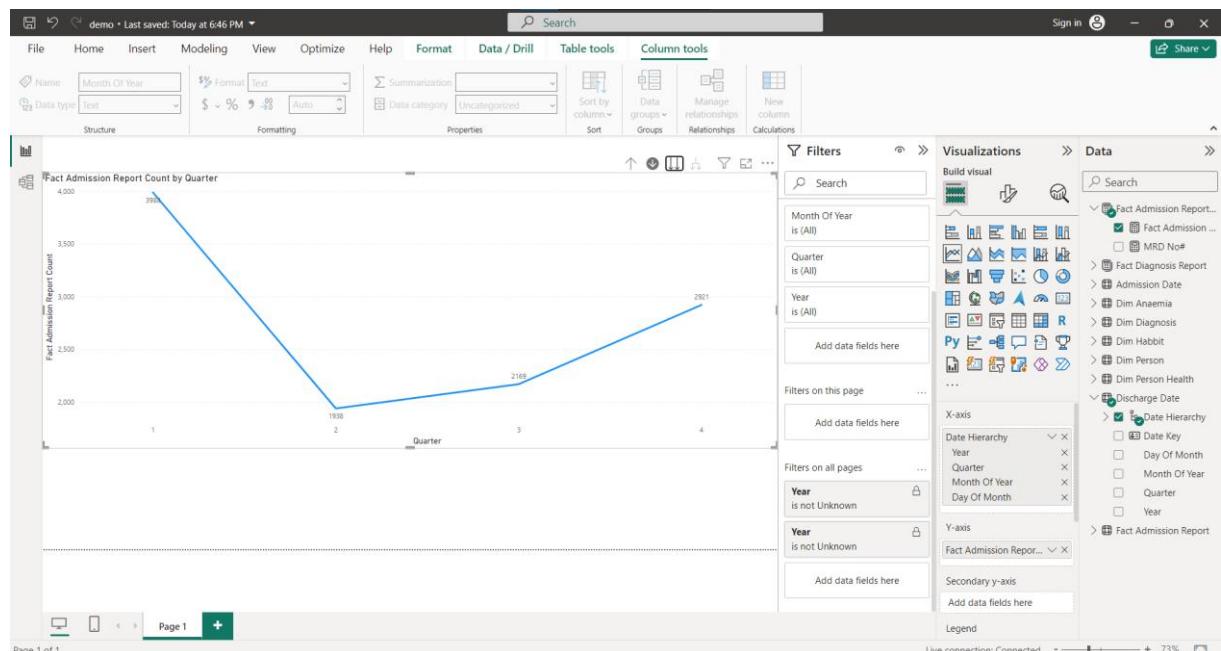
Ta thấy số lượng bệnh nhân nhập viện nhiều nhất là vào năm 2018. Bệnh nhân cũng nhập viện nhiều nhất vào quý I tập trung vào tháng 1 và 2, đặc biệt vào các ngày đầu tháng (5-9).

Câu 2: Số lượng bệnh nhân xuất viện mỗi ngày?

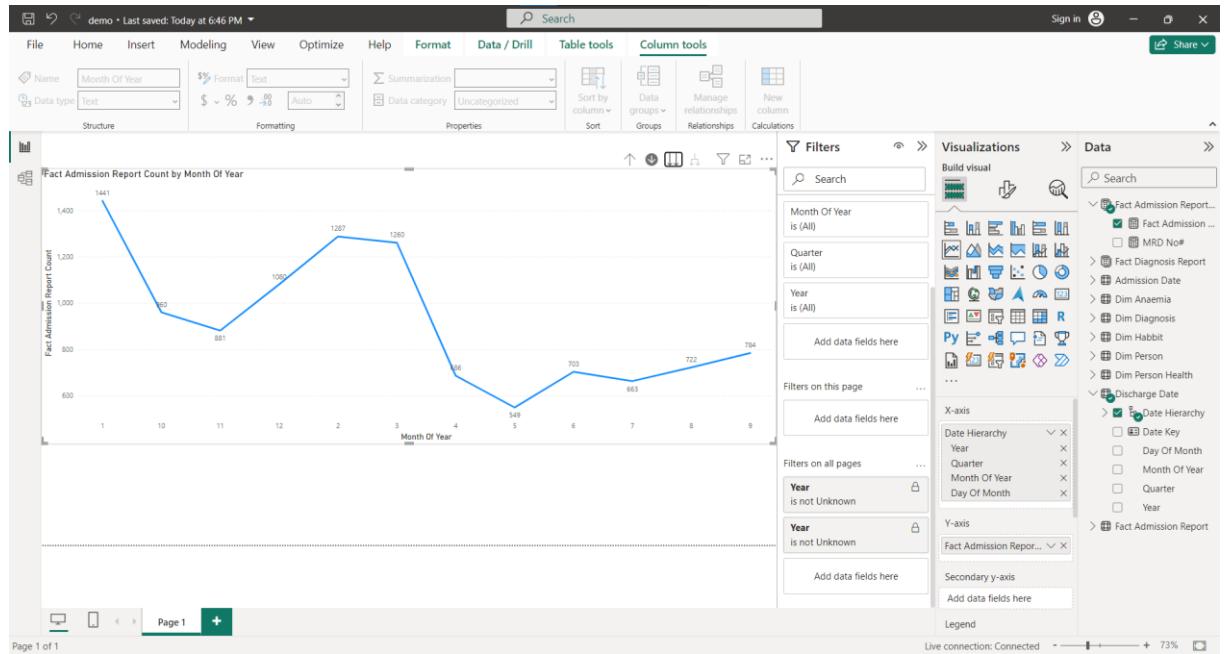
- Theo năm:



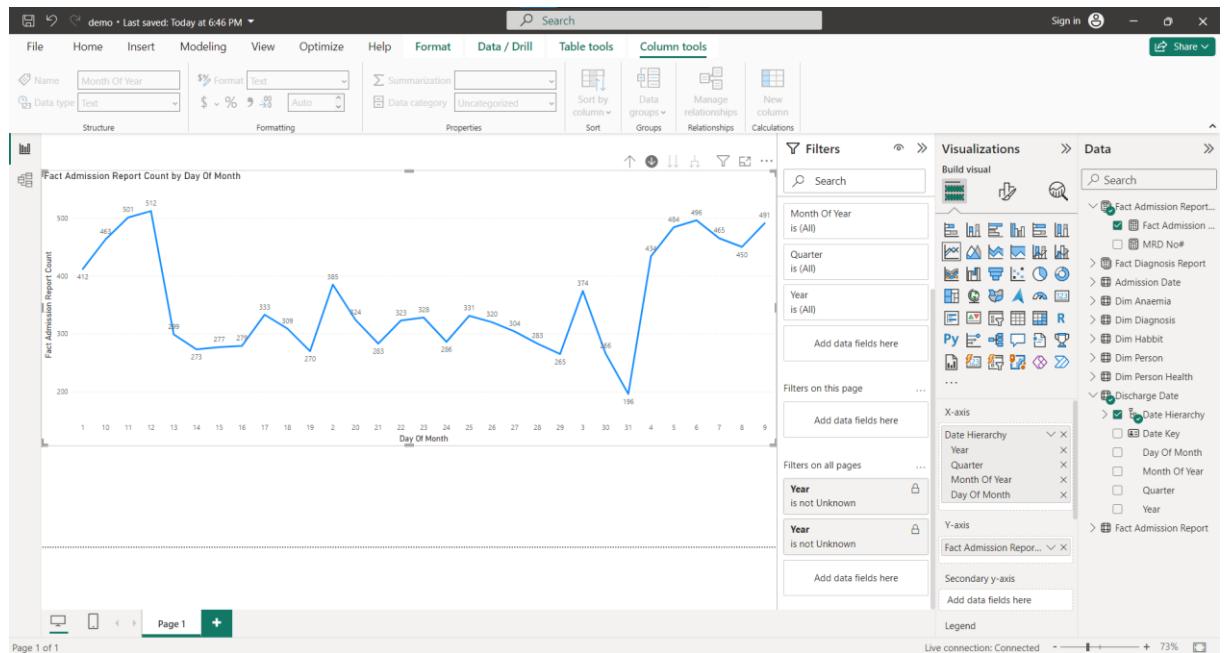
- Theo quý:



- Theo tháng:

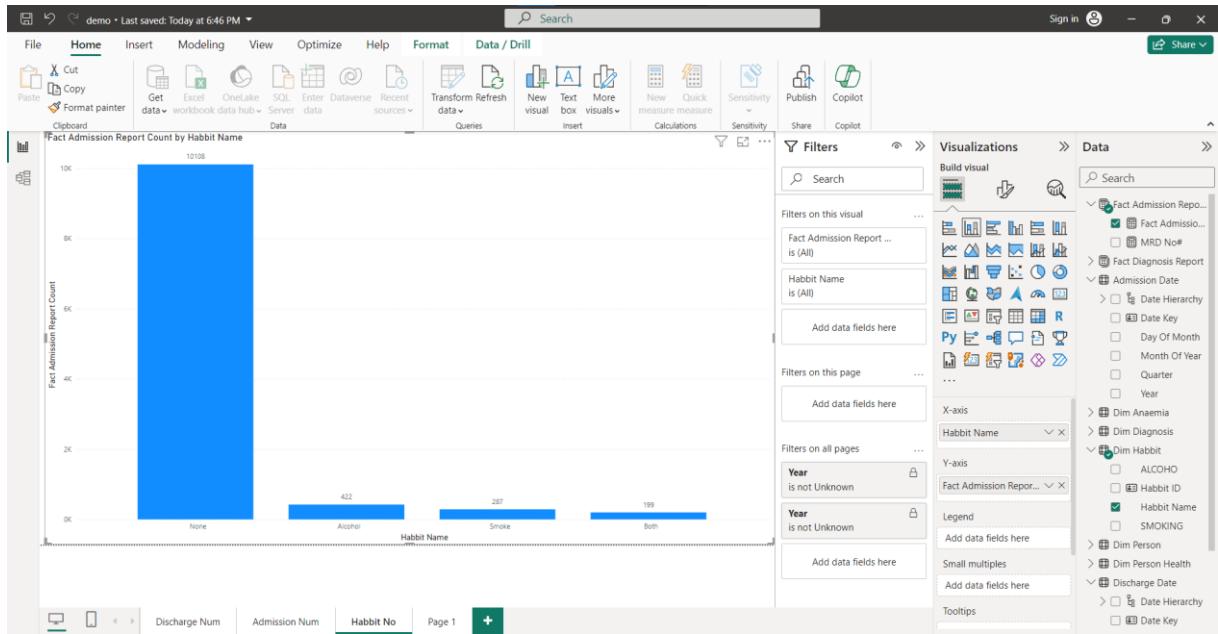


- Theo ngày:



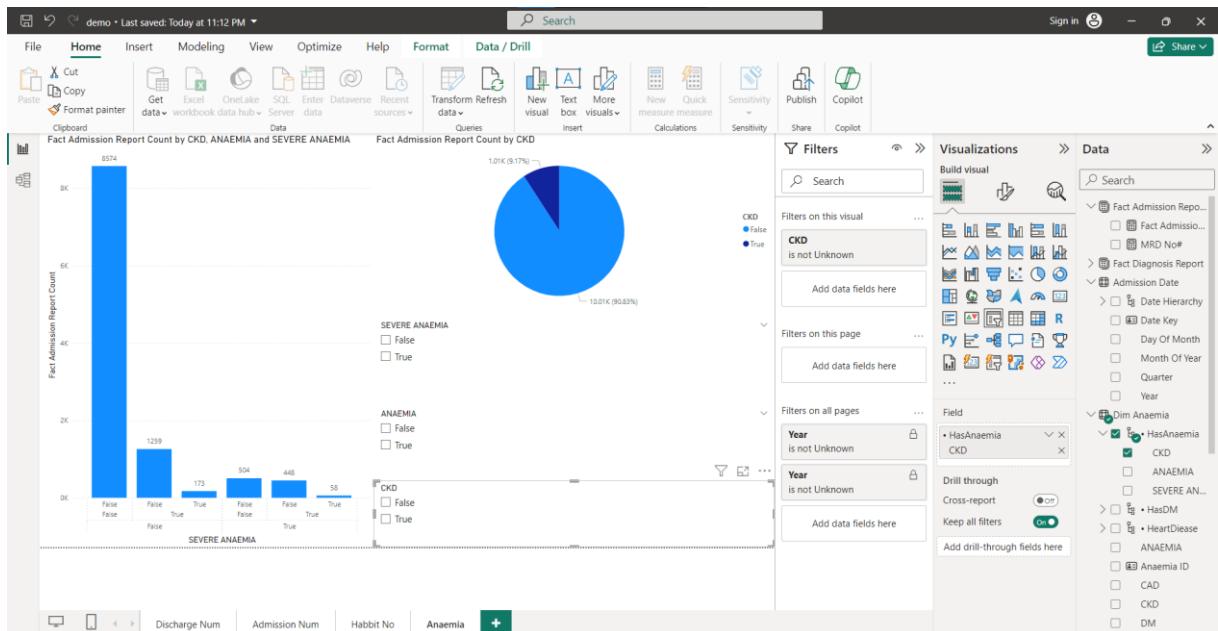
Ta cũng thấy được số lượng bệnh nhân xuất viện cũng có xu hướng như số lượng bệnh nhau nhập viện khi số lượng xuất viện nhiều nhất là vào năm 2018, các tháng đầu năm. Đặc biệt, những ngày có số lượng người xuất viện cao nhất đã mở rộng lên thành gần nửa tháng đầu (2-11).

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



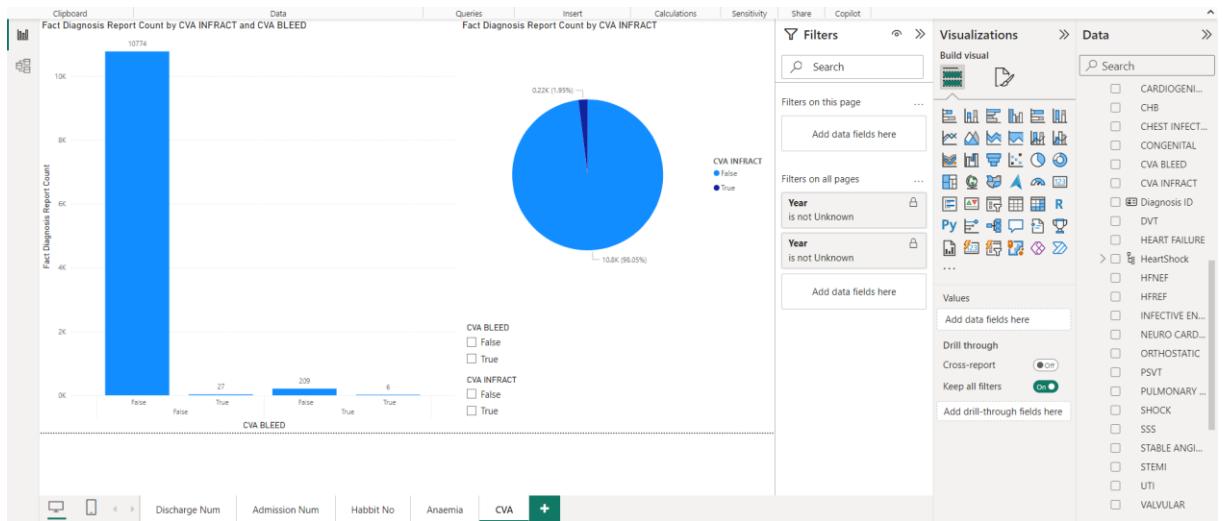
Ta thấy đa số bệnh nhân không hút thuốc và uống rượu (10119 bệnh nhân). Trong đó số lượng phụ nữ có những thói quen như hút thuốc, rượu bia thấp hơn hẳn so với đàn ông.

Câu 4: Có bao nhiêu bệnh nhân bị bệnh thiếu máu ở các tình trạng khác nhau?



Đa phần bệnh nhân không bị thiếu máu (khoảng 90%), và chỉ có khoảng 10% những người bị thiếu máu bị thiếu máu trầm trọng.

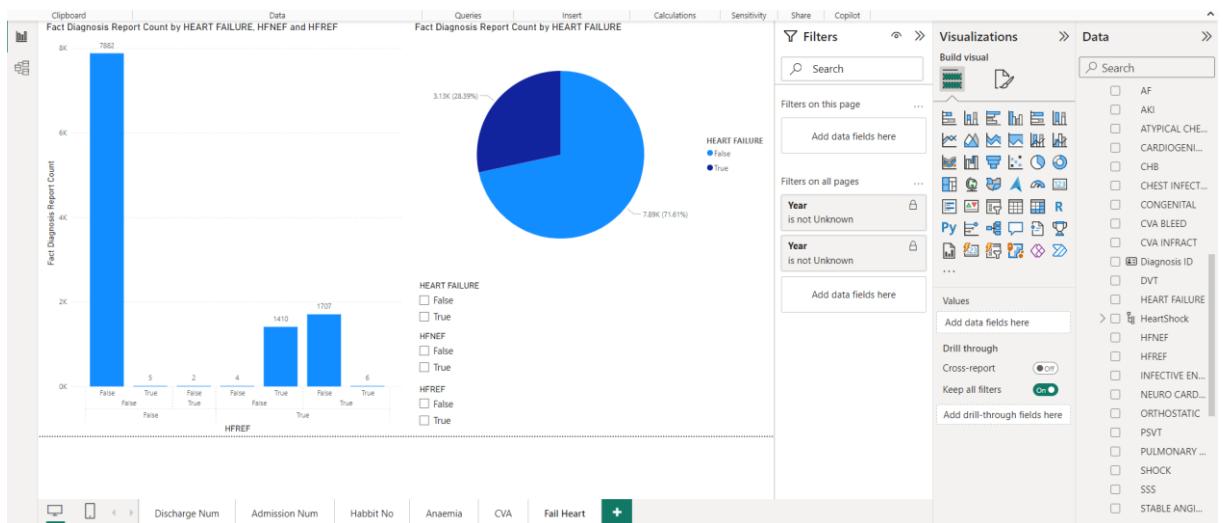
Câu 5: Nguyên nhân do đâu mà bệnh nhân bị đột quy?



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.

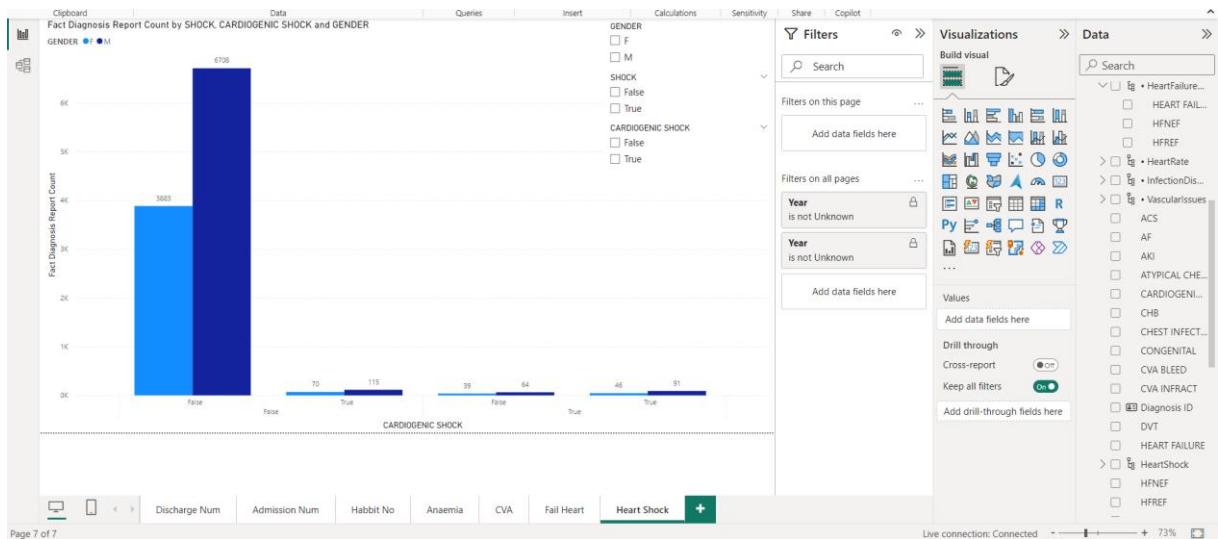
Câu 6: Có bao nhiêu bệnh nhân bị hội chứng suy tim?



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.

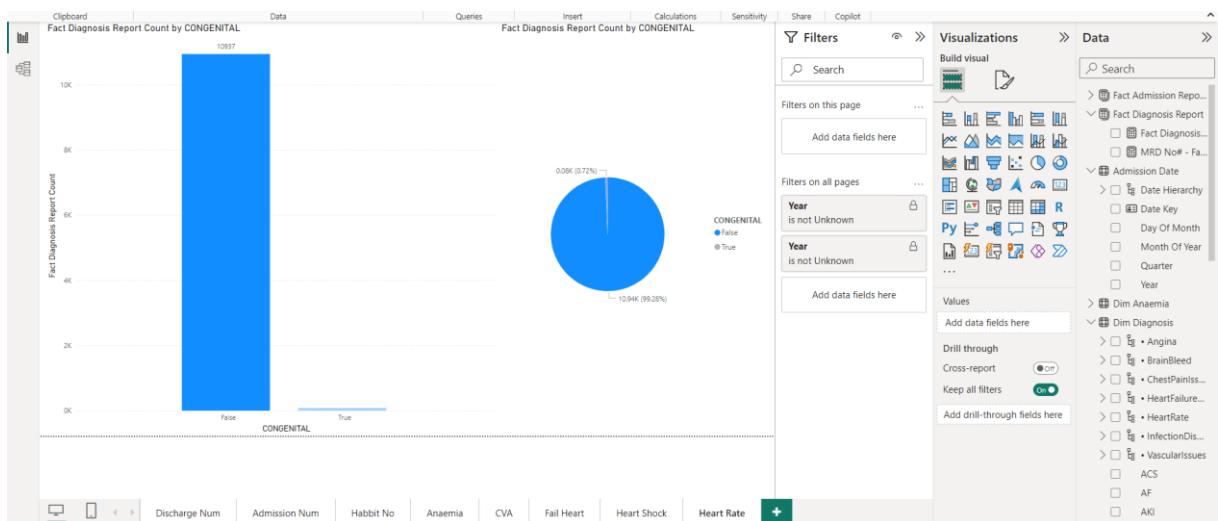
Câu 7: Có bao nhiêu bệnh nhân bị sốc tim theo giới tính?



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.

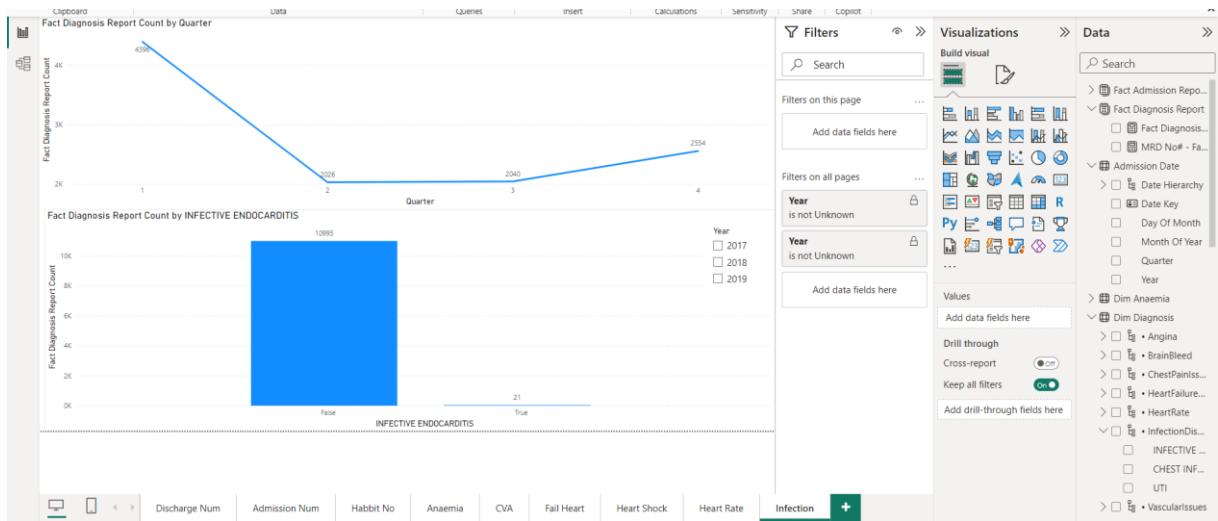
Câu 8: Có bao nhiêu bệnh nhân mắc các vấn đề về nhịp tim bất thường?



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.

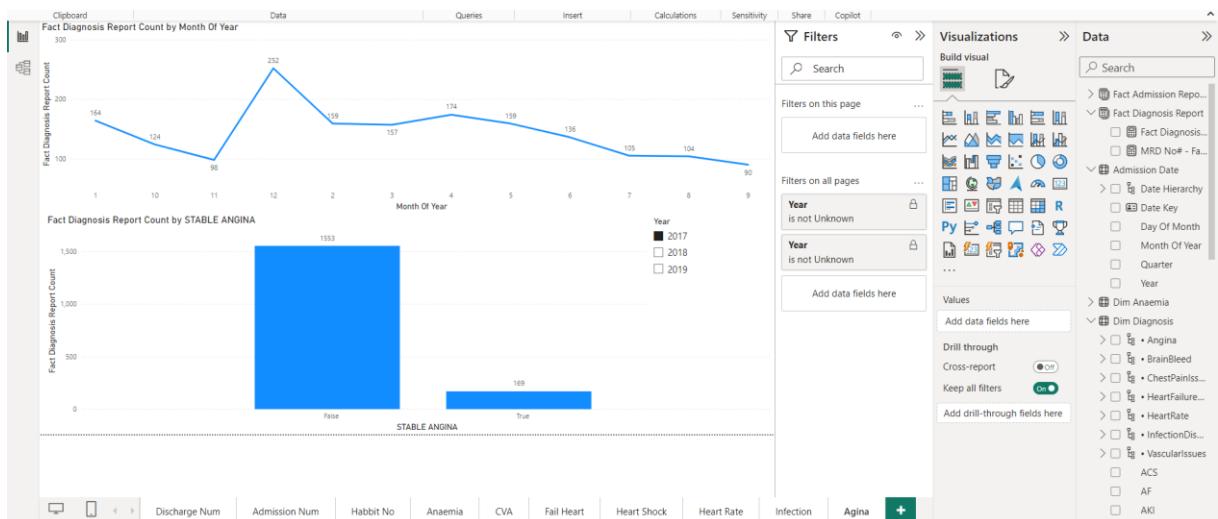
Câu 9: Có bao nhiêu bệnh nhân mắc các bệnh nhiễm trùng theo thời gian?



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



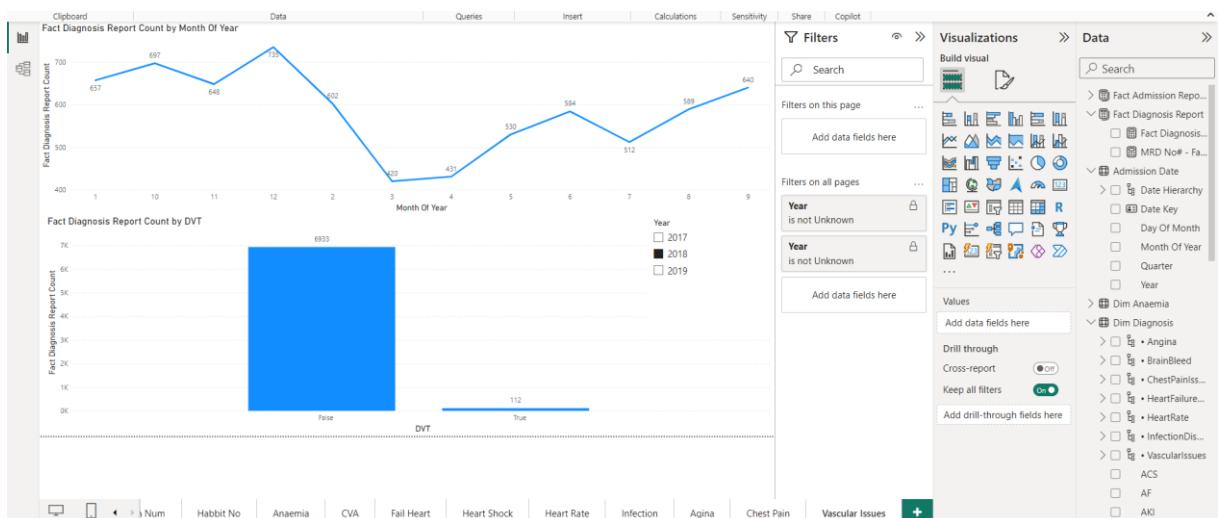
Có 1396 ca bị nhồi máu cơ tim cấp tính (STEMI). 974 ca bị đau thắt ngực liên tục. 12 ca bị cả 2 đau thắt ngực liên tục và nhồi máu cơ tim. Ta thấy đau thắt ngực liên tục (Stable Angina) và nhồi máu cơ tim (STEMI) có vẻ không liên quan nhau lắm.

Câu 11: 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 điểm hình (ACP). 309 ca đau ngực không điểm hình (ACP) nhưng không bị co thắt ngực liên tục.

Câu 12: Có bao nhiêu bệnh nhân bị các vấn đề về mạch máu như thuyên tắc tĩnh mạch sâu hay thuyên tắc phổi?



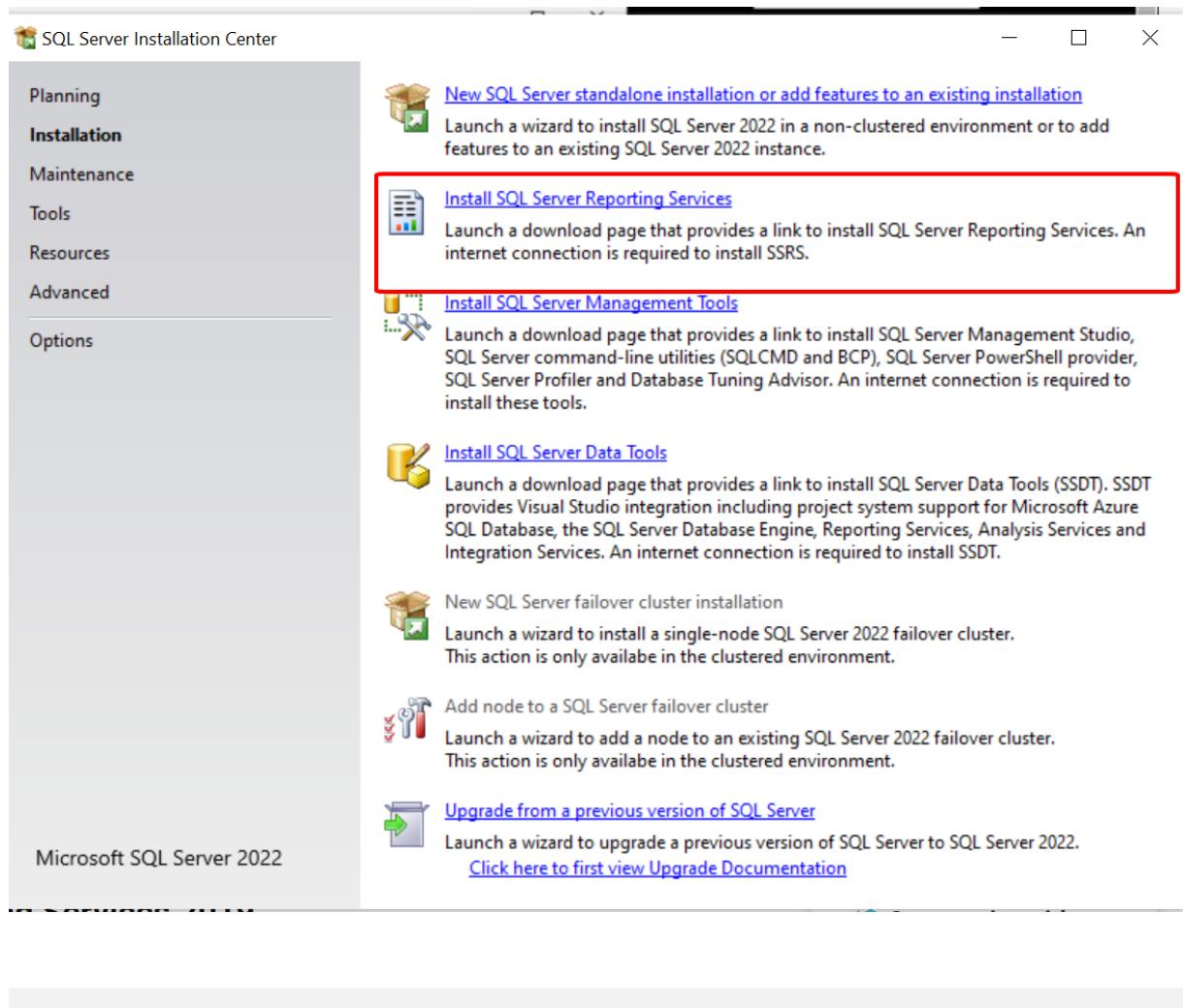
321 ca bị sốc tim (CARDIOGENIC SHOCK). 72 ca thuyên tắc phổi (PULMONARY EMBOLISM). 6 ca vừa thuyên tắc phổi vừa sốc tim. 96 ca thuyên tắc tĩnh mạch sâu (DVT). 57 ca vừa thuyên tắc tĩnh mạch sâu vừa thuyên tắc phổi.

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

CHƯƠNG 5: BÁO CÁO - REPORT

Sử dụng Reporting Service thực hiện tạo ra 3 report, trong đó có tối thiểu một report có grouping và một Report thực hiện bằng tay.

5.1. Cấu hình SSRS (SQL Server Reporting Services)



Microsoft SQL Server 2022 Reporting Services

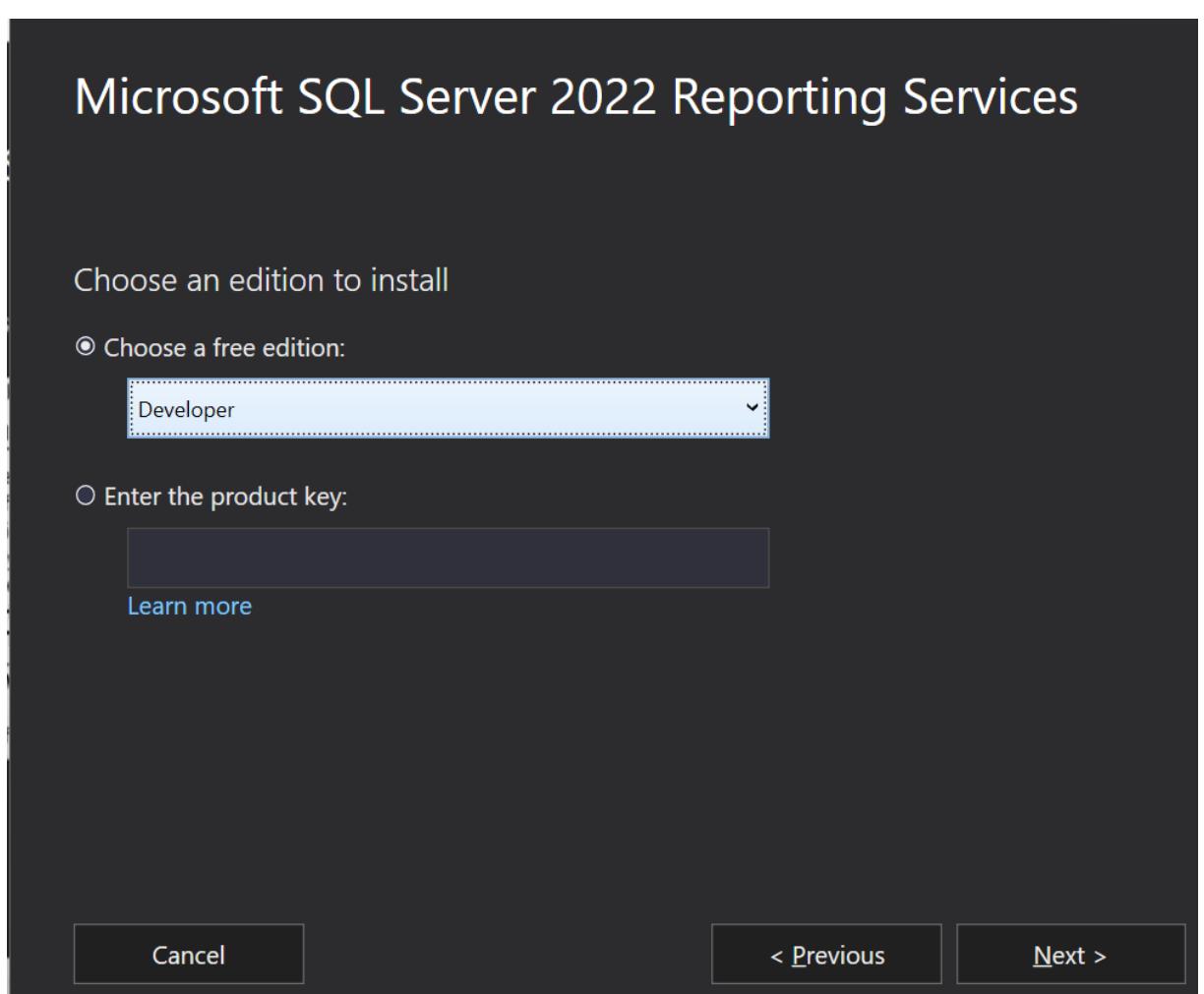
SQL Server Reporting Services is a server-based reporting platform that provides comprehensive reporting functionality.

Important! Selecting a language below will dynamically change the complete page content to that language.

Select language

English

Download



Microsoft SQL Server 2022 Reporting Services

Setup completed

Setup has installed the files you need. You're ready to configure your report server.

Configure manually and customize settings

We'll start Report Server Configuration Manager for you to configure your report server.

[Learn more](#)

[Configure report server](#)

[Close](#)

Report Server Connection Configuration

Report Server Connection

Please specify a server name, click the Find button, and select a report server instance to configure.

Server Name: [Find](#)

Report Server Instance:

[?](#) [Connect](#) [Cancel](#)

Report Server Configuration Manager

Report Server Status

Use the Report Server Configuration Manager tool to define or modify settings for the report server and web portal. Before you can use the report server, you must configure the Web Service URL, the database, and the Web Portal URL.

Current Report Server

Instance ID:	SSRS
Edition:	SQL Server Developer
Product Version:	16.0.1115.96
Report Server Database Name:	
Report Server Mode:	
Report Service Status:	Started

Report Server Configuration Manager: DESKTOP-EHURID8\SSRS

Report Server Configuration Manager

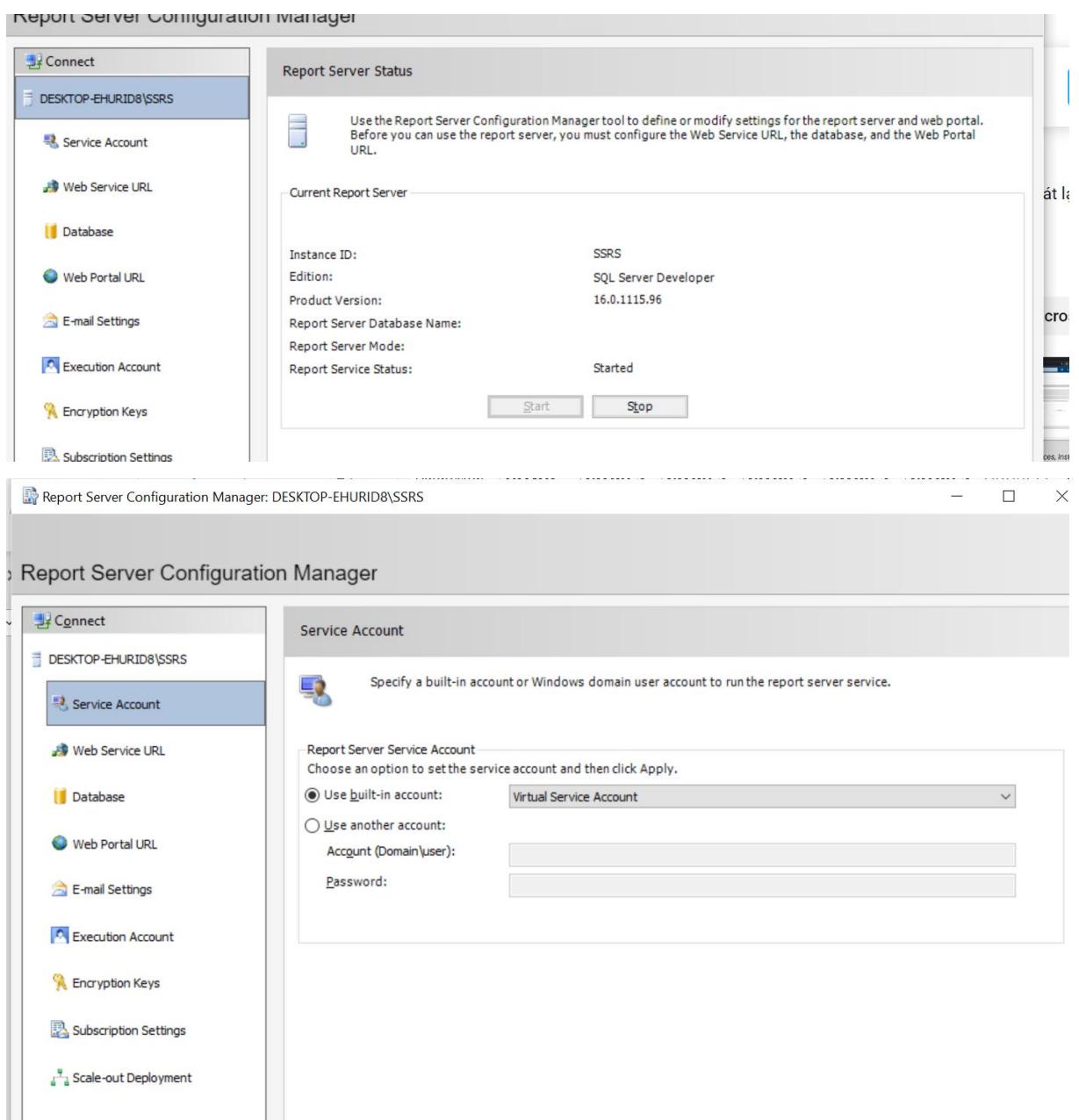
Service Account

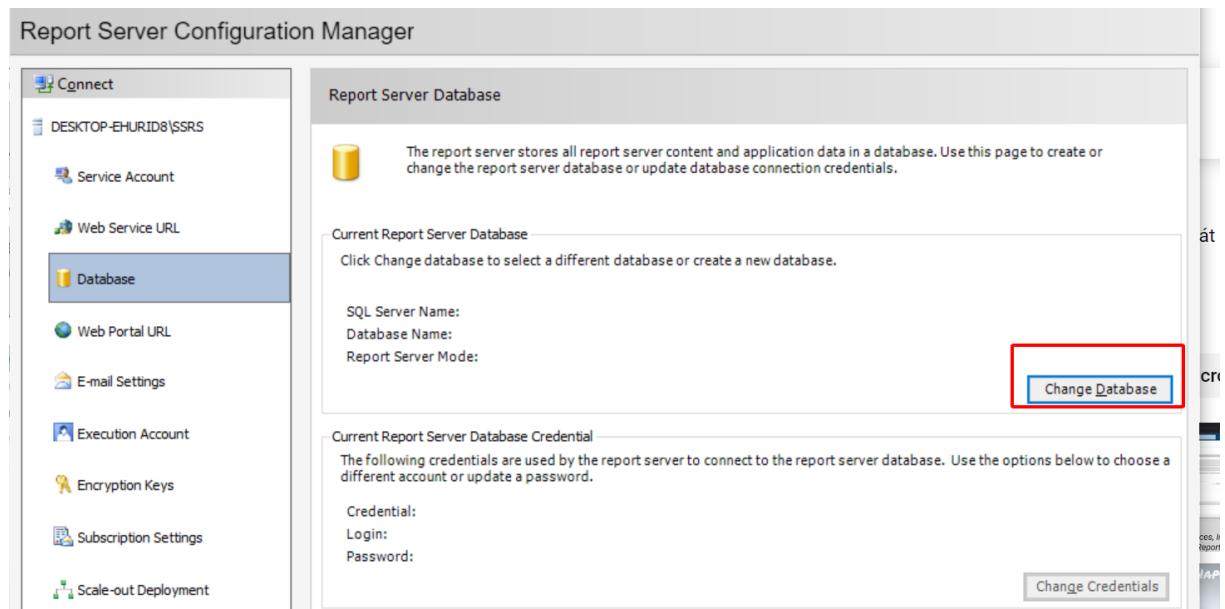
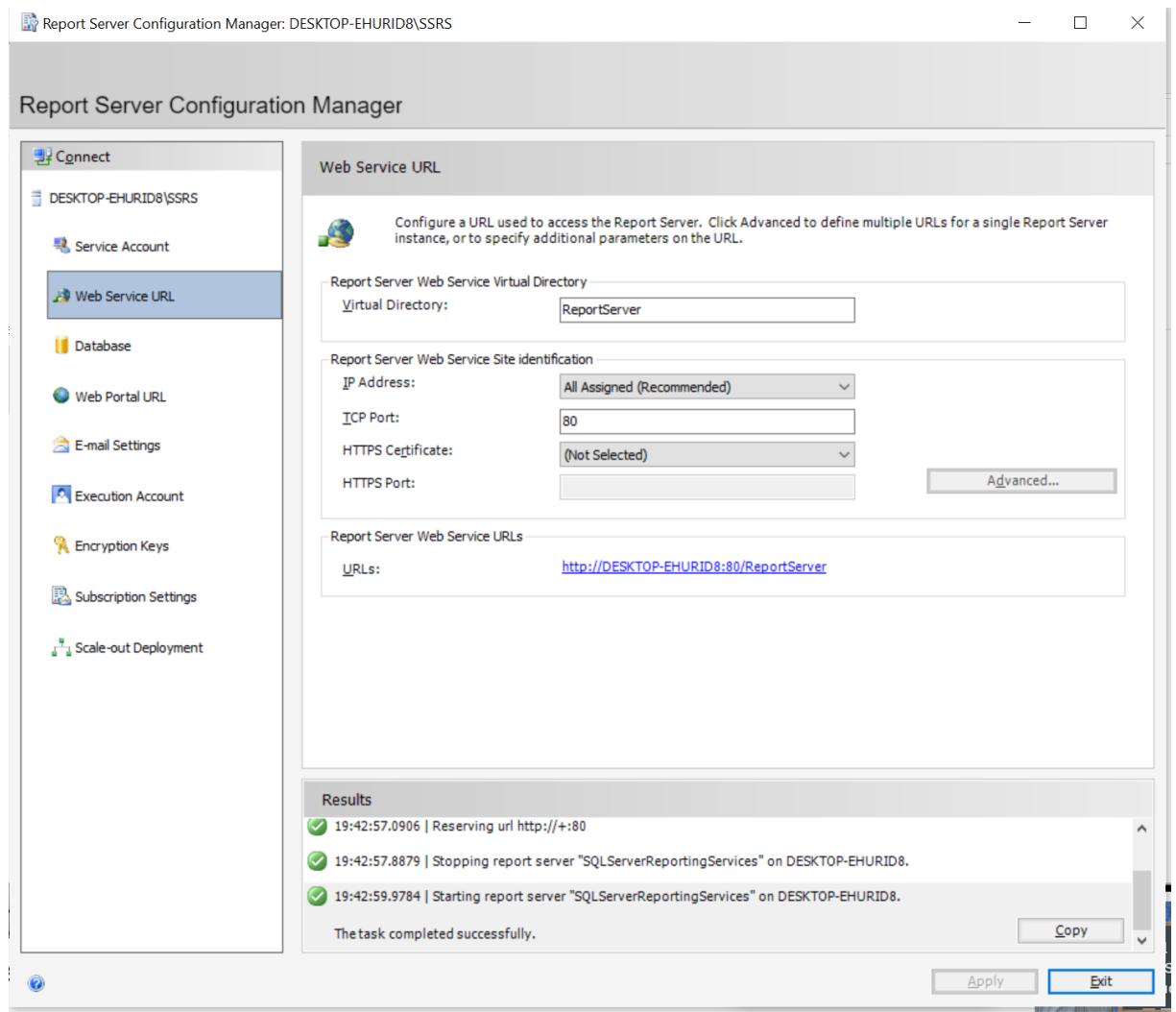
Specify a built-in account or Windows domain user account to run the report server service.

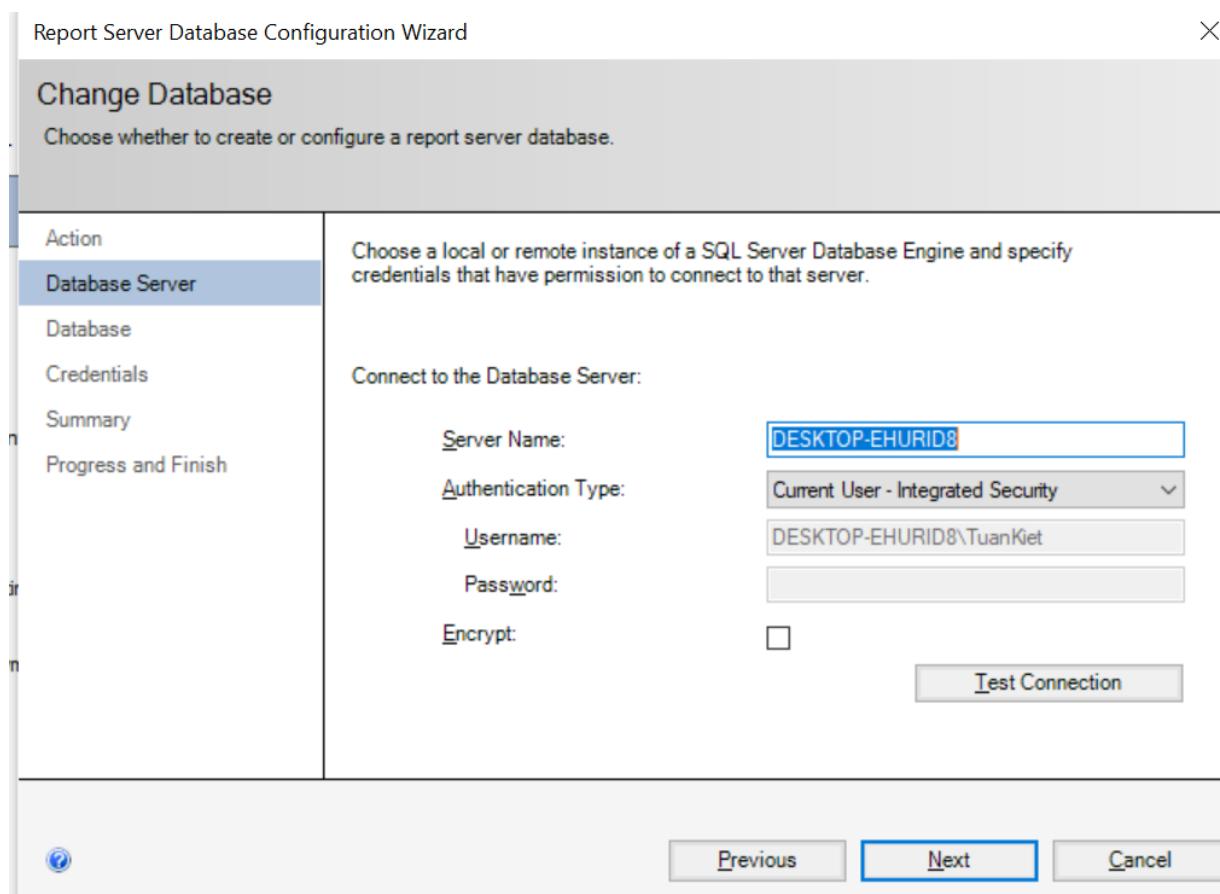
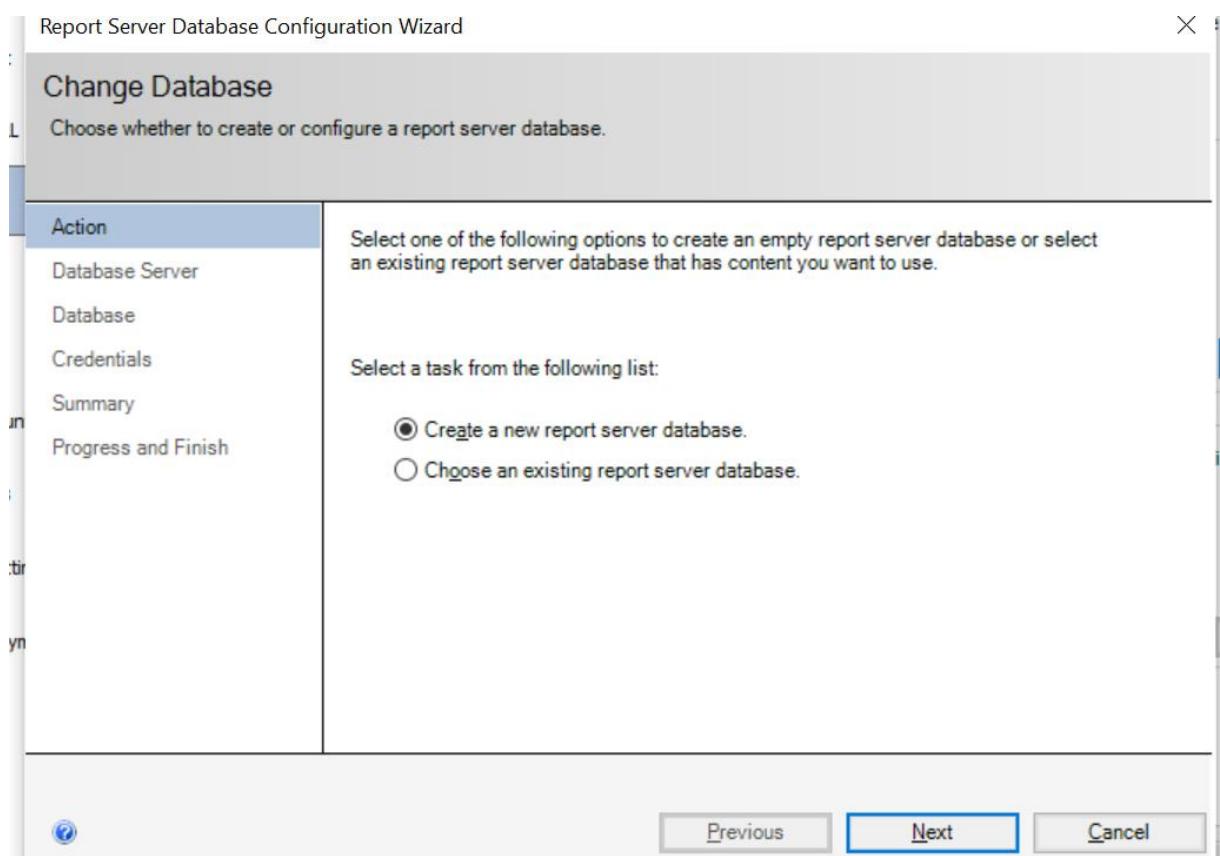
Report Server Service Account
Choose an option to set the service account and then click Apply.

Use built-in account: Virtual Service Account

Use another account:
Account (Domain\user):
Password:







Report Server Database Configuration Wizard

Change Database

Choose whether to create or configure a report server database.

Action	Enter a database name and select the language to use for running SQL scripts.
Database Server	
Database	
Credentials	Database Name: <input type="text" value="ReportServer"/>
Summary	Temp Database Name: <input type="text" value="ReportServerTemp"/>
Progress and Finish	Language: <input type="button" value="English (United States)"/>
	Report Server Mode: Native



Action	Specify the credentials of an existing account that the report server will use to connect to the report server database. Permission to access the report server database will be automatically granted to the account you specify.
Database Server	
Database	
Credentials	
Summary	Credentials:
Progress and Finish	Authentication Type: <input type="button" value="Service Credentials"/>
	User name: <input type="text" value="NT SERVICE\SQLServerReportingServices"/>
	Password: <input type="password"/>



Report Server Database Configuration Wizard

X

Change Database

Choose whether to create or configure a report server database.

Action	The following information will be used to create a new report server database. Verify this information is correct before you continue.
Database Server	
Database	
Credentials	
Summary	<p>SQL Server Instance: DESKTOP-EHURID8</p> <p>Report Server Database: ReportServer</p> <p>Temp Database: ReportServerTempDB</p> <p>Report Server Language: English (United States)</p> <p>Report Server Mode: Native</p> <p>Authentication Type: Service Account</p> <p>Username: NT SERVICE\SQLServerReportingService...</p> <p>Password: *****</p>
Progress and Finish	

Results

Report Server Configuration Manager: DESKTOP-EHURID8\SSRS

Report Server Configuration Manager

Report Server Database

The report server stores all report server content and application data in a database. Use this page to create or change the report server database or update database connection credentials.

Current Report Server Database

Click Change database to select a different database or create a new database.

SQL Server Name:	DESKTOP-EHURID8
Database Name:	ReportServer
Report Server Mode:	Native

Change Database

Current Report Server Database Credential

The following credentials are used by the report server to connect to the report server database. Use the options below to choose a different account or update a password.

Credential:	Service Account
Login:	NT SERVICE\SQLServerReportingServices
Password:	*****

Change Credentials

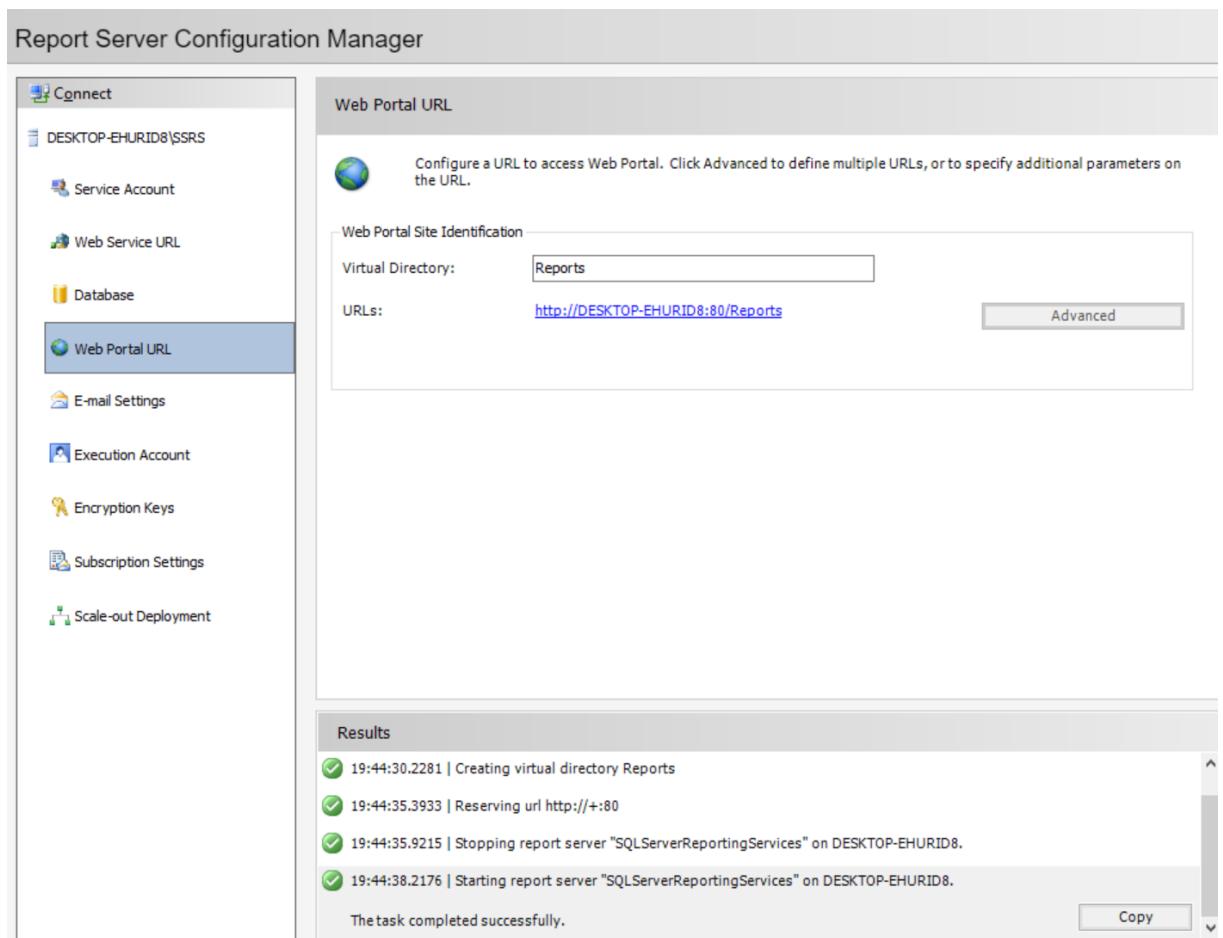
Results

- 19:39:16.2714 | Stopping report server "SQLServerReportingServices" on DESKTOP-EHURID8.
- 19:39:18.3186 | Starting report server "SQLServerReportingServices" on DESKTOP-EHURID8.

The task completed successfully.

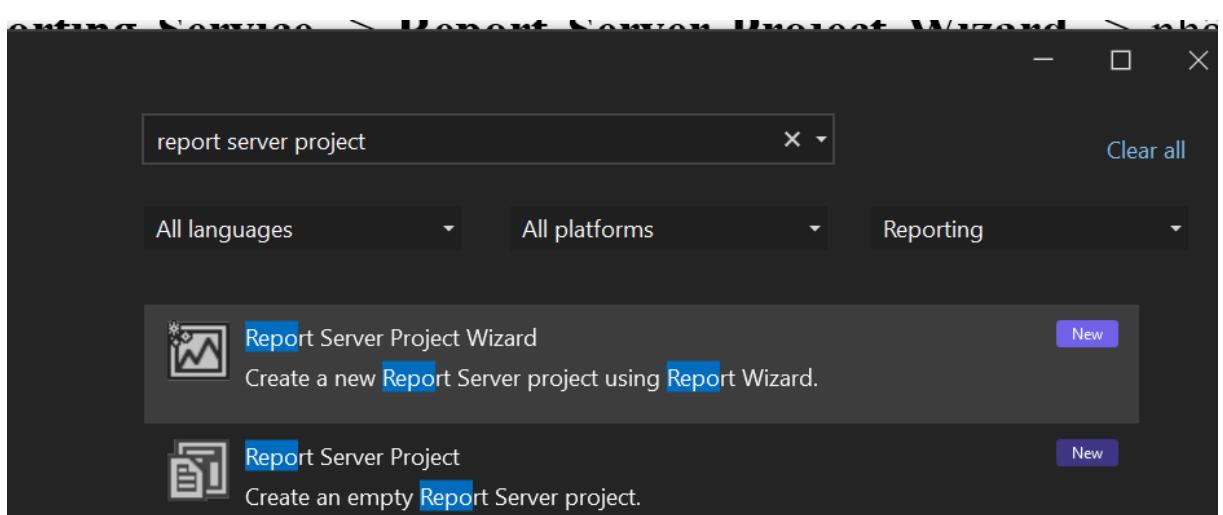
Copy

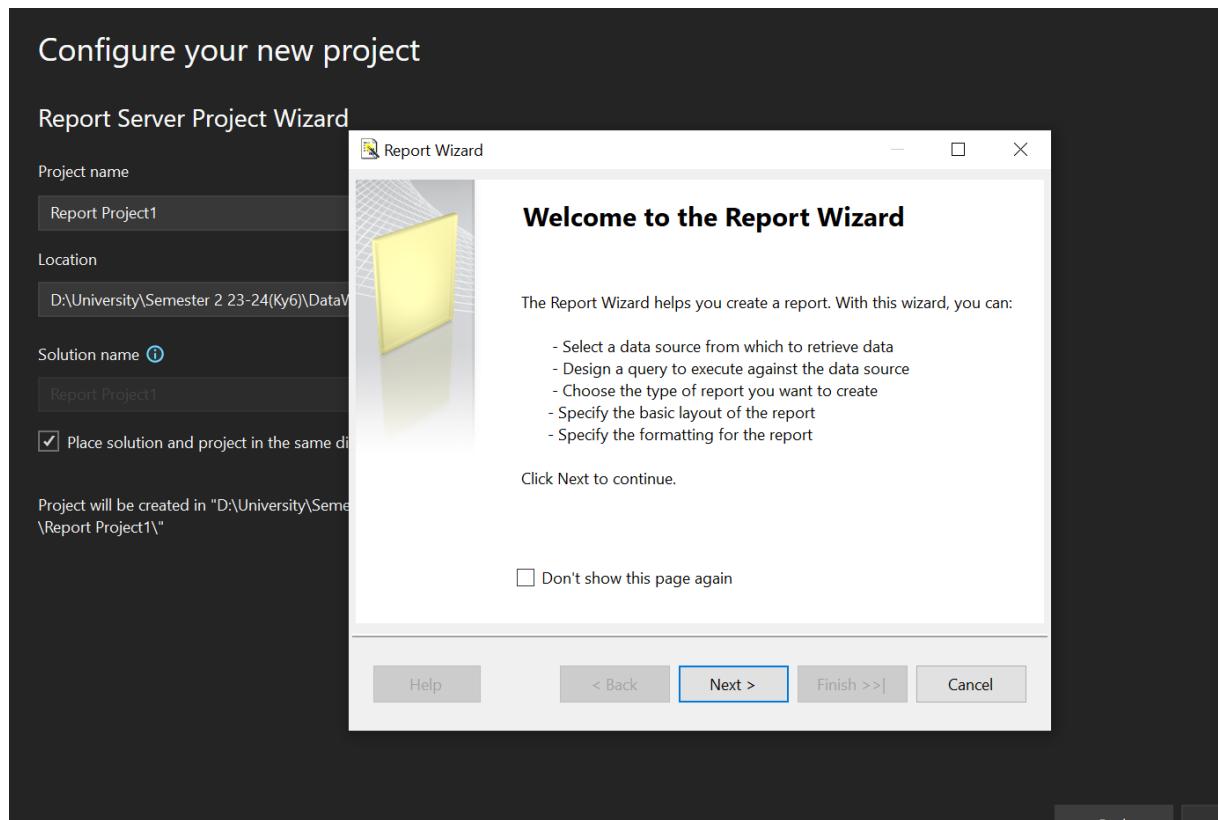
Apply Exit



5.2. Quá trình tạo Project SSRS

1. Vào Visual Studio chọn new Project và chọn **Reporting Service -> Report Server Project Wizard** -> nhập tên project -> OK:



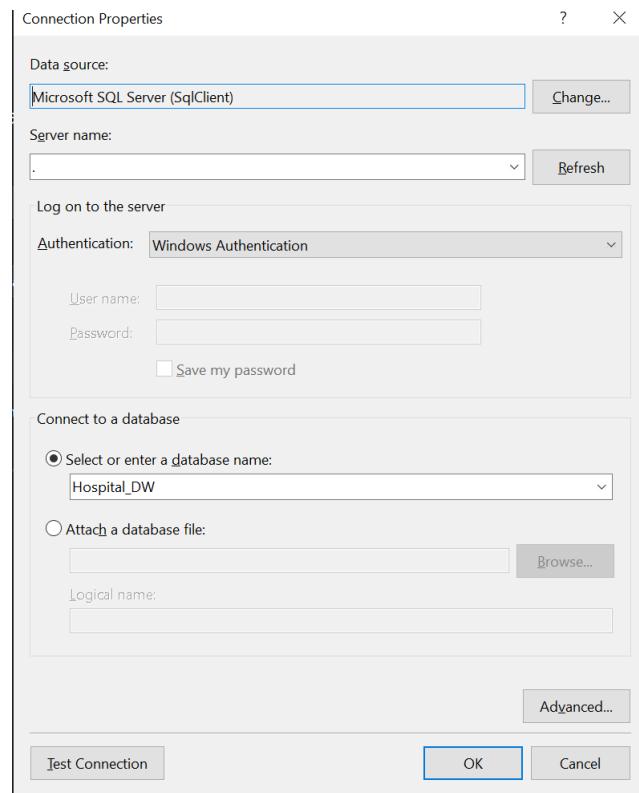


Giao diện sau khi tạo project

2. Tạo Shared Data Source:

The screenshot shows the 'Shared Data Source Properties' dialog box. The title bar includes 'Analyze', 'Tools', 'Extensions', 'Window', 'Help', 'Search', and 'Report Project1'. The main area has tabs 'General' (selected) and 'Credentials'. The 'General' tab contains a note: 'Change name, type, and connection options.' Fields include 'Name:' (set to 'DataSource1'), 'Type:' (set to 'Microsoft SQL Server'), and 'Connection string:' (set to 'Data Source=.;Initial Catalog=Hospital_DW'). A 'Build...' button is located at the bottom right of the connection string input field.

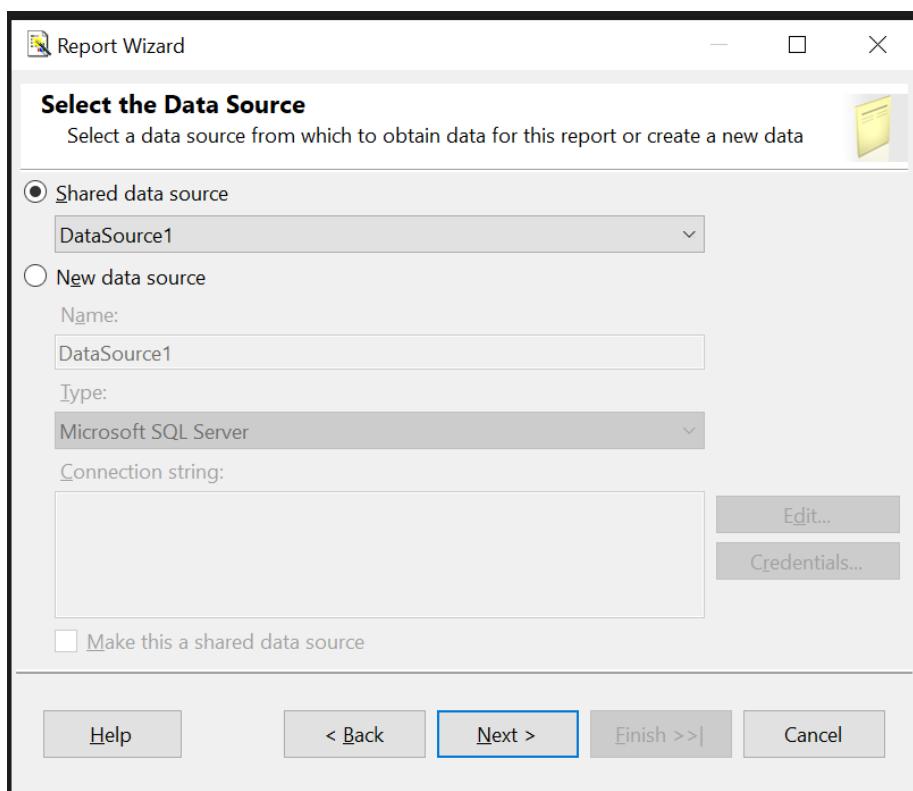
Nhấn Edit: Nhập server name và chọn database Hospital_DW -> nhấn OK



Nhập tên cho Shared Data Source và nhấn OK để hoàn tất.

5.3. Báo cáo về số lượng bệnh nhân và các vấn đề về nhịp tim theo từng năm

1. Nhấn chuột phải **Report** chọn **Add new report** -> nhấn **Next** -> chọn Data Soure đã tạo ở trên:



2. Nhấn **Query Builder** → chọn bảng **Fact** và các bảng **Dimension** → chọn các cột cần sử dụng:

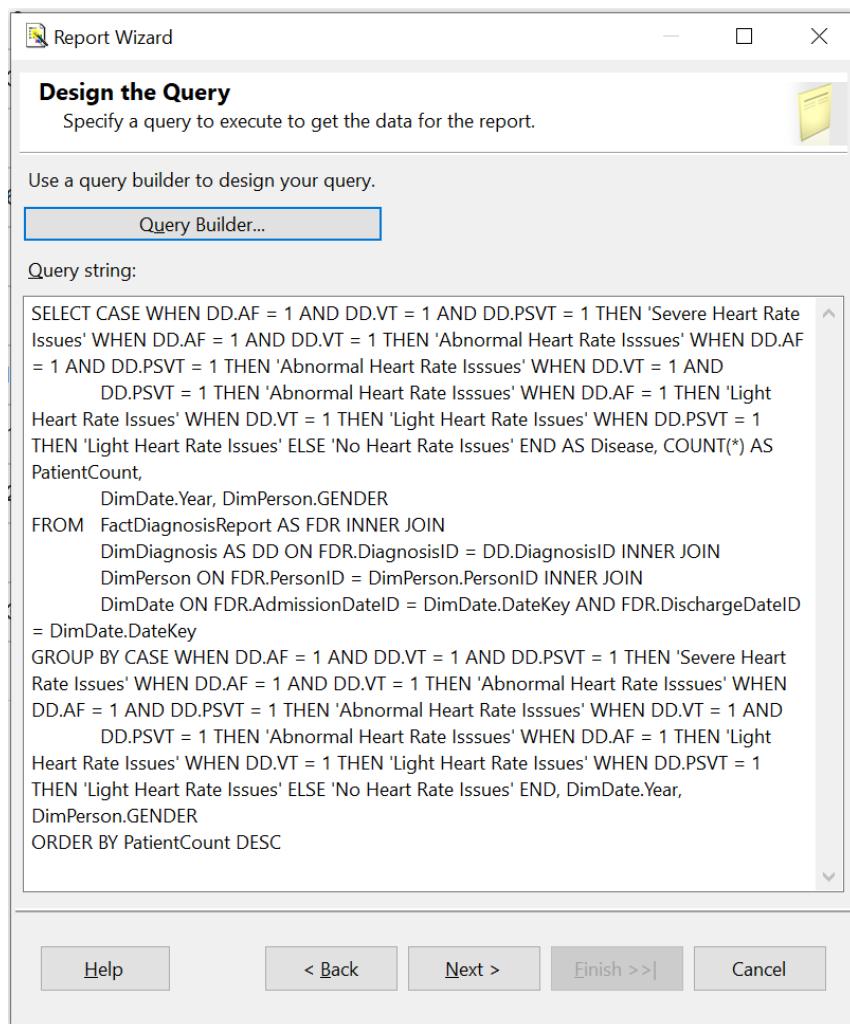
```

SELECT CASE WHEN DD.AF = 1 AND DD.VT = 1 AND DD.PSVT = 1 THEN 'Severe Heart Rate Issues' WHEN DD.AF = 1 AND DD.VT = 1 THEN 'Abnormal Heart Rate Issues' WHEN DD.PSVT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.AF = 1 THEN 'Light Heart Rate Issues' WHEN DD.VT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.PSVT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.AF = 1 AND DD.VT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.PSVT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.AF = 1 THEN 'Normal Heart Rate Issues' WHEN DD.VT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.PSVT = 1 THEN 'Normal Heart Rate Issues'
FROM FactDiagnosisReport AS FDR INNER JOIN
DimDiagnosis AS DD ON FDR.DiagnosisID = DD.DiagnosisID INNER JOIN
DimPerson ON FDR.PersonID = DimPerson.PersonID
GROUP BY CASE WHEN DD.AF = 1 AND DD.VT = 1 AND DD.PSVT = 1 THEN 'Severe Heart Rate Issues' WHEN DD.AF = 1 AND DD.VT = 1 THEN 'Abnormal Heart Rate Issues' WHEN DD.PSVT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.AF = 1 THEN 'Light Heart Rate Issues' WHEN DD.VT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.PSVT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.AF = 1 AND DD.VT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.PSVT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.AF = 1 THEN 'Normal Heart Rate Issues' WHEN DD.VT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.PSVT = 1 THEN 'Normal Heart Rate Issues'
ORDER BY PatientCount DESC
  
```

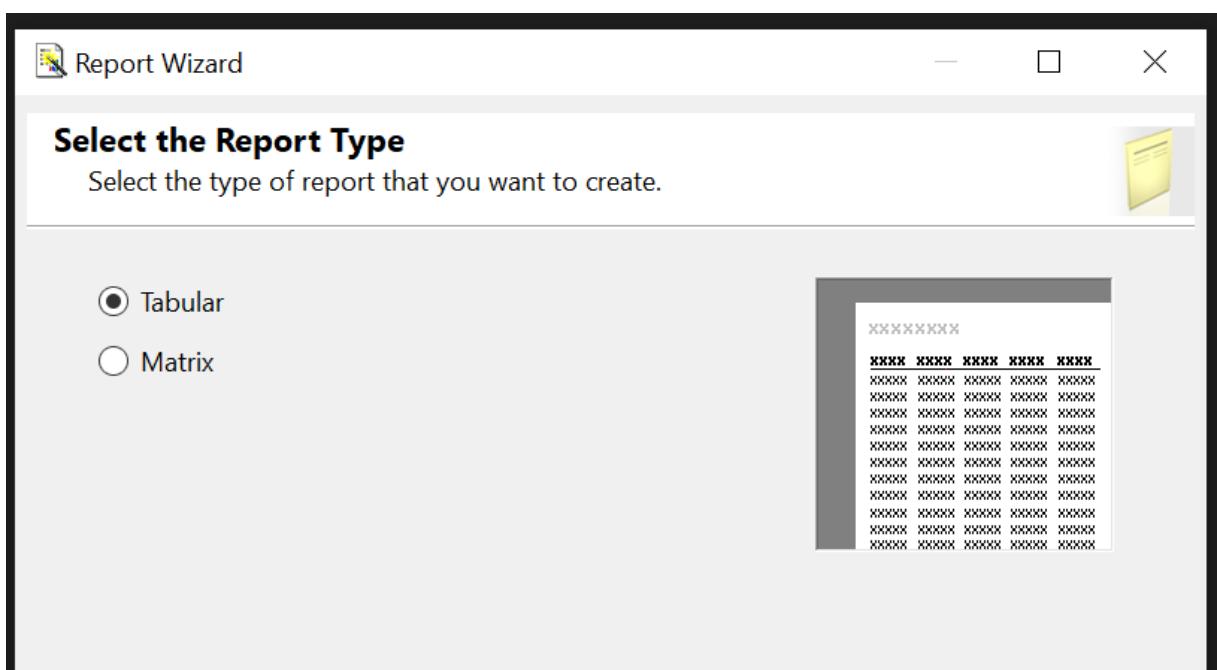
Đếm các ca bệnh về nhịp tim. Nếu chỉ mắc 1 trong 3 ta đưa vào ‘Light Heart Rate Issues’. Mắc 2 trong 3 đưa vào ‘Abnormal Heart Rate Issues’. Mắc cả 3 đưa vào ‘Severe Heart Rate Issues’.

```

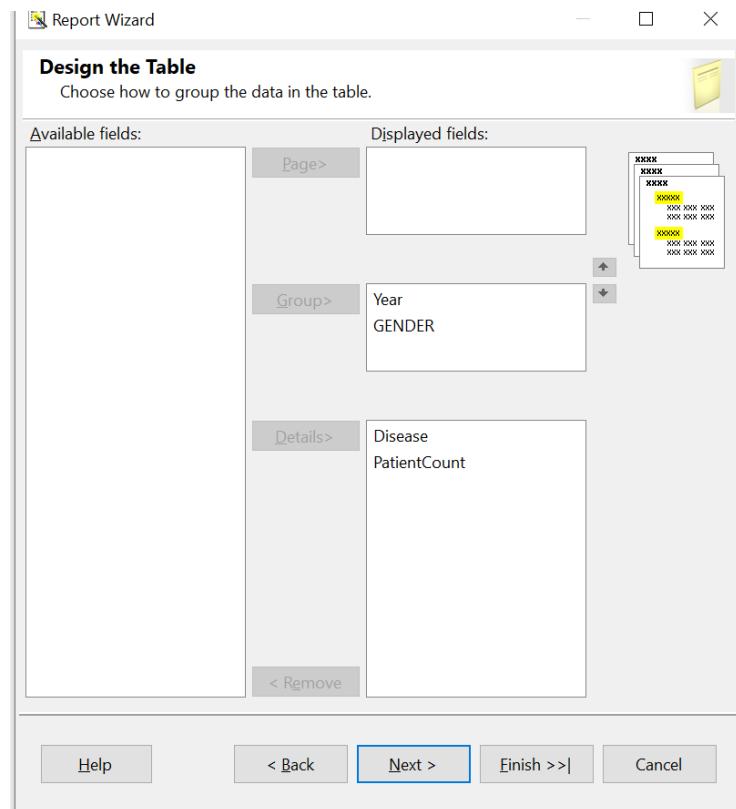
SELECT CASE WHEN DD.AF = 1 AND DD.VT = 1 AND DD.PSVT = 1 THEN 'Severe Heart Rate Issues' WHEN DD.AF = 1 AND DD.VT = 1 THEN 'Abnormal Heart Rate Issues' WHEN DD.PSVT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.AF = 1 THEN 'Light Heart Rate Issues' WHEN DD.VT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.PSVT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.AF = 1 AND DD.VT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.PSVT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.AF = 1 THEN 'Normal Heart Rate Issues' WHEN DD.VT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.PSVT = 1 THEN 'Normal Heart Rate Issues'
FROM FactDiagnosisReport AS FDR INNER JOIN
DimDiagnosis AS DD ON FDR.DiagnosisID = DD.DiagnosisID INNER JOIN
DimPerson ON FDR.PersonID = DimPerson.PersonID
GROUP BY CASE WHEN DD.AF = 1 AND DD.VT = 1 AND DD.PSVT = 1 THEN 'Severe Heart Rate Issues' WHEN DD.AF = 1 AND DD.VT = 1 THEN 'Abnormal Heart Rate Issues' WHEN DD.PSVT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.AF = 1 THEN 'Light Heart Rate Issues' WHEN DD.VT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.PSVT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.AF = 1 AND DD.VT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.PSVT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.AF = 1 THEN 'Normal Heart Rate Issues' WHEN DD.VT = 1 THEN 'Normal Heart Rate Issues' WHEN DD.PSVT = 1 THEN 'Normal Heart Rate Issues'
ORDER BY PatientCount DESC
  
```



3. Nhấn Ok → Nhấn Next để tiếp tục → chọn Tabular.



4. Nhấn Next → chọn các cột vào Detail và nhấn Next:



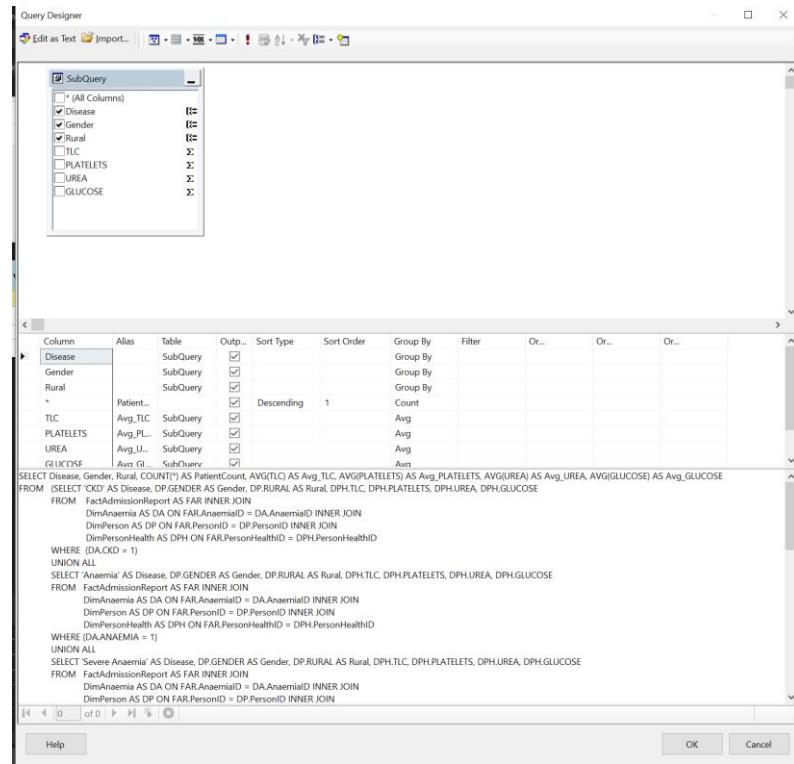
Nhập tên và nhấn finish để hoàn tất.

Ván đề về nhịp tim theo từng năm			
Year	GENDER	Disease	Patient Count
2017	F	No Heart Rate Issues	21
		Light Heart Rate Issues	3
2017	M	No Heart Rate Issues	39
		Light Heart Rate Issues	8
2018	F	No Heart Rate Issues	60
		Light Heart Rate Issues	7
		Abnormal Heart Rate	1
2018	M	No Heart Rate Issues	144
		Light Heart Rate Issues	10
		Abnormal Heart Rate	1
2019	F	No Heart Rate Issues	25
		Light Heart Rate Issues	3
		No Heart Rate Issues	37
		Light Heart Rate Issues	1

Nhấn vào tag Preview để xem trước báo cáo:

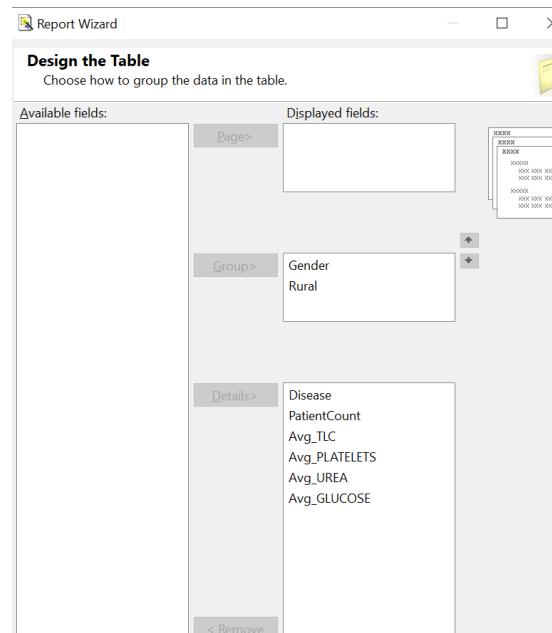
5.4. Báo cáo về các chỉ số trung bình của các bệnh nhân

Thực hiện tương tự Báo cáo 1 cho đến Query Builder. Chọn các bảng sau:

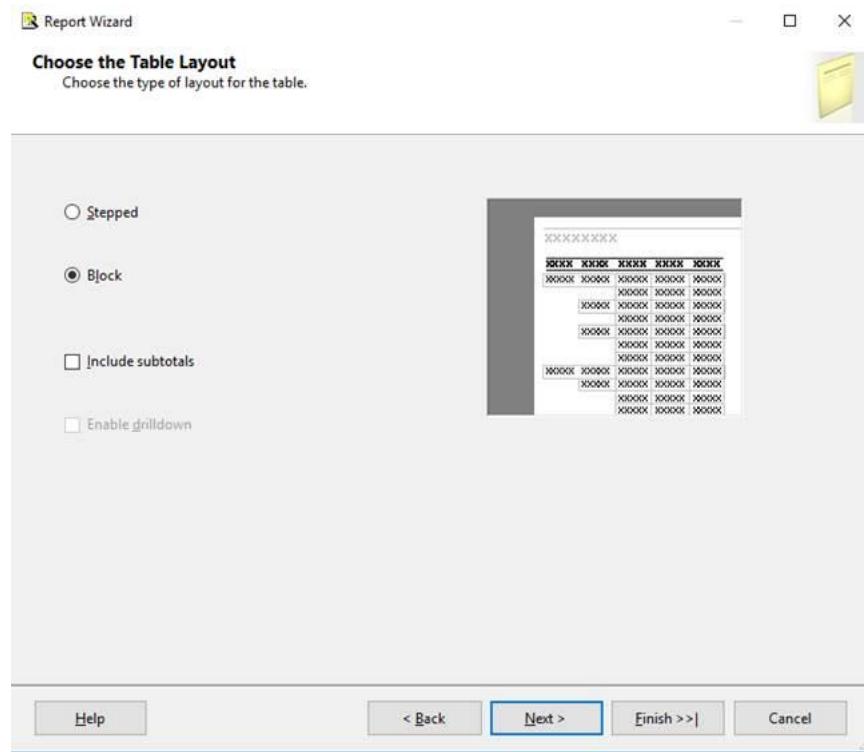


Thực hiện đếm số ca bệnh của các bệnh phổi biến như Suy thận, tiêu đường, thiếu máu, thiếu máu trầm trọng, rung tâm nhĩ (AF), nhanh tâm thất (VT), viêm phổi (Chest Infection), rối loạn nhịp tim (PSVT)... và tính các trung bình các chỉ số TLC (tiêu cầu), Platelets (bạch cầu), UREA, Glucose.

Design The Table như sau:



Nhấn Next → chọn Block:



Nhấn Next → nhập tên báo cáo → nhấn Finish để hoàn tất:

Báo cáo về các chỉ số trung bình của những nhóm bệnh							
Gender	Rural	Disease	Patient Count	Avg TLC	Avg PLATELETS	Avg UREA	Avg GLUCOSE
F	R	Diabetes Mellitus	237	12.092340	269.857203	62.279661	208.815789
		Anaemia	187	11.439784	270.953475	72.526881	176.108571
		CKD	67	13.246969	253.511940	136.417910	188.435483
		Atrial Fibrillation	43	10.734883	240.976744	59.395348	195.232558
		Ventricular Tachycardia	22	12.618181	231.636363	52.136363	158.000000
		Severe Anaemia	21	9.437619	249.619047	57.952380	162.523809
		Chest Infection	18	13.994444	299.444444	52.777777	141.941176
		Paroxysmal Supra Ventricular Tachycardia	2	7.100000	189.000000	19.000000	128.000000
		Infective Endocarditis	1	10.900000	373.000000	27.000000	182.000000
U	U	Diabetes Mellitus	957	12.254344	256.830293	57.629979	217.776923
		Anaemia	806	11.913965	272.325031	67.131467	178.513054
		CKD	251	14.428800	233.538152	130.844621	177.688311
		Atrial Fibrillation	133	11.865116	243.456692	57.065151	177.064516
		Ventricular Tachycardia	91	15.815116	247.449411	62.770114	163.025316
		Severe Anaemia	90	11.735555	269.166666	78.444444	177.905882
		Chest Infection	71	12.225352	271.732394	49.746478	176.515151
		Paroxysmal Supra Ventricular Tachycardia	44	9.554545	258.681818	44.181818	149.772727
		Infective Endocarditis	2	8.200000	336.000000	26.000000	206.000000
M	R	Diabetes Mellitus	388	11.186093	227.438903	59.449218	201.686863
		Anaemia	191	12.088631	267.765789	82.531413	170.337016
		CKD	144	13.007560	221.474624	142.246527	170.124444

Kết quả Preview

- Bệnh tiểu đường và thiếu máu là hai nhóm bệnh phổ biến nhất, đặc biệt là ở nữ giới và ở khu vực urban. Điều này có thể phản ánh một số yếu tố lối sống và môi trường ở khu vực đô thị.
- Chỉ số TLC (bạch cầu) và Platelets (tiểu cầu): Các chỉ số này thường nằm trong khoảng giới hạn bình thường nhưng có sự tăng ở những bệnh nhân mắc bệnh nhiễm trùng và bệnh tim.
- Chỉ số Urea: Bệnh nhân suy thận (CKD) có chỉ số Urea cao nhất, phản ánh tình trạng bệnh lý của họ.

Chỉ số Glucose: Bệnh nhân mắc bệnh tiểu đường có chỉ số đường trong máu cao nhất, điều này phản ánh đúng tình trạng bệnh của họ, một số trường hợp glucose cao ở bệnh nhân mắc bệnh Atrial Fibrillation (rung tâm nhĩ).

CHƯƠNG 6: KẾT LUẬN

Báo cáo kết quả được sau khi hoàn thành đồ án, nêu ra những hạn chế còn tồn tại, định hướng phát triển trong tương lai, các tài liệu tham khảo. Bảng phân công công việc cụ thể và mức độ hoàn thành công việc.

6.1. Kết quả đạt được

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

6.2. Những hạn chế

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

6.3. Hướng phát triển

Từ những hạn chế được nêu trên, chúng em có thể đưa ra các đường hướng, mục tiêu cho tương lai như: Xây dựng tập dữ liệu phục vụ cho việc khai thác dữ liệu (data mining) – có thể tiến hành chạy các mô hình học máy trên đó.

6.4. Tài liệu tham khảo

- [1] Tài liệu các file PDF hướng dẫn về Data Integration với SQL Server, SSIS, SSAS của GVHD Ths. Nguyễn Văn Thành.
- [2] Minewiskan(no date), *Lesson 3: Modifying Measures, Attributes and Hierarchies*. [online] learn.microsoft.com. Available at: <https://learn.microsoft.com/en-us/analysis-services/multidimensional-tutorial/lesson-3-modifying-measures-attributes-and-hierarchies?view=asallproducts-allversions>. [Accessed 12 May 2023].
- [3] Minewiskan(no date), *Lesson 4: Defining Advanced Attribute and Dimension Properties*. [online] learn.microsoft.com. Available at: <https://learn.microsoft.com/en-us/analysis-services/multidimensional-tutorial/lesson-4-defining-advanced-attribute-and-dimension-properties?view=asallproducts-allversions>.

- us/analysis-services/multidimensional-tutorial/lesson-4-defining-advanced-attribute-and-dimension-properties?view=asallproducts-allversions. [Accessed 12 May 2023].
- [4] Nguyễn Văn Chúc (no date). *ETL Project From Excel Data Source to Star Schema with SSIS*. [online] youtube.com. Available at: https://www.youtube.com/watch?v=Yp8fXLnVCp8&ab_channel=ChucNguyenVan. [Accessed 10 May 2023].
- [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. An Optimized Machine Learning Model Accurately Predicts In-Hospital Outcomes at Admission to a Cardiac Unit. *Diagnostics* 2022, 12, 241. <https://doi.org/10.3390/diagnostics12020241>