

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



HCMUTE

**BÁO CÁO ĐÖ ÁN CUỐI KỲ
ĐỀ TÀI: XÂY DỰNG KHO DỮ LIỆU
PHÂN TÍCH GIÁO DỤC ĐẠI HỌC**

Môn học: KHO DỮ LIỆU

Mã lớp học phần: DAWH430784_23_2_02

GVHD: ThS. Nguyễn Văn Thành

Nhóm sinh viên thực hiện: Nhóm 15

TP. Hồ Chí Minh, tháng 05 năm 2024

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

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

STT	HỌ VÀ TÊN	MSSV	TỶ LỆ ĐÓNG GÓP
1	Lai Khánh Tùng	19110496	100%
2	Nguyễn Hữu Quang	21110917	100%
3	Nguyễn Quỳnh Nhi	21110582	100%
4	Trần Ngọc Nhựt Thanh	21110647	100%

Nhận xét của giảng viên:

Ngày ... tháng 05 năm 2024

Giảng viên chấm điểm

Ths. Nguyễn Văn Thành

LỜI CẢM ƠN

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

Nhóm chúng em xin chân thành cảm ơn!

MỤC LỤC

CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI	1
1.1. Lý do chọn đề tài.....	1
1.2. Tổng quan về tập dữ liệu	1
1.2.1. Nguồn dữ liệu	1
1.2.2. Mô tả chi tiết tập dữ liệu	2
1.2.2.1. Bảng COURSES	3
1.2.2.2. Bảng Assessments.....	3
1.2.2.3. Bảng Vle	4
1.2.2.4. Bảng StudentInfo	4
1.2.2.5. Bảng StudentRegistration	5
1.2.2.6. Bảng StudentAssessment.....	6
1.2.2.7. Bảng StudentVle	6
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: Student Assessment.....	7
2.1.2. Business Process: Student Courses.....	7
2.1.3. Business Process: Student Vle:	8
2.1.4. Business Process: Student Registration:	8
2.2. Xây dựng các bảng Dimension	8
2.2.1. DimDate	8
2.2.2. DimStudentInfo.....	9
2.2.3. DimTerms	9
2.2.4. DimCourses.....	9
2.2.5. DimAssessmentTypes	9
2.2.6. DimVle.....	9
2.3. Galaxy Schema (Lược đồ galaxy)	10
CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS).....	11
3.1. Tạo ETL project và solution	11
3.2. Thiết lập các Connection	13
3.3. Import Dữ liệu vào các bảng dimension:	18
3.3.1.Date Dimension	18
3.3.1.1: Tạo Package.....	19
3.3.1.2: Thiết lập Control Flow.....	21
3.3.1.3: Load dữ liệu từ Source vào Stage	22
3.3.1.4: Thiết lập truncate table SQL task	34
3.3.1.5: Load dữ liệu từ Stage vào Dimension.....	36
3.3.2.StudentInfo Dimension	45
3.3.2.1. Tạo Package	45
3.3.2.2. Thiết lập Control Flow	45
3.3.2.3. Load dữ liệu từ Source vào Stage	47

3.3.2.4. Load dữ liệu từ Stage vào Dimension.....	51
3.3.3.Courses Dimension	52
3.3.3.1. Load dữ liệu từ Source vào Stage	52
3.3.3.2. Load dữ liệu từ Stage vào Dimension.....	56
3.3.4.Term Dimension	60
3.3.4.1. Load dữ liệu từ Source vào Stage	60
3.3.4.2. Load dữ liệu từ Stage vào Dimension.....	64
3.3.5.AssessmentTypes Dimension	67
3.3.5.1. Load dữ liệu từ Source vào Stage	67
3.3.5.2. Load dữ liệu từ Stage vào Dimension.....	70
3.3.6.Vle Dimension	73
3.3.6.1. Load dữ liệu từ Source vào Stage	73
3.3.6.2. Load dữ liệu từ Stage vào Dimension.....	76
3.4. Import dữ liệu vào bảng các bảng fact.....	79
3.4.1. Fact Student Assessment.....	79
3.4.1.1. Tạo Package	79
3.4.1.2. Thiết lập Control Flow.....	80
3.4.1.3. Load dữ liệu từ Source vào Stage	82
3.4.1.4. Load dữ liệu từ Stage vào Fact	85
3.4.2. Fact Student Courses.....	90
3.4.2.1. Load dữ liệu từ Source vào Stage	90
3.4.2.2. Load dữ liệu từ Stage vào Fact	94
3.4.3. Fact Student Registration.....	100
3.4.3.1. Load dữ liệu từ Source vào Stage	100
3.4.3.2. Load dữ liệu từ Stage vào Fact	103
3.4.4. Fact Student Vle.....	109
3.4.4.1. Load dữ liệu từ Source vào Stage	109
3.4.4.2. Load dữ liệu từ Stage vào Fact	112
CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS).....	119
4.1. Quá trình xây dựng mô hình	119
4.1.1. Tạo Data Source.....	119
4.1.2. Tạo Data Source View	120
4.2. Quá trình xây dựng khối Cube.....	122
4.3. Cấu hình Hierarchy	126
4.3.1. Tạo Hierarchy cho Dim Date	126
4.4. Thực hiện phân tích dữ liệu	127
4.4.1. Câu hỏi: Cho biết kết quả học tập của học sinh theo loại đánh giá .	127
4.4.1.1. Sử dụng Pivot Table trong Excel	127
4.4.1.2. Sử dụng Power BI.....	129
4.4.2. Câu hỏi: Cho biết kết quả học tập của học sinh theo bậc học.....	130
4.4.2.1. Sử dụng Pivot Table trong Excel	130
4.4.2.2. Sử dụng Power BI.....	132
4.4.3. Câu hỏi: Cho biết tổng số lượng thao tác trên ứng dụng của học viên trong cùng một mức học vấn	132

4.4.3.1. Sử dụng Pivot Table trong Excel	132
4.4.3.2. Sử dụng Power BI.....	134
4.4.4. Câu hỏi: Cho biết tổng số lượng thao tác và điểm số của học viên trên ứng dụng theo kết quả học tập	134
4.4.4.1. Sử dụng Pivot Table trong Excel	134
4.4.4.2. Sử dụng Power BI.....	136
CHƯƠNG 5: KẾT LUẬN	137
5.1. Kết quả đạt được	137
5.2. Những hạn chế	137
5.3. Tài liệu tham khảo	137

PHÂN CÔNG NHIỆM VỤ

Nhiệm vụ	Khánh Tùng	Quỳnh Nhi	Nhụt Thanh	Hữu Quang
1.Thiết kế và xử lý tập dữ liệu	x	x	x	x
+Tìm kiếm tập dữ liệu				x
+Phân tích nghiệp vụ				x
+Mô hình hóa dữ liệu vào workbook	x			
+Lọc dữ liệu đầu vào		x	x	
+Đẩy dữ liệu vào source	x	x	x	x
2.Thực hiện tích hợp dữ liệu vào kho bằng SSIS	x	x	x	x
2,1 Thực hiện trích xuất đưa dữ liệu vào CSDL Stage	x	x	x	x
+Course, Terms, AssessmentTypes, Date				x
+StudentInfo, StudentAssessment			x	

+StudentCourses, +StudentRegistration		x		
+StudentVle, Vle	x			
2.2 Trích xuất dữ liệu vào kho	x	x	x	x
+DimCourse, DimTerms, DimAssessmentTypes, DimDate				x
+DimStudentInfo, DimStudentAssessment	x			
+DimStudentCourses, DimStudentRegistration		x		
+DimStudentVle, DimVle			x	
Tạo cây phân cấp cho DimDate	x			
3.Trả lời câu hỏi theo từng business process bằng các công cụ SSAS, Power BI và Pivot Table.	x	x	x	x

CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI

1.1. Lý do chọn đề tài

- Việc chọn đề tài "Education" trong việc xây dựng kho dữ liệu xuất phát từ nhiều lý do quan trọng. Giáo dục là nền tảng của sự phát triển cá nhân và xã hội, giúp nâng cao chất lượng nguồn nhân lực và đáp ứng nhu cầu thị trường lao động. Một kho dữ liệu giáo dục chi tiết là nguồn tài nguyên vô giá cho nghiên cứu, hỗ trợ giảng viên và sinh viên trong việc phân tích và cải tiến chất lượng dạy học. Nó còn giúp xác định và giải quyết bất bình đẳng trong tiếp cận giáo dục, đảm bảo cơ hội học tập công bằng cho mọi người. Dữ liệu giáo dục cũng là cơ sở để hoạch định chính sách hiệu quả, theo dõi xu hướng mới và thích ứng với các thay đổi trong giáo dục hiện đại. Hơn nữa, kho dữ liệu này thúc đẩy hợp tác quốc tế, chia sẻ kinh nghiệm và học hỏi từ các hệ thống giáo dục tiên tiến. Như vậy, xây dựng kho dữ liệu giáo dục là một bước đi quan trọng, không chỉ hỗ trợ nghiên cứu và quản lý, mà còn thúc đẩy sự phát triển bền vững của giáo dục và xã hội.

1.2. Tổng quan về tập dữ liệu

1.2.1. Nguồn dữ liệu

- Để hỗ trợ nghiên cứu trong lĩnh vực này, nhóm đã sử dụng một bộ dữ liệu phân tích học tập của trường đại học mở (OULAD) của Đại học mở của Vương quốc Anh). Điều làm cho tập dữ liệu trở nên độc đáo là nó chứa dữ liệu nhân khẩu học cùng với dữ liệu dòng nhập chuột tổng hợp về các tương tác của học sinh trong Môi trường học tập ảo (VLE)
- Đường dẫn tải tập dữ liệu: [OULAD](#)

1.2.2. Mô tả chi tiết tập dữ liệu

- Tập dữ liệu ‘OULAD’ chứa dữ liệu về các khóa học, sinh viên và sự tương tác của họ với Môi trường học tập ảo (VLE) cho bảy khóa học được chọn (được gọi là mô-đun). Các bài thuyết trình về các khóa học bắt đầu vào tháng 2 và tháng 10 - chúng được đánh dấu lần lượt bằng “B” và “J”. Tập dữ liệu bao gồm các bảng được kết nối bằng mã định danh duy nhất. Tất cả các bảng được lưu trữ ở định dạng csv. Dữ liệu bao gồm các 7 bảng với mô tả như sau:

Tên bảng	Mô tả	Số dòng
courses	Tệp này chứa danh sách tất cả các mô-đun có sẵn và bản trình bày của chúng.	319 dòng
assessments	Tệp này chứa thông tin về các đánh giá trong bài thuyết trình mô-đun.	7 dòng
vle	Tệp csv chứa thông tin về các tài liệu có sẵn trong VLE.	212 dòng
studentInfo	Tệp này chứa thông tin nhân khẩu học về học sinh cùng với kết quả của họ.	384 dòng
studentRegistration	Tệp này chứa thông tin về thời gian sinh viên đăng ký thuyết trình mô-đun.	384 dòng
studentAssessment	Tệp này chứa kết quả đánh giá của học sinh.	1634 dòng
studentVle	Tệp sinh viênVle.csv chứa thông tin về tương tác của mỗi học sinh với các tài liệu trong VLE.	5000 dòng

1.2.2.1. Bảng COURSES

Biến	Mô tả
code_module	tên mã của mô-đun, dùng làm mã định danh.
code_presentation	tên mã của bài thuyết trình. Nó bao gồm năm và “B” cho bài thuyết trình bắt đầu vào tháng 2 và “J” cho bài thuyết trình bắt đầu vào tháng 10.
length	độ dài của phần trình bày mô-đun tính bằng ngày.

1.2.2.2. Bảng Assessments

Biến	Mô tả
code_module	mã nhận dạng của mô-đun chưa đánh giá.
code_presentation	mã nhận dạng của bản trình bày mà đánh giá thuộc về.
id_assessment	số nhận dạng của đánh giá.
assessment_type	loại đánh giá. Có ba loại bài đánh giá: Đánh giá do giáo viên chấm (TMA), Đánh giá do máy tính chấm (CMA) và Kỳ thi cuối kỳ (Bài kiểm tra).
date	thông tin về ngày nộp bài đánh giá cuối cùng được tính bằng số ngày kể từ khi bắt đầu trình bày mô-đun. Ngày bắt đầu của bài thuyết trình có số 0 (không).
weight	trọng số của đánh giá tính bằng %. Thông thường, Bài kiểm tra được xử lý riêng biệt và có trọng số 100%; tổng của tất cả các đánh giá khác là 100%.

1.2.2.3. Bảng Vle

Biến	Mô tả
id_site	số nhận dạng của tài liệu.
code_module	mã nhận dạng cho mô-đun.
code_presentation	mã nhận dạng của trình bày.
activity_type	vai trò liên quan đến tài liệu mô-đun.
week_from	tuần mà tài liệu dự kiến sẽ được sử dụng.
week_to	tuần cho đến khi tài liệu được lên kế hoạch sử dụng.

1.2.2.4. Bảng StudentInfo

Biến	Mô tả
code_module	mã nhận dạng cho mô-đun mà sinh viên đã đăng ký.
code_presentation	mã nhận dạng của bài thuyết trình mà sinh viên được đăng ký trên mô-đun.
id_student	mã số nhận dạng duy nhất của sinh viên.
gender	giới tính – giới tính của học sinh.
region	xác định khu vực địa lý, nơi sinh viên sống trong khi tham gia bài thuyết trình mô-đun.
highest_education	trình độ học vấn cao nhất của sinh viên khi bắt đầu thuyết trình mô-đun.
imd_band	chỉ định chỉ số phạm vi mức độ sa đọa đa dạng của nơi sinh viên sống trong buổi thuyết trình mô-đun.
age_band	nhóm tuổi của học sinh.
num_of_prev_attempts	số lần học sinh đã thử mô-đun này.
studied_credits	tổng số tín chỉ cho các học phần mà sinh viên hiện đang học.
disability	cho biết liệu học sinh có tuyên bố khuyết tật hay không.
final_result	kết quả cuối cùng của học sinh trong phần trình bày mô-đun.

1.2.2.5. Bảng StudentRegistration

Biến	Mô tả
code_module	mã nhận dạng cho một mô-đun.
code_presentation	mã nhận dạng của bài thuyết trình.
id_student	mã số nhận dạng duy nhất của sinh viên.
date_registration	ngày sinh viên đăng ký thuyết trình mô-đun, đây là số ngày được tính tương ứng với thời điểm bắt đầu trình bày mô-đun (ví dụ: giá trị âm -30 có nghĩa là sinh viên đã đăng ký trình bày mô-đun 30 ngày trước khi bắt đầu).
date_unregistration	ngày sinh viên hủy đăng ký thuyết trình mô-đun, đây là số ngày được tính tương ứng với thời điểm bắt đầu thuyết trình mô-đun. Những sinh viên đã hoàn thành khóa học sẽ để trống trường này. Những sinh viên chưa đăng ký có giá trị Rút tiền là giá trị của cột Final_result trong tệp sinh viênInfo.csv.

1.2.2.6. Bảng StudentAssessment

Biến	Mô tả
id_assessment	số nhận dạng của đánh giá.
id_student	mã số nhận dạng duy nhất của sinh viên.
date_submitted	ngày nộp bài của sinh viên, được tính bằng số ngày kể từ khi bắt đầu trình bày mô-đun.
is_banked	còn trạng thái cho biết kết quả đánh giá đã được chuyển từ bản trình bày trước đó.
score	điểm của học sinh trong đánh giá này. Phạm vi từ 0 đến 100. Điểm thấp hơn 40 được hiểu là Thất bại. Các điểm nằm trong khoảng từ 0 đến 100.

1.2.2.7. Bảng StudentVle

Biến	Mô tả
code_module	mã nhận dạng cho một mô-đun.
code_presentation	mã nhận dạng của bản trình bày mô-đun.
id_student	mã số nhận dạng duy nhất của sinh viên.
id_site	số nhận dạng cho vật liệu VLE.
date	ngày tương tác của sinh viên với tài liệu được tính bằng số ngày kể từ khi bắt đầu trình bày mô-đun.
sum_click	số lần học sinh tương tác với tài liệu trong ngày đó.

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)

2.1. Xác định các Business Process và bảng Fact

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

Instructions!		Business Process Name	Fact Table	Fact Grain Type	Granularity	Facts	studentinfo	Courses	Terms	AssessmentTypes	VLE	Date
studentAssessment	studentAssessment_fact	Transaction	one row per studentinfo			date_submitted, is_banked, score	x	x	x	x		x
studentCourses	studentCourses_fact	Accumulating snapshot	one row per terms			module_presentation_length		x	x			x
studentVle	studentVle_fact	Transaction	one row per student with VLE			interaction_date, sum_click, sum_from, sum_to	x	x	x		x	x
studentRegistration	studentRegistration_fact	Periodic Snapshot	one row per student registration event			date_registration, date_unregistration	x	x	x			x

2.1.1. Business Process: Student Assessment

- Câu hỏi cụ thể được đặt ra: Thông tin về kết quả của các đánh giá của sinh viên theo khóa học trong học kì khác nhau ảnh hưởng như thế nào đến kết quả học tập để tối ưu hóa chúng cho phù hợp, đánh giá và theo dõi xu hướng về kết quả học tập của sinh viên theo khu vực, tuổi, trình độ học.

- Bảng FactStudentAssessment

Instructions!		Attribute / Fact Name	Description	Alternate Names	Sample Values
StudentAssessment	StudentAssmtKey				1
StudentAssessment	assessment_typeKey				1
StudentAssessment	code_module_key				AAA
StudentAssessment	code_presentation_key				2013J
StudentAssessment	studentInfoKey				11391
StudentAssessment	date_submitted				30/4/2024
StudentAssessment	id_assessment				1
StudentAssessment	is_banked				1
StudentAssessment	score				10

2.1.2. Business Process: Student Courses

- Câu hỏi cụ thể được đặt ra: Theo dõi việc đăng ký và hủy đăng ký của sinh viên trong các khóa học, phân tích tỷ lệ hủy đăng ký và ổn định của sinh viên trong khóa học.

- Bảng FactStudentCourses

Instructions!		Attribute / Fact Name	Description	Alternate Names	Sample Values
StudentCourses	code_module				AAA
StudentCourses	code_presentation_key				2013J
StudentCourses	module_presentation_length				12

2.1.3. Business Process: Student Vle:

- Câu hỏi cụ thể được đặt ra: đánh giá mức độ tham gia của sinh viên trong hoạt động trực tuyến, phát hiện xu hướng học tập trực tuyến, và đánh giá hiệu suất của các hóa học trên VLE theo kỳ học.

- **Bảng FactStudentVle**

Instructions!	Attribute / Fact Name	Description	Alternate Names	Sample Values
Dimension / Fact Table				
StudentVle	studentVleKey			
StudentVle	code_module_key			AAA
StudentVle	code_presentation_key			2013J
StudentVle	studentInfoKey			11391
StudentVle	vle_activity_type_key			1
StudentVle	id_site			1
StudentVle	interaction_date			28/04/2024
StudentVle	sum_click			3
StudentVle	week_from			1
StudentVle	week_to			2

2.1.4. Business Process: Student Registration:

- Câu hỏi cụ thể được đặt ra:
 - + Theo dõi quá trình đăng ký và hủy đăng ký của sinh viên vào các khóa học theo kỳ.
 - + Cung cấp thông tin về số lần đăng ký lại của sinh viên (nếu có) và tốc độ hủy đăng ký trong các khóa học cụ thể.

- **Bảng FactStudentRegistration:**

Instructions!	Attribute / Fact Name	Description	Alternate Names	Sample Values
Dimension / Fact Table				
StudentRegistration	StudentRegKey			1
StudentRegistration	code_module_key			AAA
StudentRegistration	code_presentation_key			2013J
StudentRegistration	studentInfoKey			11391
StudentRegistration	date_registration			28/04/2024
StudentRegistration	date_unregistration			29/04/2024

2.2. Xây dựng các bảng Dimension

2.2.1. DimDate

2.2.2. DimStudentInfo

Instructions!	Dimension / Fact Table	Attribute / Fact Name	Description	Alternate Names	Sample Values
	StudentInfo	disability			N
	StudentInfo	final_result			Pass
	StudentInfo	gender			M
	StudentInfo	highest_education			HE Qualification
	StudentInfo	id_student			11391
	StudentInfo	imd_band			imd_band
	StudentInfo	num_of_prev_attempts			0
	StudentInfo	region			East Anglian Region
	StudentInfo	studied_redits			240

2.2.3. DimTerms

Instructions!	Dimension / Fact Table	Attribute / Fact Name	Description	Alternate Names	Sample Values
	Terms	code_presentation_key			2013J
	Terms	code_presentation			AEHIU234GG

2.2.4. DimCourses

Instructions!	Dimension / Fact Table	Attribute / Fact Name	Description	Alternate Names	Sample Values
	Courses	code_module_key			AAA
	Courses	code_module			ABCCC

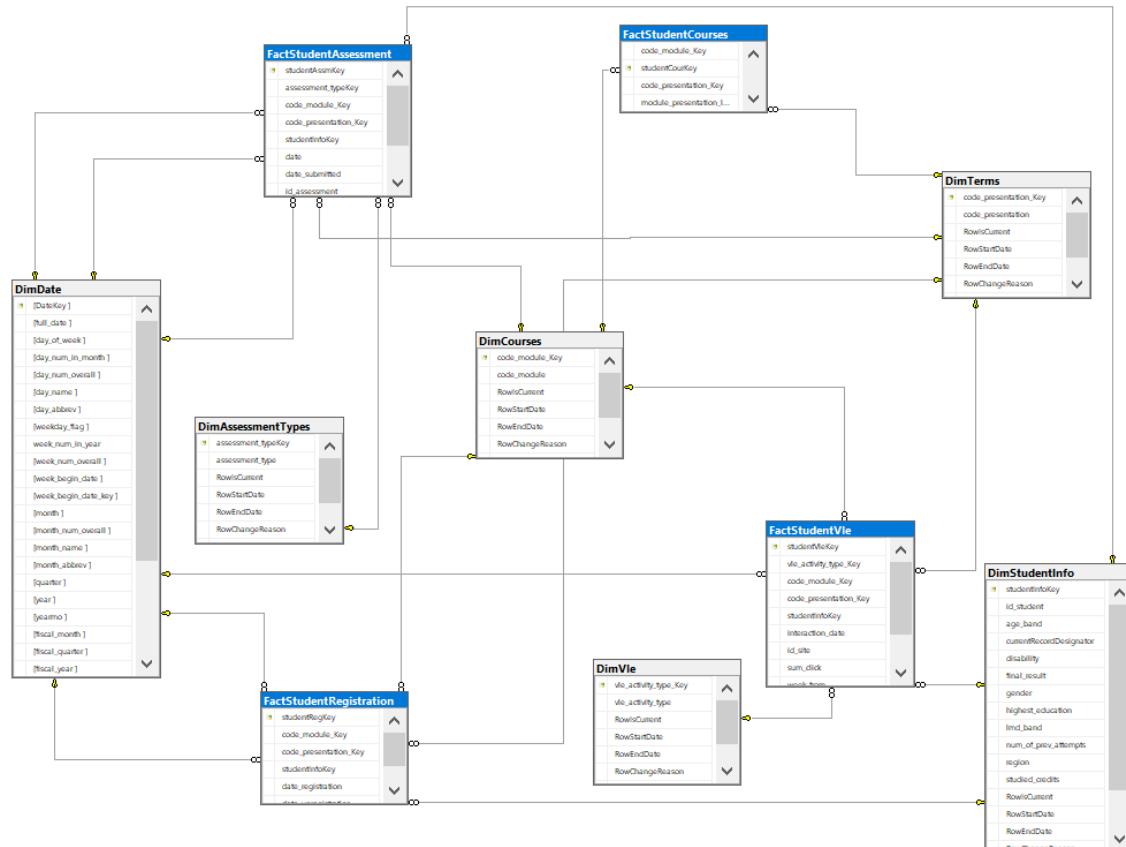
2.2.5. DimAssessmentTypes

Instructions!	Dimension / Fact Table	Attribute / Fact Name	Description	Alternate Names	Sample Values
	AssessmentTypes	assessment_typekey			1
	AssessmentTypes	assessment_type			TMA

2.2.6. DimVle

Instructions!	Dimension / Fact Table	Attribute / Fact Name	Description	Alternate Names	Sample Values
	Vle	vle_activity_type_key			1
	Vle	vle_activity_type			resource

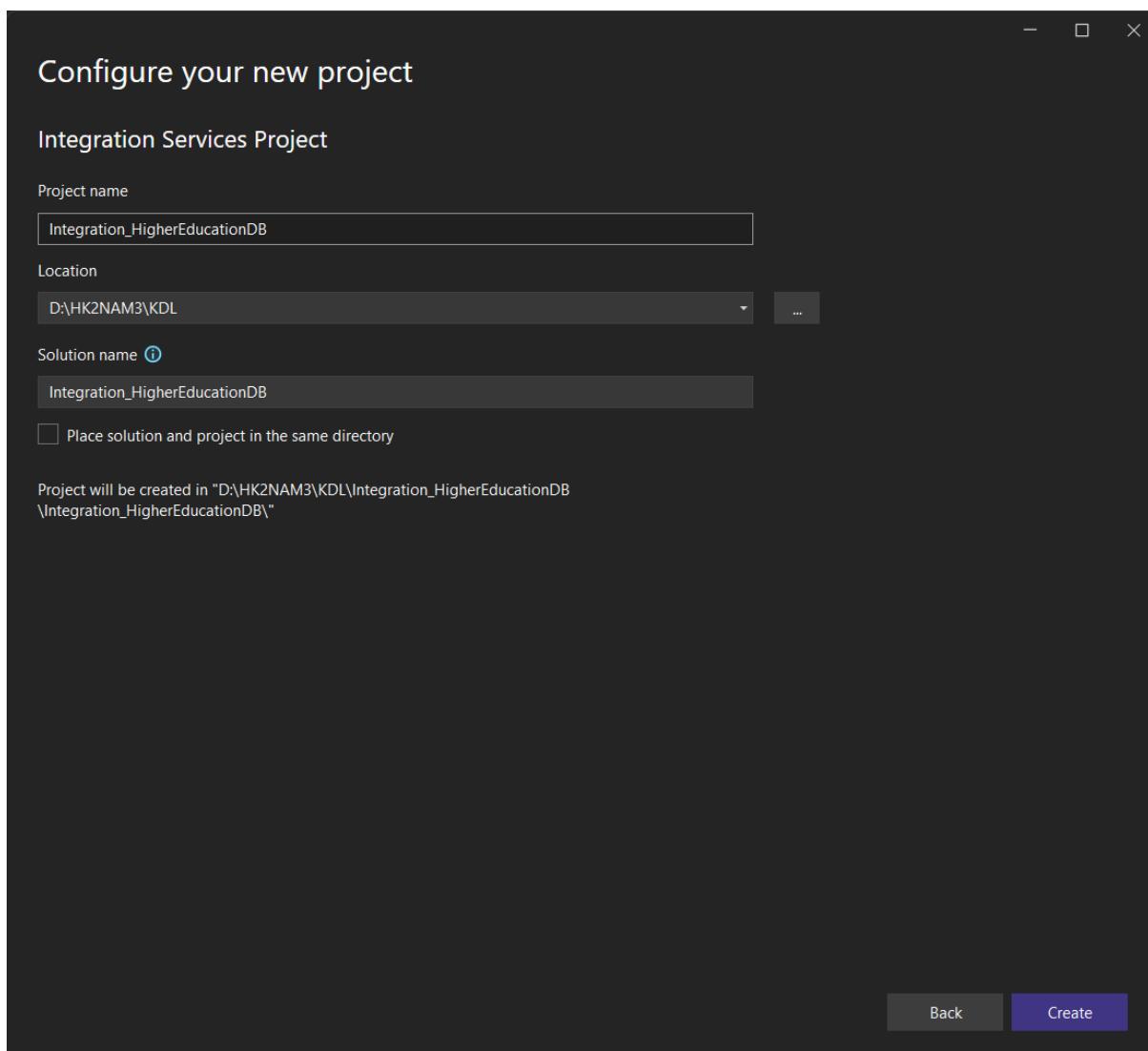
2.3. Galaxy Schema (Lược đồ galaxy)

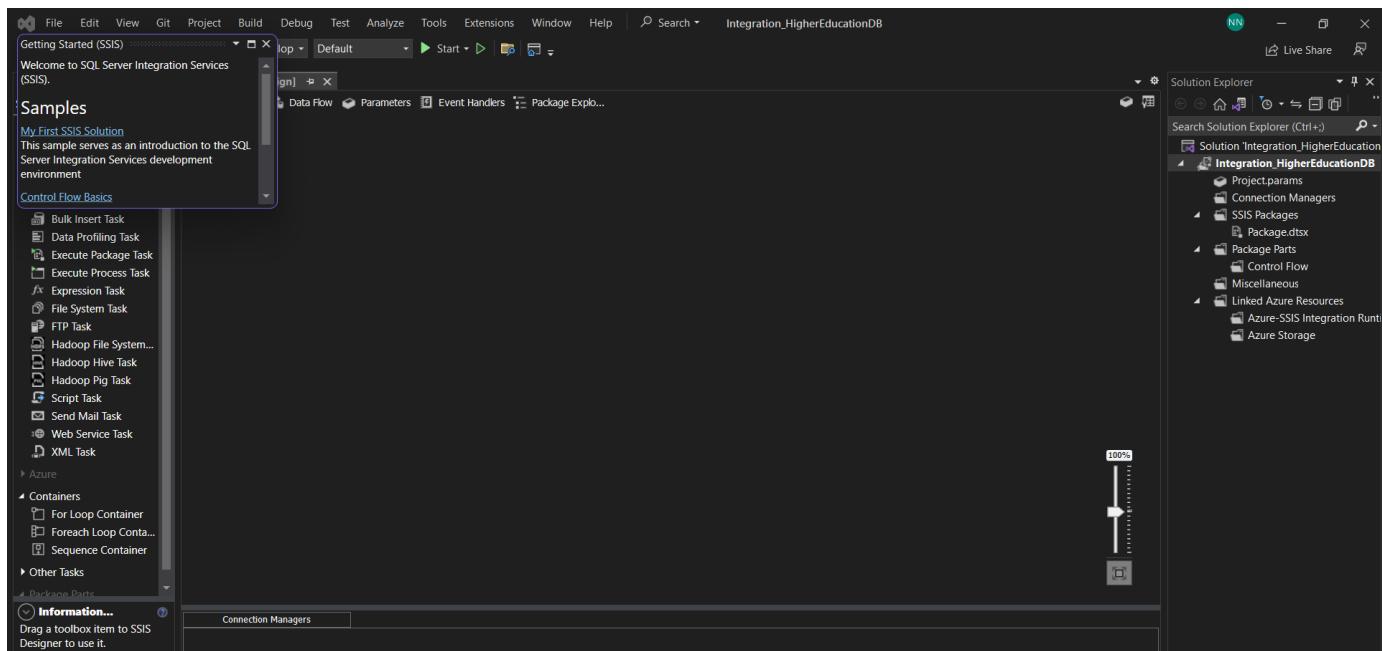


CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS)

3.1. Tạo ETL project và solution

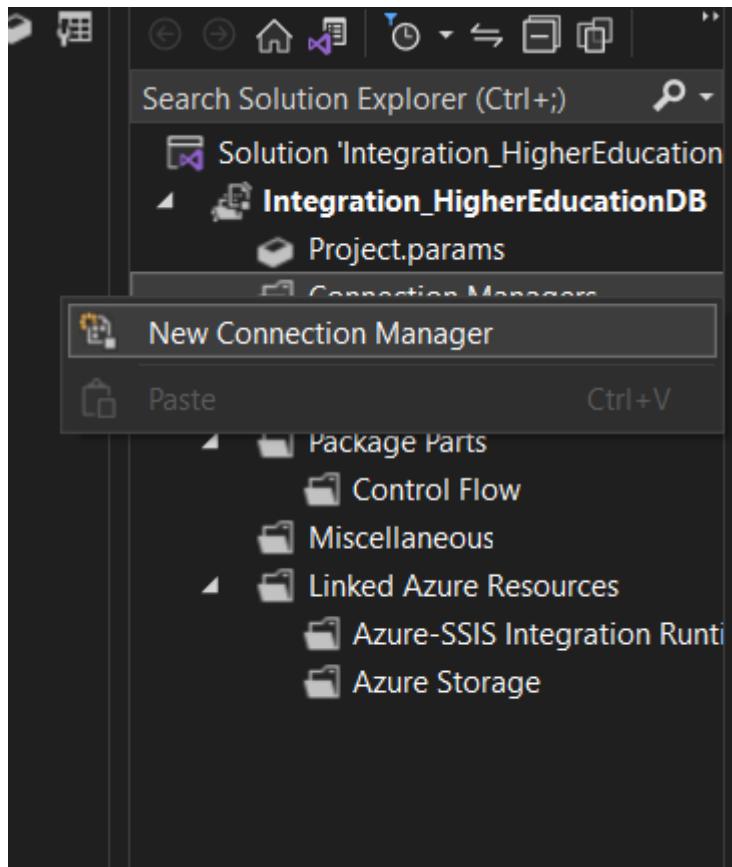
1. Từ **Visual Studio** menu, chọn **File -> New -> Project**.
2. Từ **dialog**:
 - a. Chọn **Integration Services Project**
 - b. Đặt tên: **Integration_HigherEducationDB**
3. Click **OK**



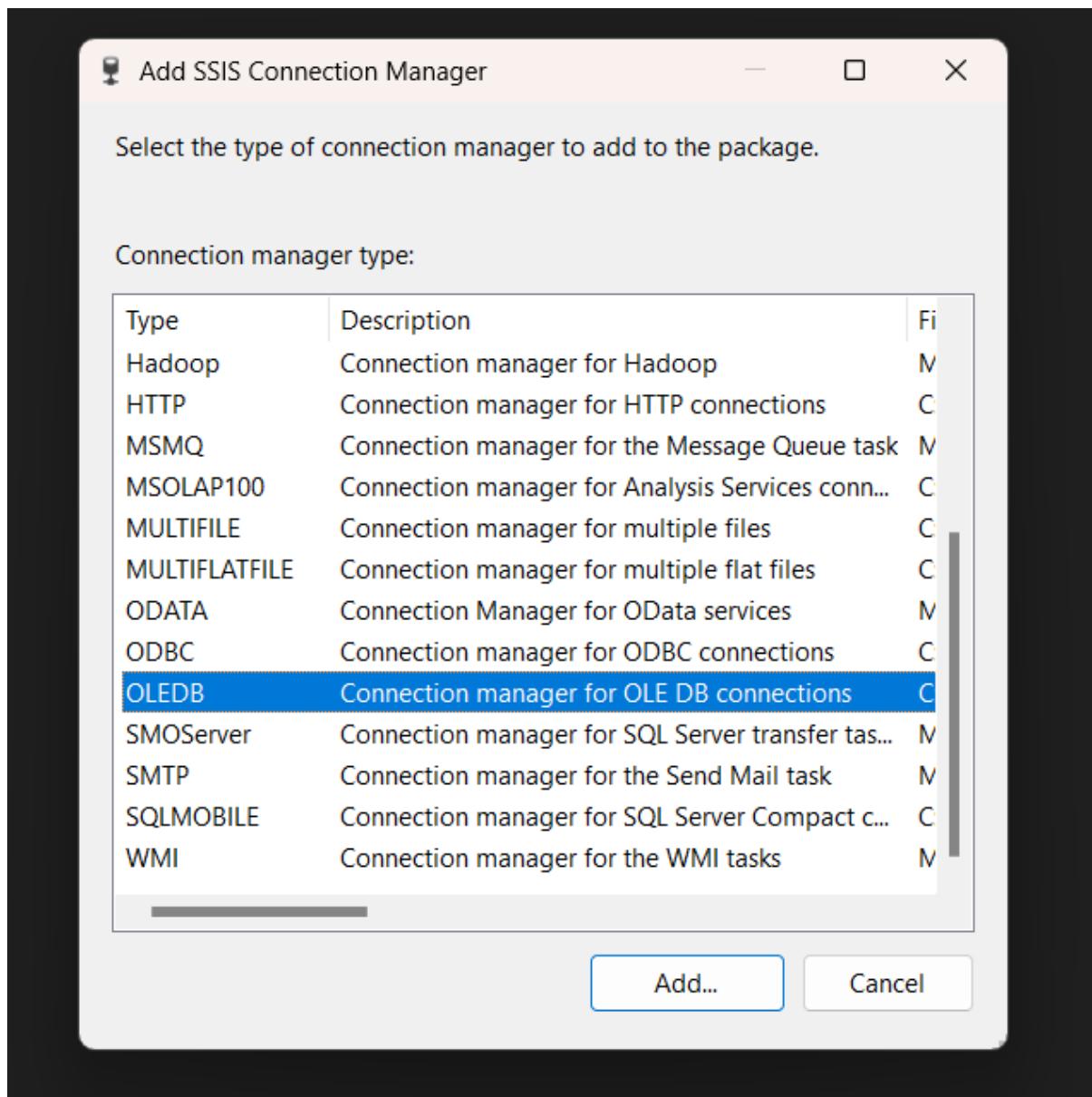


3.2. Thiết lập các Connection

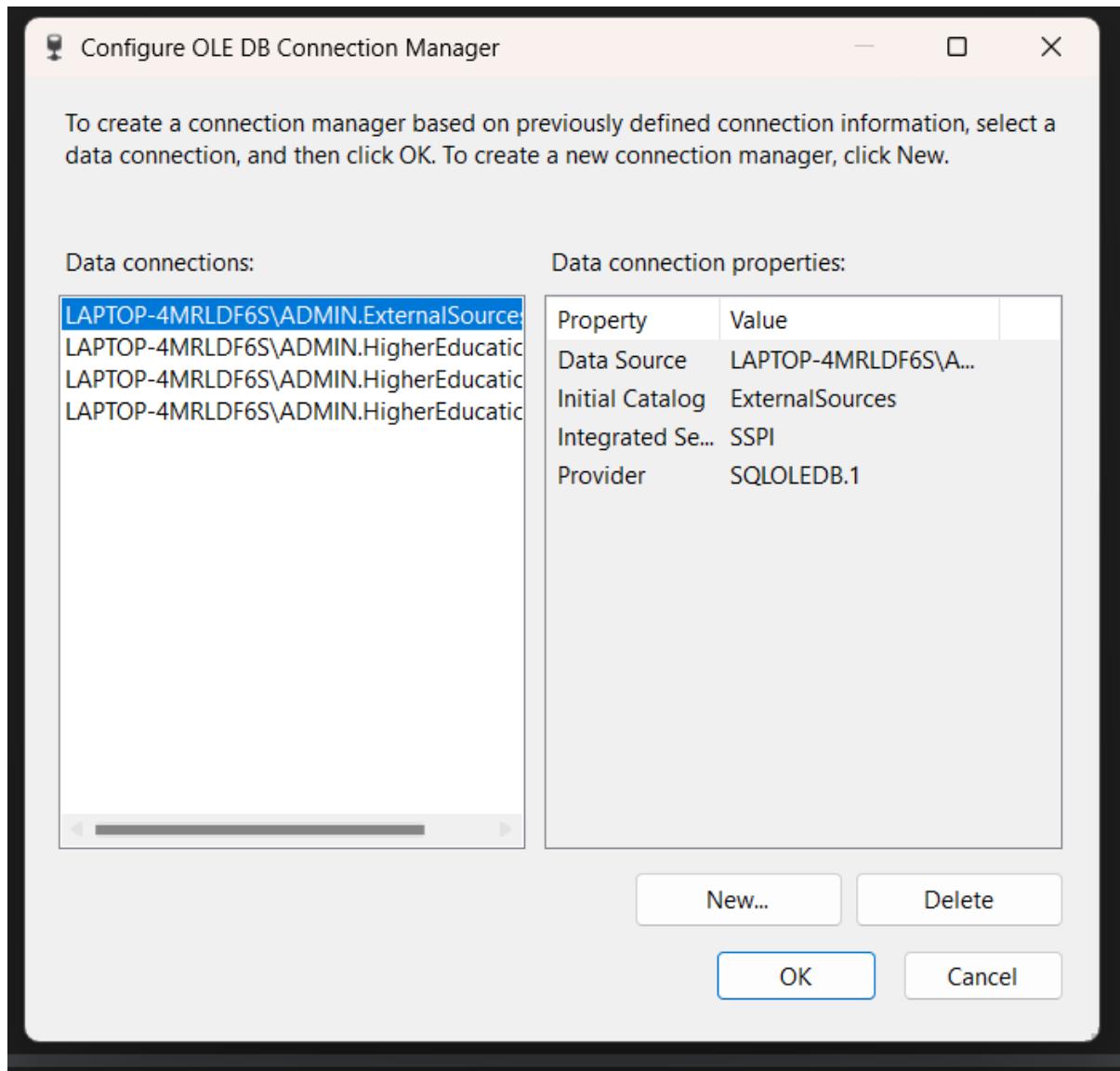
1. Tại **solution explorer**, right-click vào **Connection Managers**, chọn **New Connection Manager**.



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



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



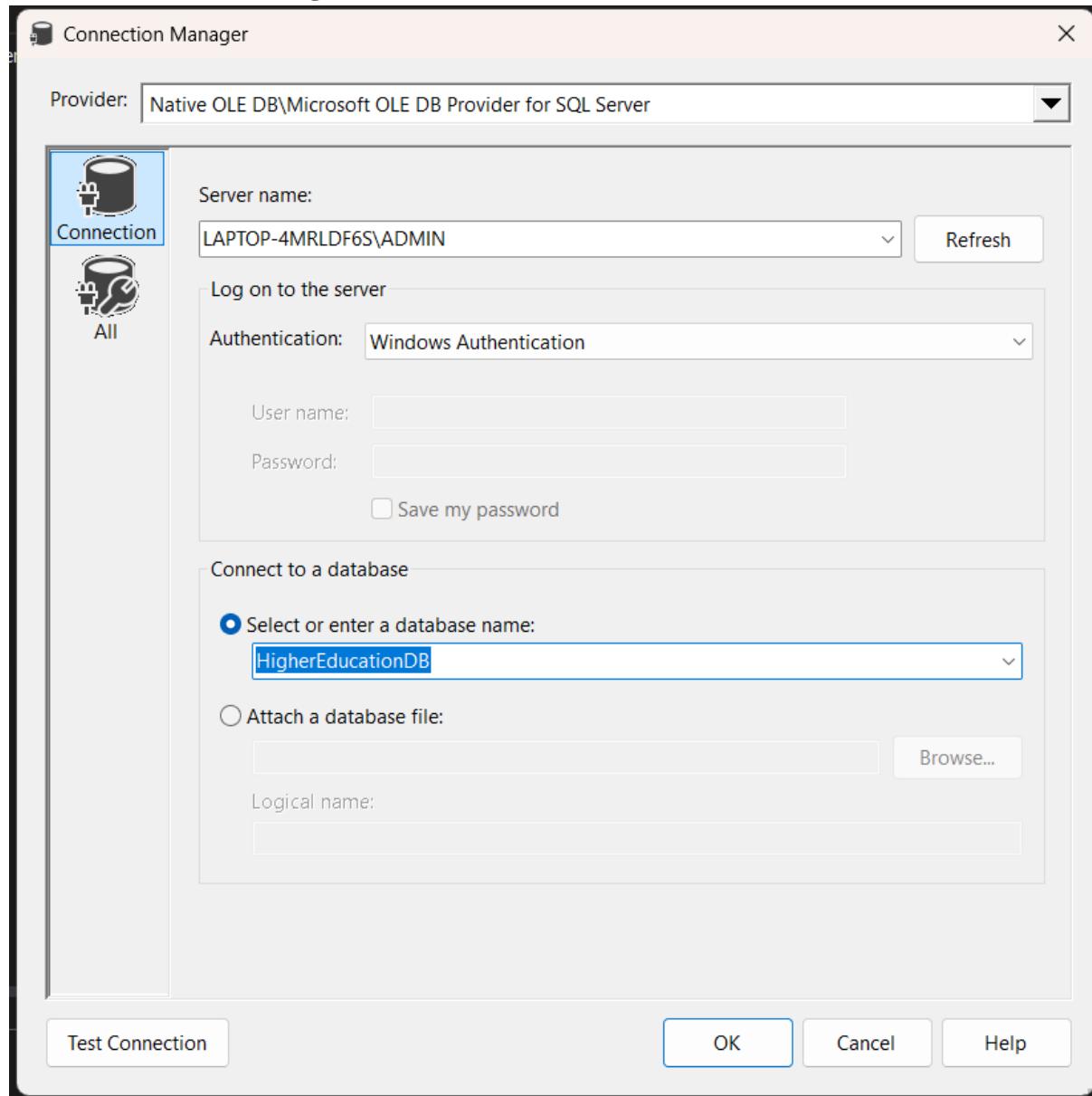
Đặt tên cho các thông tin như bên dưới:

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

Server Name: **LAPTOP-4MRLDF6S\ADMIN**

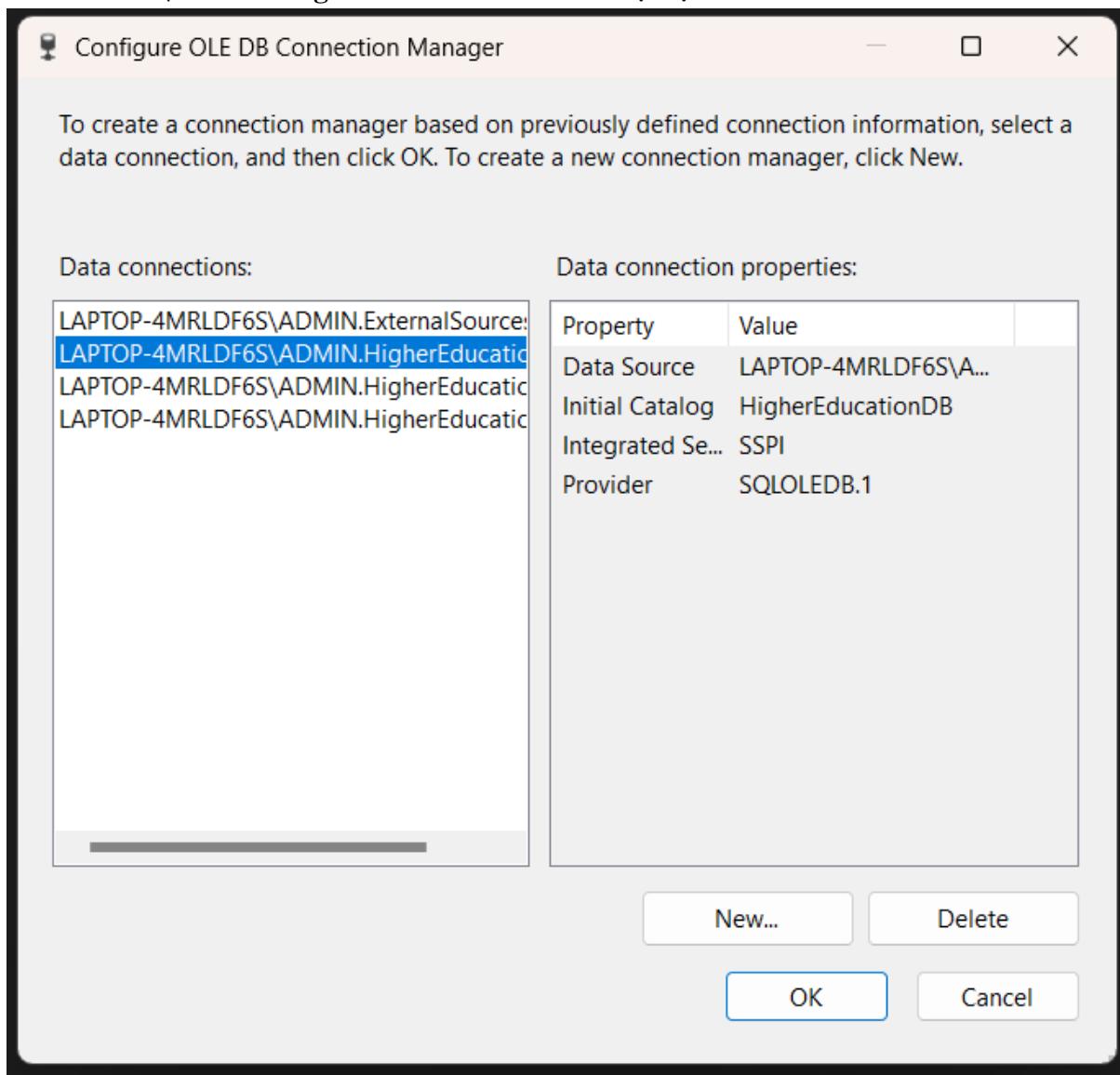
Log on to the Server: **Use Windows Authentication**

Connect to a database: **HigherEducationDB**



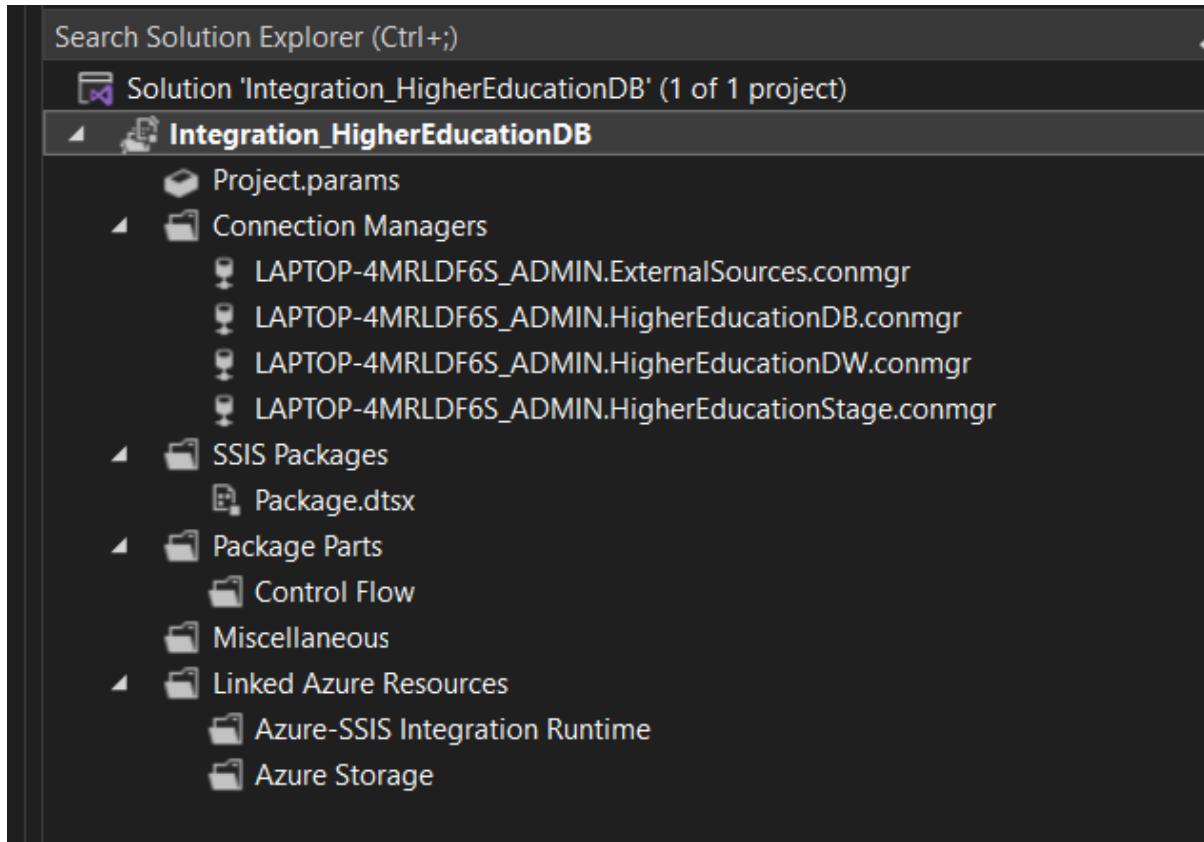
Click **Test Connection** để kiểm tra connection hoạt động, sau đó click **OK** để lưu connection.

4. Trở lại màn hình **Configure OLE DB Connection Manager**, ta được **LAPTOP-4MRLDF6S\ADMIN.HigherEducationDB** vừa được tạo:



5. Chọn **HigherEducationDB** connection và click **OK** để thêm connection này vào project.

6. Lặp lại các bước trên (bước 1->5) ba lần cho lần lượt **HigherEducationDW**, **HigherEducationStage** và **ExternalSources**. Sau khi tạo xong **Connection Managers** folder sẽ như bên dưới.



3.3. Import Dữ liệu vào các bảng dimension:

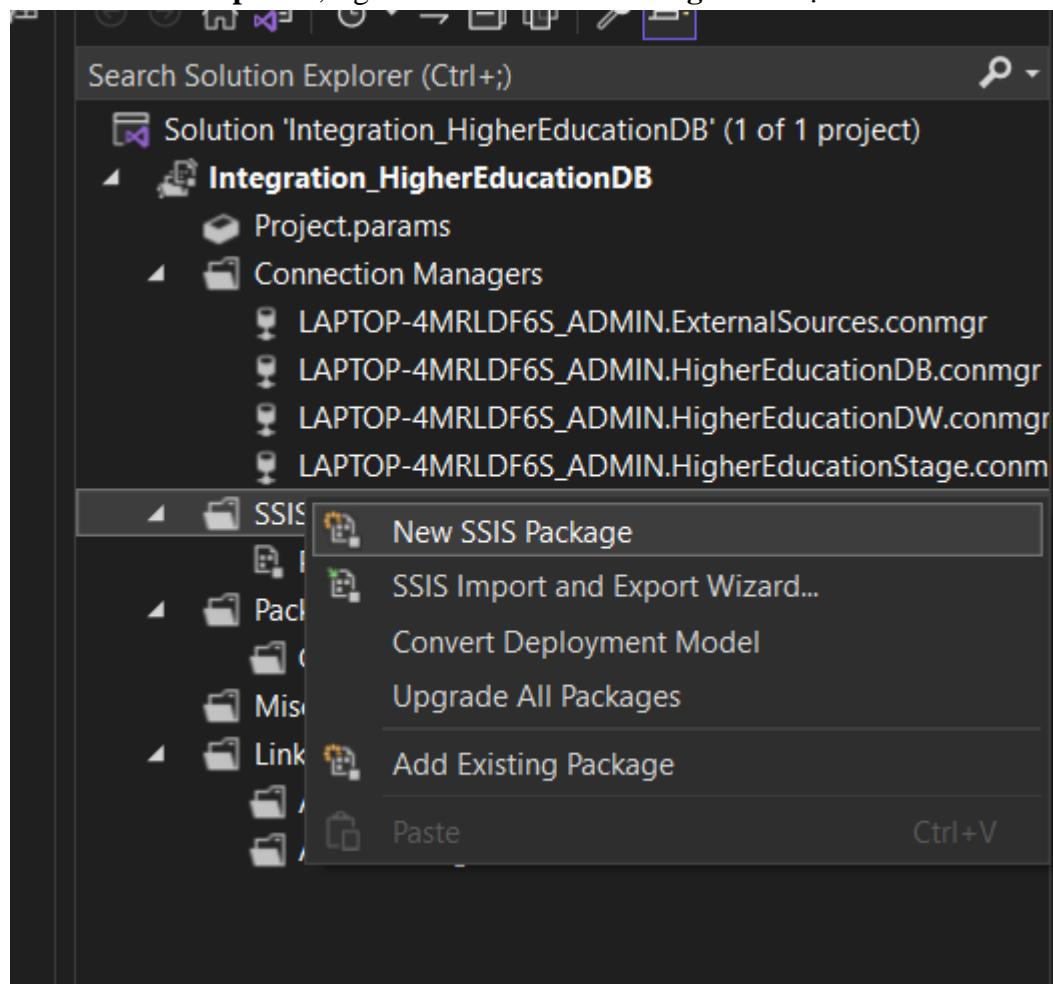
3.3.1. Date Dimension

- Đầu tiên chép dữ liệu từ bảng **[Date]** của CSDL **ExternalSources** sang bảng **stgDate** CSDL **HigherEducationStage**. Sử dụng SSIS để tạo destination table, truncate table trước khi staging dữ liệu.
- Ánh xạ các cột từ bảng **stgDate** sang bảng **DimDate**.

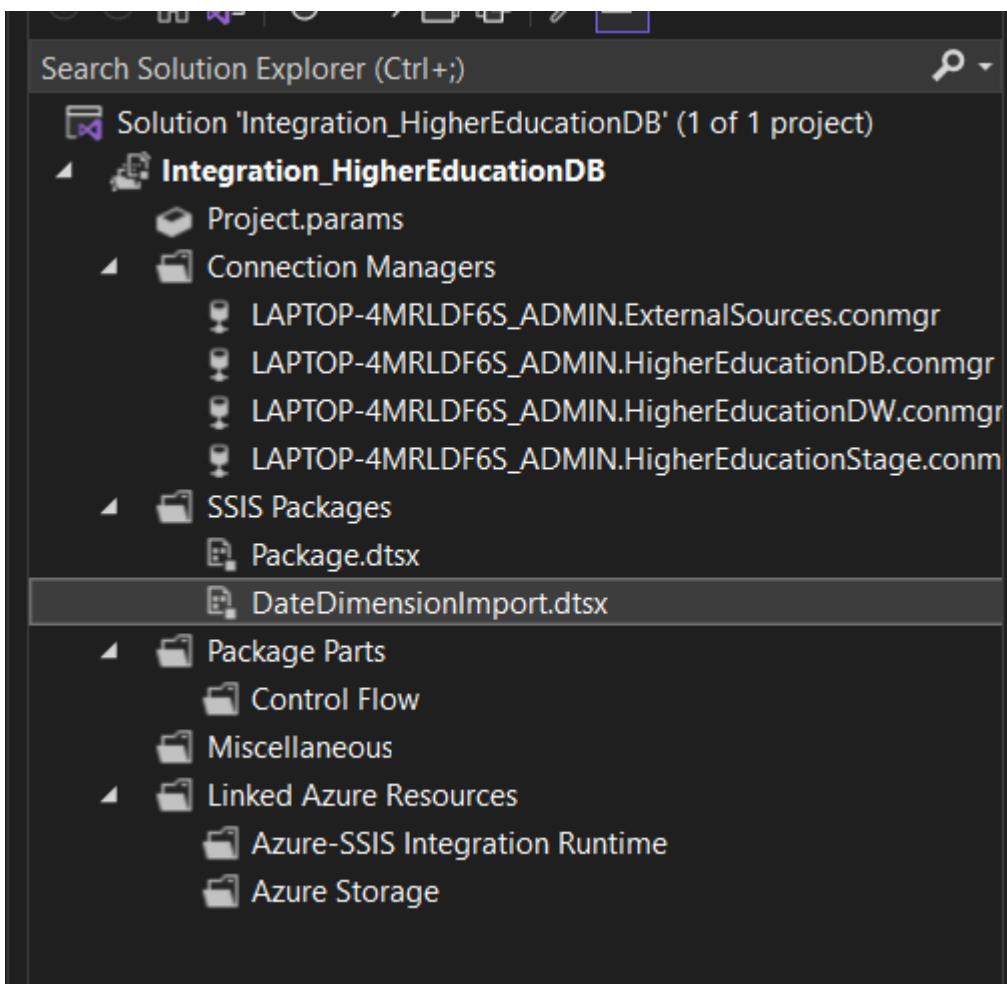
3.3.1.1: Tạo Package

Tạo một package mới, **DateDimensionImport**.

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



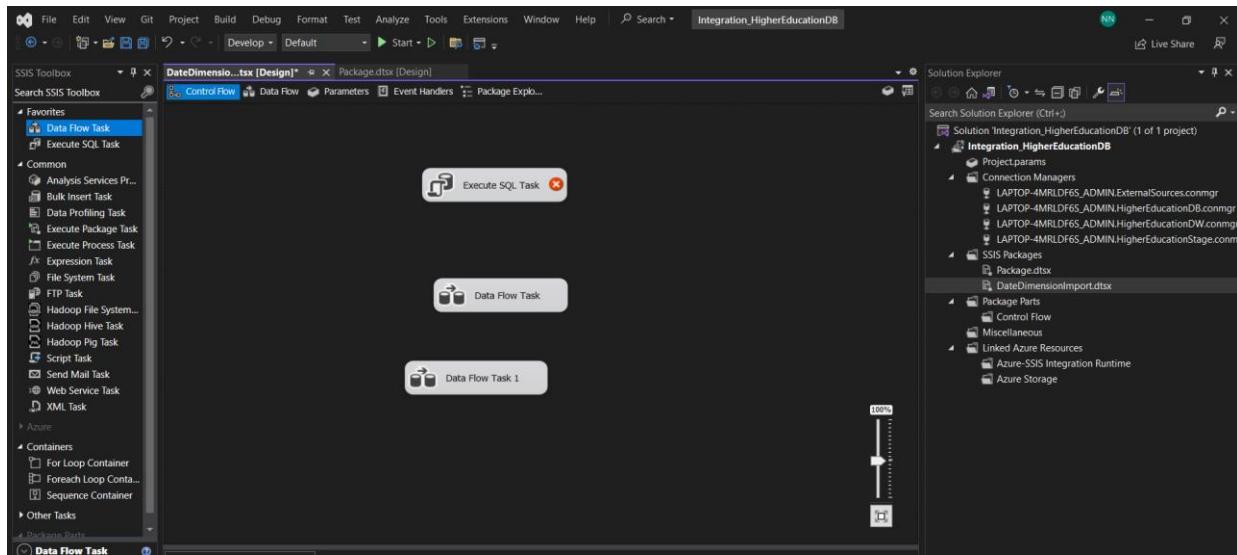
2. Right-Click vào package có tên **Package1.dtsx** và chọn **Rename**
3. Đặt tên package là **DateDimensionImport**
4. Sau khi thực hiện xong ta được **Package.dtsx** và **DateDimensionImport.dtsx** như hình sau:



5. Double-click vào **DateDimensionImport.dtsx** package để mở nó.

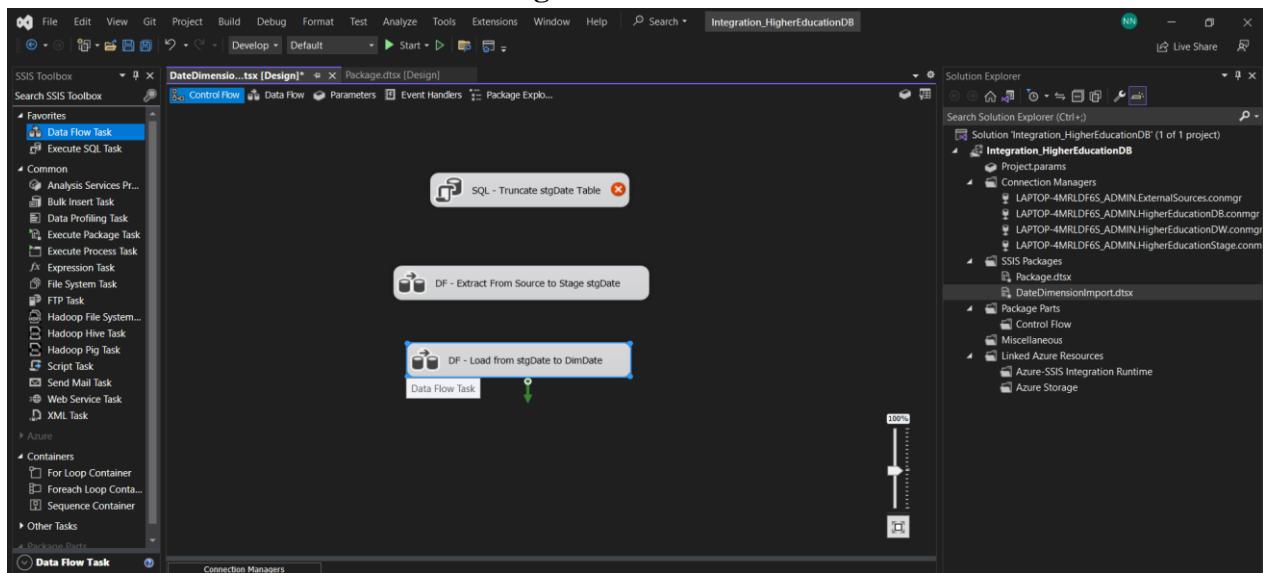
3.3.1.2: Thiết lập Control Flow

1. Từ SSIS Toolbox tạo 1 **Execute SQL task** và 2 **Data flow tasks** vào surface design như hình sau:

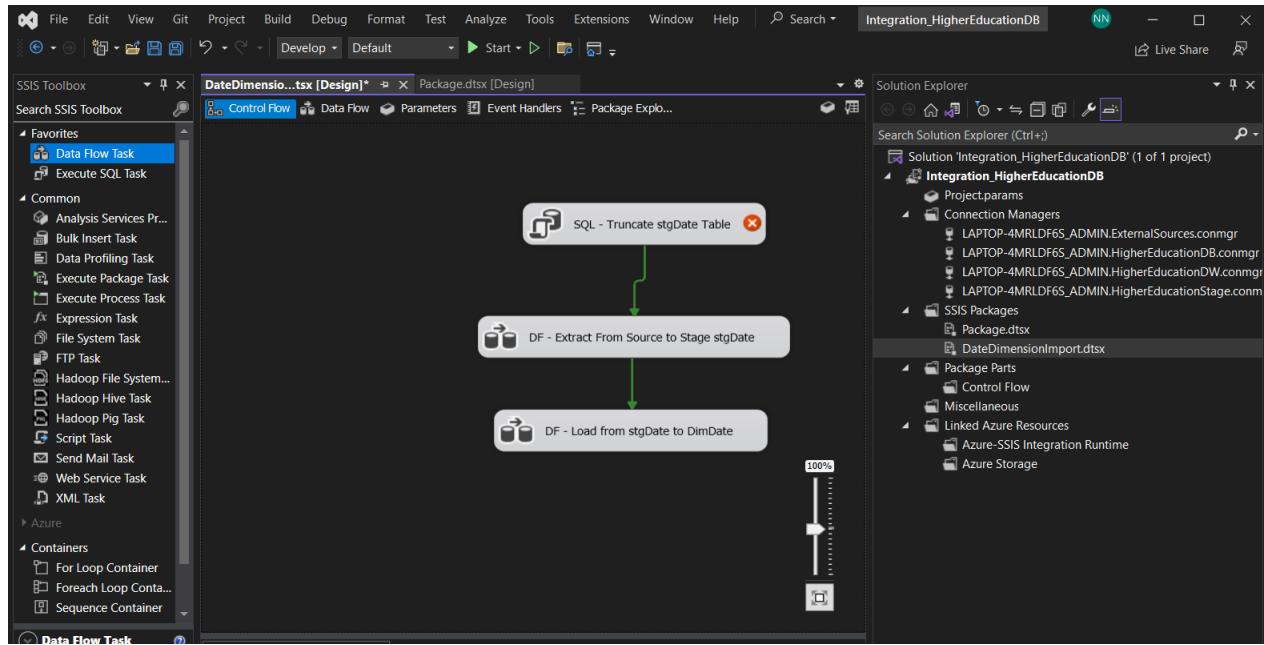


2. Đặt tên cho các task.

- Execute SQL Task SQL - Truncate stgDate Table**
- Data Flow Task DF - Extract From Source to Stage stgDate**
- Data Flow Task 1 DF - Load from stgDate to DimDate**

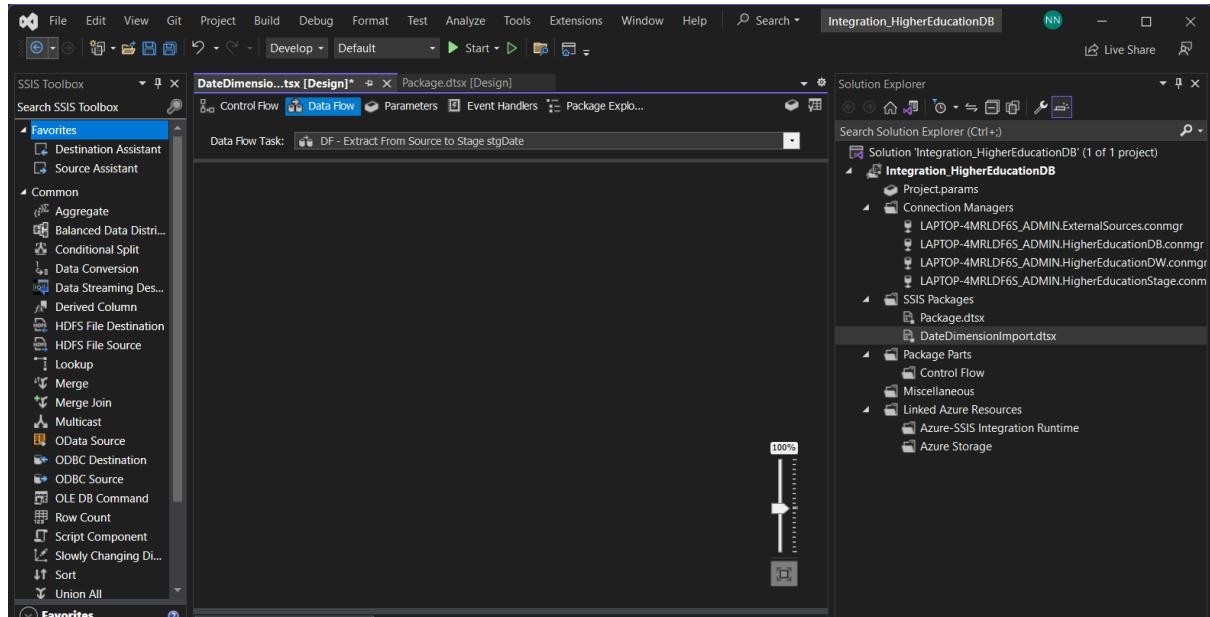


3. Connect các **tasks** để chúng có thể chạy theo thứ tự. Click vào **SQL - Truncate stgDate Table** task và một mũi tên xanh sẽ xuất hiện. Kéo rê nó vào **DF - Extract From Source To Stage stgDate** task. Tương tự, connect hai task còn lại.

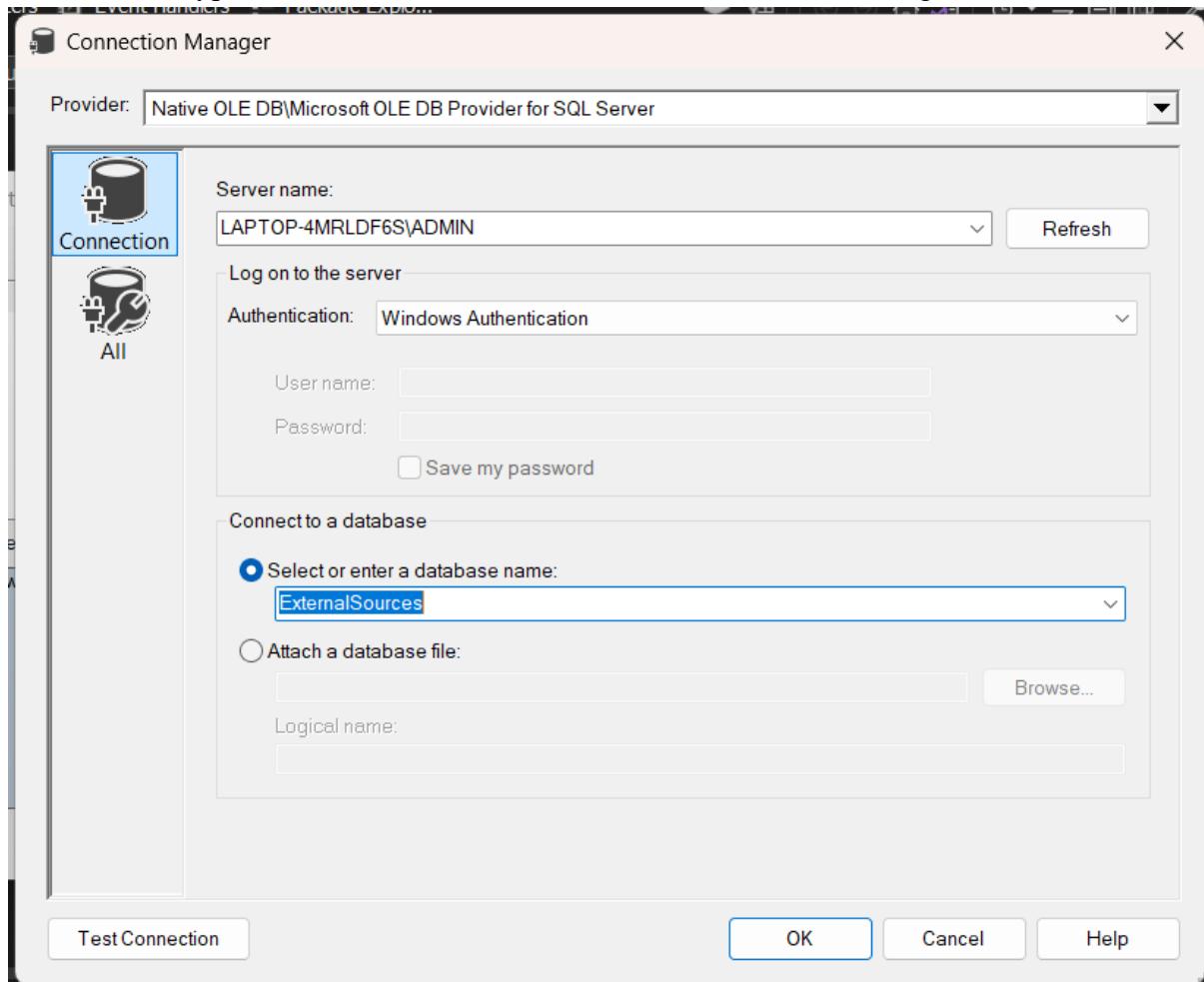


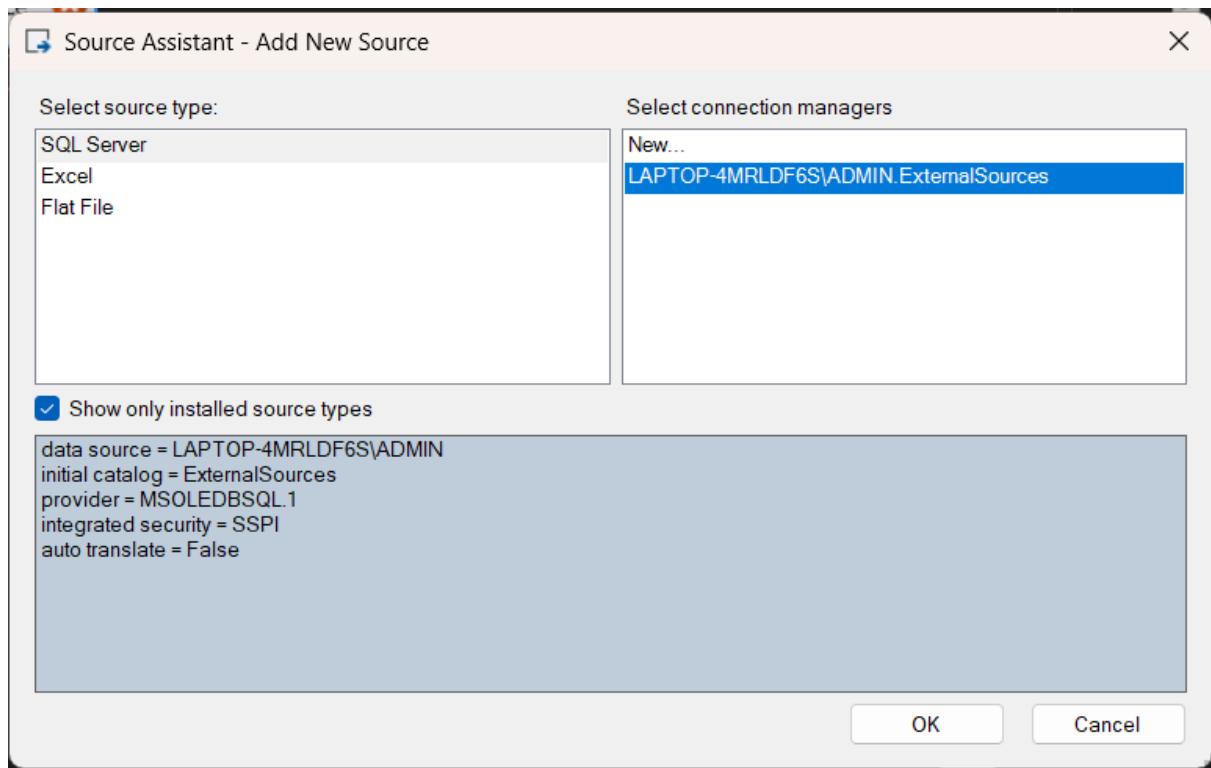
3.3.1.3: Load dữ liệu từ Source vào Stage

1. Double-click **DF - Extract From Source to Stage stgDate** task để mở nó trong **Data Flow design surface**:



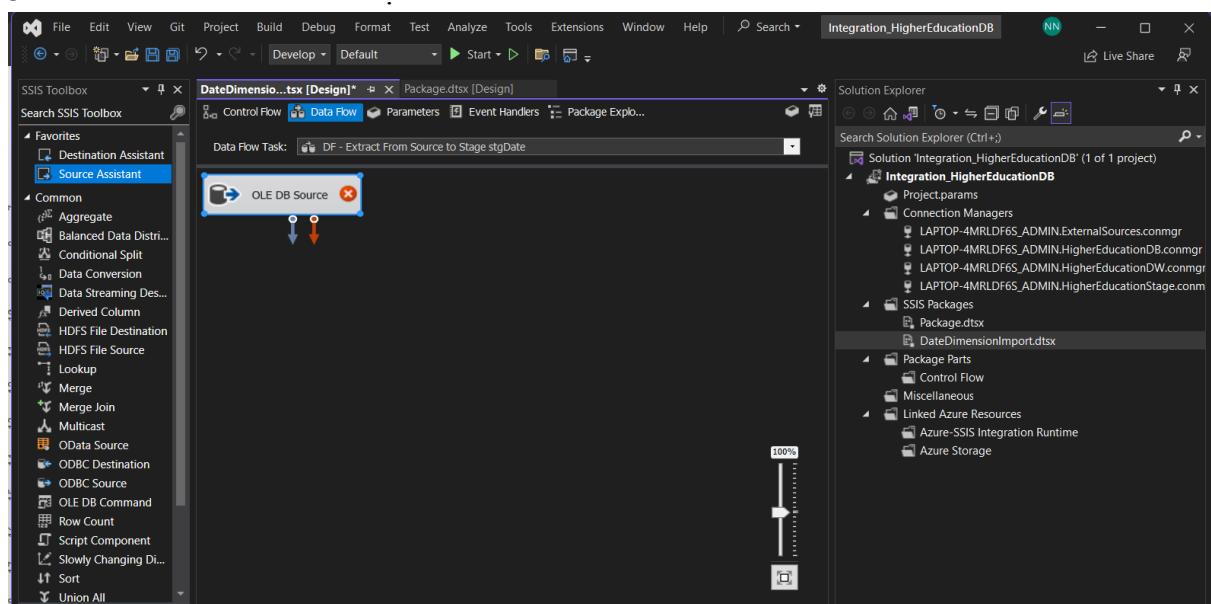
2. Từ **SSIS Toolbox**, kéo rê **Source Assistant** vào design surface, một dialog mở ra, Chon source Type: **SQL Server** và **ExternalSources** connection manager.



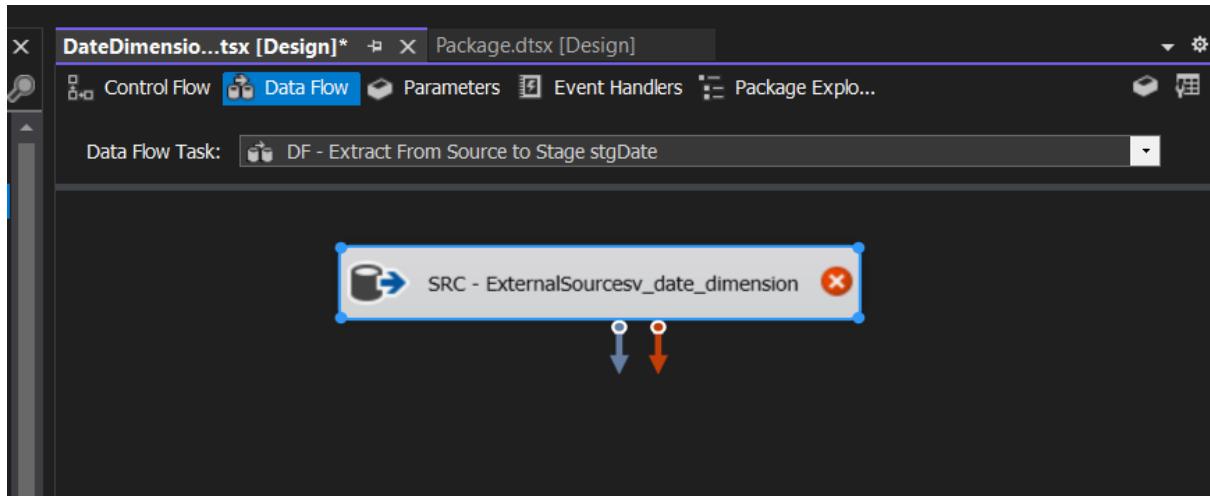


Click OK.

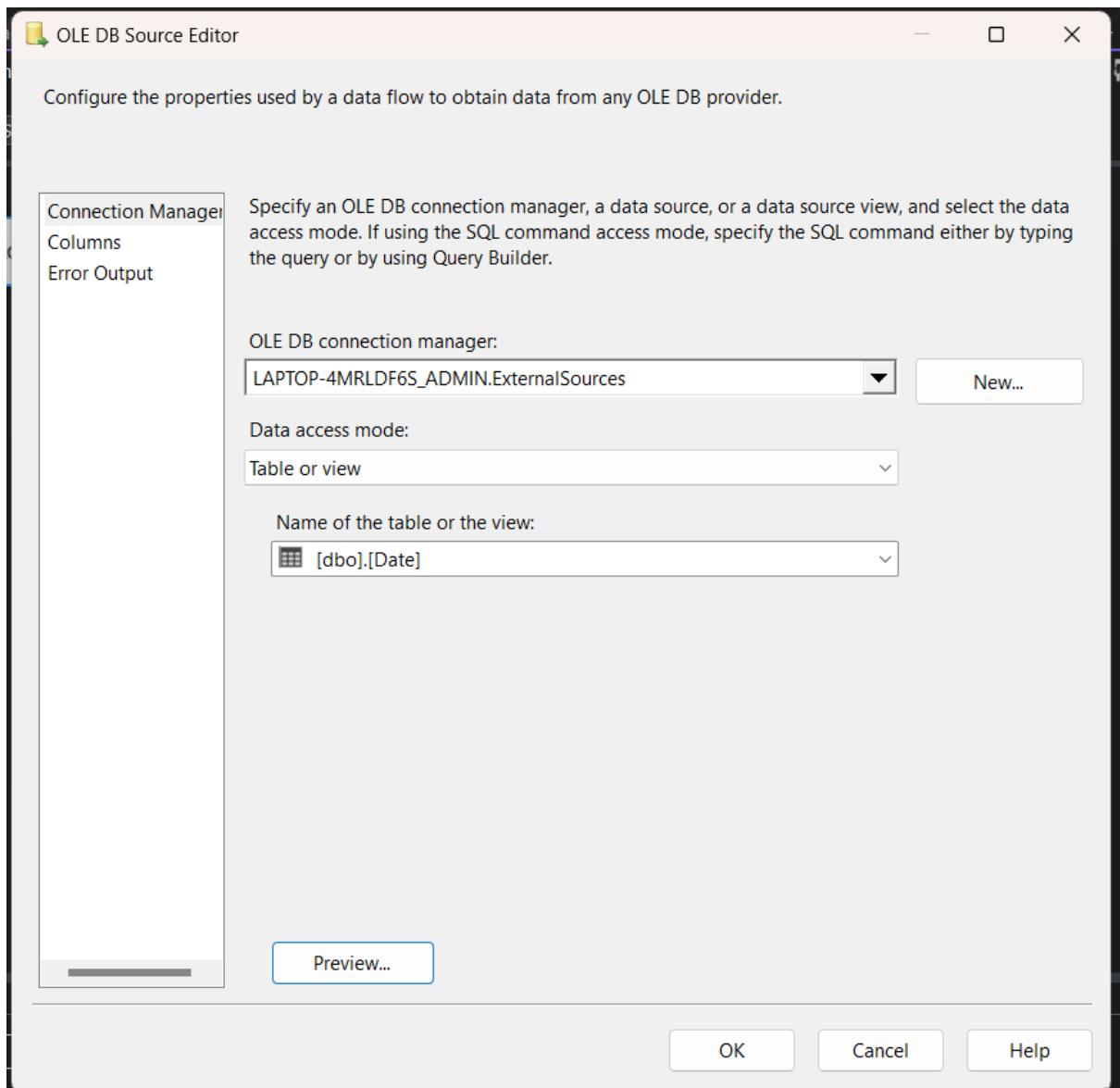
3. OLE DB Source sẽ xuất hiện như sau:



4. Đổi tên nó thành SRC - ExternalSourcesv_date_dimension

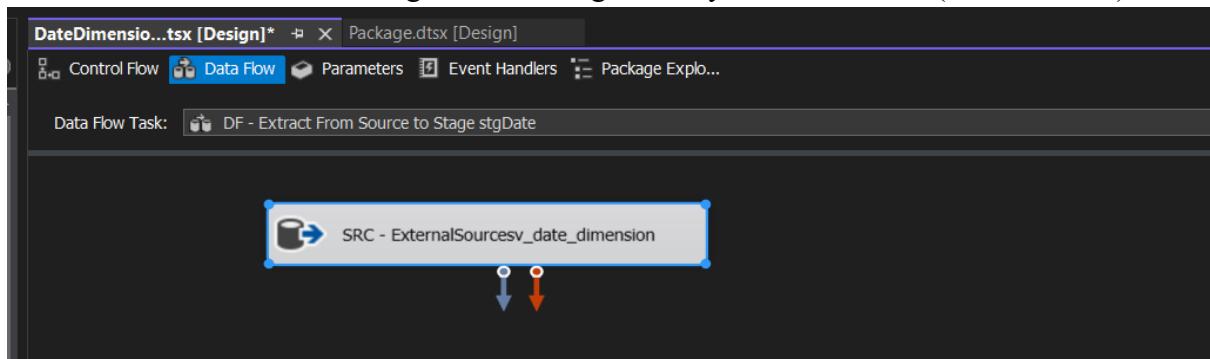


5. Double click vào source để cấu hình nó, **OLE DB Source Editor** sẽ mở ra. Từ dropdown **name of the table or view** chọn [**dbo].[Date]** là source.
Click **Preview...** để xem source data.

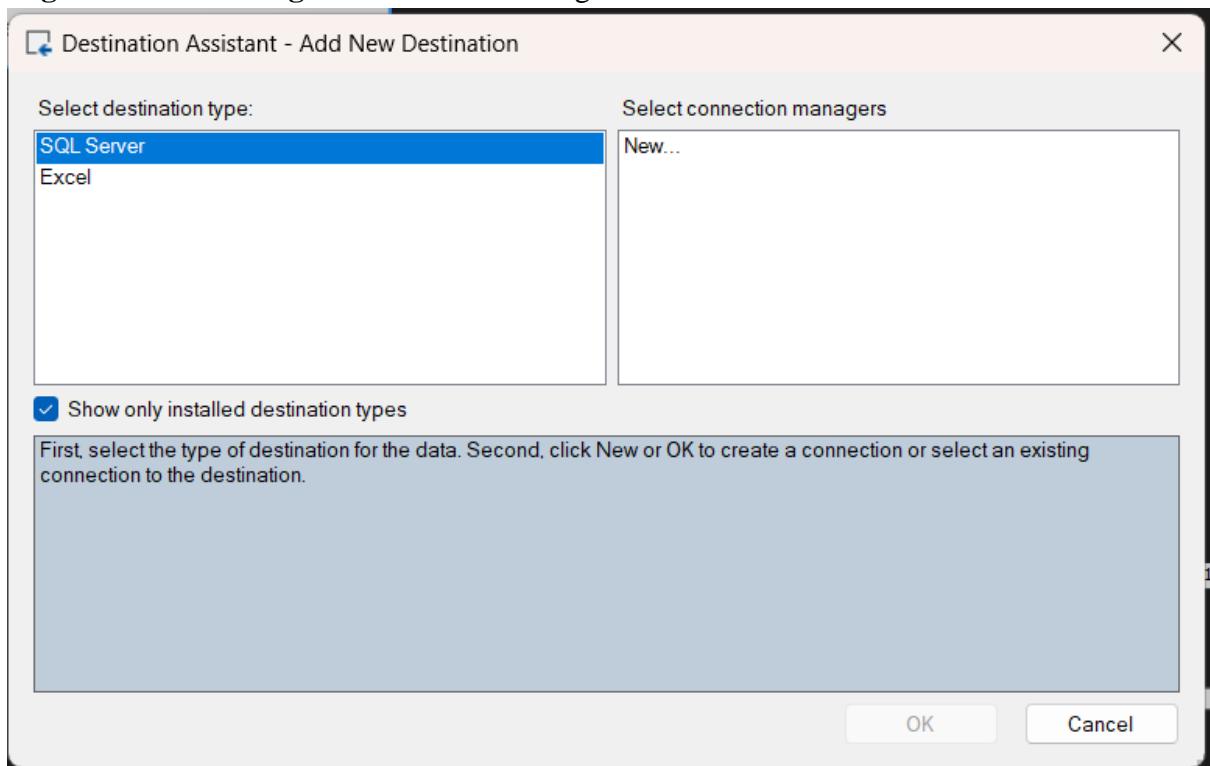


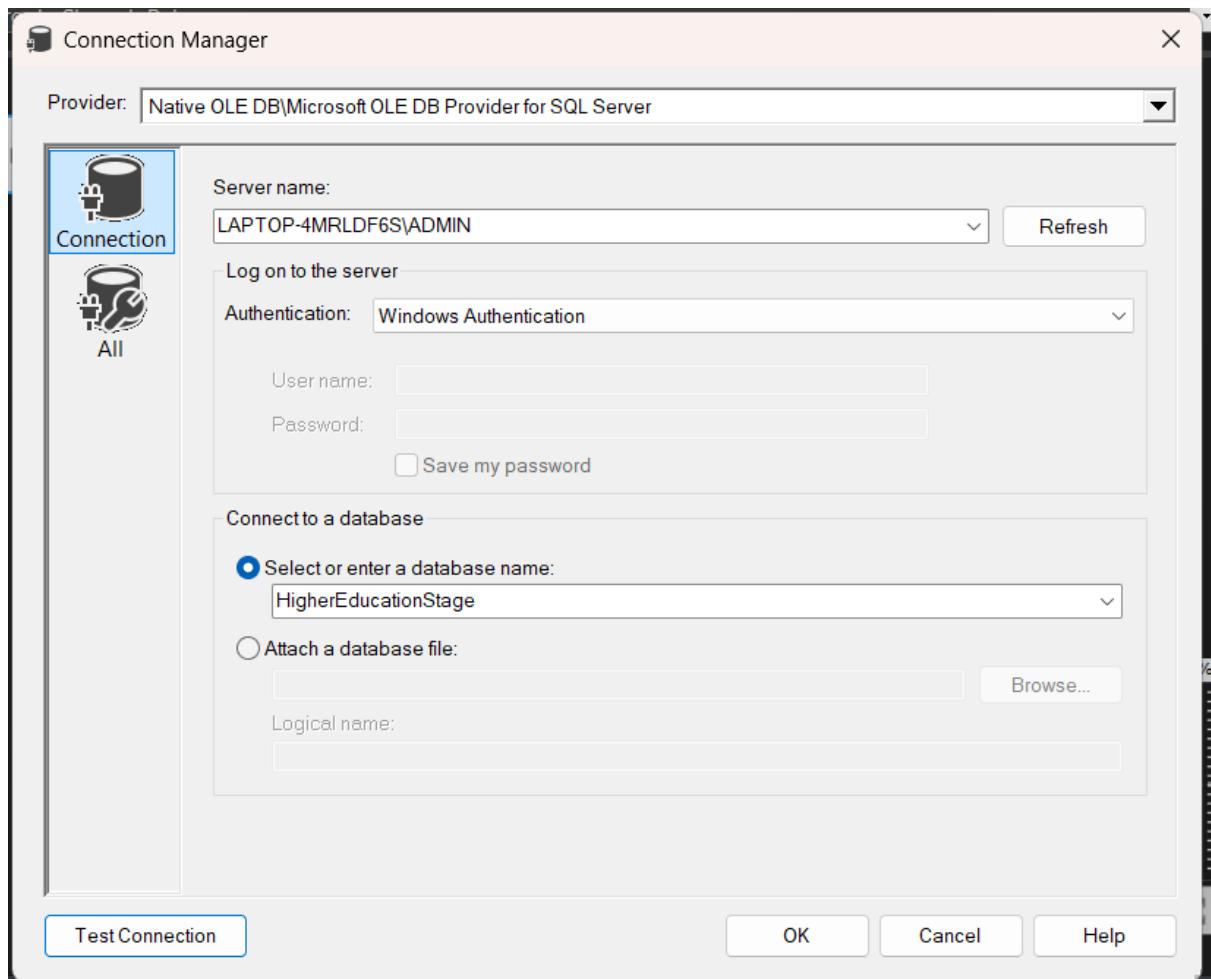
Click **OK** để lưu.

6. Sau khi cấu hình source xong, bạn sẽ không còn thấy dấu [x] màu đỏ (như bên dưới):



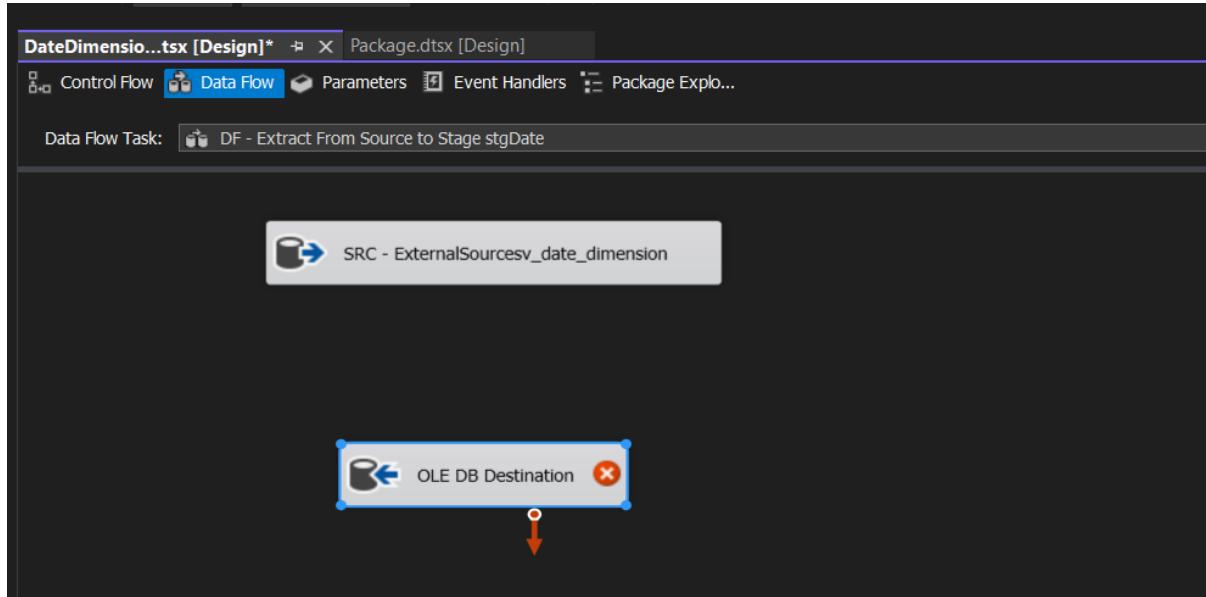
7. Tiếp tục, cấu hình destination. Từ **SSIS Toolbox**, kéo rê **Destination Assistant** vào design surface, một dialog xuất hiện. Chọn source Type: **SQL Server** và **HigherEducationStage** là connection manager.



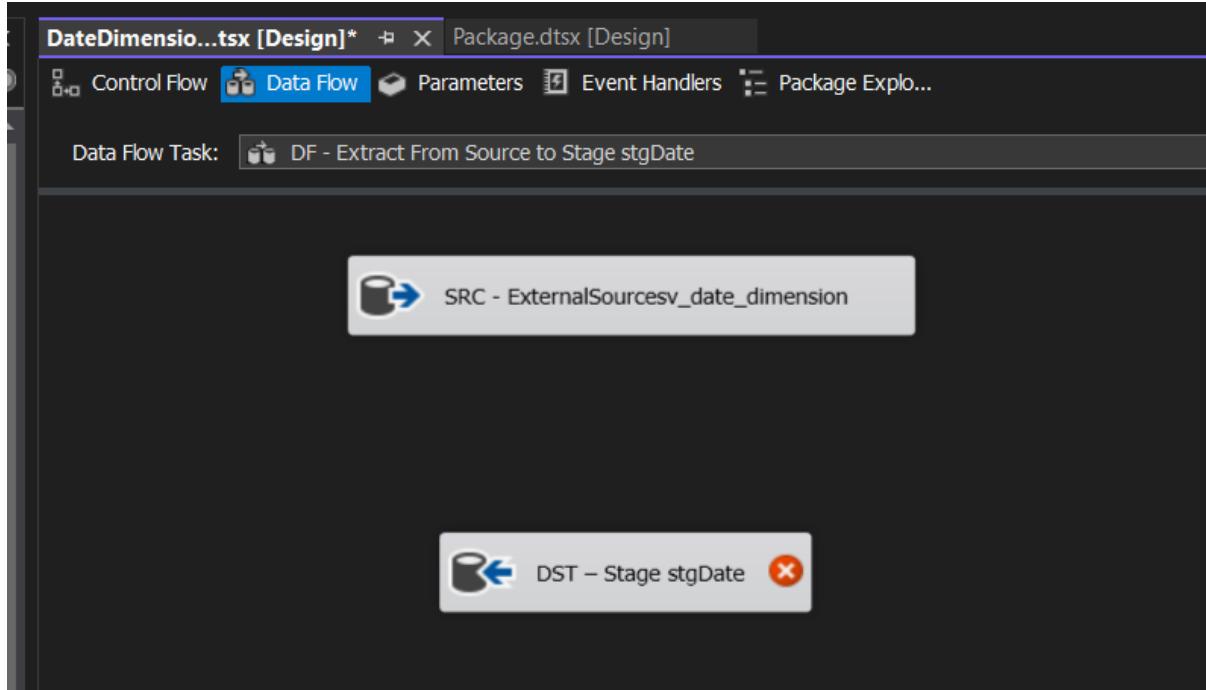


Click **OK**.

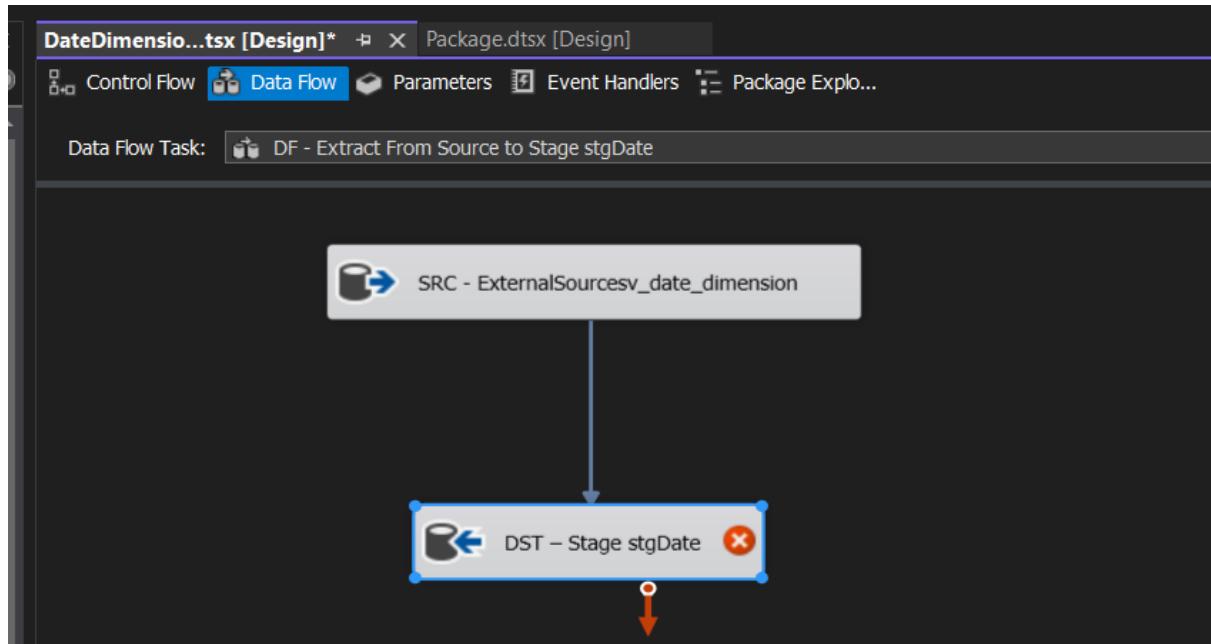
8. Ta có một **destination** như bên dưới:



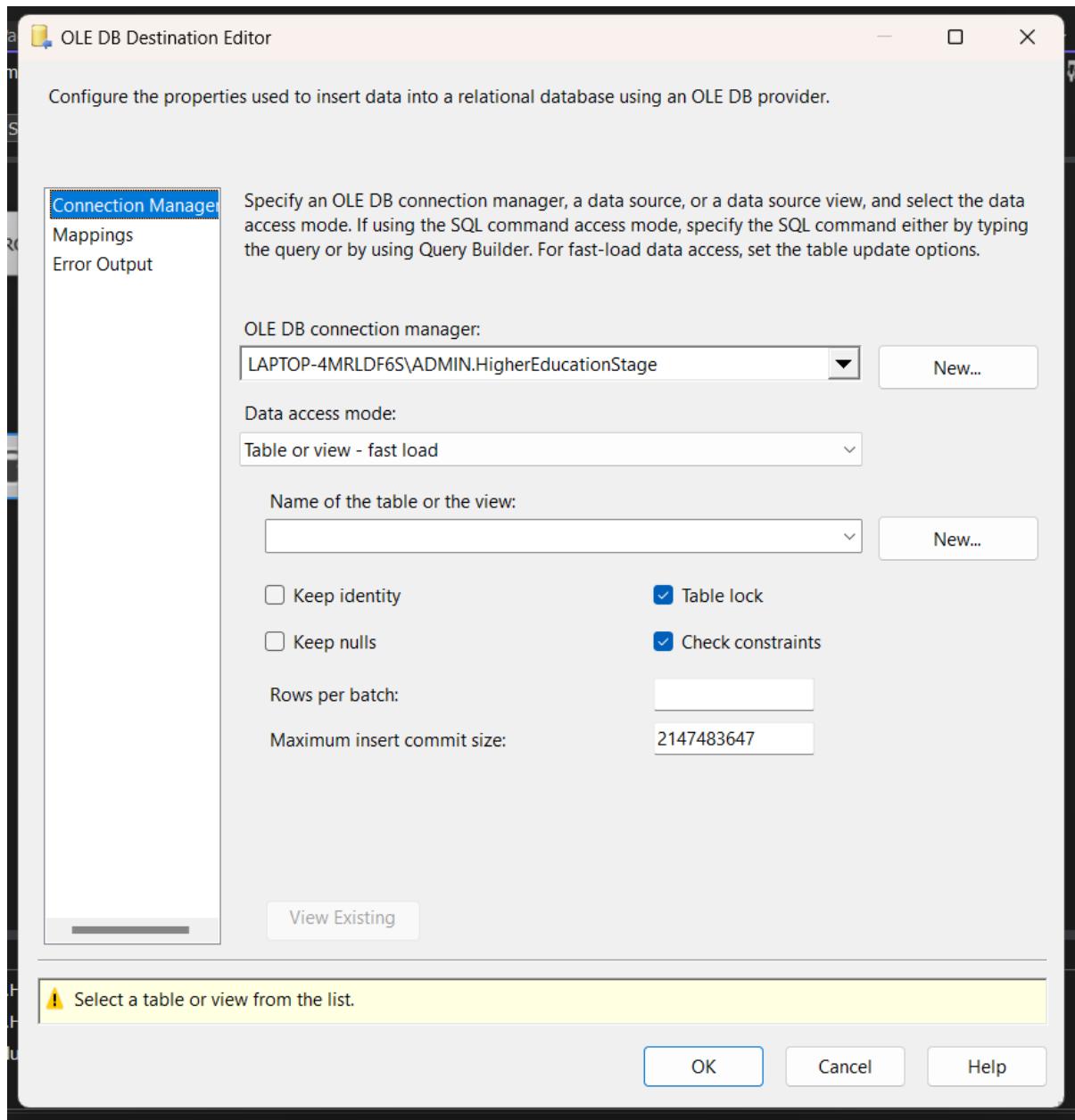
9. Đổi tên lại thành **DST – Stage stgDate**



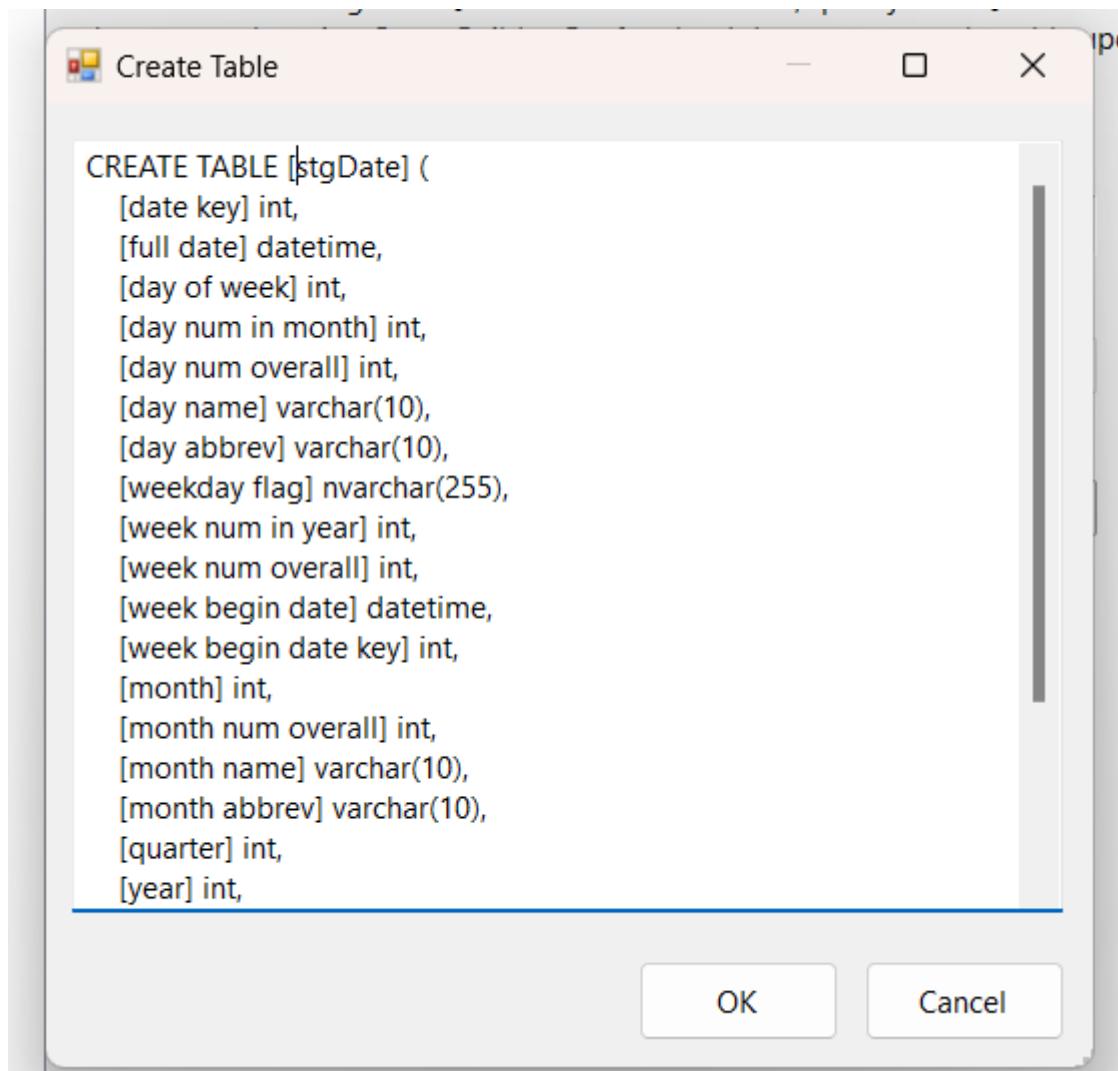
10. Click vào **source** và kéo rê nó vào **destination**.



11. Double click vào destination để cấu hình nó, **OLE DB Destination Editor** sẽ xuất hiện như sau:



Click nút **New...** để tạo bảng trong CSDL stage, ta sẽ thấy window **Create Table** như sau:

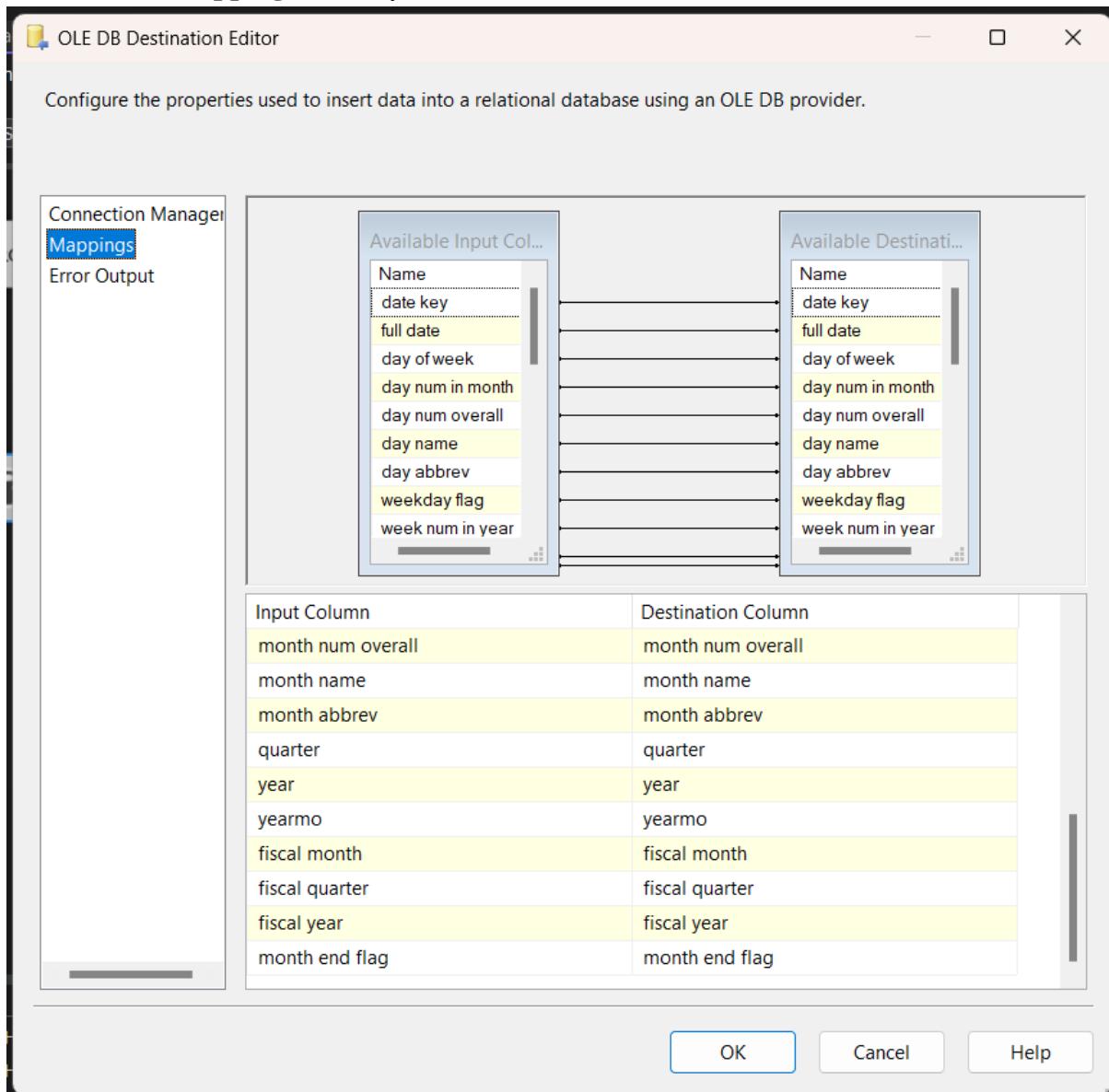


Đổi tên bảng lại thành **stgDate**. (**CREATE TABLE [stgDate]**).

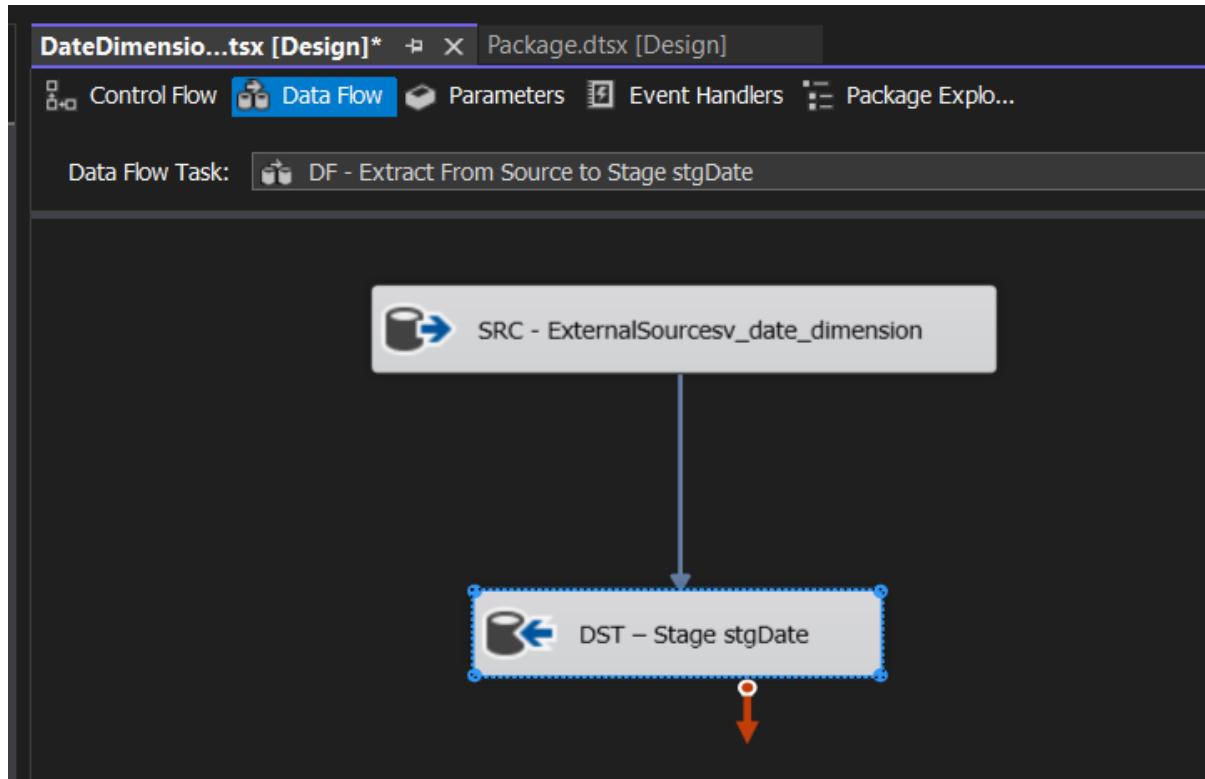
Click OK.

Trở lại màn hình **OLEDB Destination Editor**, tại mục **Name of the table or view** bạn sẽ thấy **[stgDate]**.

Click vào tab **Mapping** ta sẽ thấy như hình bên dưới:



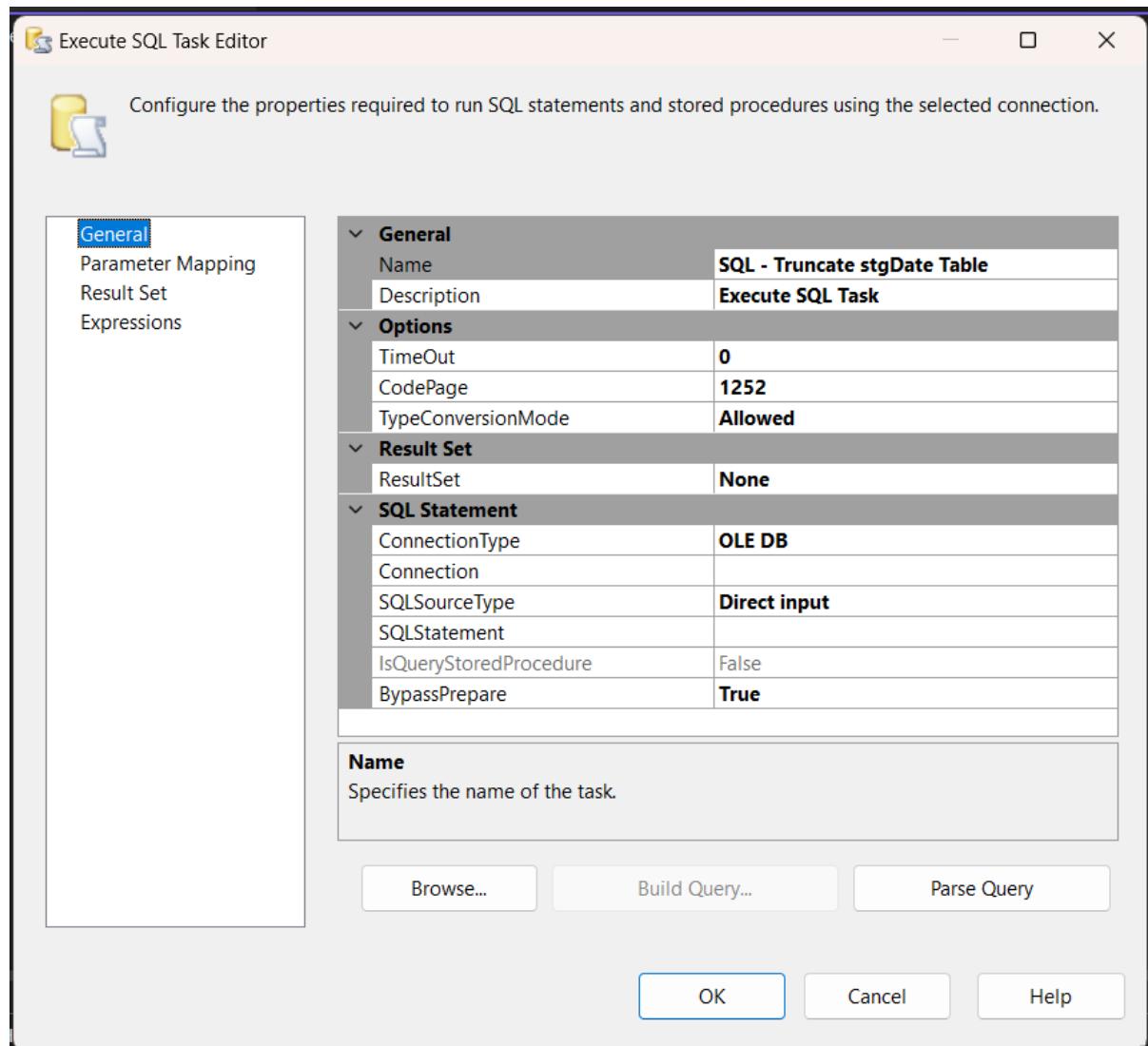
Click **OK** để hoàn tất việc cấu hình. Màn hình data flow design surface sẽ như bên dưới.



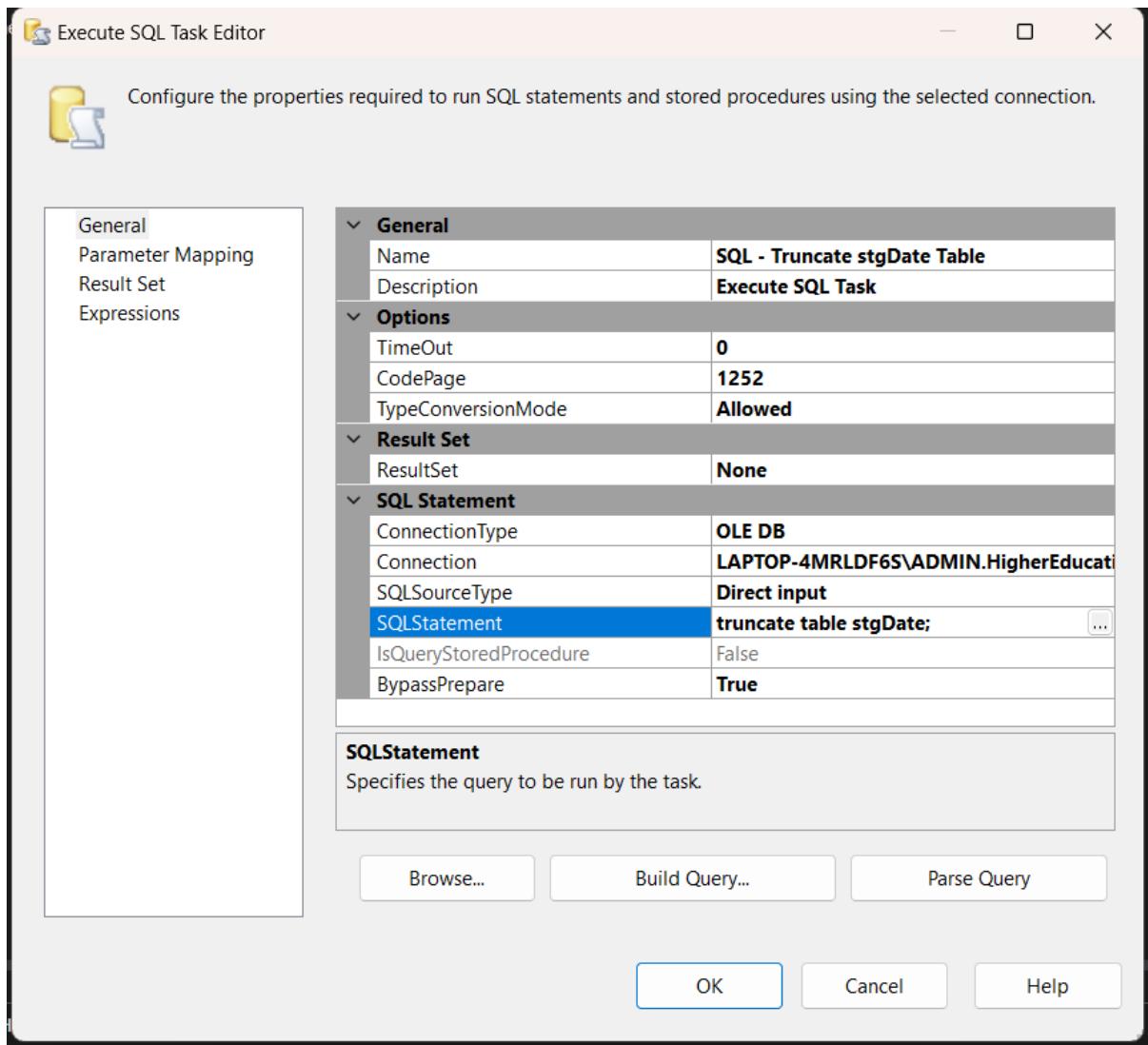
3.3.1.4: Thiết lập truncate table SQL task

Click **Control Flow** tab.

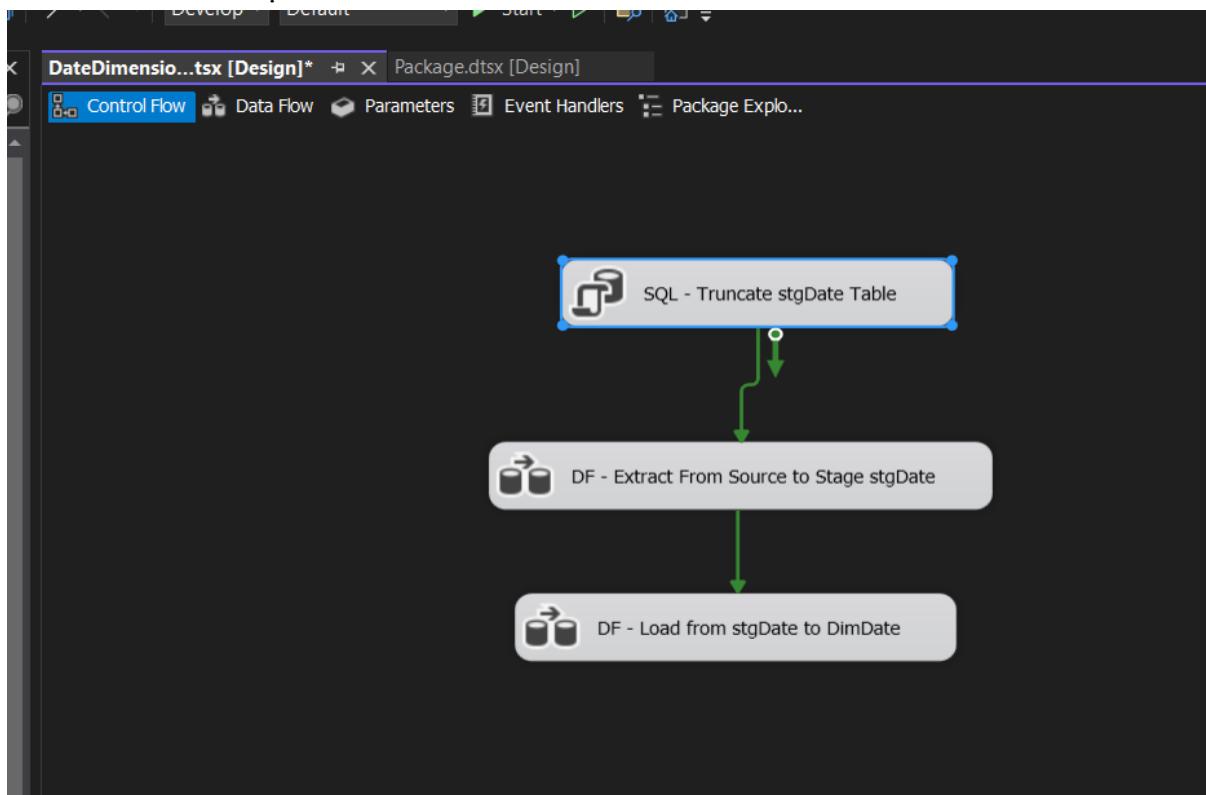
1. Double-click vào **SQL - Truncate stgDate Table** task để cấu hình nó, **Execute SQL Task Editor** sẽ mở ra.



2. Ở mục **SQL Statement**, chọn **Connection** là **LAPTOP-4MRLDF6S_ADMIN.HigherEducationStage** và **SQLStatement** là **truncate table stgDate;**

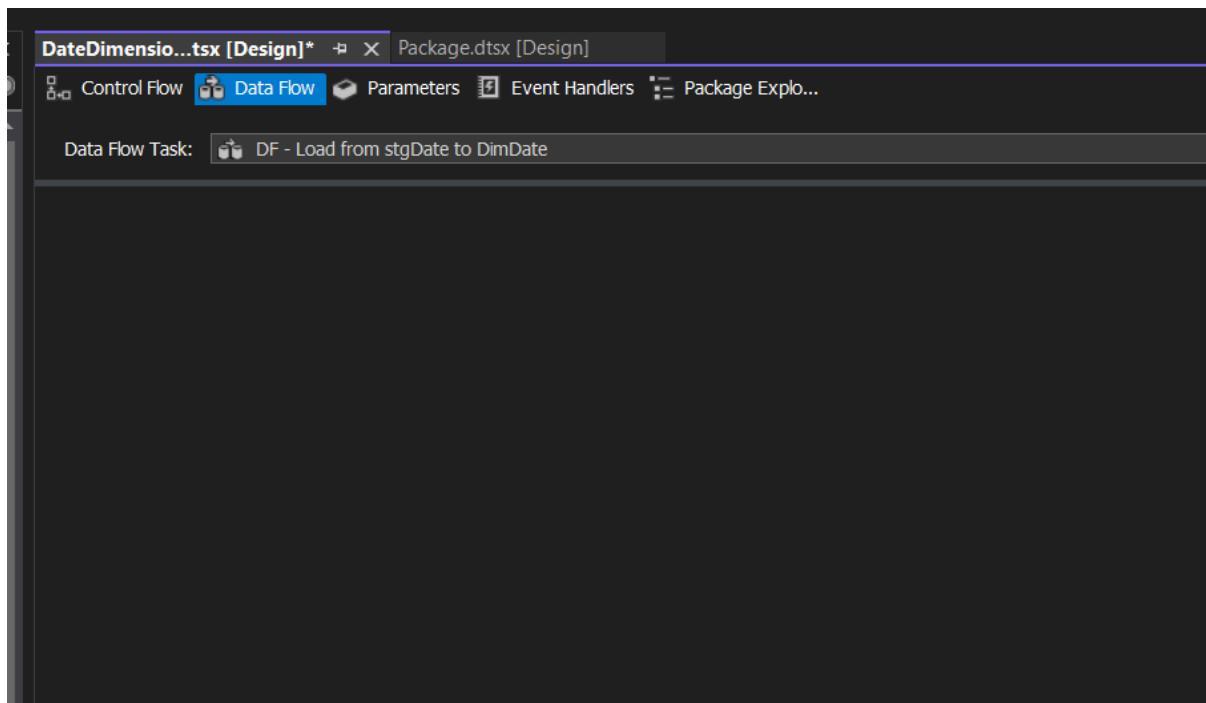


3. Click **OK** và trở lại màn hình **Control Flow**

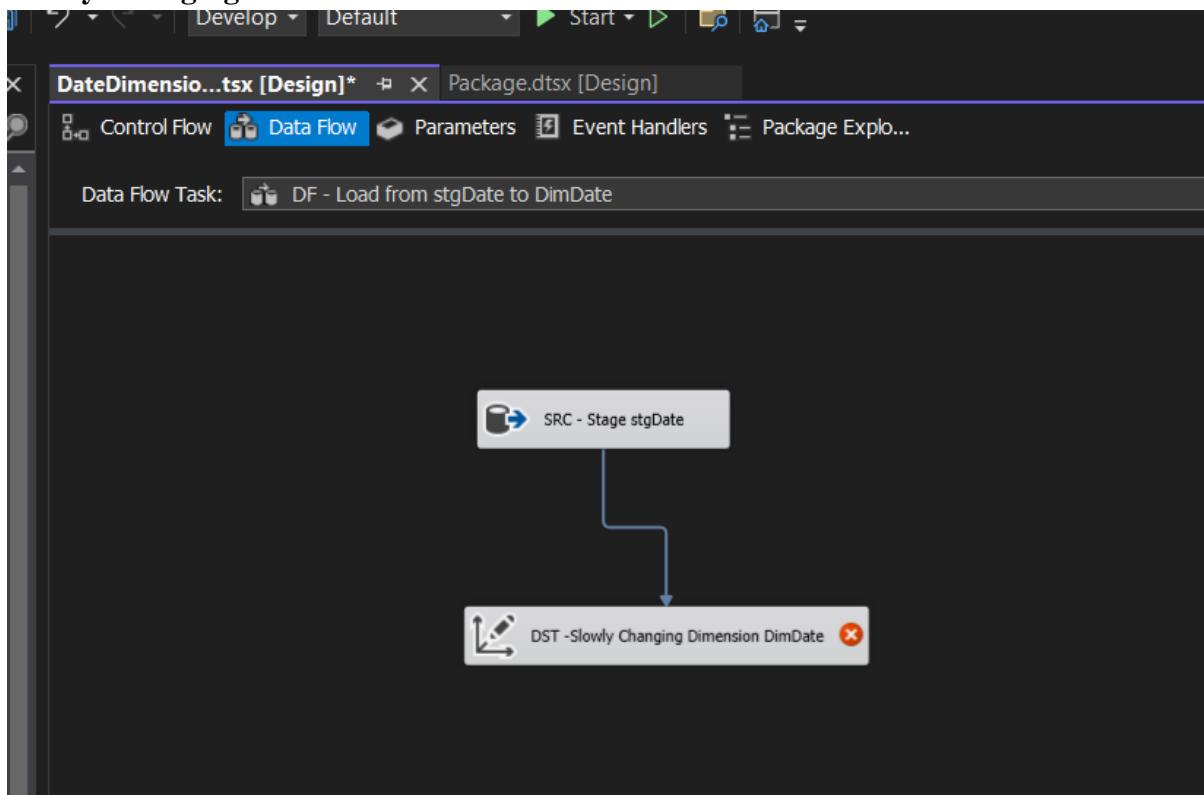


3.3.1.5: Load dữ liệu từ Stage vào Dimension

1. Double-click vào **DF - Load from stgDate to DimDate** task để mở data flow design surface.



2. Kéo rê **Source Assistant** vào design surface, chọn **source type** là **SQL Server** và connection manager là **HigherEducationStage**.
3. Đổi tên source thành **SRC - Stage stgDate**, và double-click vào nó để cấu hình.
4. Từ **OLE DB Source Editor** chọn **[dbo].[stgDate]** table ở mục **name of the table or view**. Click OK để hoàn tất việc cấu hình.
5. Kéo rê **Slowly Changing Dimension** vào design surface. Đặt tên nó lại thành **DST - Slowly Changing Dimension DimDate** và connect **SRC -> DST**.



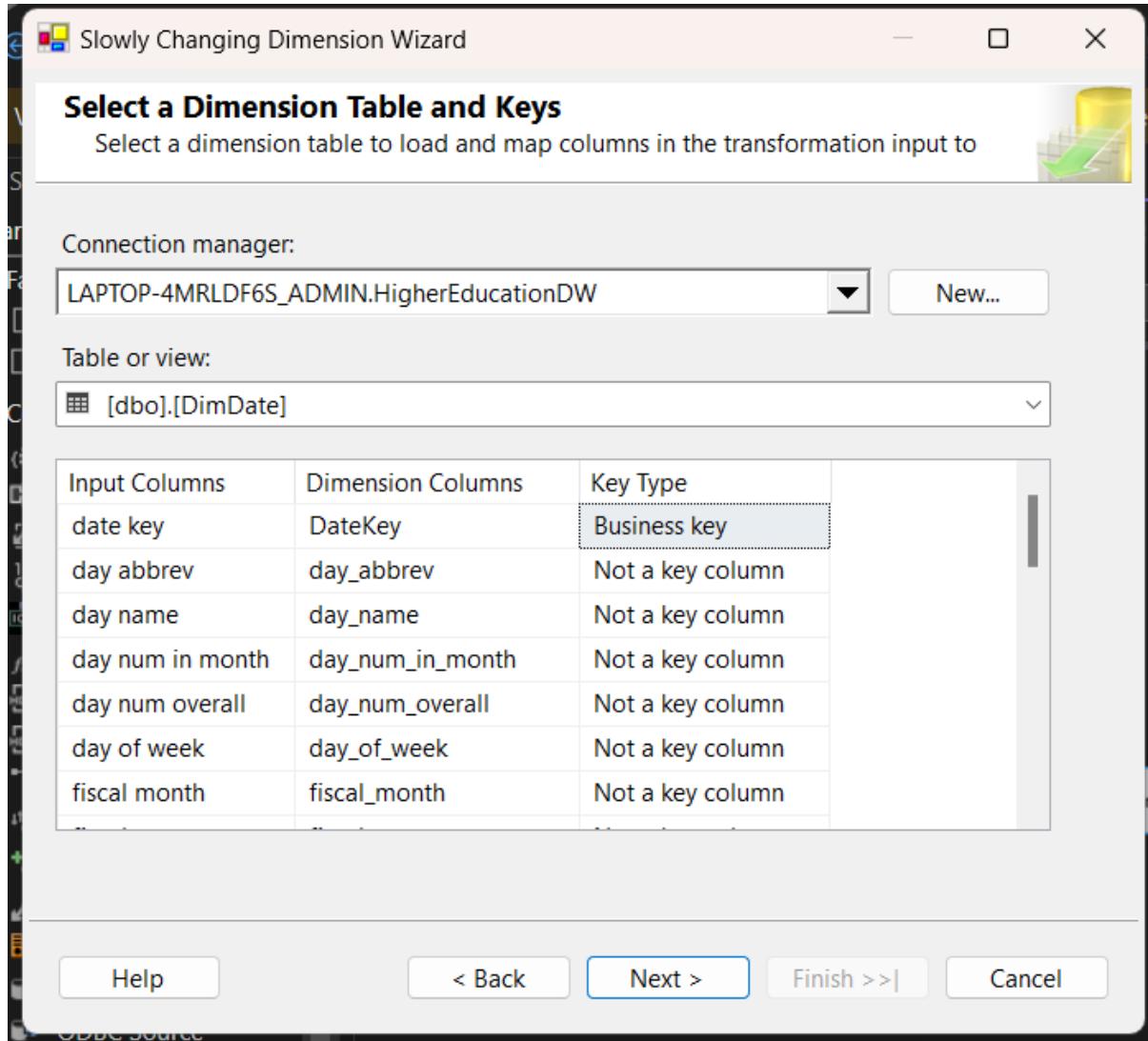
6. Double click vào **DST - Slowly Changing Dimension DimDate** để cấu hình.

7. Slowly Changing Dimension Wizard xuất hiện



Click **Next**.

8. Chọn bảng **DimDate** ở CSDL **HigherEducationDW**. Ở mục **connection manager** chọn CSDL **HigherEducationDW**.
9. Ở mục **Table or view** chọn bảng [**DimDate**].
10. Chọn **DateKey** là **business key**. Tất cả các cột khác là “**Not a key column**”.



Click **Next**.

11. Cấu hình SCD type cho dimension. Các lựa chọn là:

- **Fixed** **Không thay đổi**
- **Changing** **Type 1 (Update)**
- **Historical** **Type 2 (thêm dòng mới, dòng cũ inactive)**

Slowly Changing Dimension Wizard

Slowly Changing Dimension Columns

Manage the changes to column data in your slowly changing dimensions by setting

Fixed Attribute

Select this type when the value in a column should not change. Changes are treated as errors.

Changing Attribute

Select this type when changed values should overwrite existing values. This is a Type 1 change.

Historical Attribute

Select this type when changes in column values are saved in new records. Previous values are saved in records marked as outdated. This is a Type 2 change.

Select a change type for slowly changing dimension columns:

Dimension Columns	Change Type
day_abbrev	Changing a...
day_name	Changing a...
day_num_in_month	Changing a...
day_num_overall	Changing a...
day_of_week	Changing a...
fiscal_month	Changing a...
fiscal_quarter	Changing a...
fiscal_year	Changing a...
full_date	Changing a...
last_day_in_month...	Changing a...
month	Changing a...
month_abbrev	Changing a...
month_name	Changing a...
month_num_overall	Changing a...
quarter	Changing a...
week_begin_date	Changing a...
week_begin_date_...	Changing a...
week_num_in_year	Changing a...
week_num_overall	Changing a...
weekday_flag	Changing a...
year	Changing a...
yearmo	Changing a...

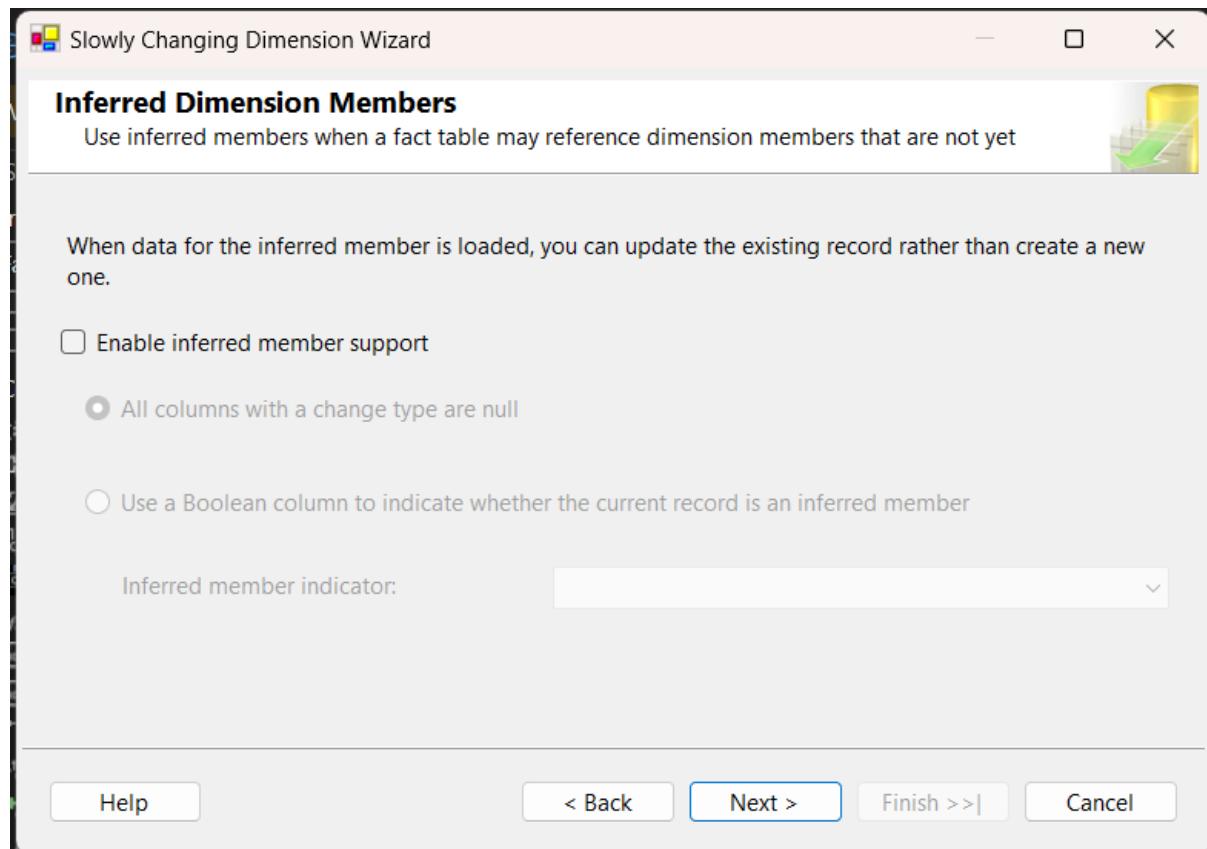
Help < Back Next > Finish >> Cancel

Cáu hình tất cả non-business key là changing attribute.

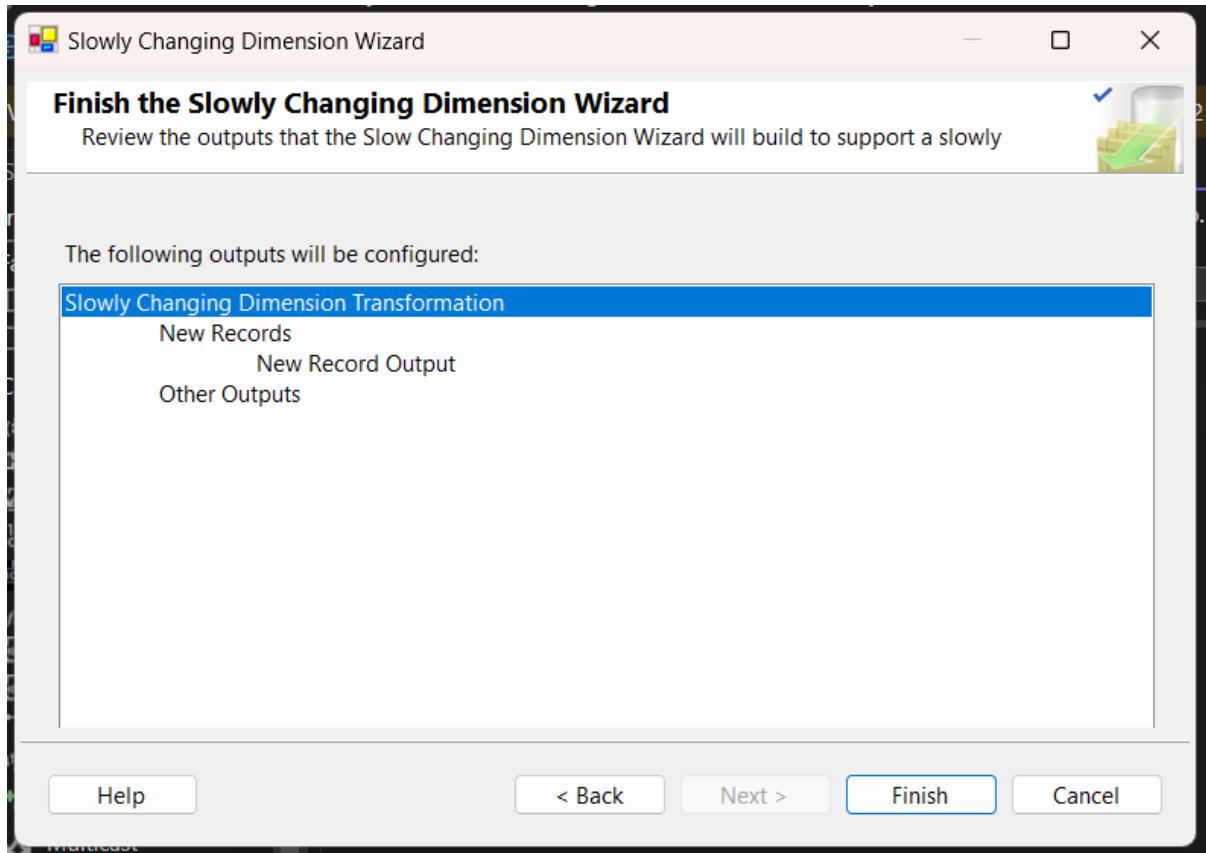
12. Click Next, xuất hiện màn hình **Fixed and Changing Attribute Options**. Click Next.

13. Xuất hiện màn hình **Inferred Dimension Members**. Bỏ chọn check box và click

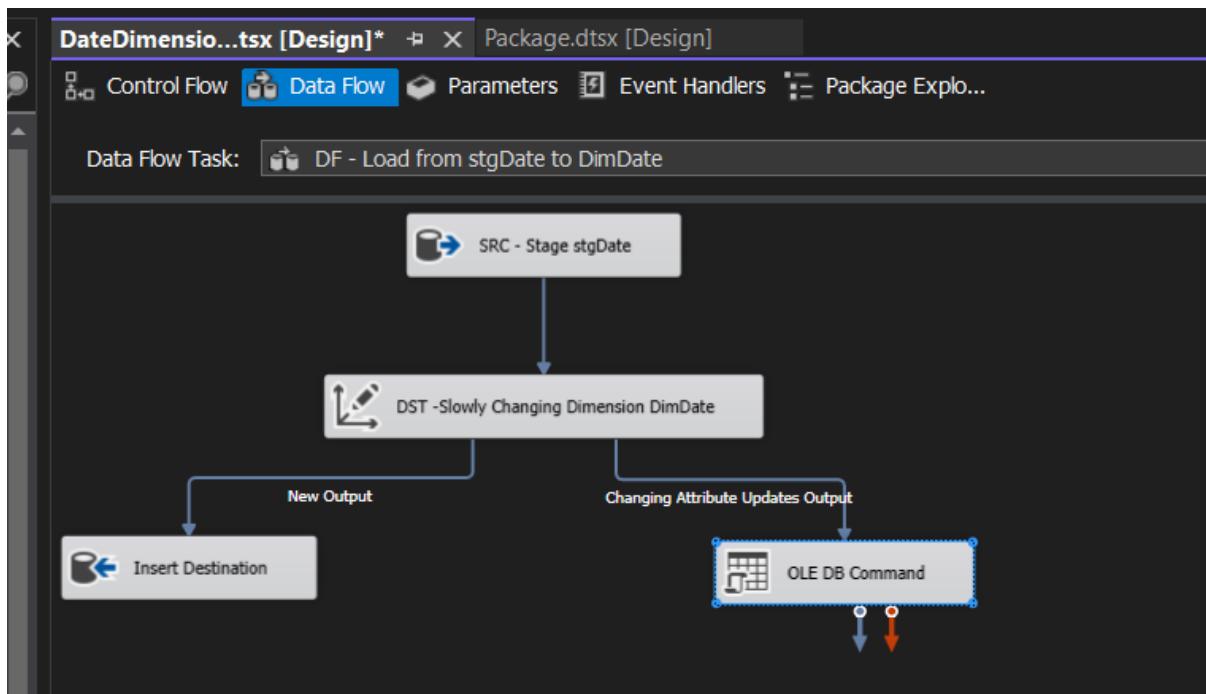
Next.



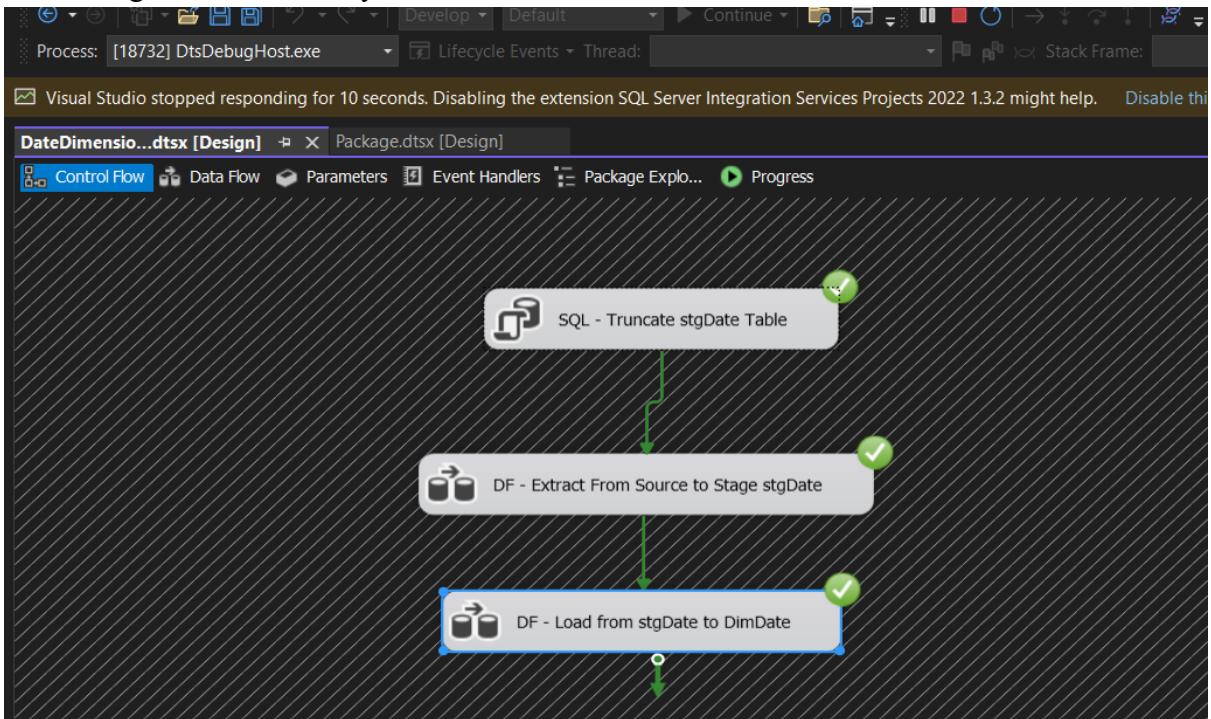
14. Xuất hiện màn hình Finish



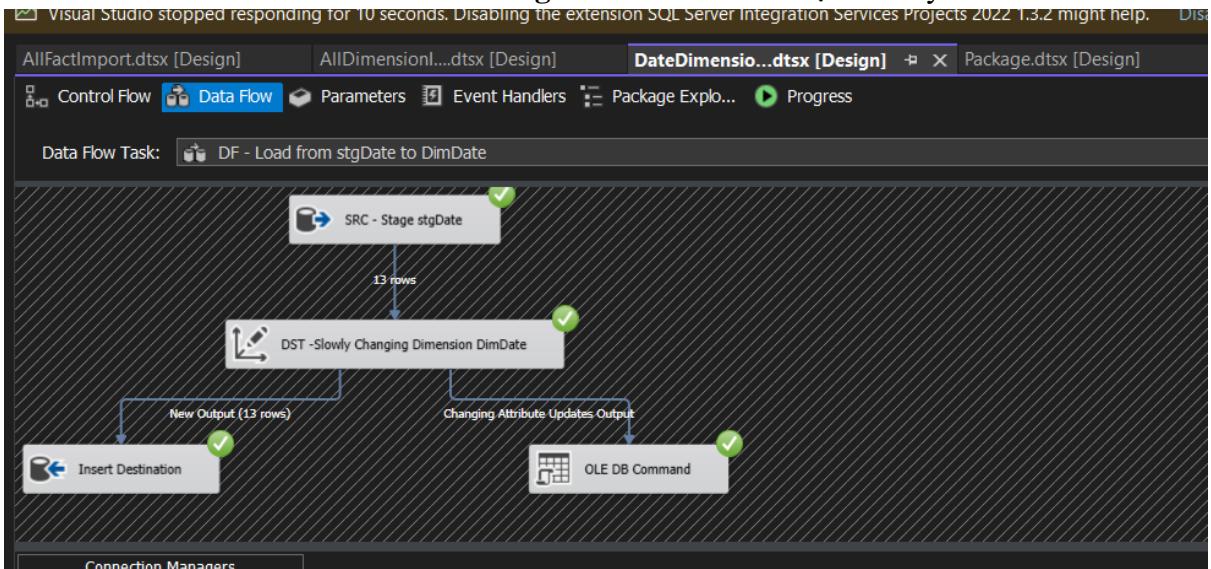
Click **Finish**.



15. Bấm **F5** để chạy. Nếu xảy ra lỗi, cố gắng đọc hiểu thông báo lỗi để tìm cách sửa. Nếu không có lỗi bạn sẽ thấy màn hình như sau:

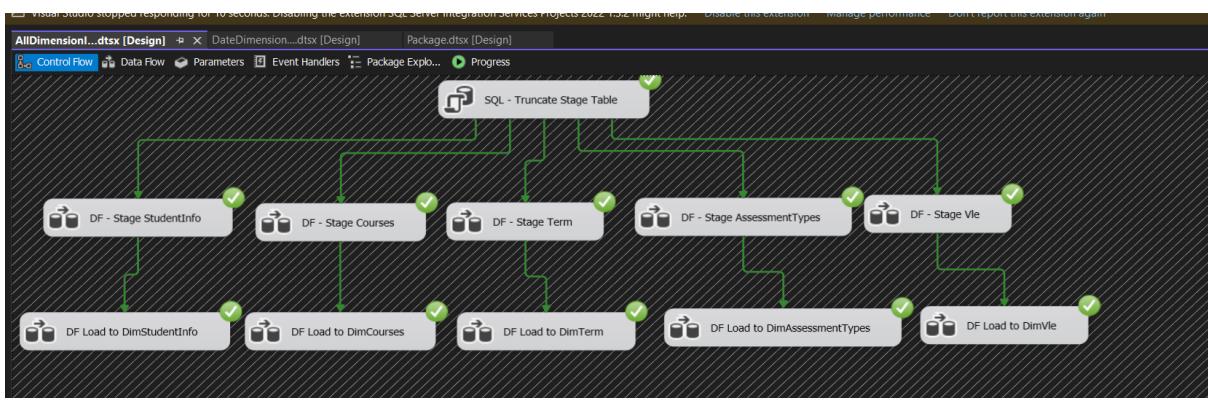


16. Double click vào **DF - Load From stgDate To DimDate** bạn sẽ thấy màn hình sau:



17. Nhấn **[Shift] + [F5]** để kết thúc thực thi.

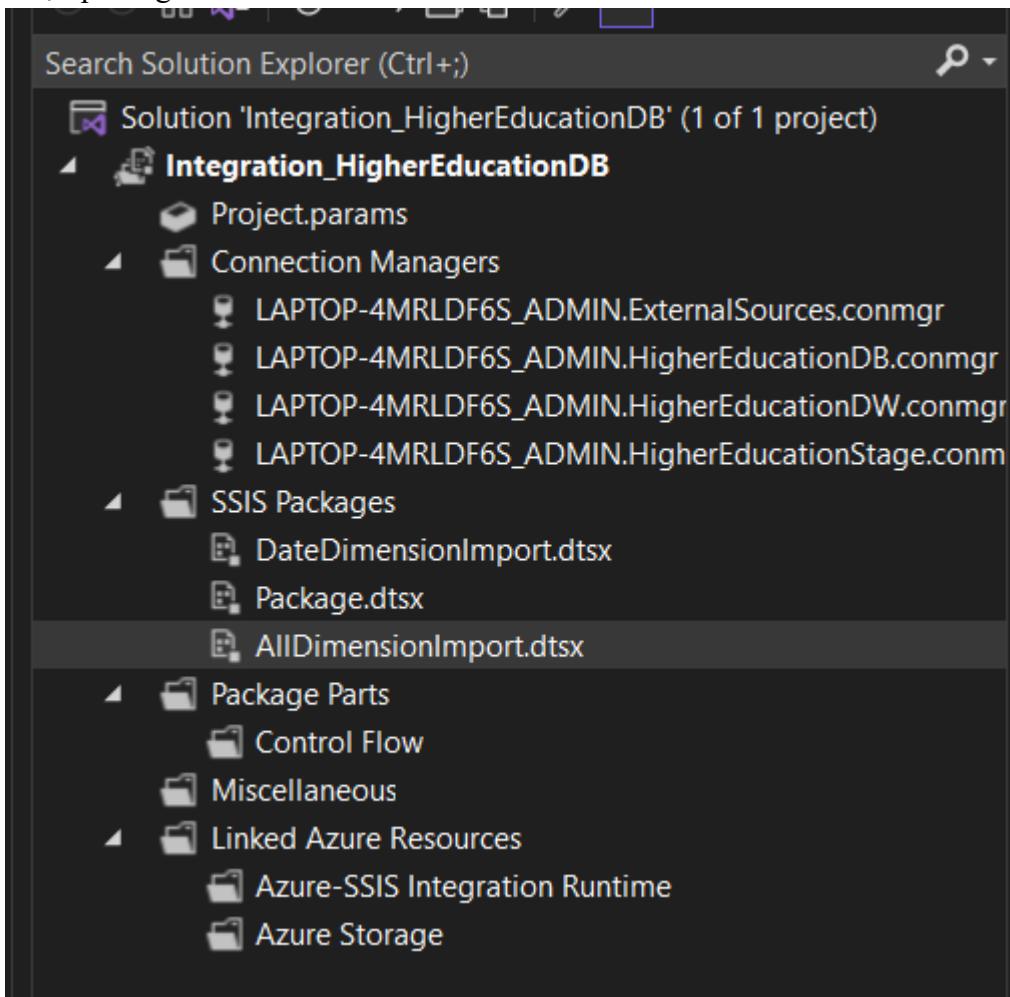
Thực hiện tương tự các bước trên với các bảng Dimension còn lại.



3.3.2. StudentInfo Dimension

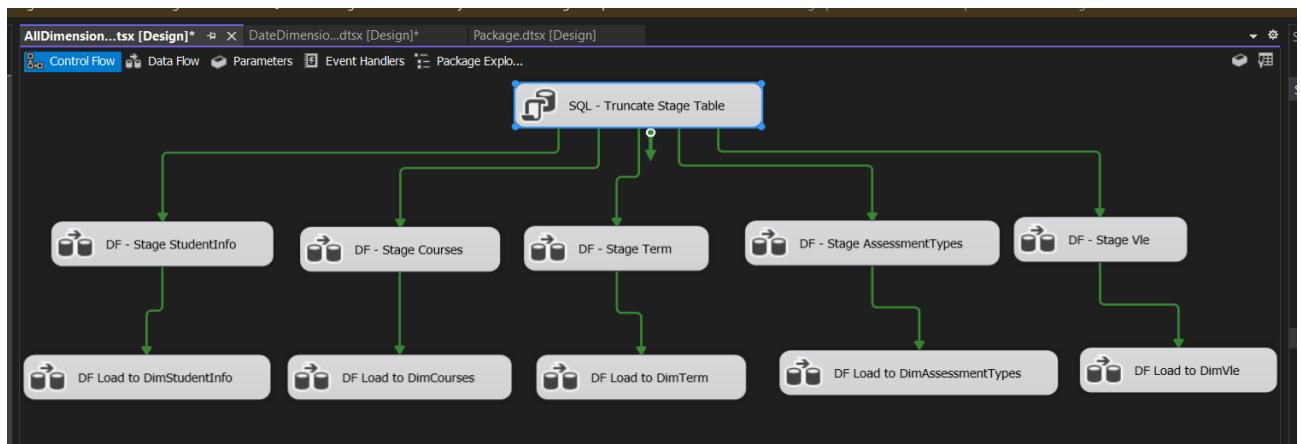
3.3.2.1. Tạo Package

- Tạo package mới và đổi tên như hình bên dưới.

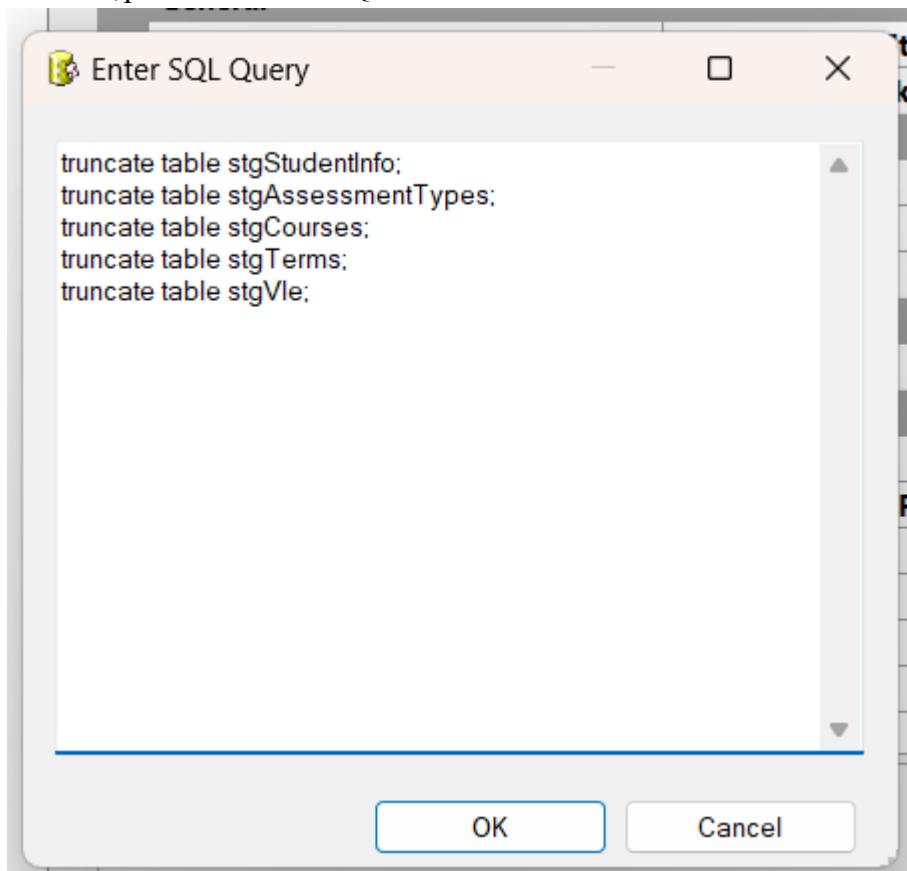


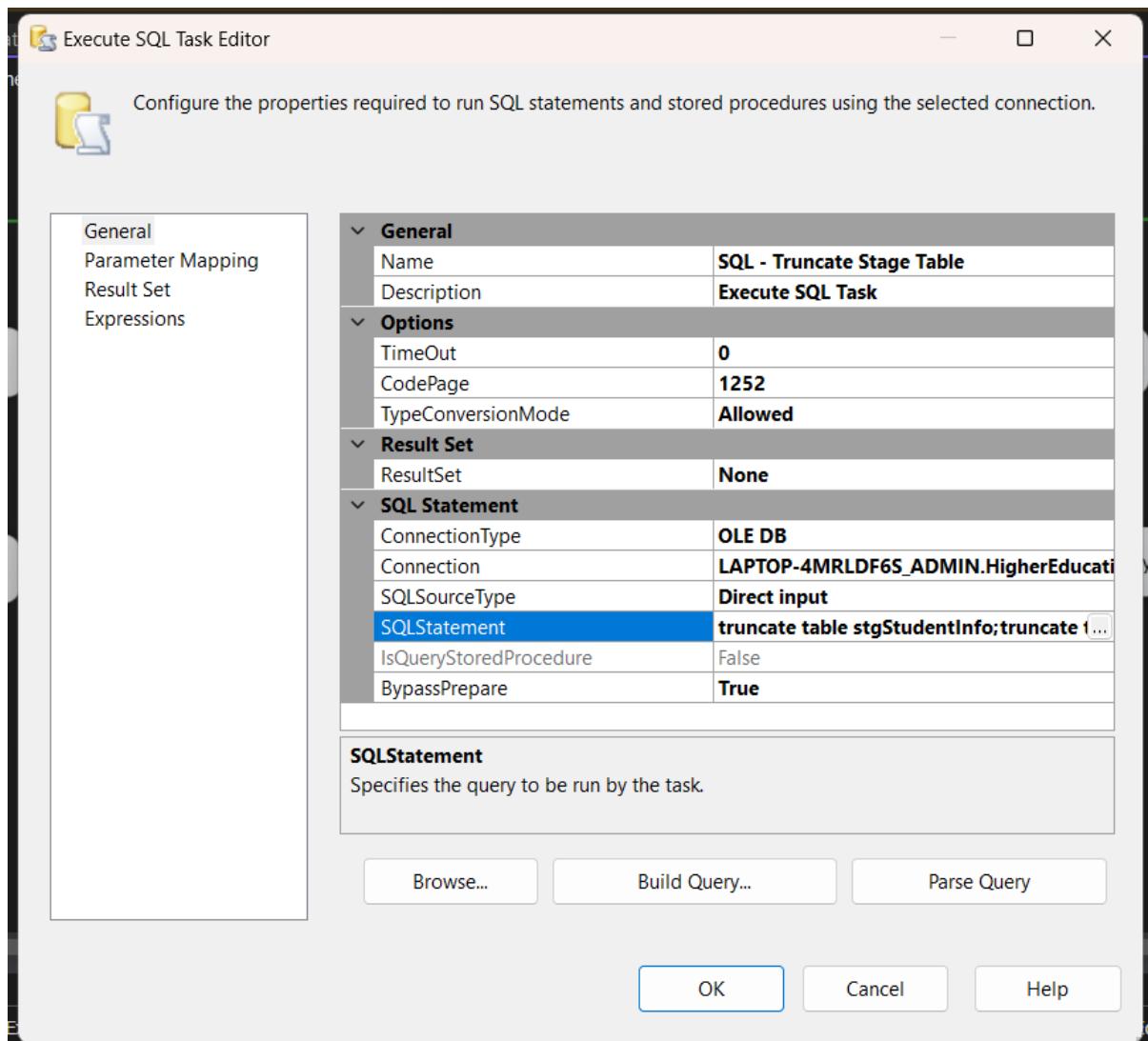
3.3.2.2. Thiết lập Control Flow

- Tạo và đặt tên tương ứng cho các task, tạo các connect bằng cách kéo mũi tên vào các task phù hợp.



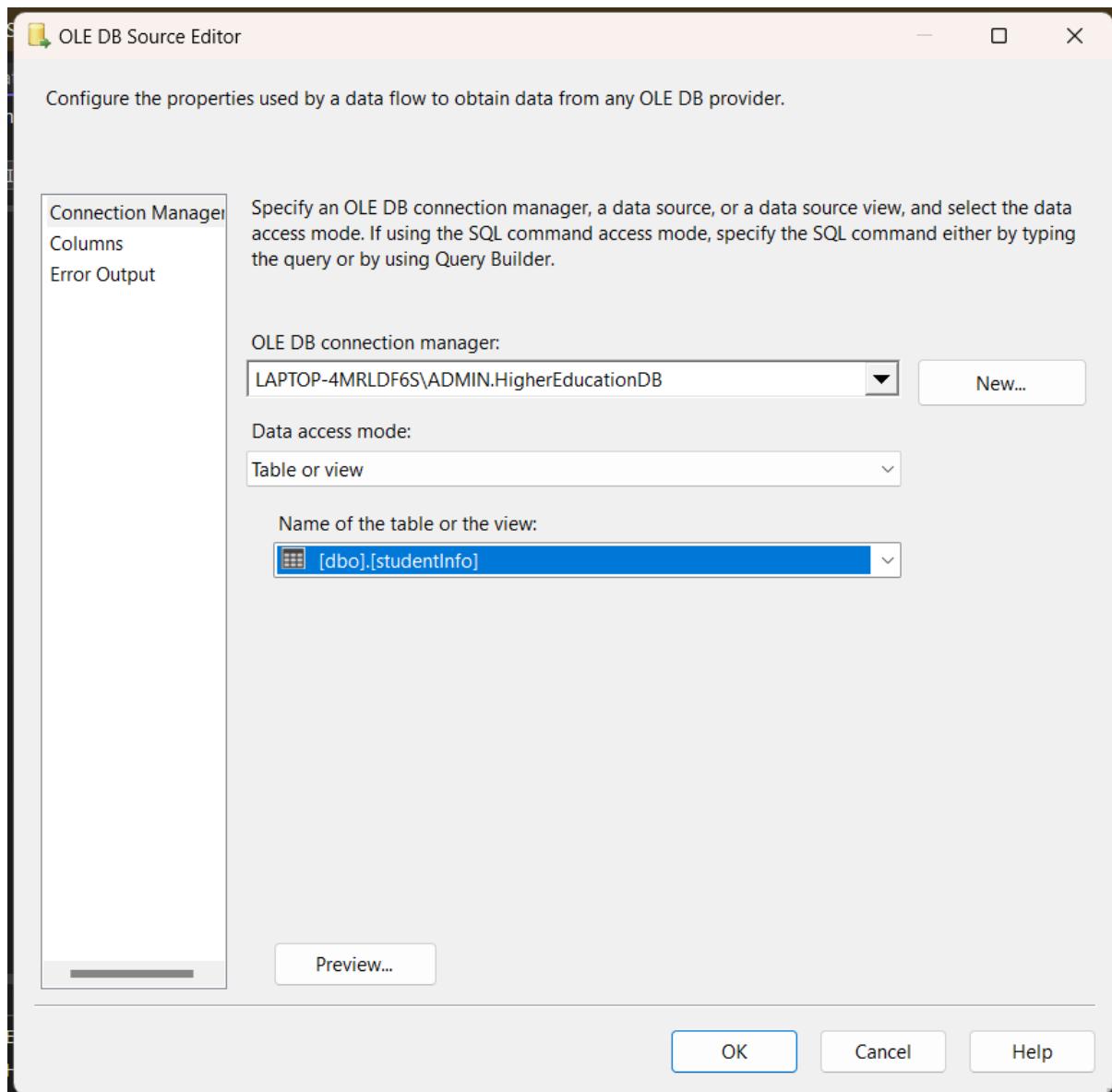
- Thiết lập truncate table SQL task



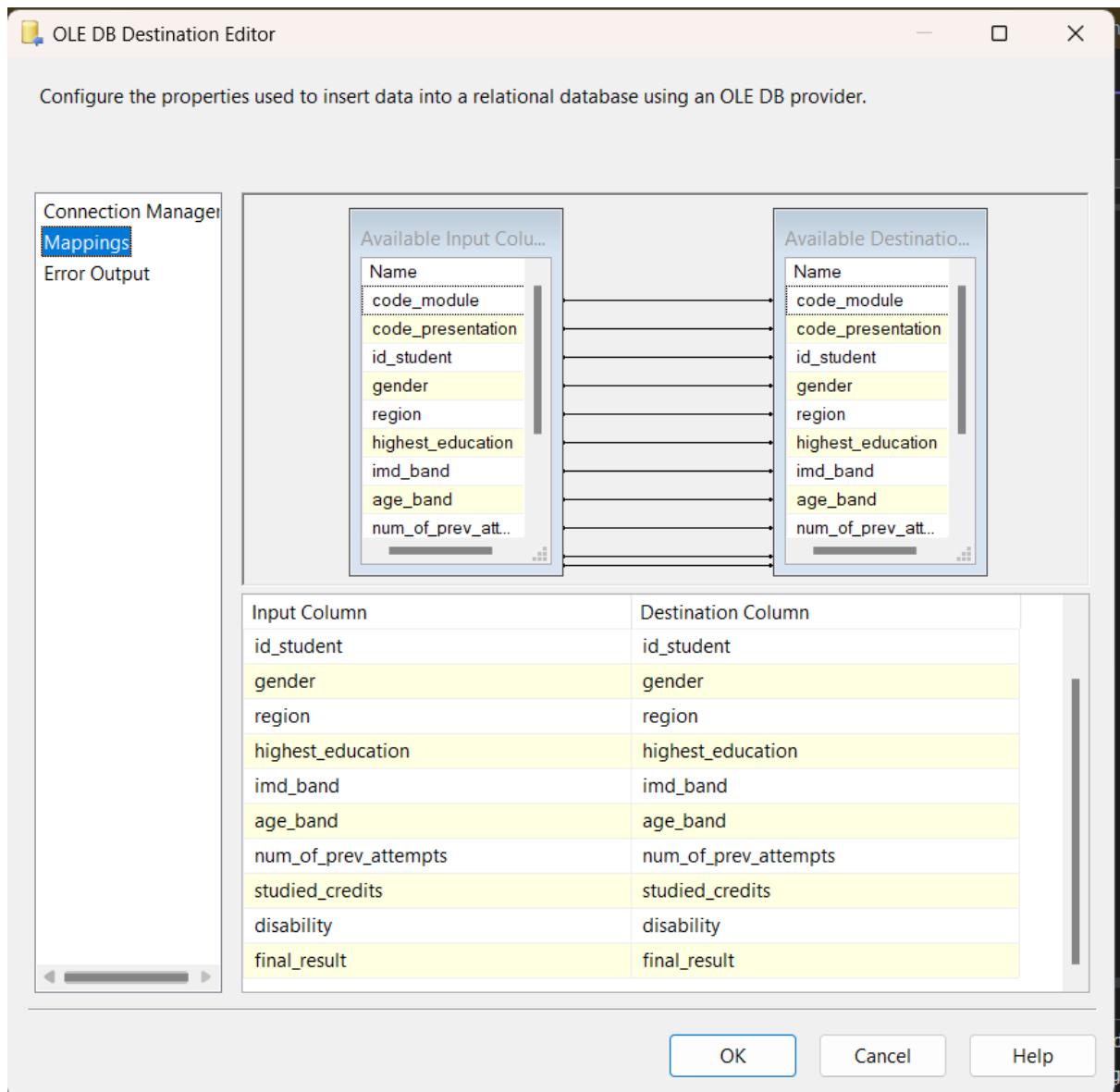


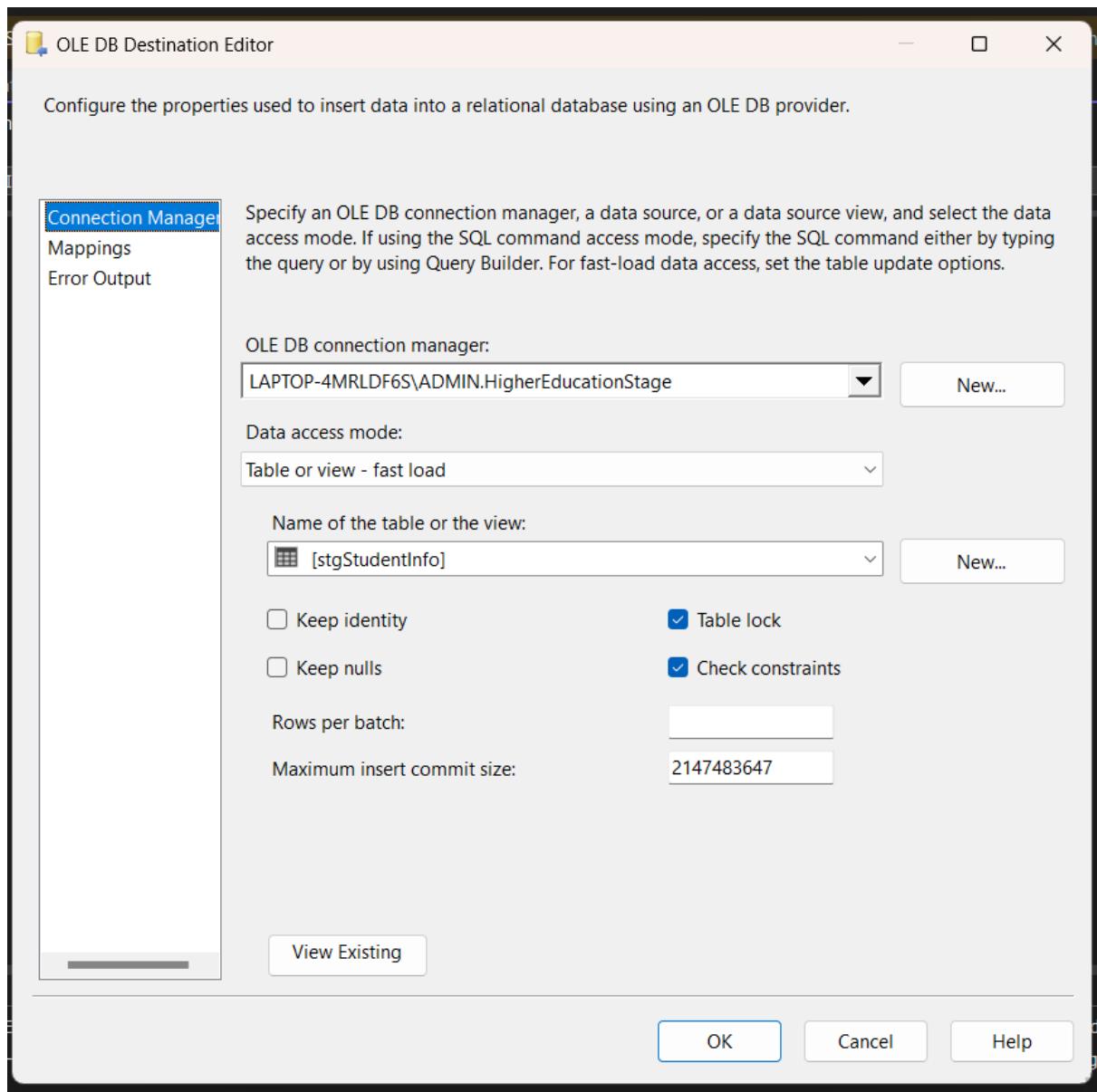
3.3.2.3. Load dữ liệu từ Source vào Stage

- Double-click **DF - Stage StudentInfo** task để mở nó trong **Data Flow design surface**:
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:
 - + Source Assistant:

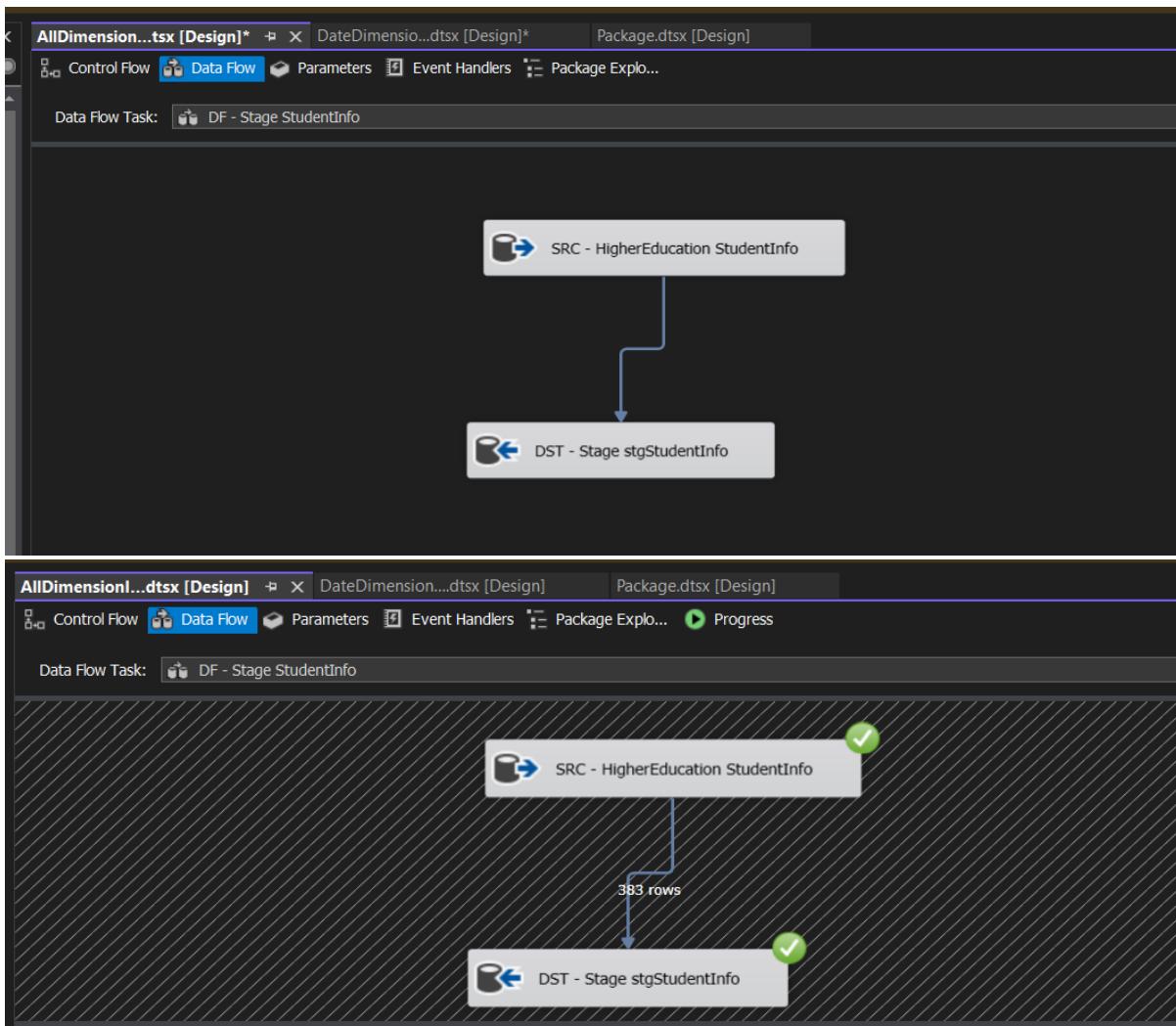


+ Destination Assistant:



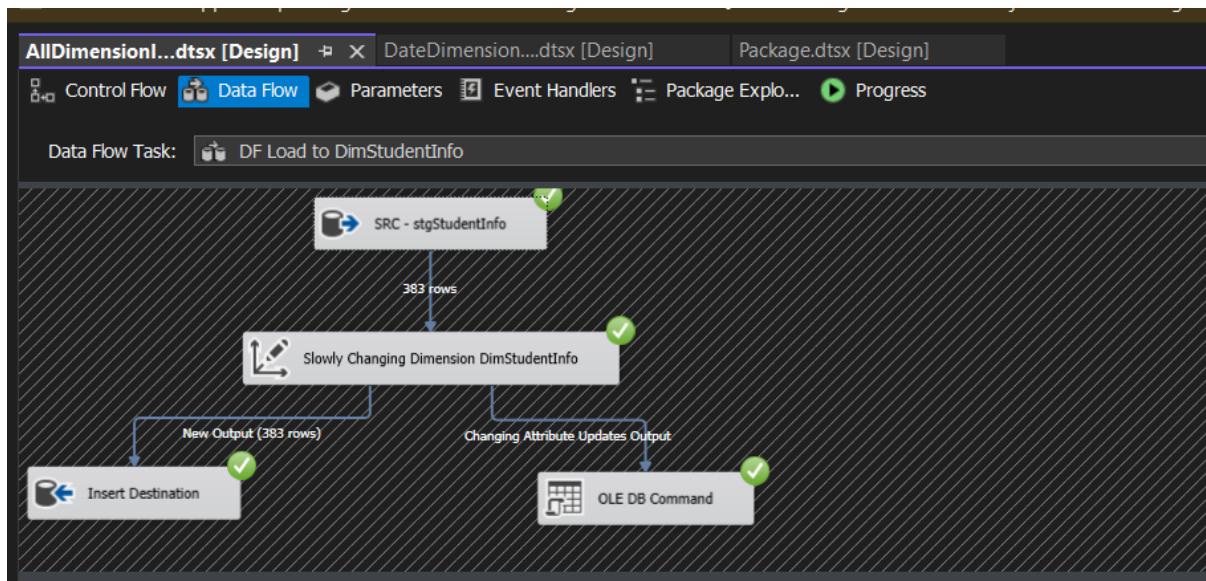


+ Sau khi cấu hình xong ta được:



3.3.2.4. Load dữ liệu từ Stage vào Dimension

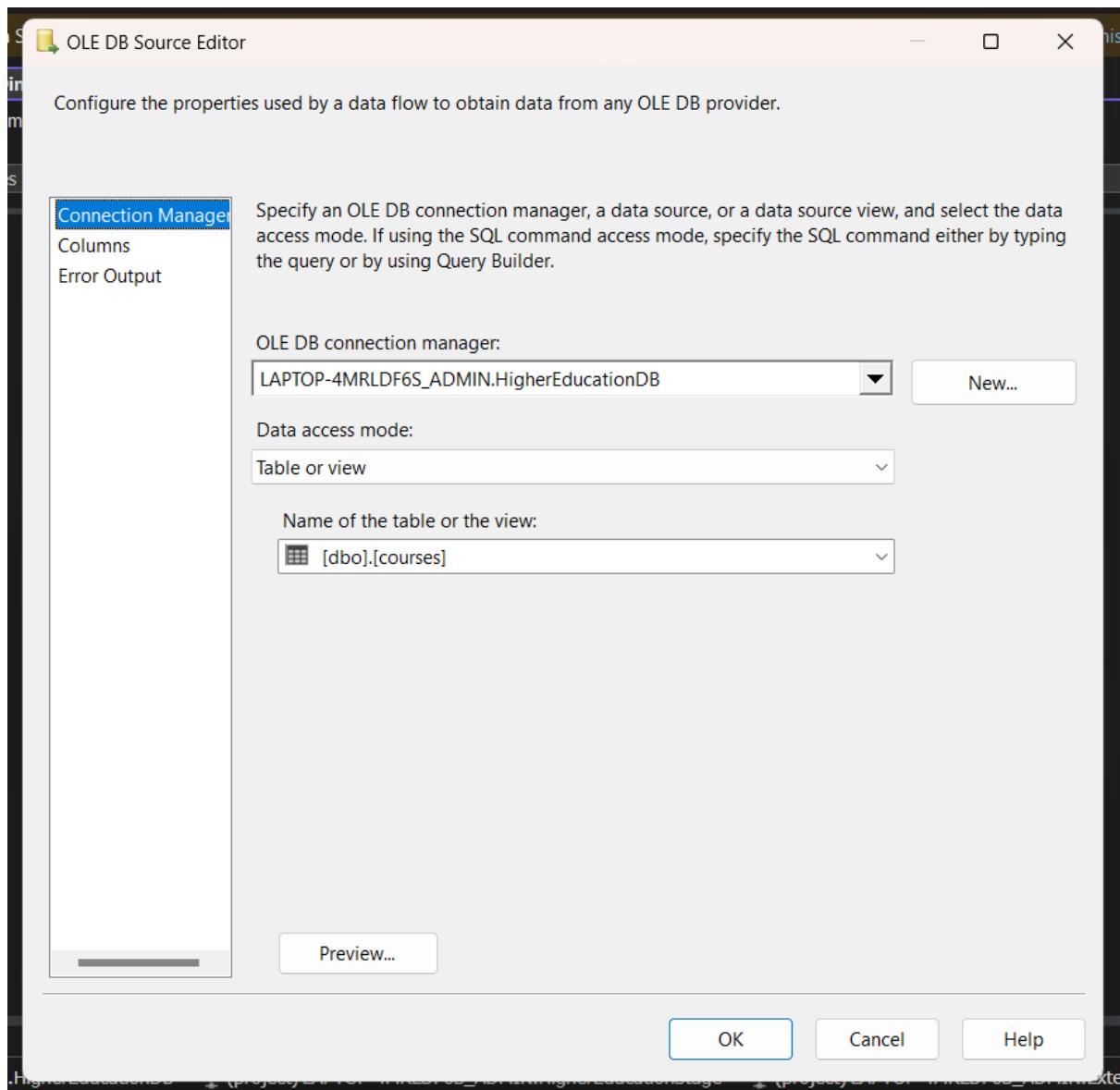
- Double-click vào **DF - Load to DimStudentInfo** task để mở data flow design Surface và thực hiện các bước cấu hình tương tự mục 3.3.



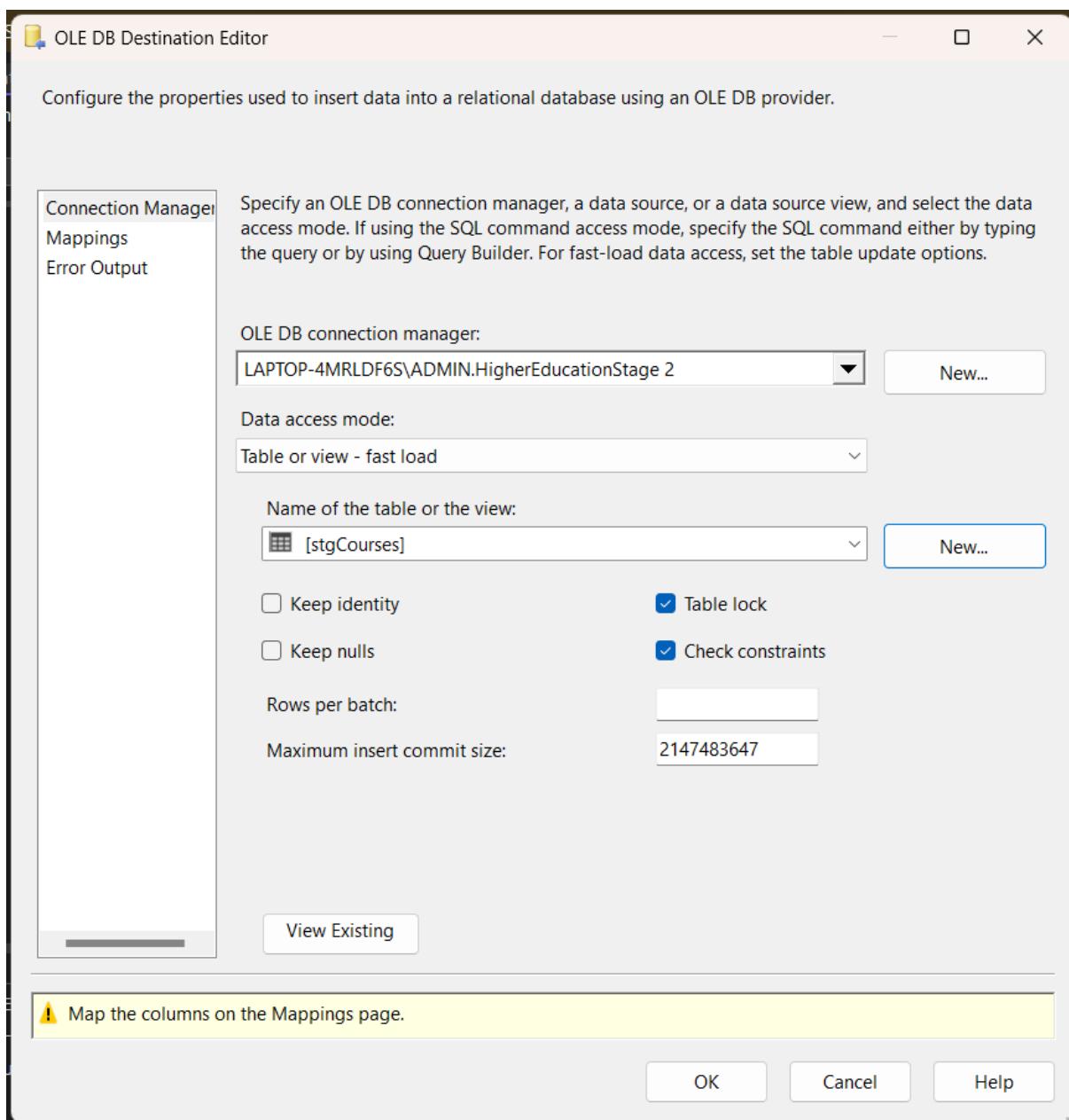
3.3.3. Courses Dimension

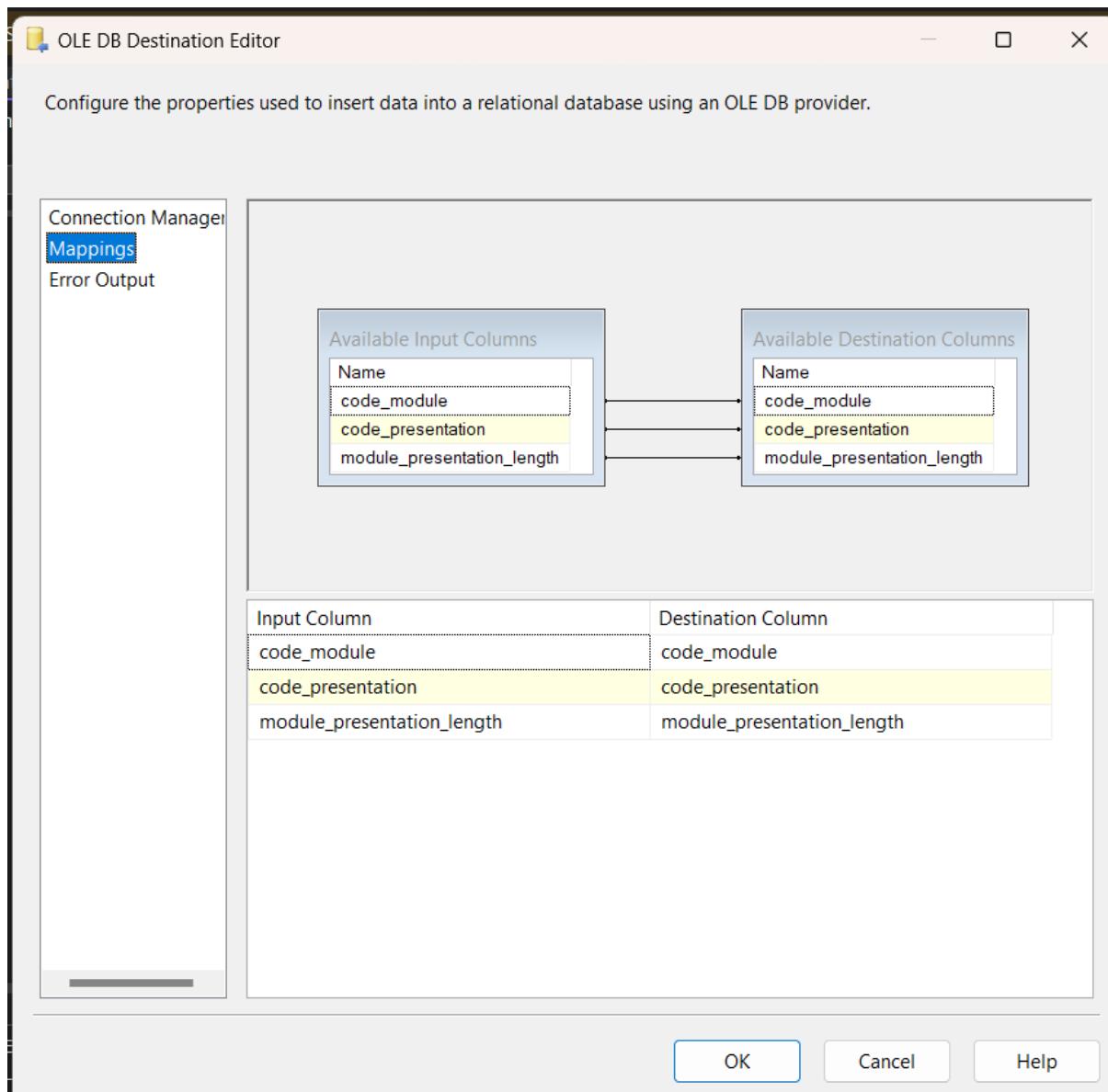
3.3.3.1. Load dữ liệu từ Source vào Stage

- Double-click **DF - Stage Courses** task để mở nó trong **Data Flow design surface**:
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:
- + Source Assistant:

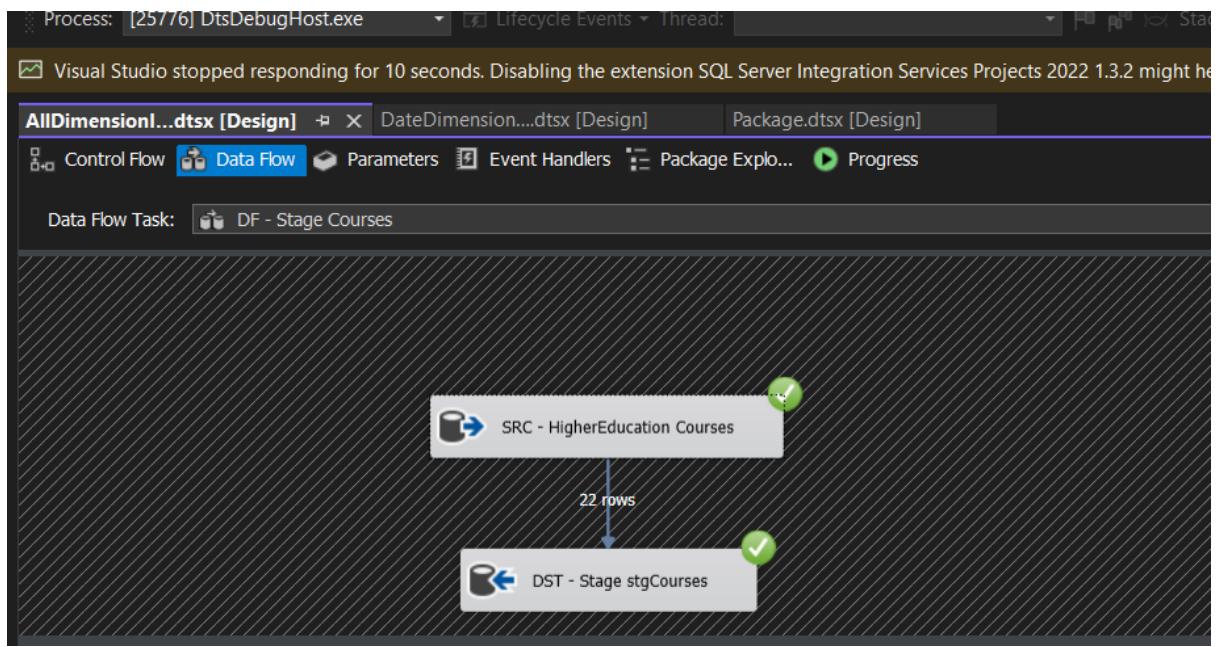


+ Destination Assistant:



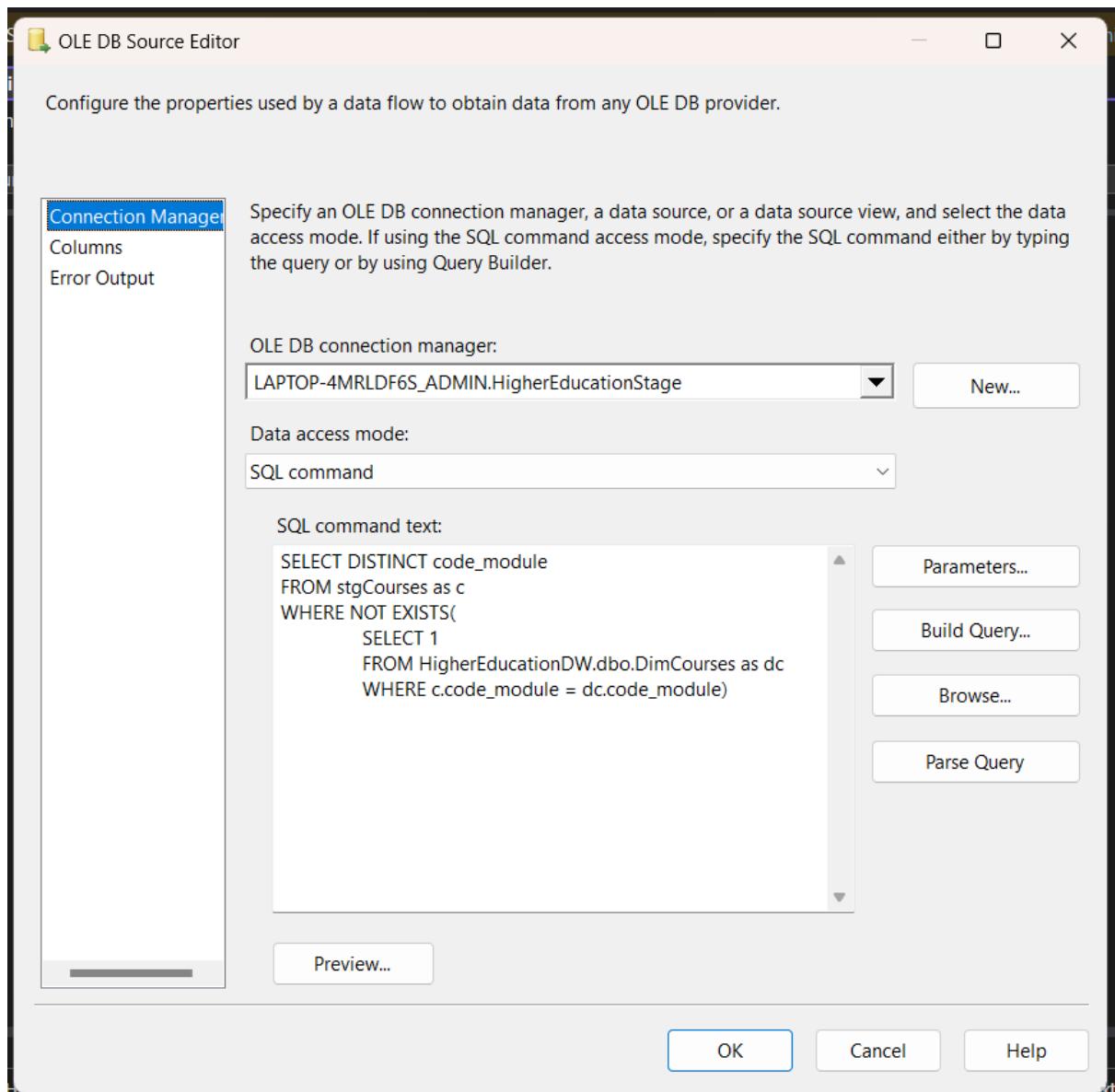


+ Sau khi cấu hình xong ta được:

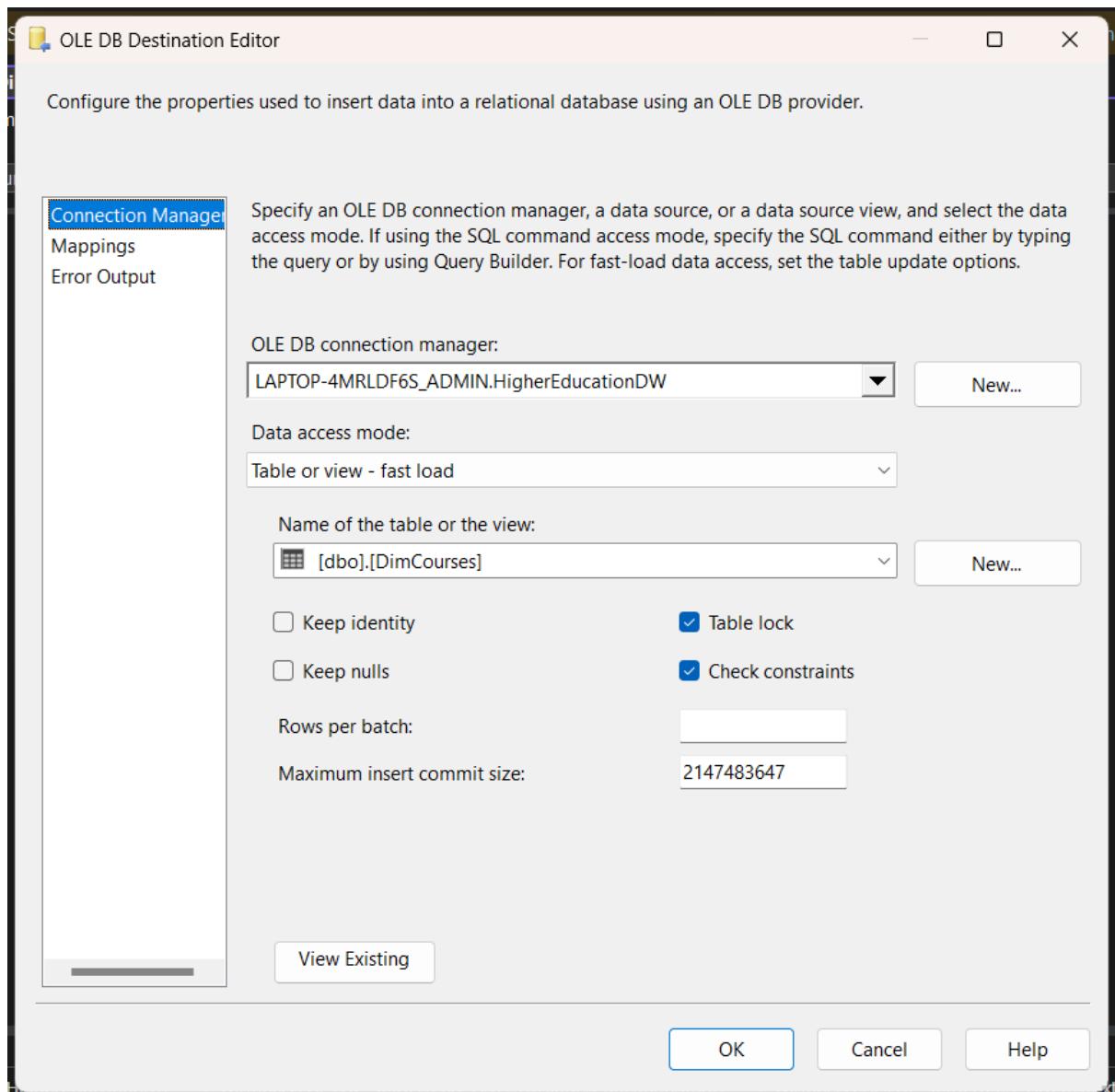


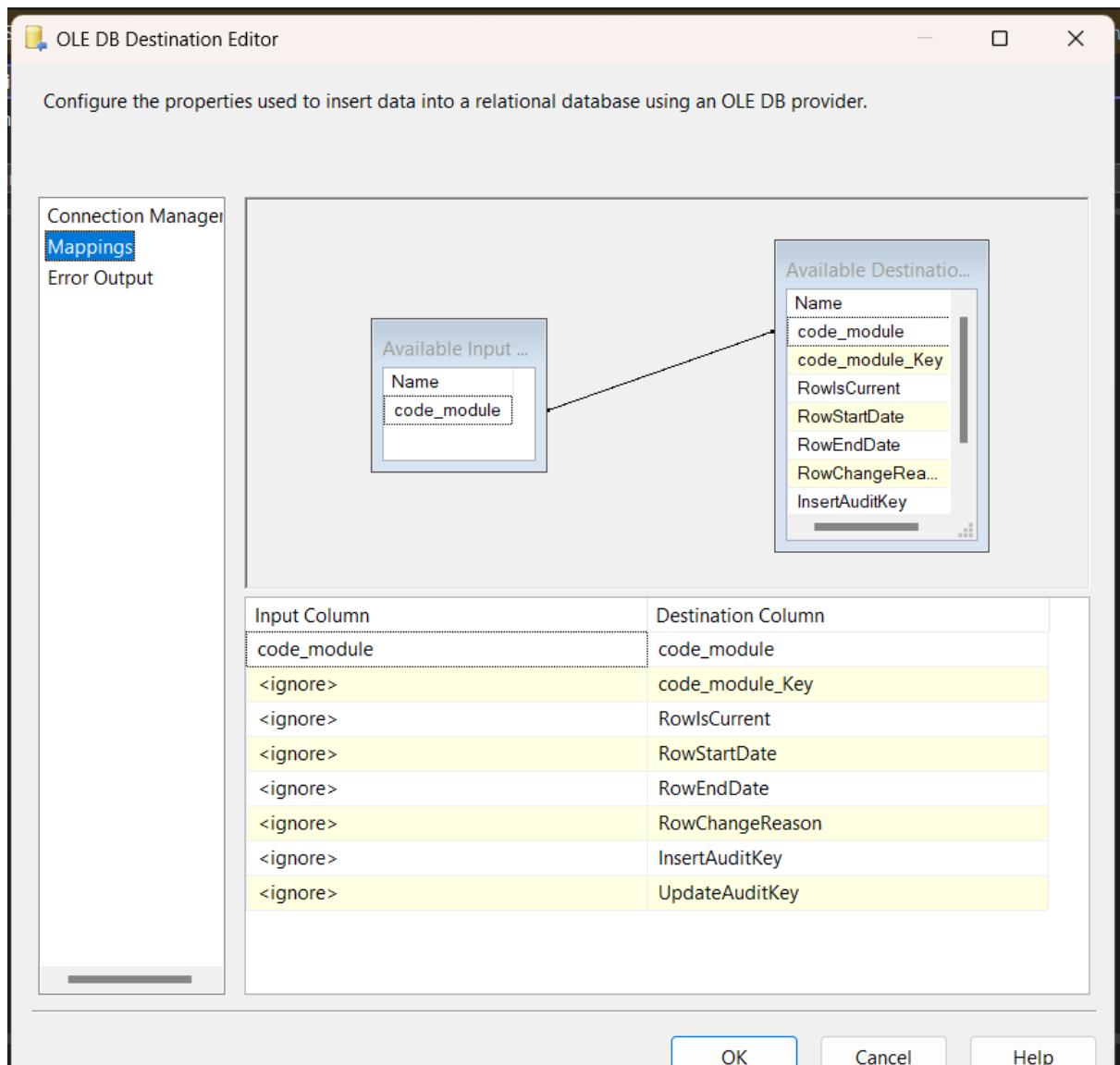
3.3.3.2. Load dữ liệu từ Stage vào Dimension

- Double-click vào **DF - Load to DimCourses** task để mở data flow design surface.
- + Source Assistant:

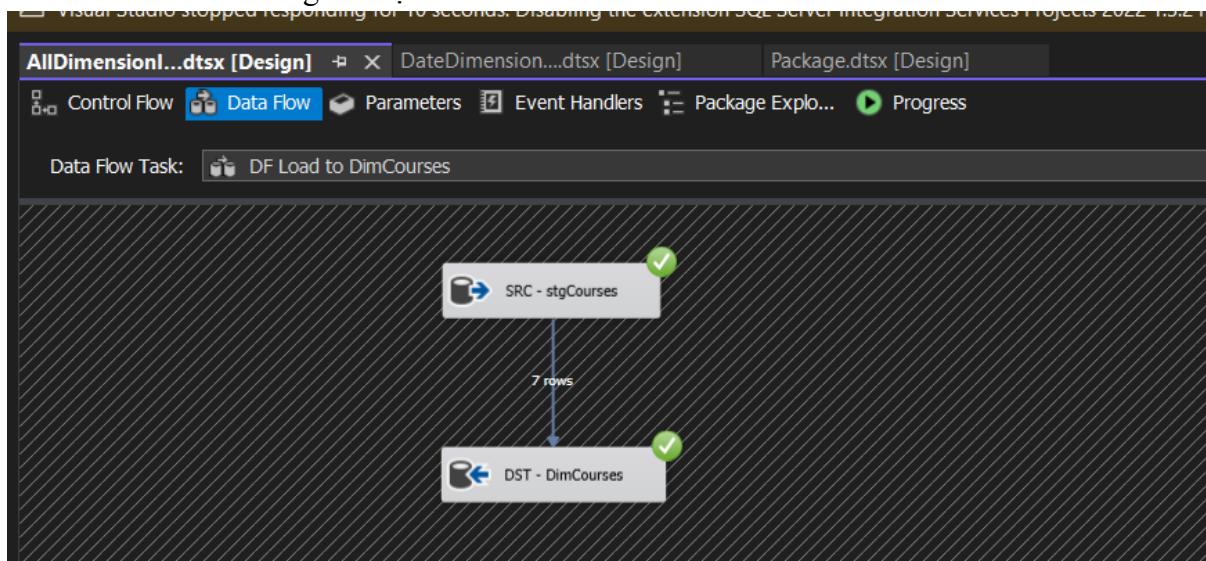


+ Destination Assistant:





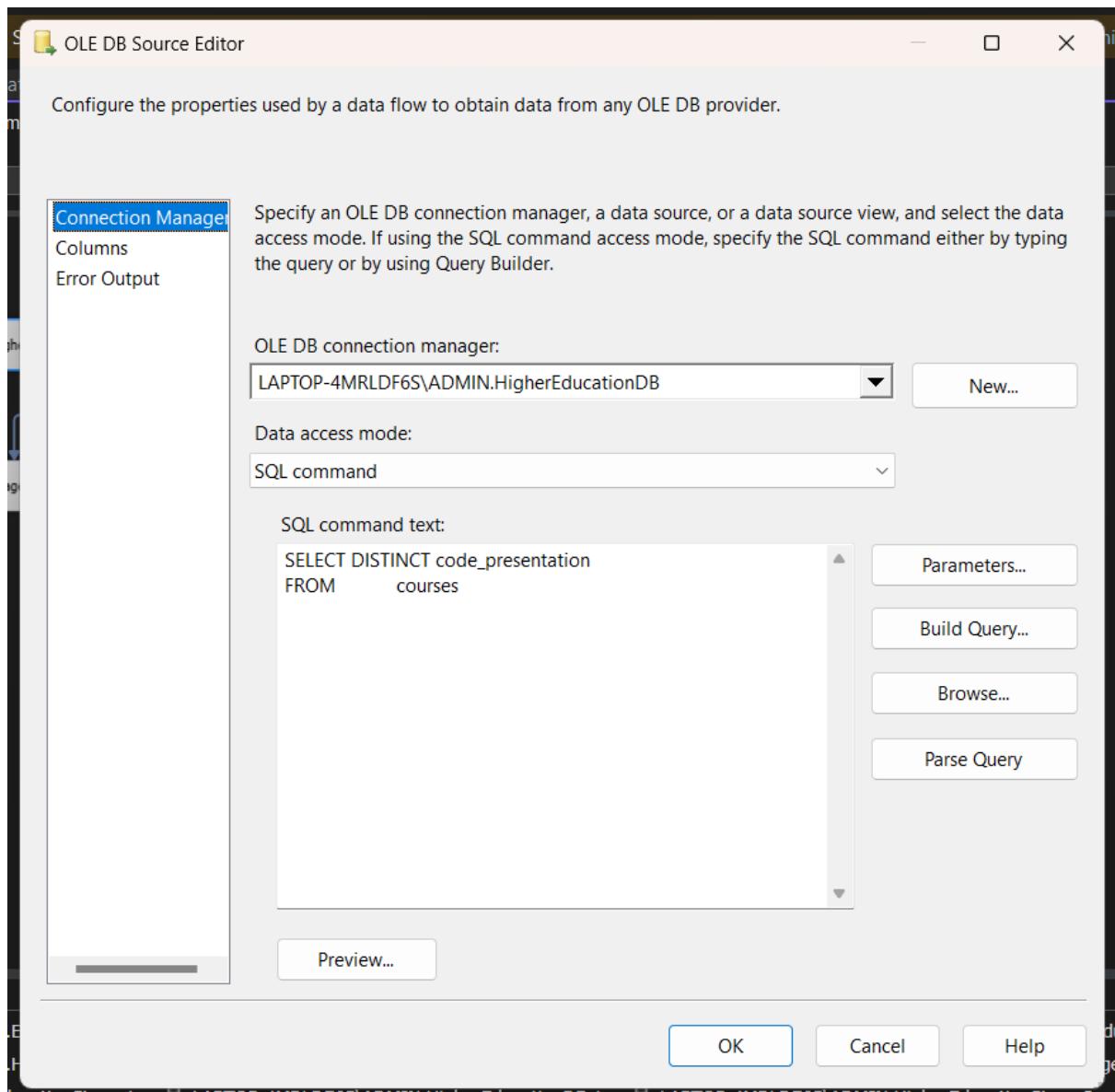
+ Sau khi cấu hình xong ta được:



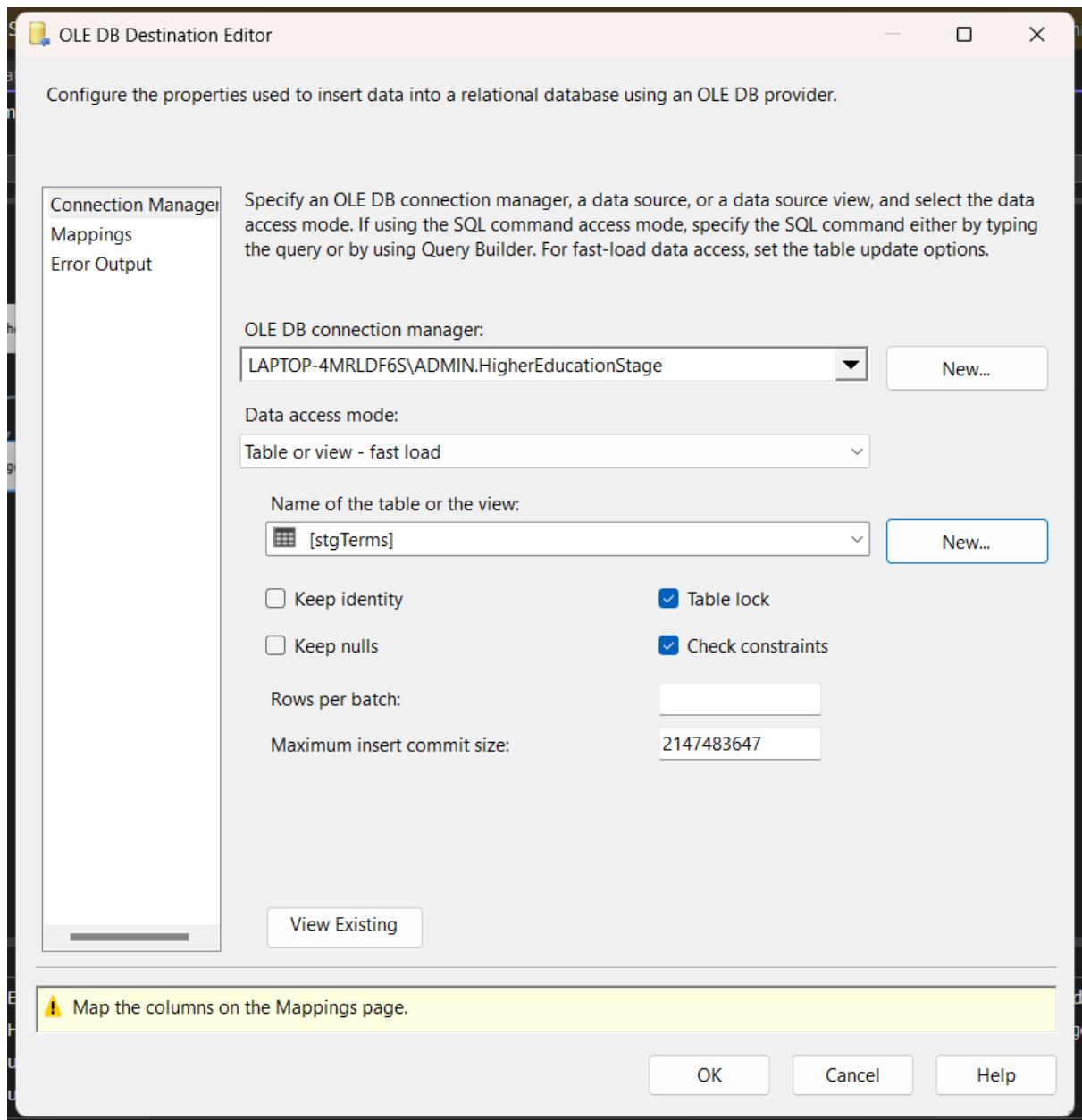
3.3.4.Term Dimension

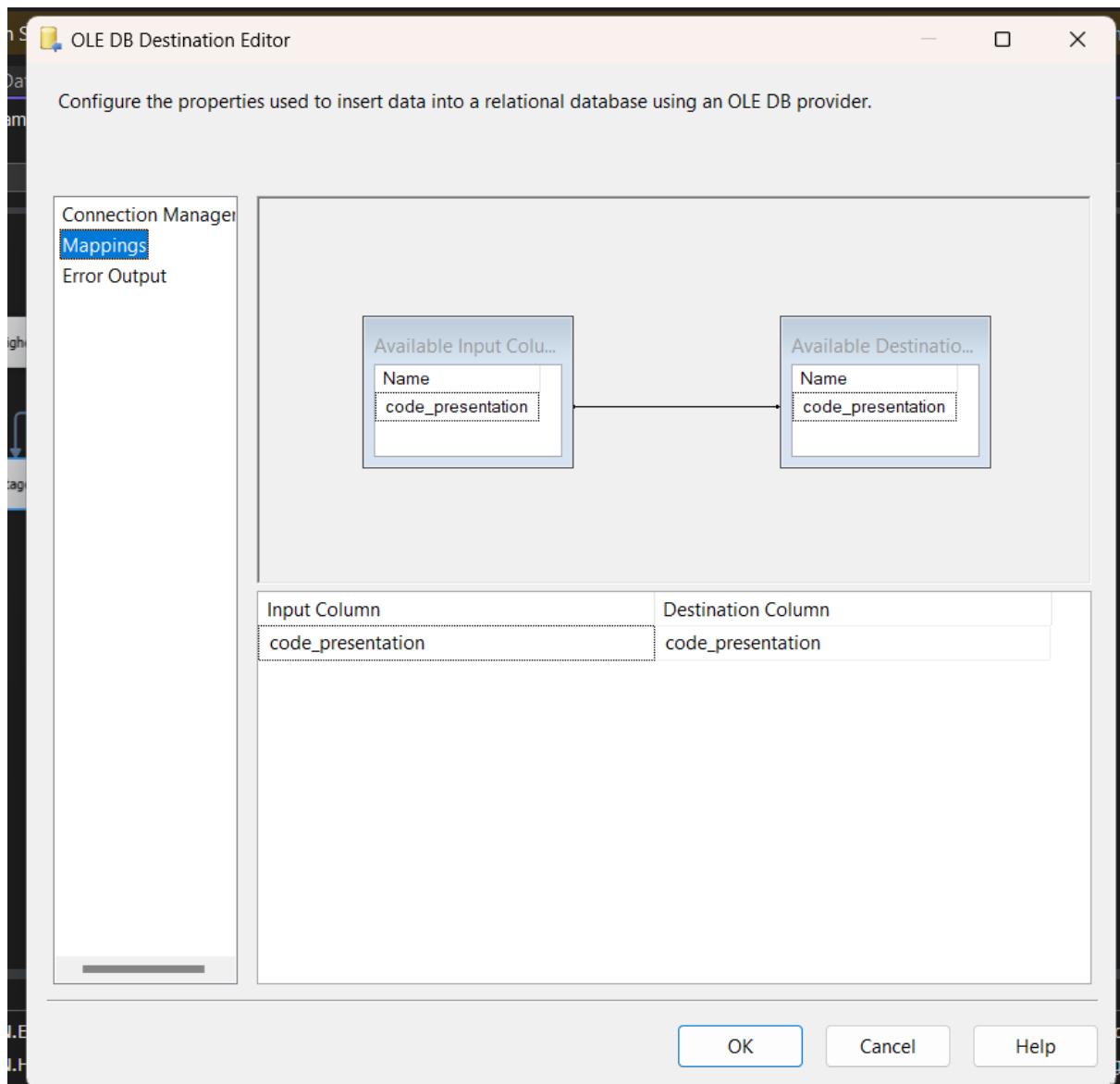
3.3.4.1. Load dữ liệu từ Source vào Stage

- Double-click **DF - Stage Term** task để mở nó trong **Data Flow design surface**:
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:
 - + Source Assistant:

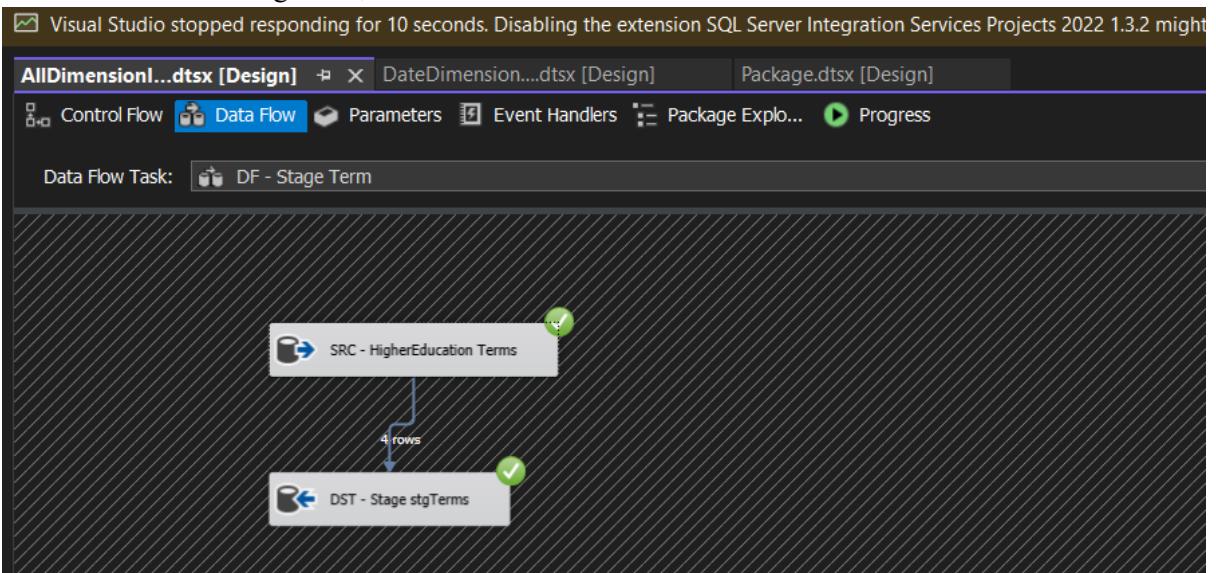


+ Destination Assistant:



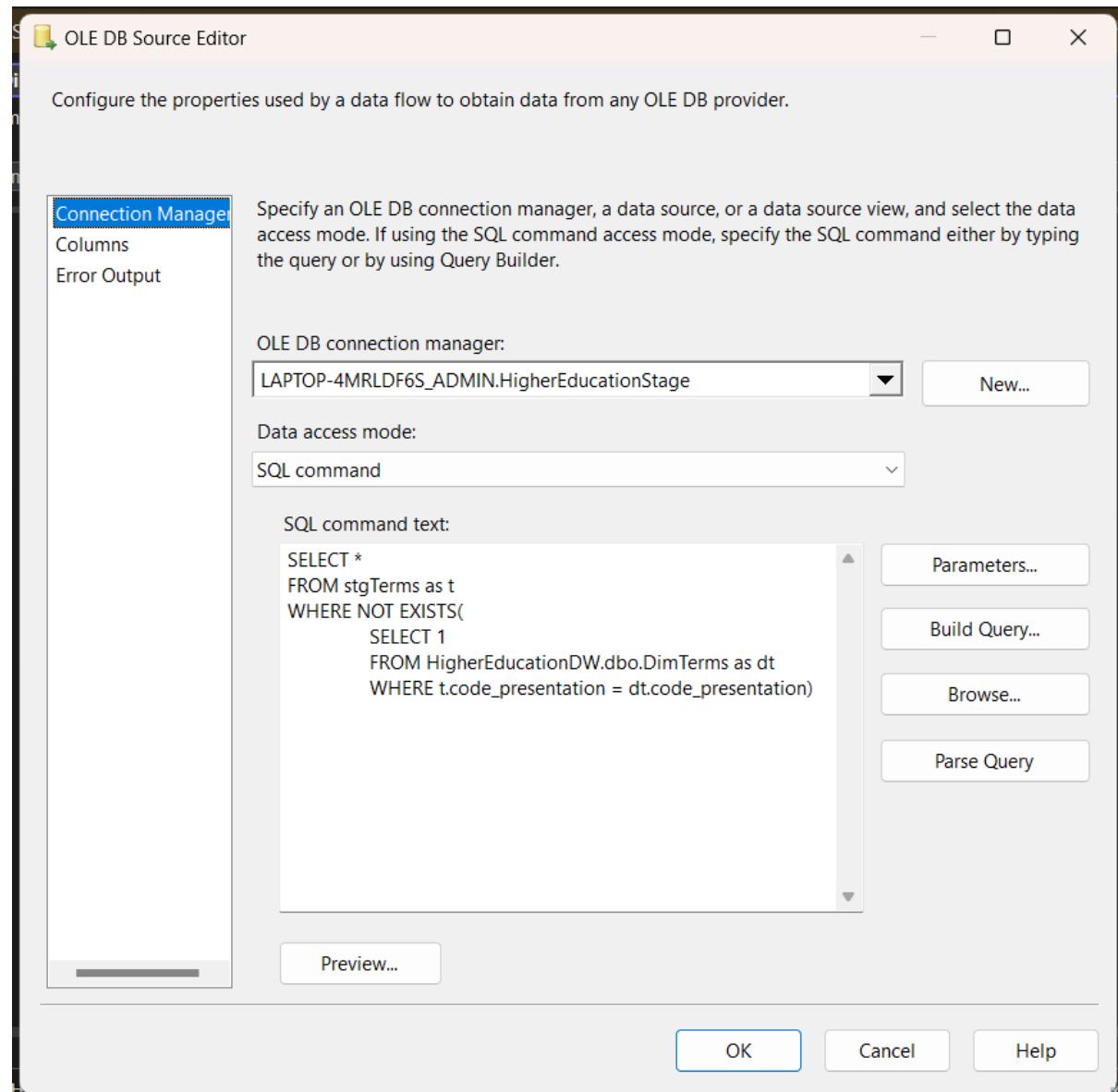


+ Sau khi cấu hình xong ta được:

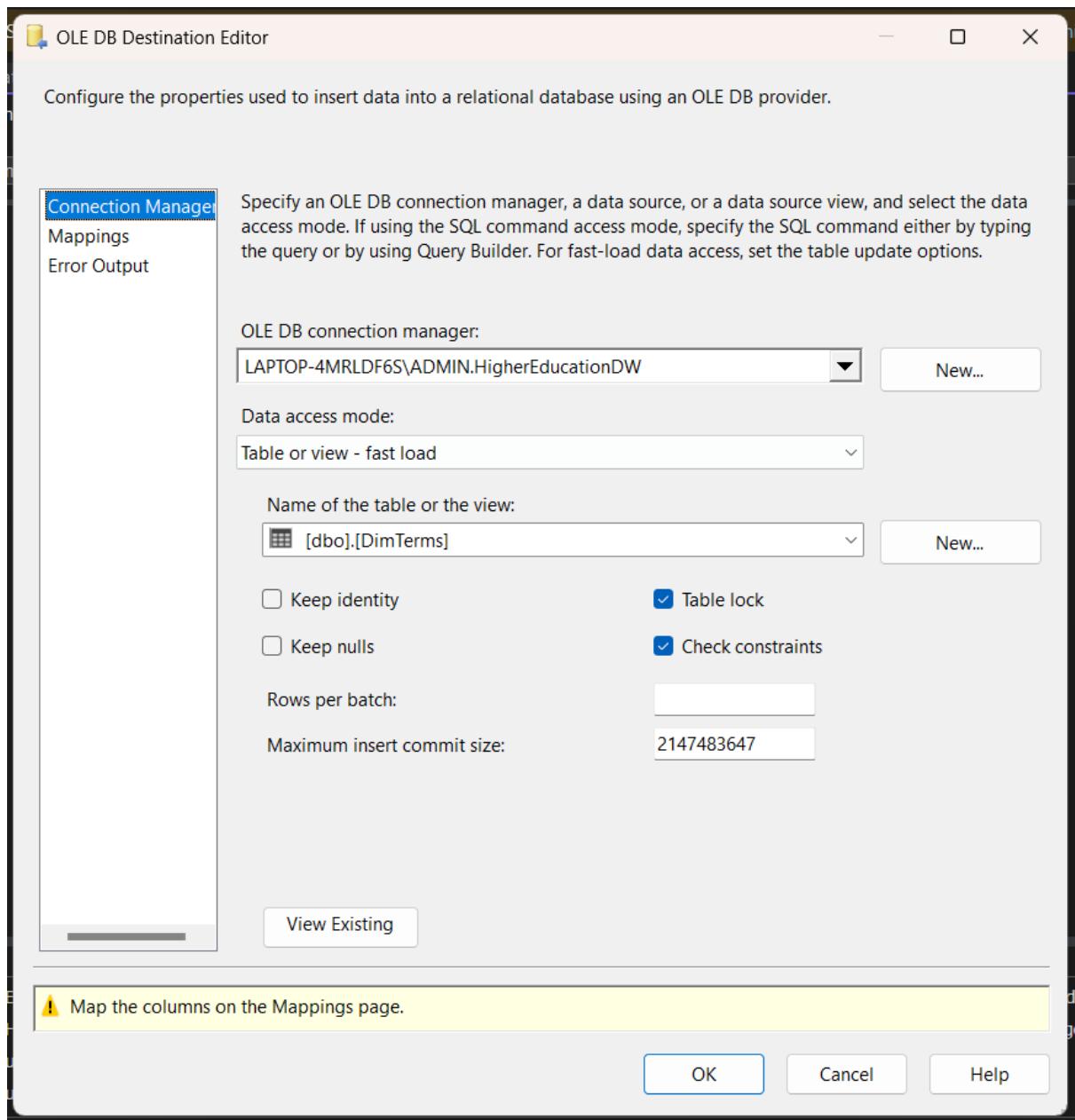


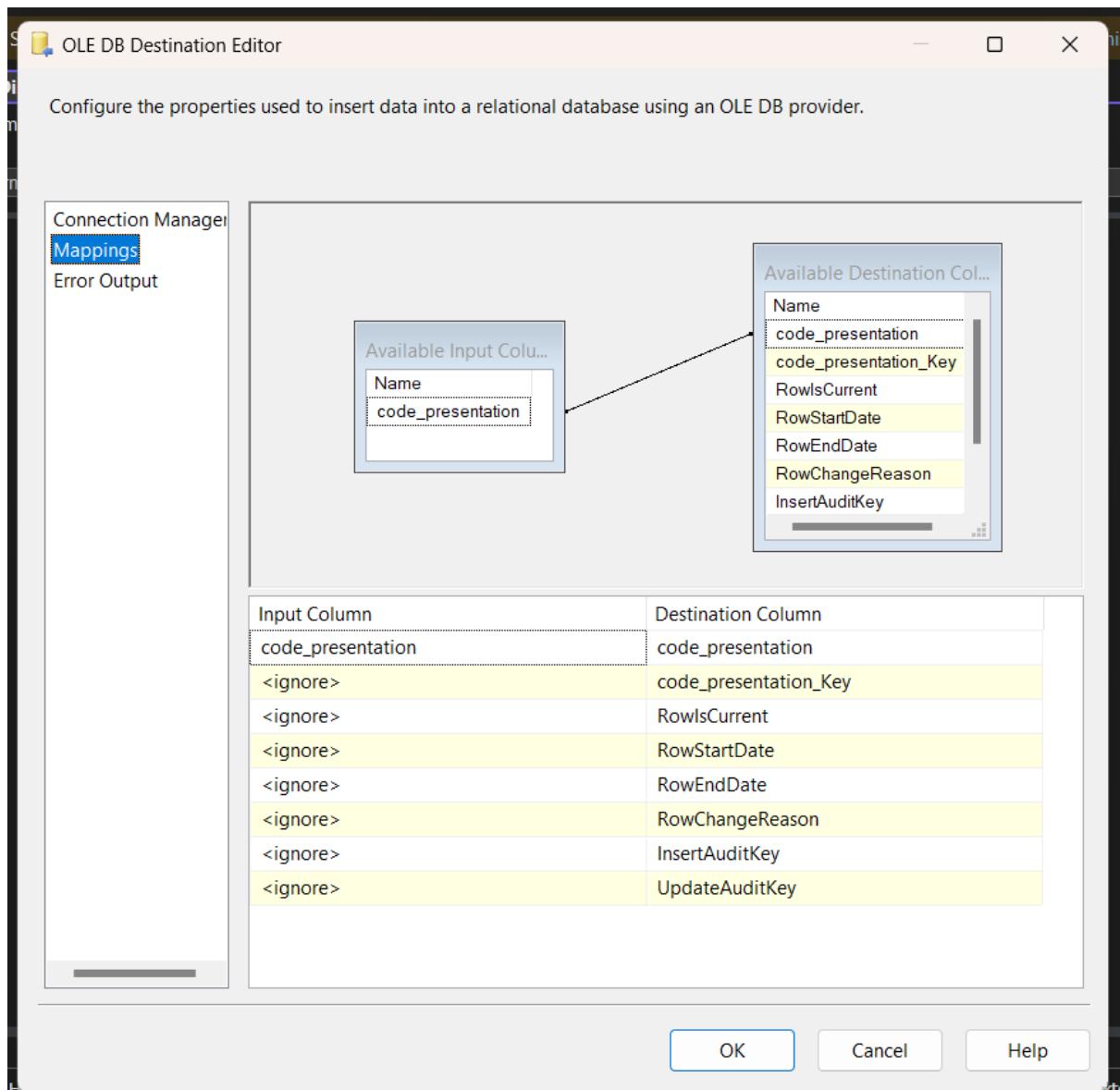
3.3.4.2. Load dữ liệu từ Stage vào Dimension

- Double-click vào **DF - Load to DimTerms** task để mở data flow design surface.
- + Source Assistant:

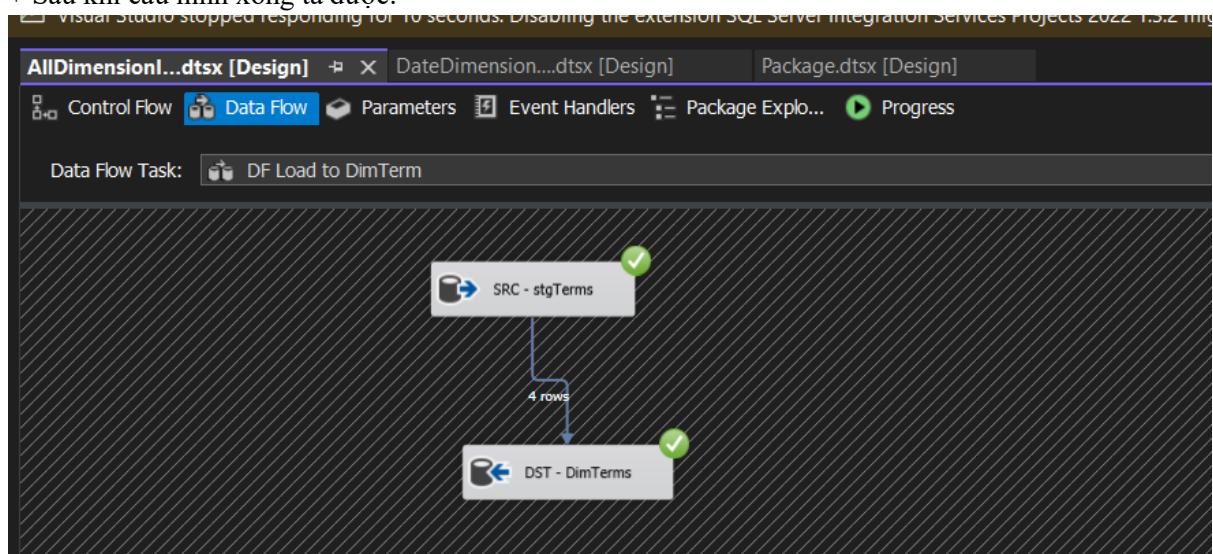


- + Destination Assistant:





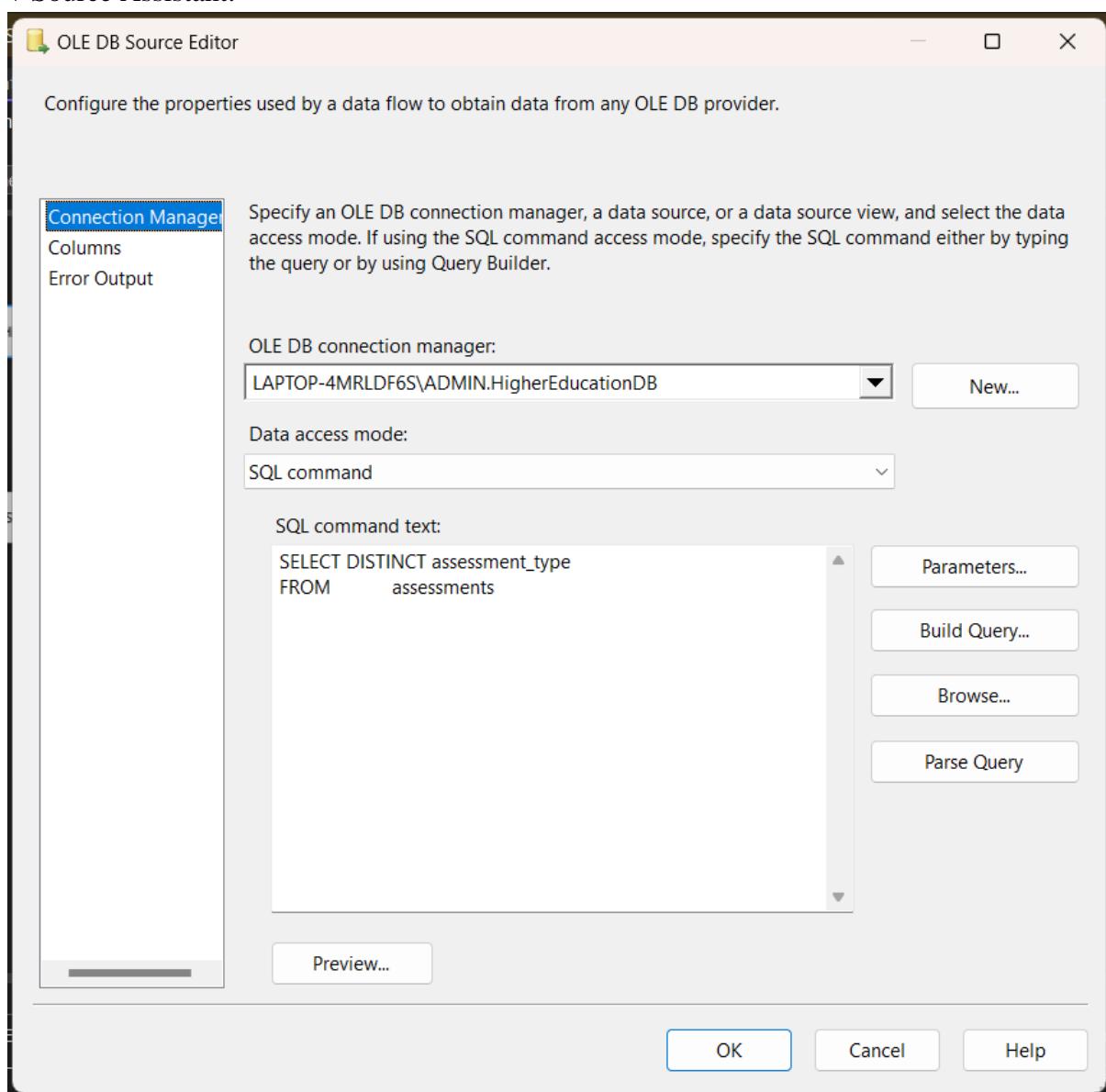
+ Sau khi cấu hình xong ta được:



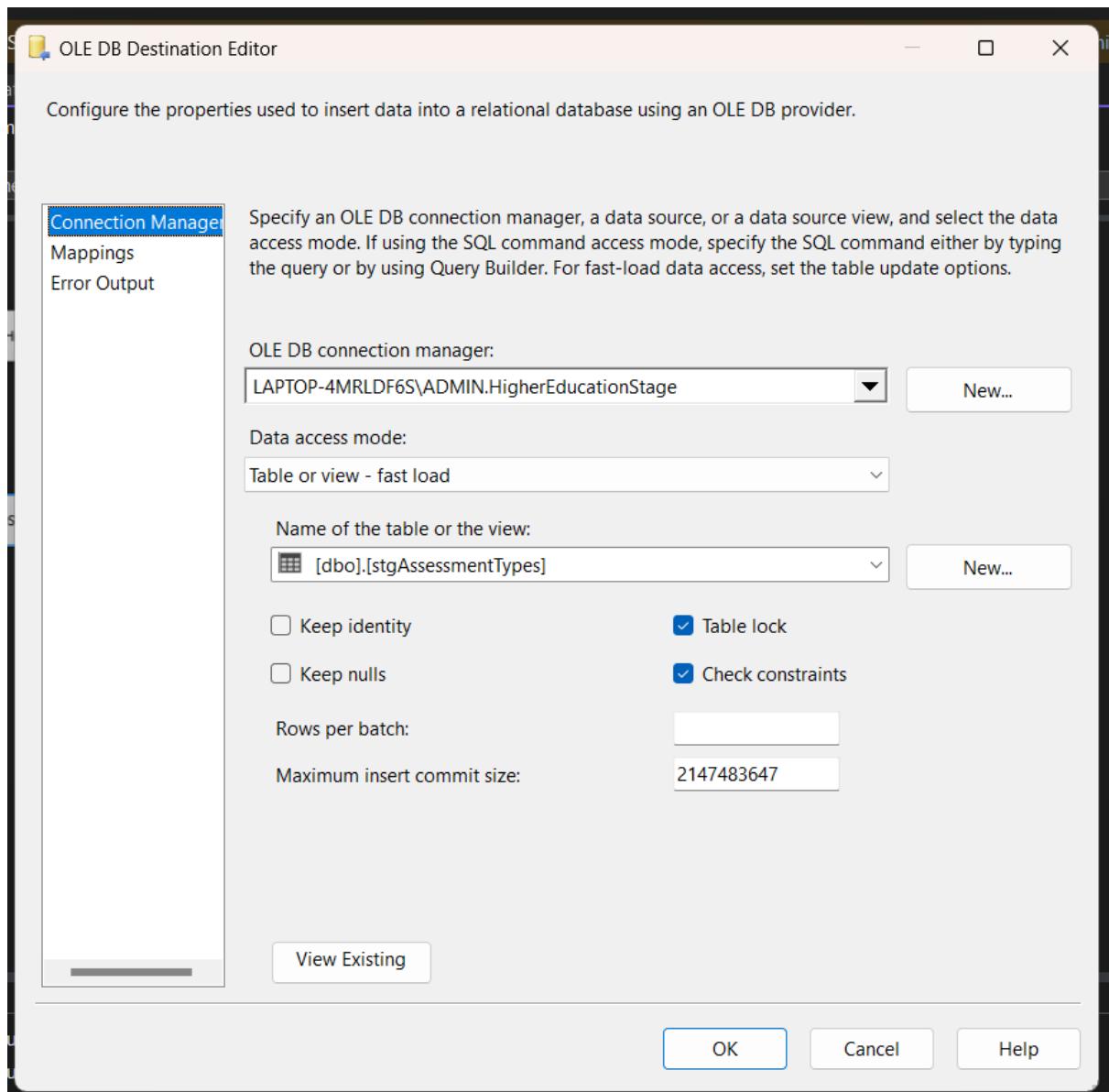
3.3.5.AssessmentTypes Dimension

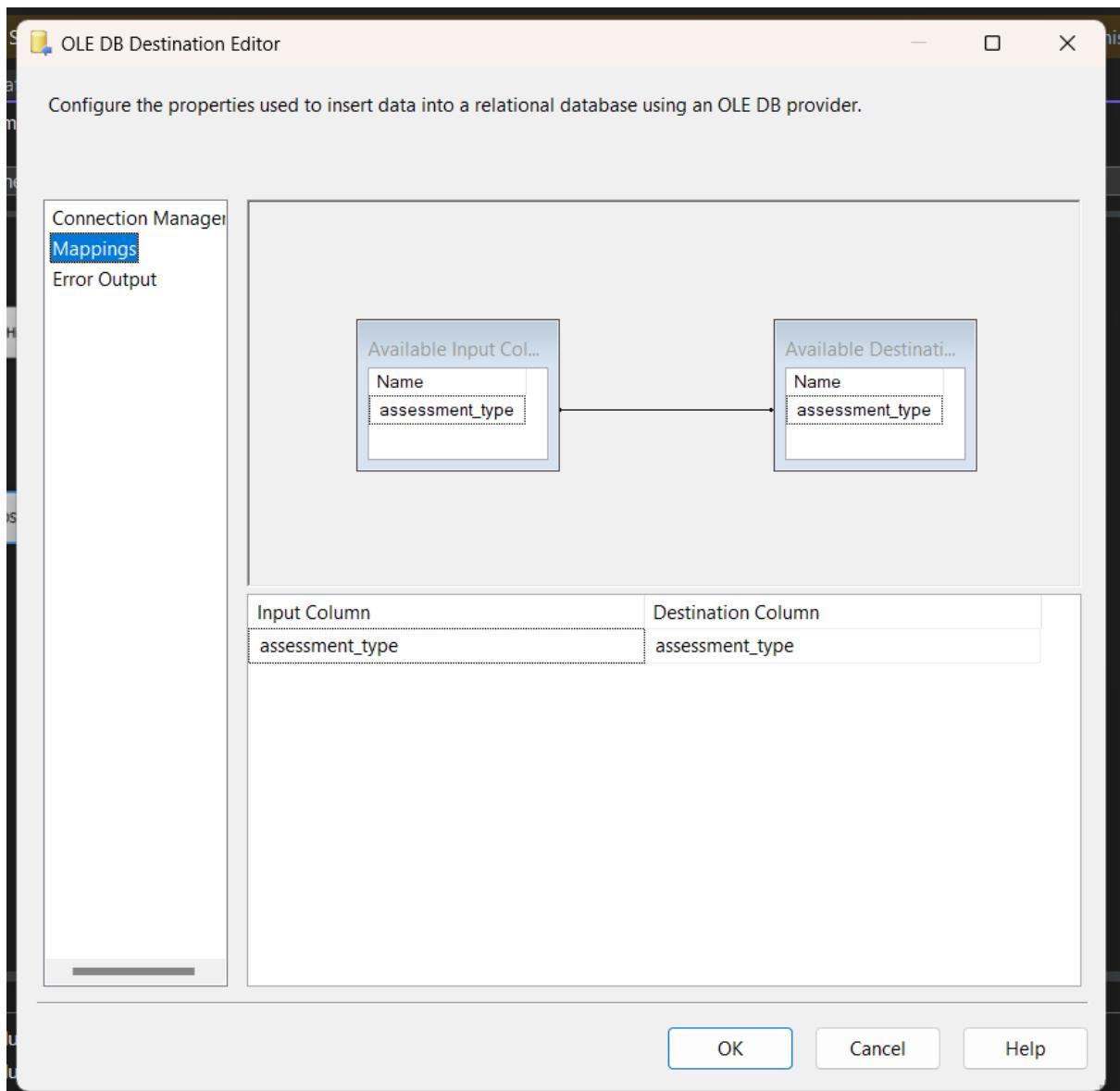
3.3.5.1. Load dữ liệu từ Source vào Stage

- Double-click DF - Stage AssessmentTypes task để mở nó trong Data Flow design surface:
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:
- + **Source Assistant:**

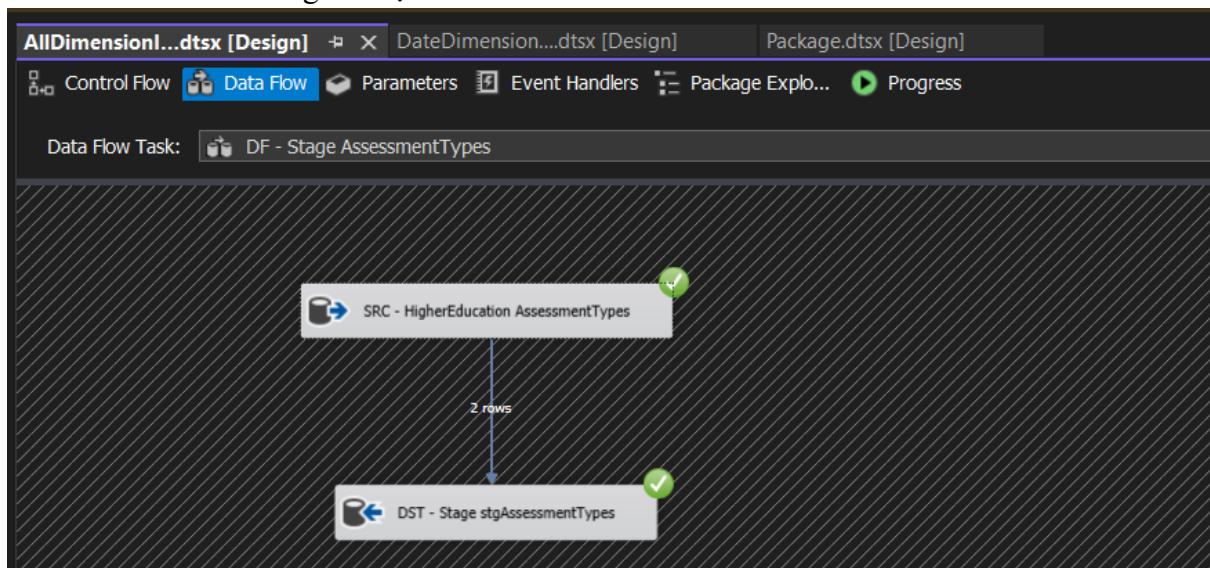


- + **Destination Assistant:**



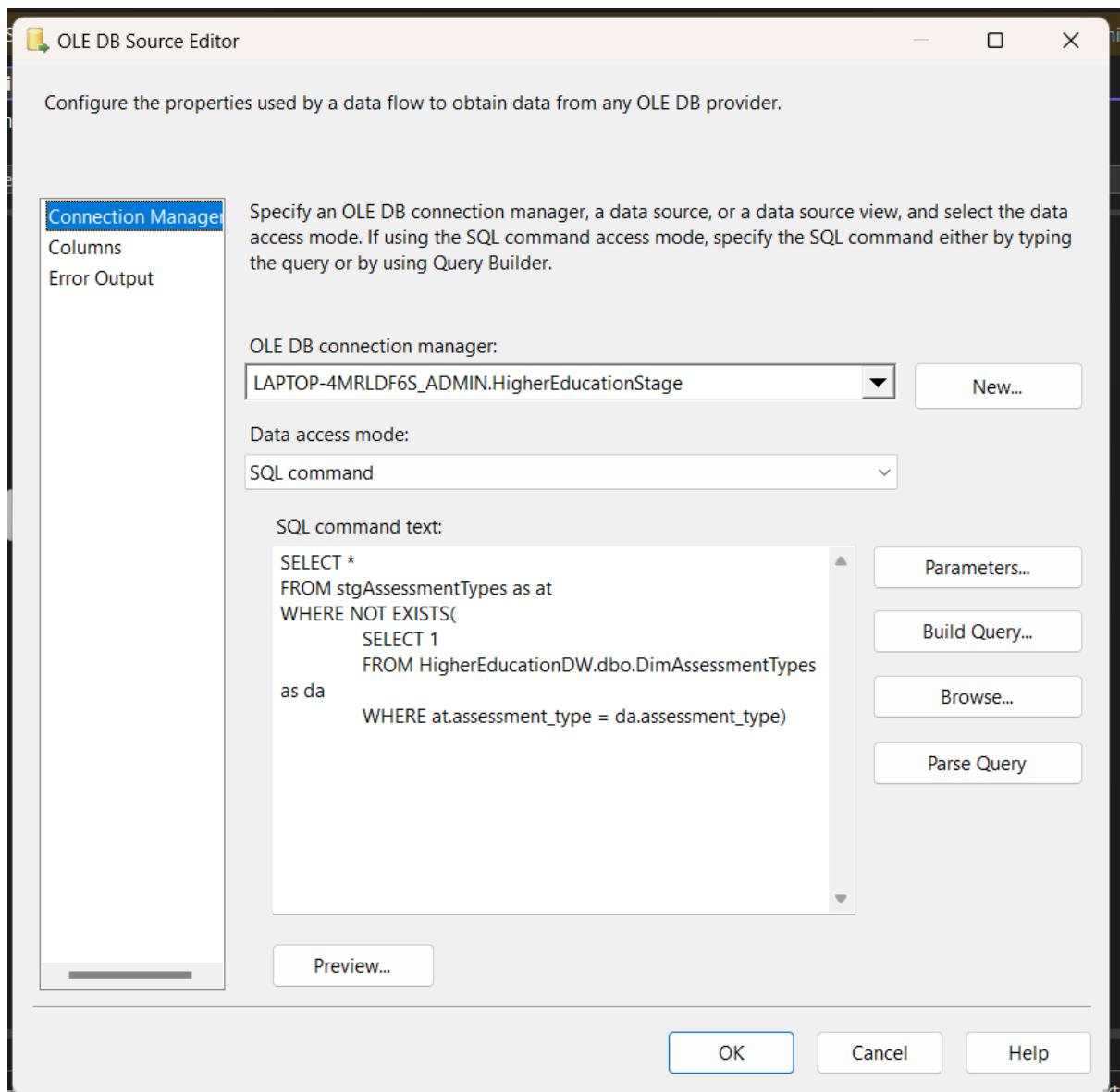


+ Sau khi cấu hình xong ta được:

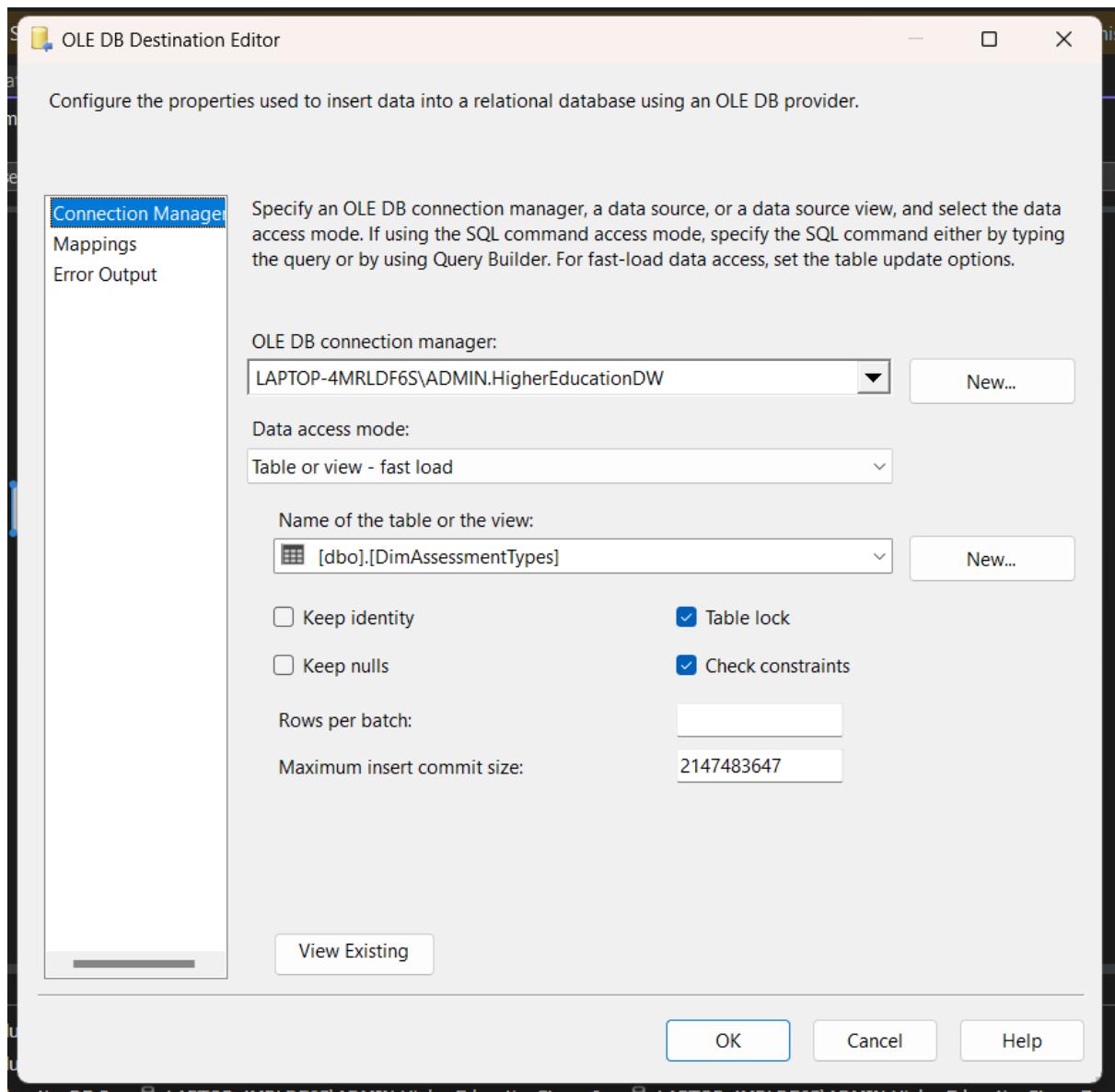


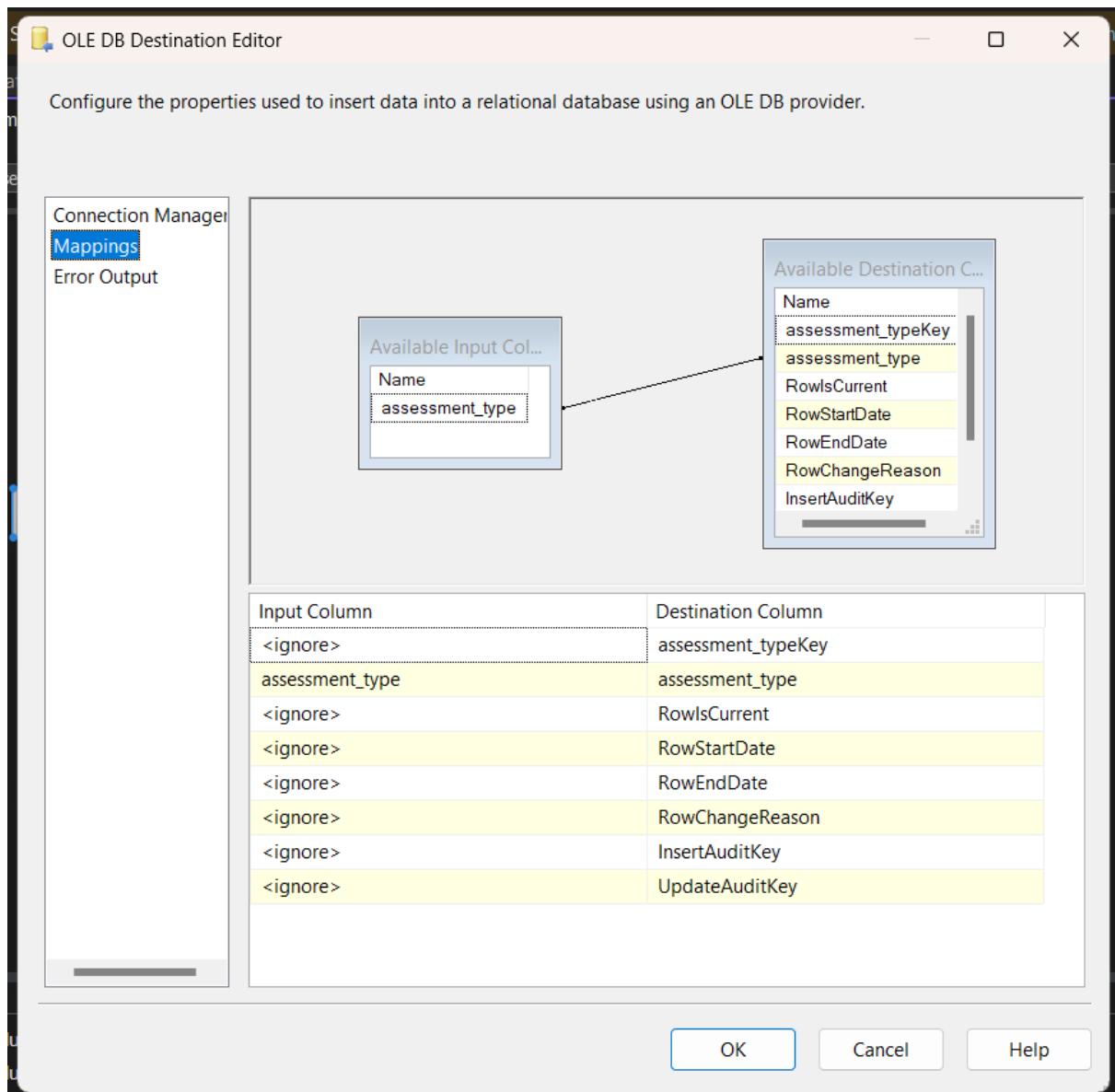
3.3.5.2. Load dữ liệu từ Stage vào Dimension

- Double-click vào **DF - Load to DimAssessmentTypes** task để mở data flow design surface.
- + Source Assistant:

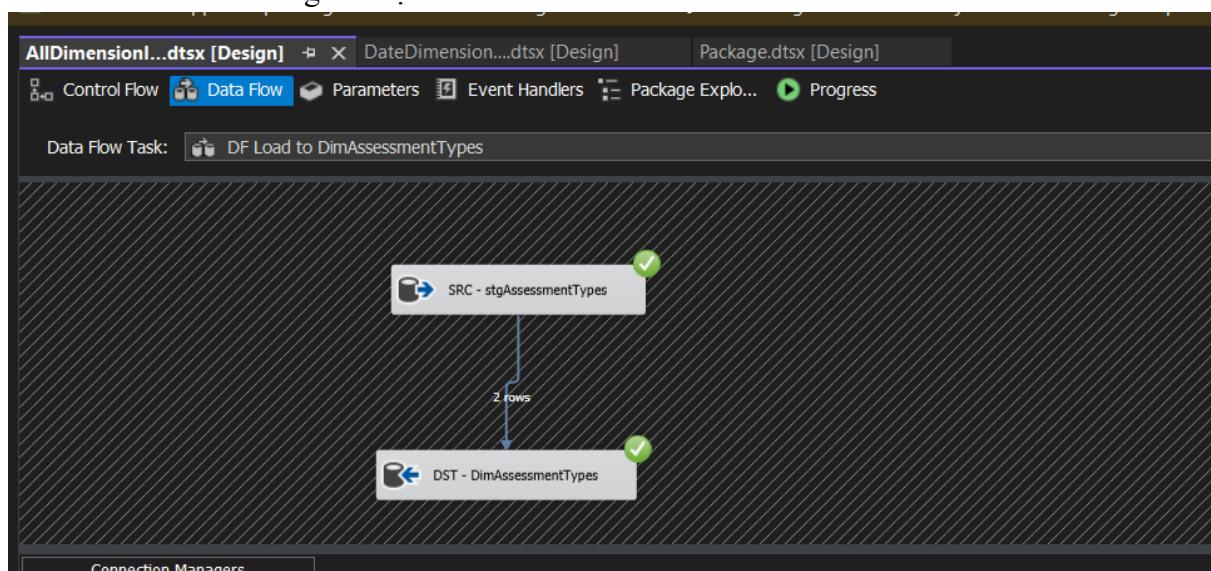


- + Destination Assistant:





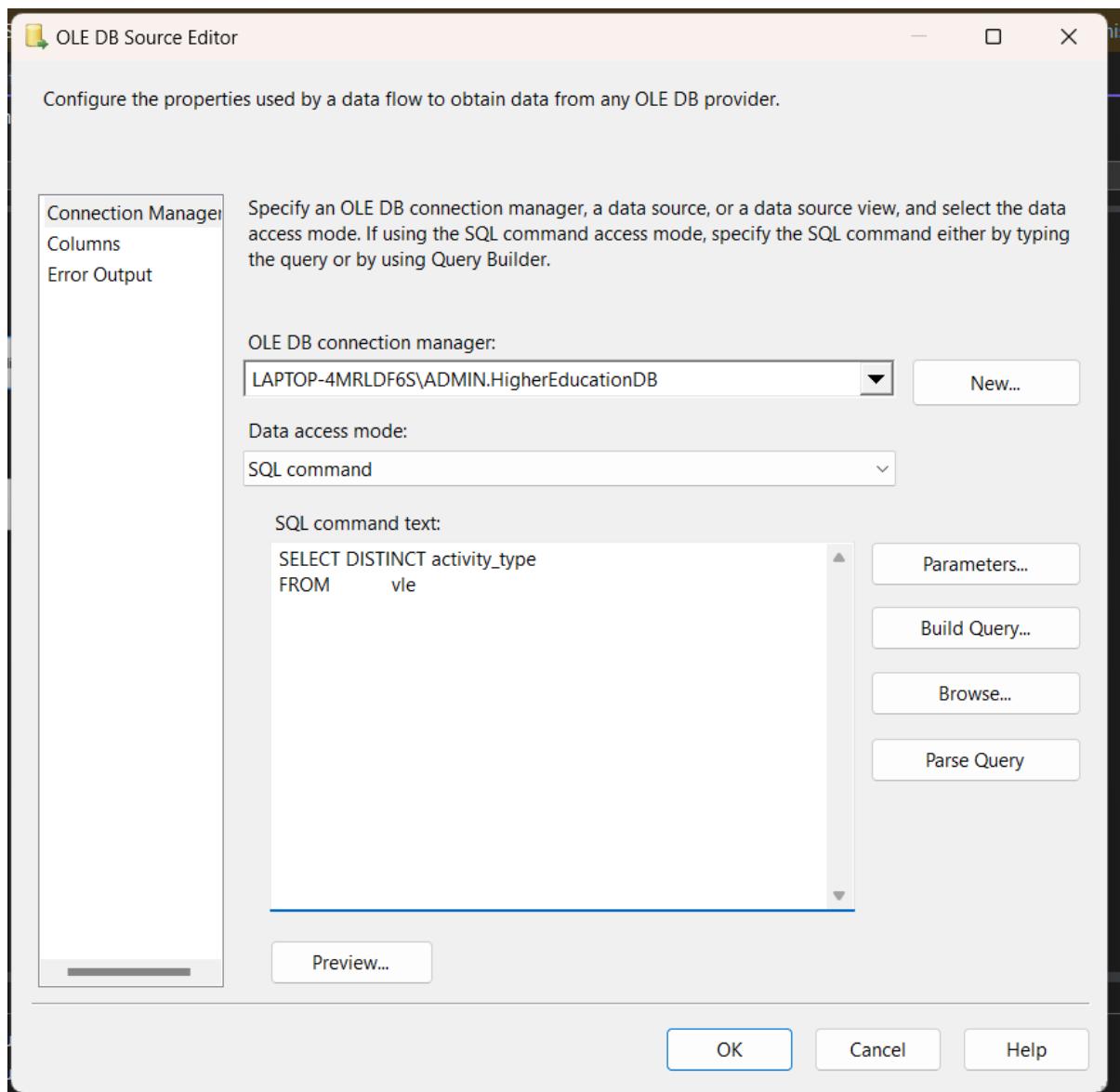
+ Sau khi cấu hình xong ta được:



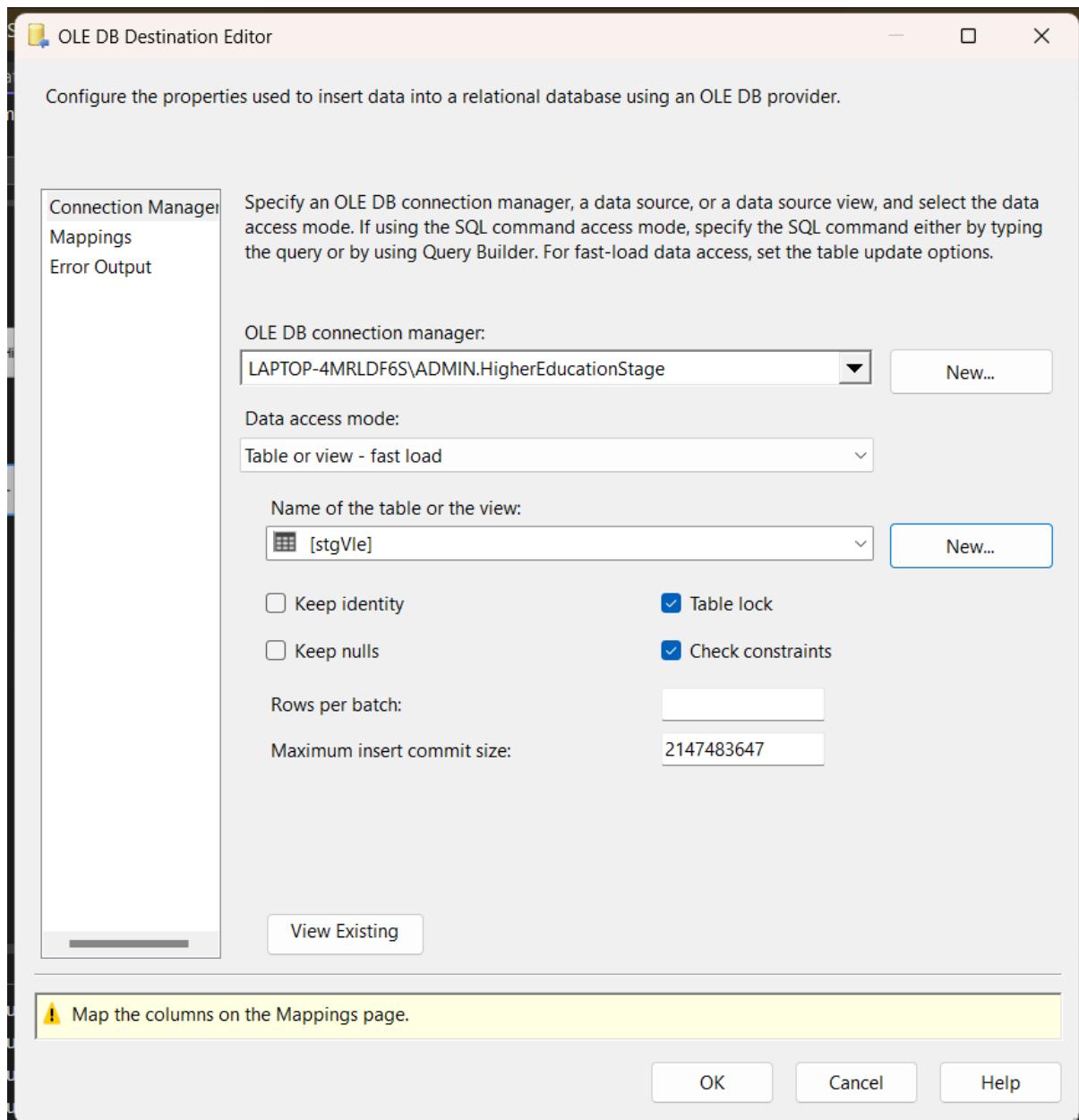
3.3.6. Vle Dimension

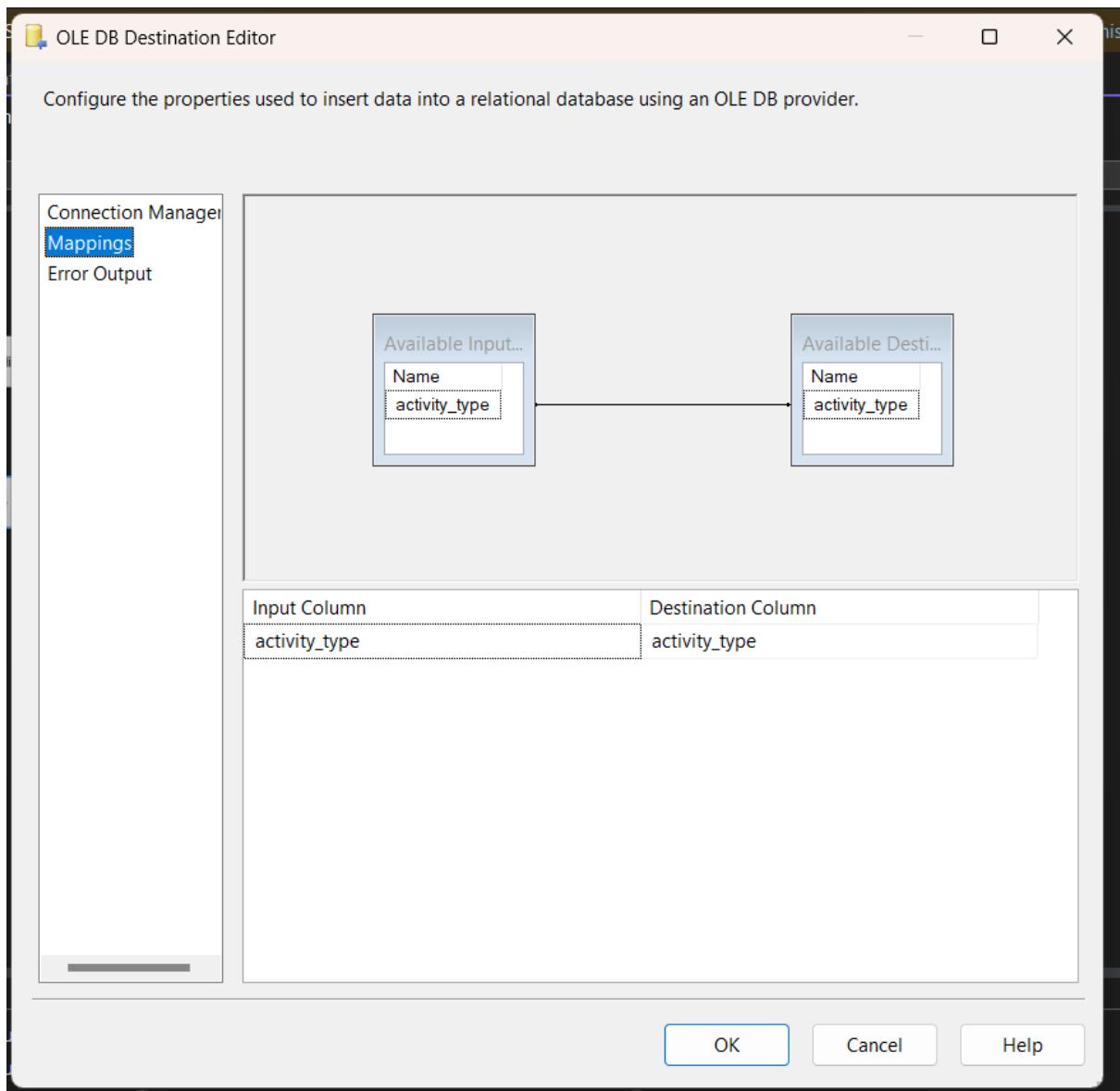
3.3.6.1. Load dữ liệu từ Source vào Stage

- Double-click DF - Stage Vle task để mở nó trong **Data Flow design surface**:
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:
- + Source Assistant:

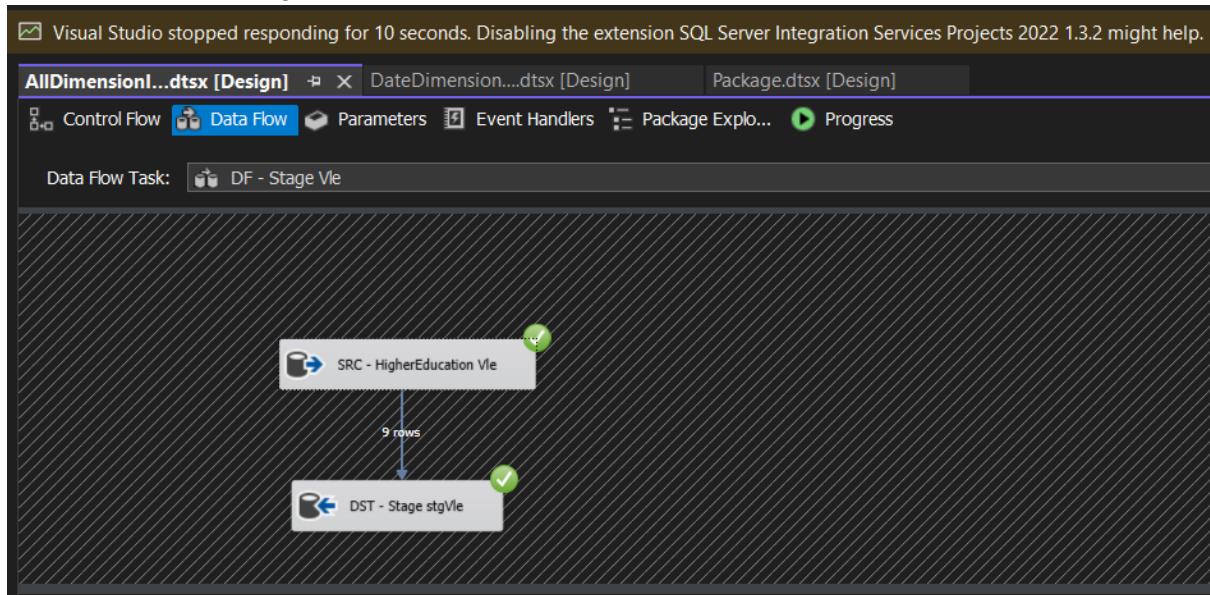


- + Destination Assistant:





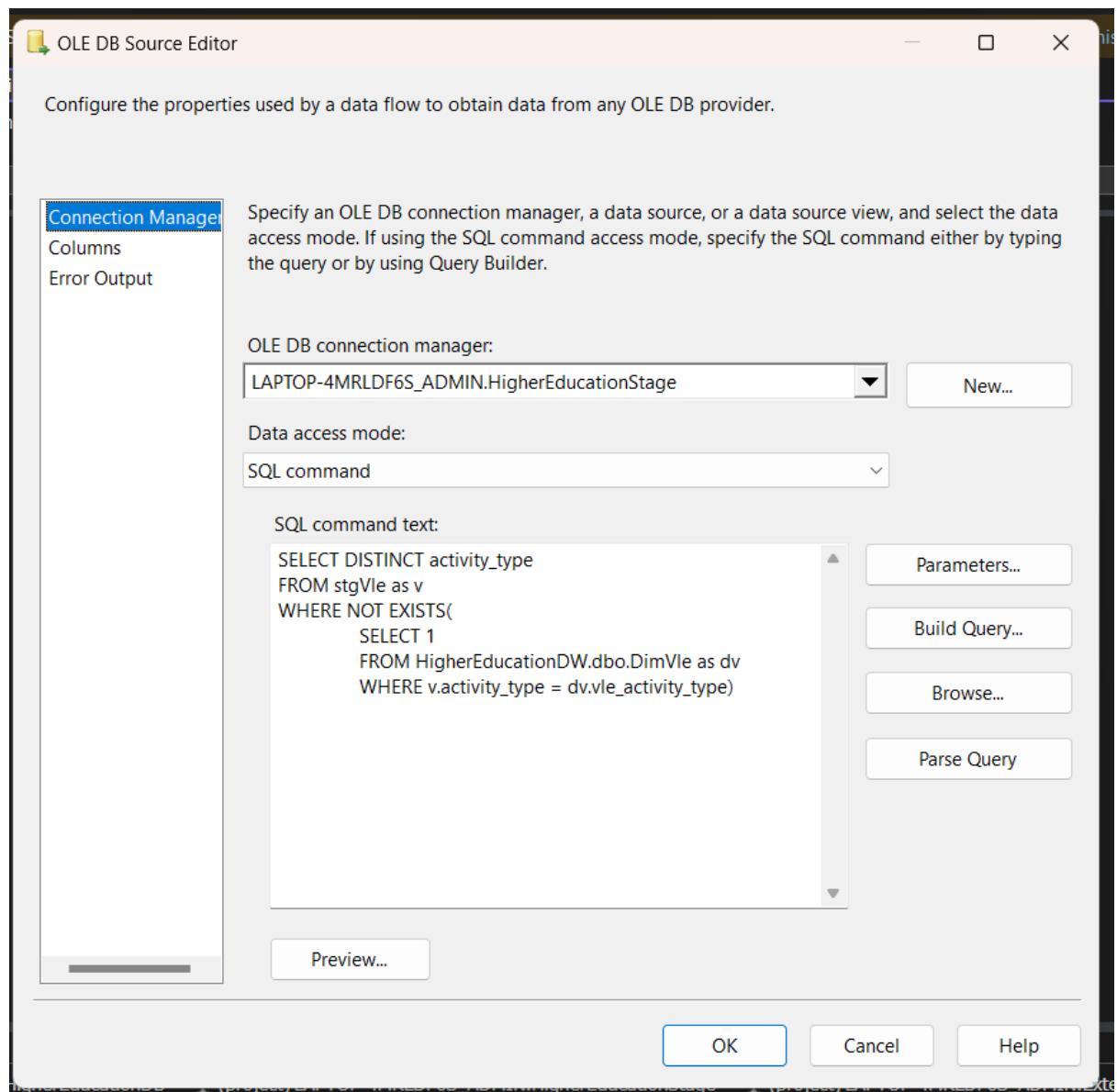
+ Sau khi cấu hình xong ta được:



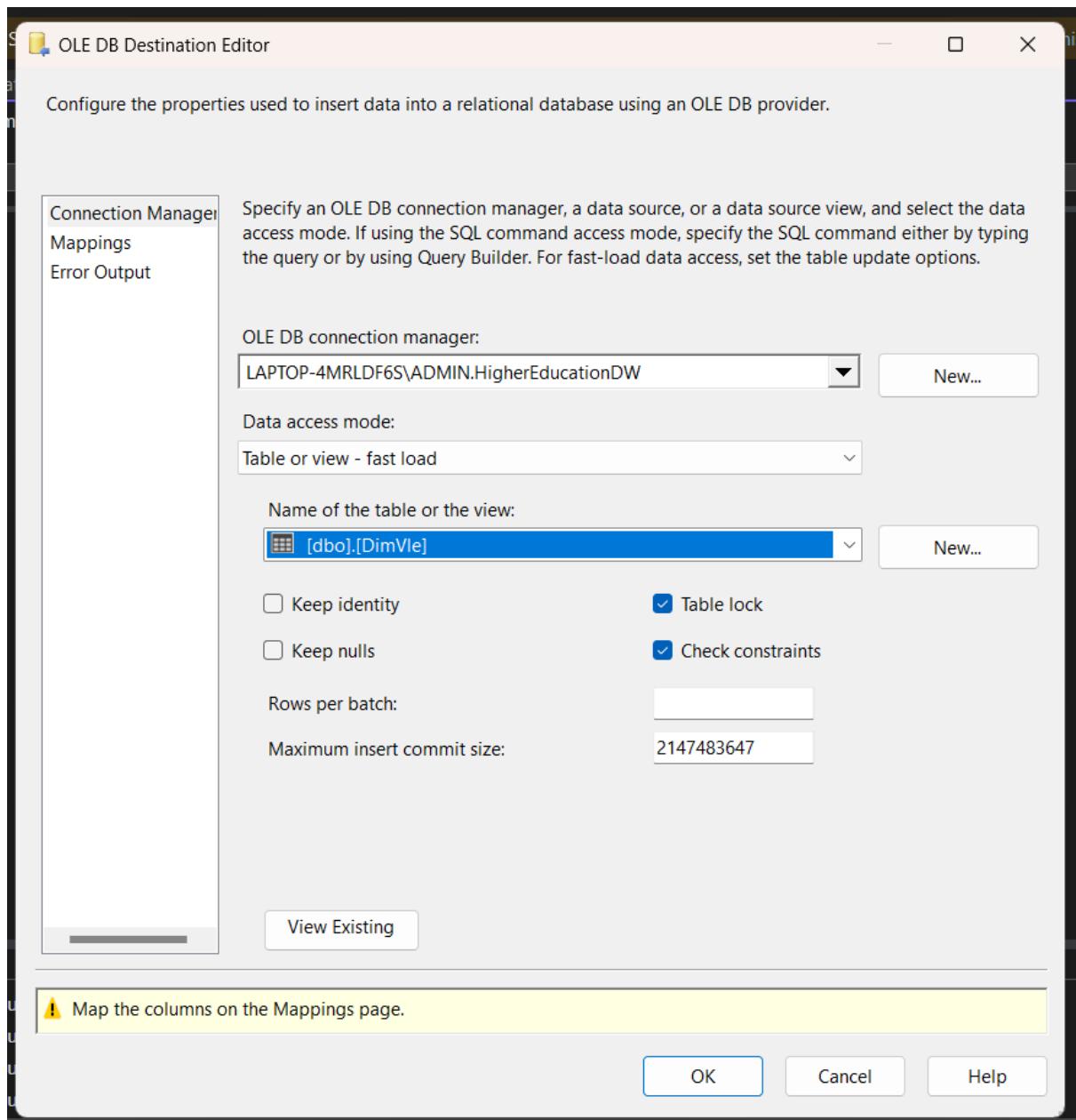
3.3.6.2. Load dữ liệu từ Stage vào Dimension

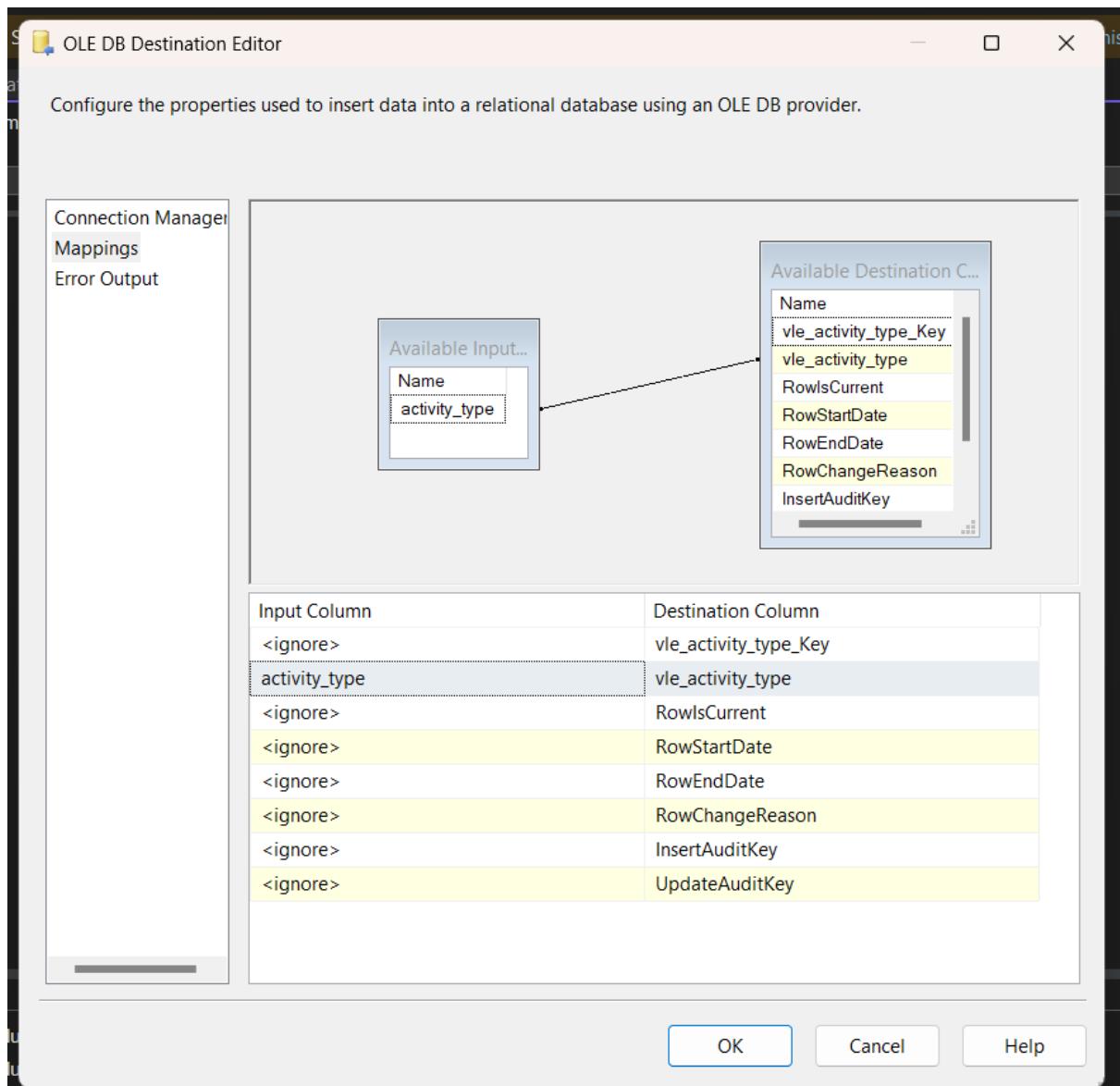
- Double-click vào **DF - Load to DimVle** task để mở data flow design surface.

+ Source Assistant:

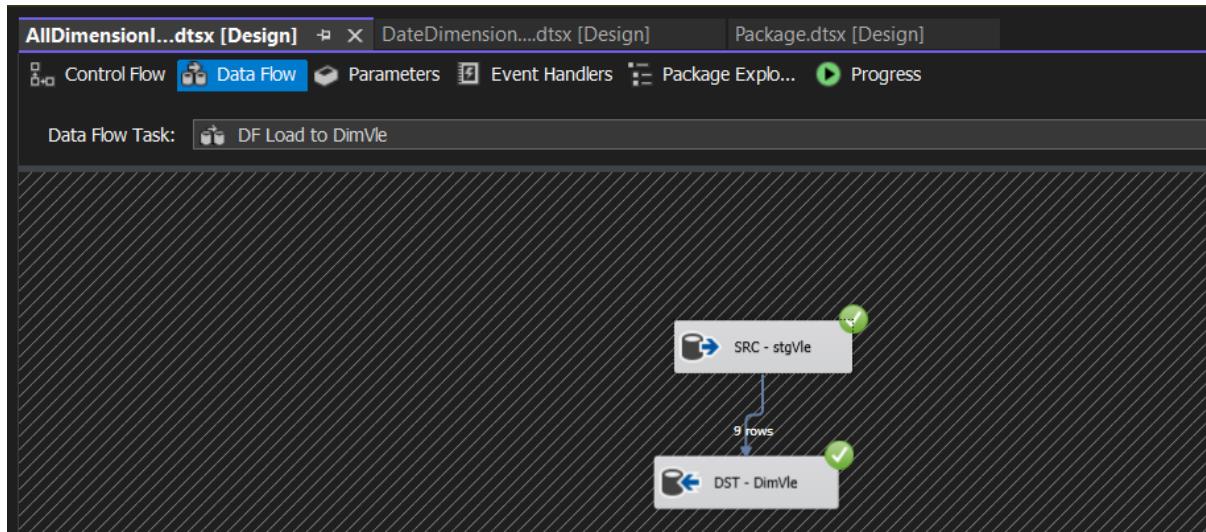


+ Destination Assistant:





+ Sau khi cấu hình xong ta được:

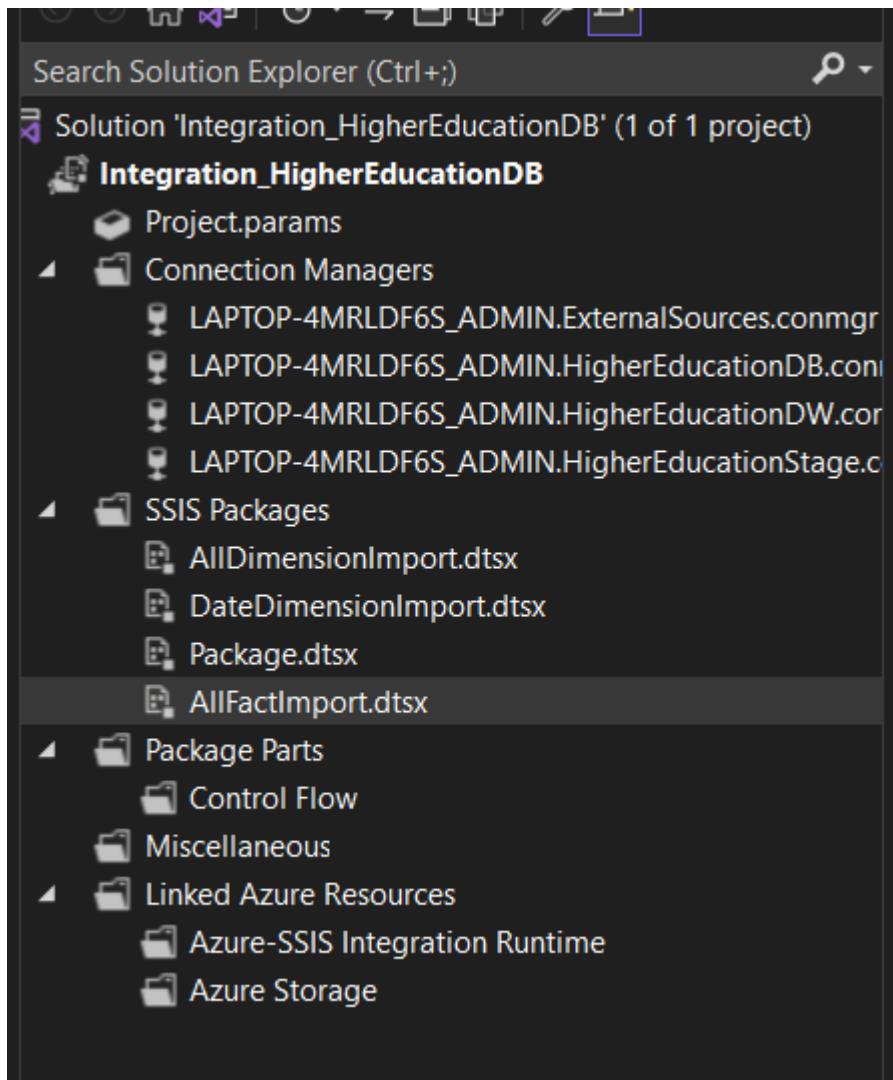


3.4. Import dữ liệu vào bảng các bảng fact

3.4.1. Fact Student Assessment

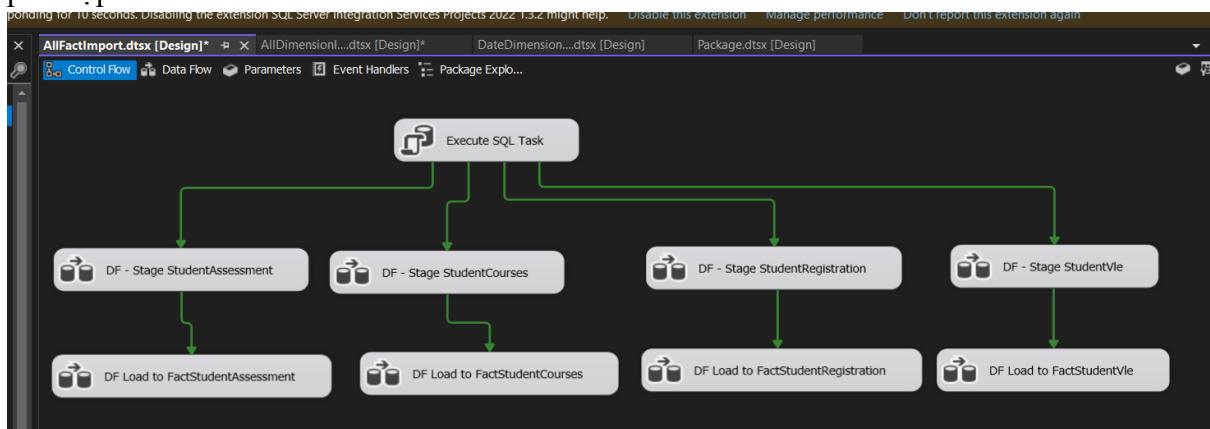
3.4.1.1. Tạo Package

- Tạo package mới và đổi tên như hình bên dưới.

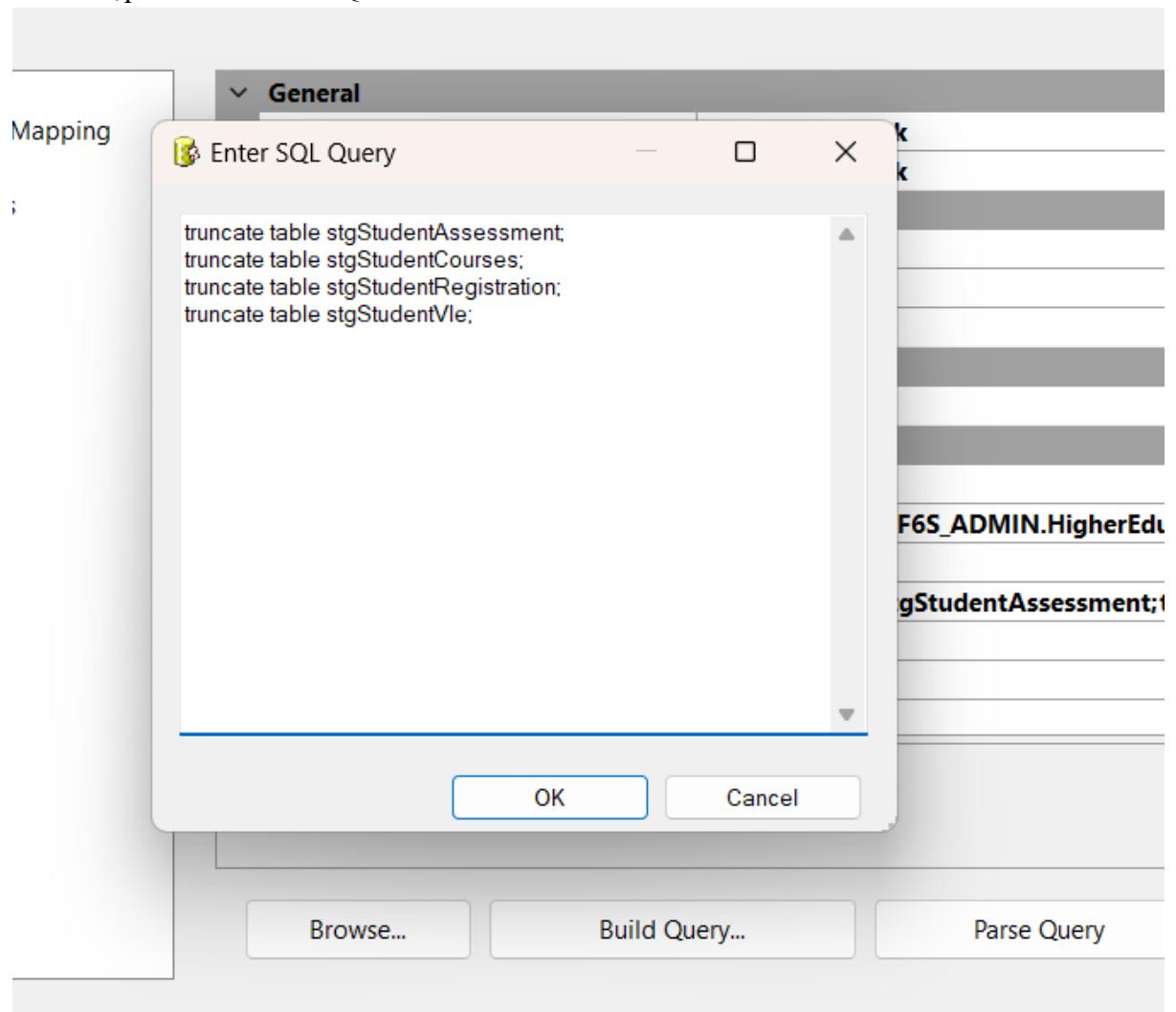


3.4.1.2. Thiết lập Control Flow

- Tạo và đặt tên tương ứng cho các task, tạo các connect bằng cách kéo mũi tên vào các task phù hợp.



- Thiết lập truncate table SQL task

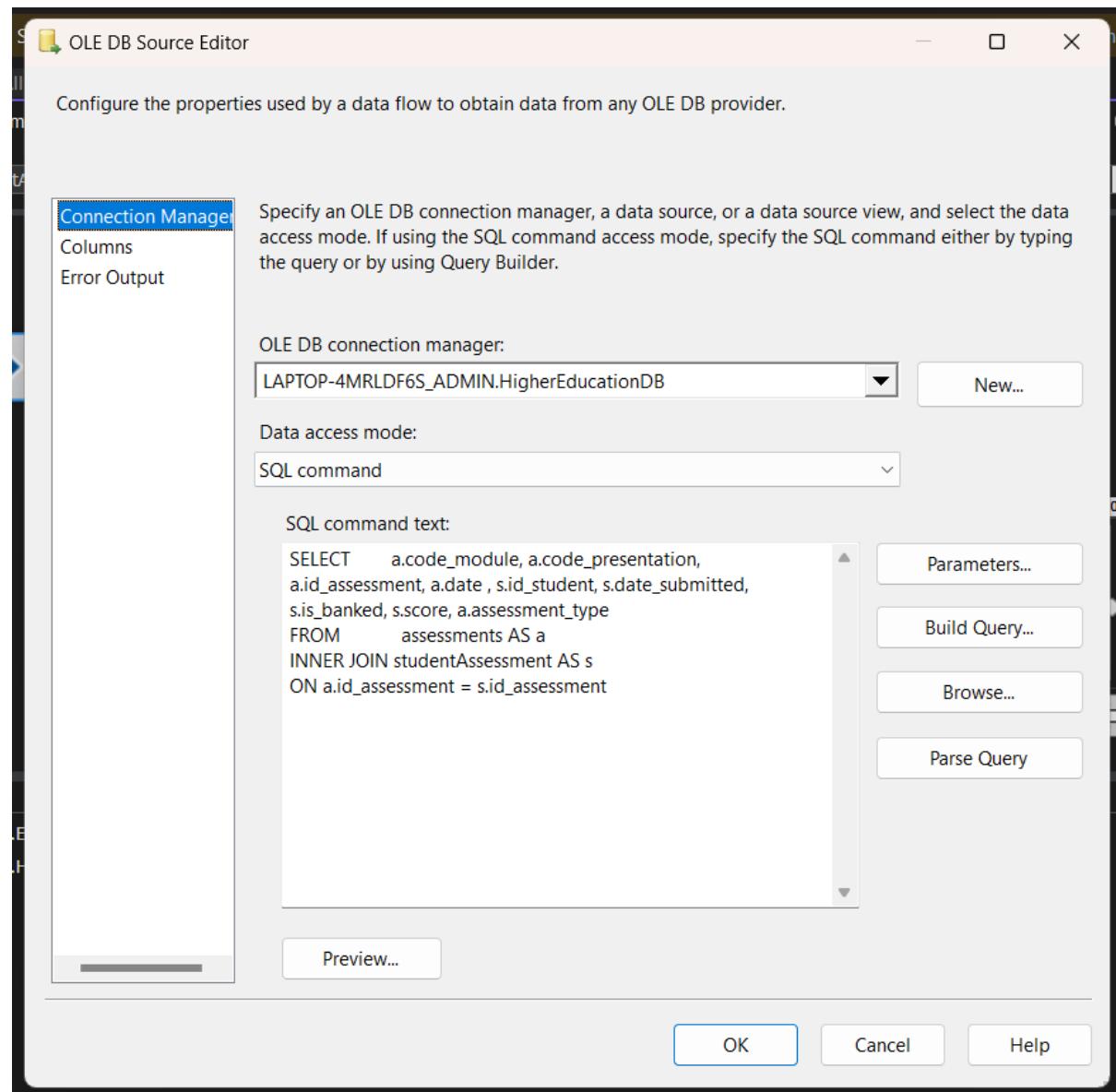


3.4.1.3. Load dữ liệu từ Source vào Stage

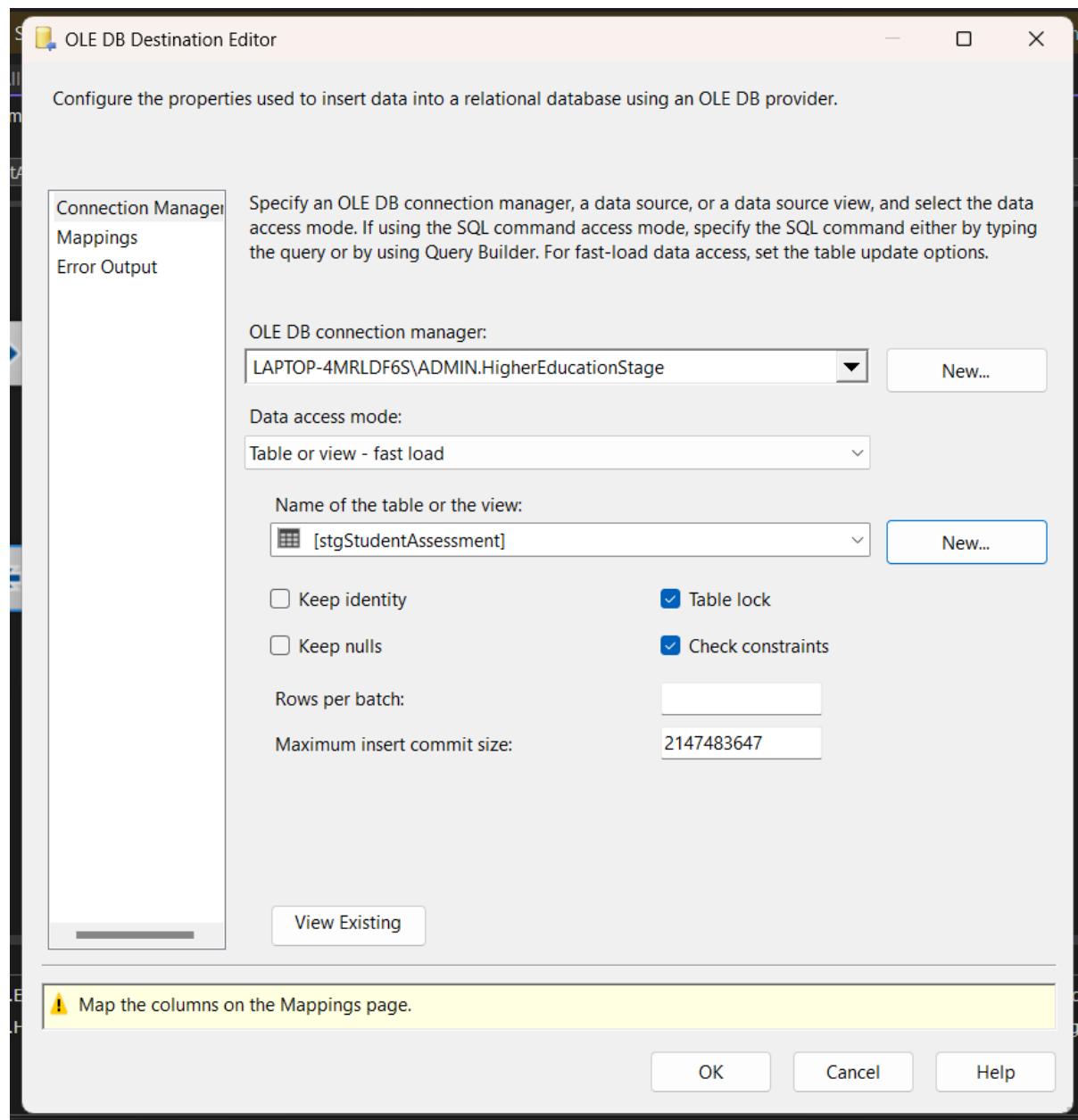
- Double-click **DF - Stage StudentAssessment** task để mở nó trong **Data Flow design surface**:

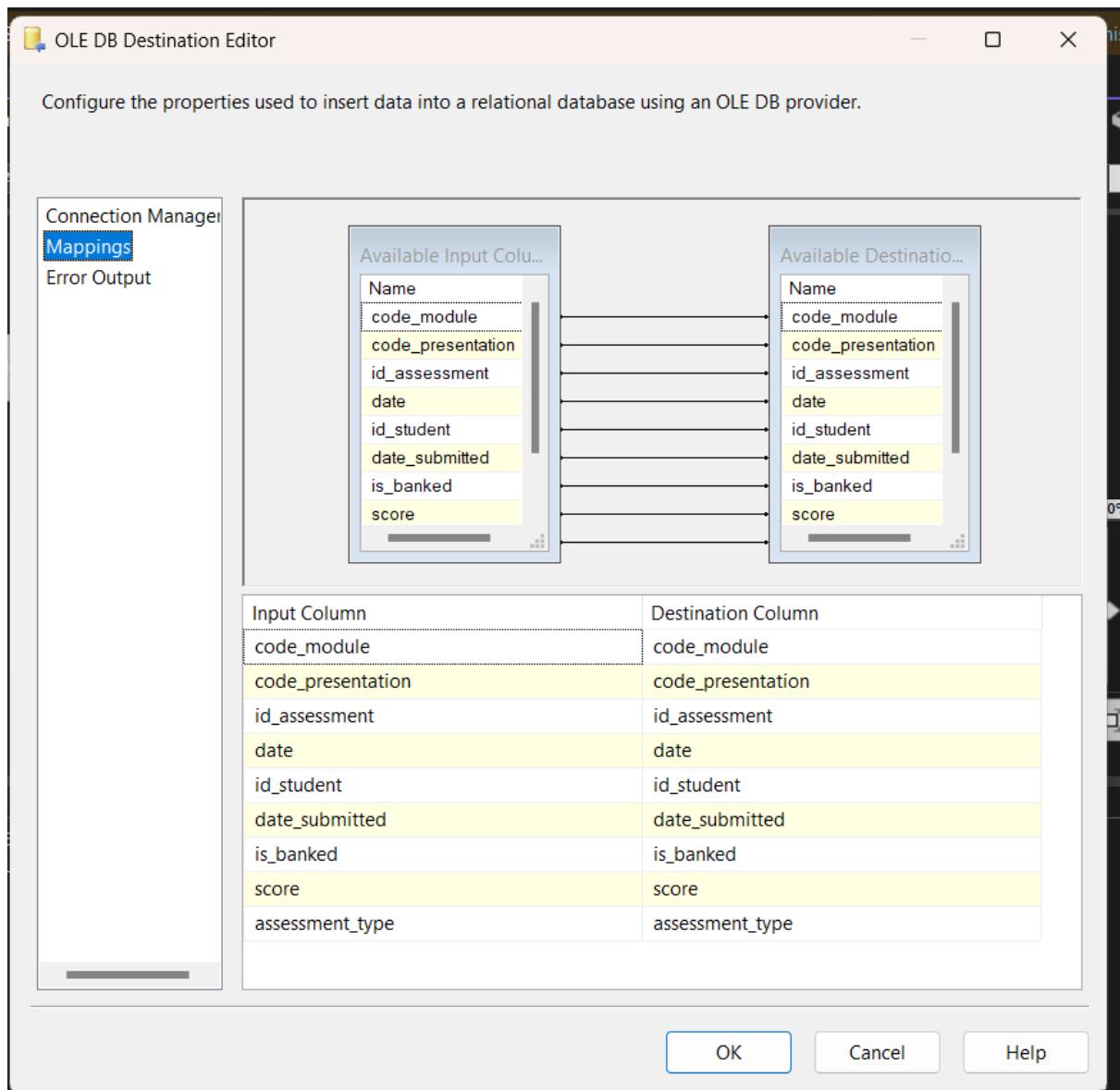
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:

+ Source Assistant:

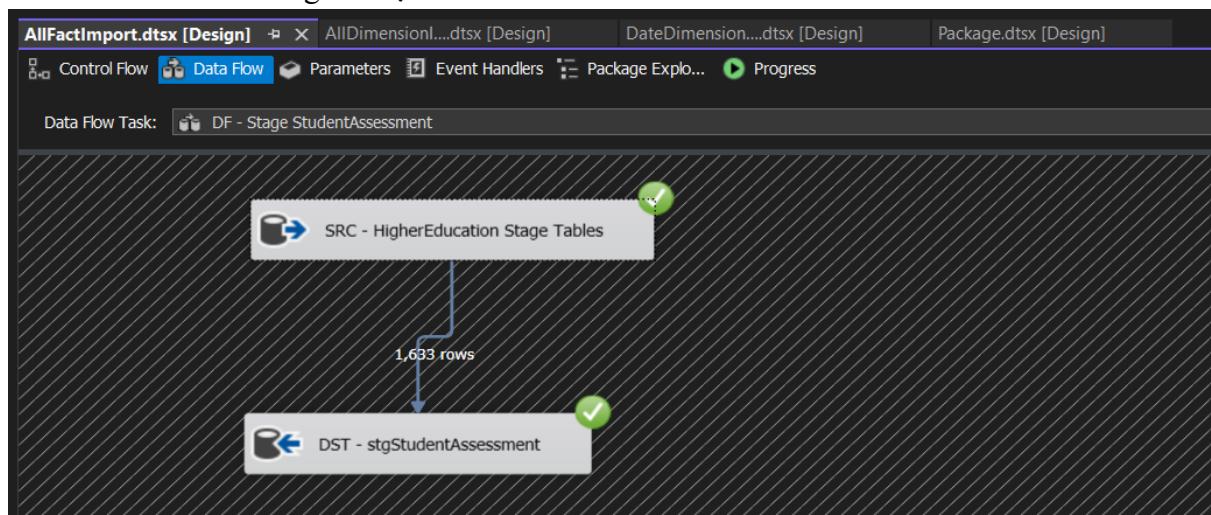


+ Destination Assist



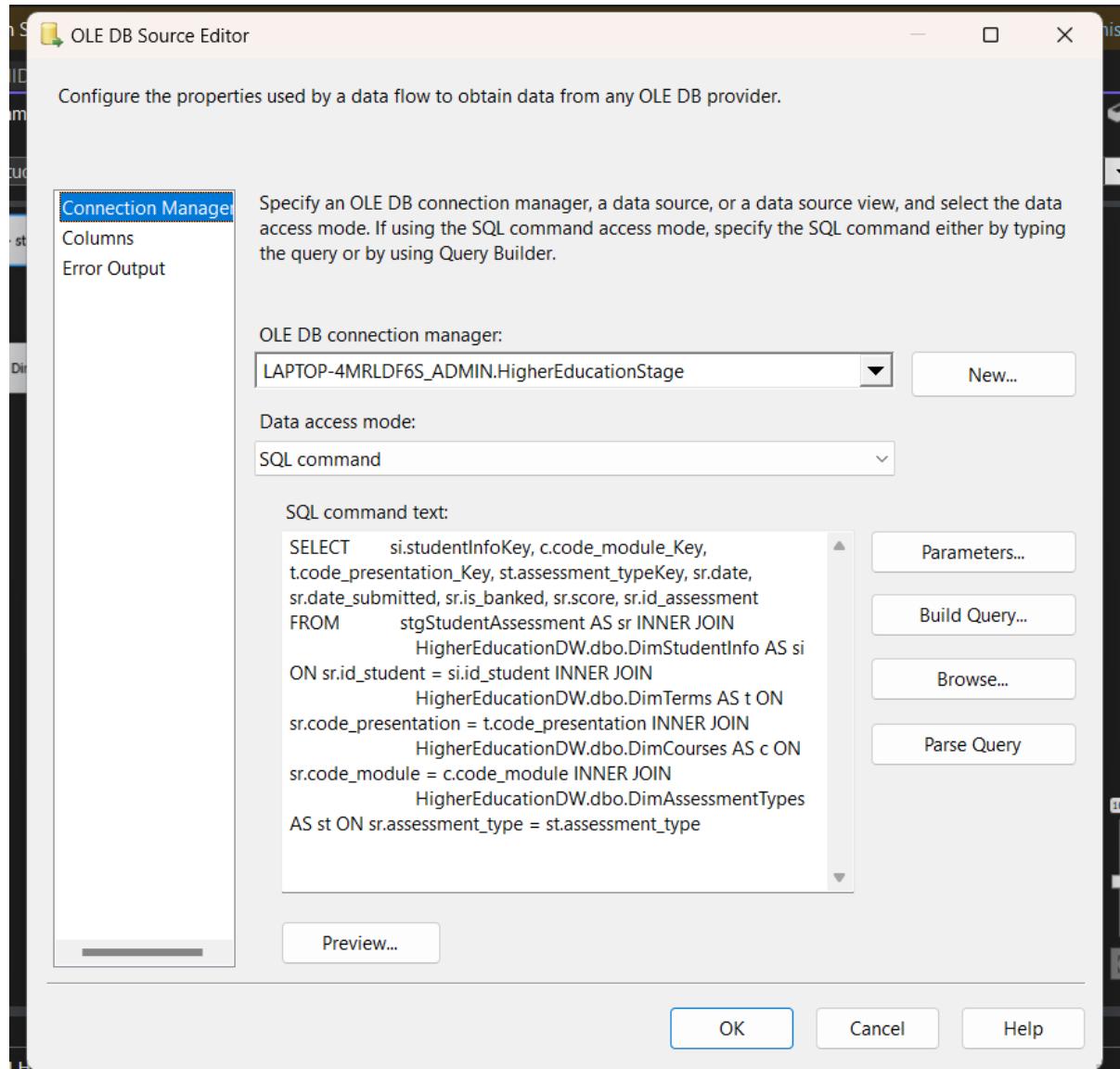


+ Sau khi cấu hình xong ta được:

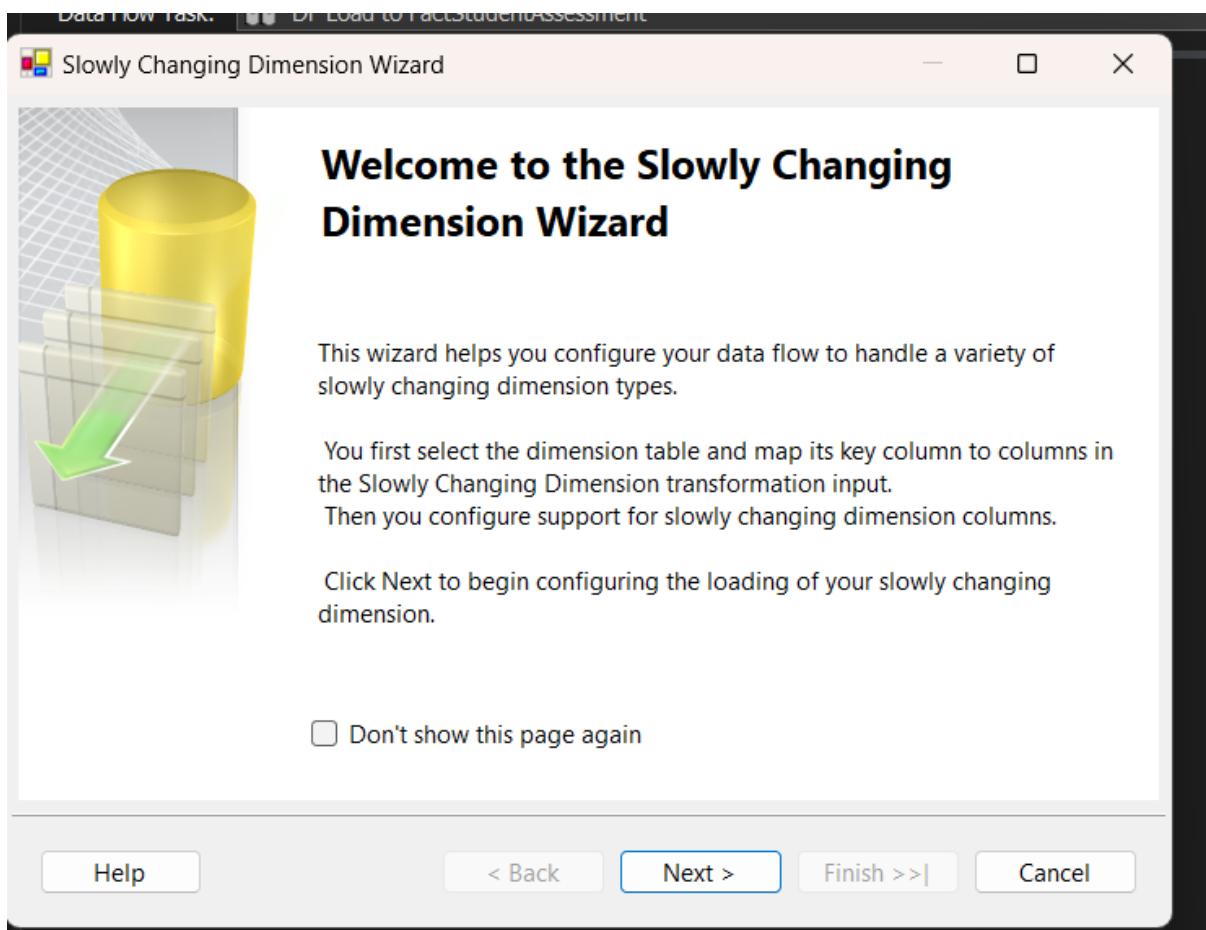


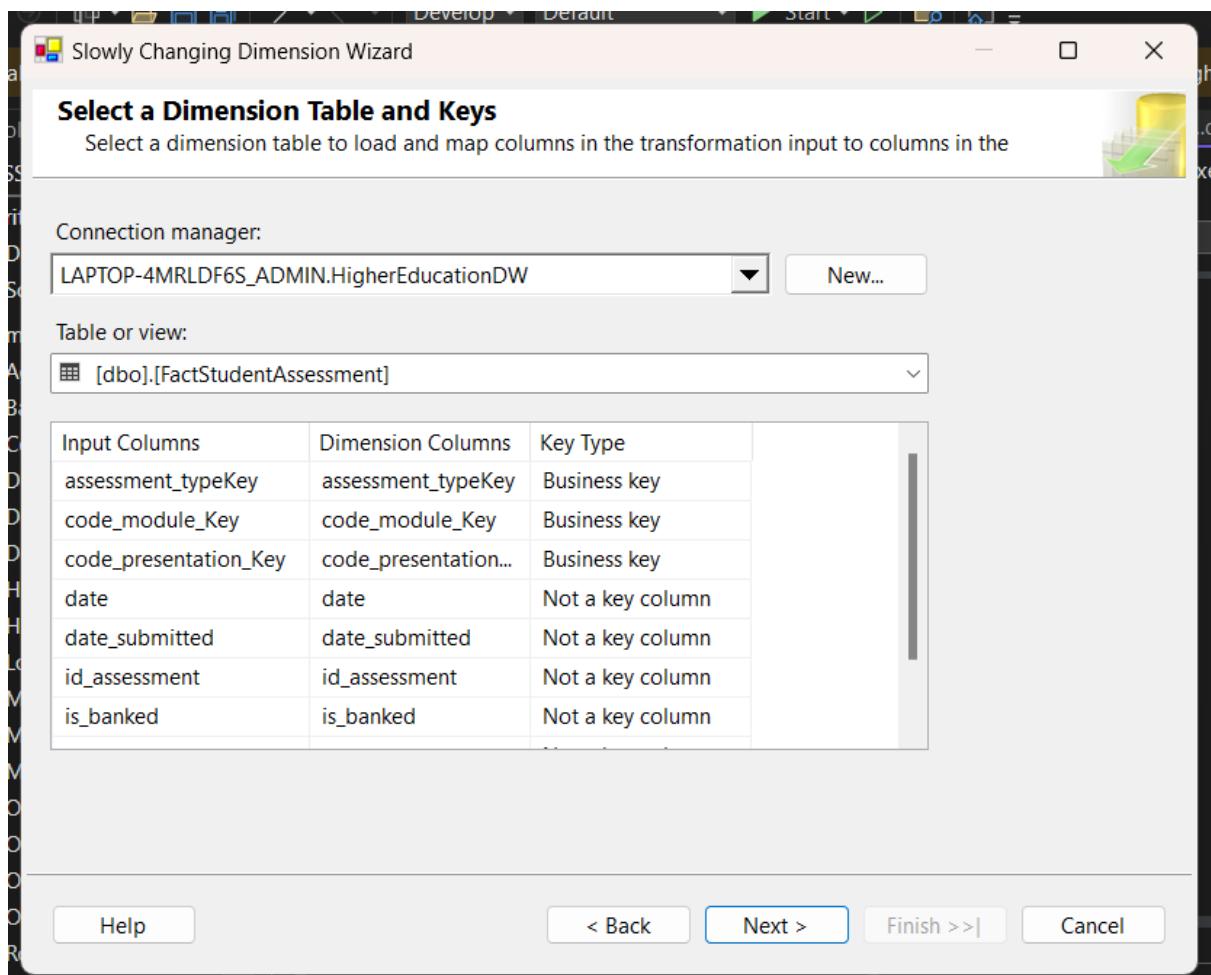
3.4.1.4. Load dữ liệu từ Stage vào Fact

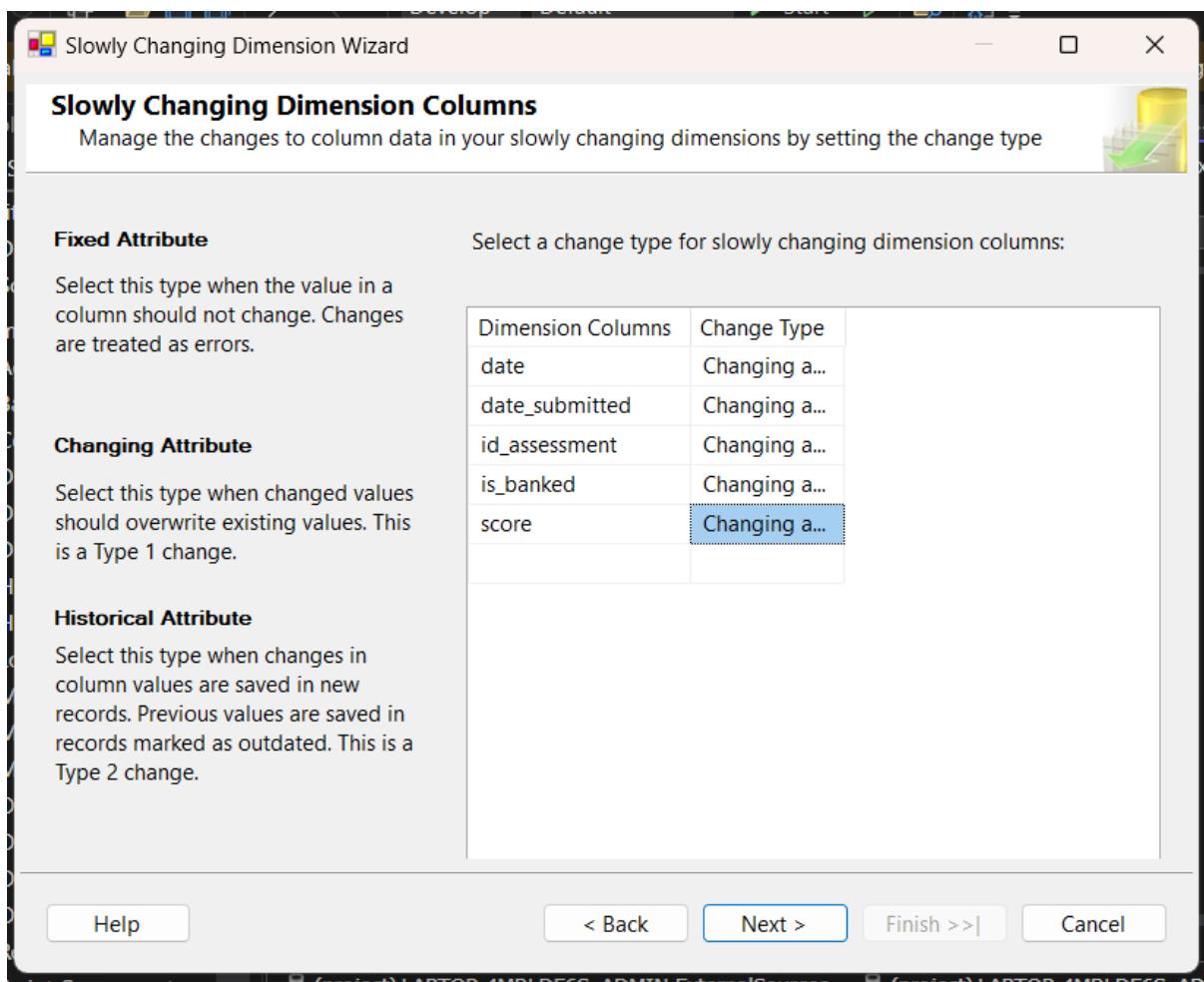
- Double-click vào **DF - Load to FactStudentAssessment** task để mở data flow design Surface và thiết lập các cấu hình sau
 - + Source Assistant:

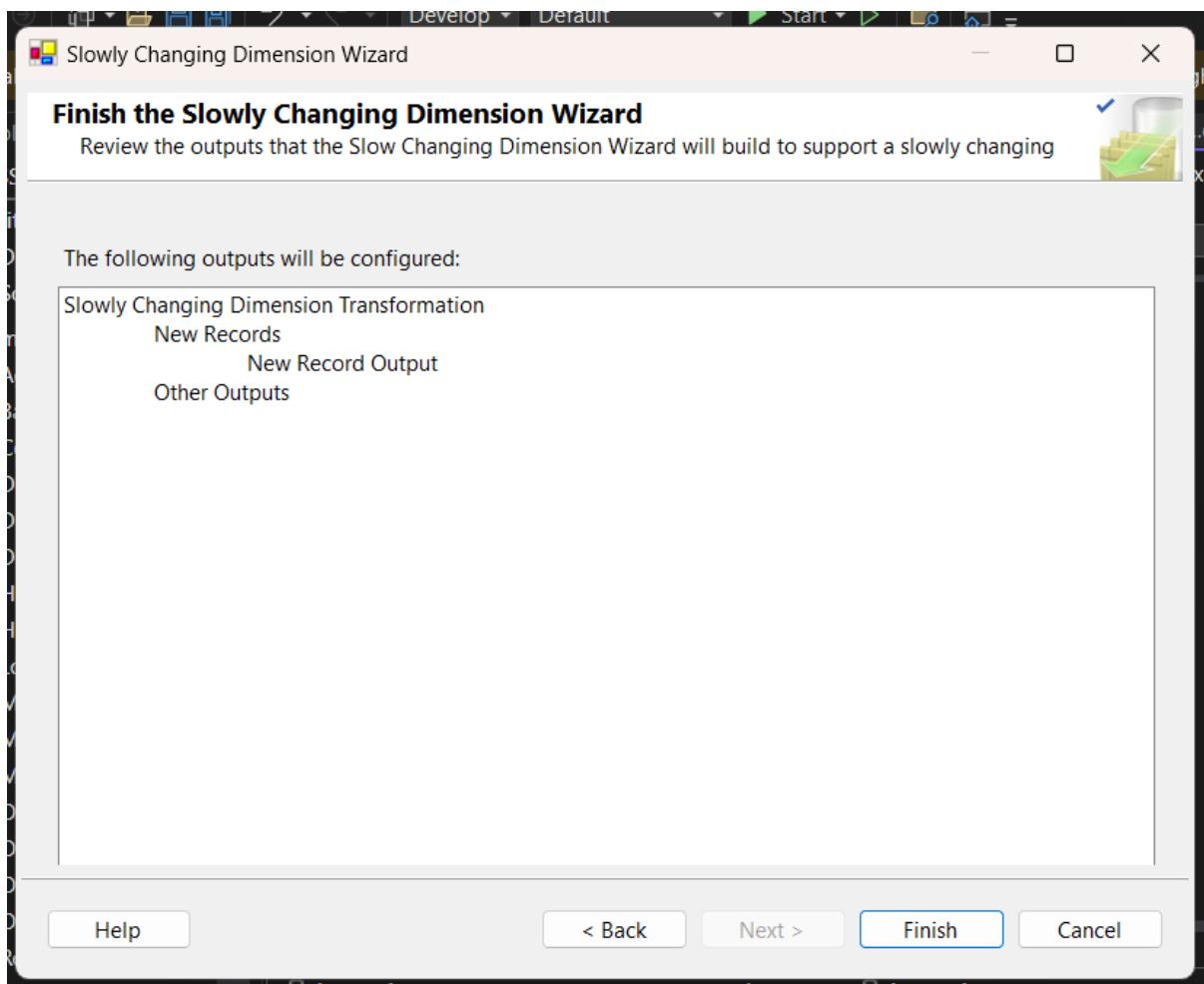


- + Slowly Changing Dimension:

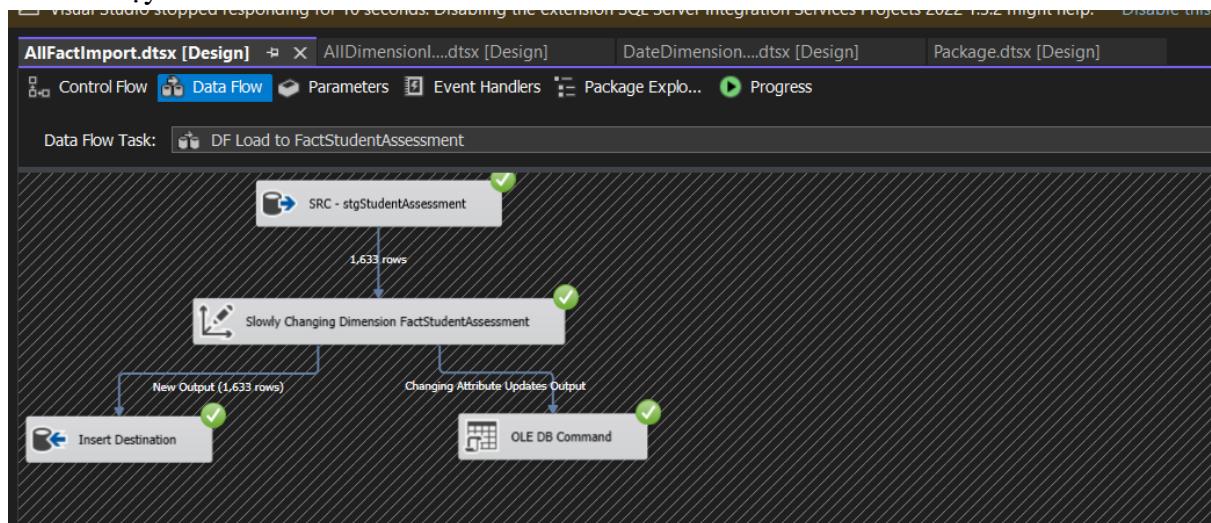








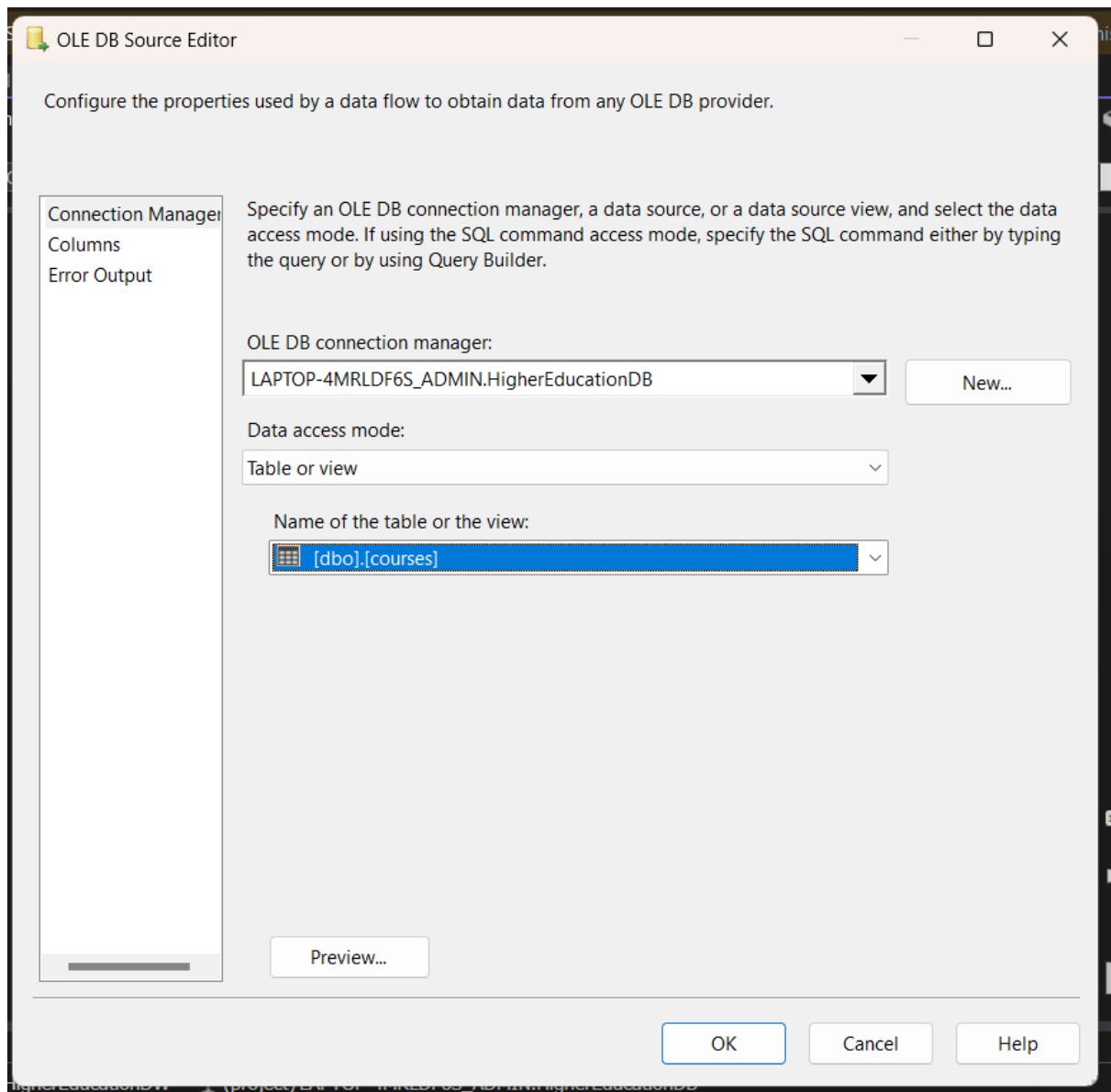
- Khởi chạy



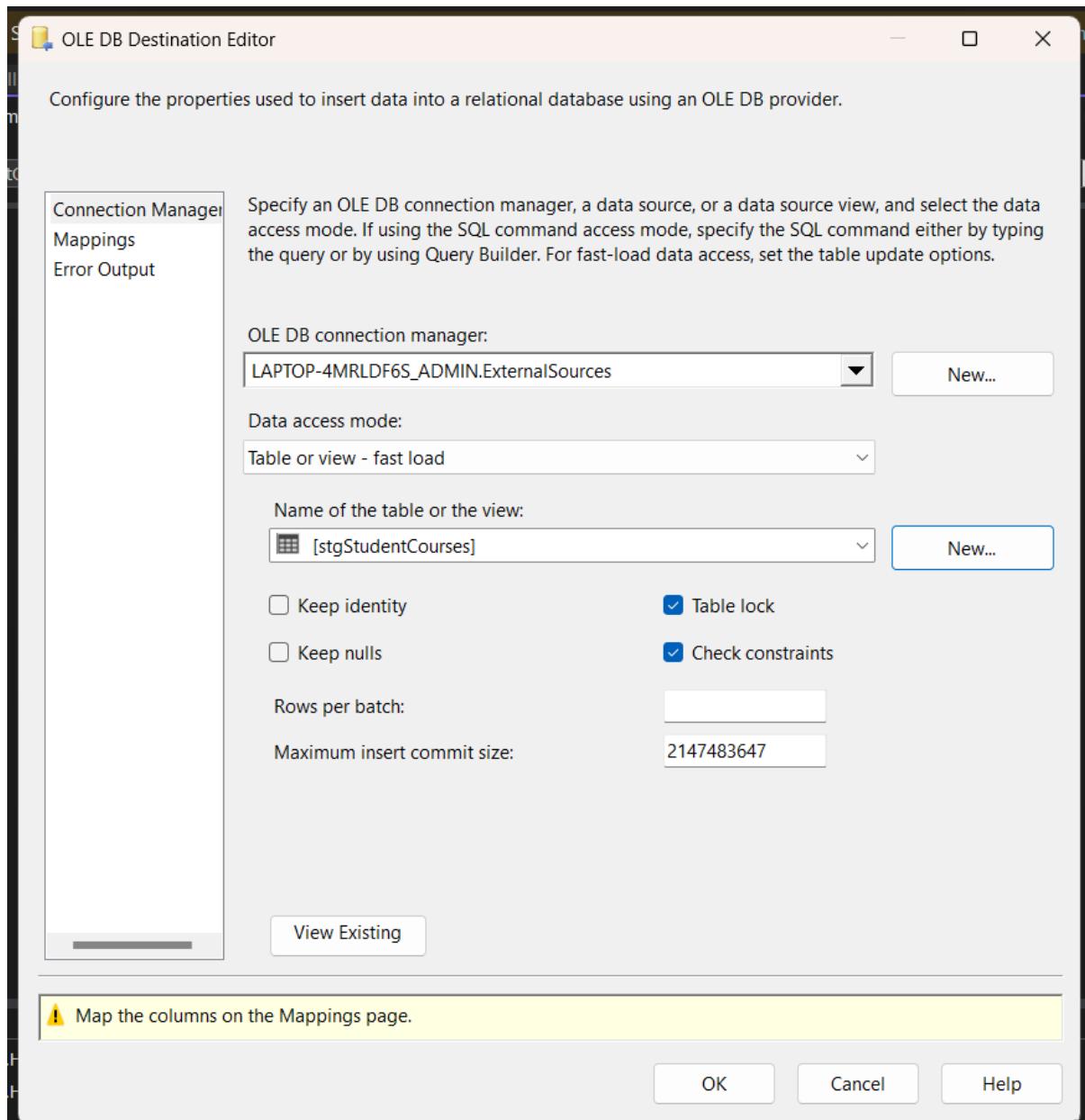
3.4.2. Fact Student Courses

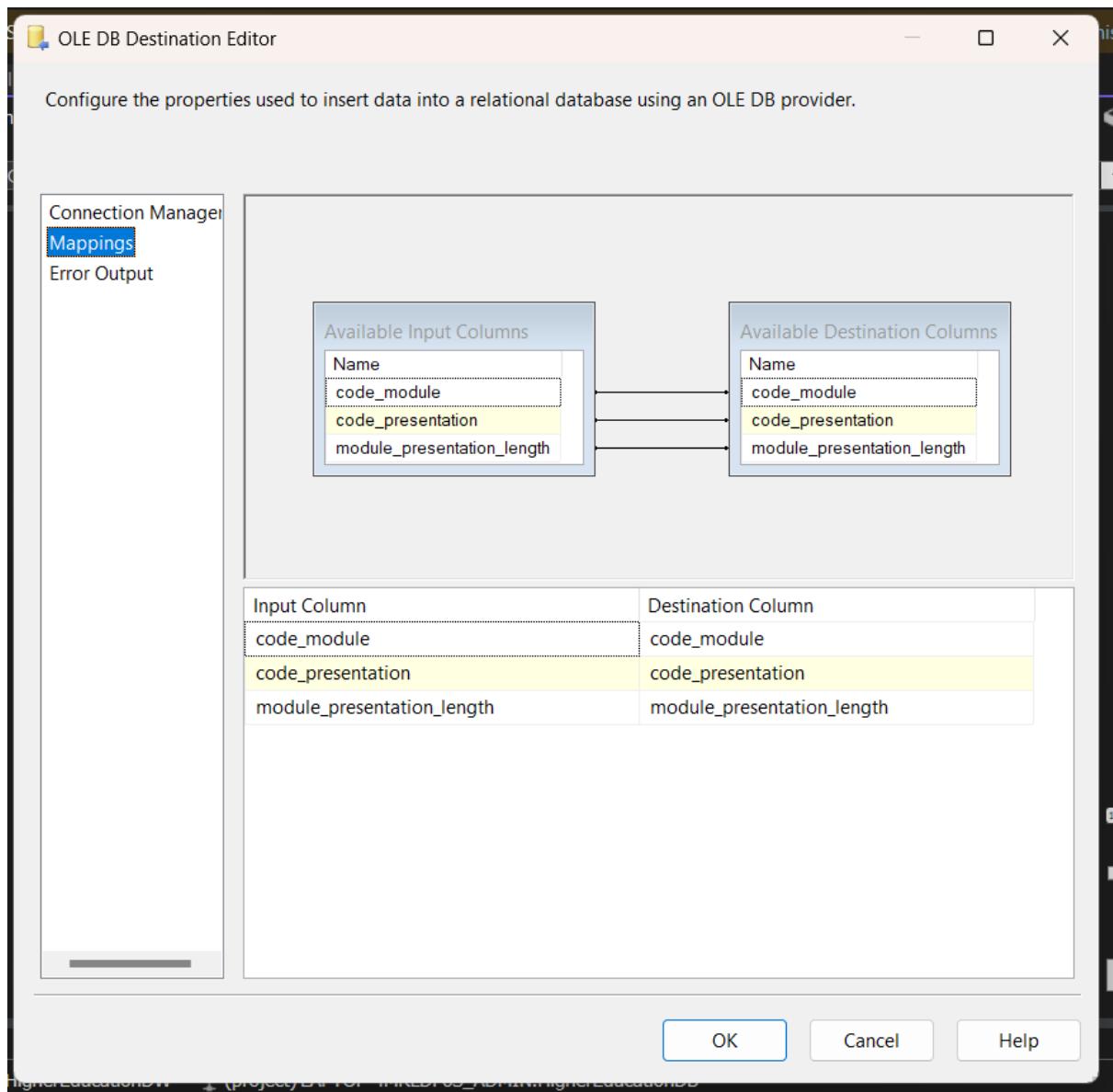
3.4.2.1. Load dữ liệu từ Source vào Stage

- Double-click **DF - Stage StudentCourses** task để mở nó trong **Data Flow design surface**:
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:
 - + Source Assistant:

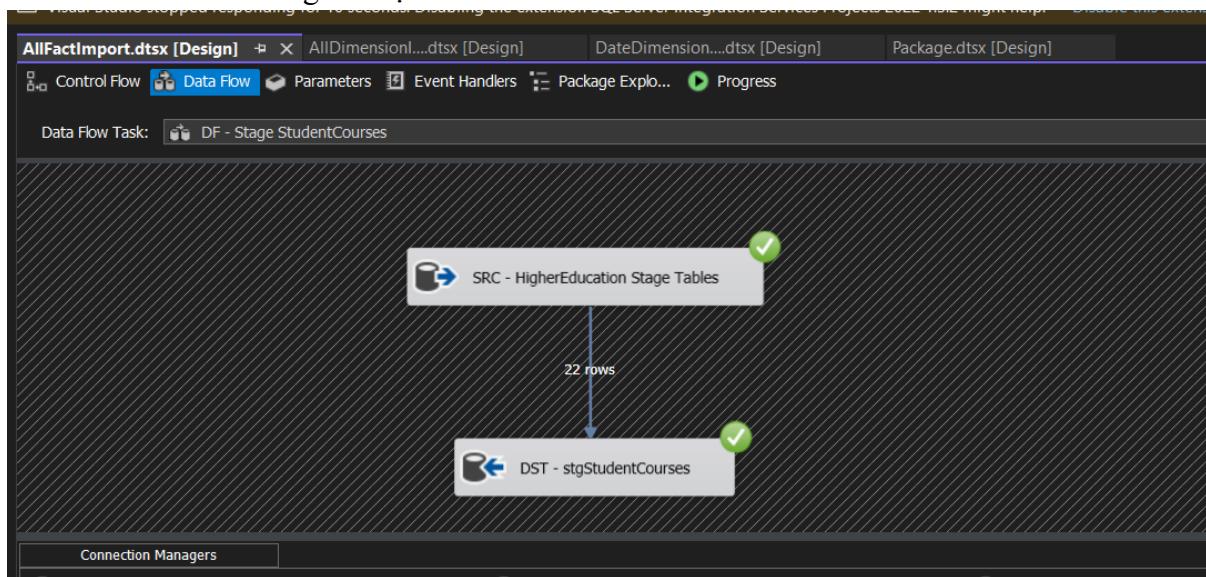


+ Destination Assist





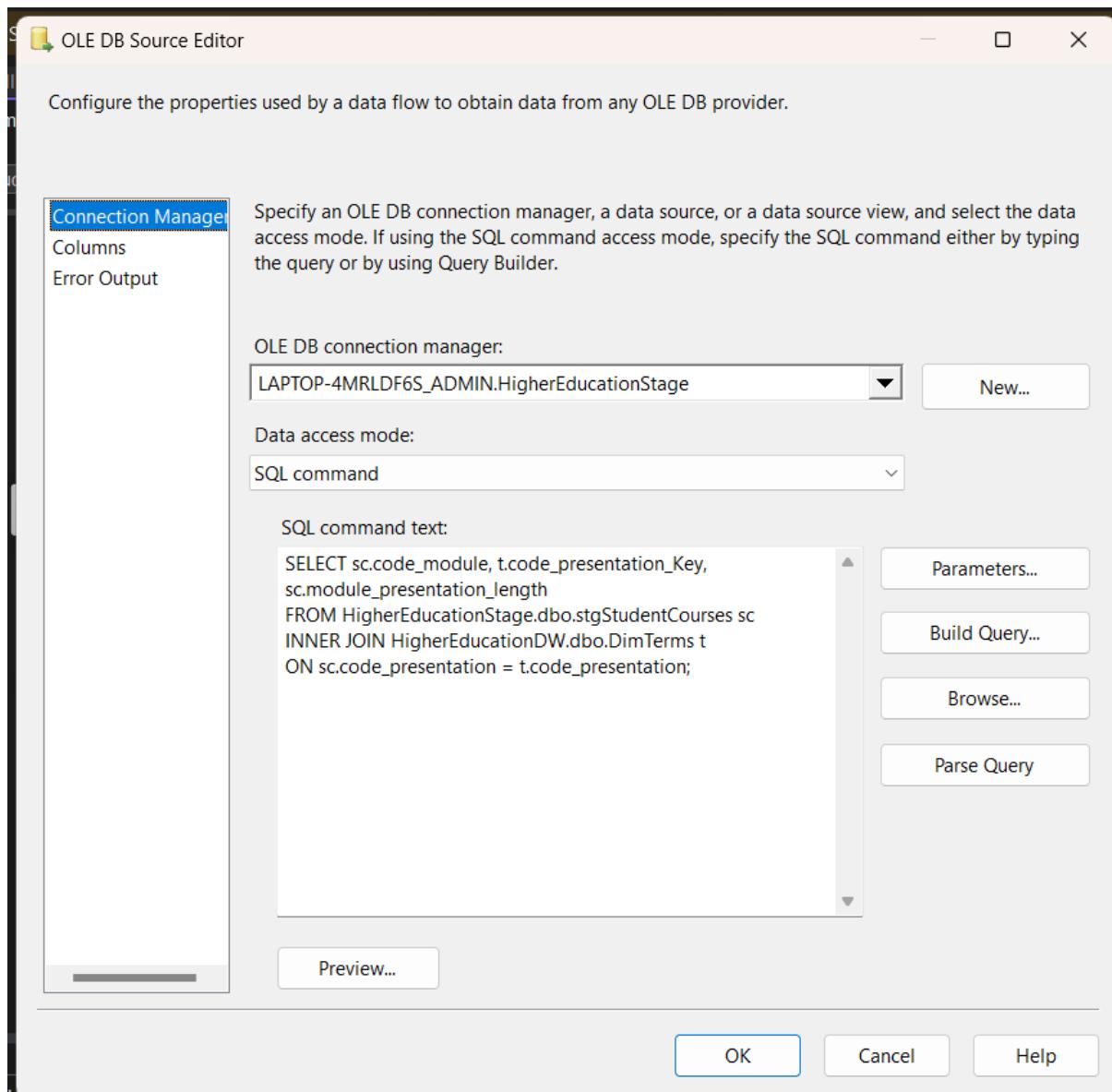
+ Sau khi cấu hình xong ta được:



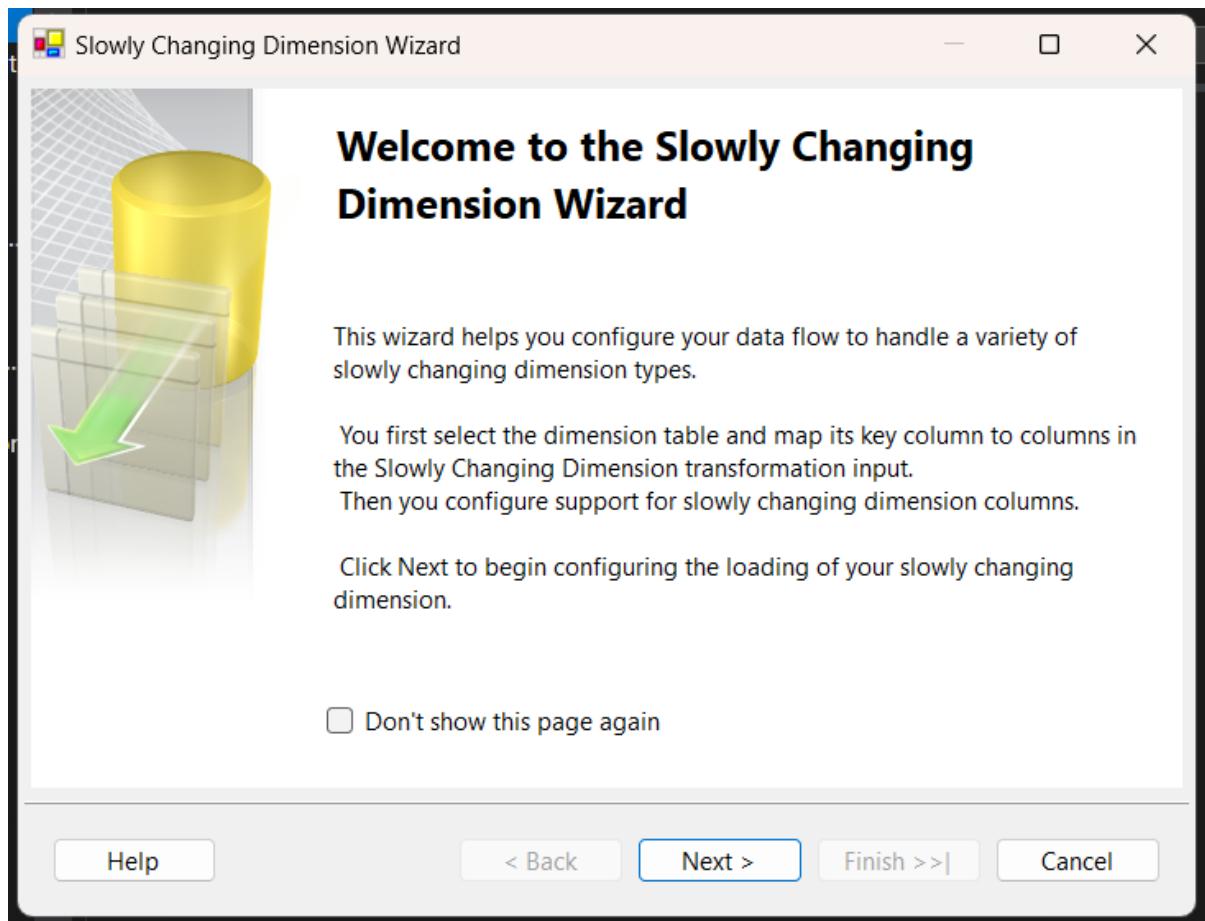
3.4.2.2. Load dữ liệu từ Stage vào Fact

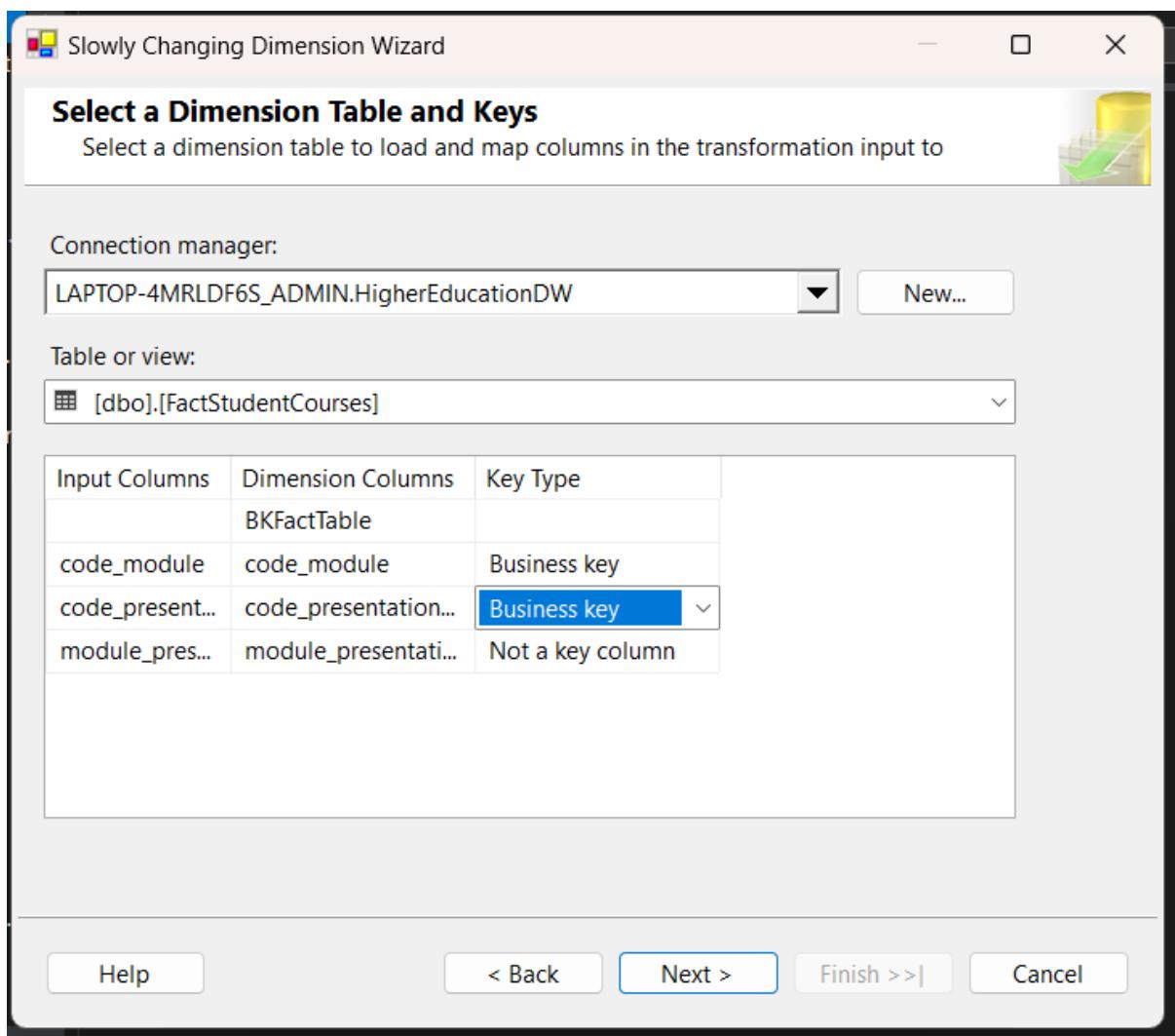
- Double-click vào **DF - Load to FactStudentCourses** task để mở data flow design Surface và thiết lập các cấu hình sau

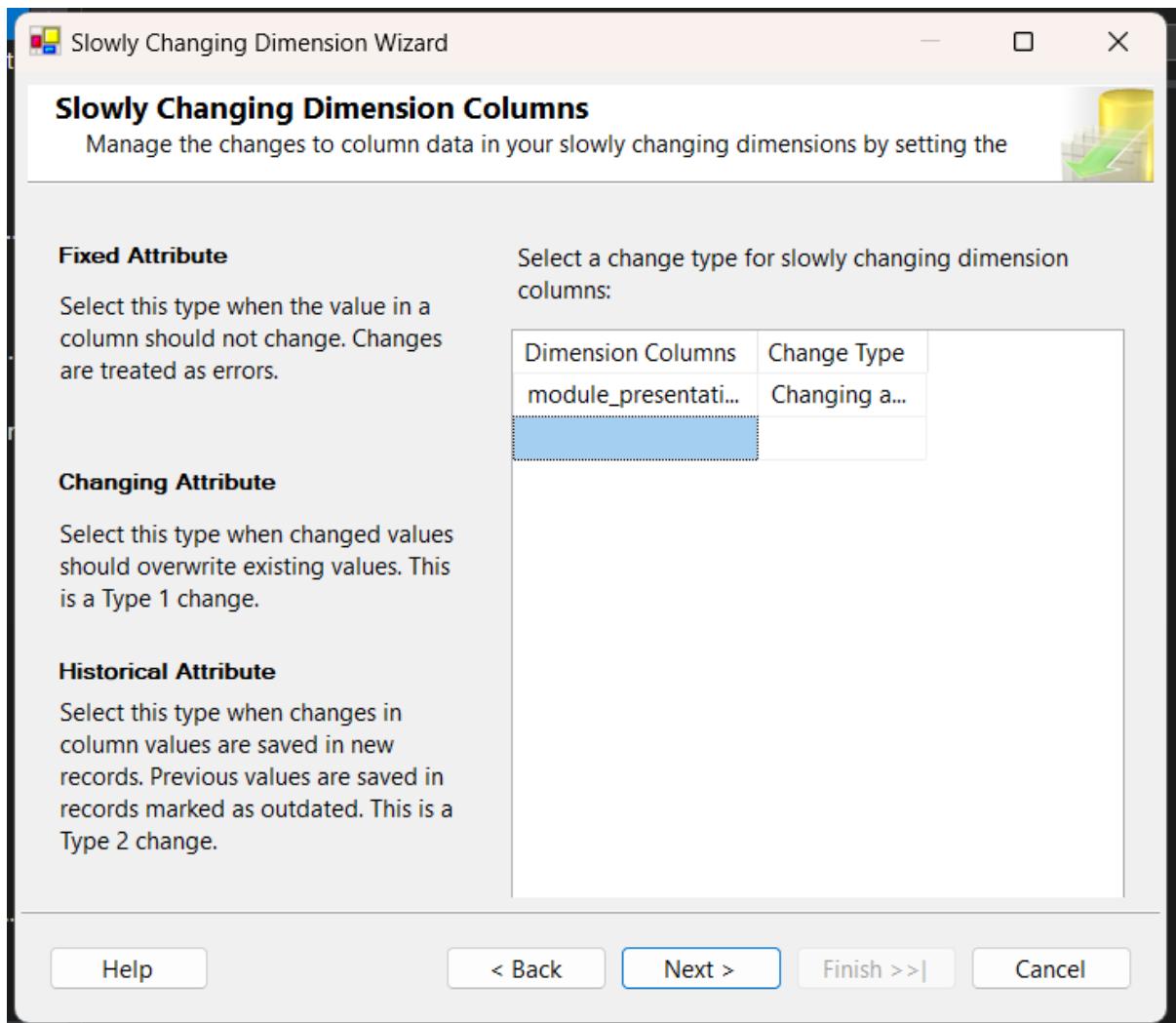
+ Source Assistant:

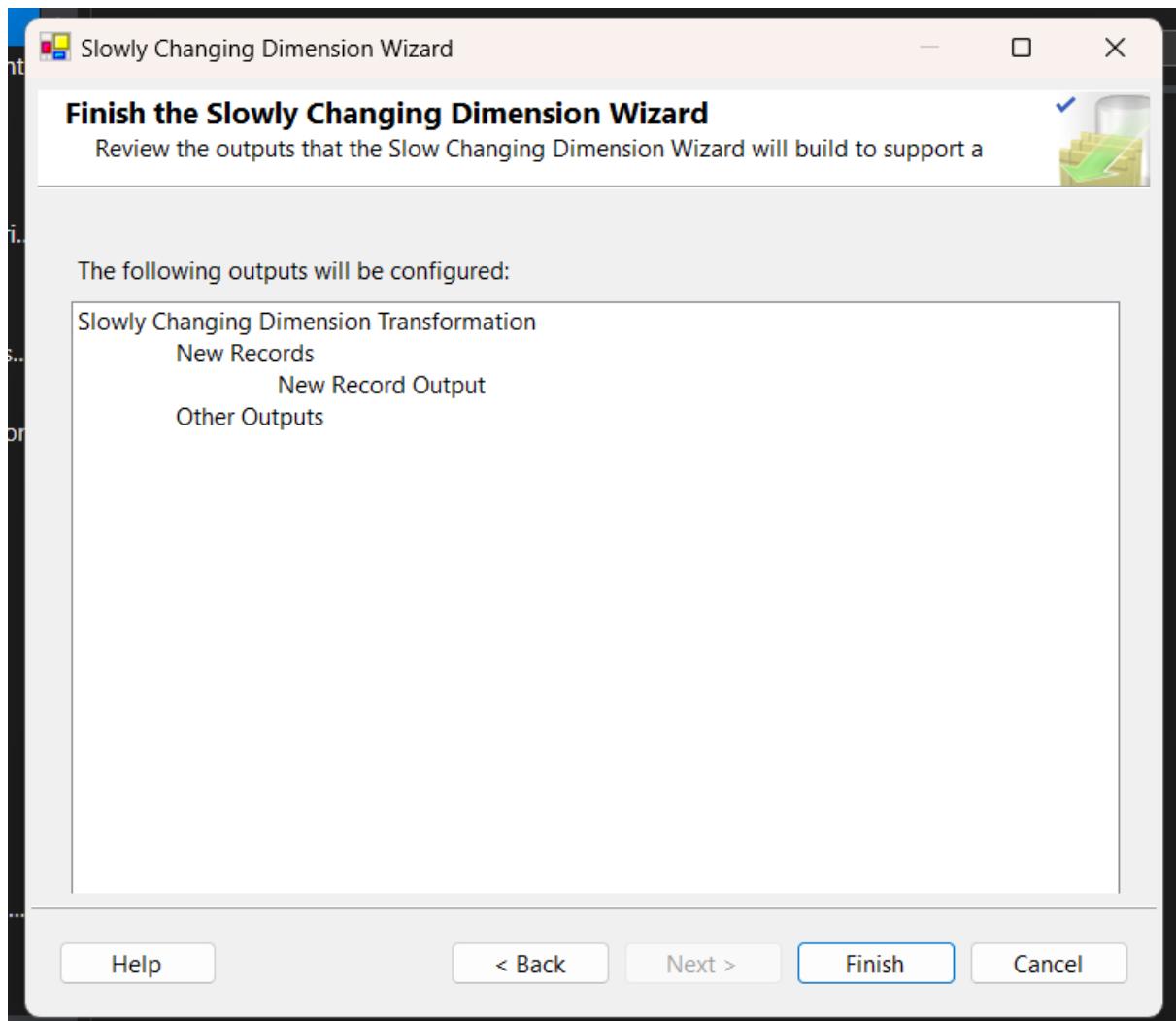


+ Slowly Changing Dimension:

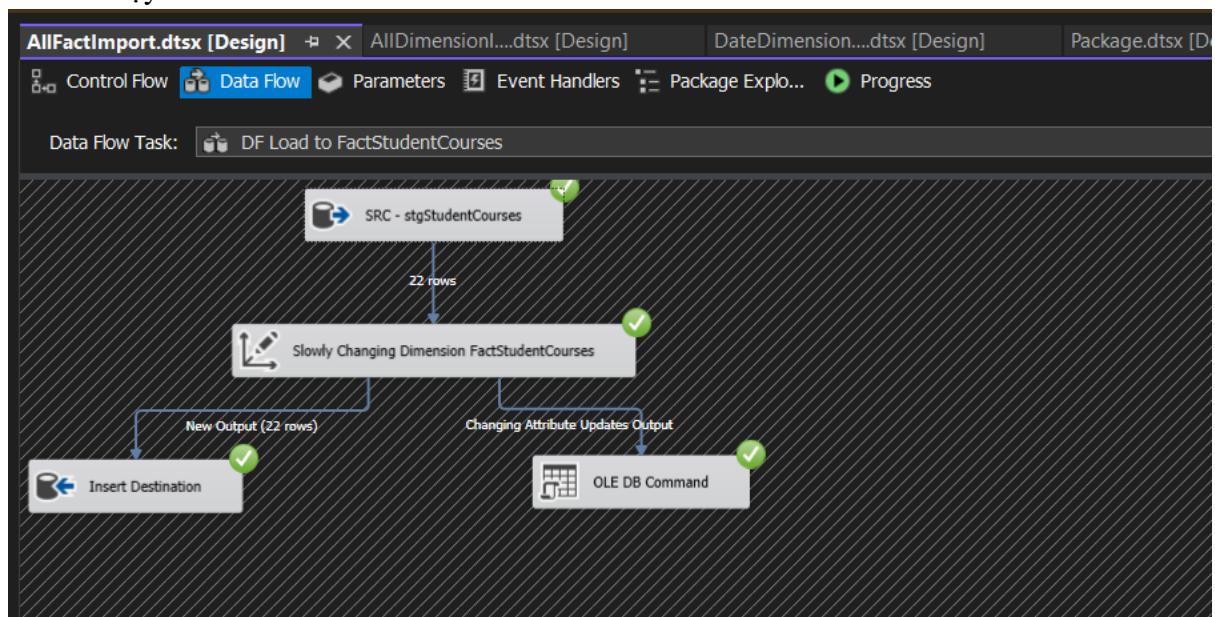








- Khởi chạy



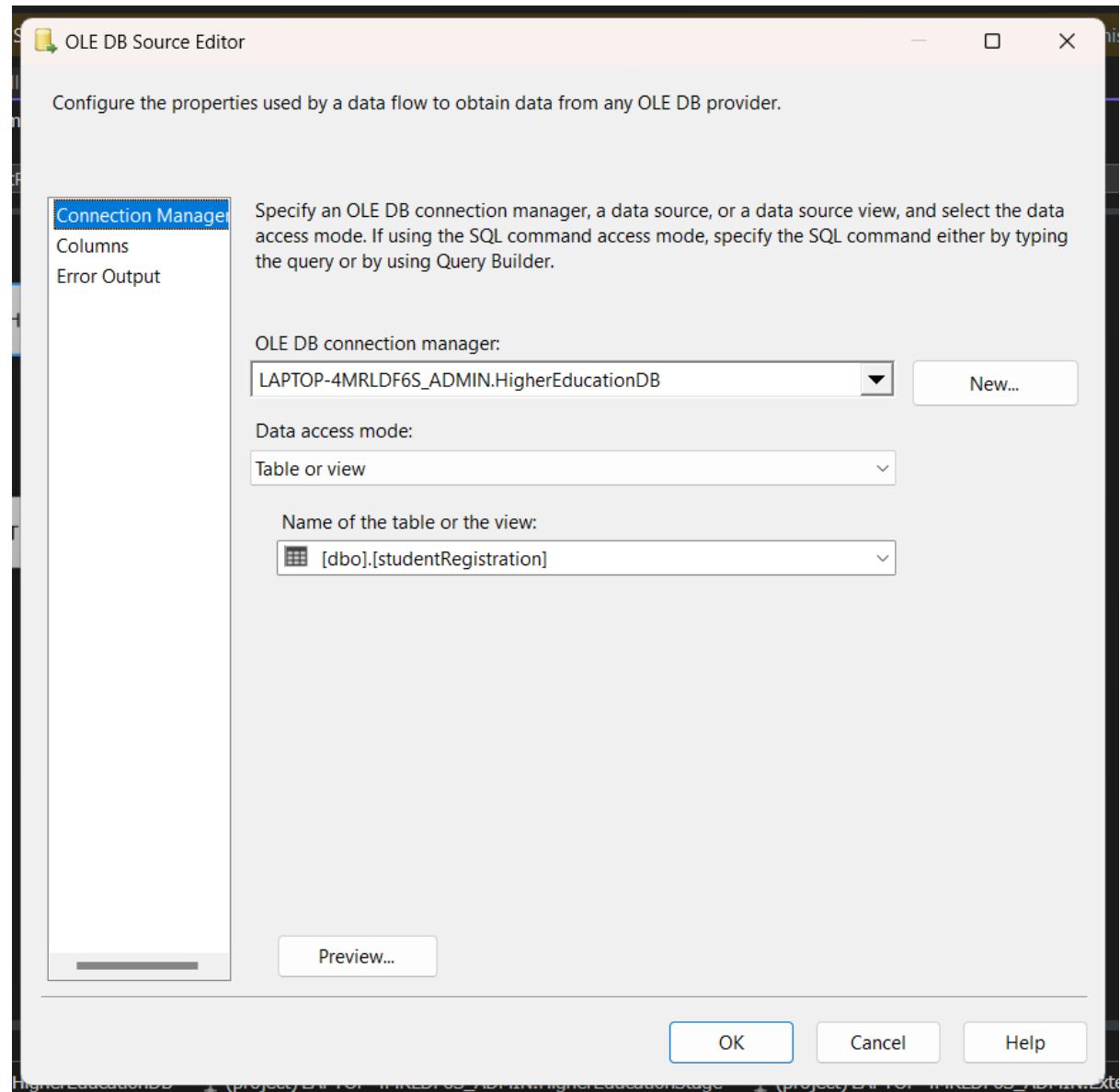
3.4.3. Fact Student Registration

3.4.3.1. Load dữ liệu từ Source vào Stage

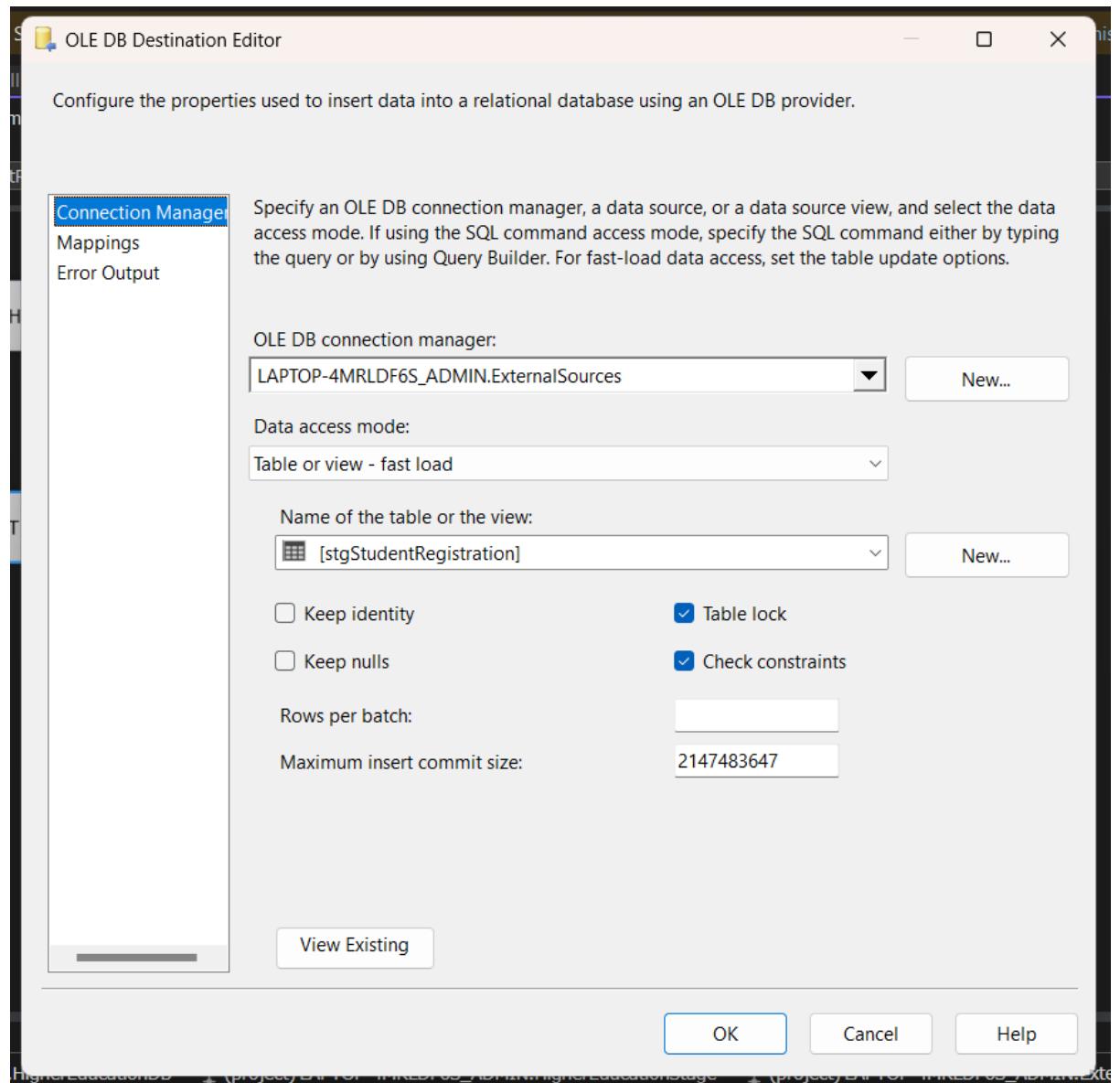
- Double-click **DF - Stage Student Registration** task để mở nó trong **Data Flow design surface**:

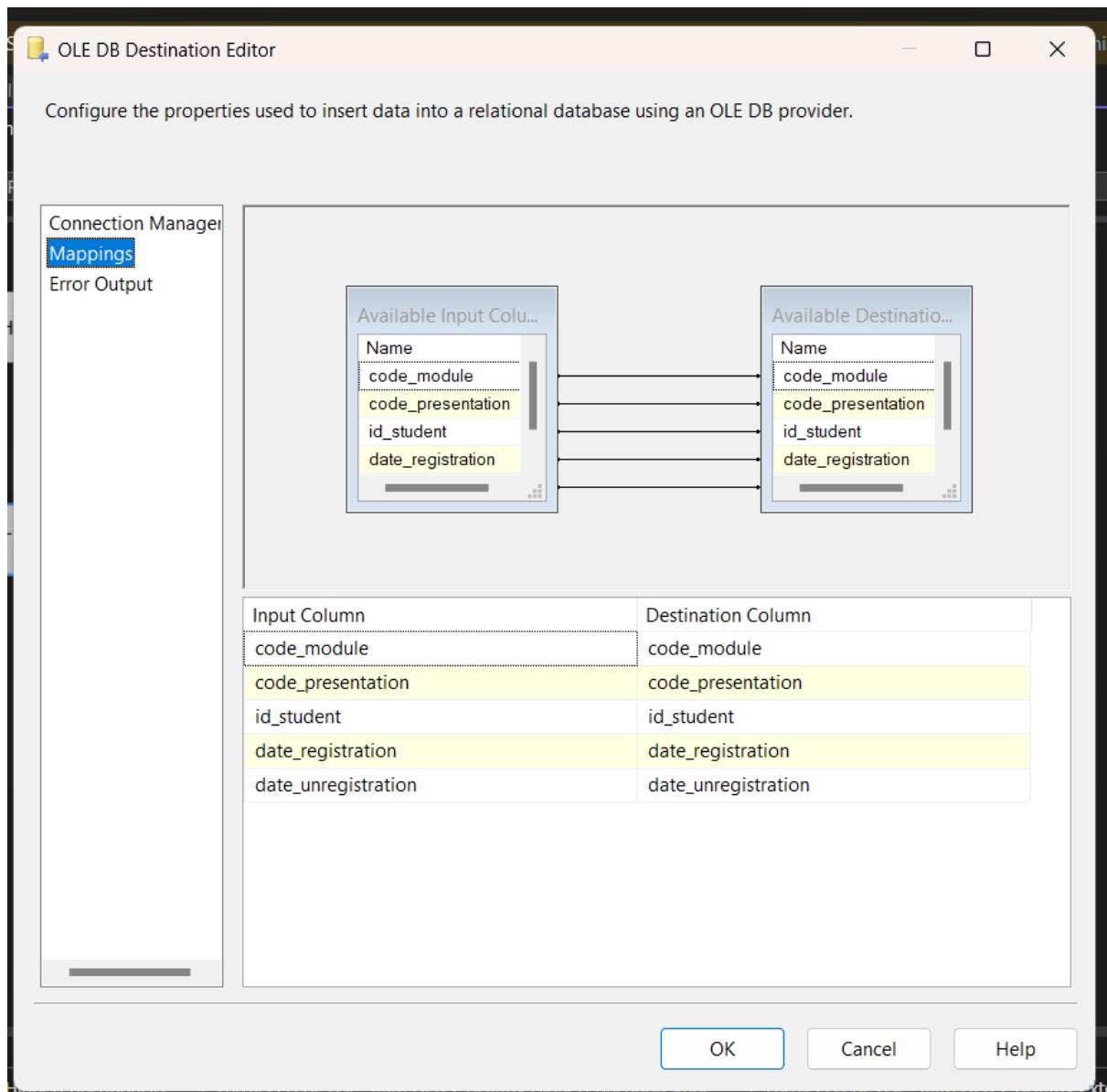
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:

+ Source Assistant:

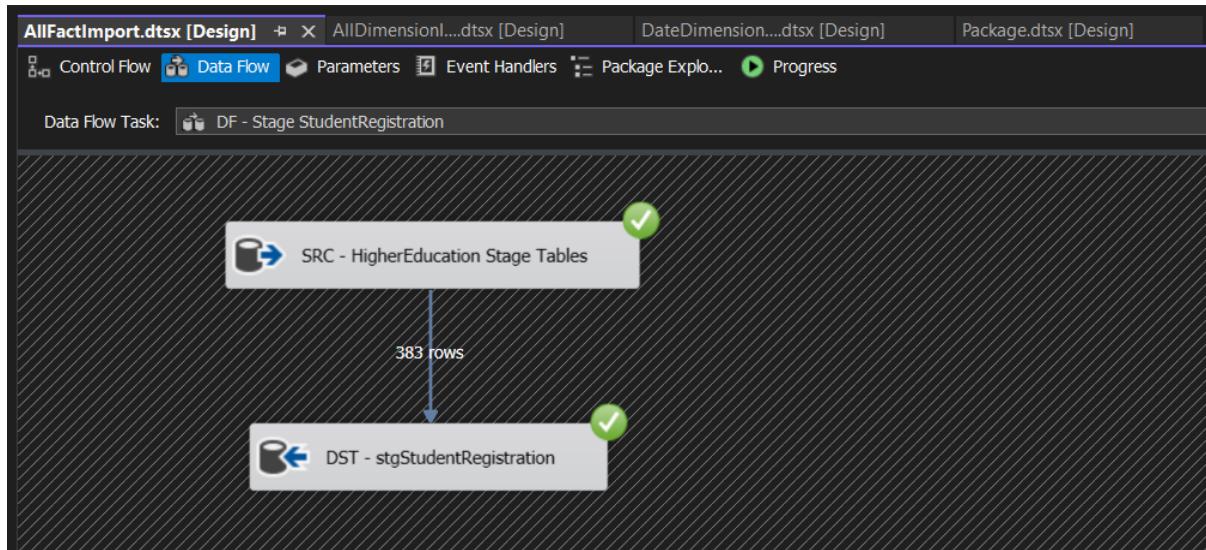


+ Destination Assist



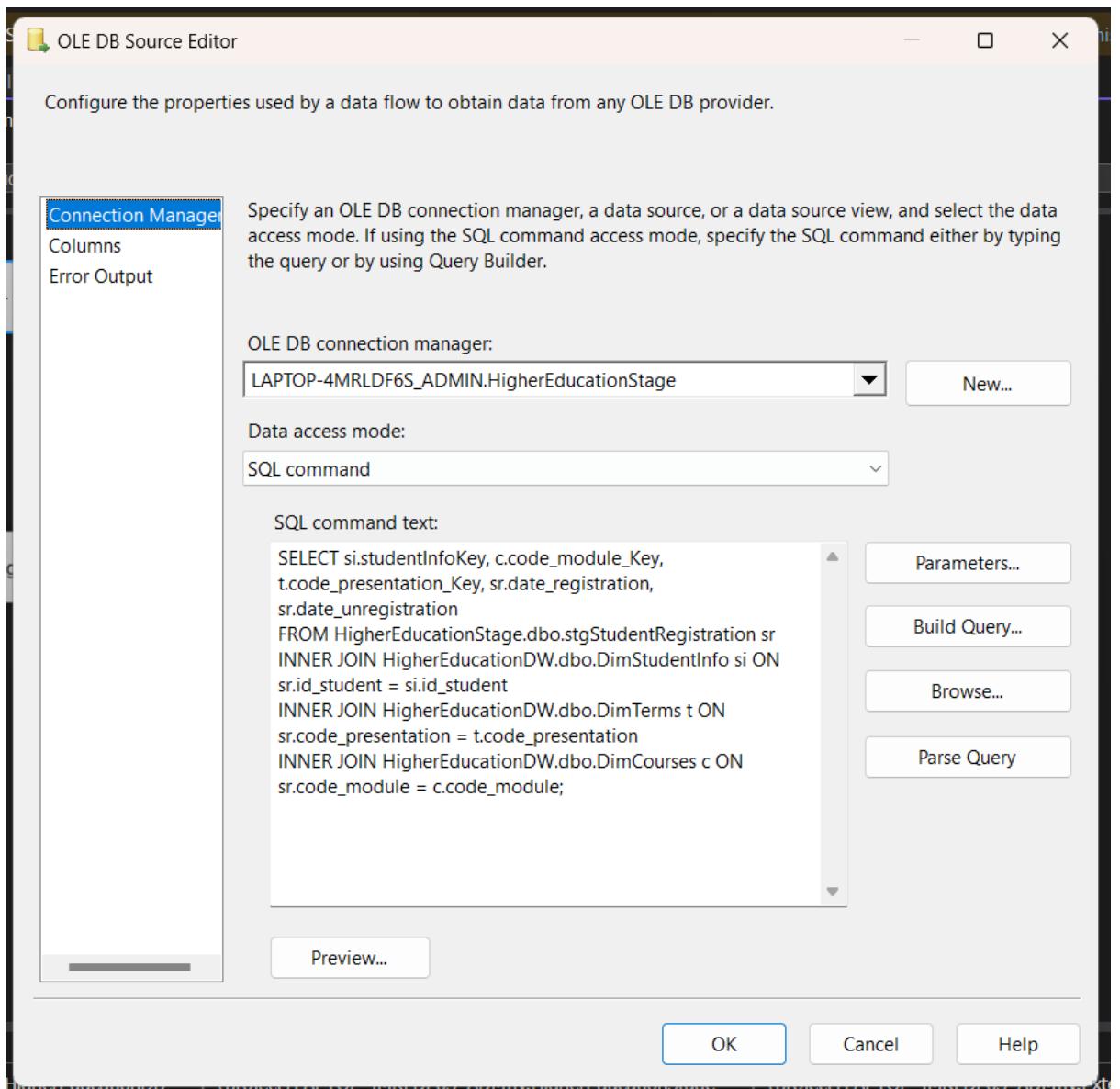


+ Sau khi cấu hình xong ta được:

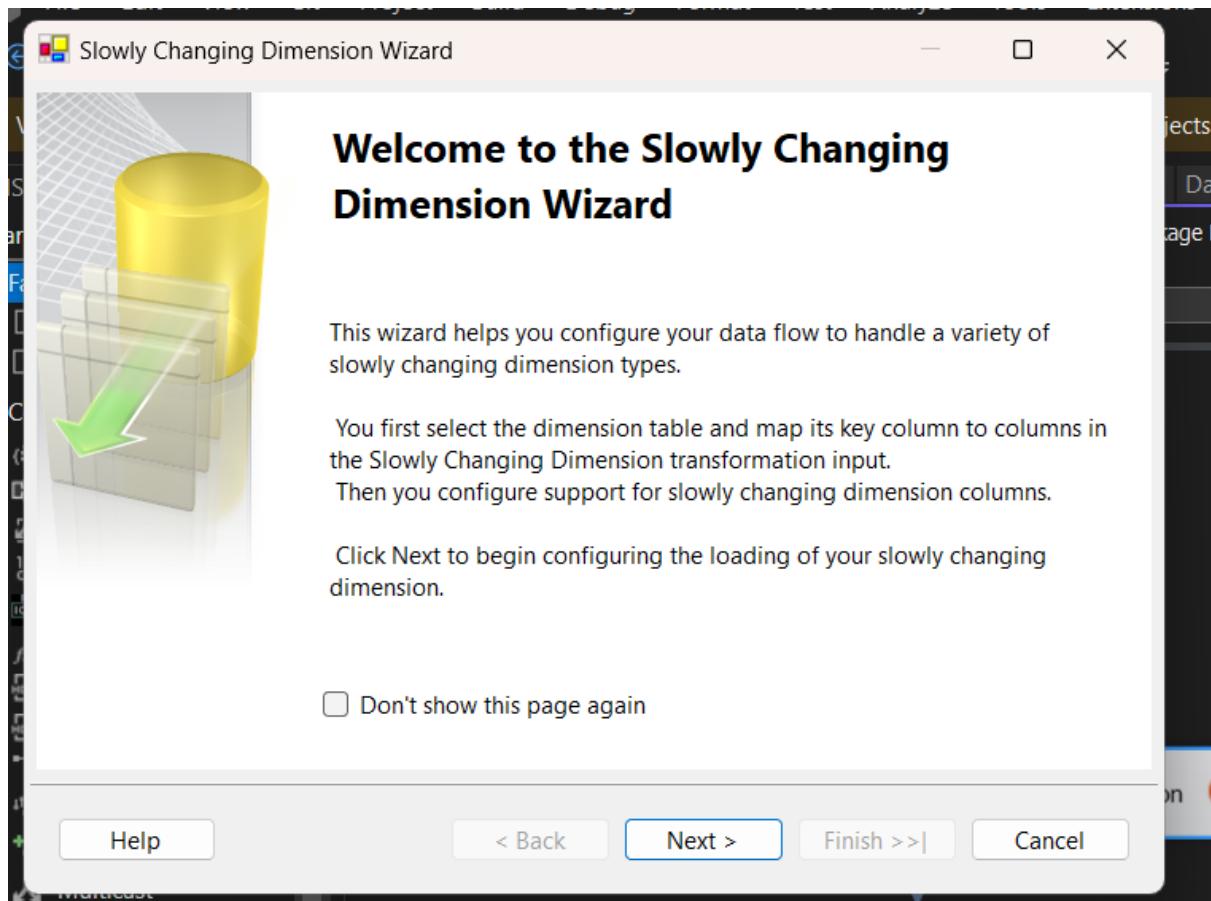


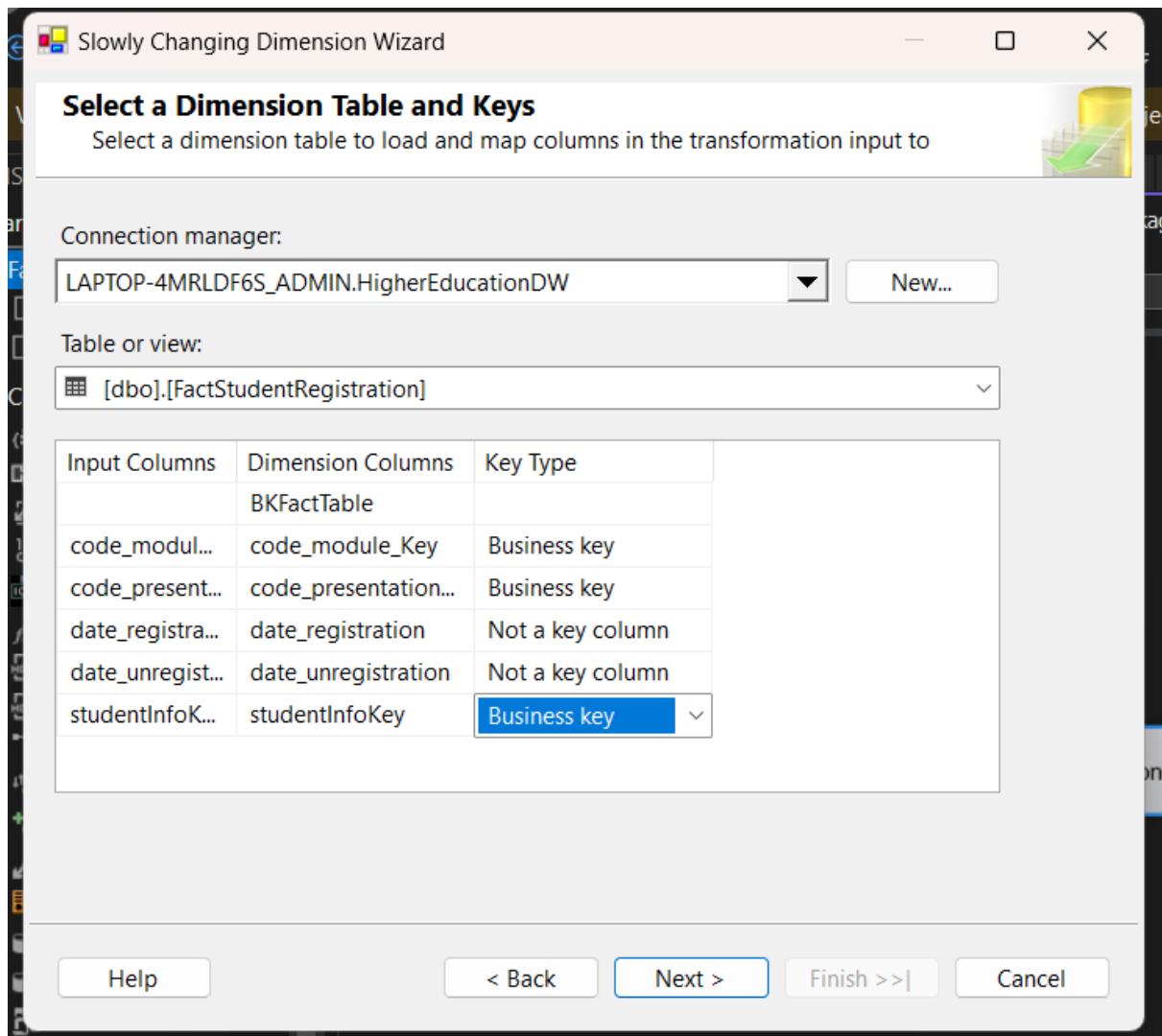
3.4.3.2. Load dữ liệu từ Stage vào Fact

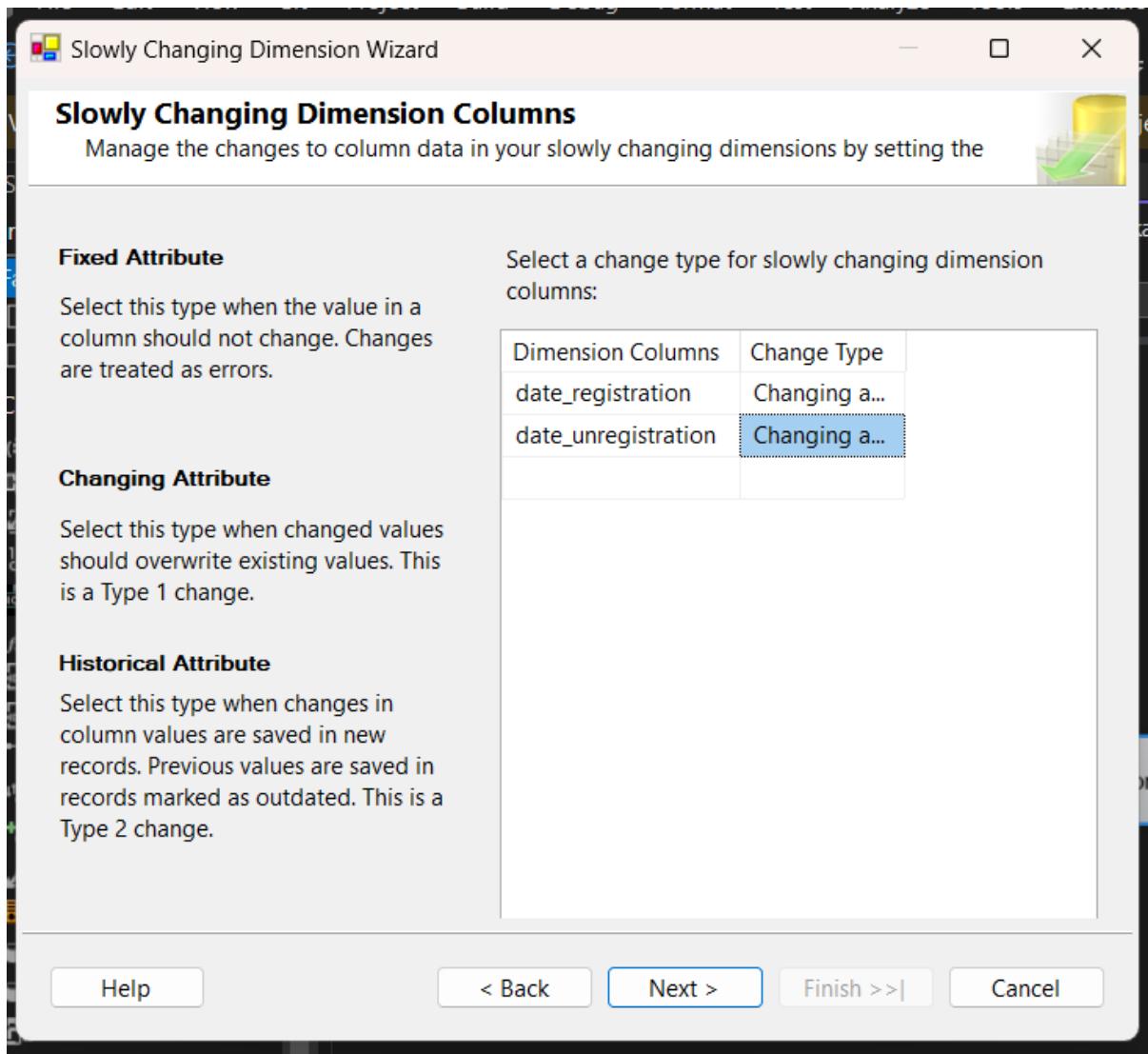
- Double-click vào **DF - Load to FactStudentRegistration** task để mở data flow design Surface và thiết lập các cấu hình sau
 - + Source Assistant:

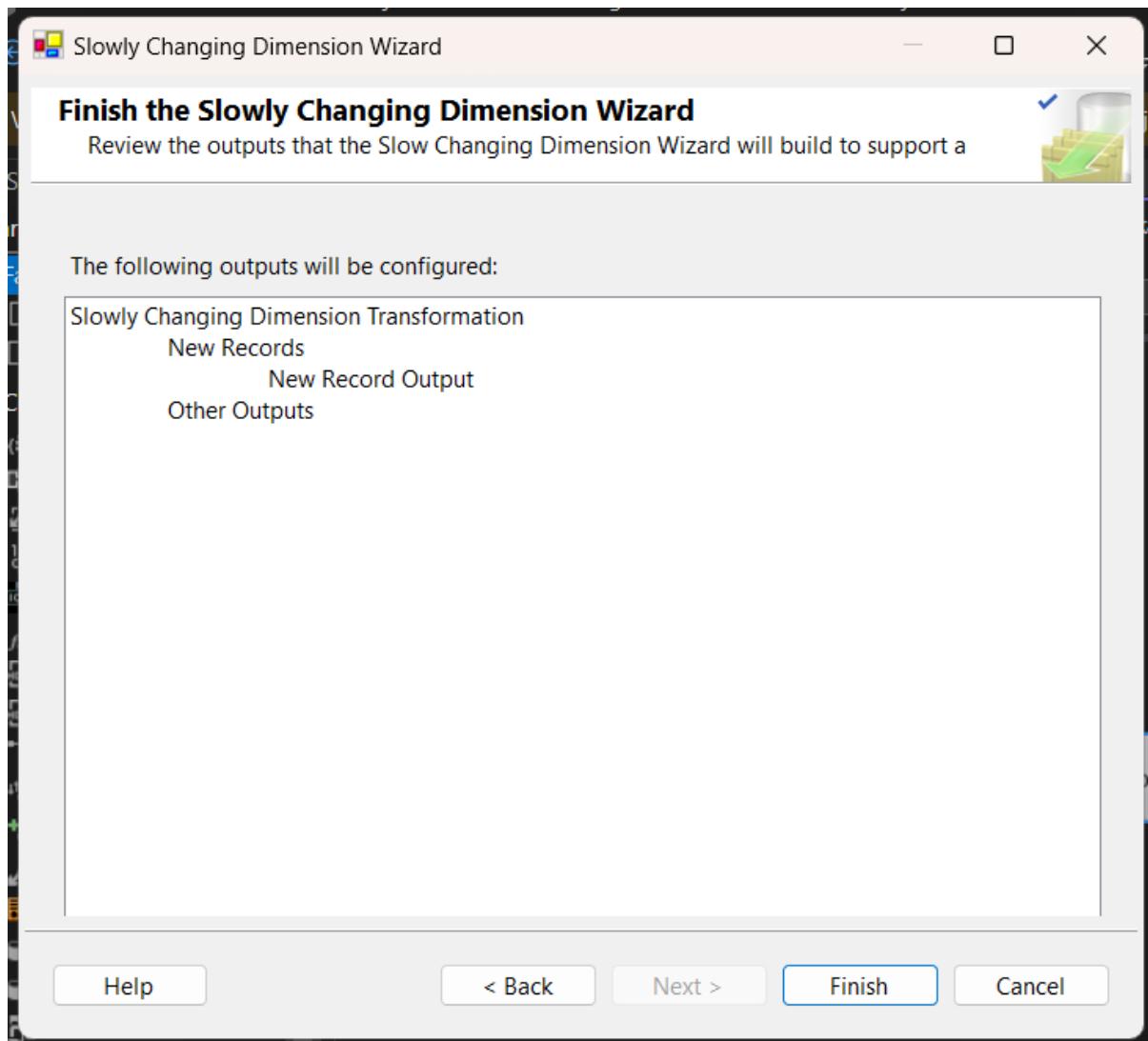


- + Slowly Changing Dimension:

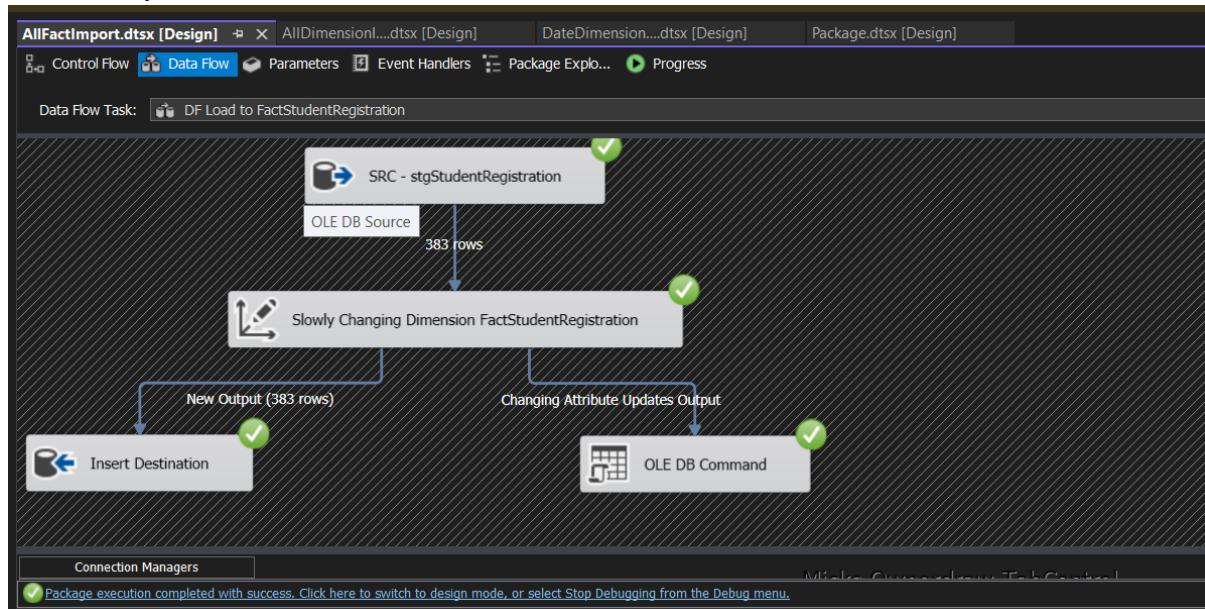








- Khởi chạy



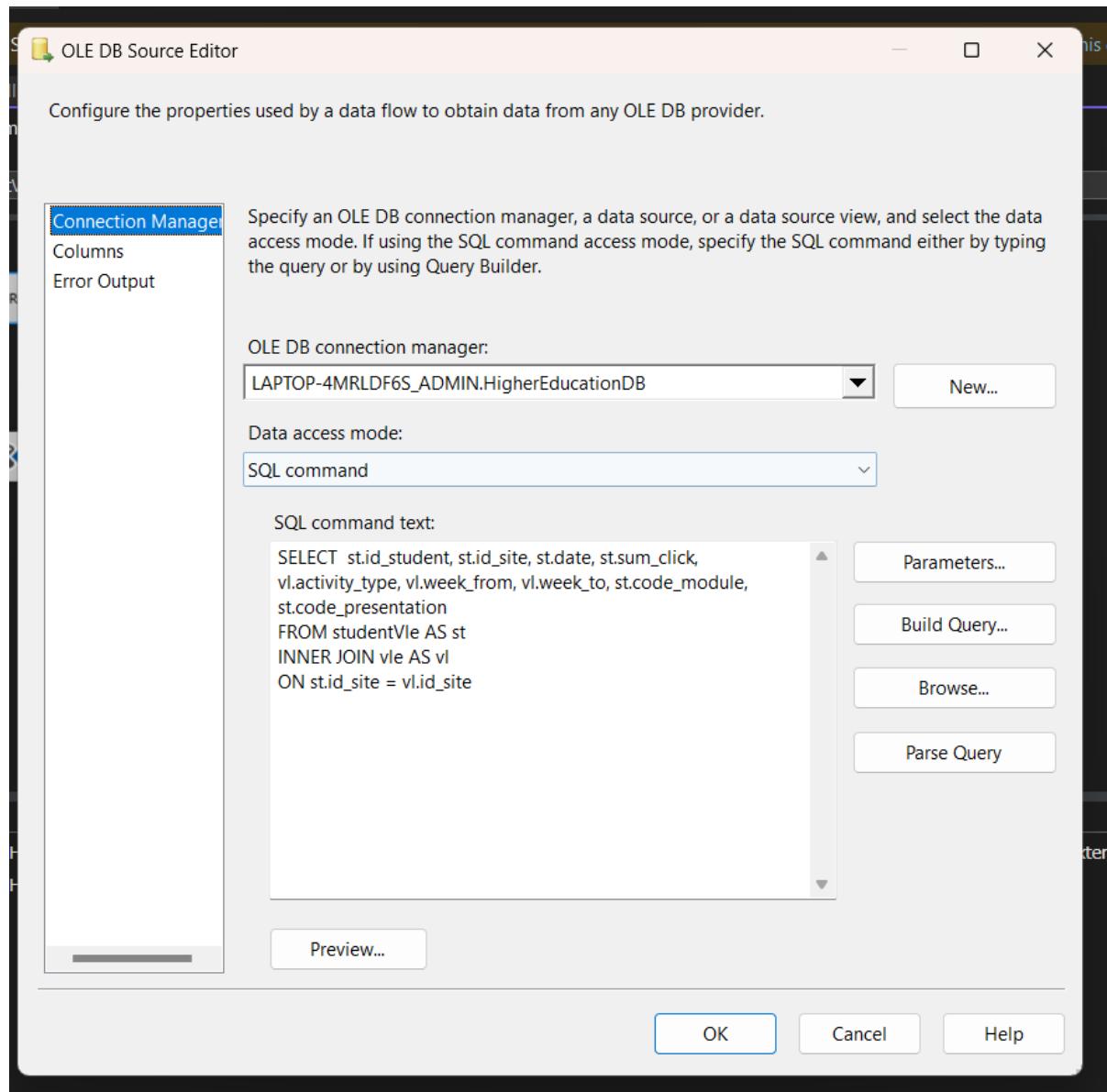
3.4.4. Fact Student Vle

3.4.4.1. Load dữ liệu từ Source vào Stage

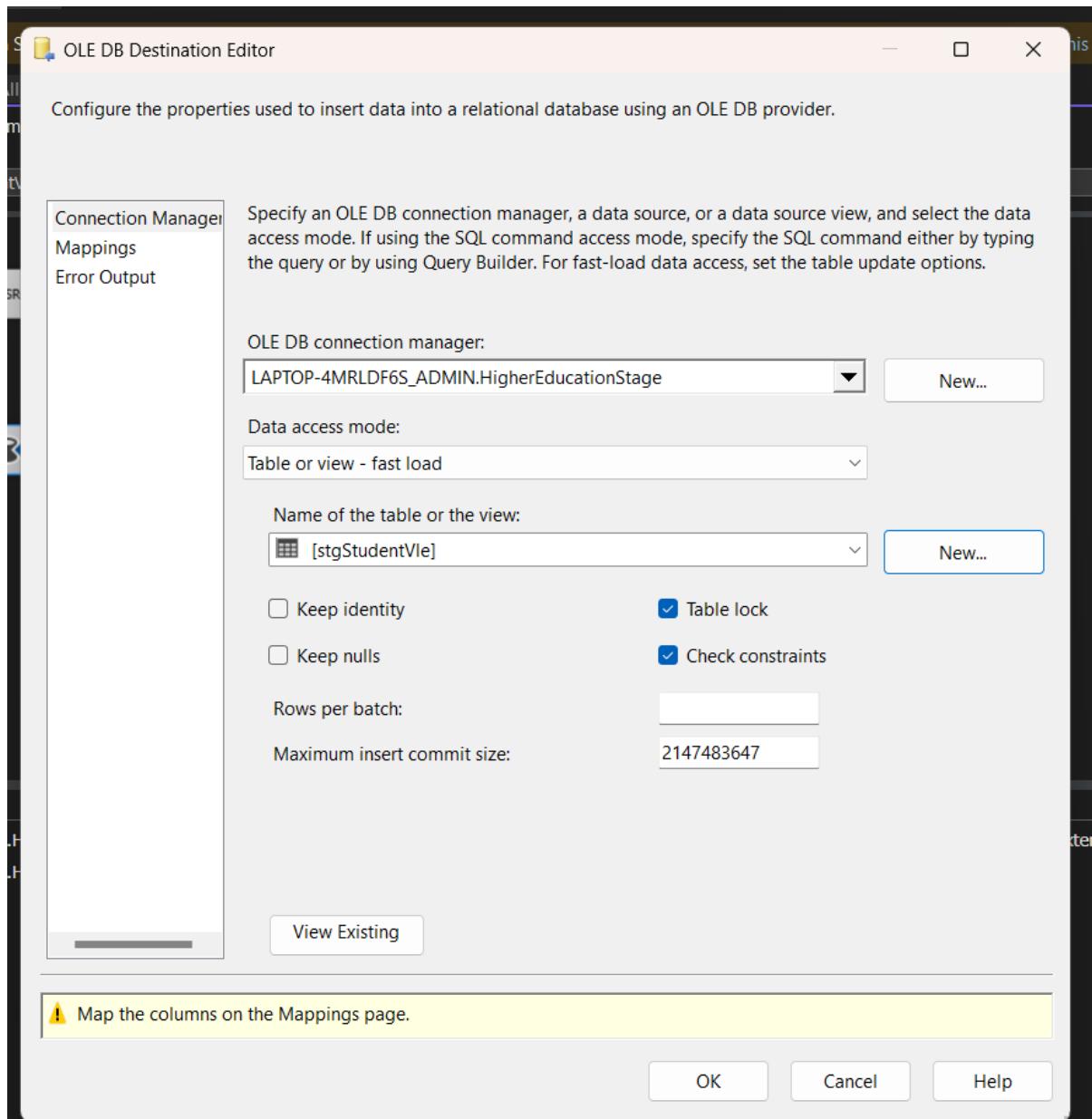
- Double-click DF - Stage Student Vle task để mở nó trong **Data Flow design surface**:

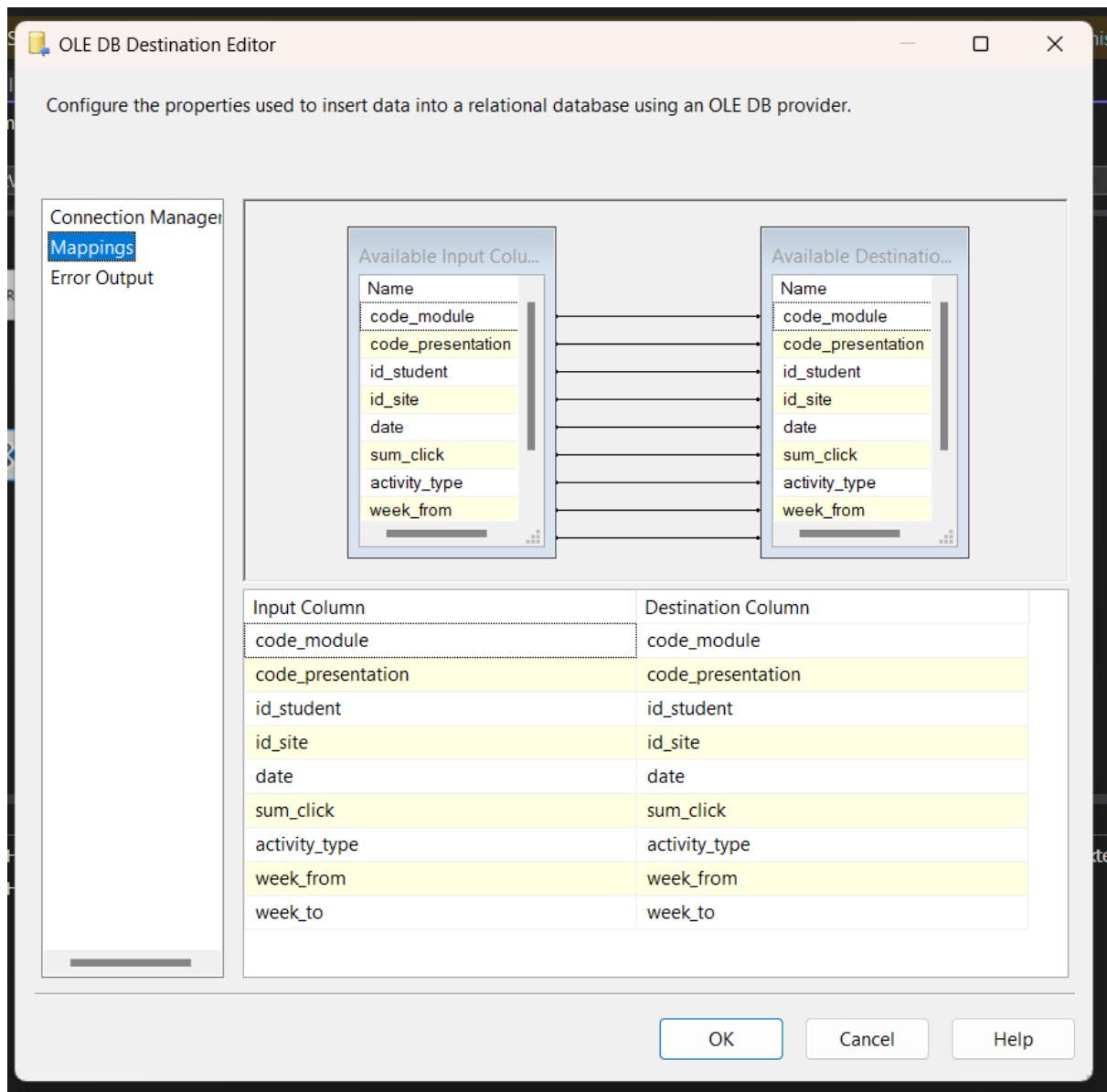
- Tạo **Source Assistant** và **Destination Assistant** và cấu hình chúng như sau:

+ Source Assistant:

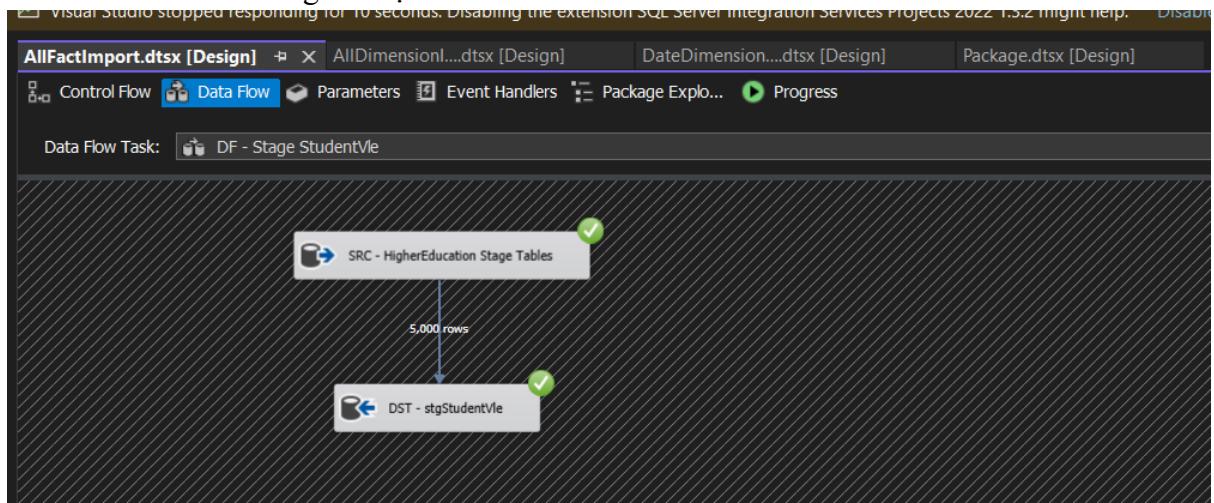


+ Destination Assist



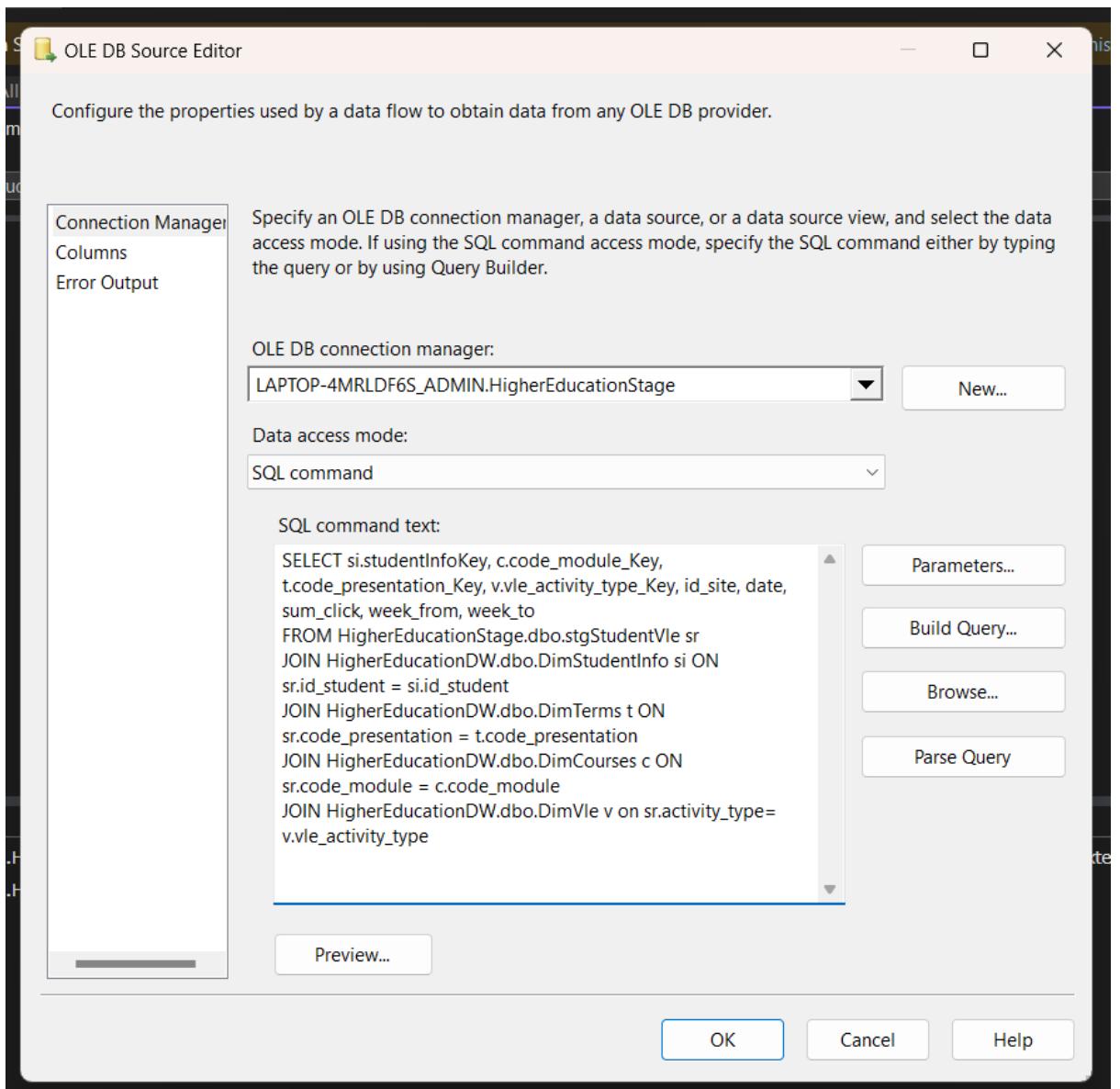


+ Sau khi cấu hình xong ta được:

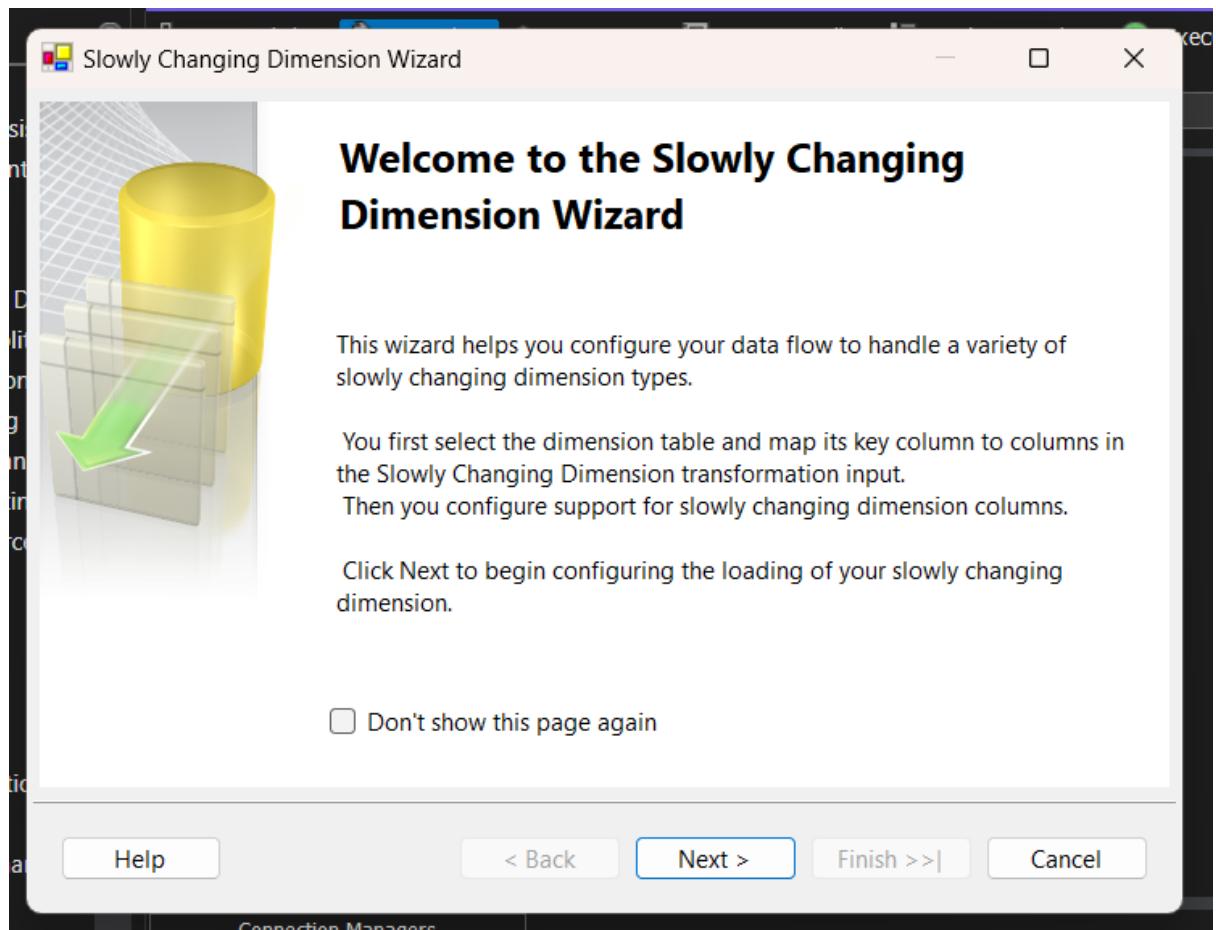


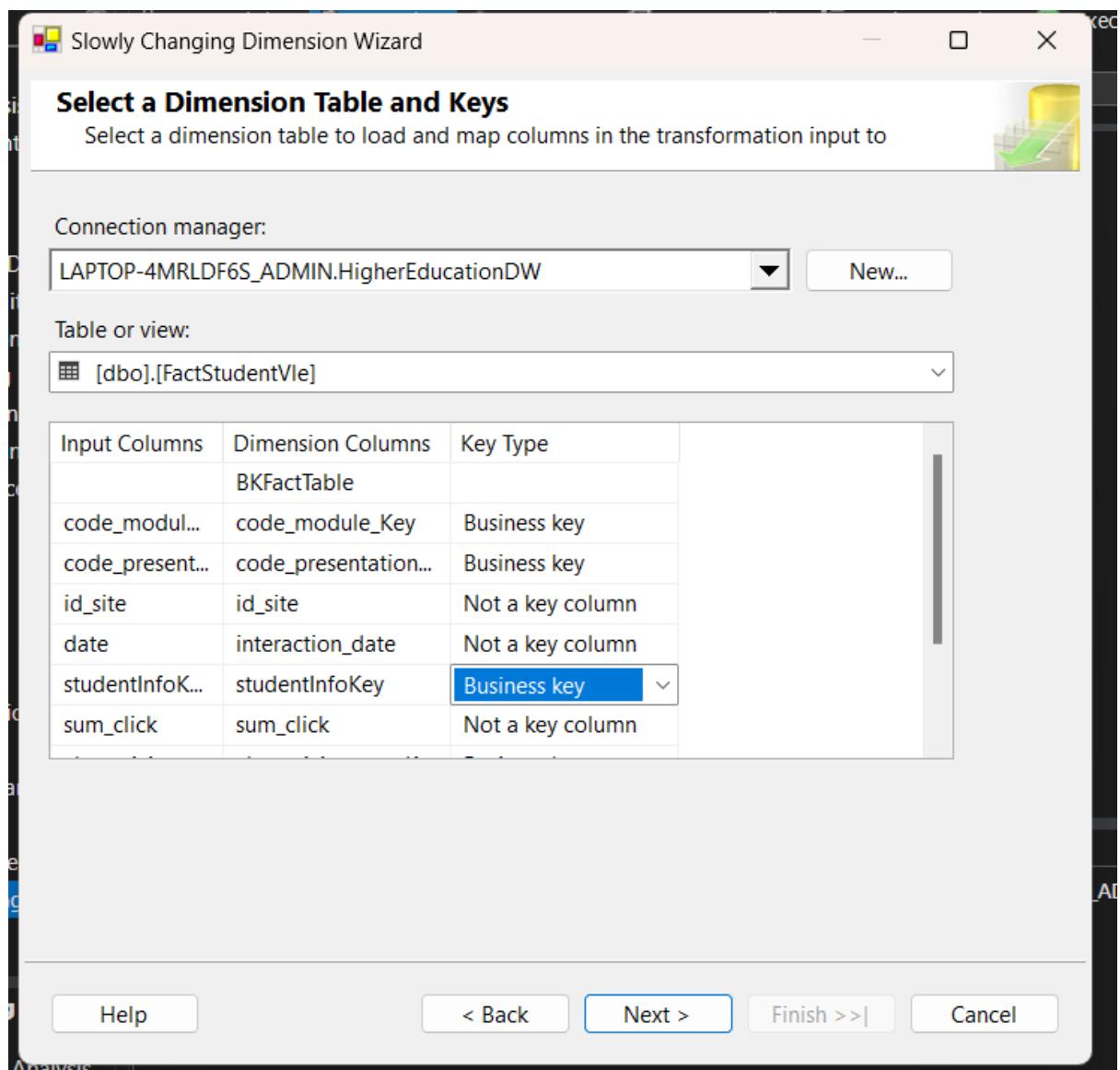
3.4.4.2. Load dữ liệu từ Stage vào Fact

- Double-click vào **DF - Load to FactStudentVle** task để mở data flow design Surface và thiết lập các cấu hình sau
 - + Source Assistant:



- + Slowly Changing Dimension:





Slowly Changing Dimension Wizard

Slowly Changing Dimension Columns

Manage the changes to column data in your slowly changing dimensions by setting the

Fixed Attribute

Select this type when the value in a column should not change. Changes are treated as errors.

Changing Attribute

Select this type when changed values should overwrite existing values. This is a Type 1 change.

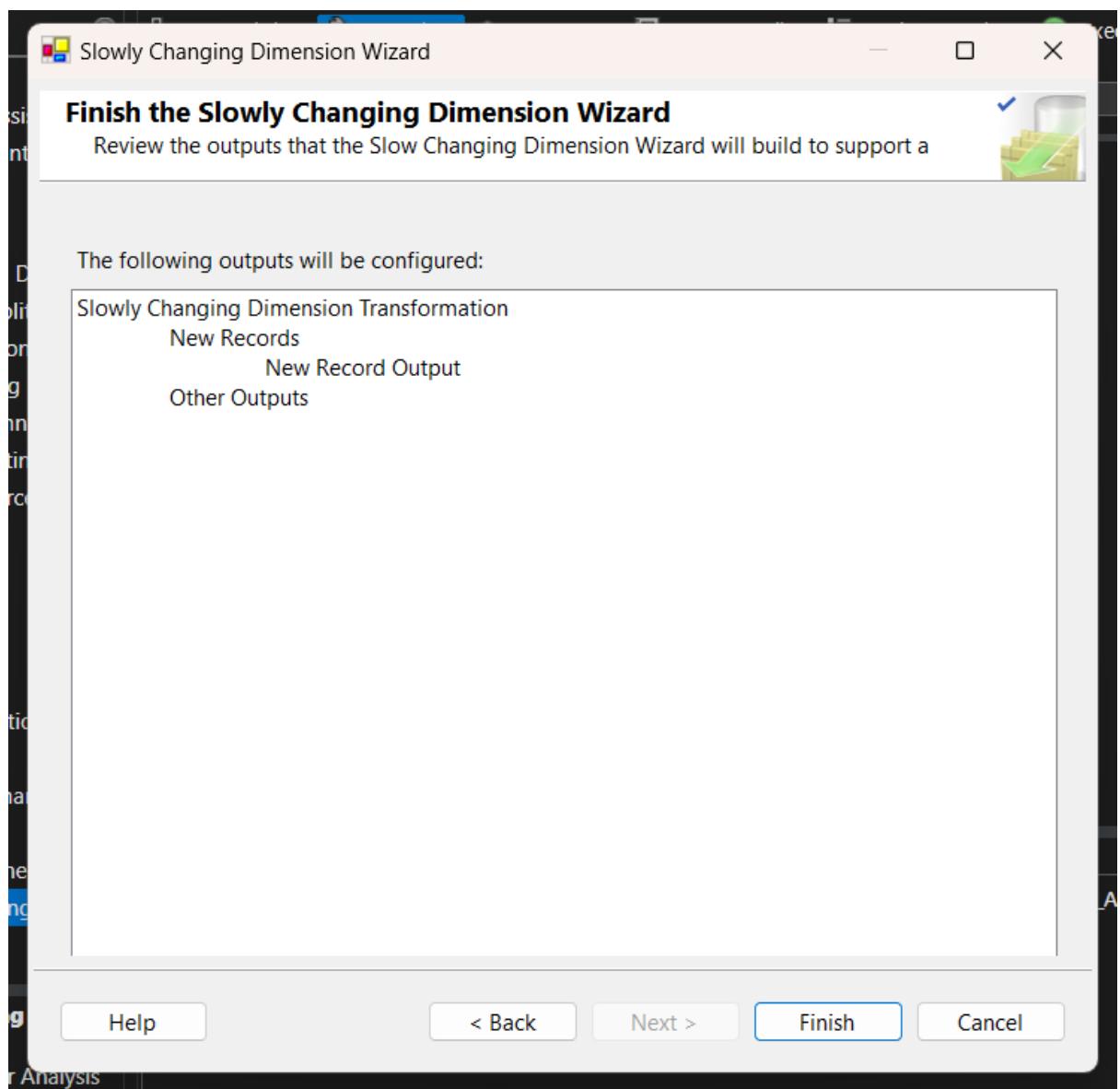
Historical Attribute

Select this type when changes in column values are saved in new records. Previous values are saved in records marked as outdated. This is a Type 2 change.

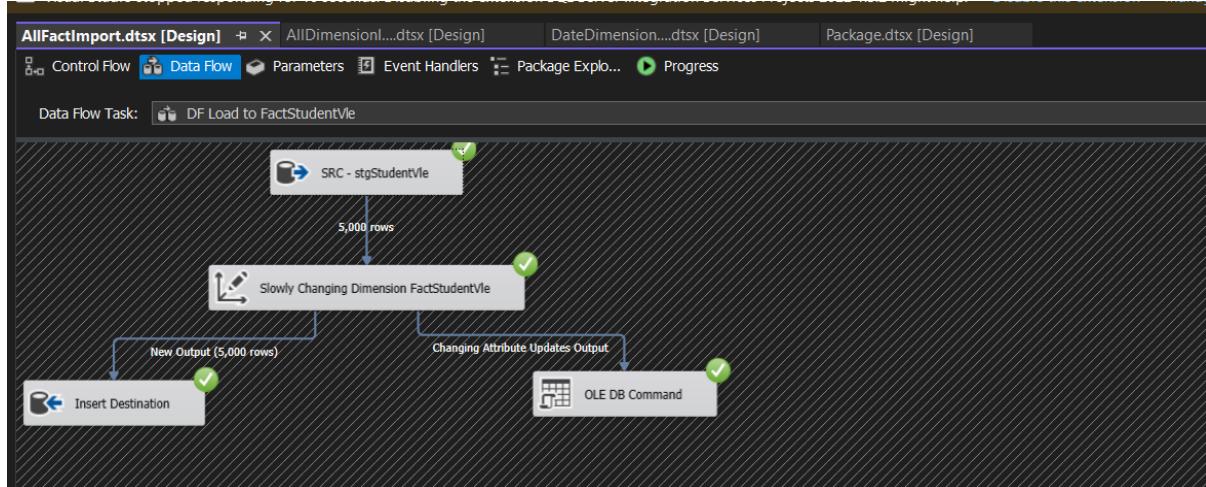
Select a change type for slowly changing dimension columns:

Dimension Columns	Change Type
id_site	Changing a...
interaction_date	Changing a...
sum_click	Changing a...
week_from	Changing a...
week_to	Changing a...

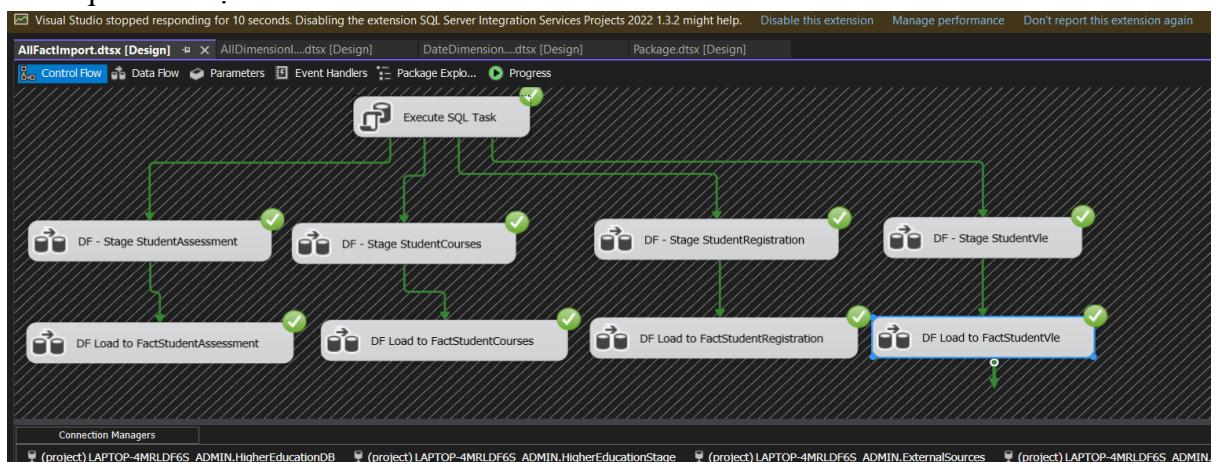
Help < Back Next > Finish >> Cancel



- Khởi chạy



- Kết quả thu được

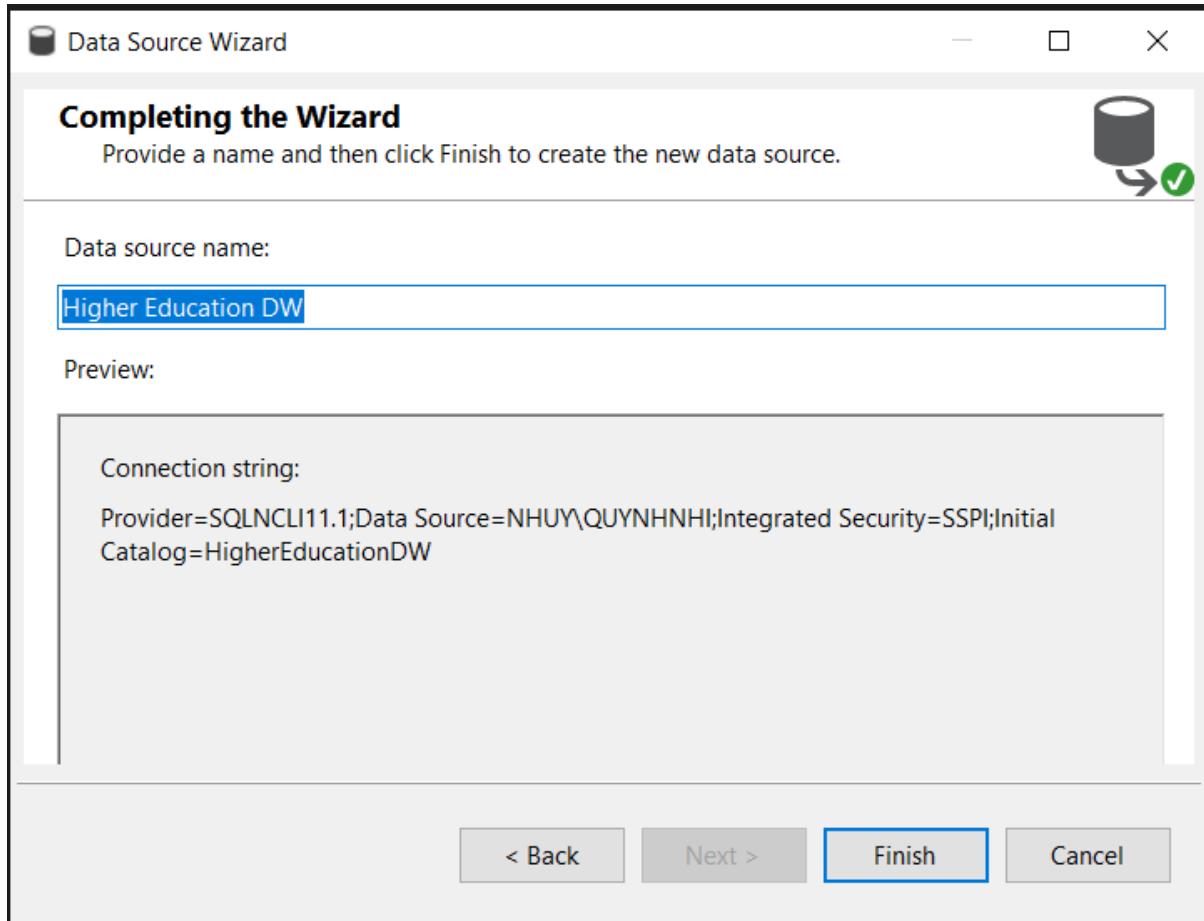


CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS)

4.1. Quá trình xây dựng mô hình

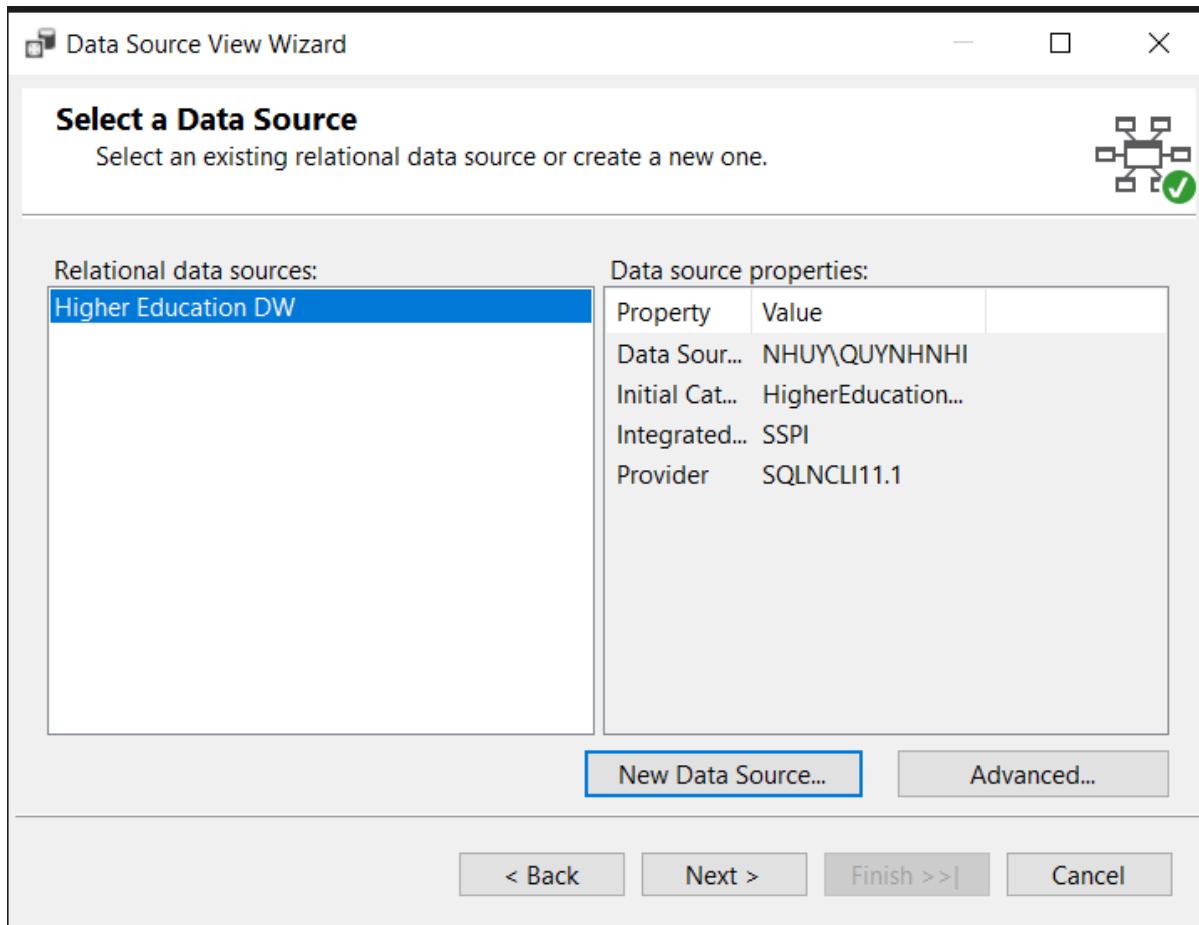
4.1.1. Tạo Data Source

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

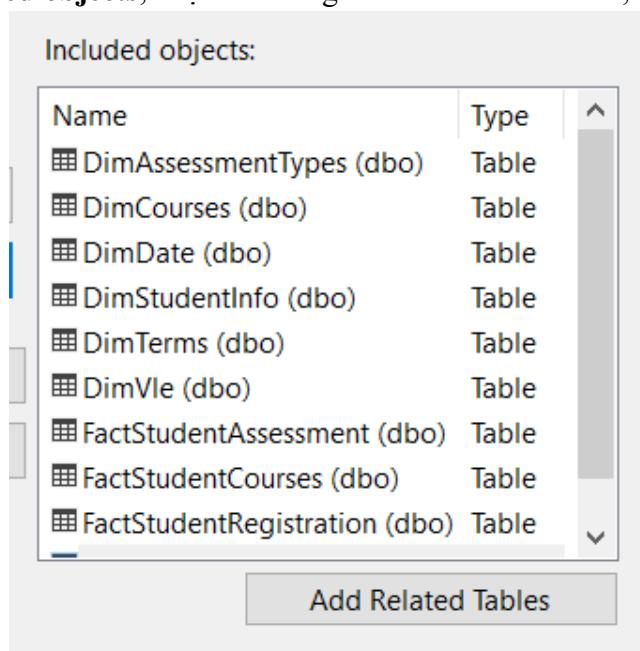


4.1.2. Tạo Data Source View

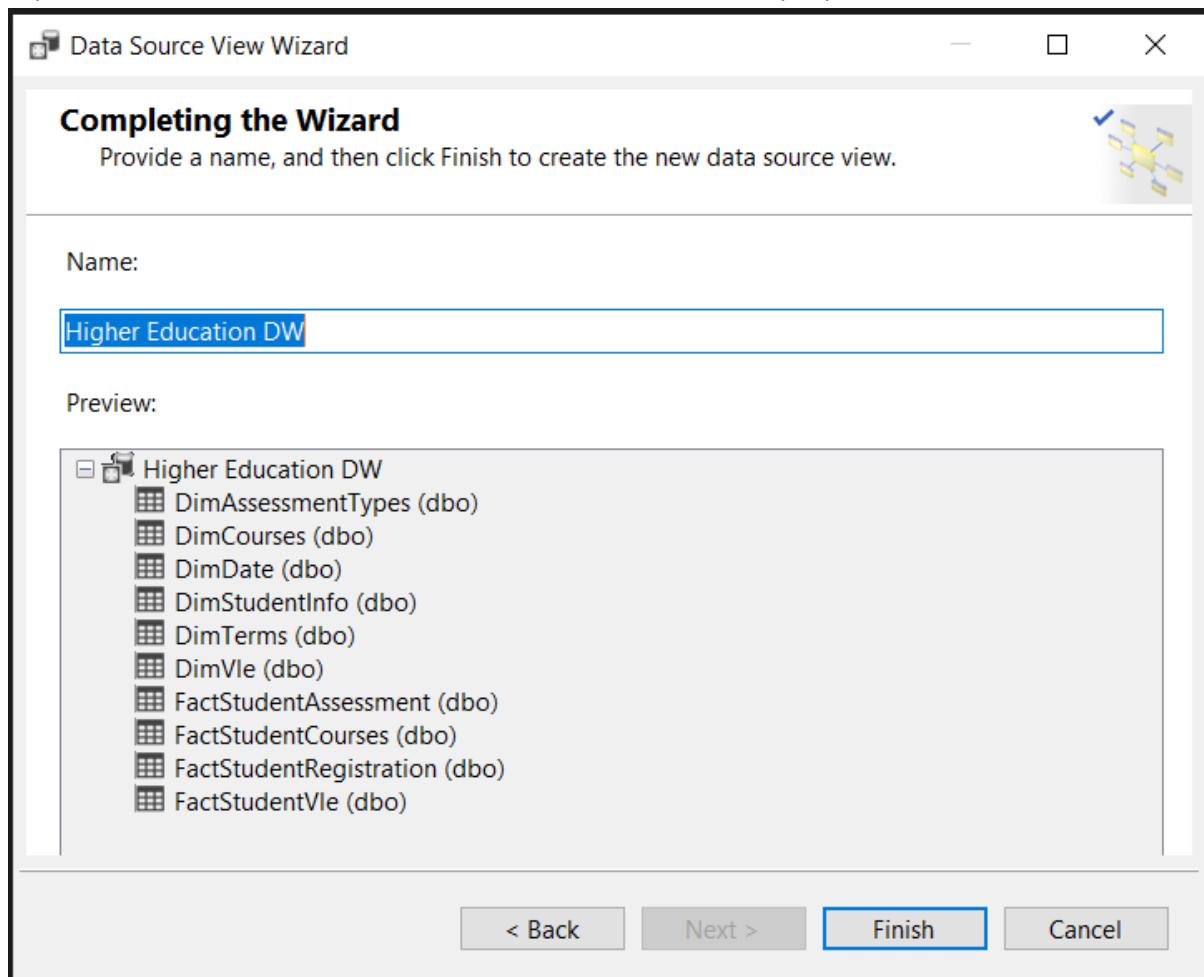
- Right-click vào Data Source Views ở khung Solution Explorer. Chon New Data Source View để tạo Data Source View mới. Chọn Data Soure là **Higher Education DW** và chọn Next.



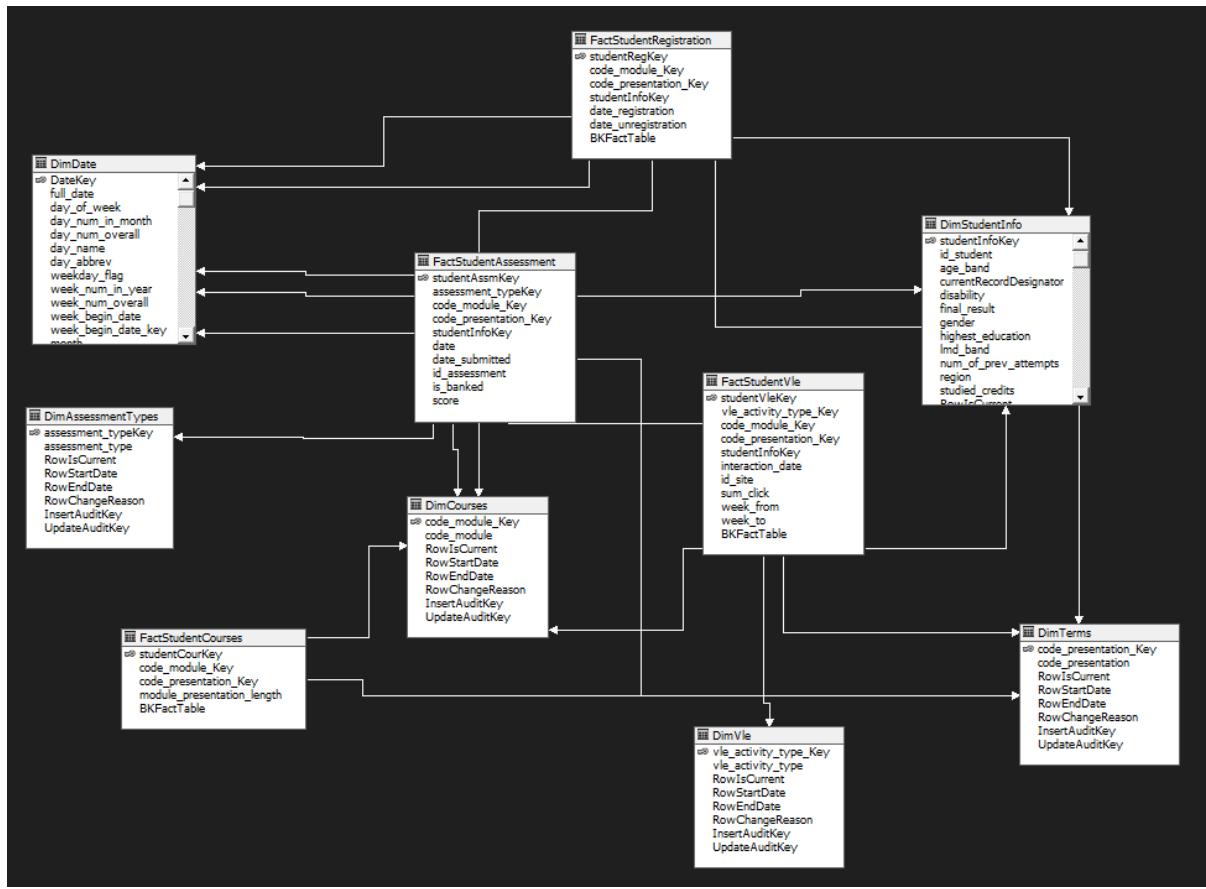
- Trong mục **Included objects**, chọn các bảng Dim và Fact cần thiết, nhấn Next



- Đặt tên cho Data Source View và nhấn Finish để hoàn tất việc tạo Data Source View

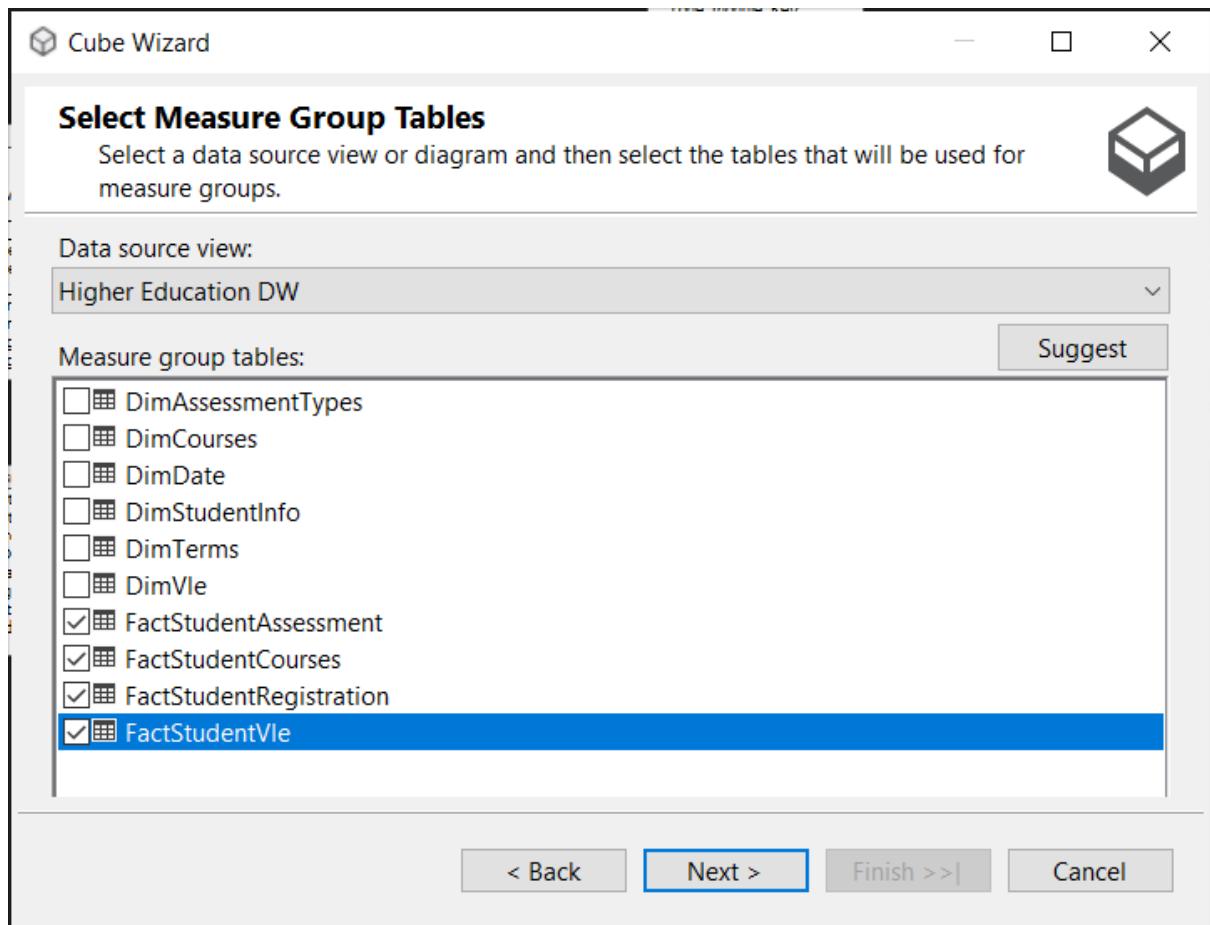


- Double-Click vào Data Source View vừa tạo để hiển thị diagram

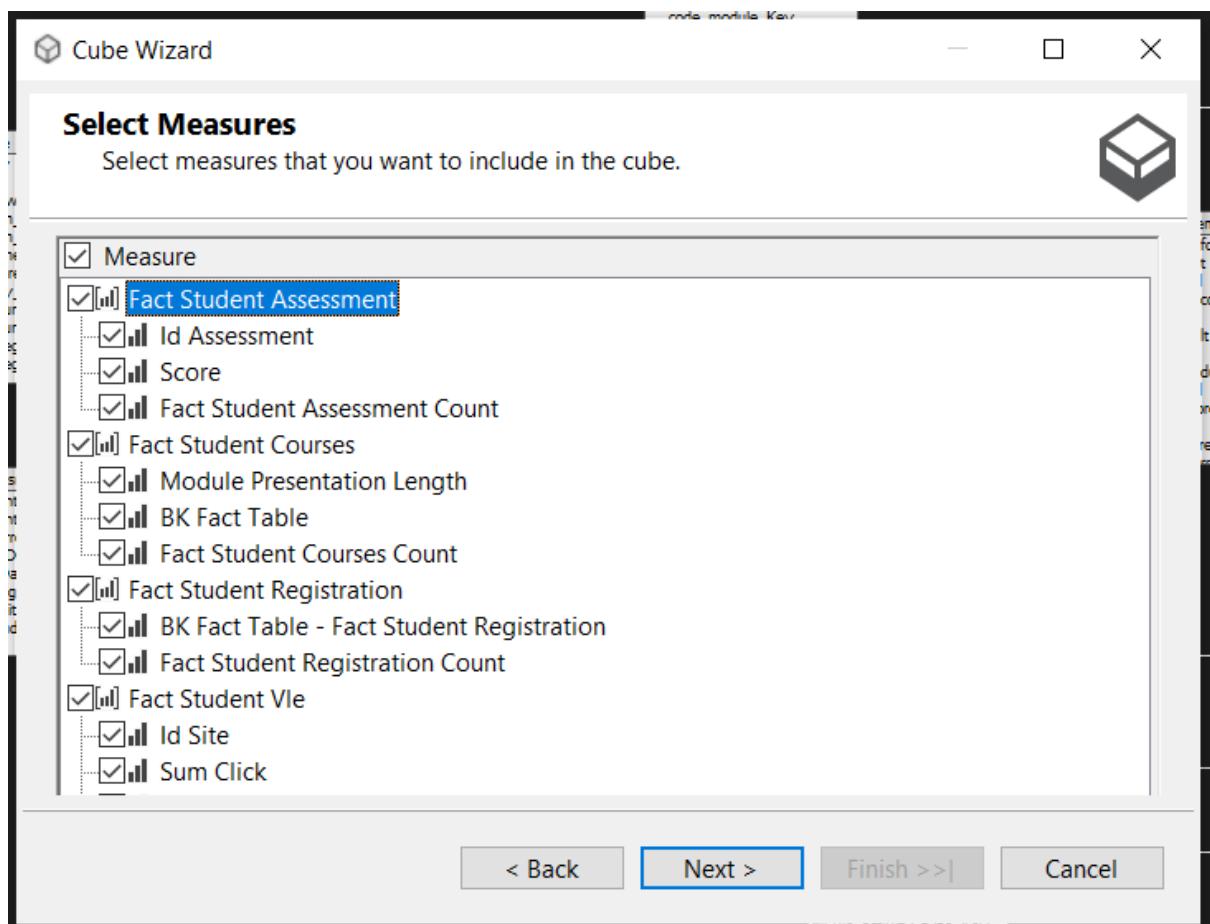


4.2. Quá trình xây dựng khối Cube

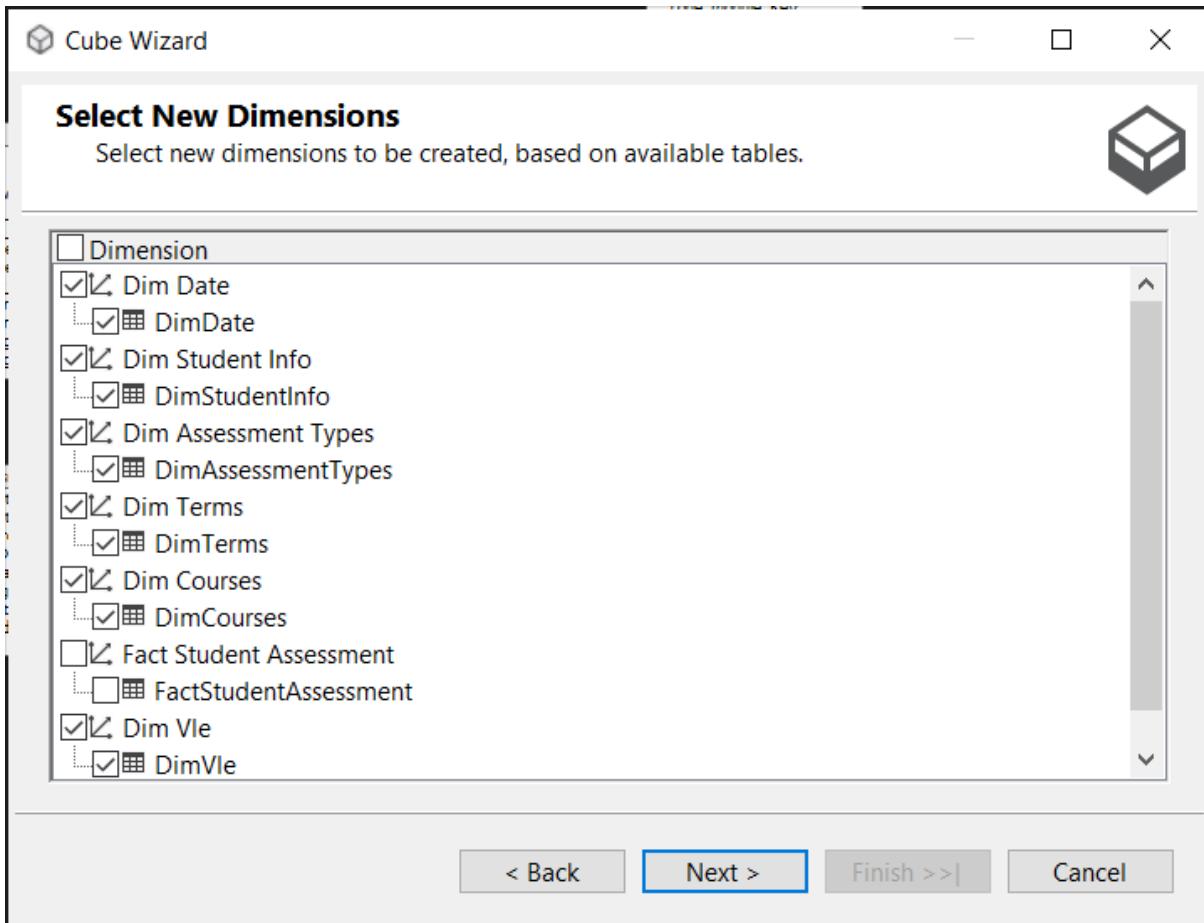
- Right-click vào **Cubes** ở khung Solution Explorer. Chon New Cube để tạo Cube mới.
- Chọn Data Source View là **Higher Education DW**. Tích chọn 4 bảng **FactStudentAssessment**, **FactStudentVle**, **FactStudentCourses** và **FactStudentRegistration** nhán Next



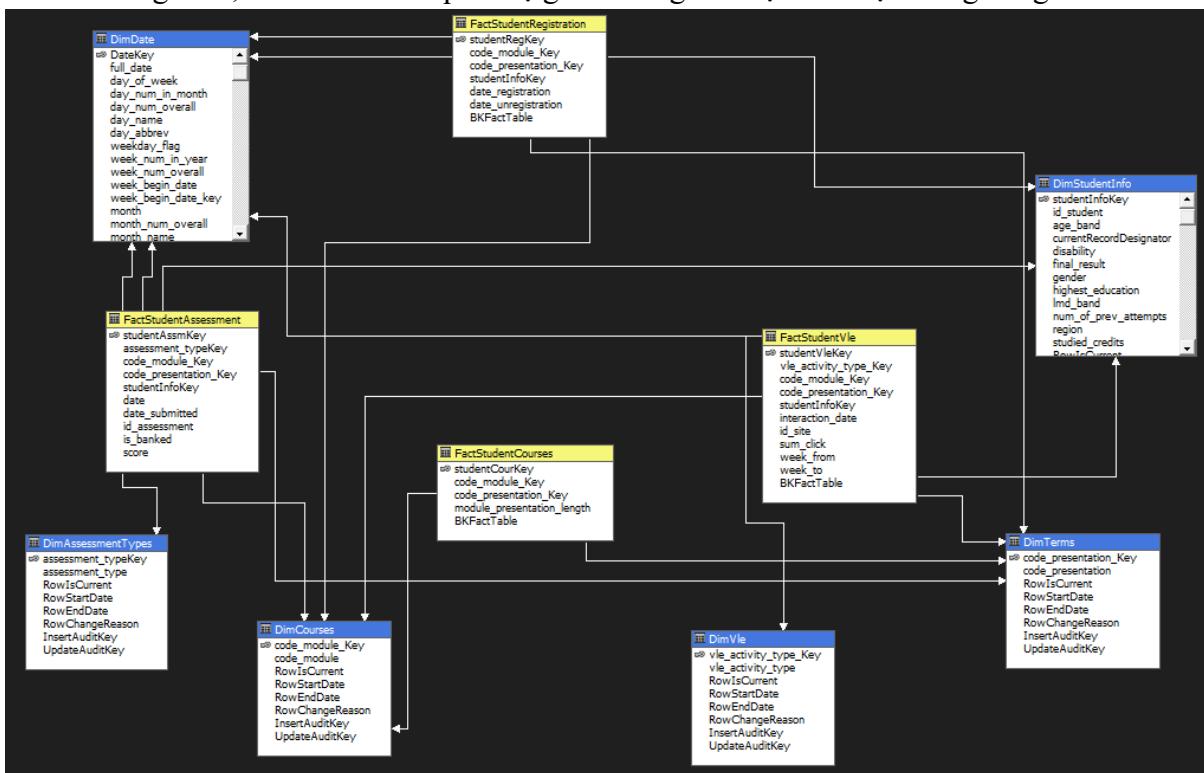
- Ở dialog Select Measures, tích chọn những measure cần thiết các mục sau



- Khởi tạo các bảng Dimension



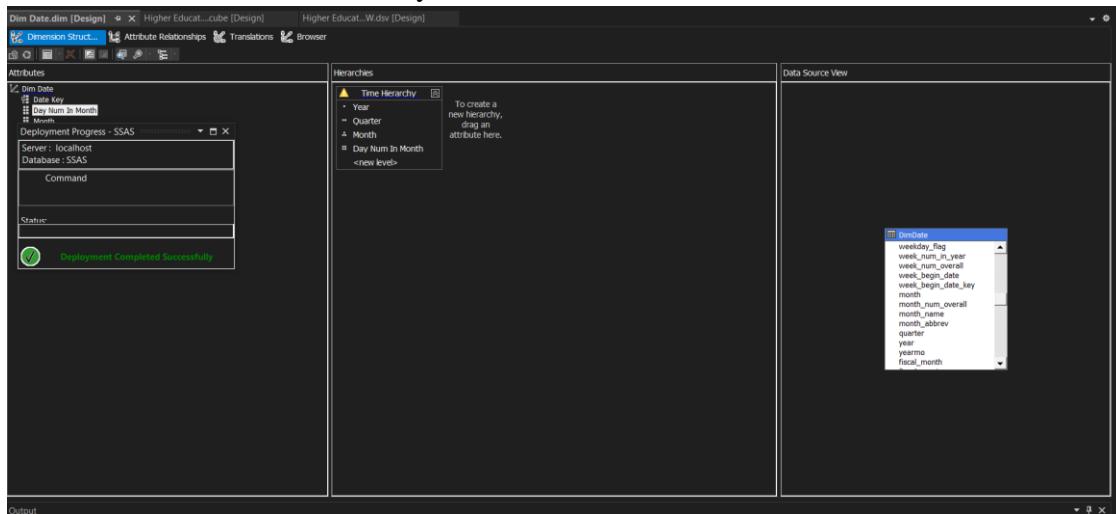
- Đặt tên **Higher Education DW Cube** để hoàn tất khởi tạo.
- Các bảng Fact, Dimension và quan hệ giữa chúng sẽ được thể hiện trong Diagram



4.3. Cấu hình Hierarchy

4.3.1. Tạo Hierarchy cho Dim Date

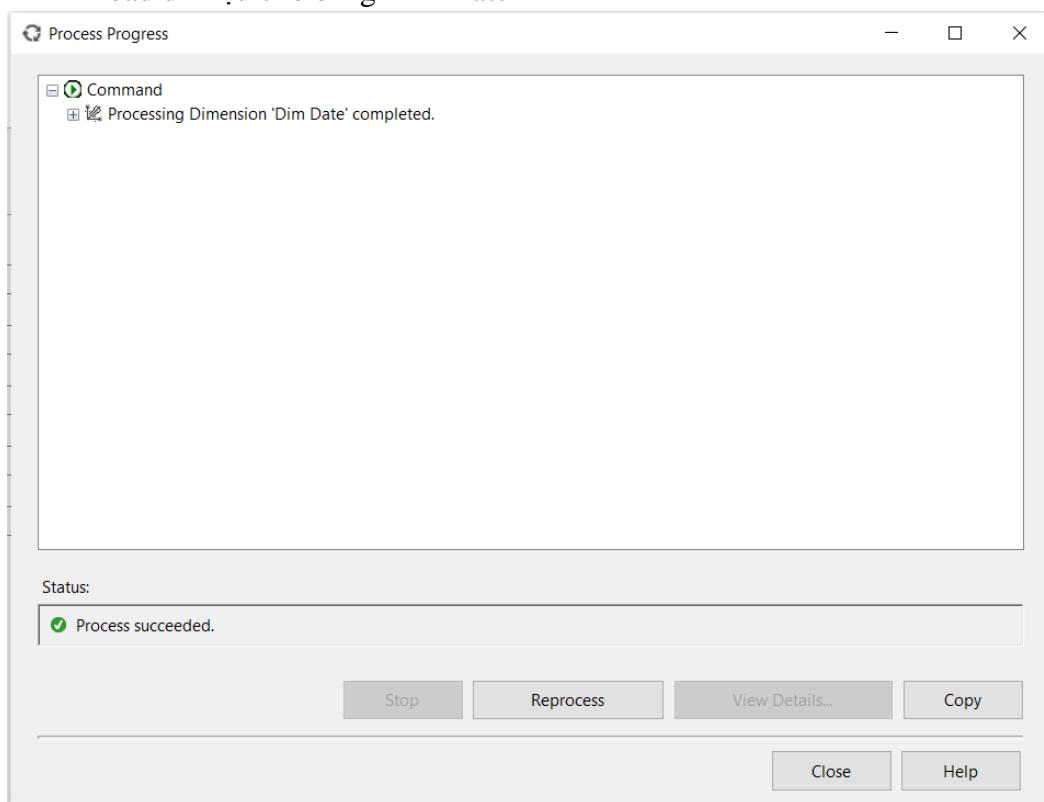
- Thêm thuộc tính vào bảng Dim: Tại table DimDate trong panel Data Source View, kéo thuộc tính cần thiết vào panel Attributes.
- Kéo thả 4 thuộc tính *Year, Quarter, Month, Day Num In Month* từ Attributes vào panel Hierarchies để tạo ra Time Hierarchy



- Tạo Attribute Relationship



- Process để load dữ liệu cho bảng Dim Date



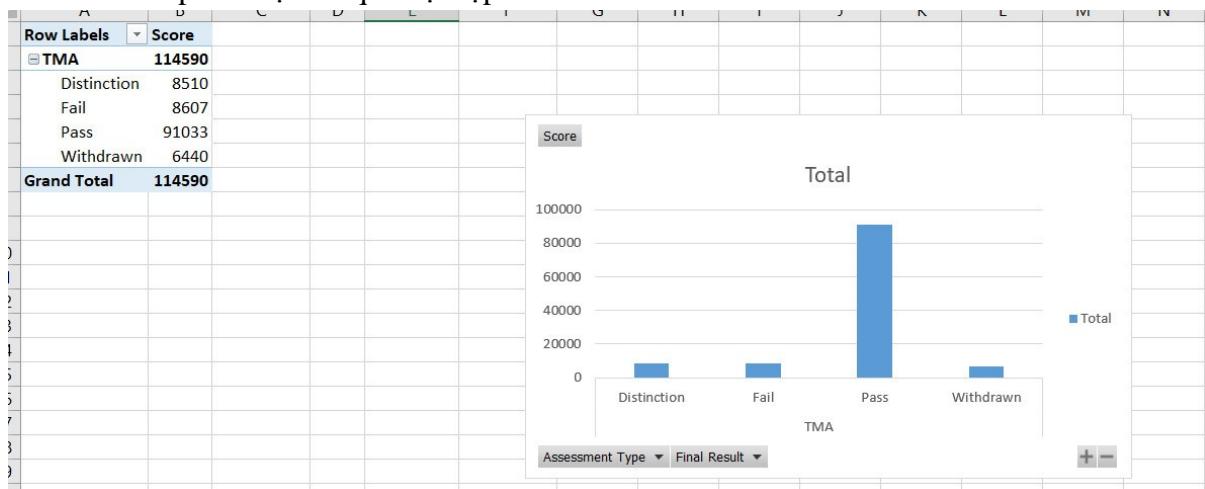
- Chuyển đến tab Browser. Ta thấy thời gian được phân cấp theo năm → quý → tháng.

4.4. Thực hiện phân tích dữ liệu

4.4.1. Câu hỏi: Cho biết kết quả học tập của học sinh theo loại đánh giá

4.4.1.1. Sử dụng Pivot Table trong Excel

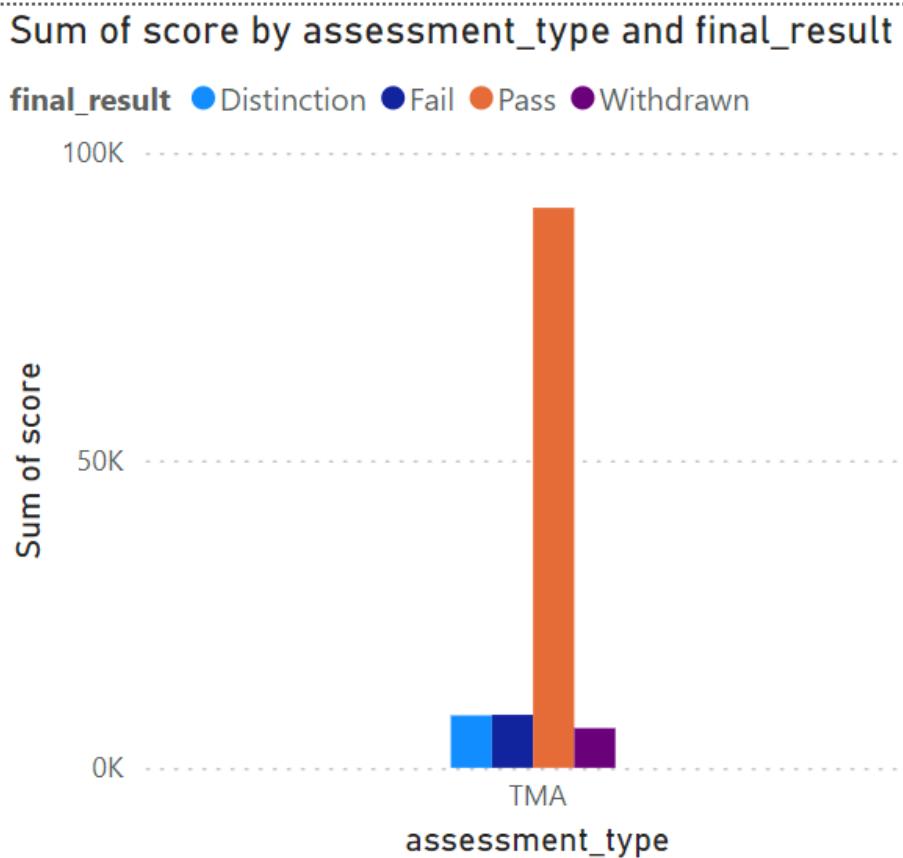
- Hiển thị kết quả học tập của học sinh theo loại đánh giá, ở đây sẽ thống kê số lượng sinh viên theo phân loại kết quả học tập



A	B	C	D	E	F	G
Score	Column Labels					
Row Labels	Distinction	Fail	Pass	Withdrawn	Grand Total	
3 TMA	8510 8607 91033		6440	114590		
4 100893		370		370		
5 101781		330		330		
5 102806		390		390		
7 102952		365		365		
3 1035023		340		340		
9 104476		385		385		
0 106247			198	198		
1 106577		190		190		
2 110175		265		265		
3 1105478		410		410		
4 111717		240		240		
5 113295		375		375		
6 11391		425		425		
7 114017		360		360		
8 114999		400		400		
9 116541		300		300		
0 116692		390		390		
1 118983		315		315		
2 123044		350		350		
3 127582		375		375		
4 129955			246	246		
5 132976		360		360		
6 134143	430			430		
7 1352868			132	132		
8 135400			144	144		
9 137873		390		390		

- Với 2 hình trên thì ta cũng có thể thấy được loại đánh giá và đồng thời hiển thị số lượng học viên theo loại đánh giá.

4.4.1.2. Sử dụng Power BI

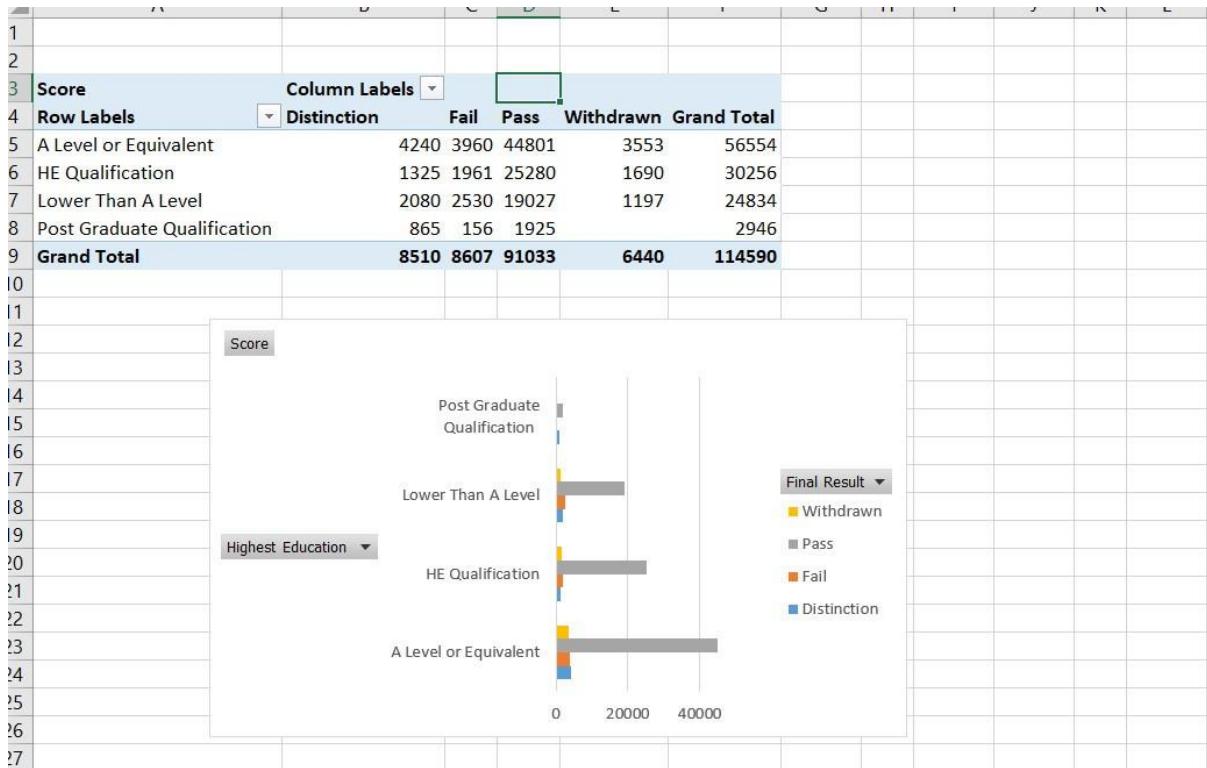


- Nhìn biểu đồ có thể thấy rằng số lượng học viên thông qua 2 loại bài đánh giá đều có số lượng cao hơn hẳn so với rớt. Thể hiện được chất lượng giảng dạy của VLE.

4.4.2. Câu hỏi: Cho biết kết quả học tập của học sinh theo bậc học

4.4.2.1. Sử dụng Pivot Table trong Excel

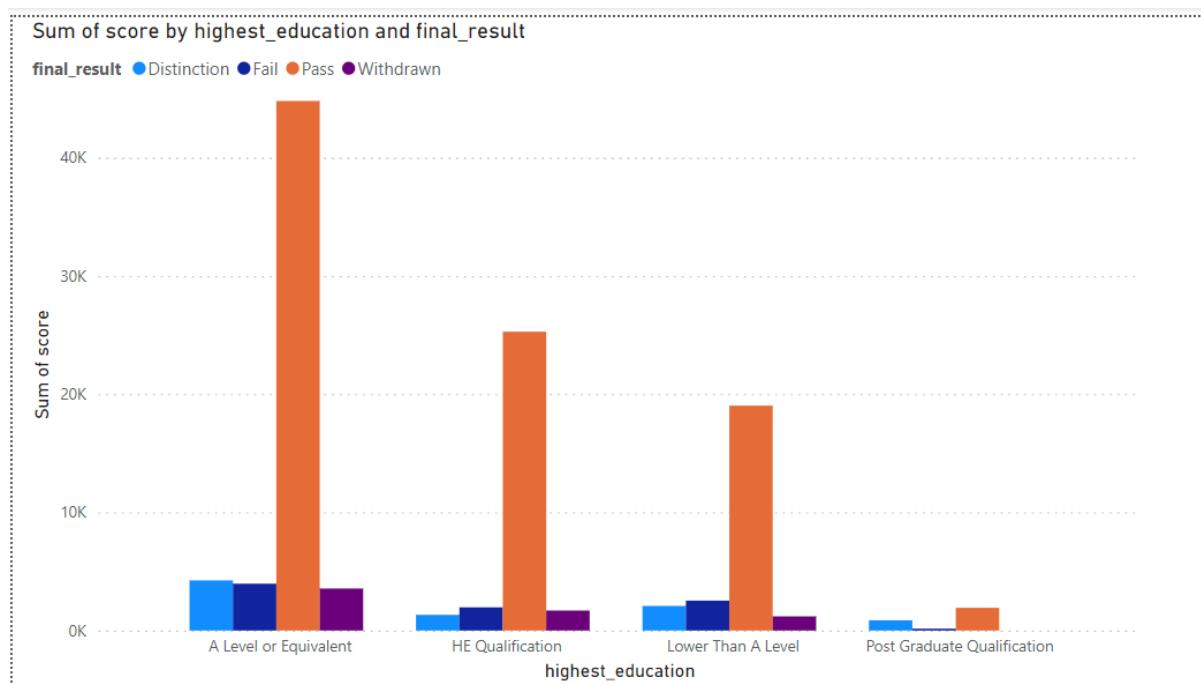
- Hiển thị thống kê kết quả học tập của học viên theo kết quả đánh giá phân loại theo các bậc học



1	Score	Column Labels				
2	Row Labels	Distinction	Fail	Pass	Withdrawn	Grand Total
3	☒ A Level or Equivalent		4240	3960	44801	3553
4	100893		370			370
5	102806		390			390
6	116692		390			390
7	123044		350			350
8	127582		375			375
9	129955			246		246
10	134143	430				430
11	1352868			132		132
12	137873		390			390
13	1401935	90				90
14	141355		355			355
15	141377			45		45
16	1414443	116				116
17	142326		355			355
18	146188	215				215
19	1488959		320			320
20	148993			210		210
21	151358		265			265
22	155550		305			305
23	155854		405			405
24	157568		300			300
25	159954			130		130
26	1626710	425				425
27	1640609		425			425
28	164259		365			365
29	167906		360			360

- Hình trên thể hiện số lượng các học viên có cùng đánh giá theo phân loại kết quả học tập và theo cấp bậc học.

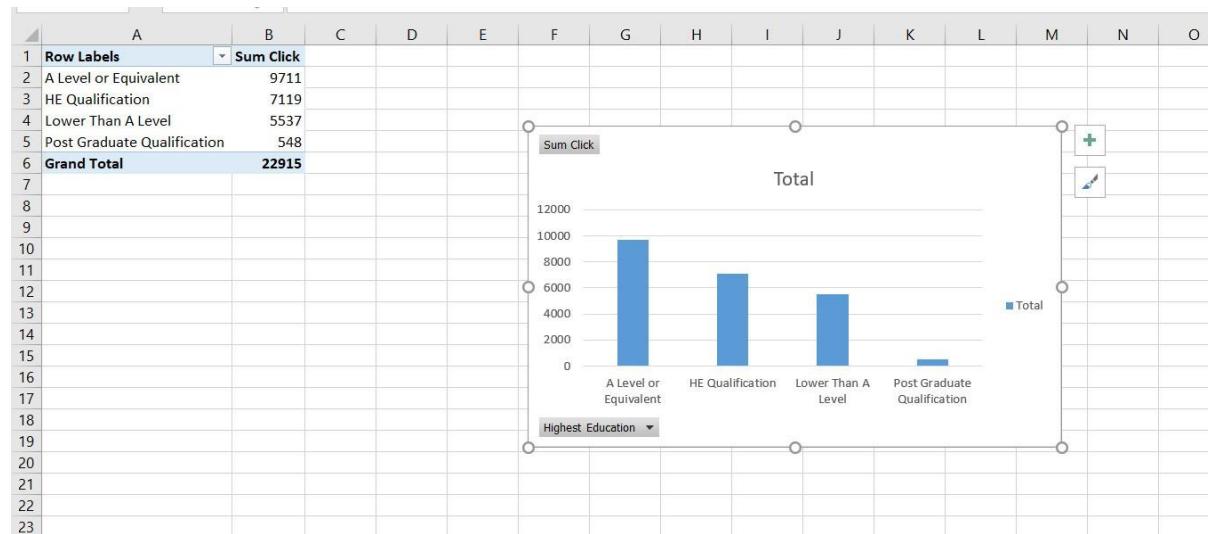
4.4.2.2. Sử dụng Power BI



- Dựa vào biểu đồ trên có thể so sánh sự chênh lệch giữa các loại đánh giá kết quả của các học viên theo cấp bậc học, từ đó có thể đưa ra được các đánh giá và biện pháp cải thiện kết quả học tập cho học viên một cách hiệu quả nhất.

4.4.3. Câu hỏi: Cho biết tổng số lượng thao tác trên ứng dụng của học viên trong cùng một mức học vấn

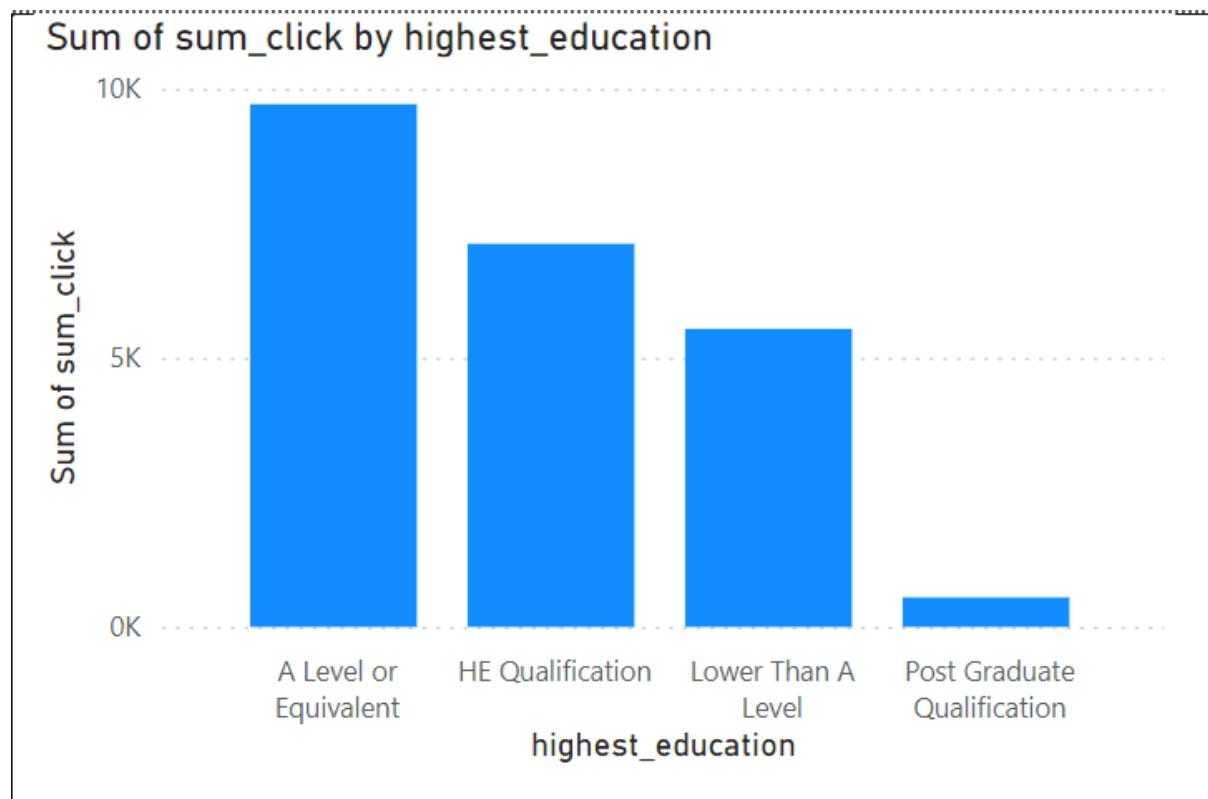
4.4.3.1. Sử dụng Pivot Table trong Excel



	A	B	C
1	Row Labels	Sum Click	
2	▪ A Level or Equivalent	9711	
3	100893	14	
4	102806	35	
5	116692	135	
6	123044	3	
7	127582	23	
8	129955	18	
9	134143	7	
10	1352868	35	
11	137873	48	
12	1401935	20	
13	141355	19	
14	1414443	25	
15	1456619	75	
16	146188	28	
17	1488959	112	
18	148993	136	
19	155550	30	
20	155854	162	
21	157568	28	
22	159954	25	
23	1626710	468	
24	1640609	130	
25	164259	6	
26	167906	136	
27	172797	136	
28	1746134	55	
29	1801641	17	

- Hình trên biểu diễn tổng số lần thao tác lên hệ thống theo học viên và tính tổng các lần thao tác của tất cả học viên trong cùng một cấp bậc học để thống kê.

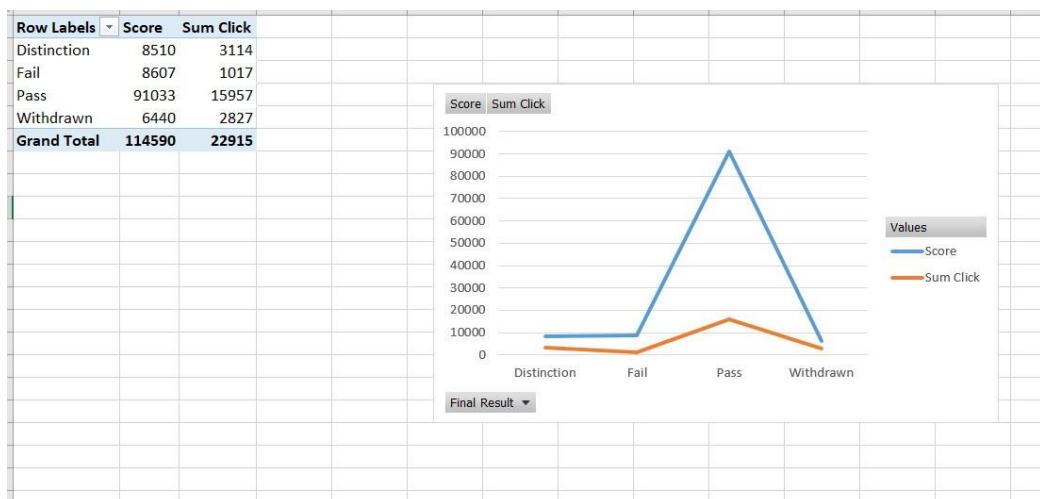
4.4.3.2. Sử dụng Power BI



- Dựa vào số lượng thao tác có thể thống kê được mức độ thao tác trên hệ thống theo trình độ học vấn để xác định số lượng học viên ở trình độ nào nhiều hơn.
- Dựa vào biểu đồ trên có thể thấy được trình độ học vấn thuộc mức độ A có tổng số lần click/thao tác trên hệ thống cao nhất.

4.4.4. Câu hỏi: Cho biết tổng số lượng thao tác và điểm số của học viên trên ứng dụng theo kết quả học tập

4.4.4.1. Sử dụng Pivot Table trong Excel

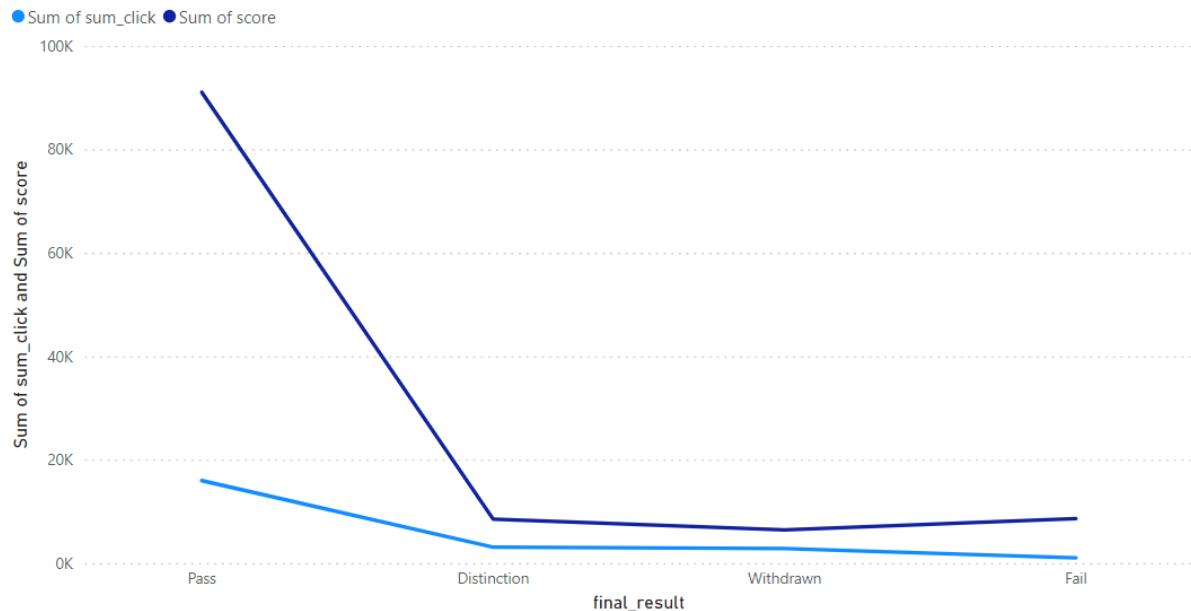


1	Row Labels	B	C	D
2	☒ Distinction	8510	3114	
3	134143	430	7	
4	1626710	425	468	
5	187100	450	353	
6	2122180	395	166	
7	227499	420	213	
8	2339528	425		
9	2412002	430	177	
10	2536991	410	30	
11	2649826	435	309	
12	2689958	410	95	
13	2691206	450	193	
14	279572	395	114	
15	291334	430	157	
16	296332	440	21	
17	312537	425	204	
18	345357	425		
19	376439	470	205	
20	444677	445	6	
21	705379	390	158	
22	760729	410	238	
23	☒ Fail	8607	1017	
24	1401935	90	20	
25	1414443	116	25	
26	146188	215	28	
27	1472925	310		
28	147756	300	58	
29	1620151	170	101	

- Hình trên biểu diễn tổng số lần thao tác lên hệ thống và điểm số theo học viên, tính tổng các lần thao tác và điểm số của tất cả học viên theo phân loại kết quả đánh giá

4.4.4.2. Sử dụng Power BI

Sum of sum_click and Sum of score by final_result



- Dựa vào biểu đồ trên có thể thấy được số lượng học viên thông qua bài kiểm tra có số lượng cao nhất kể cả theo tổng số lượng click hay điểm số.

CHƯƠNG 5: KẾT LUẬN

5.1. Kết quả đạt được

Thực hiện dự án 'KHO DỮ LIỆU GIÁO DỤC CHO TRƯỜNG ĐẠI HỌC' giúp chúng em nắm vững hơn kiến thức về môn học Kho Dữ liệu, đồng thời thành thạo trong việc sử dụng các công cụ hỗ trợ quá trình ETL như SSIS và SSAS. Chúng em cũng học được cách sử dụng các công cụ trực quan hóa dữ liệu như Excel Pivot và Power BI Desktop. Qua dự án này, kỹ năng phân tích, tiếp nhận và xử lý thông tin đầu vào của chúng em được cải thiện, cùng với khả năng đặt câu hỏi nền tảng và thảo luận để giải quyết các vấn đề và trả lời câu hỏi một cách hiệu quả.

5.2. Những hạn chế

Trong quá trình xây dựng kho dữ liệu, nhóm chúng em gặp nhiều khó khăn do hạn chế về kiến thức. Tập dữ liệu vẫn còn nhiều thông tin chưa được khai thác đầy đủ. Những nỗ lực của chúng em chưa đạt đến mức nâng cao trong toàn bộ quá trình xây dựng một kho dữ liệu hoàn chỉnh.

5.3. 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>. [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].