

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO ĐỒ ÁN THỰC HÀNH

Bộ môn: Hệ thống thông tin phục vụ trí tuệ kinh doanh

Giảng viên hướng dẫn:

Hồ Thị Hoàng Vy

Tiết Gia Hồng

TP. HỒ CHÍ MINH, 5 tháng 12 năm 2023

MỤC LỤC

I. Thông tin nhóm 6	3
II. Bảng phân công	3
III. Supermarket sales	3\
A) Quá trình ETL	3
1. Main flow	3
2. Từ Source vào stage	6
3. Từ Stage vào NDS	14
4. Từ NDS vào DDS	32
B) Report	43
C) Olap, mdx	49
D) Dashboard	56
E) Mining	57
IV. Plato's Pizza	59
A) Quá trình ETL	59
1. Main flow	59
2. Từ Source vào Stage	61
3. Stage vào NDS	67
4. NDS vào DDS	84
B) Report	93
C) Olap, mdx	96
D) Dashboard	103
E) Mining	104

I. Thông tin nhóm 6

Họ tên	MSSV
Phạm Viết Hoàng	20127170
Phạm Cao Minh Tâm	20127320
Nguyễn Phú Trí Nhân	20127580
Phan Gia Huy	20127400

II. Bảng phân công

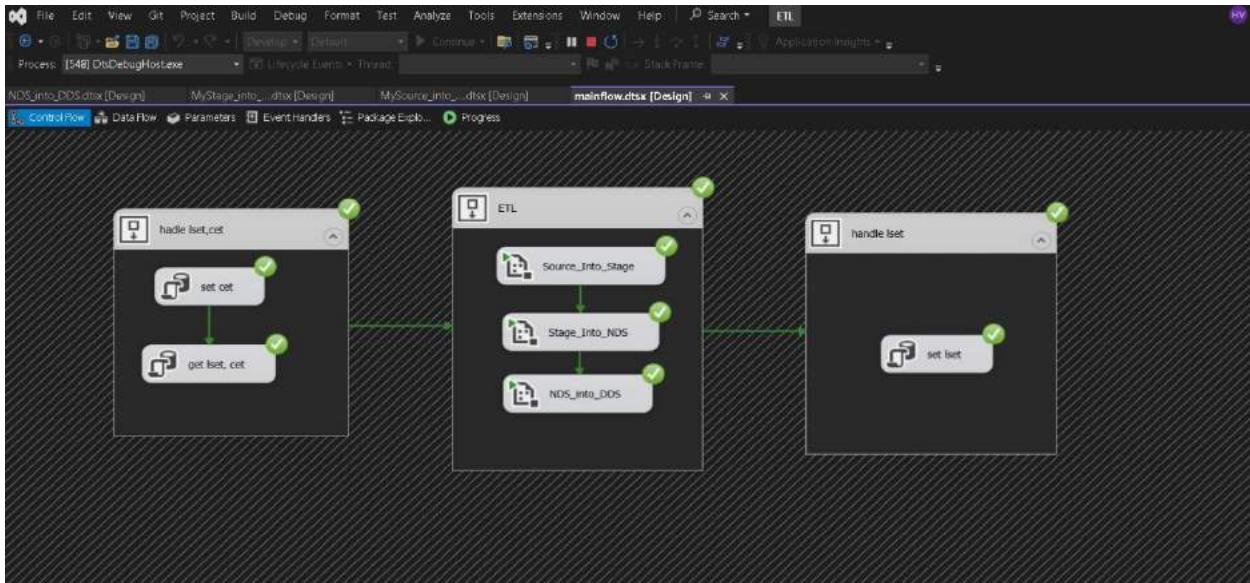
Họ tên	Phân công	Hoàn thành
Phạm Viết Hoàng	Thiết kế NDS, DDS Qui trình ETL Quay video	100%
Phạm Cao Minh Tâm	Thiết kế NDS, DDS Report Viết báo cáo	100%
Nguyễn Phú Trí Nhân	Thiết kế NDS, DDS Mining Viết báo cáo	100%
Phan Gia Huy	Thiết kế NDS, DDS Olap. Mdx Viết báo cáo	100%

III. Supermarket sales

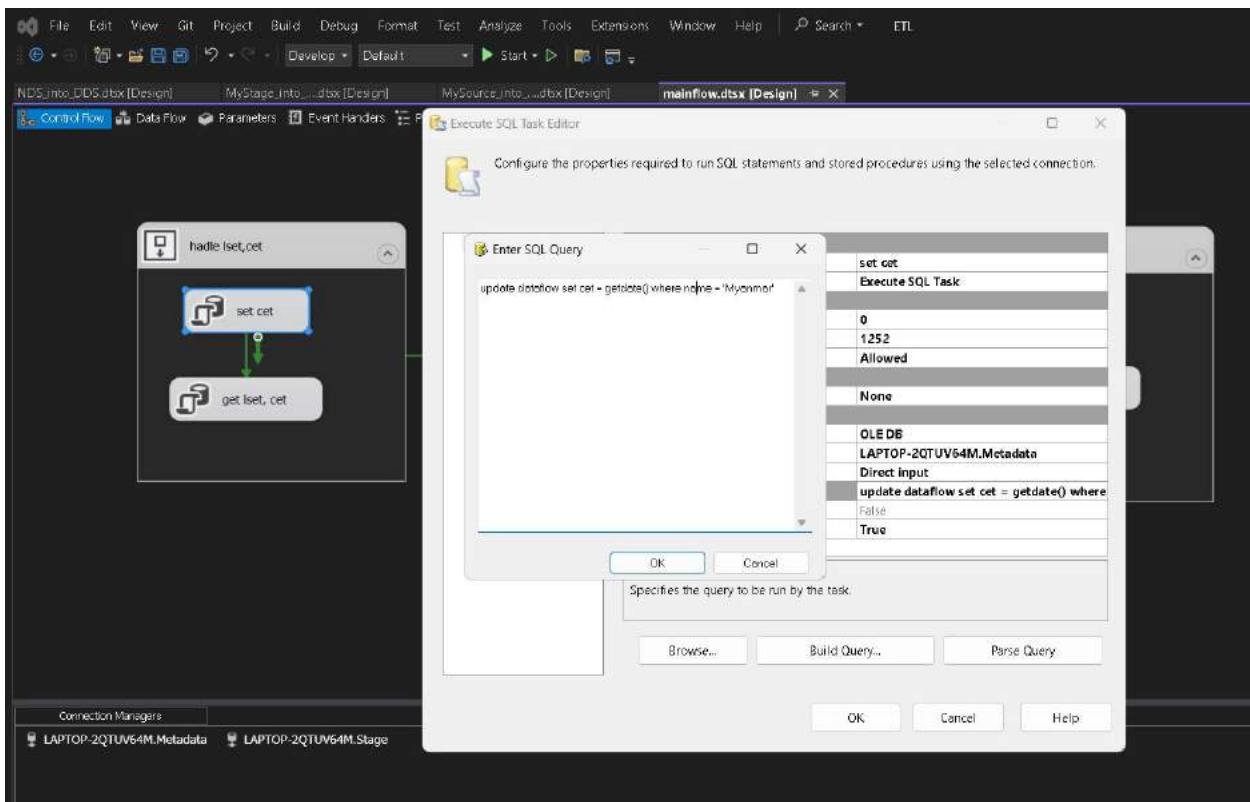
A) Quá trình ETL

1. Main flow

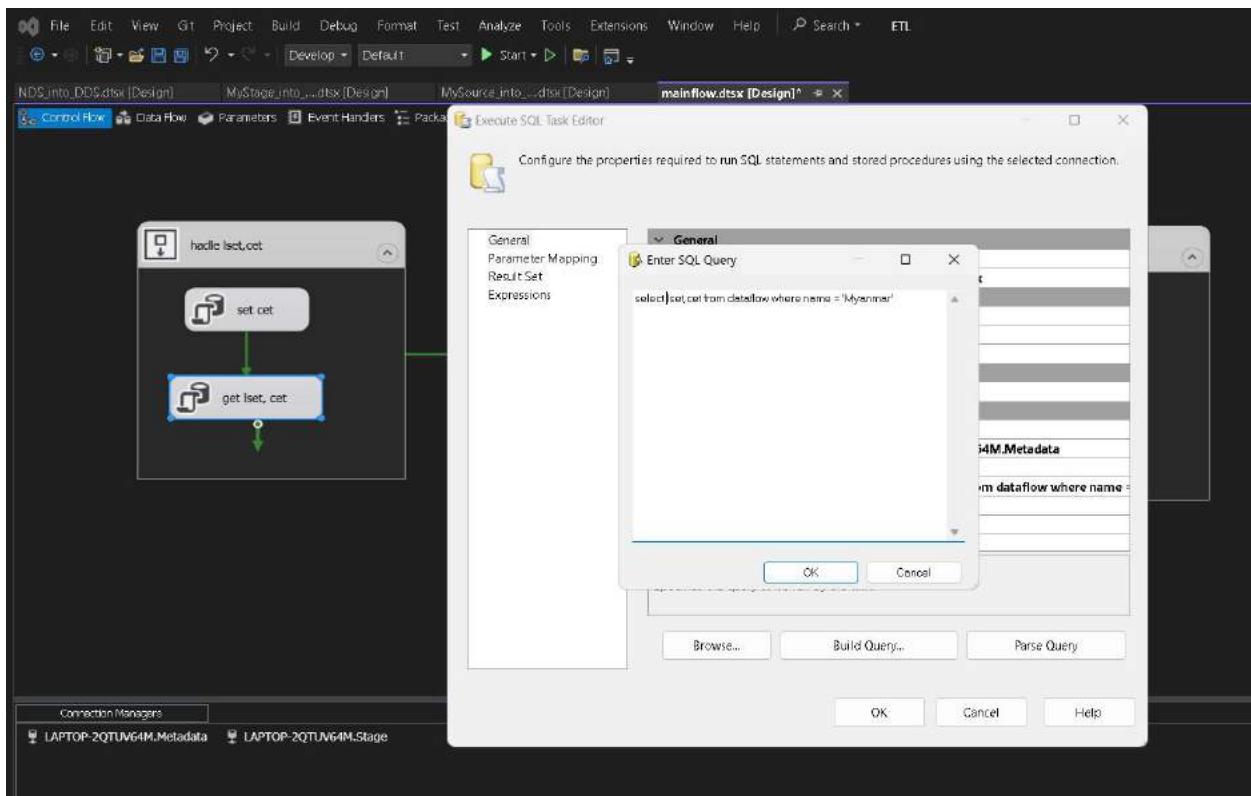
Là main flow quản lí cả quá trình ETL



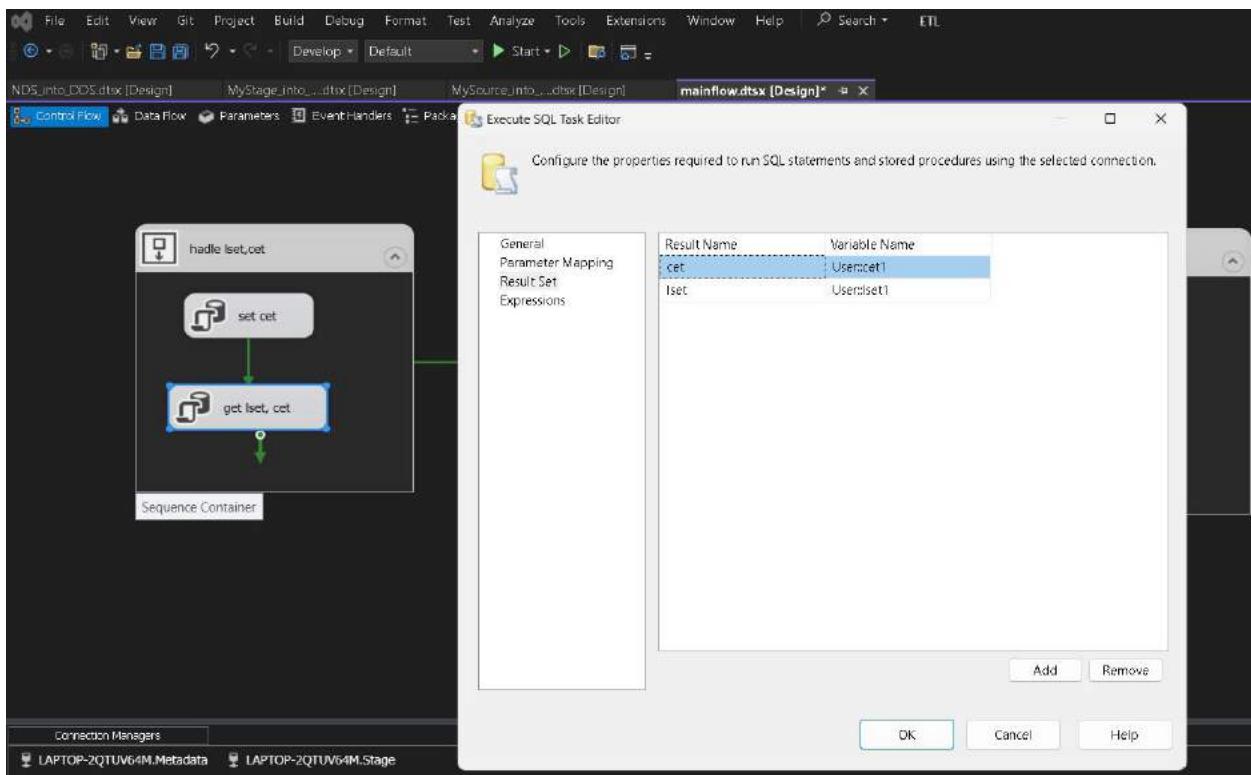
Đầu tiên , update cet=getdate() trong metadata



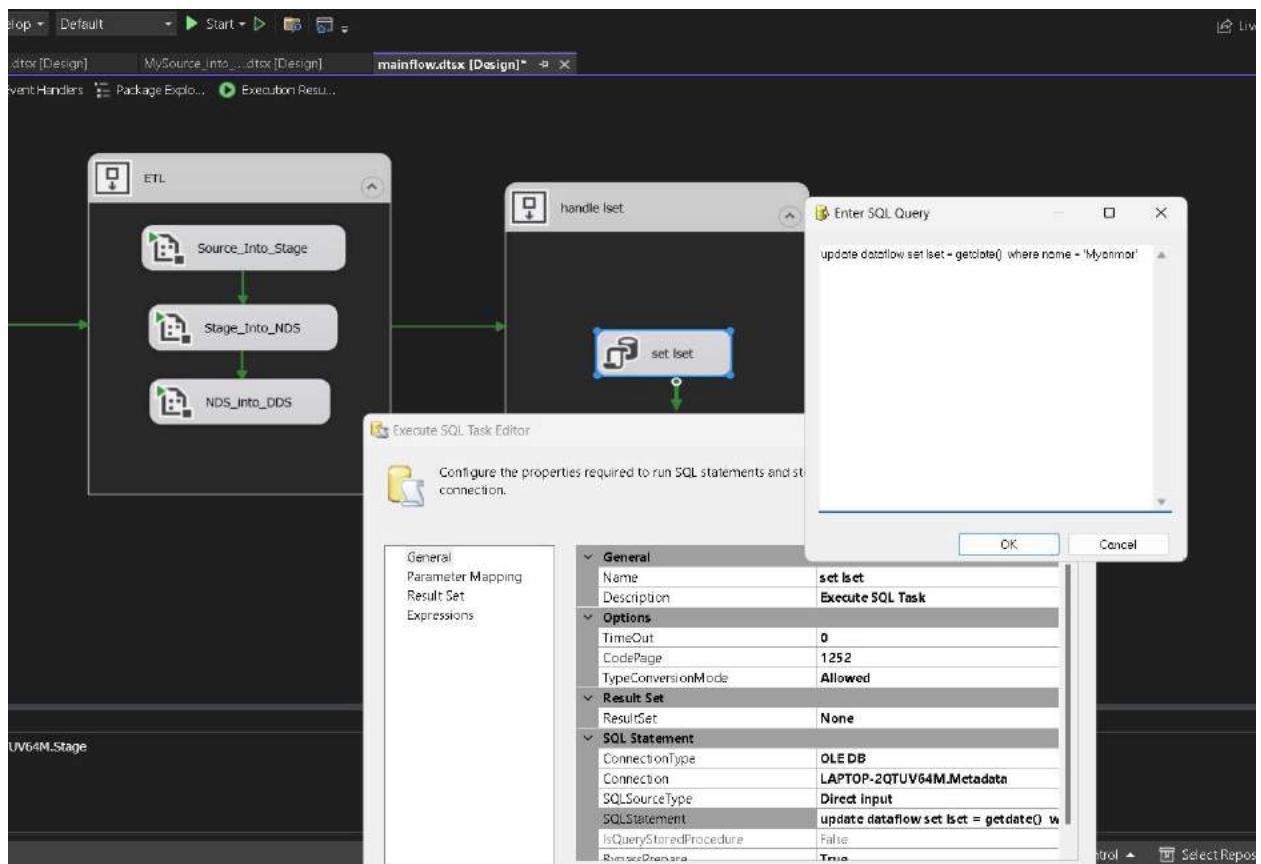
Select lset cet để sử dụng cho quá trình ETL



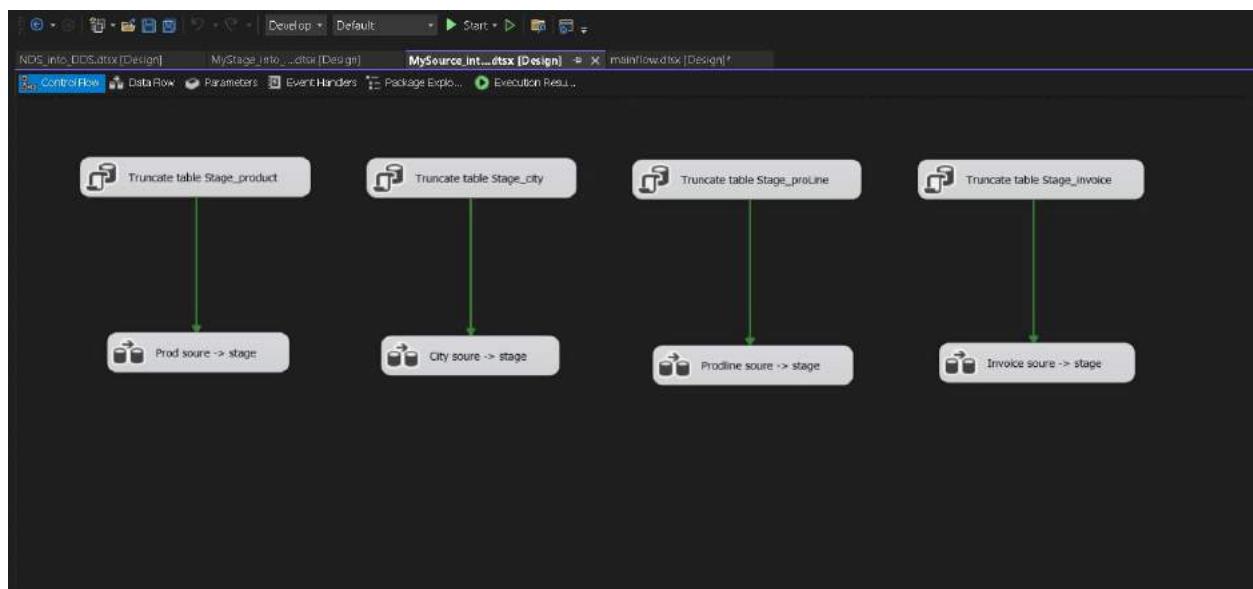
Lưu trữ lset ,cet vào 2 variable



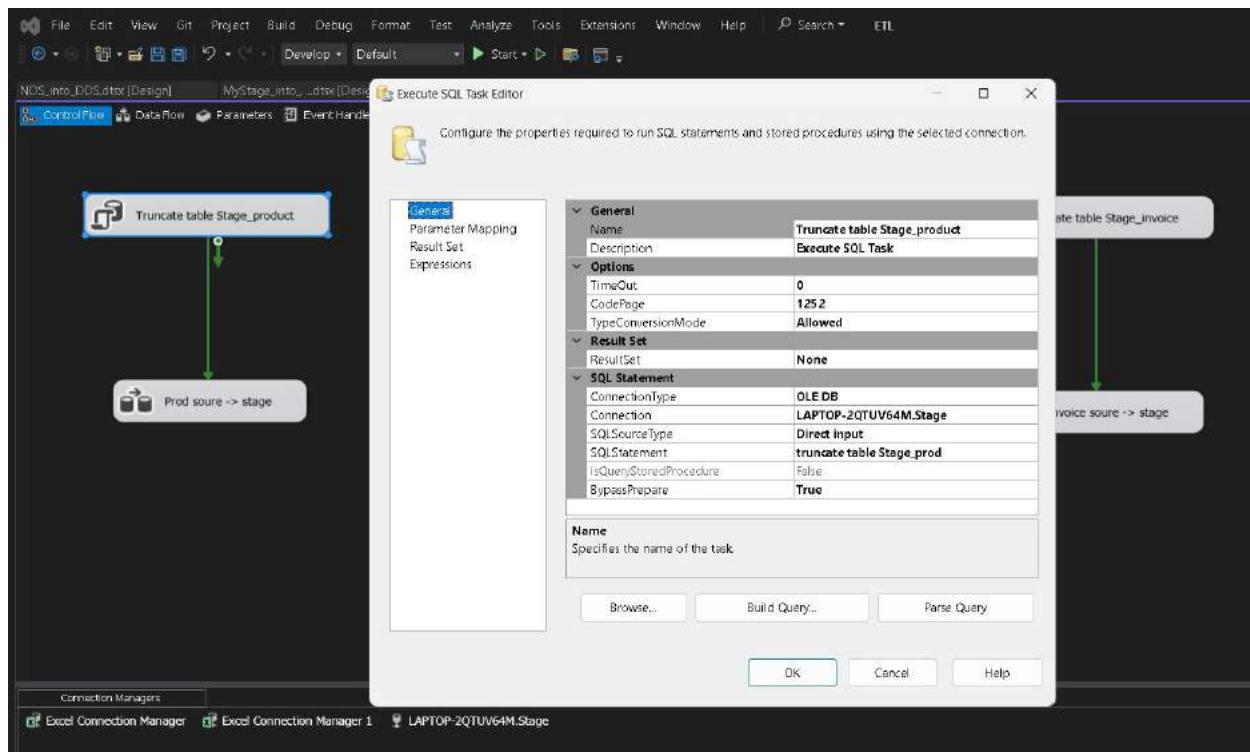
Sau khi quá trình ETL hoàn tất thì update lset = getdate() trong metadata



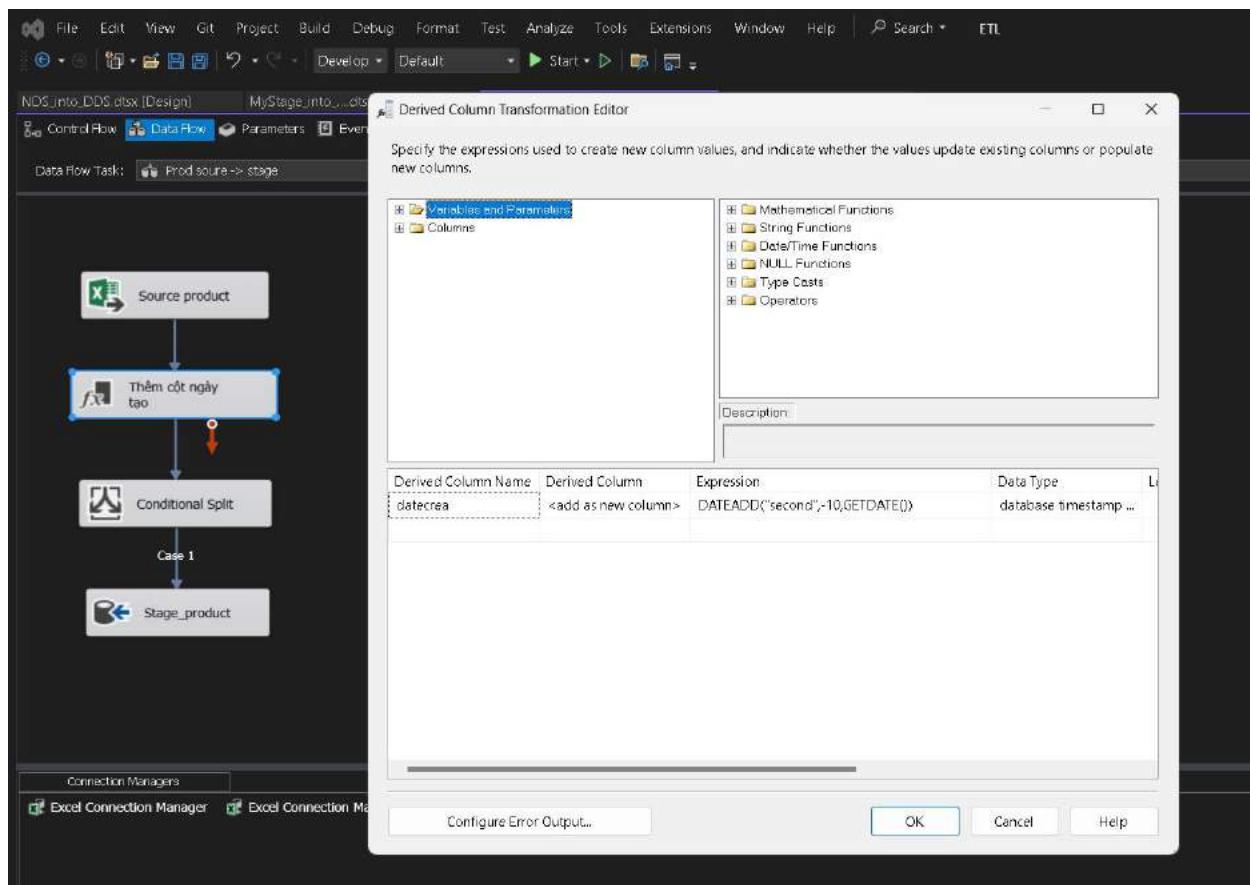
2. Từ Source vào stage



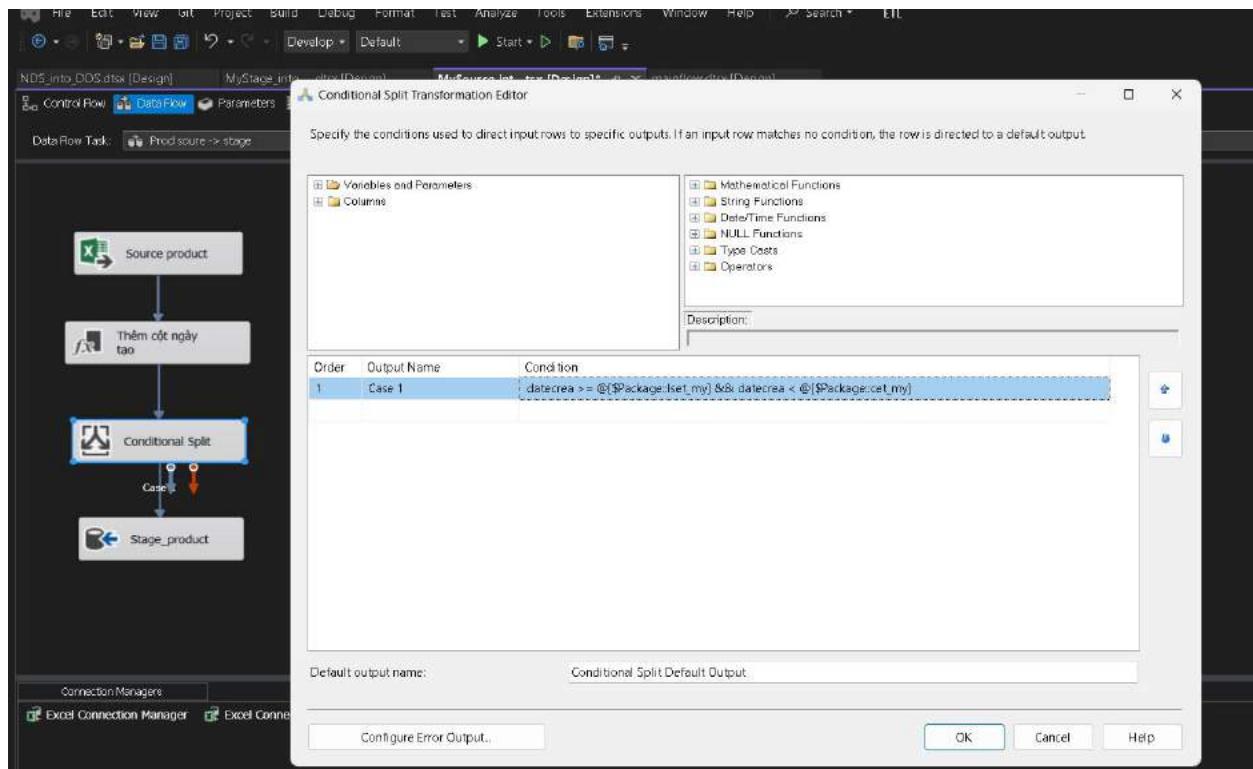
Đầu tiên truncate table stage_product để đảm bảo stage sẽ trống rỗng và không chứa dữ liệu lỗi hoặc cũ từ quá trình etl trước đó.



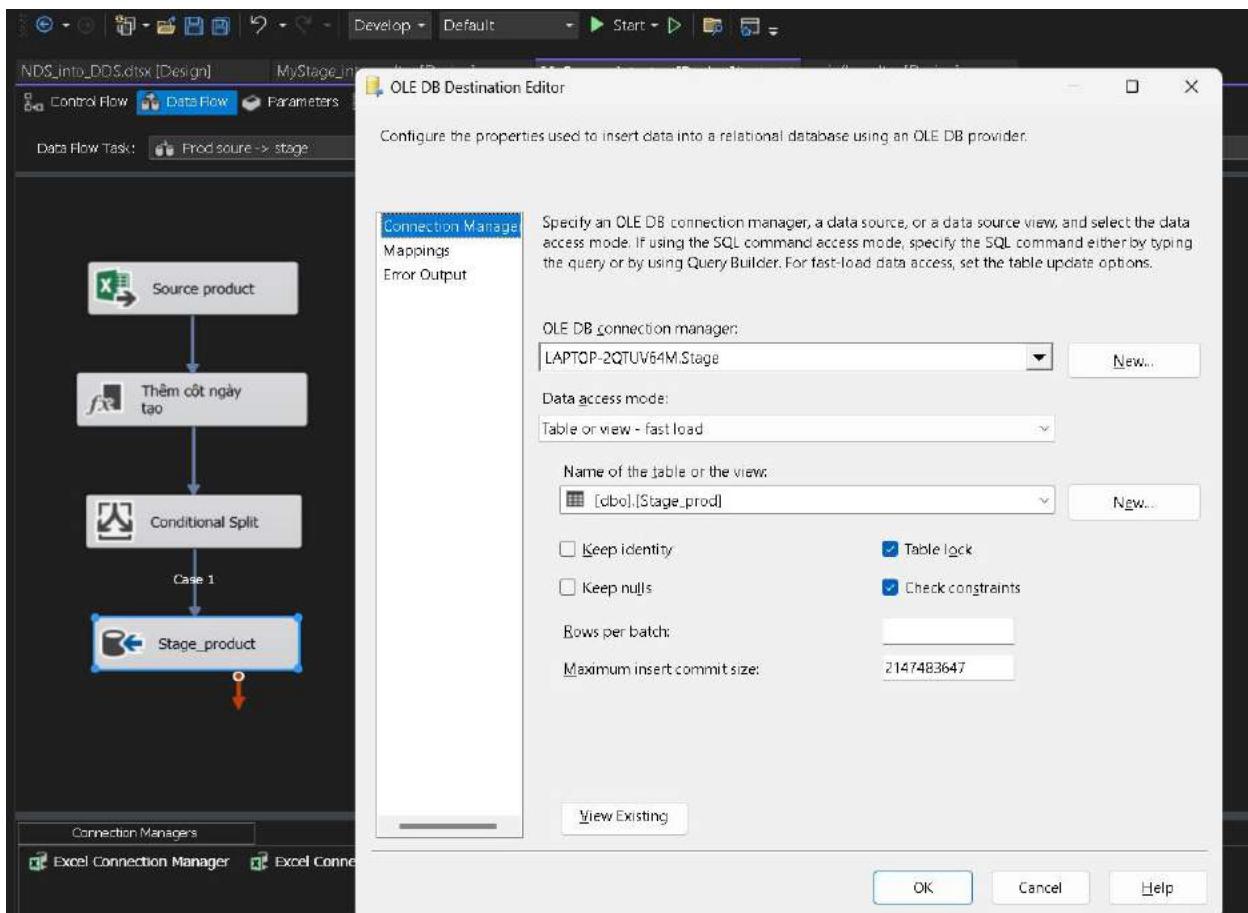
Đầu tiên chọn source product, tiếp theo ví dụ thêm cột date create = getdate() - 10s để bé hơn cete thì dữ liệu mới lưu vào stage

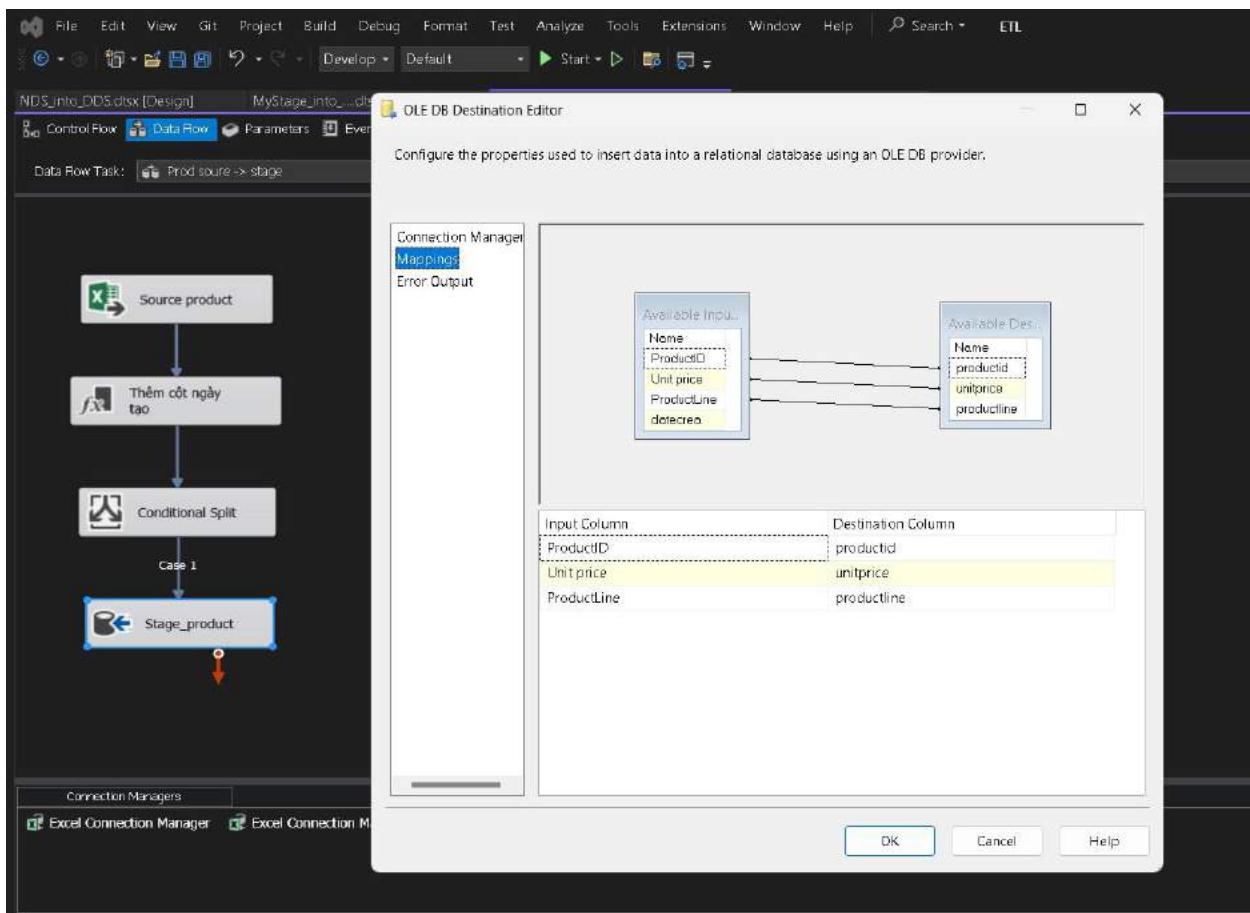


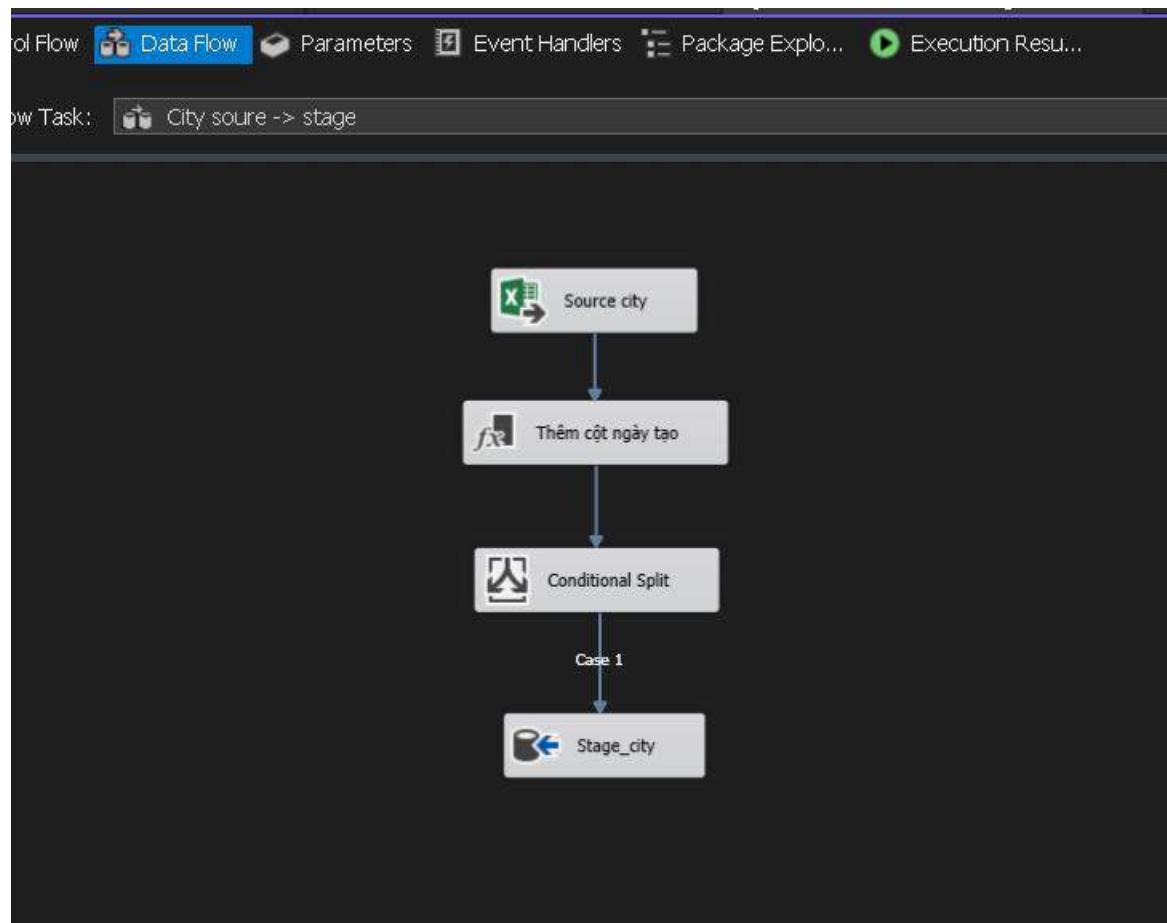
Đặt điều kiện datecrea \geq lset và datecrea $<$ cet , nếu thỏa điều kiện trên thì mới lưu dữ liệu vào stage_product

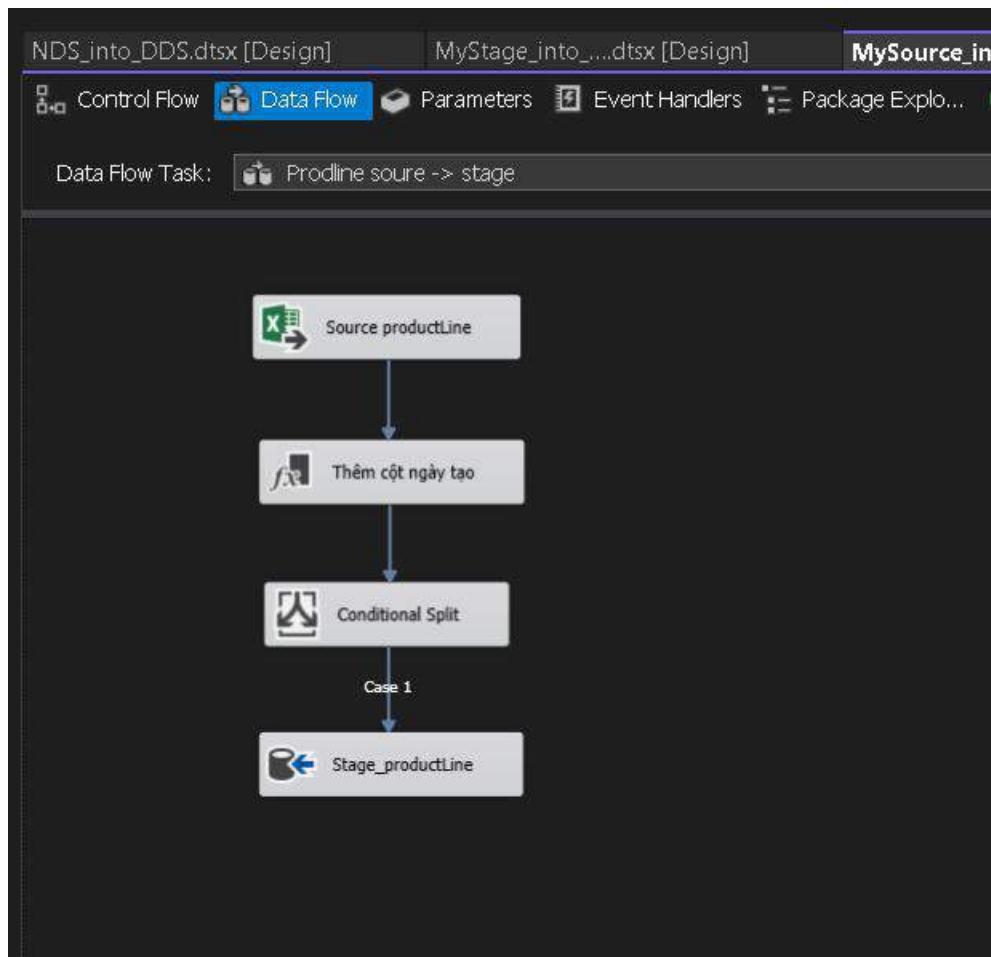


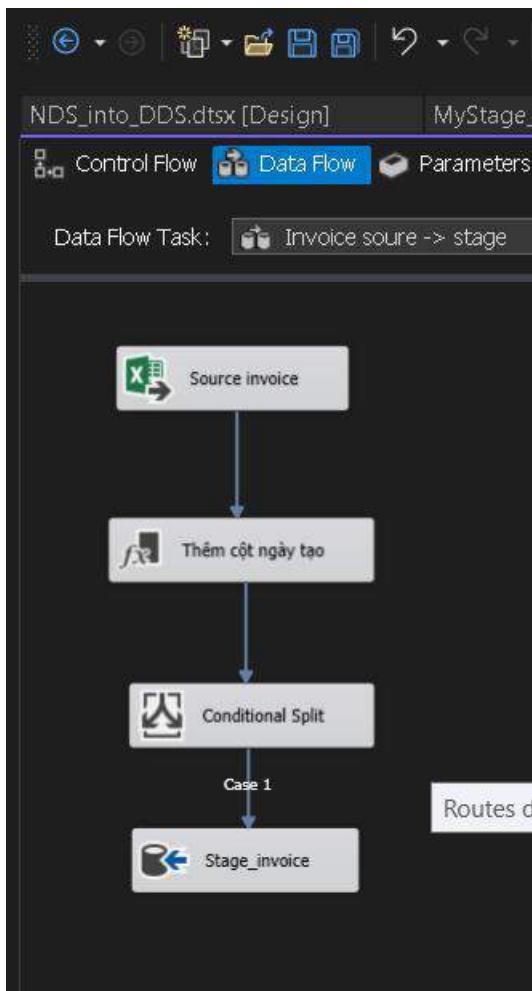
Tiếp theo mapping các thuộc tính với nhau. Tương tự các bảng còn lại cũng làm như vậy







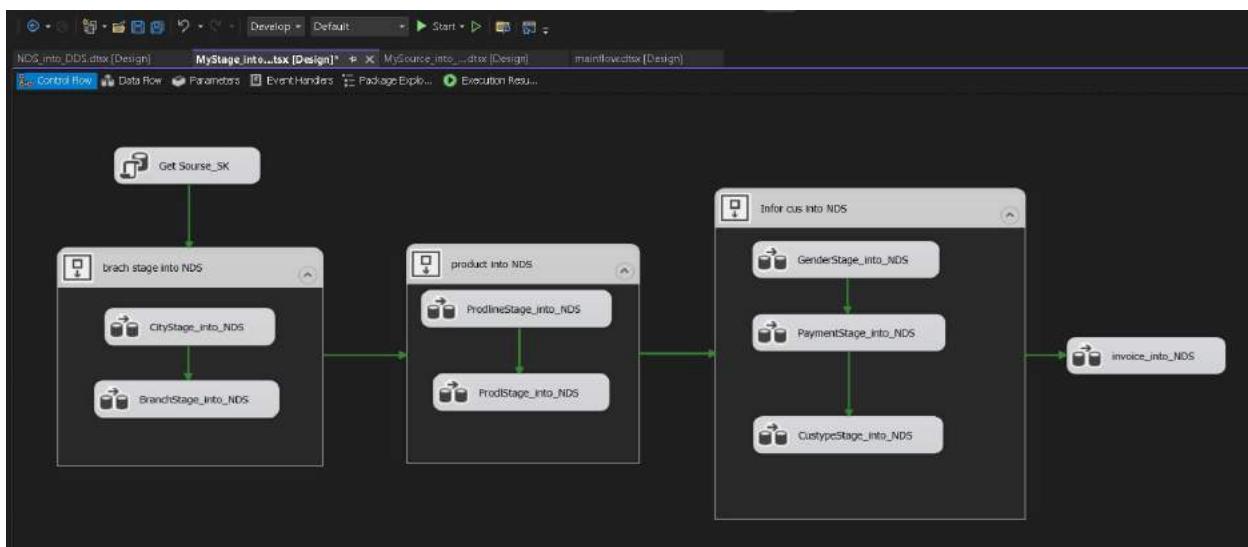
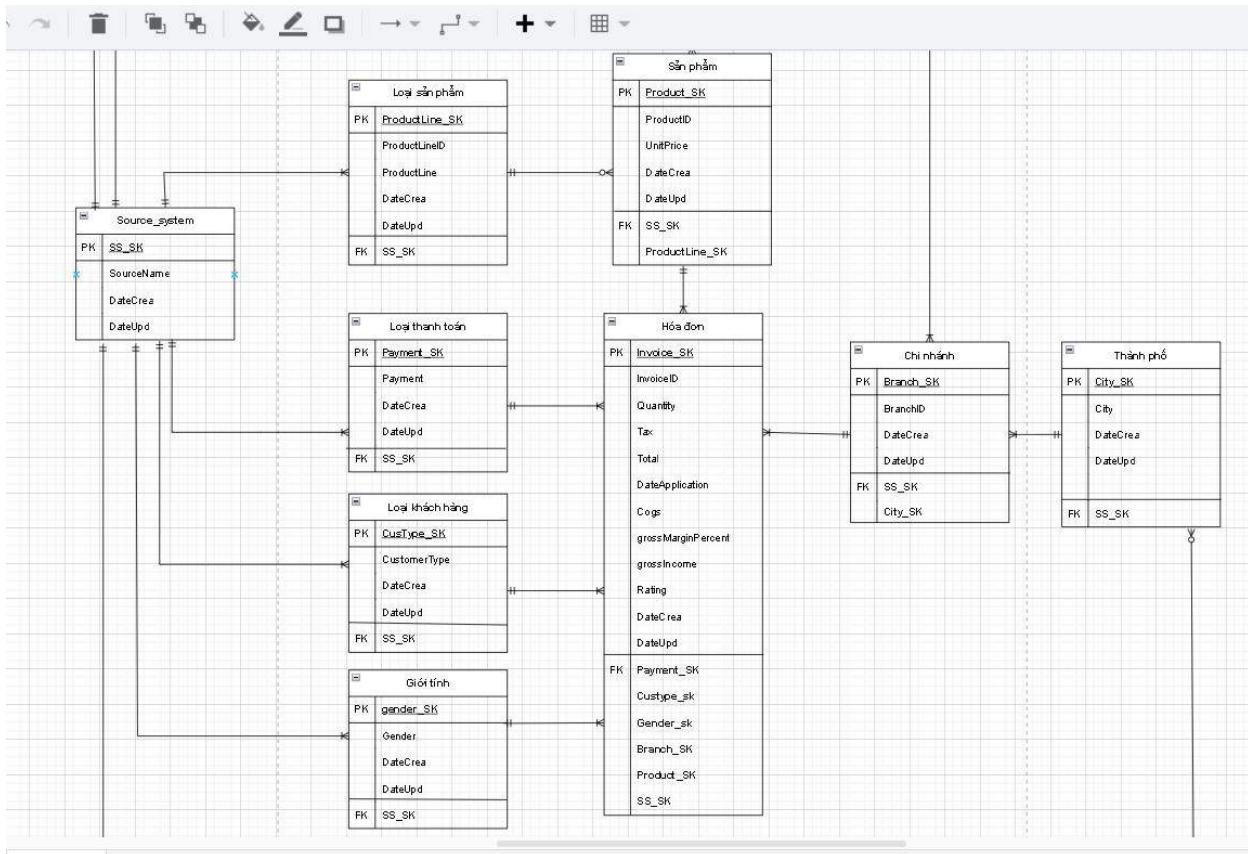




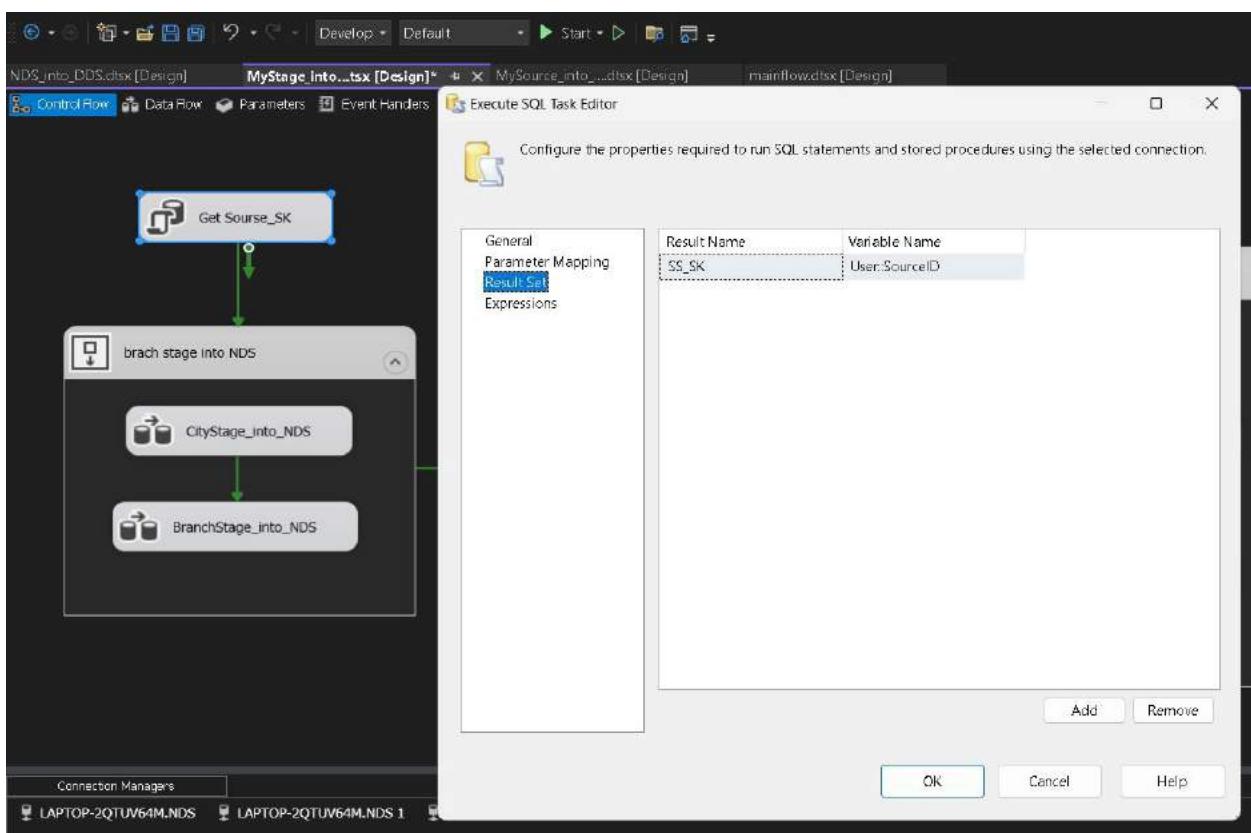
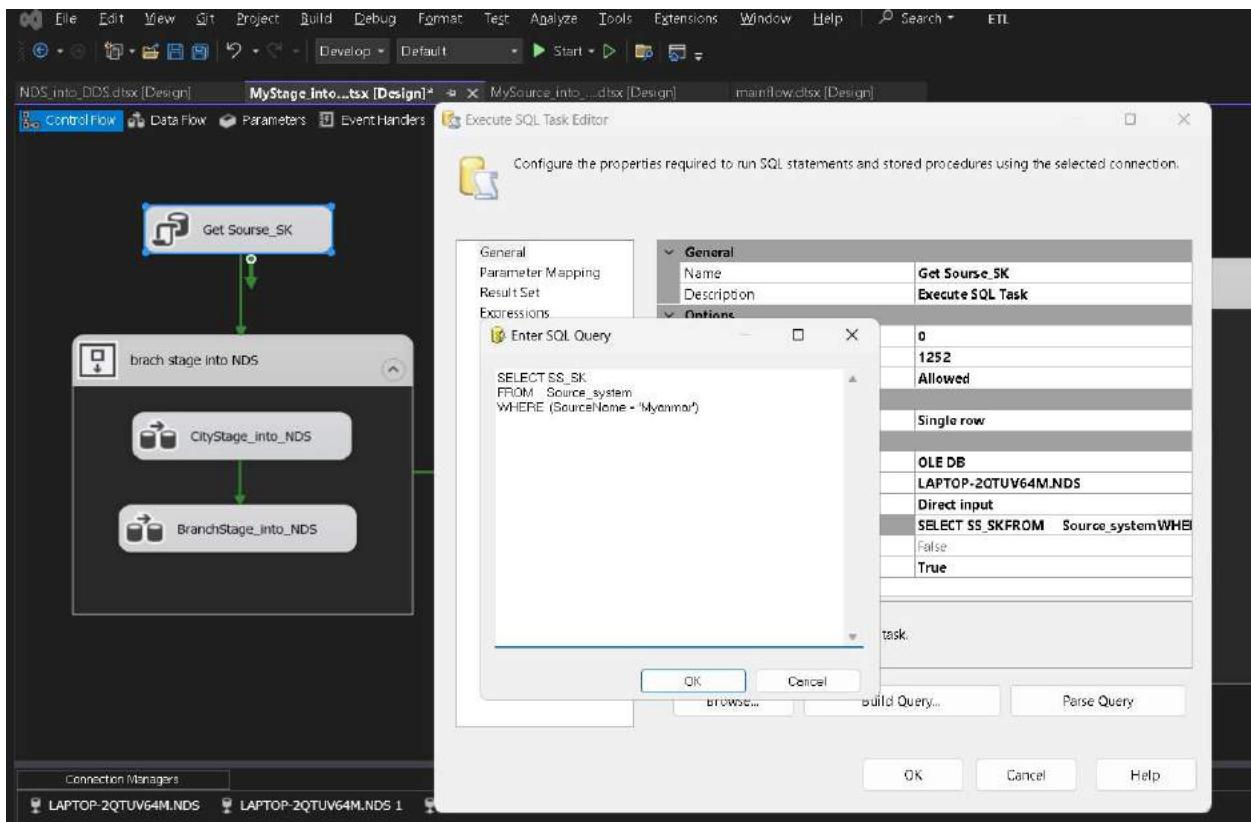
3. Từ Stage vào NDS

Thiết kế NDS:

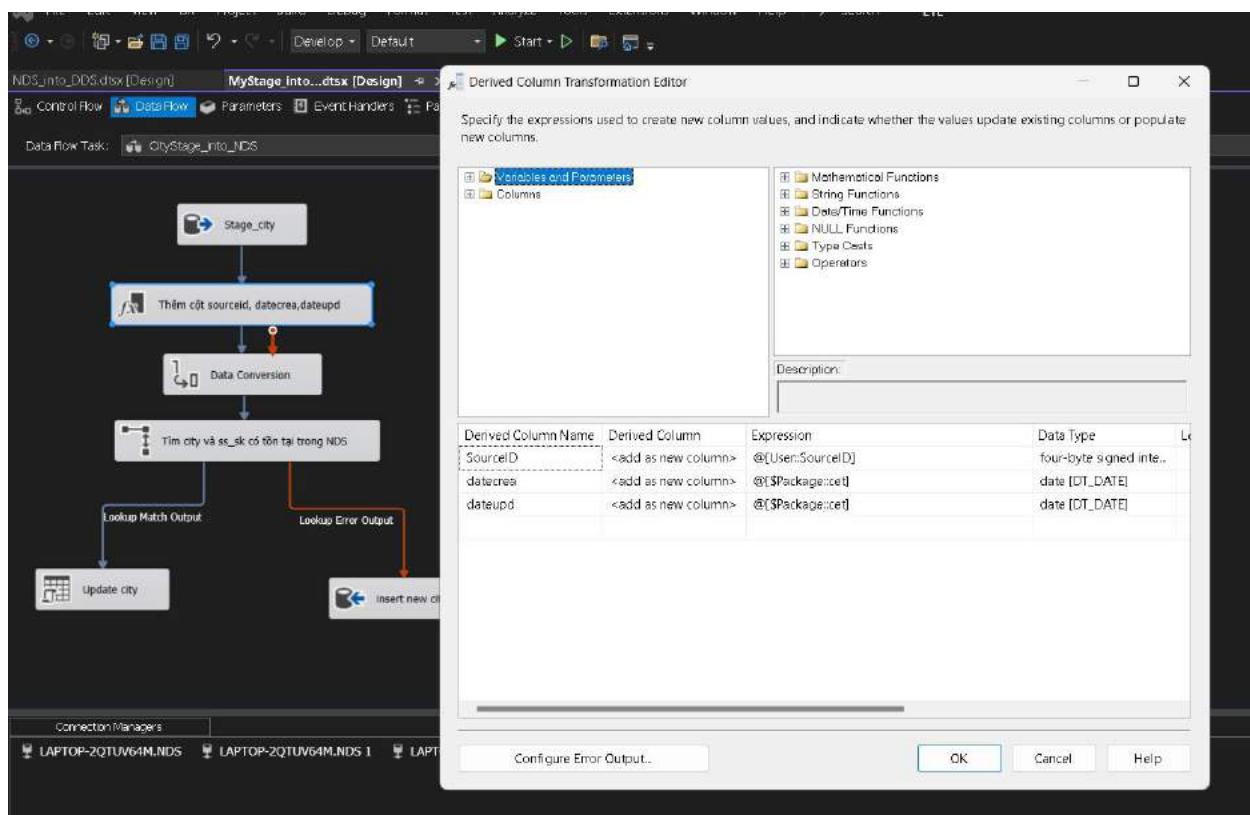
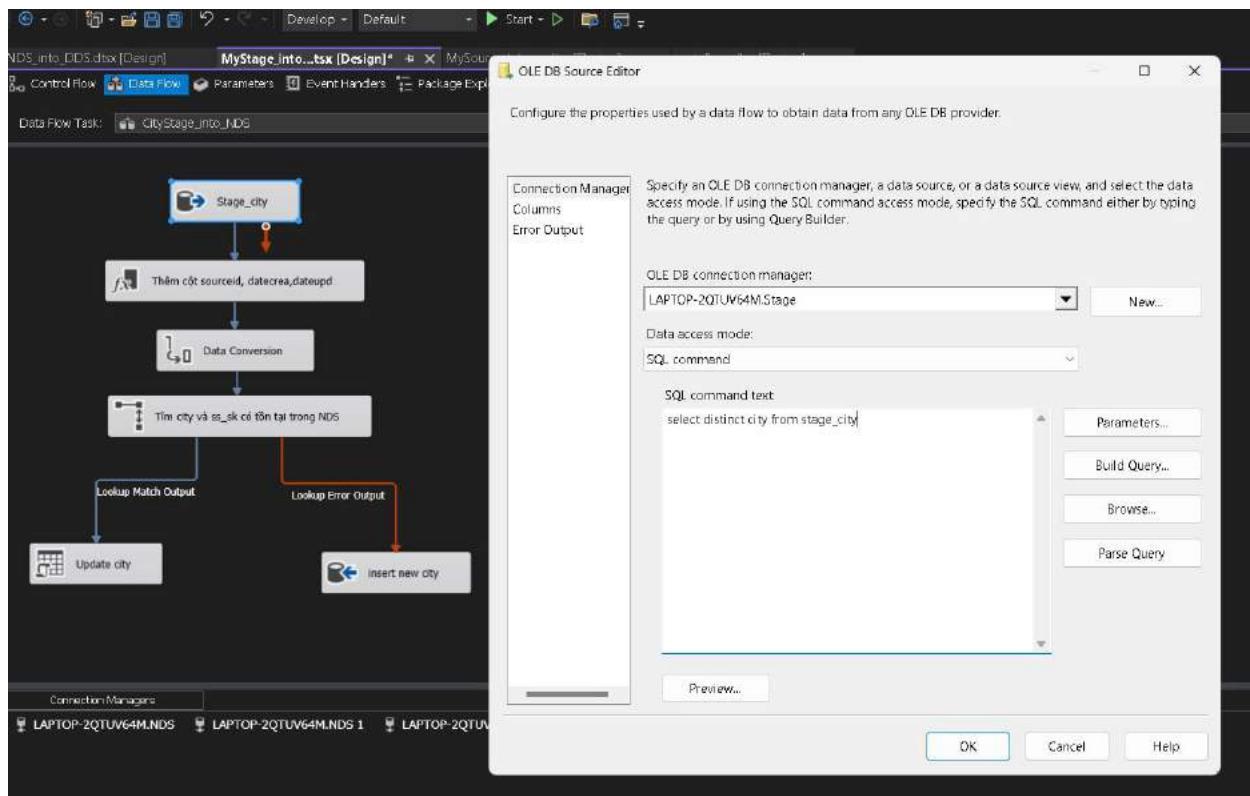
https://drive.google.com/file/d/1HXpG7uiR_AeH8G80lPRWfESjbTjA3LnD/view?usp=sharing



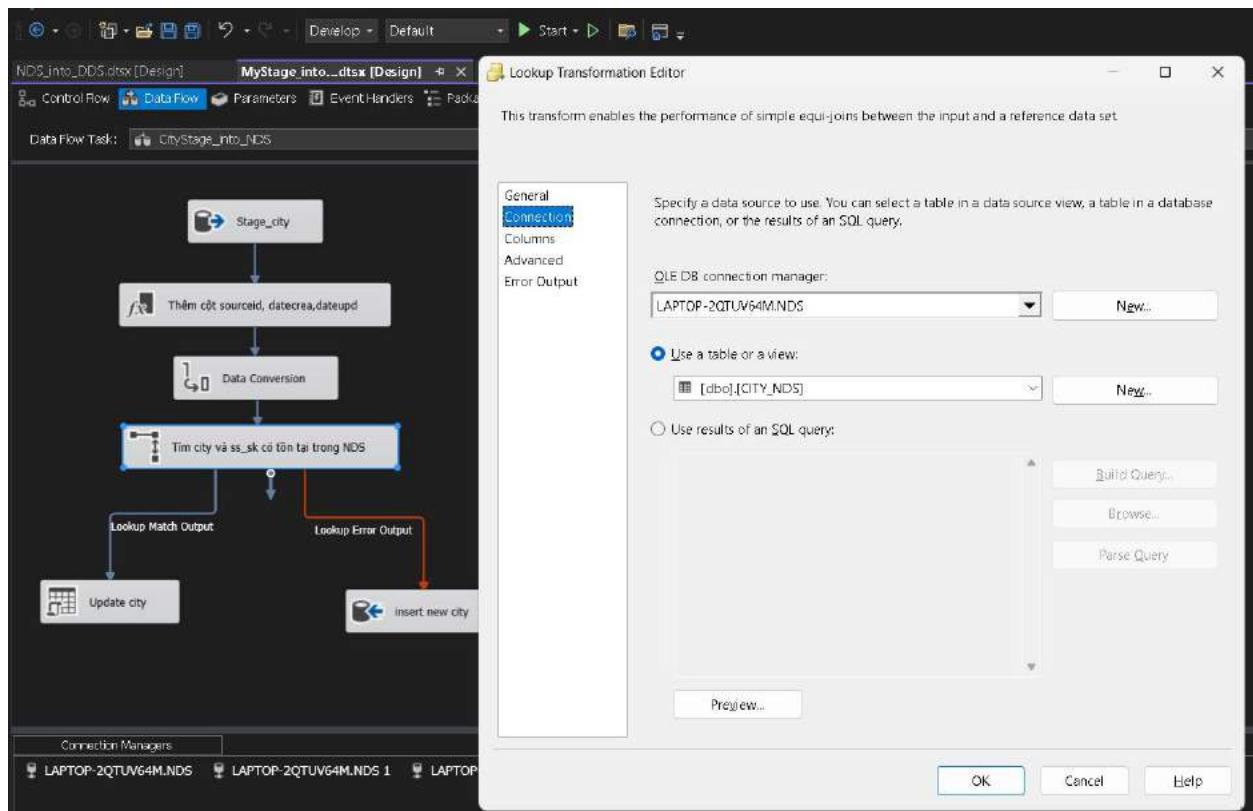
Đầu tiên select lấy source_sk và lưu ss_sk vào variable

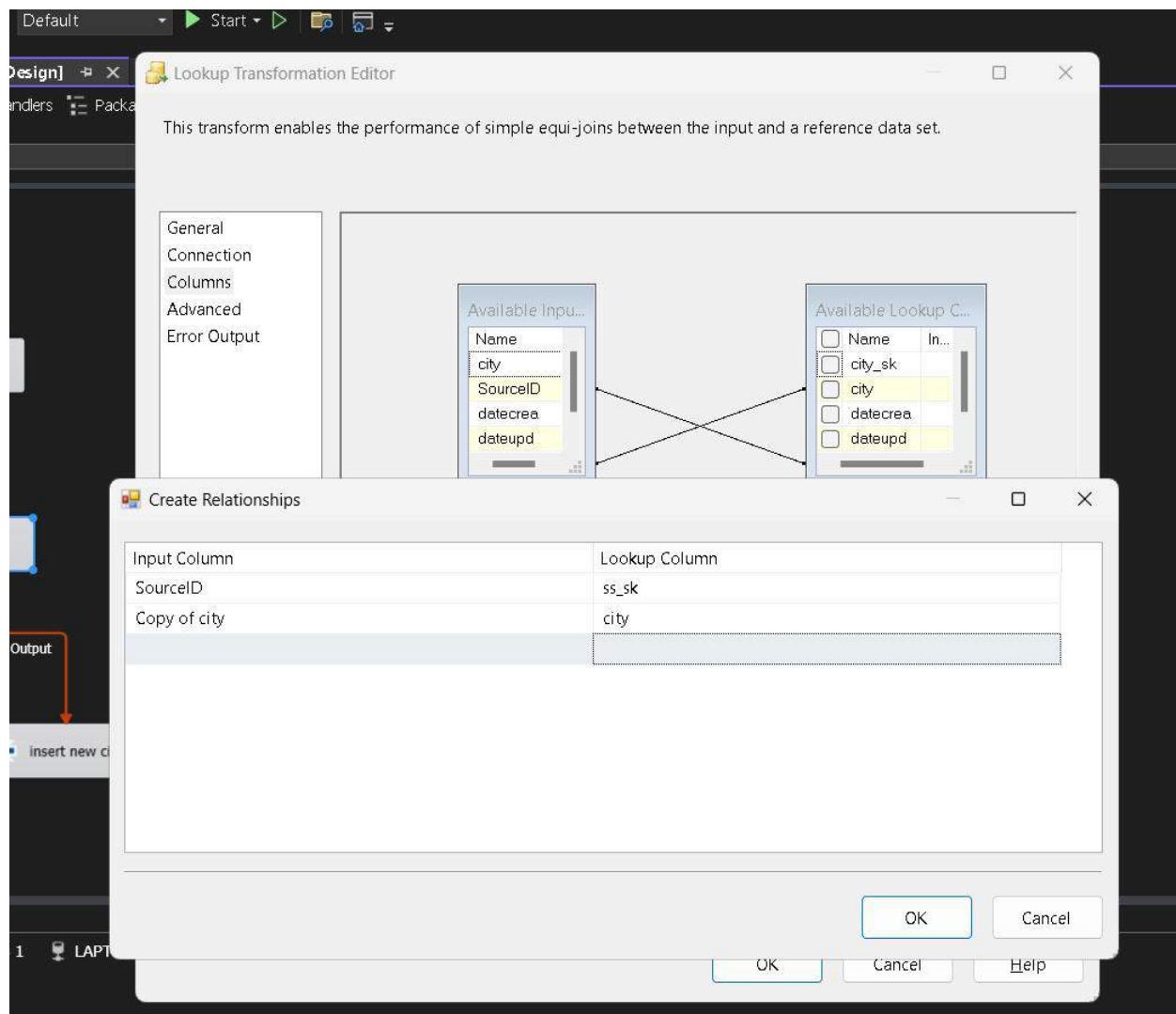


Tiếp theo stage_city vào city_nds, thêm cột source_id, datecreate, dateupdate

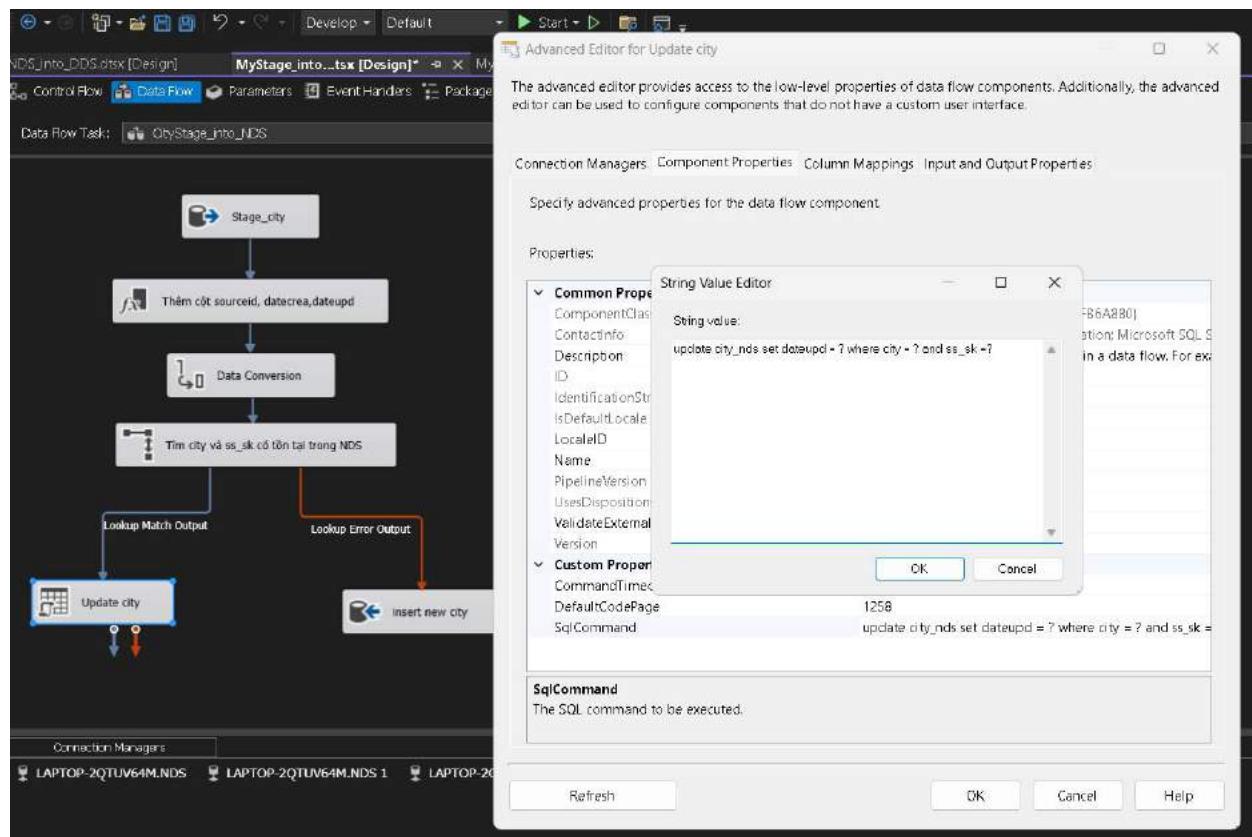


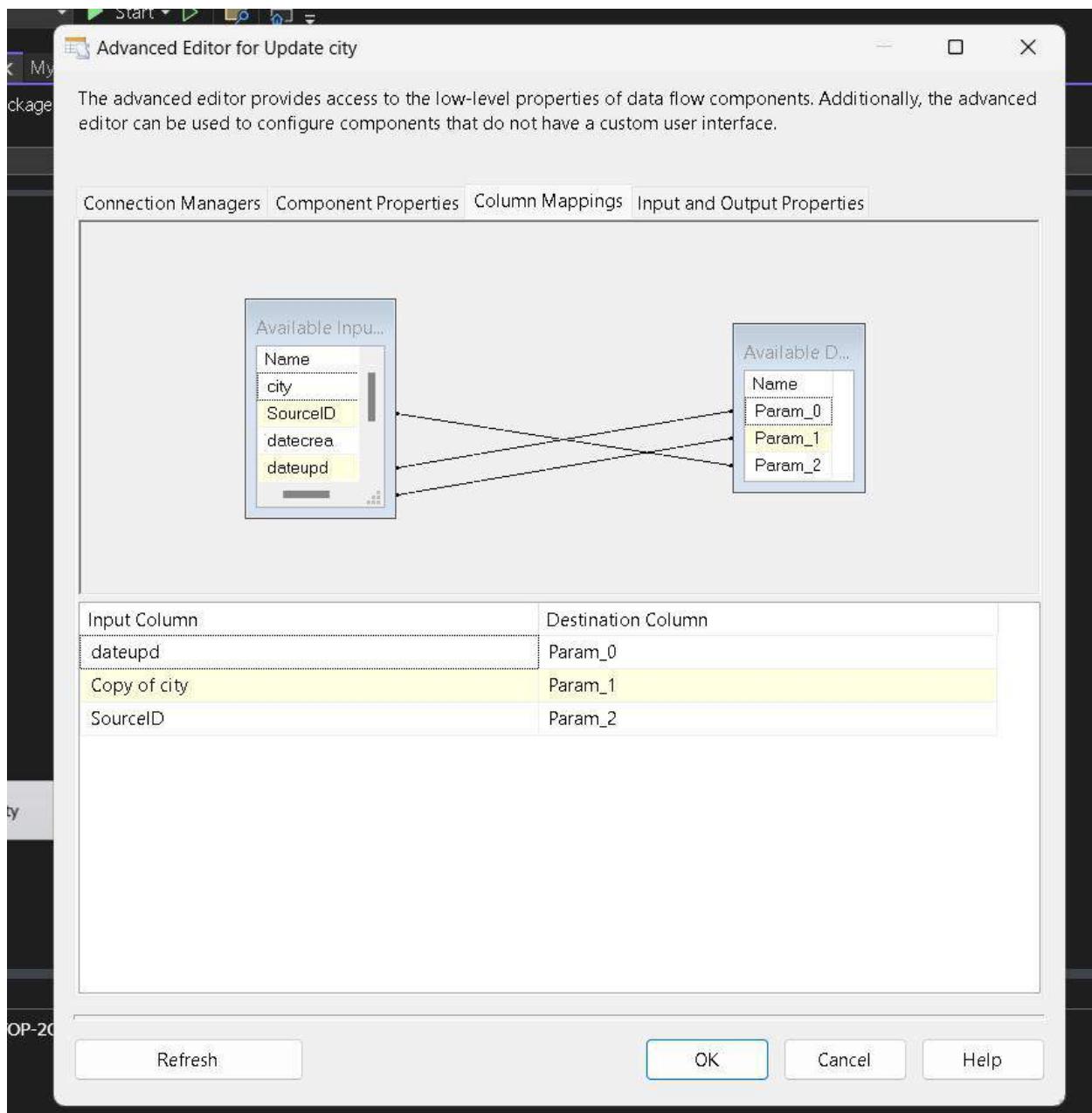
Dùng lookup để tìm city và sourceid có tồn tại trong City_nds



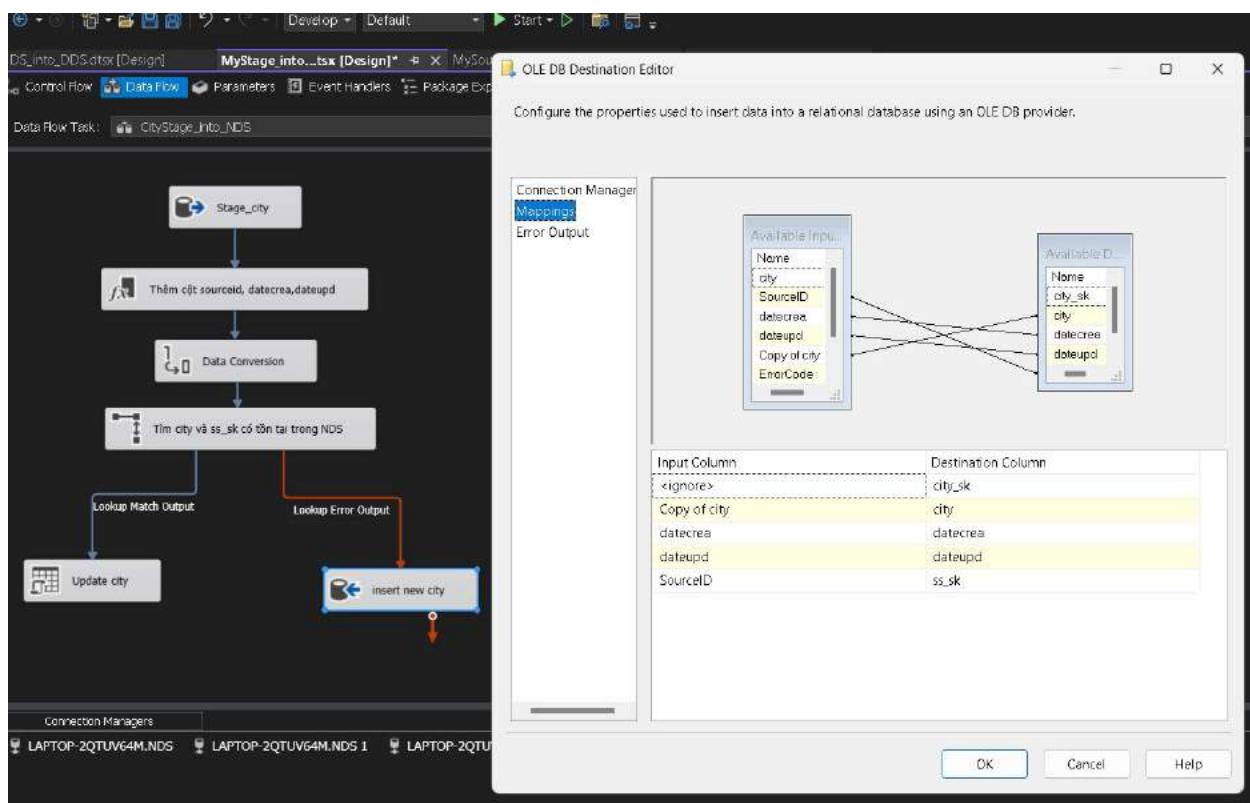
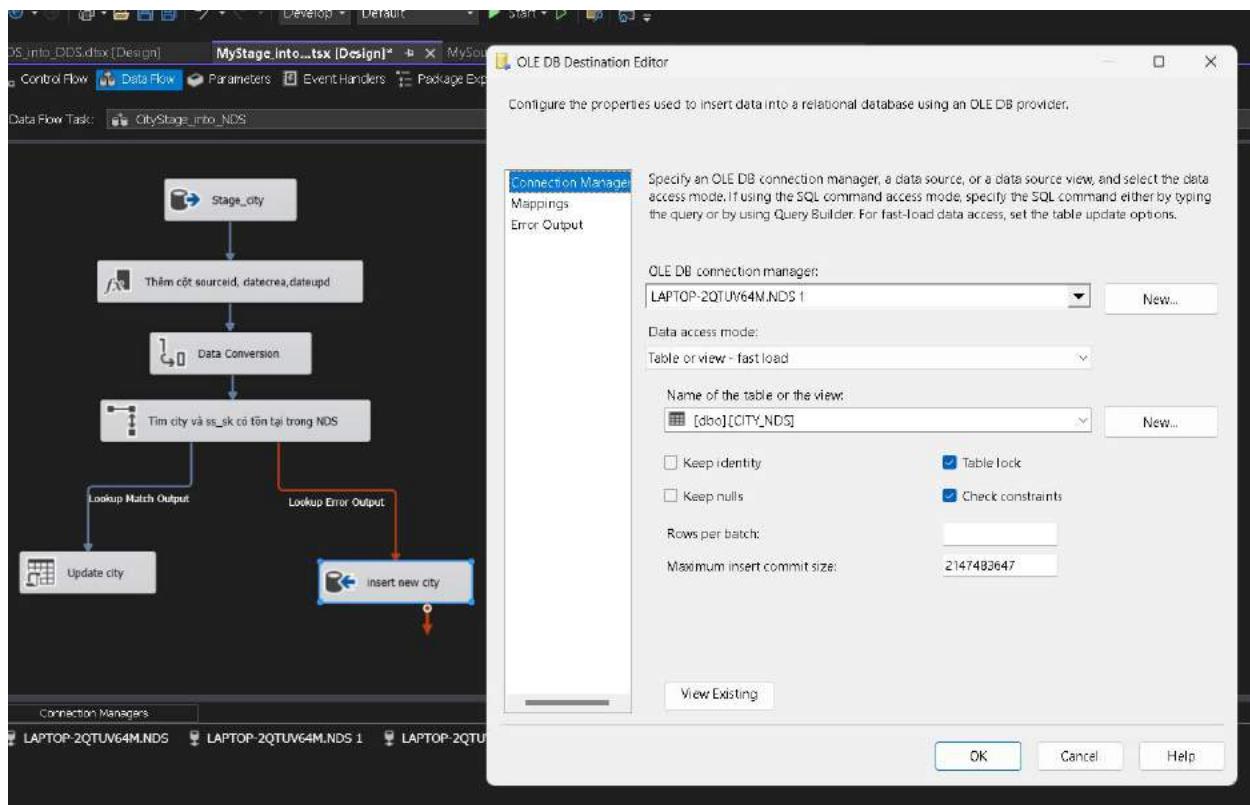


Nếu tồn tại thì update city

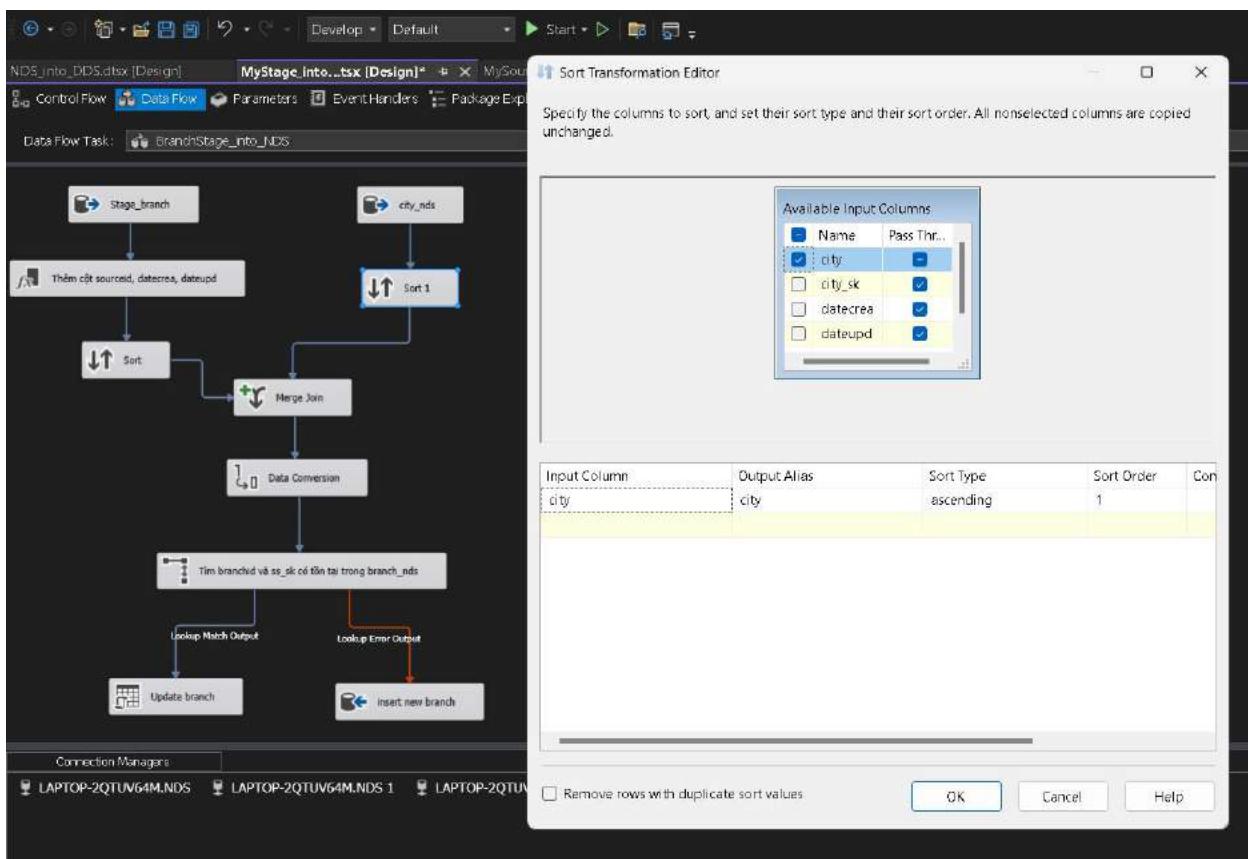
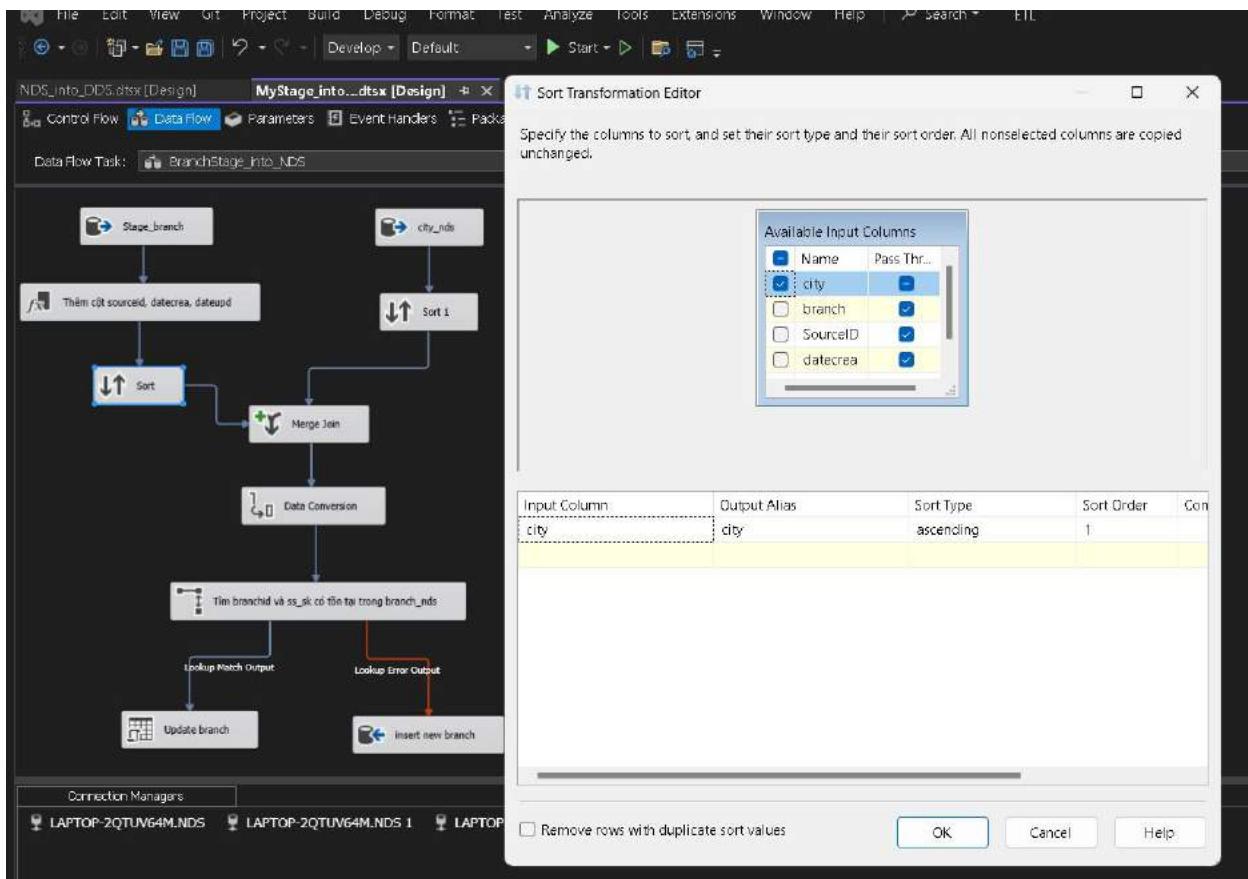


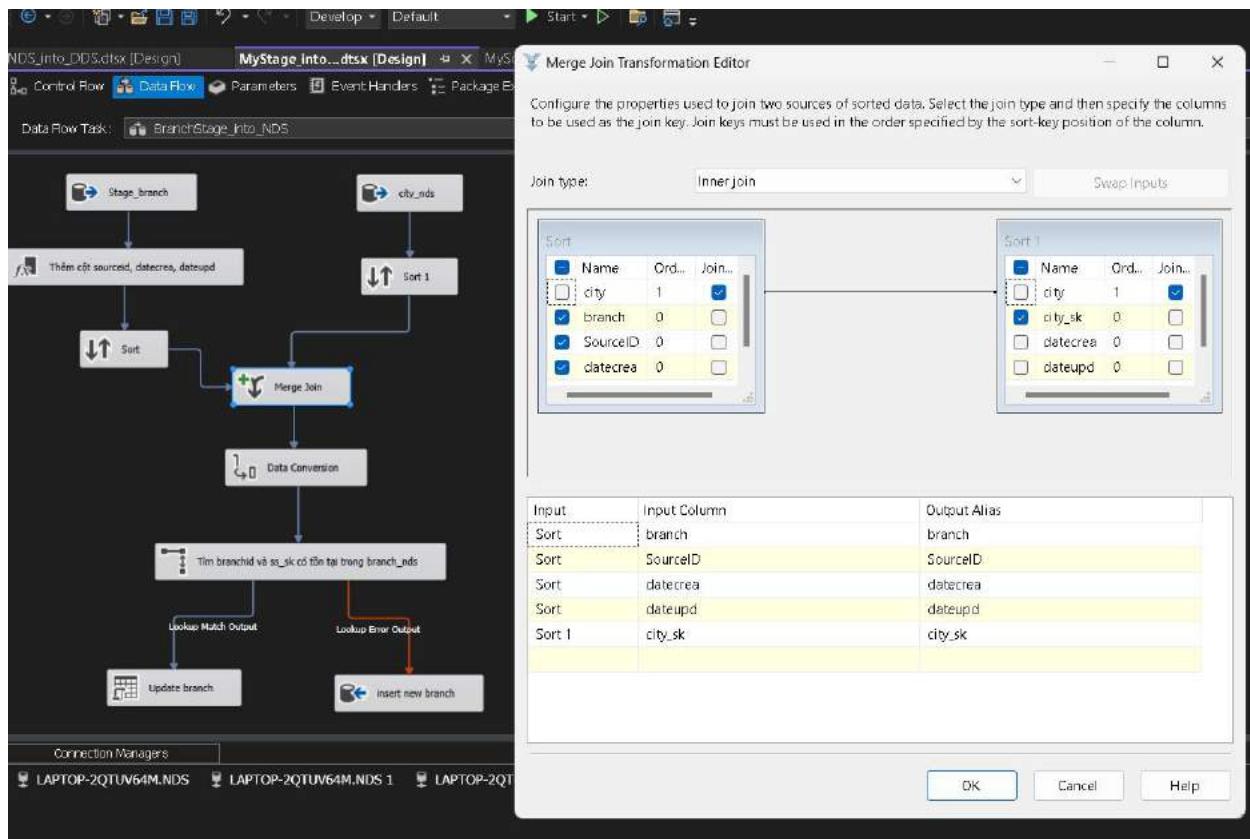


Còn không tồn tại thì thêm city mới vào City_nds



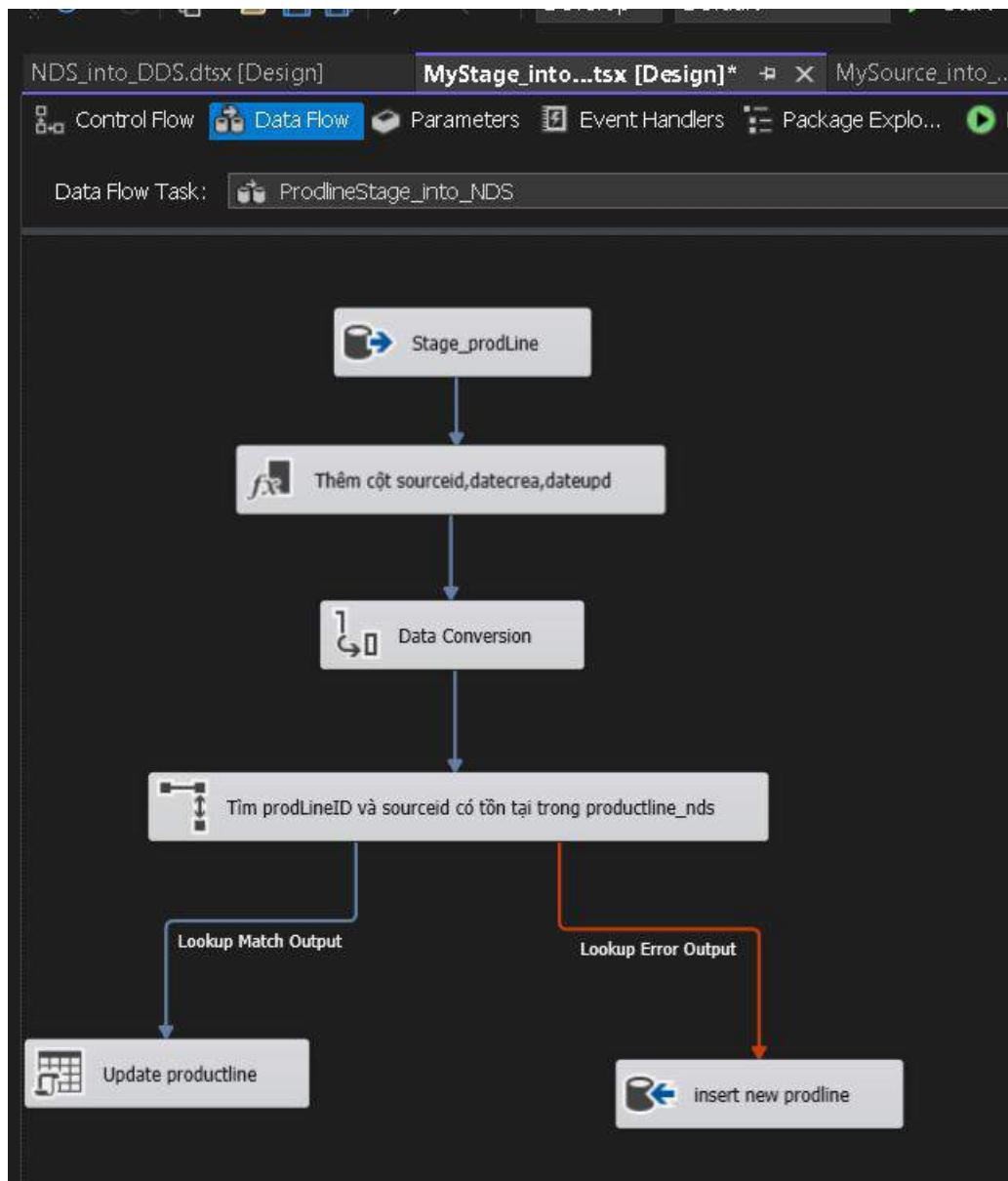
Tiếp theo Branch_stage vào Branch_nds, sau khi có dữ liệu ở city_nds thì branch_stage merge join với city_nds theo city để lấy surrogate key



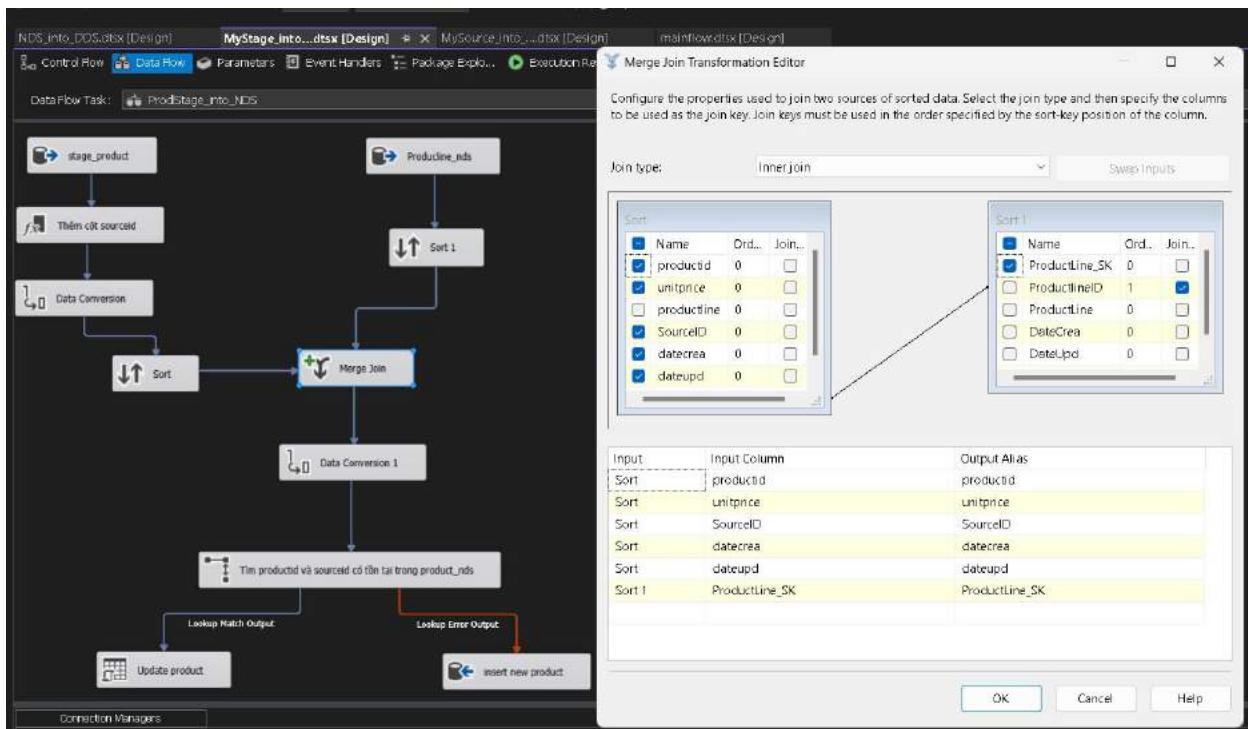


Tương tự như city, tìm branchid và sourceid có tồn tại trong branch_nds không. Nếu tồn tại thì update branch, còn chưa có thì thêm mới vào branch_nds

Productline_stage lưu vào Productline_nds

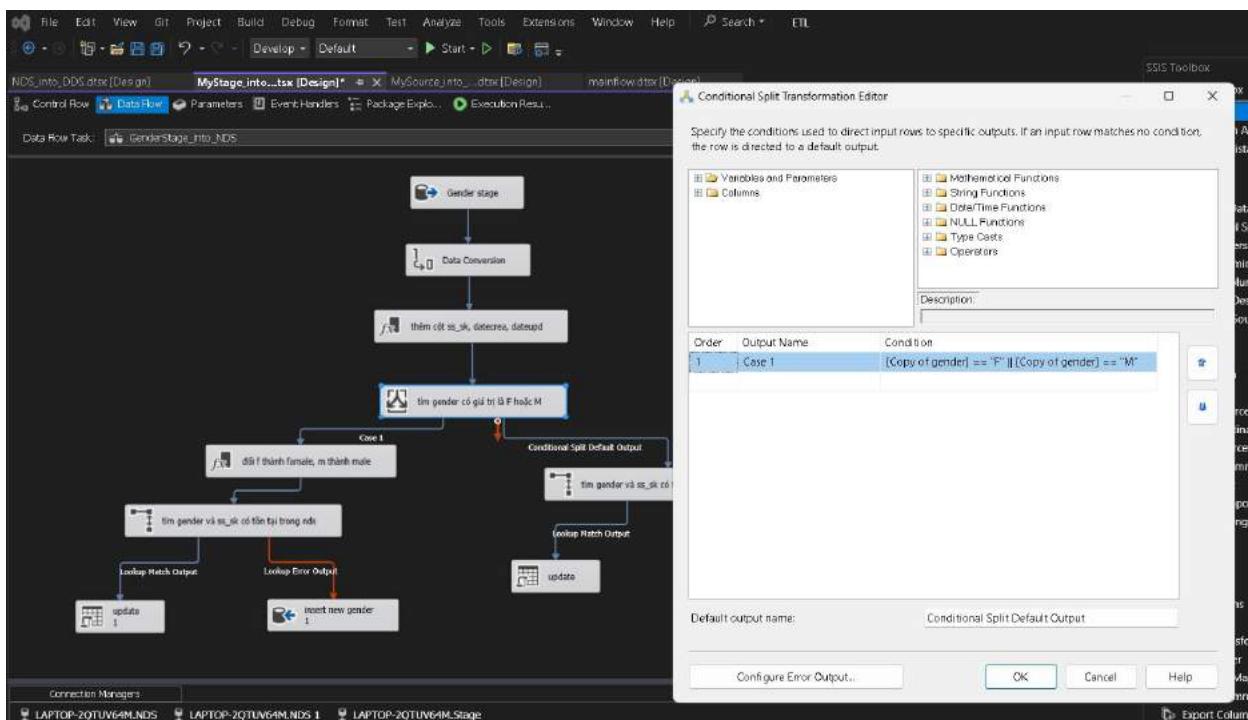


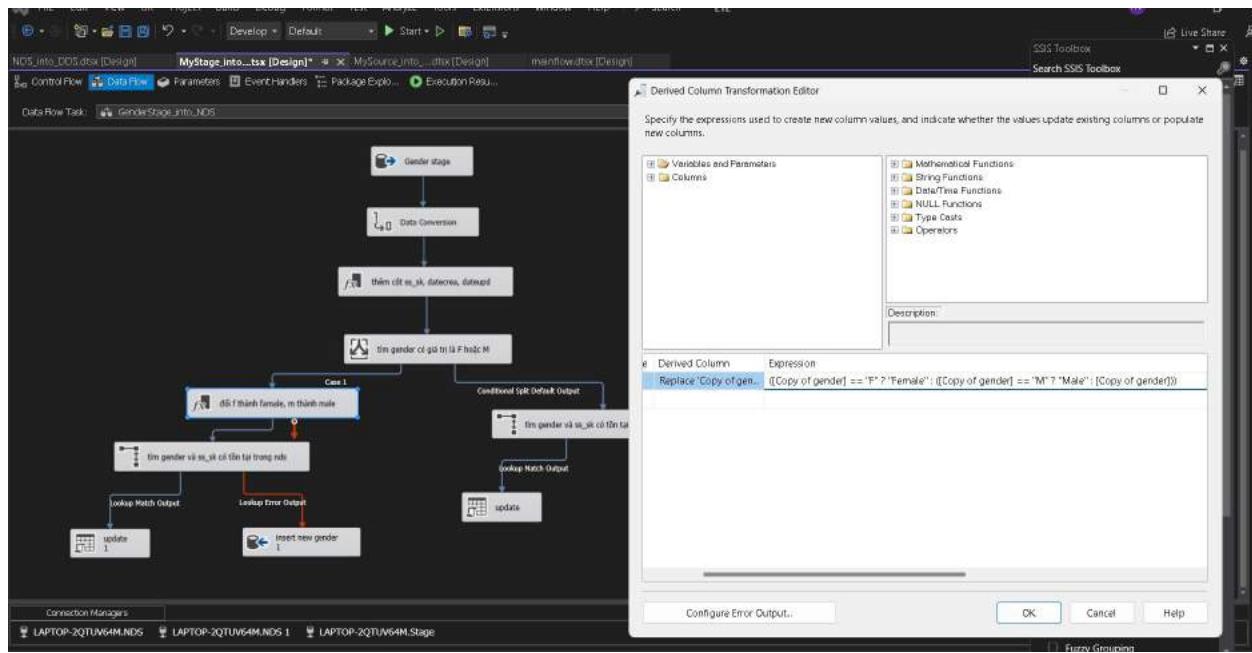
Product_stage lưu vào product_nds, merge join với prductline_nds theo productlineid lấy surrogatekey của productline_nds



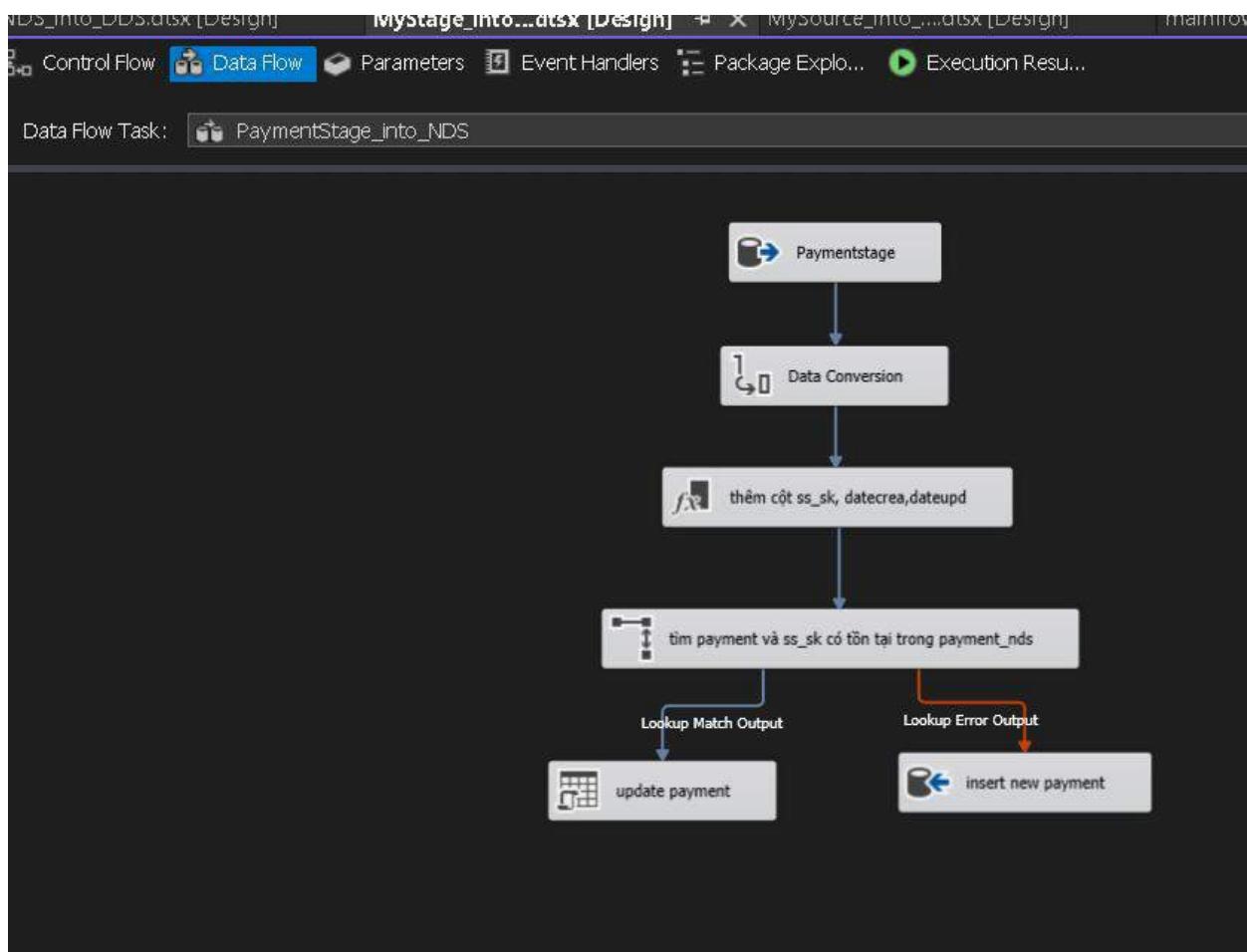
Gender_stage vào gender_nds

Tìm gender có giá trị F hoặc M, sau đó chuyển f thành famale, m thành male

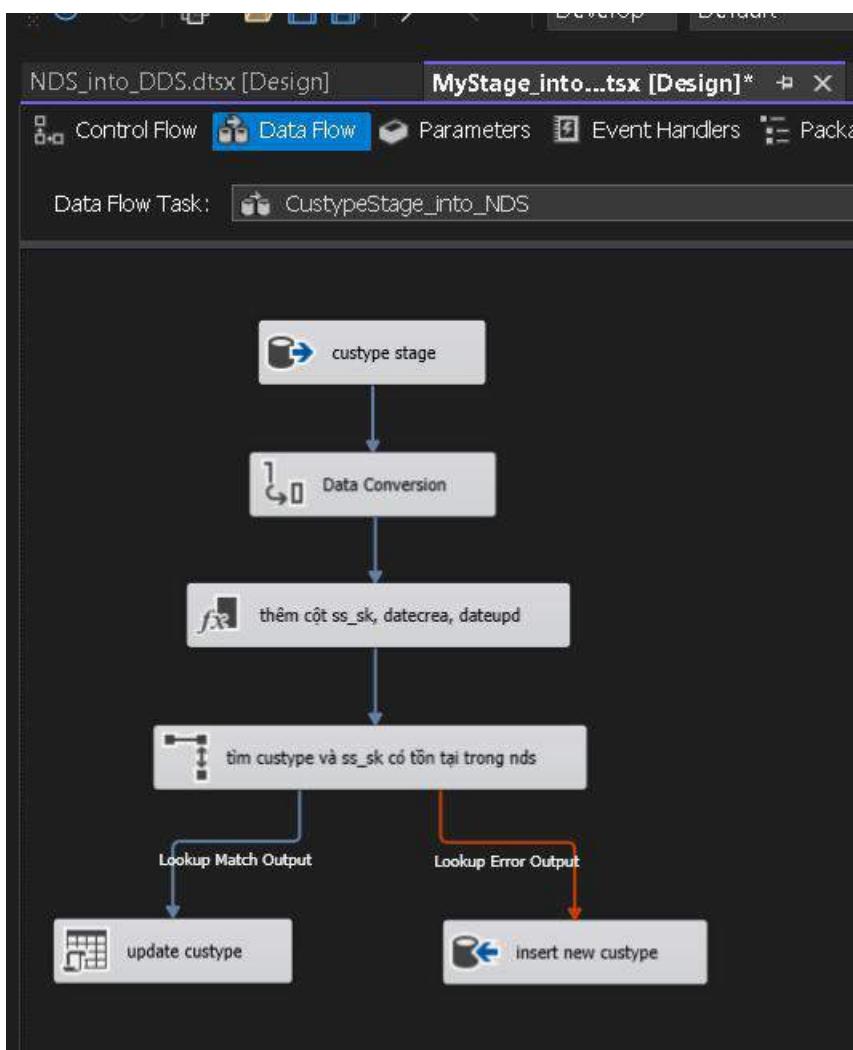




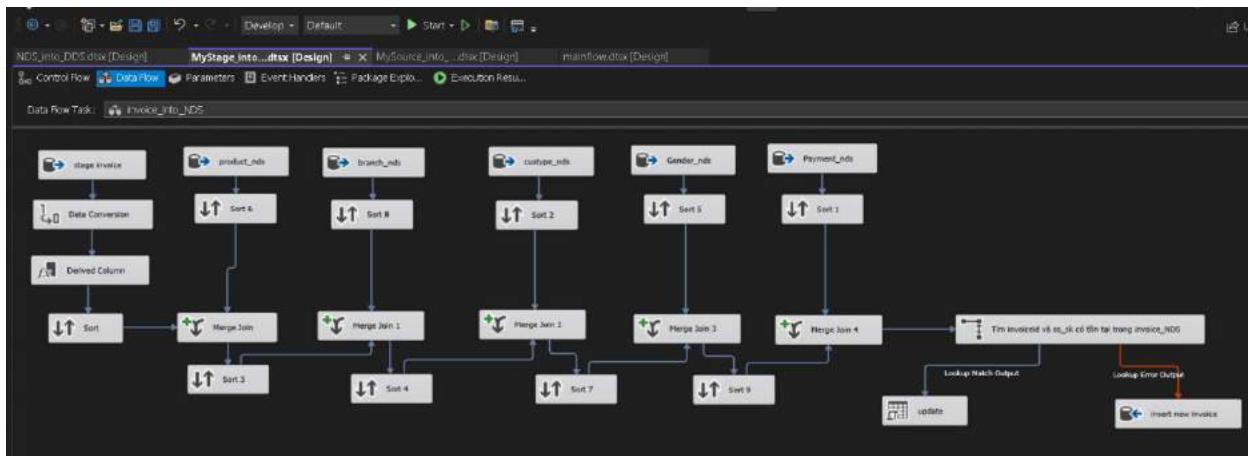
Payment_stage vào payment_nds



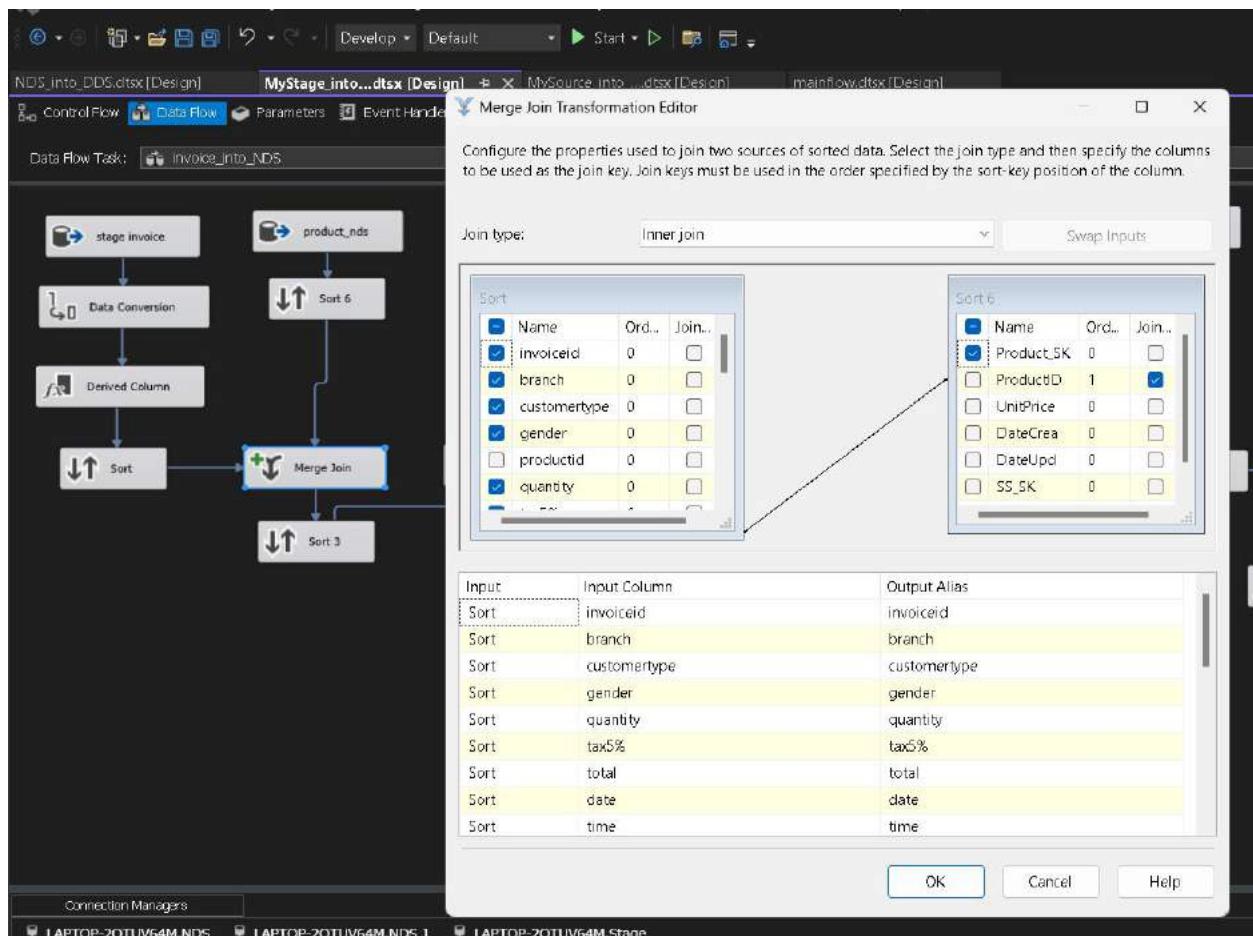
Custype_stage vào custype_nds



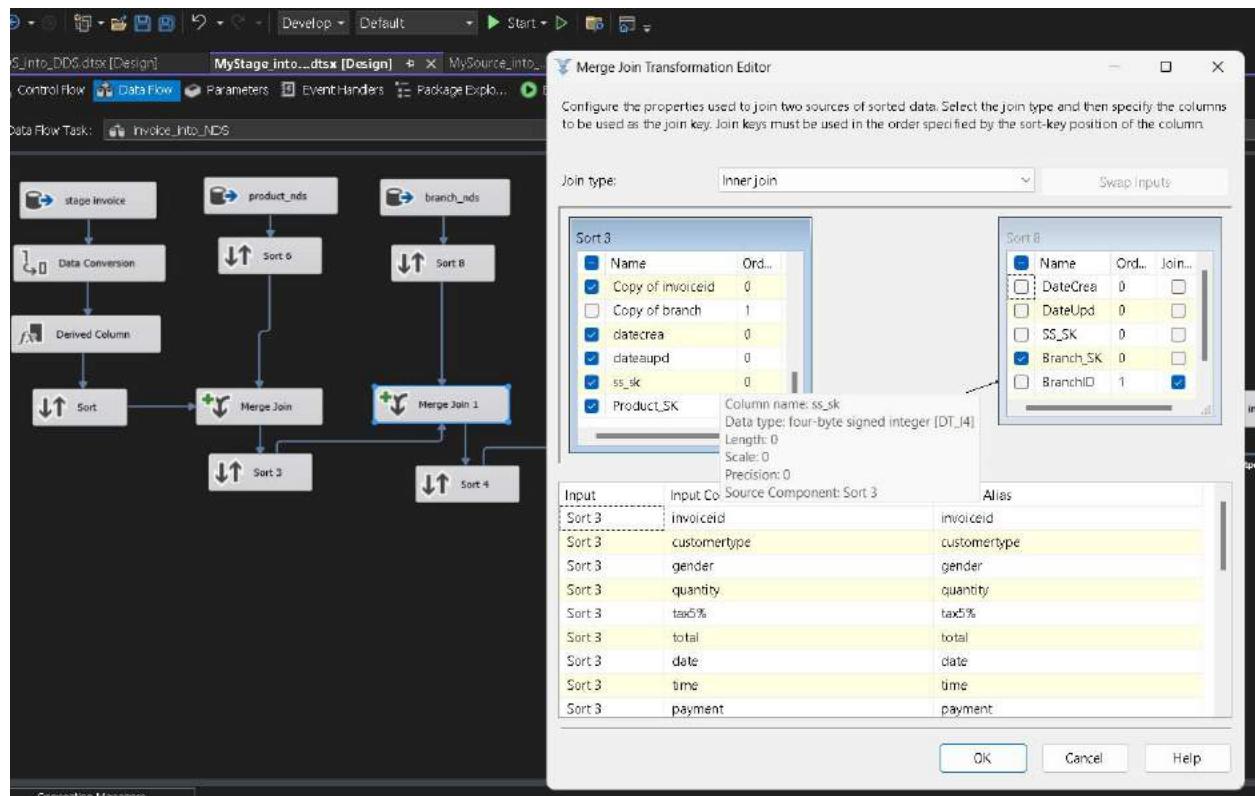
Invoice_stage vào Invoice_nds



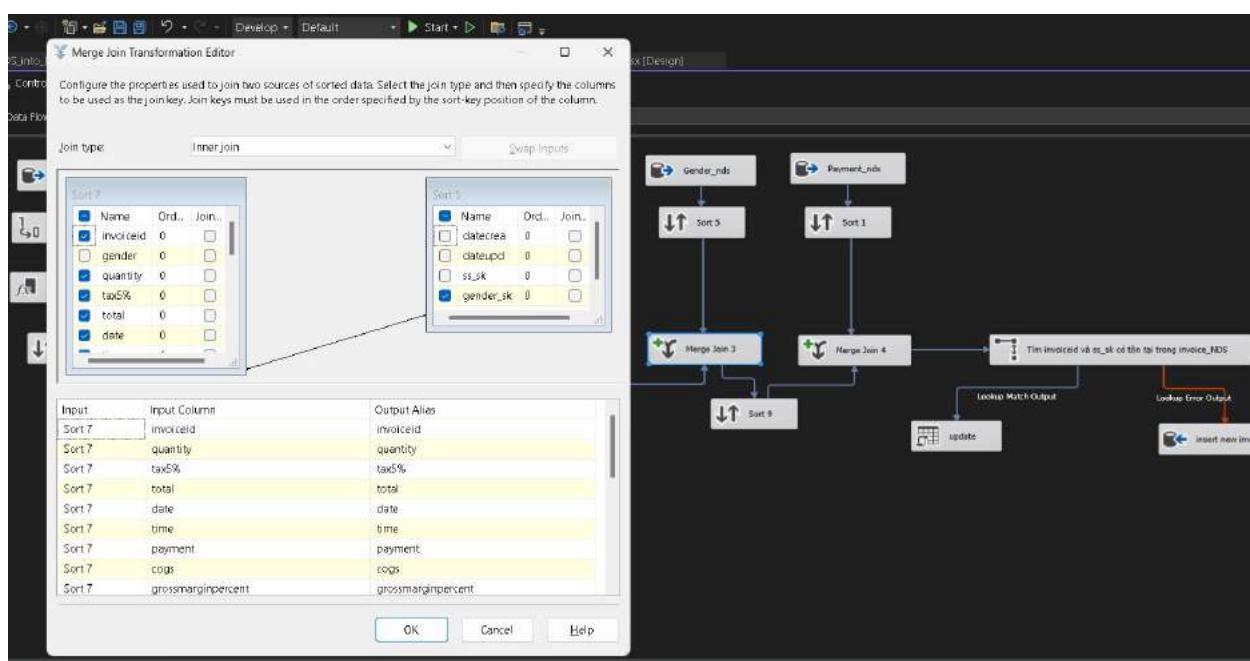
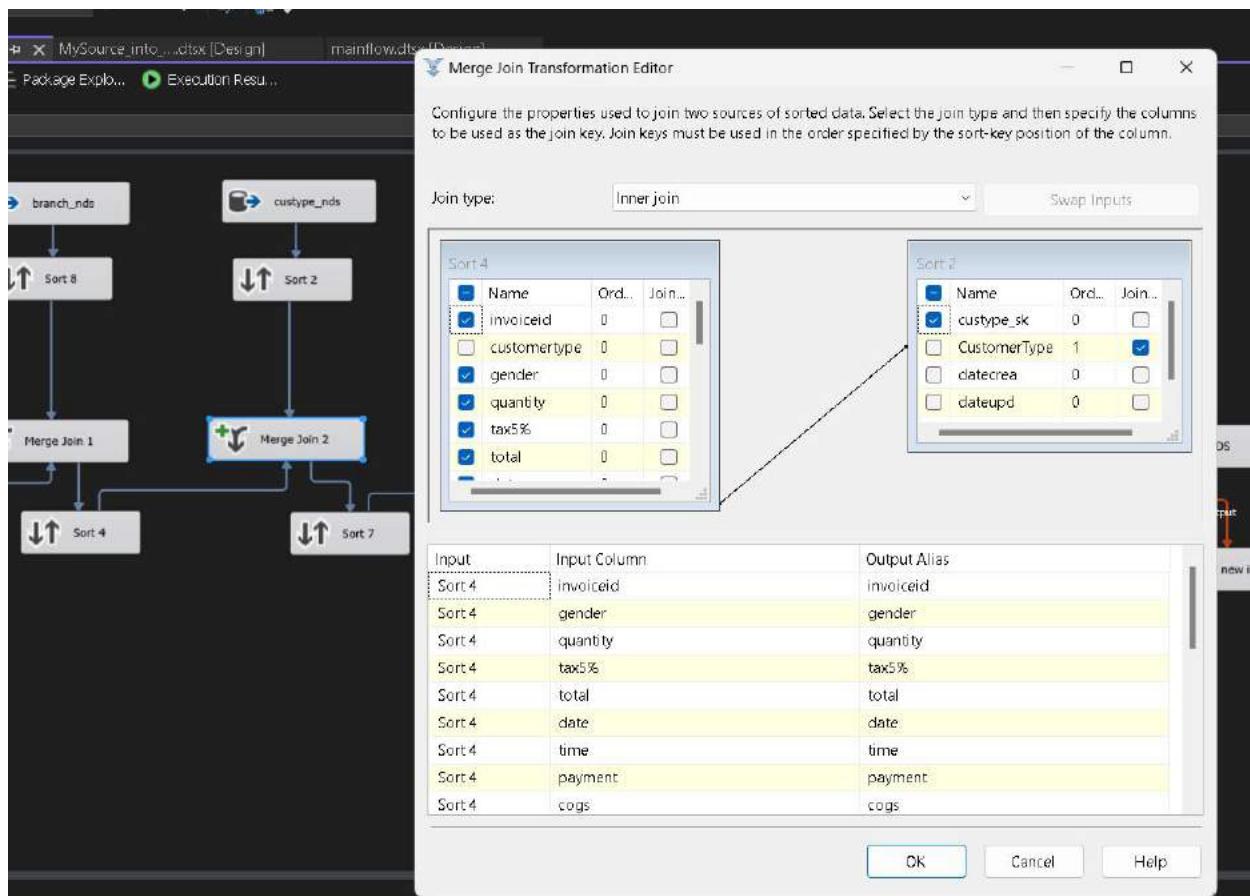
Invoice_stage merge join product_nds theo productid và lấy surrogate key của product_nds

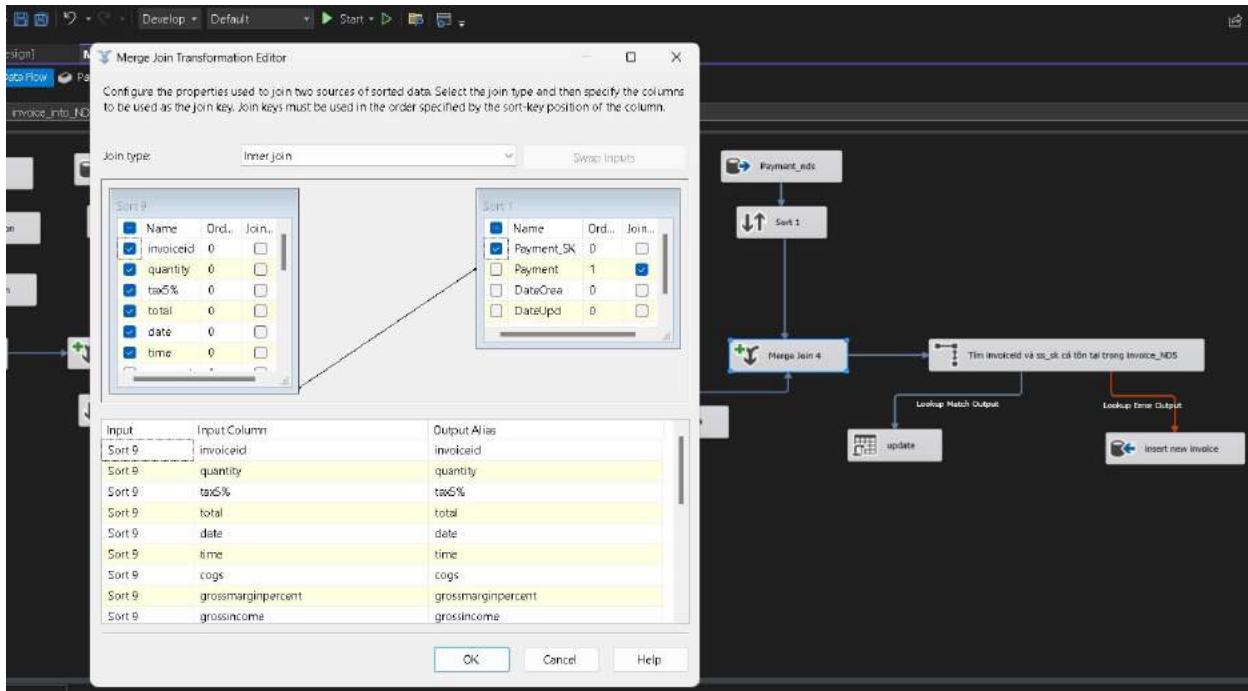


Invoice_stage merge join branch_nds theo branchid và lấy surrogate key của branch_nds



Tương tự invoice_stage merge join customertype_nds theo customertype, merger join với payment_nds theo payment, merge join với gender_nds theo gender





4. Từ NDS vào DDS

Thiết kế DDS:

1. Thống kê số thành viên mua hàng theo ngày, tháng, năm
2. Cho biết doanh thu của từng chi nhánh của các năm
3. cho biết lượng khách hàng thanh toán theo mỗi hình thức cash, debit... của mỗi chi nhánh / của tất cả chi nhánh theo từng năm...
4. Thống kê lượng rating của khách hàng (member, normal) theo từng loại sản phẩm (ProductLine)
5. Thống kê số lượng sản phẩm bán được theo từng thời điểm (time / date)
6. Thống kê số lượng khách hàng nữ đã mua theo từng loại sản phẩm

Sự kiện:

- Khi khách hàng mua sản phẩm

Bối cảnh sự kiện:

- Ai: Khách hàng
- Ở đâu: chi nhánh cửa hàng
- Cái gì: sản phẩm
- Khi nào: ngày mua hàng

Đo lường: số lượng sản phẩm , rating, doanh thu(total)

Fact table:

- Các giá trị có sẵn: quantity, rating, total
- Độ mịn: 1 dòng trong bảng fact tương ứng với 1 hóa đơn có 1 khách hàng mua 1 sản phẩm

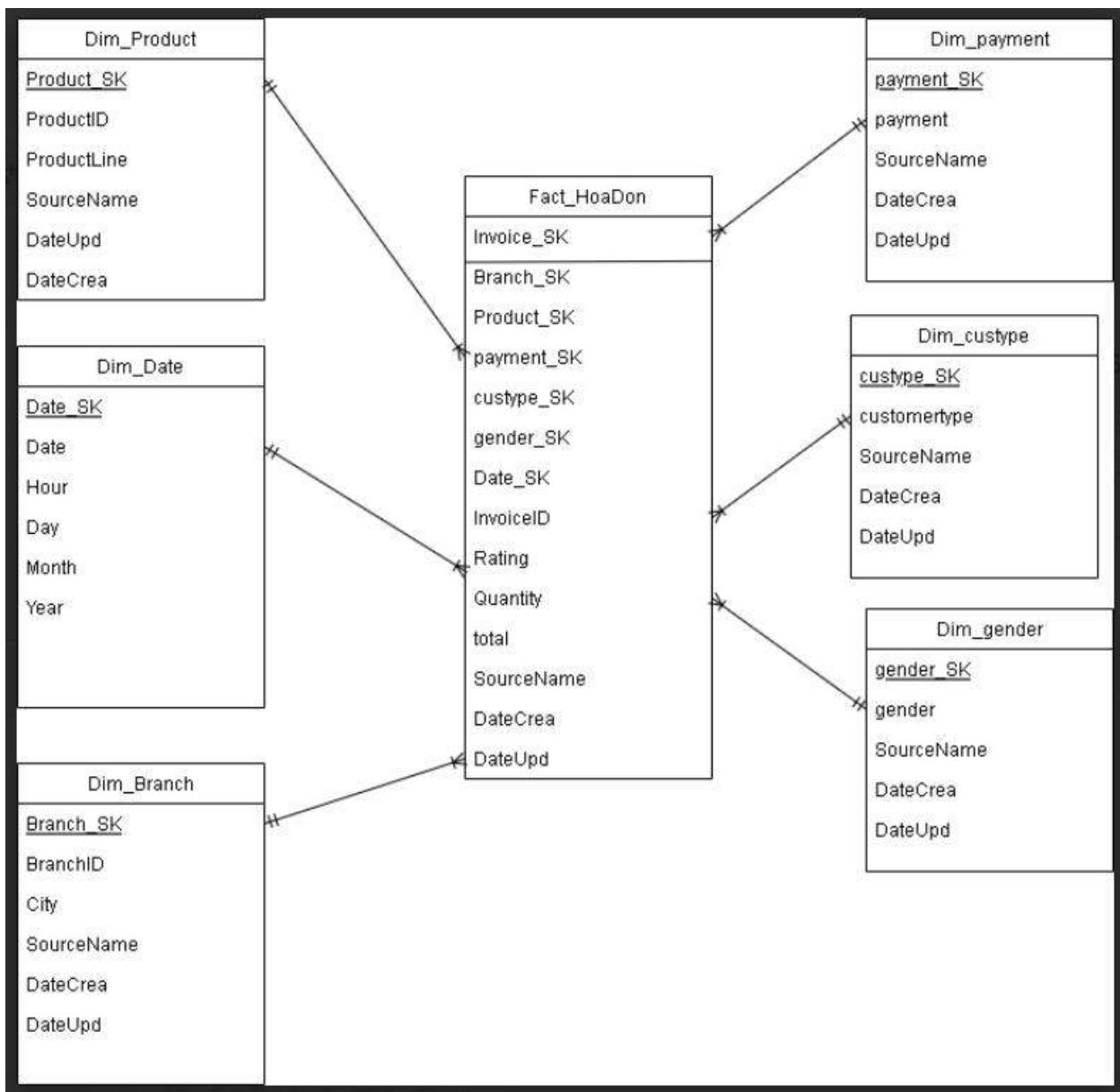
Dimention table:

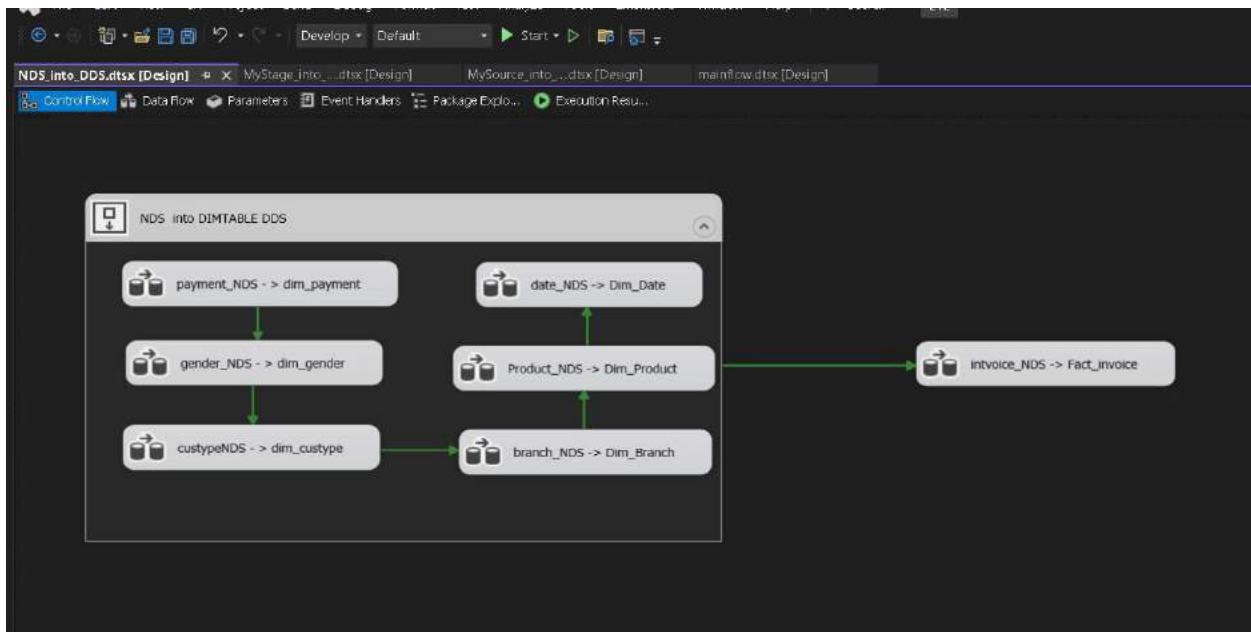
- Dim_product: không cần lưu giá trị cũ -> SCD1 : ghi chòng giá trị cũ
- Dim_payment, Dim_custype, Dim_gender: không cần lưu giá trị cũ -> SCD1: ghi chòng giá trị cũ
- Dim_branch: không cần lưu giá trị cũ -> SCD1: Ghi chòng giá trị cũ

Phân chiều:

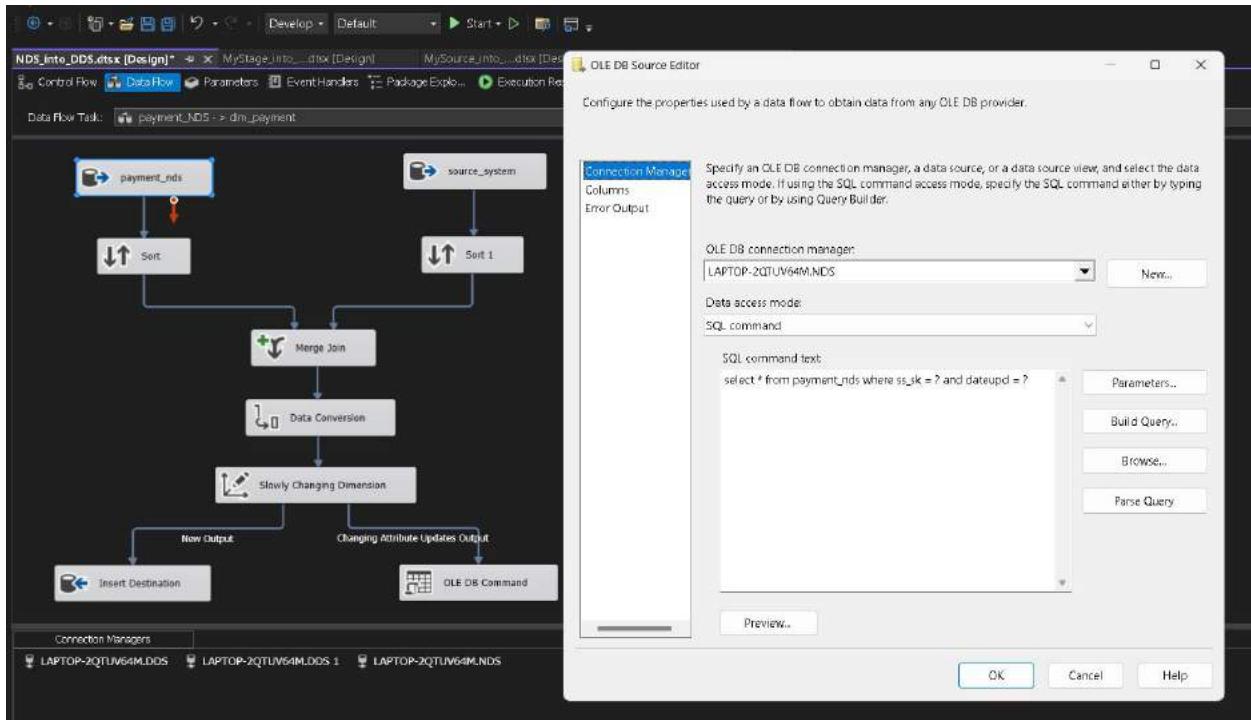
- Dim_date: năm -> tháng -> ngày -> giờ
- Dim_product: productline -> product
- Dim_branch: city -> branch

DDS: <https://drive.google.com/file/d/1qmUWVjc09mTrXnKPTGwZ1-SdCdmhXvDd/view?usp=sharing>

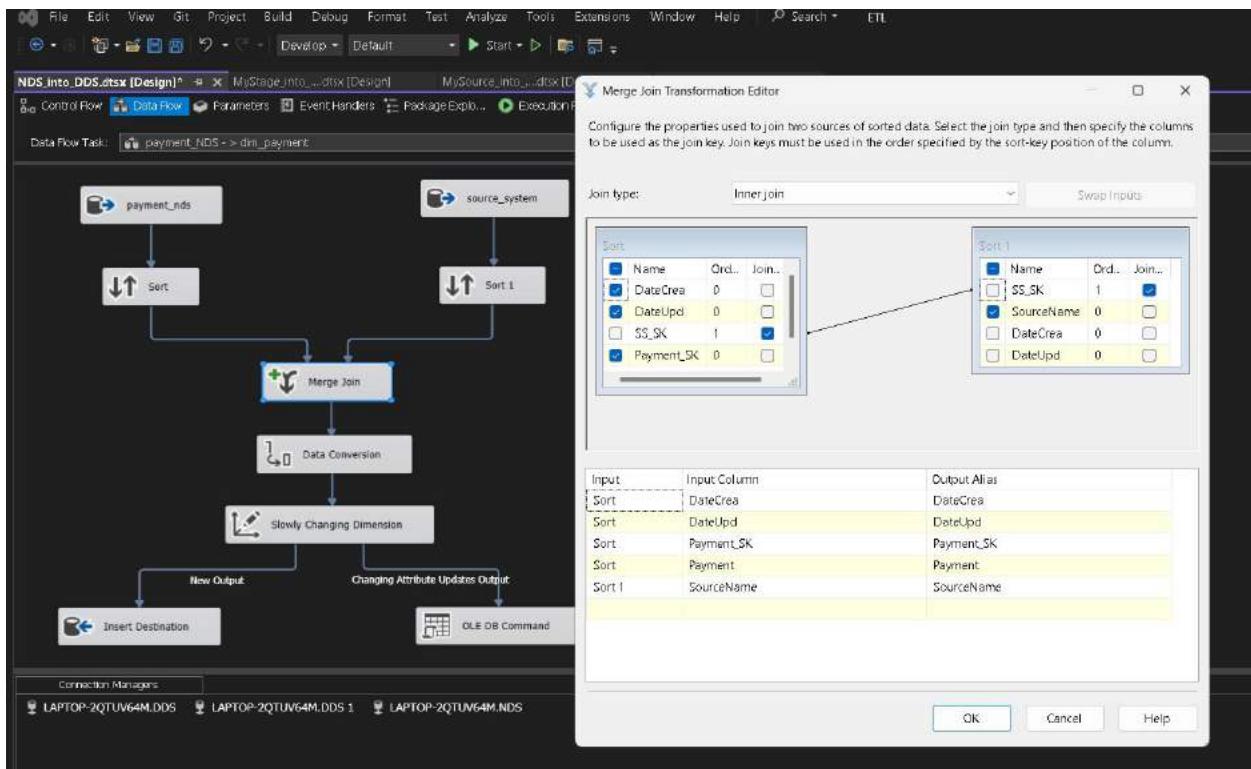




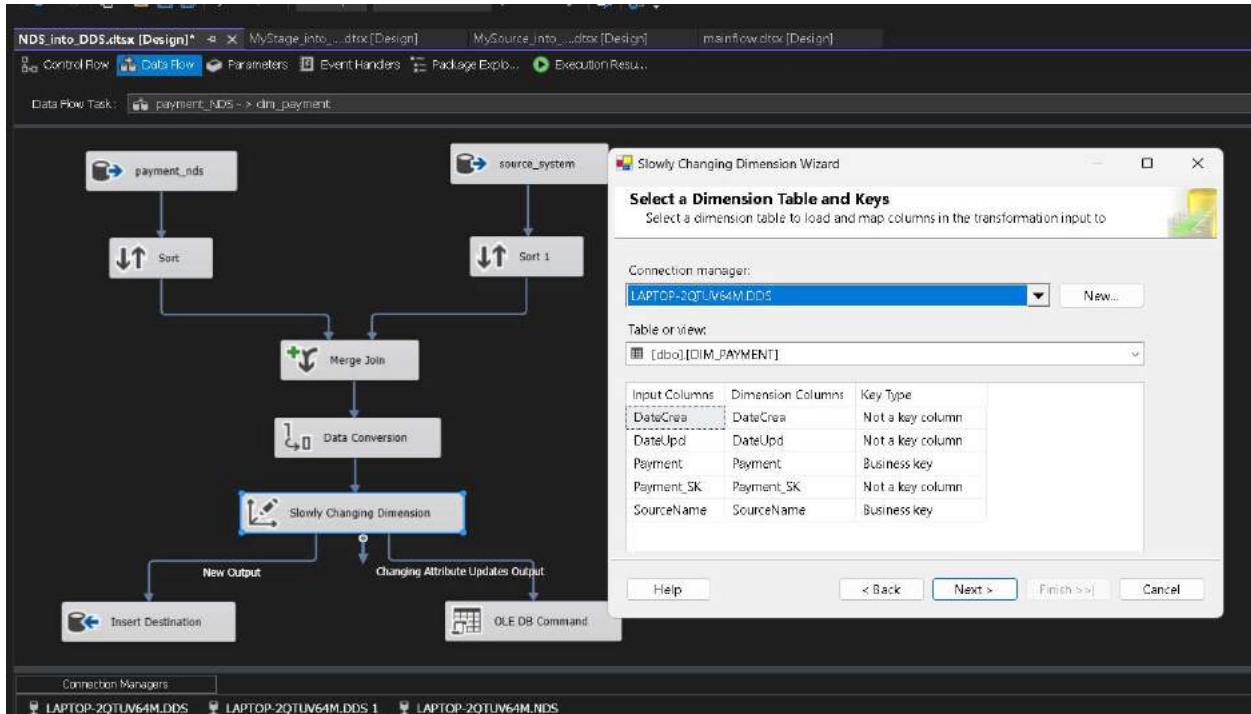
Payment_nds vào payment_dim, select payment có date update = cet để tránh load toàn bộ nds

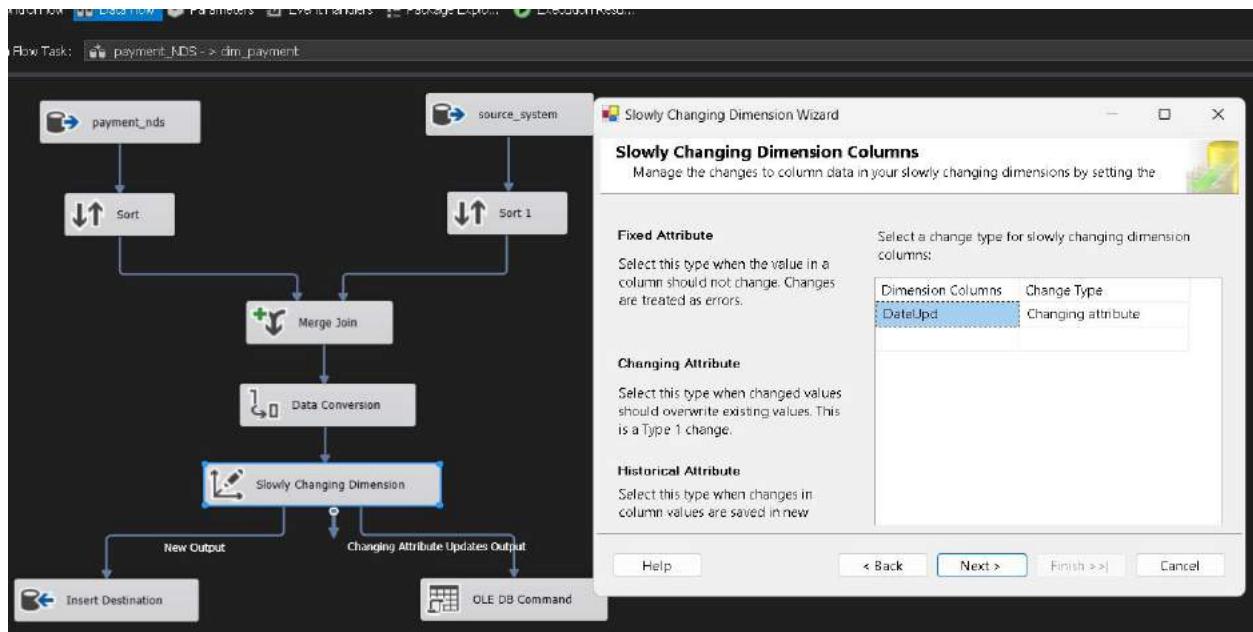


Payment_nds merge join với source_system theo ss_sk lấy sourcename

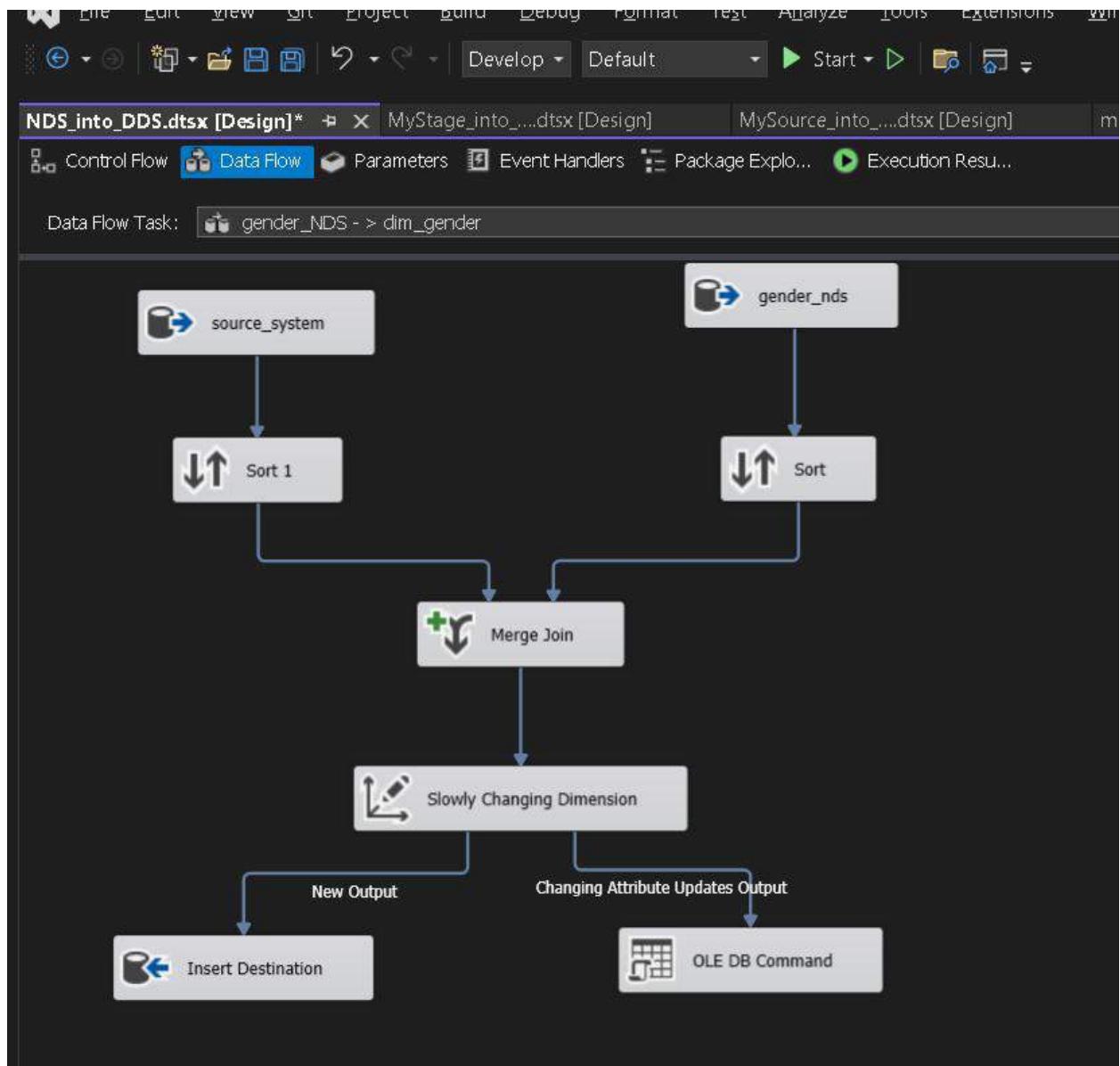


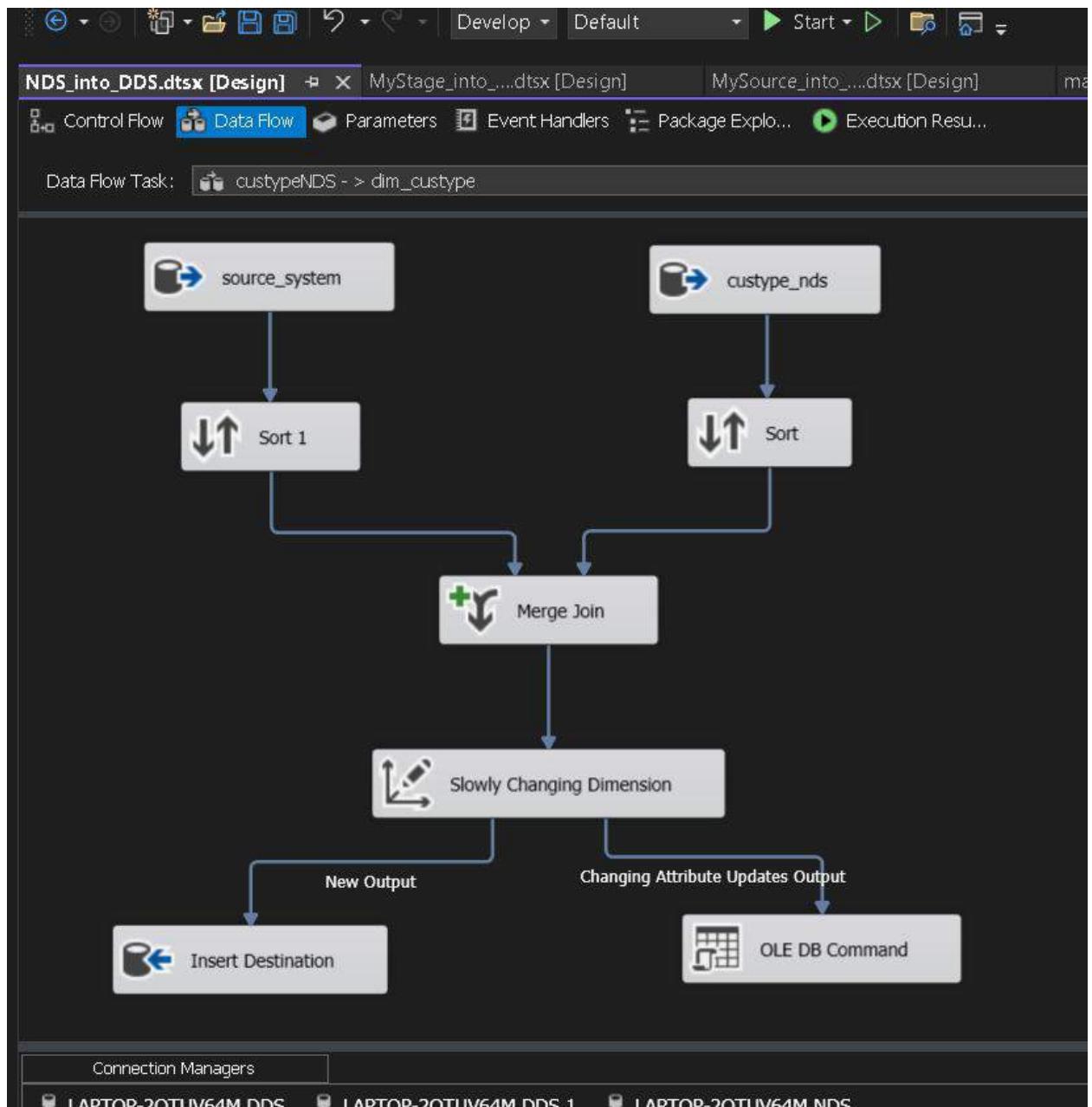
Chọn SCD set inputcolumn và business key cho payment_dds

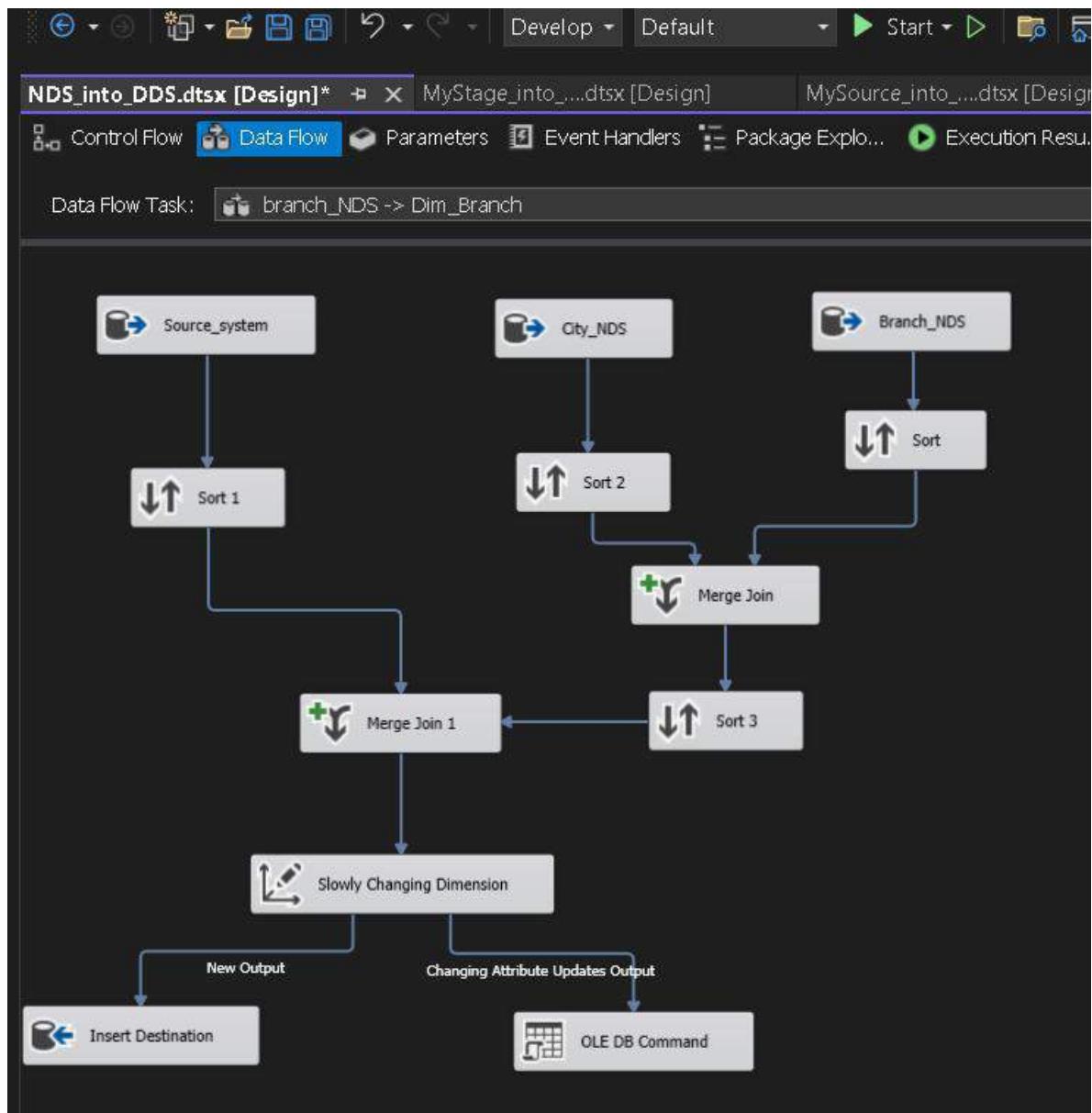


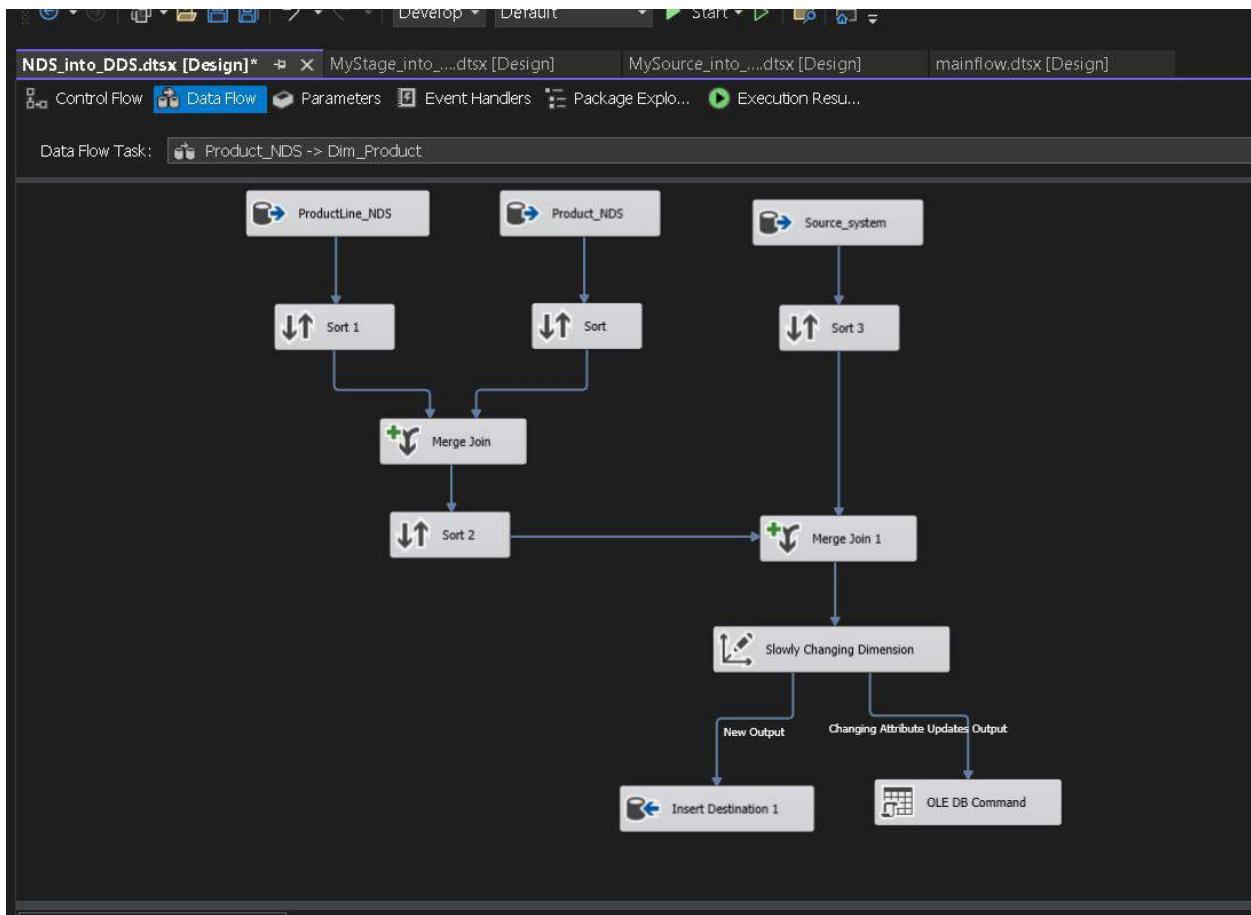


Tương tự các bảng dim_product, dim_gender, dim_custype, dim_branch làm như trên

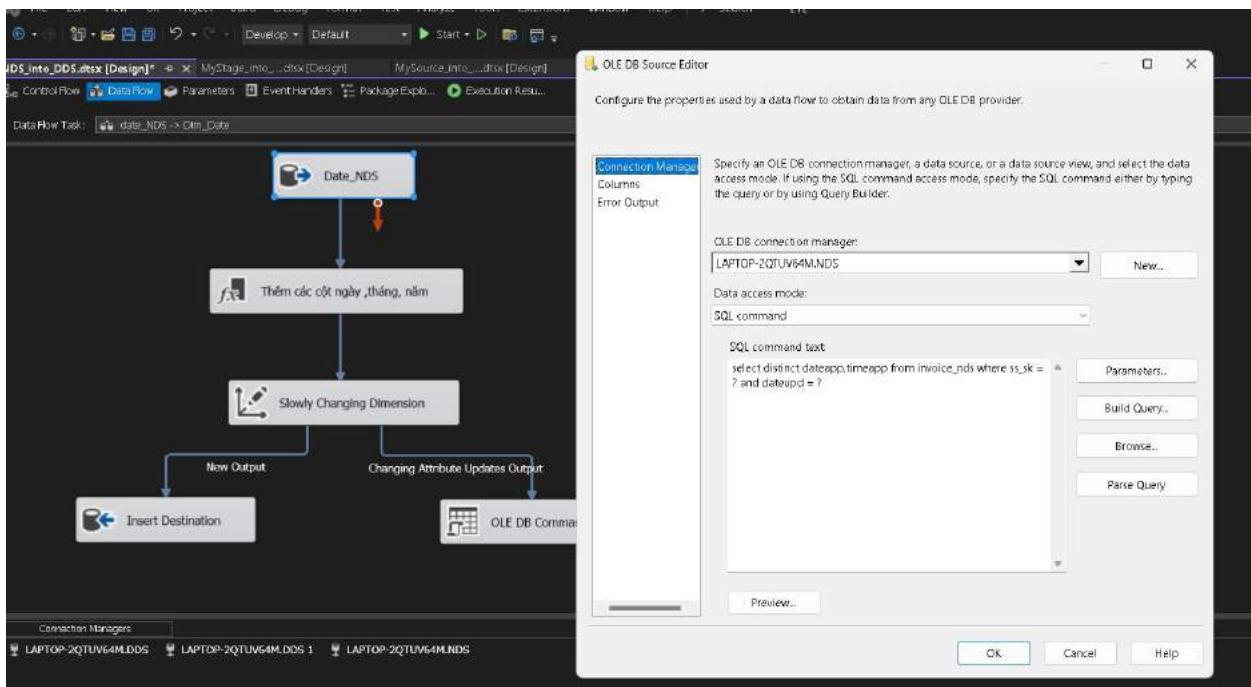


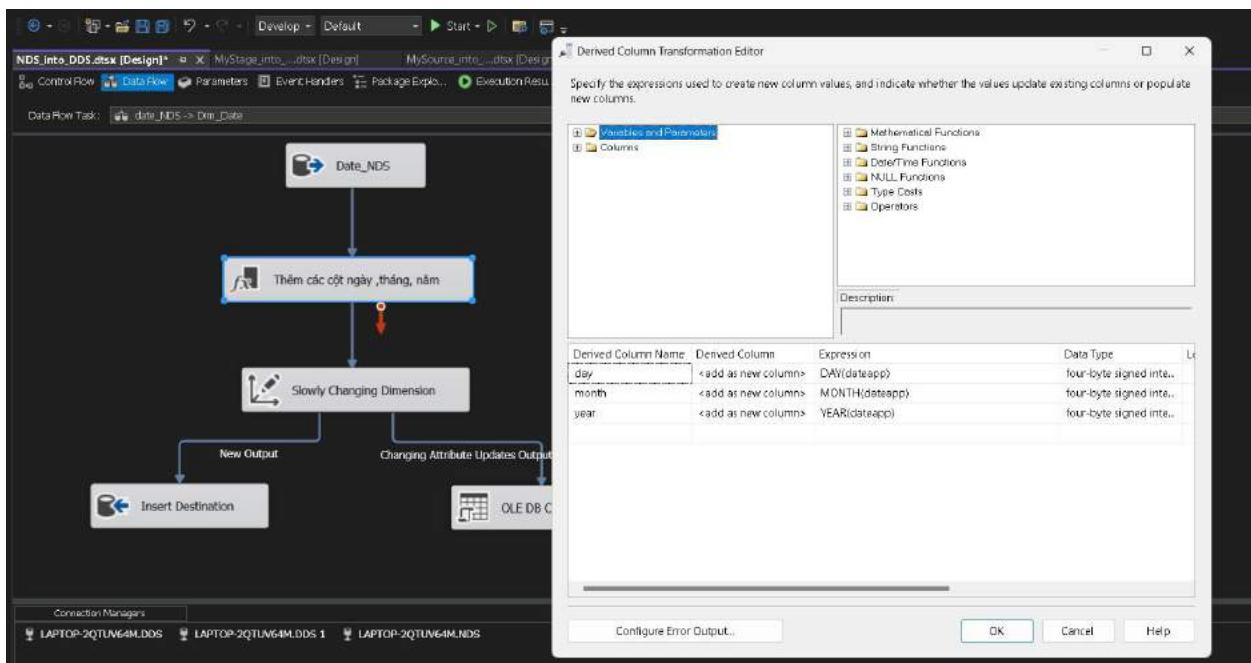




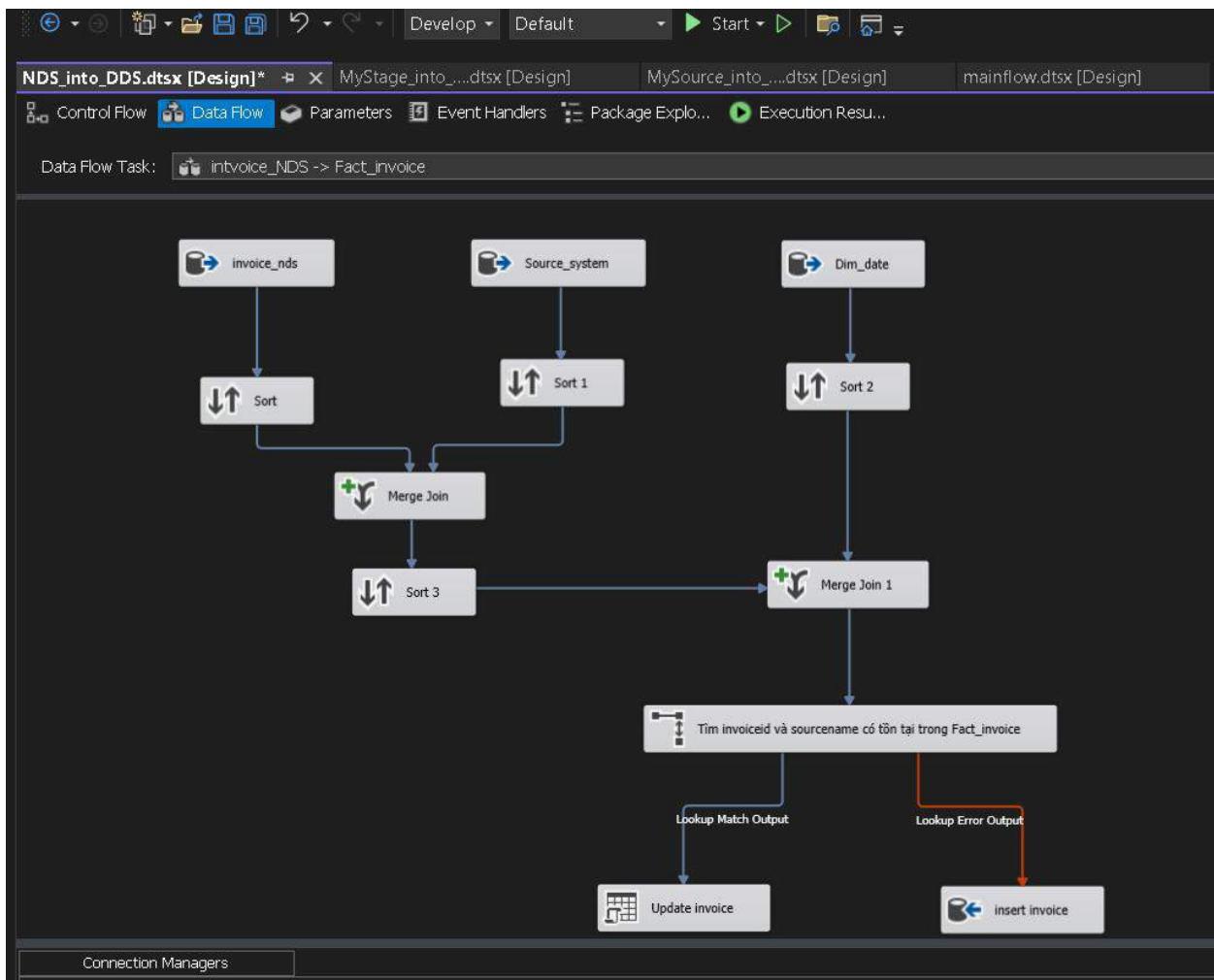


Date_nds lấy bằng cách select distinct dateapp, timeapp từ invoice_nds



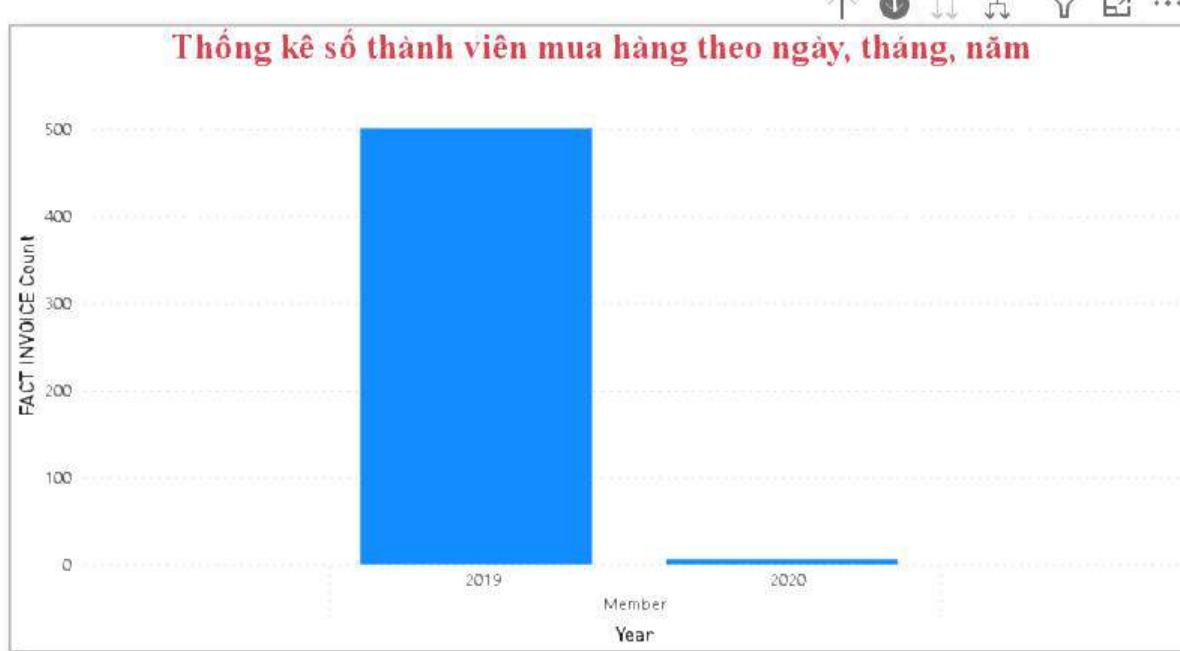


Invoice_nds vào Fact_invoice

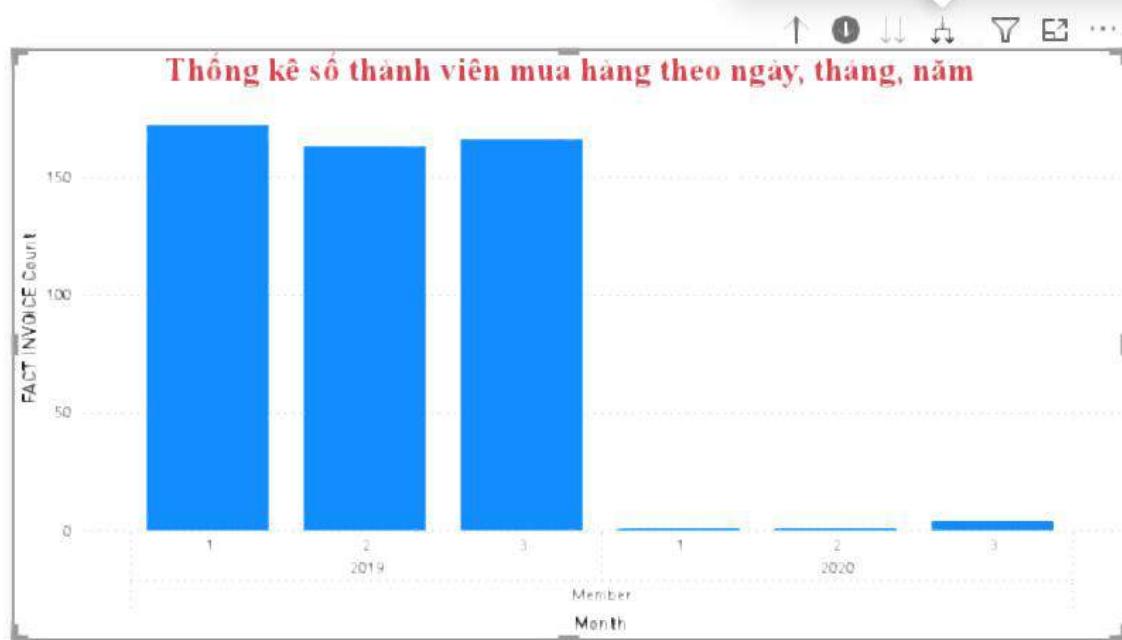


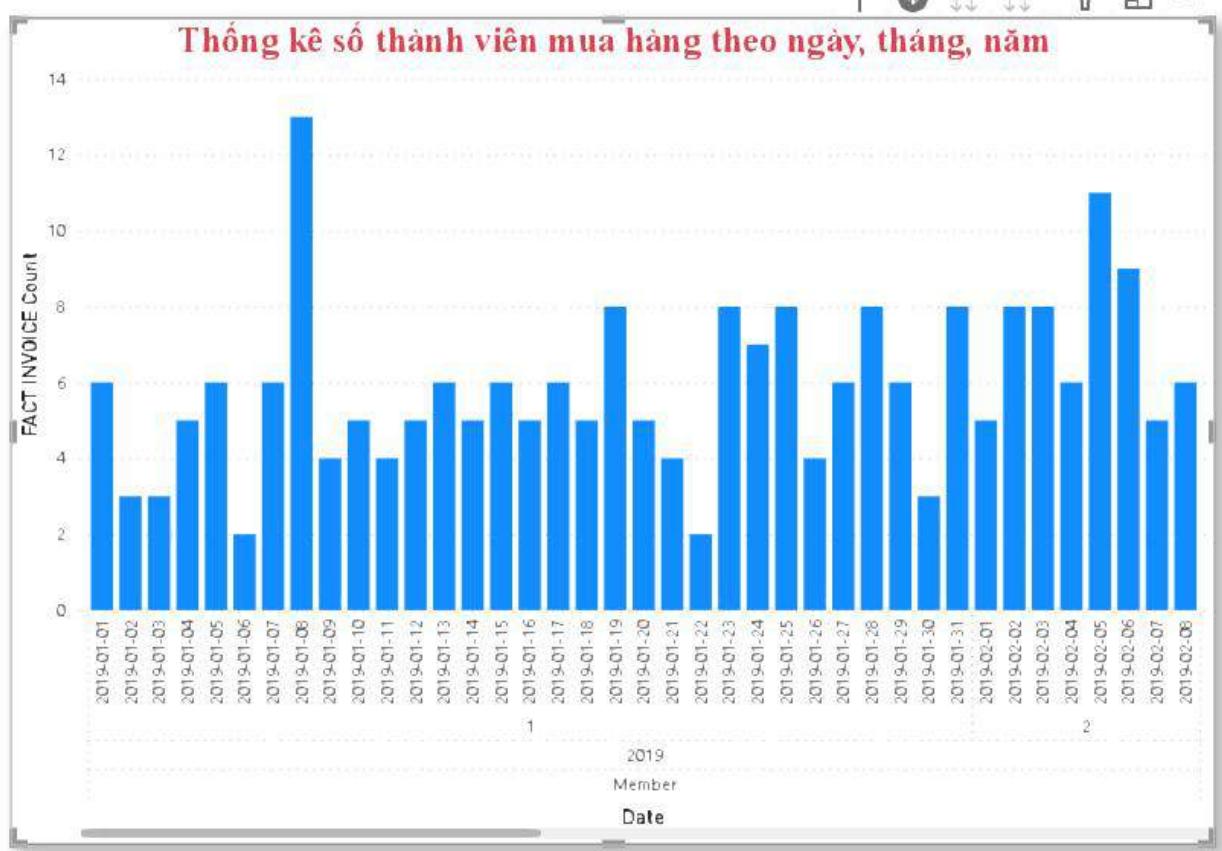
B) Report

- Thống kê số thành viên(member) mua hàng theo ngày, tháng, năm



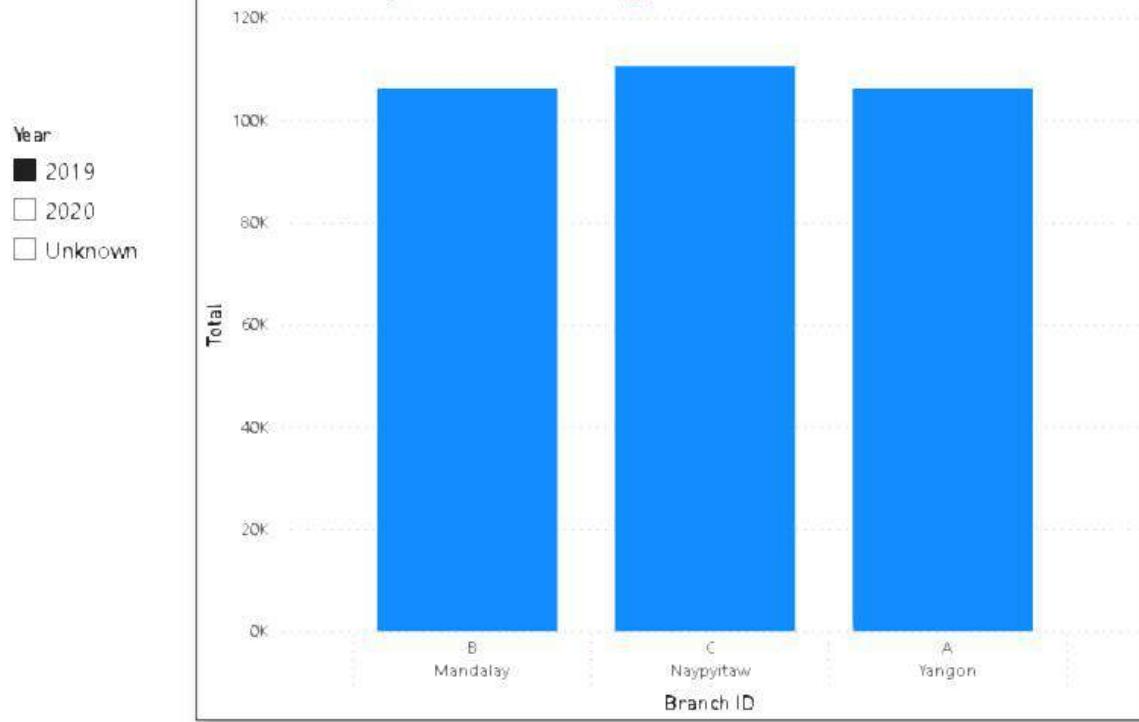
[Expand all down one level in the hierarchy](#)



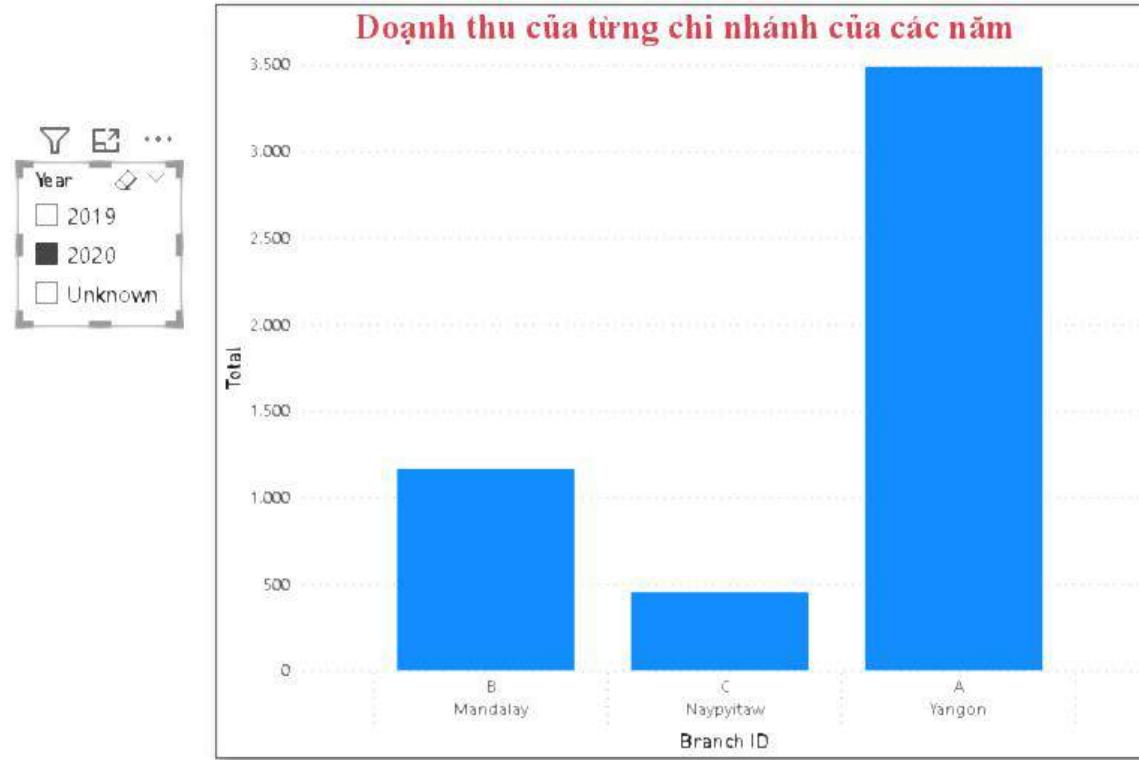


2. Doanh thu của từng chi nhánh của các năm

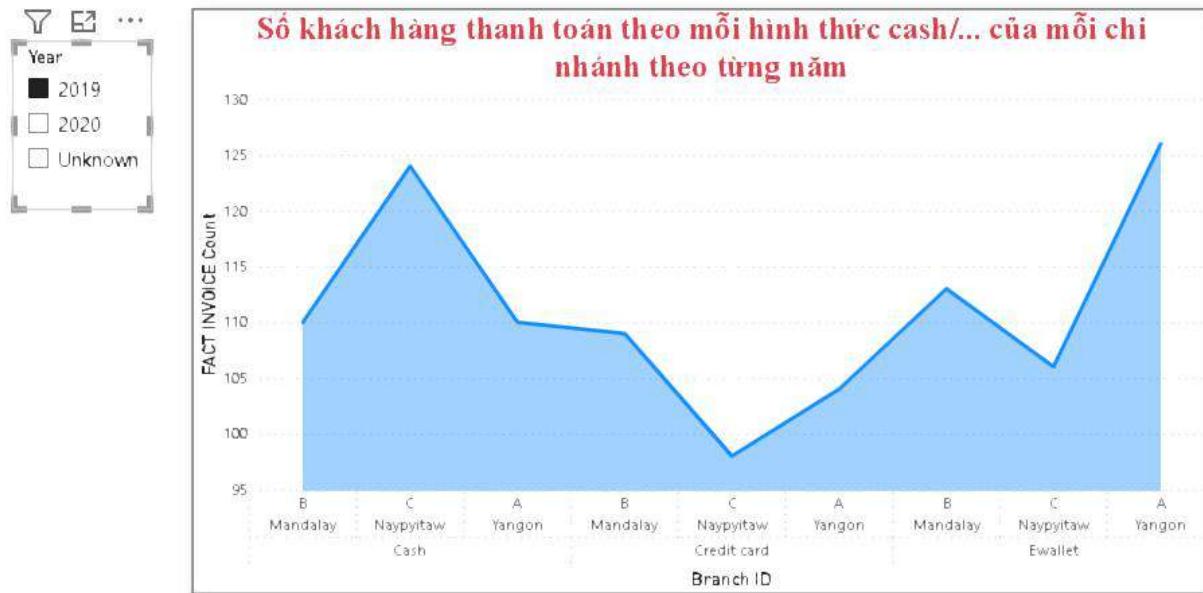
Doanh thu của từng chi nhánh của các năm



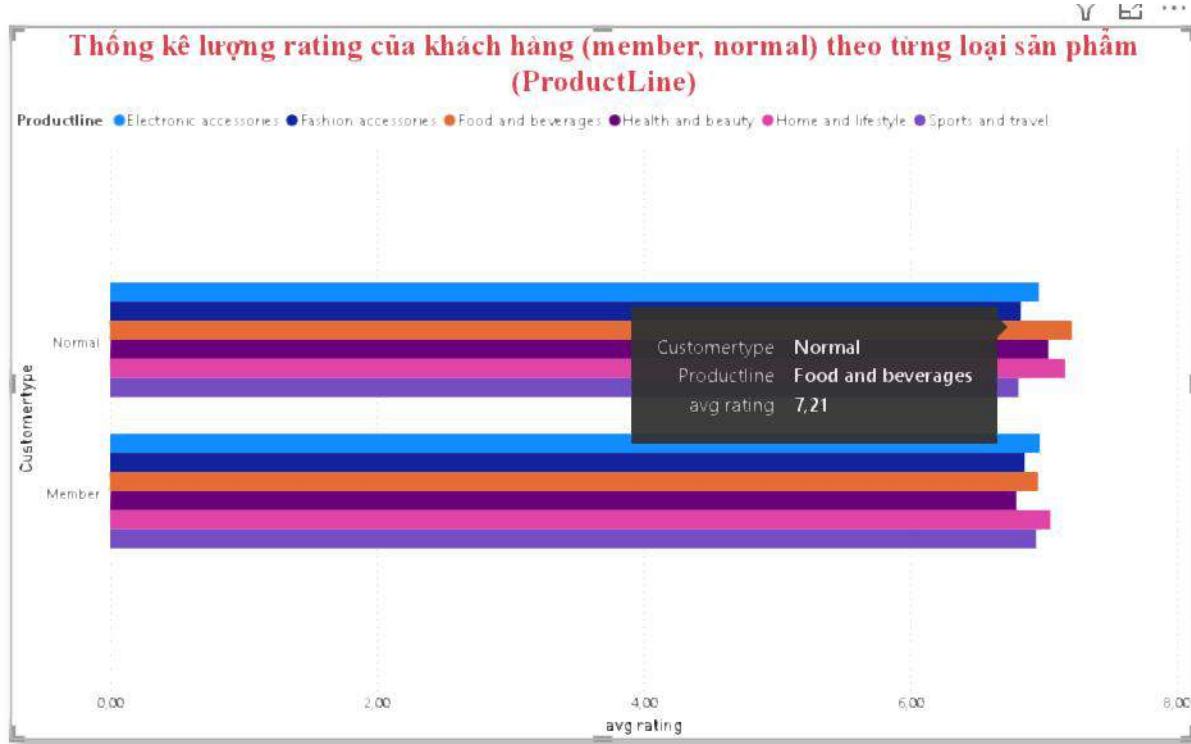
Doanh thu của từng chi nhánh của các năm



3. Số khách hàng thanh toán theo mỗi hình thức cash/debit...của mỗi chi nhánh theo từng năm



4. Lượng rating của khách hàng (member/normal) theo từng loại sản phẩm

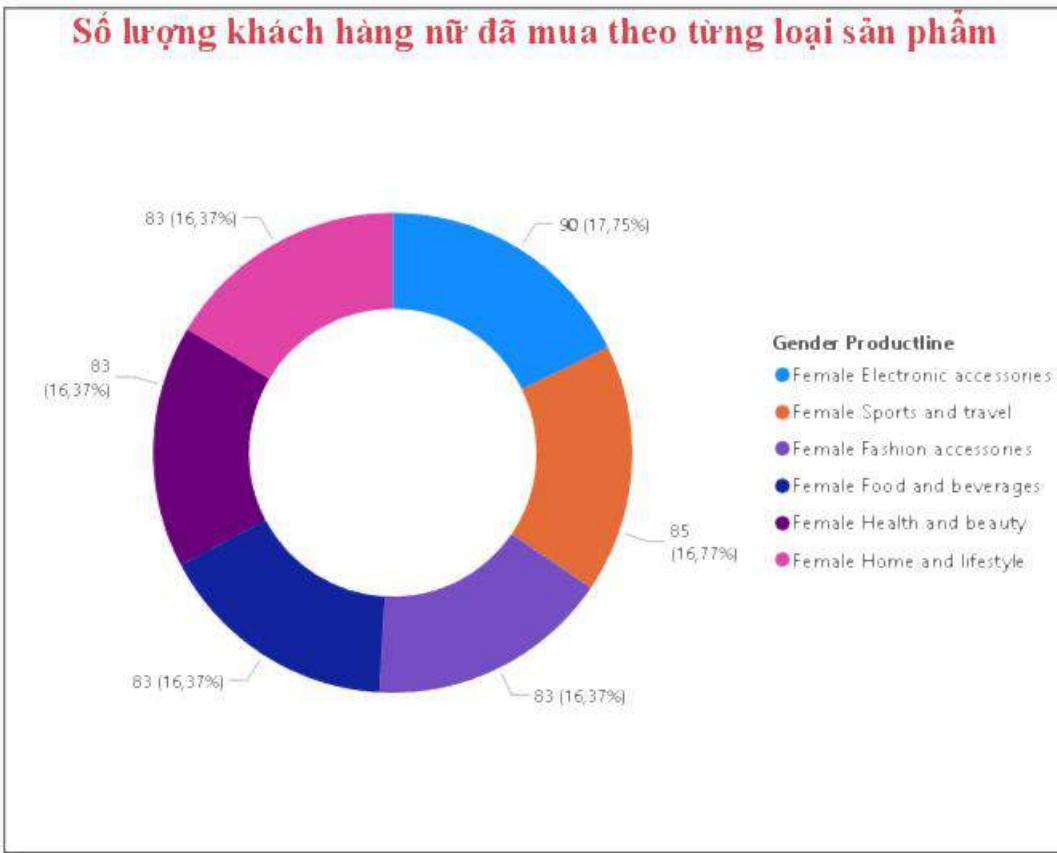


5. Số lượng sản phẩm bán được theo từng thời điểm (date/time)



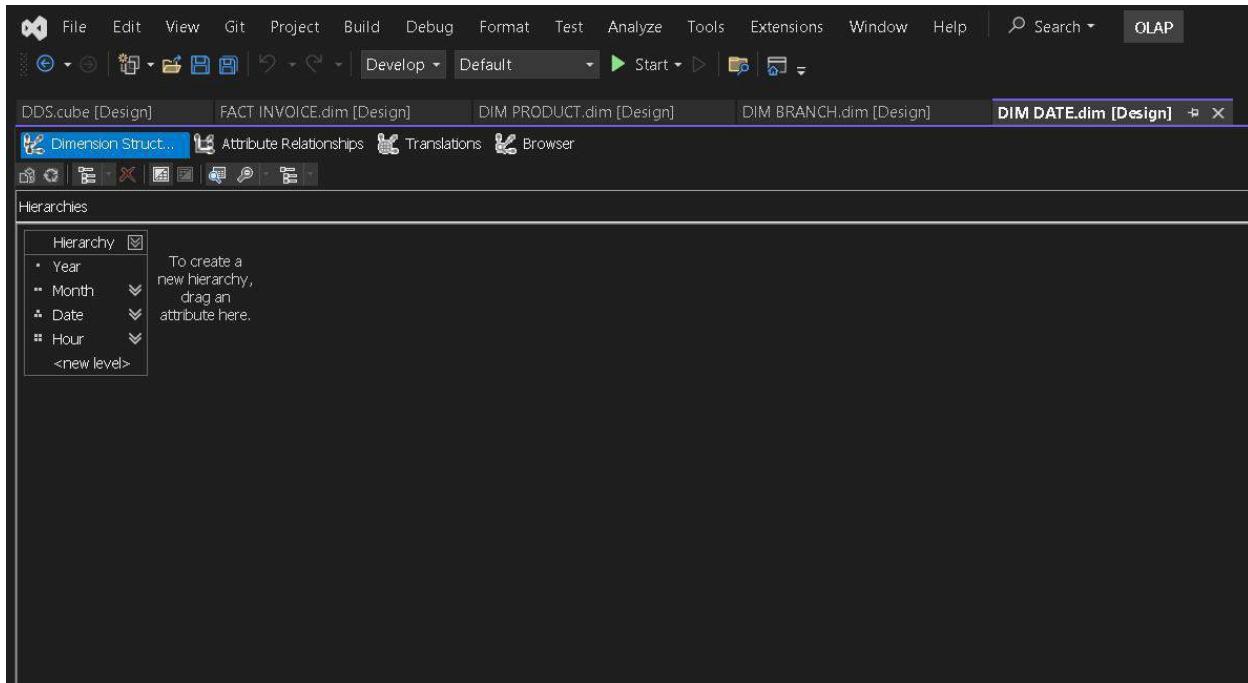
6. Số lượng khách hàng nữ đã mua theo từng loại sản phẩm

Số lượng khách hàng nữ đã mua theo từng loại sản phẩm

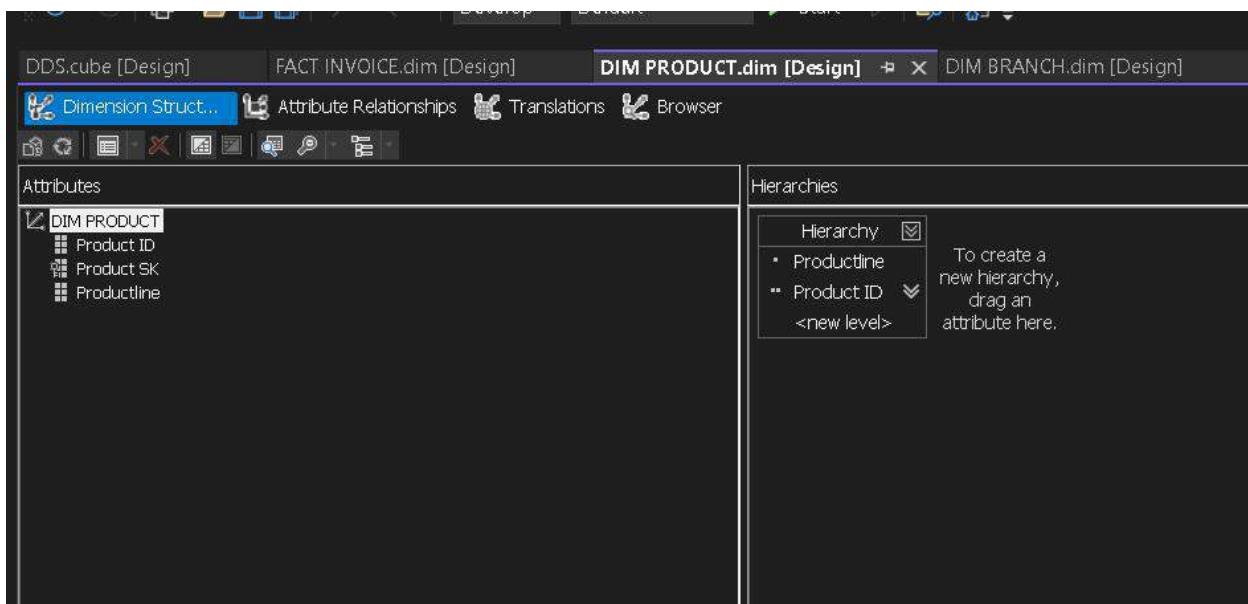


C) Olap, mdx

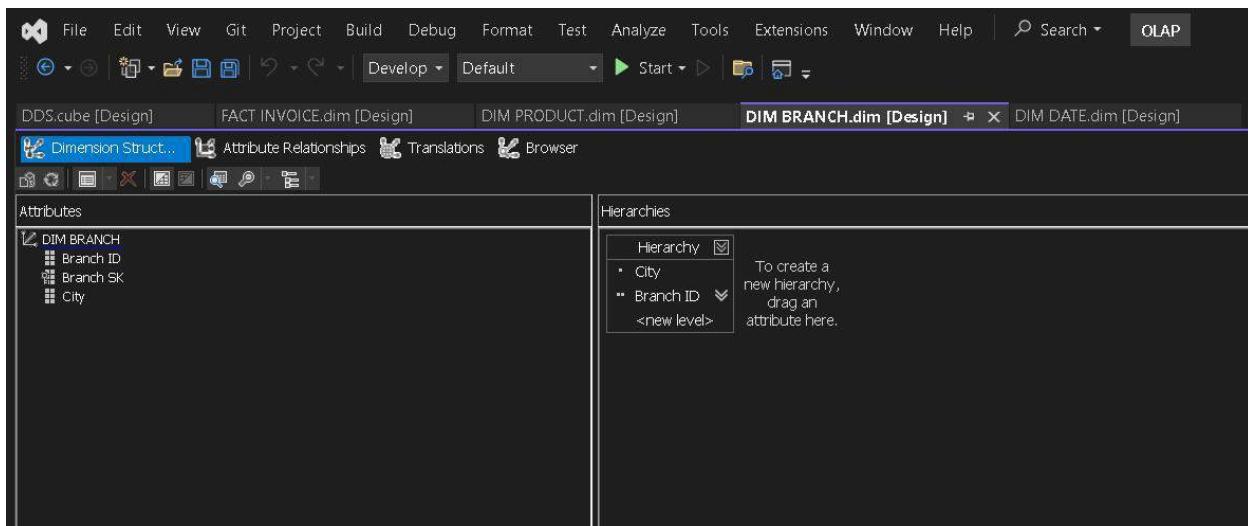
Phân chièu bảng dim_date



Phân chièu bảng Dim_product



Phân chièu bảng Dim_branch



1. Doanh thu và số lượng

- Tổng doanh thu và số lượng theo chiều phân cấp của Dim Branch và Dim Product

```
SELECT NON EMPTY {[Measures].[Total],[Quantity]} ON COLUMNS, NON EMPTY
{NONEMPTY([DIM BRANCH].[Branch Hierarchy].MEMBERS)*NONEMPTY([DIM
PRODUCT].[Product Hierarchy].MEMBERS)} ON ROWS
```

FROM [DDS];

		Total	Quantity
All	All	328070.862	5592
All	Electronic accessories	53712.4665	918
All	PID02	489.048	8
All	PID06	279.1845	3
All	PID07	441.756	8
All	PID101	182.952	9
All	PID102	764.19	10
All	PID109	45.108	2
All	PID111	510.615	10
All	PID124	536.592	8
All	PID130	462.672	9
All	PID133	92.442	4
All	PID142	118.902	4
All	PID155	145.5825	5
All	PID157	431.256	8
All	PID175	181.44	5
		325.6125	8

- Theo City nhưng slice Source Myanmar (trường hợp có nhiều source)

```
SELECT NON EMPTY {[Measures].[Total],[Quantity]} ON COLUMNS,
NON EMPTY {NONEMPTY([DIM BRANCH].[City].MEMBERS) * NONEMPTY([DIM
PRODUCT].[Product Hierarchy].MEMBERS)} ON ROWS
```

FROM [DDS]
WHERE [DIM BRANCH].[Source Name].[Myanmar];

		Total	Quantity
All	PID931	973.56	16
All	PID933	1015.959	17
All	PID936	650.58	16
All	PID943	1485.393	15
Mandalay	All	107361.933	1840
Mandalay	Electronic accessories	17614.716	297
Mandalay	PID351	189.0945	9
Mandalay	PID352	119.259	6
Mandalay	PID353	75.054	4
Mandalay	PID356	614.943	6
Mandalay	PID357	152.838	3
Mandalay	PID363	474.348	8
Mandalay	PID364	550.935	6
Mandalay	PID365	603.624	8
Mandalay	PID368	918.729	9
Mandalay	PID373	1070.853	11
-----		211.152	1

- Theo tất cả thành phố và một danh mục xác định

SELECT {[Measures].[Total],[Quantity]} ON COLUMNS,
{NONEEMPTY([DIM BRANCH].[City].MEMBERS) * [DIM
PRODUCT].[Productline].[Productline].[Electronic accessories]} ON ROWS
FROM [DDS];

		Total	Quantity
All	Electronic accessories	53712.4665	918
Mandalay	Electronic accessories	17614.716	297
Naypyitaw	Electronic accessories	19926.5325	330
Yangon	Electronic accessories	16171.218	291

2. Số lượng đánh giá và avg đánh giá theo giới tính và loại khách hàng, loại hàng

- Rating count & avg rating based on member

WITH

MEMBER [Measures].[Rating Count] AS [Measures].[FACT INVOICE Count]
SELECT NON EMPTY {[Measures].[Rating Count], [Measures].[avg rating]} ON COLUMNS,
NONEEMPTY({[DIM CUSTYPE].[Customertype].MEMBERS}) ON ROWS
FROM [DDS]
WHERE ([DIM BRANCH].[Source Name].[Myanmar]);

	Rating Count	avg rating
All	1016	6.96
Member	507	6.93
Normal	509	7.00

- Rating count & avg rating in each month by gender in specific year (slice)

WITH

```
MEMBER [Measures].[Rating Count] AS [Measures].[FACT INVOICE Count]
SELECT {[Measures].[Rating Count], [Measures].[Avg Rating]} ON COLUMNS,
{NONEMPTY([DIM GENDER].[Gender].MEMBERS) * NONEMPTY([DIM DATE].[Month].[Month].MEMBERS)} ON ROWS
FROM [DDS]
WHERE [DIM DATE].[Year].[2019];
```

	Rating Count	avg rating
All	1	352
All	2	303
All	3	345
Female	1	176
Female	2	164
Female	3	161
Male	1	176
Male	2	139

3. Average Rating by Customer Type for Health and Beauty Products in Myanmar Using slice

```
SELECT {[Measures].[avg rating]} ON COLUMNS,
{[DIM CUSTYPE].[Customertype].MEMBERS} ON ROWS
FROM [DDS]
WHERE ([DIM BRANCH].[City].Members, [DIM PRODUCT].[Productline].[Health and beauty]);
```

	avg rating
All	6.90
Member	6.80
Normal	7.04
Unknown	(null)

4. Top 5 product_id sales slice by year

```

SELECT NON EMPTY({[Measures].[Total]})  

ON COLUMNS,  

NON EMPTY(Generate([DIM PRODUCT].[Productline].[Productline].Members,  

    TopCount({[DIM PRODUCT].[Productline].CurrentMember} * [Dim Product].[Product  

ID].[Product ID].Members,  

    5,  

    [Measures].[Total]  

))  

ON ROWS  

FROM [DDS]  

WHERE [DIM DATE].[Year].[2019];

```

Messages Results

		Total
Electronic accessories	PID905	1028.685
Electronic accessories	PID876	1003.59
Electronic accessories	PID696	950.25
Electronic accessories	PID832	943.299
Electronic accessories	PID368	918.729
Fashion accessories	PID925	1531.194
Fashion accessories	PID938	1010.877
Fashion accessories	PID72	951.825
Fashion accessories	PID18	931.035
Fashion accessories	PID885	881.307
Food and beverages	PID896	1795.668
Food and beverages	PID763	1034.46
Food and beverages	PID631	1022.385
Food and beverages	PID362	944.622
Food and beverages	PID850	939.54
Health and beauty	PID915	1454.565
Health and beauty	PID731	1042.65
Health and beauty	PID695	942.9
Health and beauty	PID533	942.4485
Health and beauty	PID516	922.635
Home and lifestyle	PID940	1082.067
Home and lifestyle	PID928	1025.7135
Home and lifestyle	PID740	1020.705
Home and lifestyle	PID900	966.9555
Home and lifestyle	PID941	940.7685
Sports and travel	PID943	1485.393
Sports and travel	PID199	1039.29
Sports and travel	PID672	1022.49
Sports and travel	PID933	1015.959
Sports and travel	PID931	973.56

5. Số thành viên mua hàng theo ngày tháng năm

```
FROM [DDS]
WHERE [DIM DATE].[Year].[2019];

// số thành viên mua hàng theo ngày tháng năm
WITH
MEMBER [Measures].[Customer Count] AS [Measures].[FACT INVOICE Count]

SELECT NON EMPTY {[DIM DATE].[Year].Members* [Measures].[Customer Count]} ON 0,
NON EMPTY{
    {[DIM DATE].[Month].[Month].Members *
    [DIM DATE].[Date].[Date].Members}
} ON 1
FROM [DDS];

// Thống kê doanh thu của khách hàng (member, normal) theo ngày tháng năm và theo chi nhánh
SELECT NON EMPTY {[DIM BRANCH].[City].Members * [DIM CUSTTYPE].[Customertype].[Customert
{
```

100 %

	All	2019	2020
	Customer Count	Customer Count	Customer Count
1	2020-01-01	1	(null)
1	2020-01-15	1	(null)
1	2020-01-21	1	(null)
1	2019-01-01	12	12
1	2019-01-02	8	8
1	2019-01-03	8	8
1	2019-01-04	6	6
1	2019-01-05	12	12
1	2019-01-06	9	9
1	2019-01-07	9	9
1	2019-01-08	18	18
1	2019-01-09	8	8
1	2019-01-10	9	9
1	2019-01-11	8	8
1	2019-01-12	11	11

6. Thống kê doanh thu của khách hàng (member, normal) theo ngày tháng năm và theo chi nhánh

```

// Thống kê doanh thu của khách hàng (member, normal) theo ngày tháng năm và theo chi nhánh
SELECT NON EMPTY {[DIM BRANCH].[City].Members * [DIM CUSTYPE].[Customertype].[Customertype]* [Measures].[Total]} ON 0
{
    NONEMPTY(
        [DIM DATE].[Year].[Year] *
        [DIM DATE].[Month].[Month] *
        [DIM DATE].[Date].[Date]
    )
} ON 1
FROM [DDS];

// Thống kê số lượng khách hàng thanh toán theo cash, debit,.. ở từng chi nhánh theo từng tháng trong năm
WITH
    MEMBER [Measures].[Customer Count] AS [Measures].[FACT INVOICE Count]
SELECT NON EMPTY {[DIM BRANCH].[City].Members * [DIM PAYMENT].[Payment].[Payment] * [Measures].[Customer Count]} ON 0,
    NON EMPTY {
        [DIM DATE].[Month].Members
    } ON 1
FROM [DDS];

```

Messages Results

	All	All	Mandalay	Mandalay	Naypyitaw	Naypyitaw	Yangon	Yangon		
	Member	Normal	Member	Normal	Member	Normal	Member	Normal		
	Total	Total	Total	Total	Total	Total	Total	Total		
2019	1	2019-01-01	2753.825	1991.556	888.615	648.081	572.376	264.789	1292.634	1078.886
2019	1	2019-01-02	912.2505	1033.2525	383.7645	779.037	266.028	209.622	282.458	44.5935
2019	1	2019-01-03	644.826	1433.3025	520.8	495.894	124.026	(null)	(null)	937.4085
2019	1	2019-01-04	1477.4655	146.223	364.3605	146.223	629.8425	(null)	483.2625	(null)
2019	1	2019-01-05	1554.7035	1981.98	411.999	74.76	33.495	991.9245	1109.2095	915.2955
2019	1	2019-01-06	1255.2015	2359.0035	760.4415	261.198	(null)	1282.596	494.76	815.2095
2019	1	2019-01-07	1819.944	1014.3	989.331	352.5795	300.636	85.743	529.977	575.9775
2019	1	2019-01-08	4383.918	909.8145	2208.948	(null)	1897.4655	503.8635	277.5045	405.951
2019	1	2019-01-09	834.288	2187.0555	168.525	(null)	463.428	2187.0555	202.335	(null)
2019	1	2019-01-10	1601.271	1959.678	(null)	1348.599	869.778	611.079	731.493	(null)
2019	1	2019-01-11	1104.6315	1010.331	(null)	(null)	453.2955	(null)	651.336	1010.331
2019	1	2019-01-12	2337.489	2847.2745	797.5275	1514.6145	419.832	1023.75	1120.1295	308.91
2019	1	2019-01-13	1379.049	1072.155	583.086	152.019	166.3725	437.325	629.5905	482.811
2019	1	2019-01-14	1786.2495	2180.367	75.936	1115.5515	614.2185	1084.8155	1096.095	(null)
			2000.000	2000.000	1000.000	1500.000	2000.000	1000.000	2000.000	2000.000

7. Thống kê số lượng khách hàng thanh toán theo cash, debit,.. ở từng chi nhánh theo từng tháng trong năm

```

// Thống kê số lượng khách hàng thanh toán theo cash, debit,.. ở từng chi nhánh theo từng tháng trong năm
WITH
    MEMBER [Measures].[Customer Count] AS [Measures].[FACT INVOICE Count]
SELECT NON EMPTY {[DIM BRANCH].[City].Members * [DIM PAYMENT].[Payment].[Payment] * [Measures].[Customer Count]} ON 0,
    NON EMPTY {
        [DIM DATE].[Month].Members
    } ON 1
FROM [DDS];

```

Messages Results

	All	All	All	Mandalay	Mandalay	Mandalay	Naypyitaw	Naypyitaw	Naypyitaw
	Cash	Credit card	Ewallet	Cash	Credit card	Ewallet	Cash	Credit card	Ewallet
	Customer Count								
All	348	316	352	113	110	115	124	98	107
1	1	2	(null)	1	(null)	(null)	(null)	(null)	(null)
1	122	113	117	35	44	32	48	35	39
2	112	90	101	44	30	35	36	32	32
2	2	(null)	4	1	(null)	(null)	(null)	(null)	1
3	1	3	3	1	1	2	(null)	(null)	(null)
3	110	108	127	31	35	46	40	31	35

D) Dashboard



E) Mining

Nhóm dùng Market Basket Analysis để mining các insight của dữ liệu

Đầu tiên dùng one hot encoding để chuyển dữ liệu thành dạng binary

Feature Frequency Lookup Table		Frequency	Percentage
Feature Index	Feature		
1	Credit card	316	0.311023622
2	Ewallet	352	0.346456693
3	Cash	348	0.342519685
4	Health and beauty	153	0.150590551
5	Electronic accessories	168	0.165354331
6	Home and lifestyle	170	0.167322835
7	Sports and travel	174	0.171259843
8	Food and beverages	172	0.169291339
9	Fashion accessories	179	0.176181102
10	Quantity > 5	502	0.494094488
11	Rating > 5	837	0.823818898
12	Male	509	0.500984252
13	Member	507	0.499015748

Lấy ra các feature, Tần suất xuất hiện của mỗi feature, và phần trăm nó xuất hiện.

Sau đó dùng bảng lift để tính ra tỷ lệ tăng hay giảm của một đặc điểm khi có đặc điểm khác xuất hiện

2 Way Lift						
LHS index	RHS index					
12	13					
Feature 1	Feature 2	Occurance %	LHS Freq	RHS Freq	Lift	
Male	Member	0.239173228	0.500984252	0.499016	0.956697	

Liên kết giữa 2 biến cho thấy tỷ lệ nam giới là thành viên giảm 5%

Trong đó biến bên trái LHS là các biến độc lập, biến bên phải RHS là biến phụ thuộc.

Có thể tăng số lượng biến bên trái lên

3 Way Lift						
LHS index	LHS index 2	RHS index				
1	5	10				
A	Feature 2	Feature 3	Occurance %	LHS Freq	RHS Freq	Lift
Credit card	Electronic accessories	Quantity > 5	0.018700787	0.051181	0.494094	0.739504

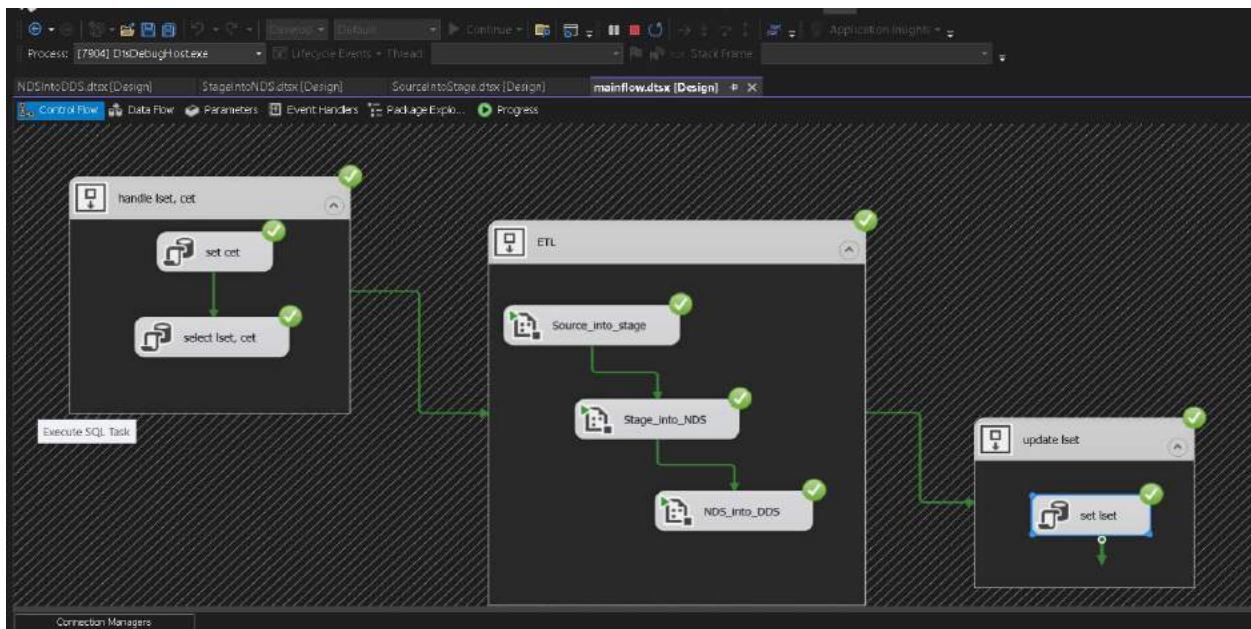
Bảng 3 way lift cho thấy khách hàng thanh toán bằng thẻ khi mua đồ điện tử thì khả năng mua ít hơn 5 món tăng gần 30%

IV. Plato's Pizza

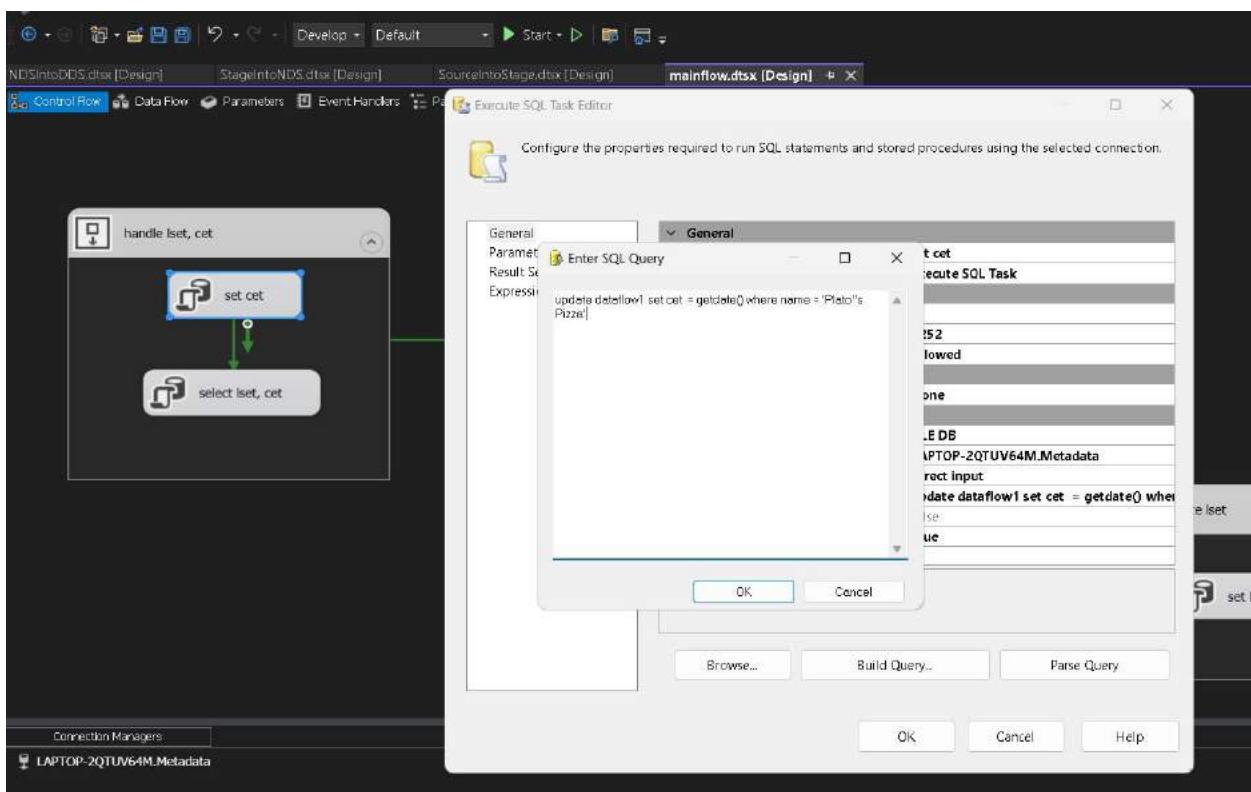
A) Quá trình ETL

1. Main flow

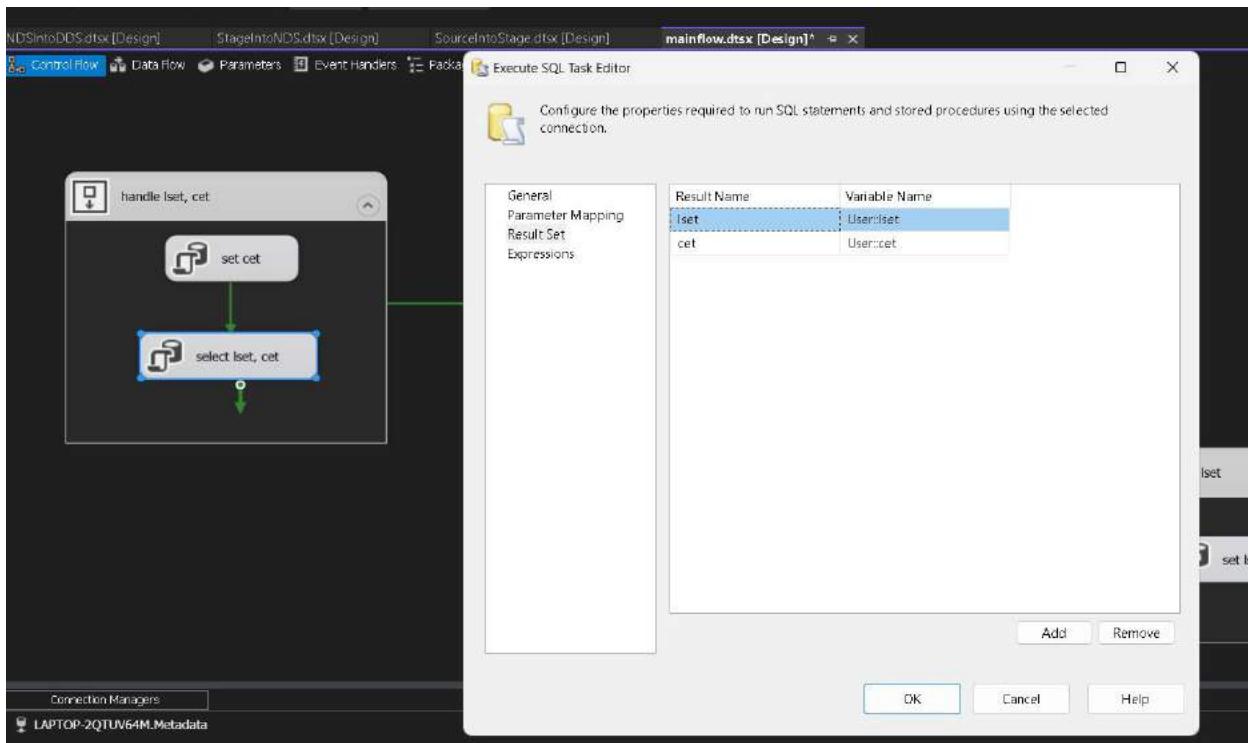
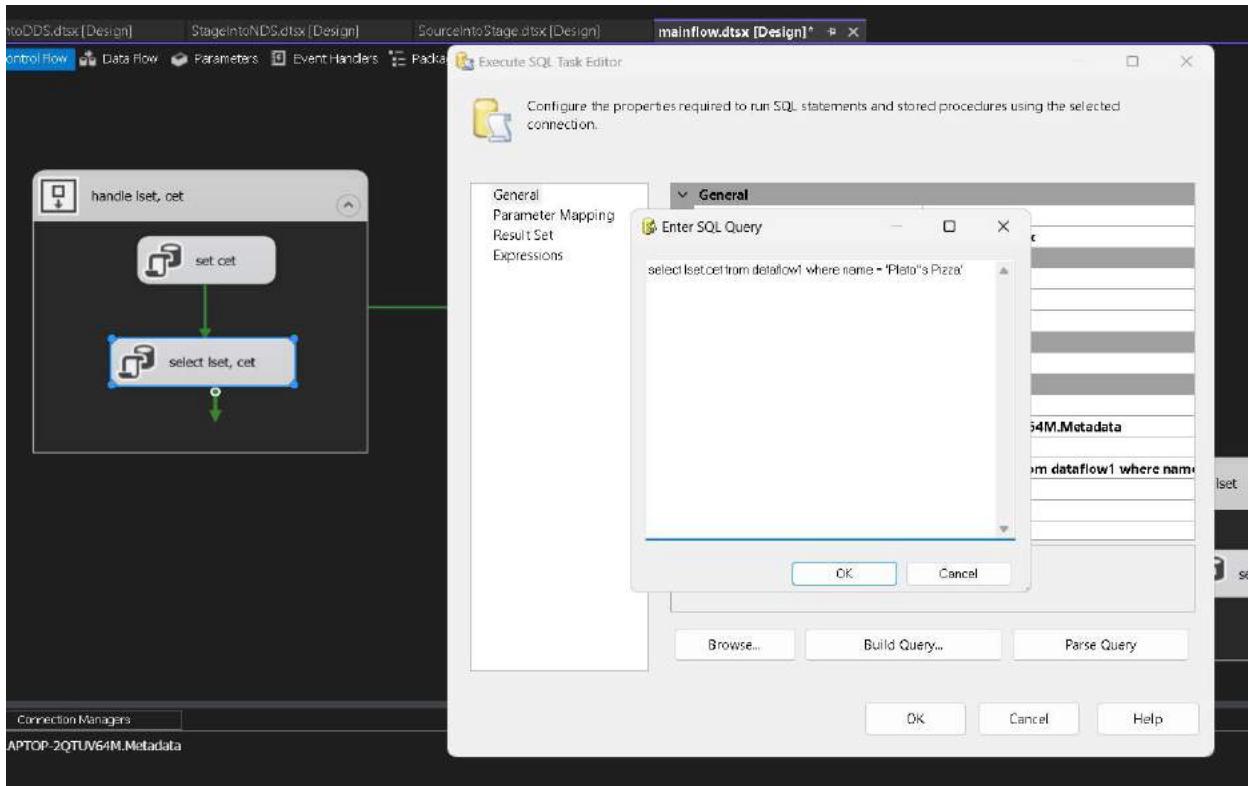
Là main flow quản lí cả quá trình etl



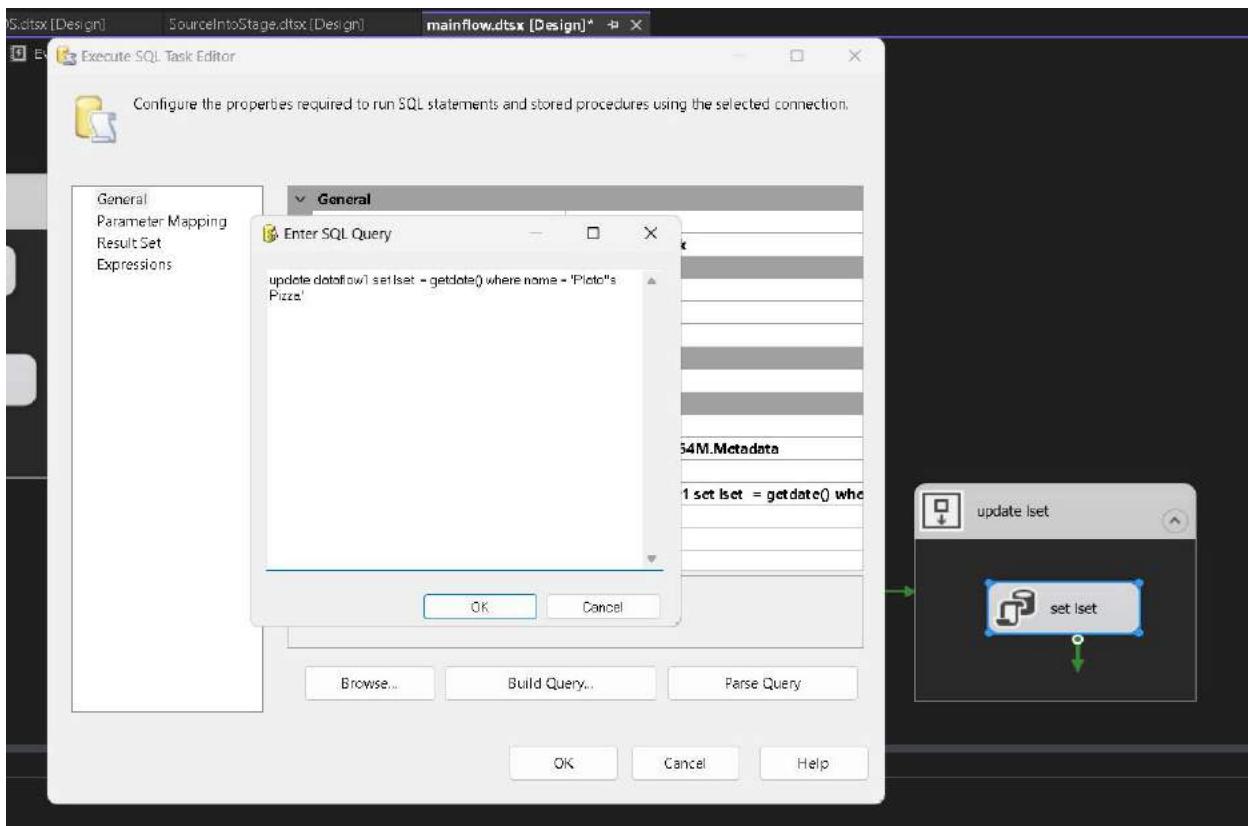
Tương tự như ở Supermarket, đầu tiên set cet trong metadata



Select lset, cet và lưu vào 2 variable để sử dụng trong các package khác



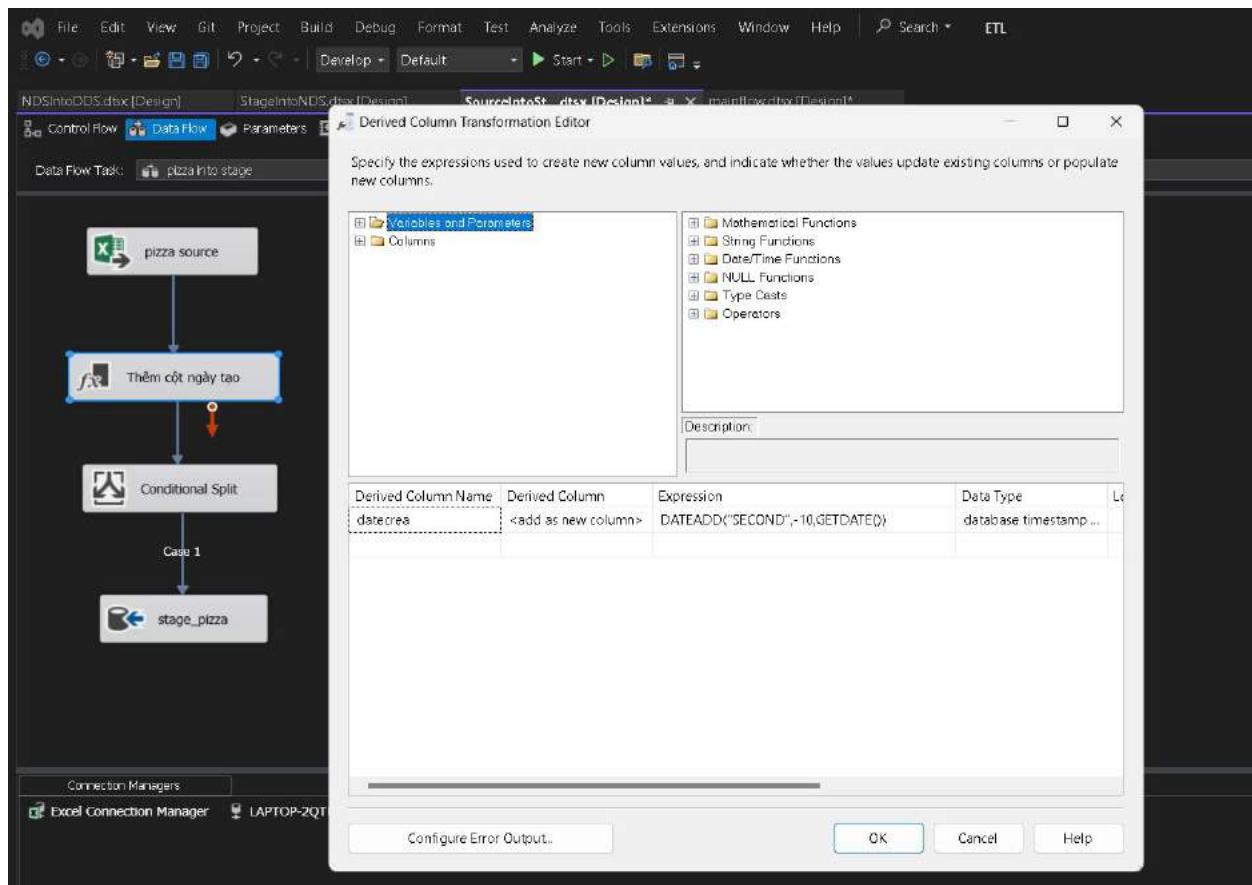
Sau khi quá trình etl hoàn tất thì update lại lset



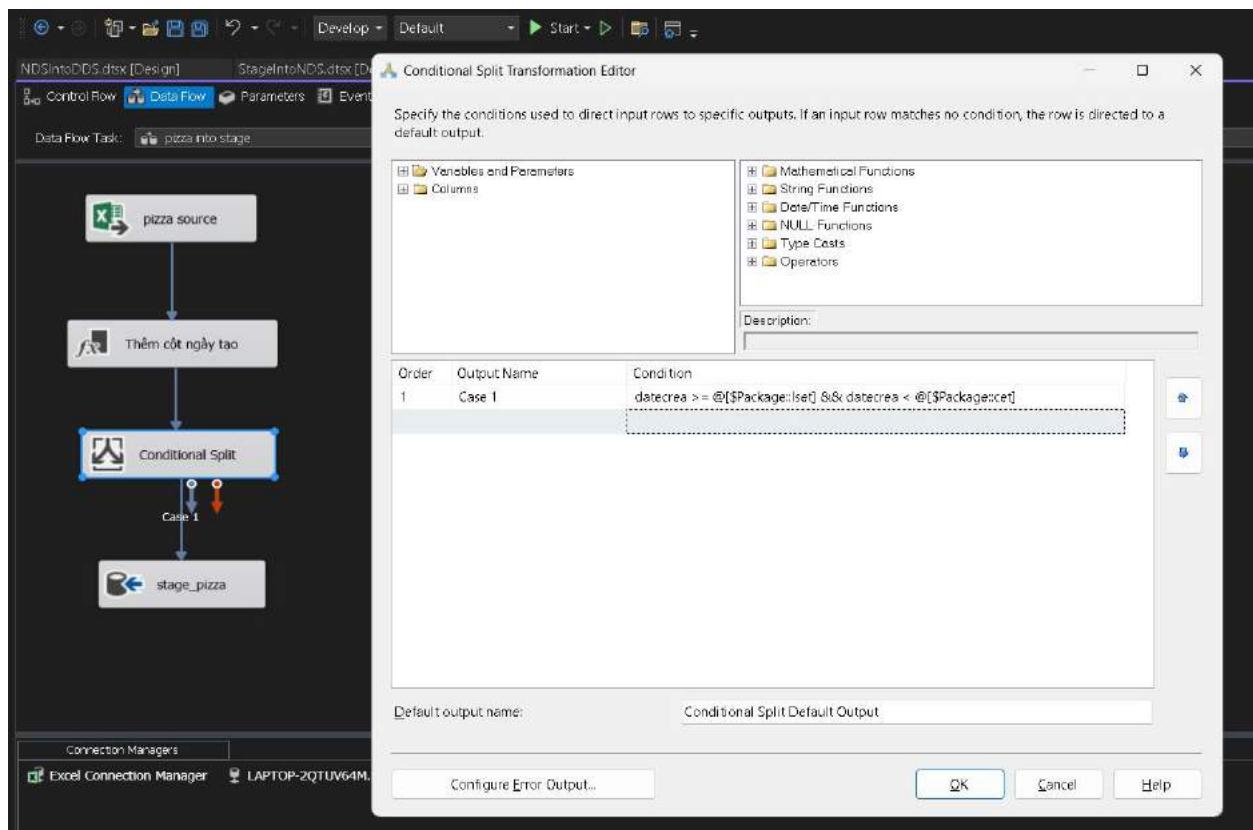
2. Từ Source vào Stage



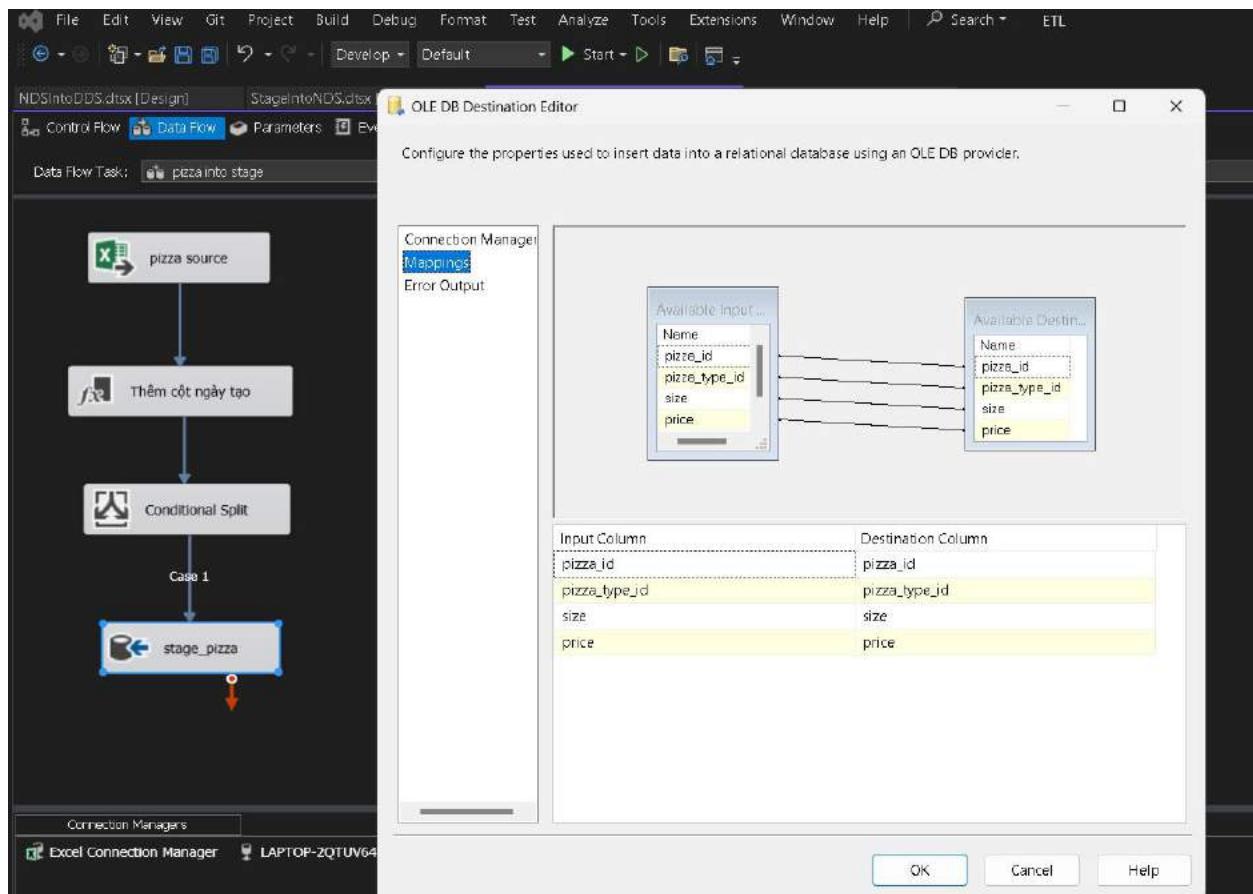
Đầu tiên truncate các bảng stage, tiếp theo thêm datecrea

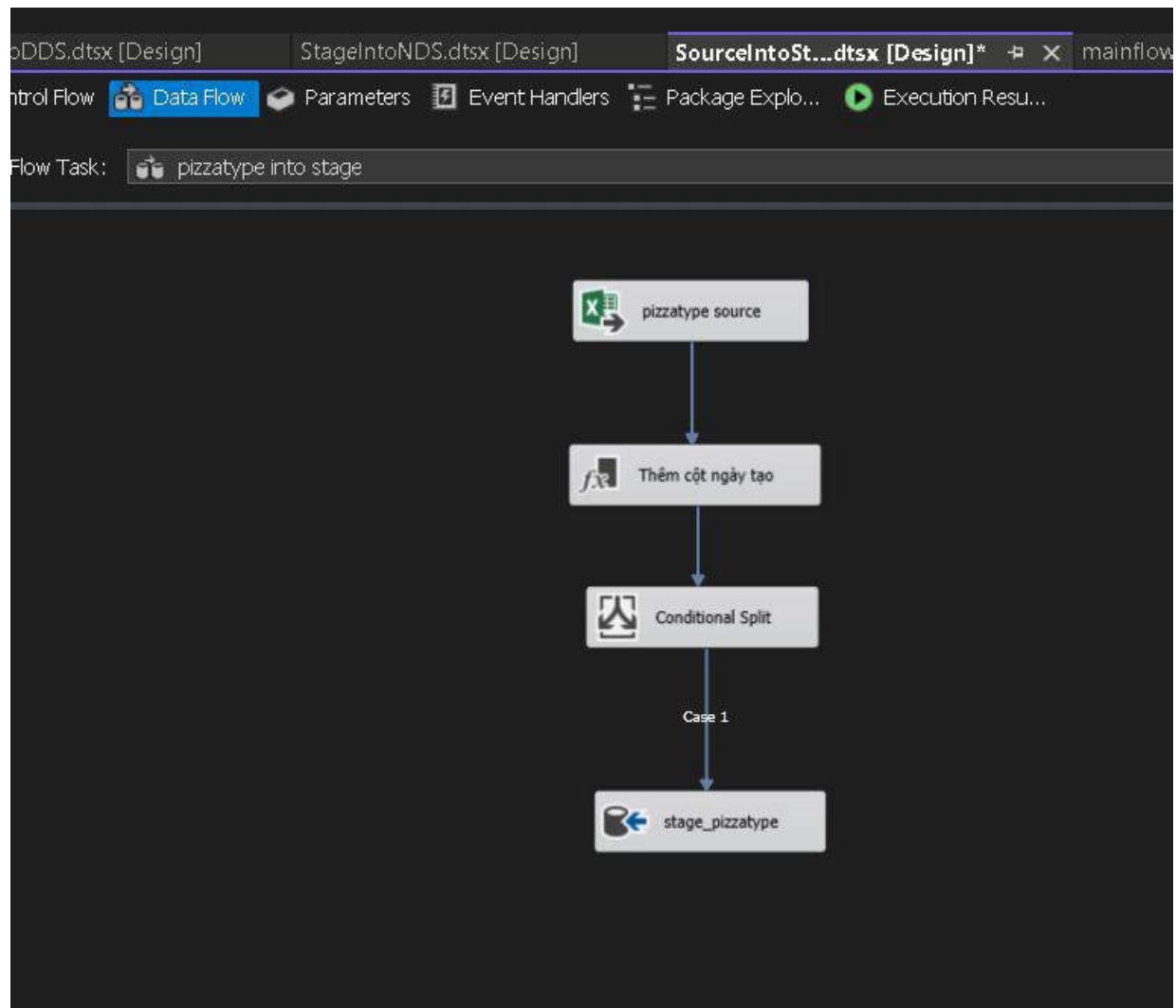


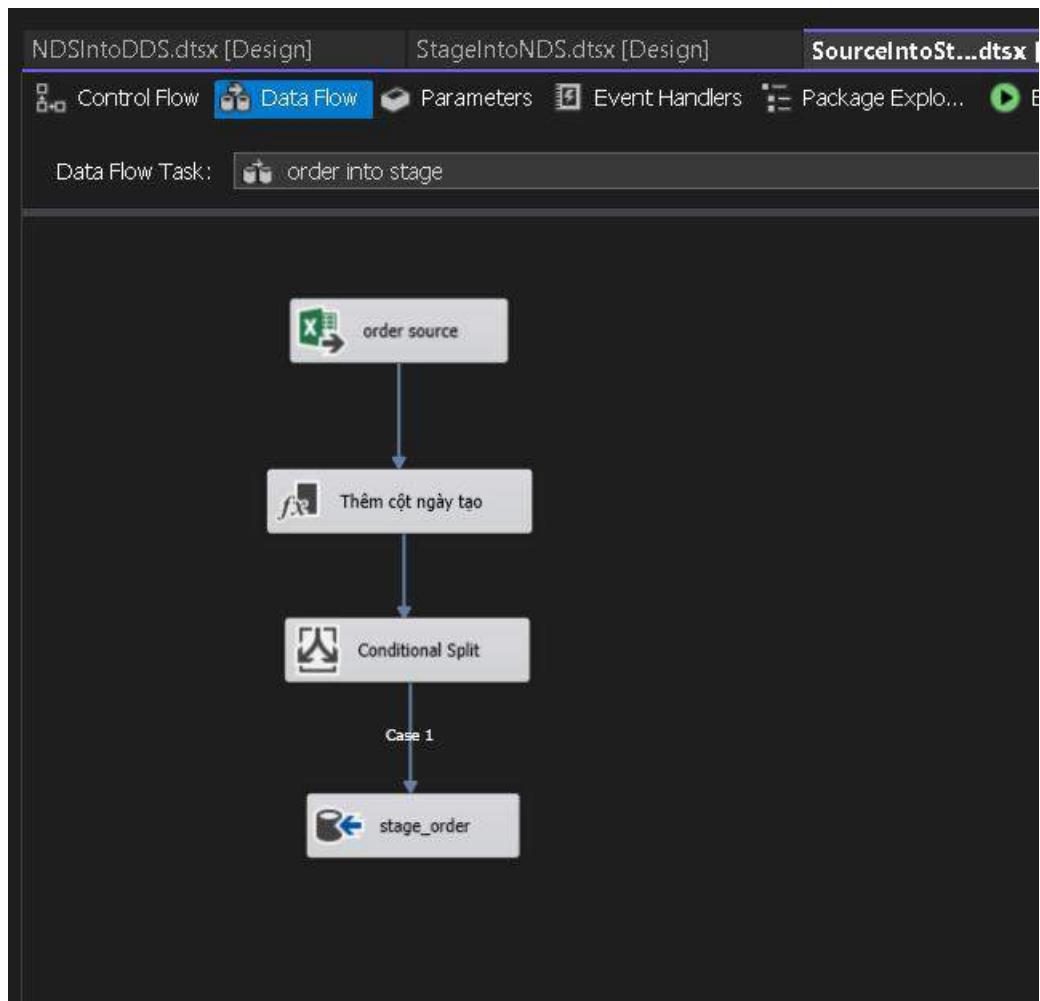
Xét điều kiện datecrea \geq lset và datecrea $<$ cet, nếu thỏa thì sẽ lưu vào stage

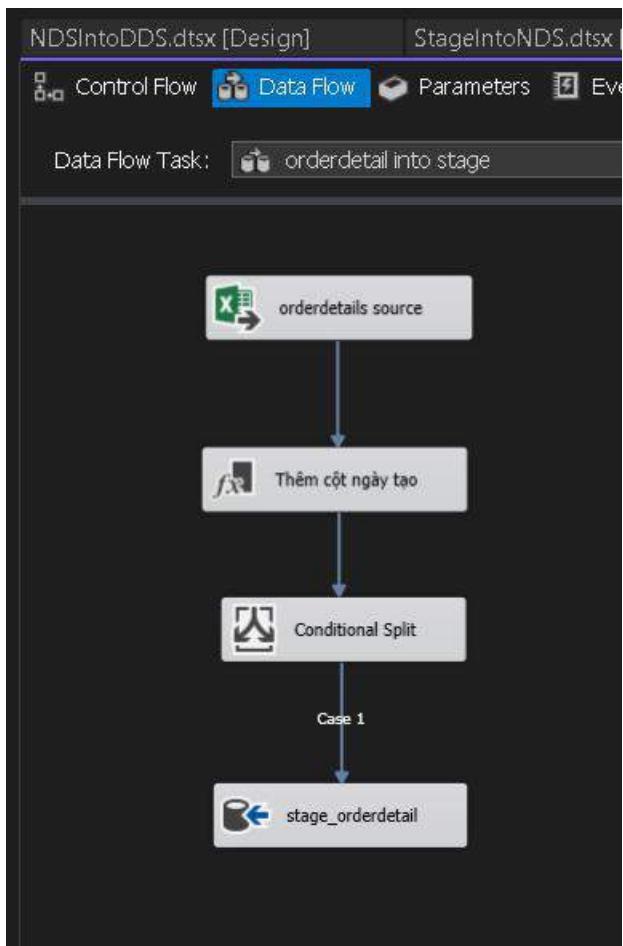


Sau đó mapping các thuộc tính với nhau, tương tự các source cũng làm tương tự





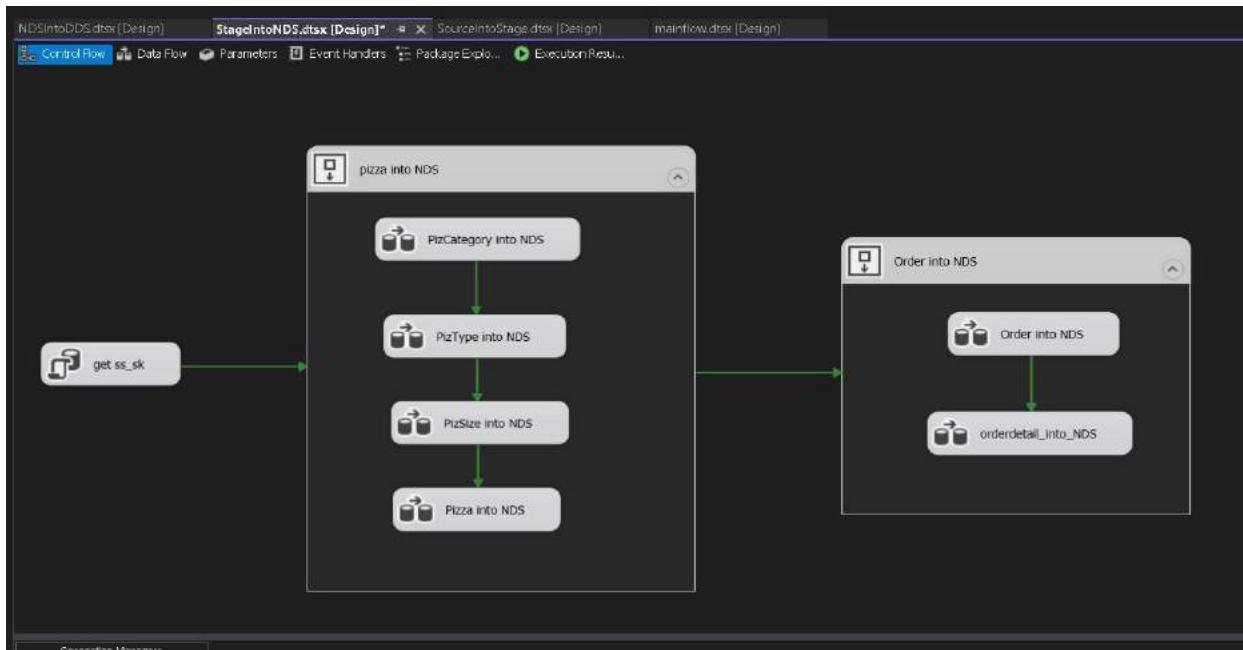
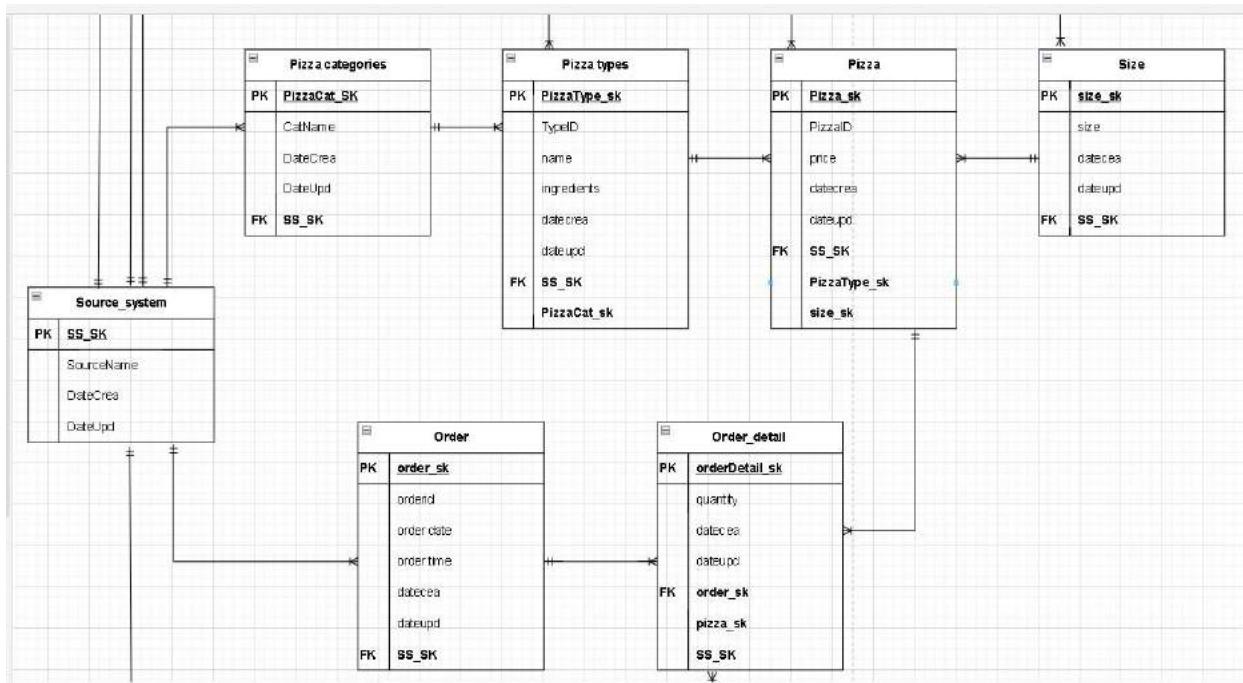




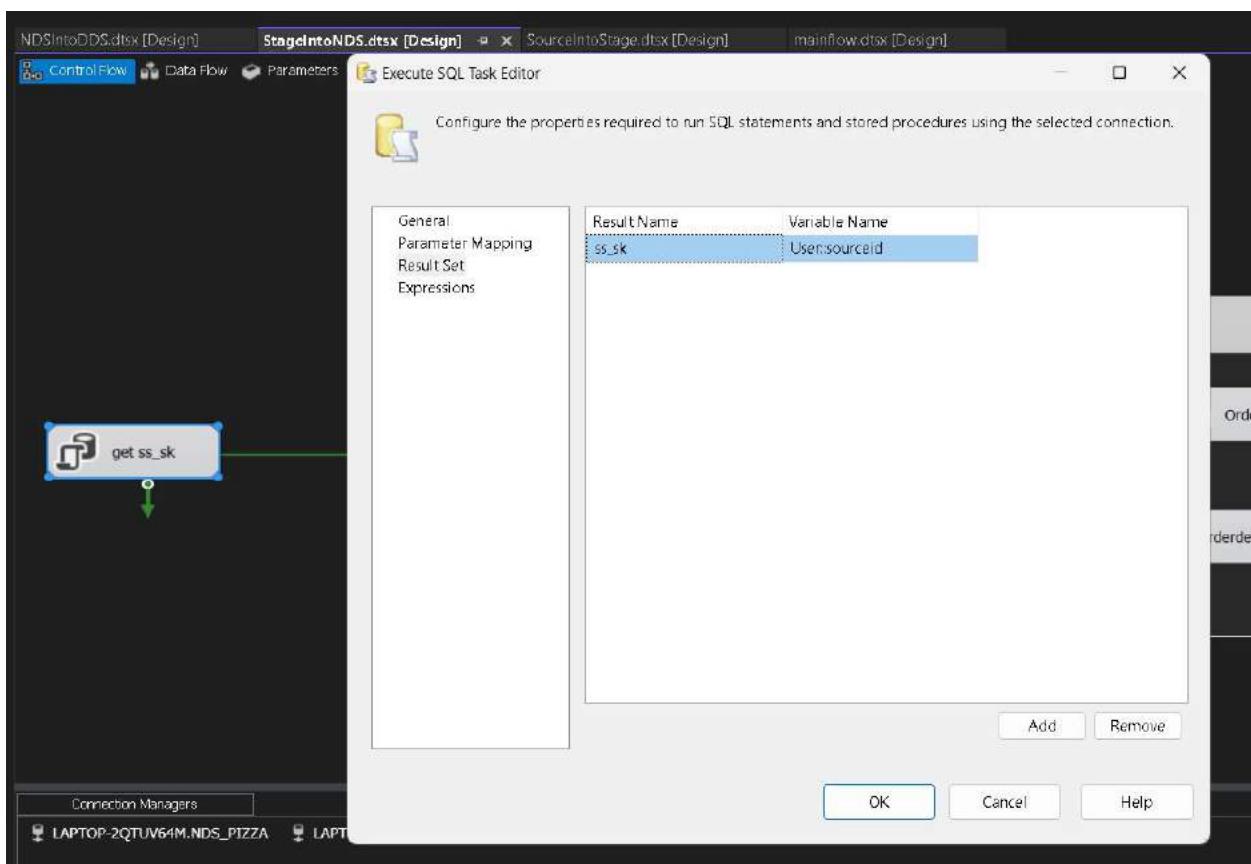
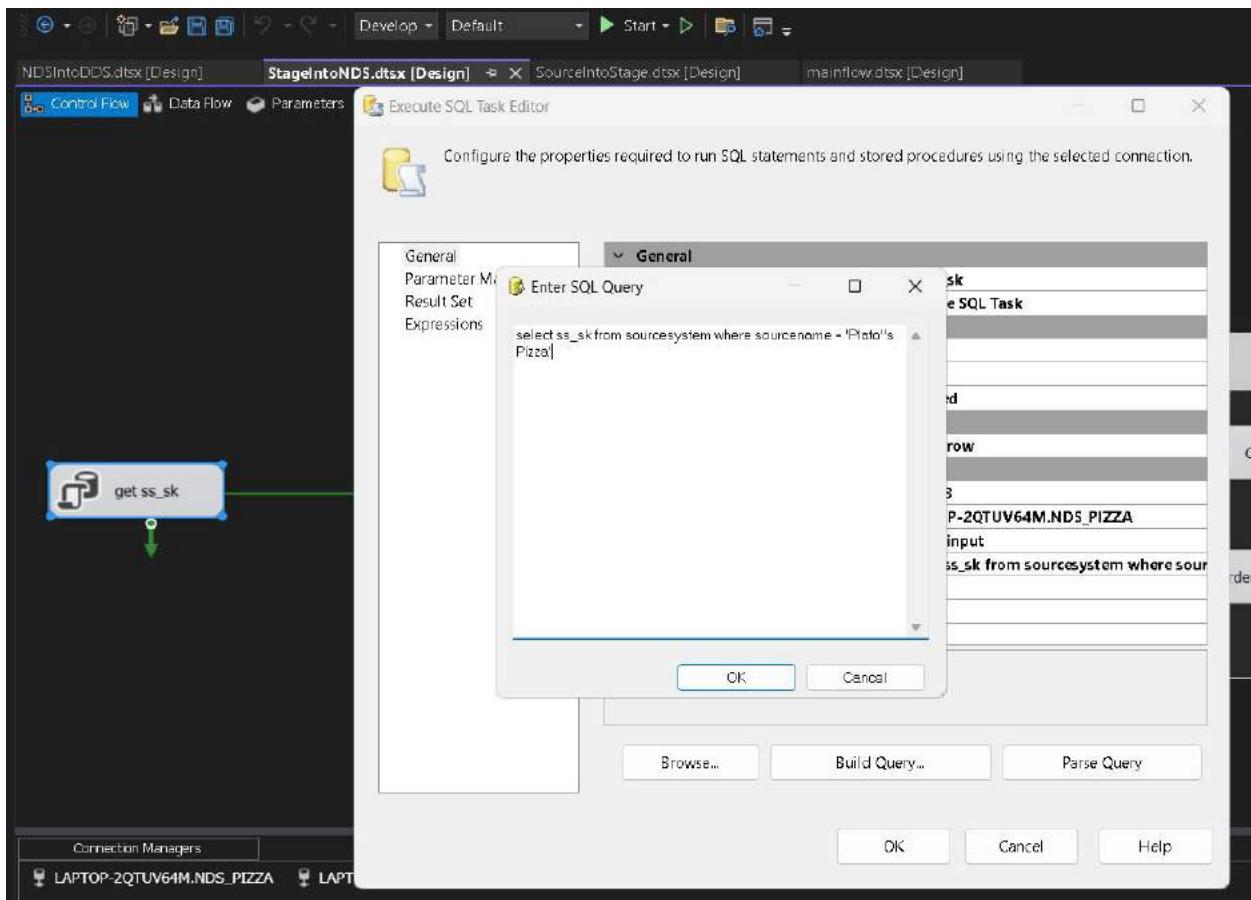
3. Stage vào NDS

Thiết kế NDS:

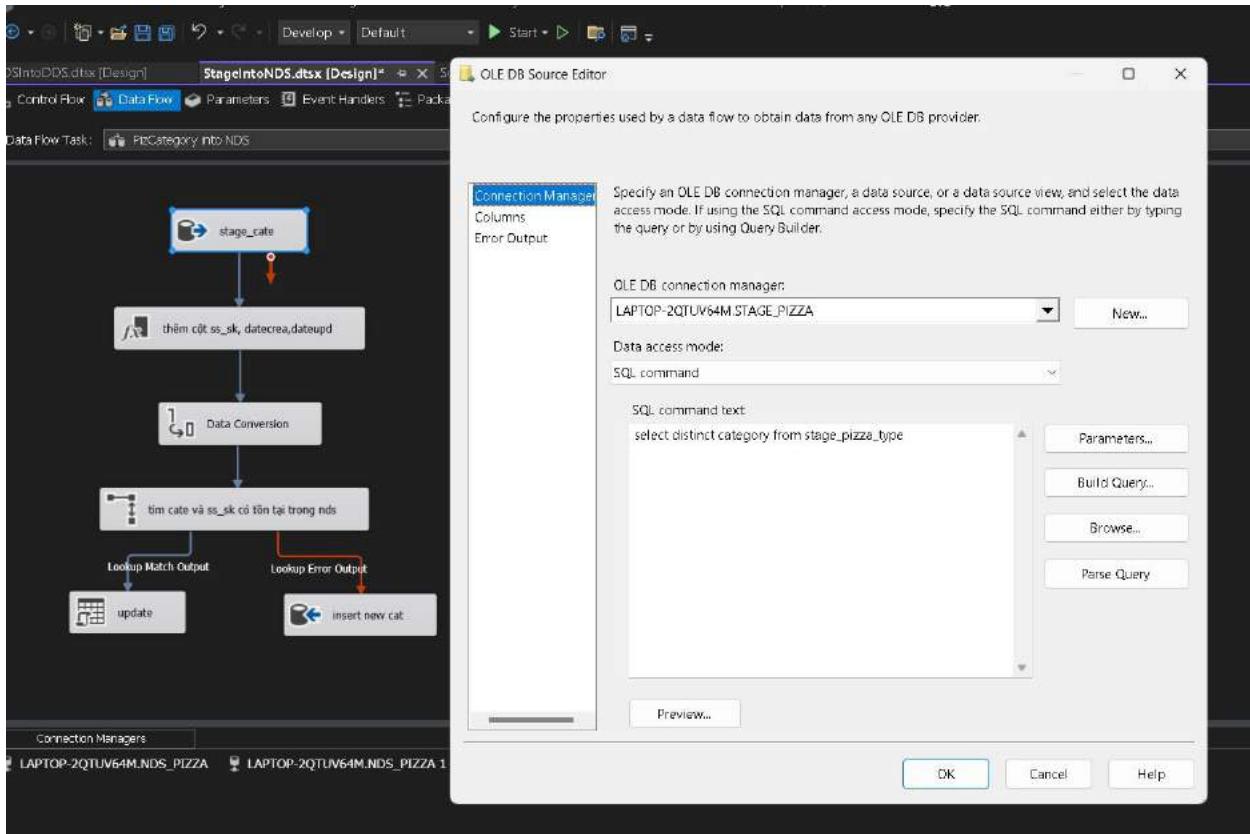
<https://drive.google.com/file/d/1qKHvbGh2l1Tor22k7Vw8ffdAW4UPzz7S/view?usp=sharing>



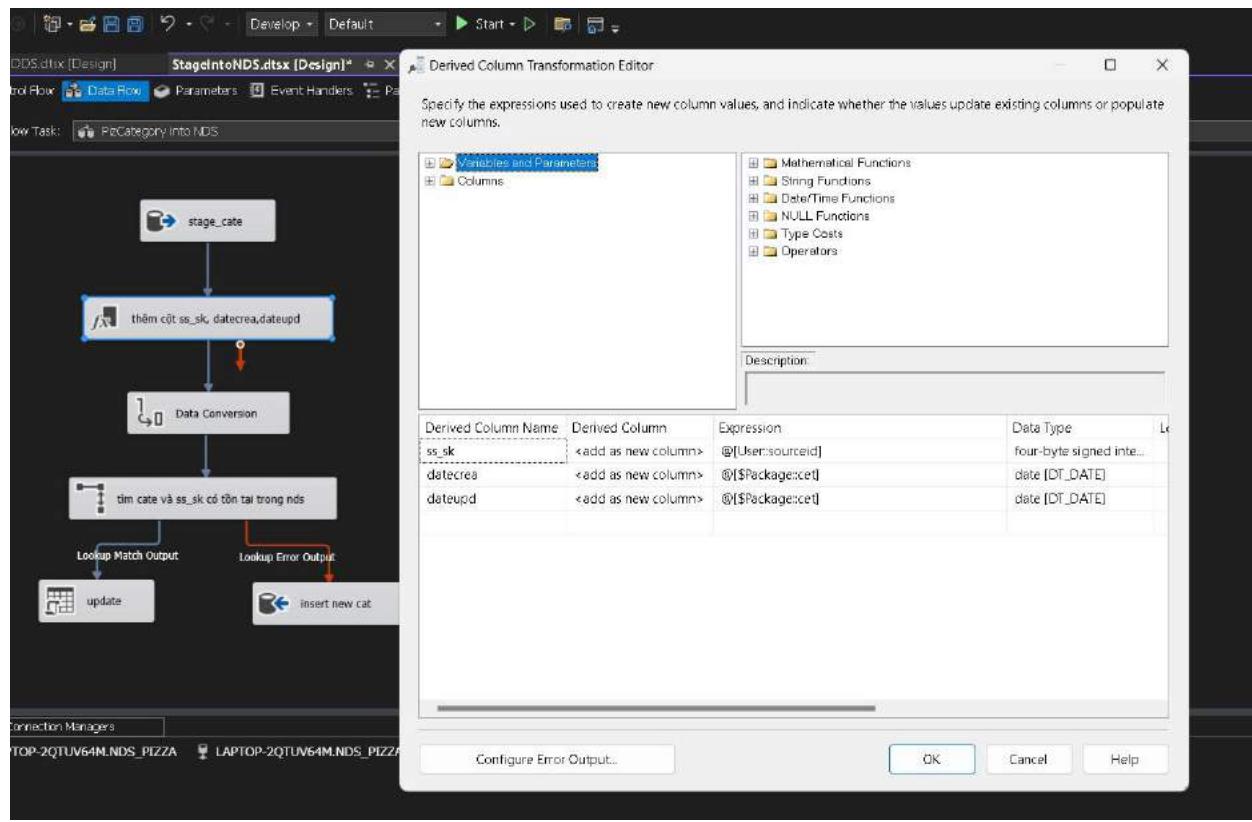
Đầu tiên select sourceid



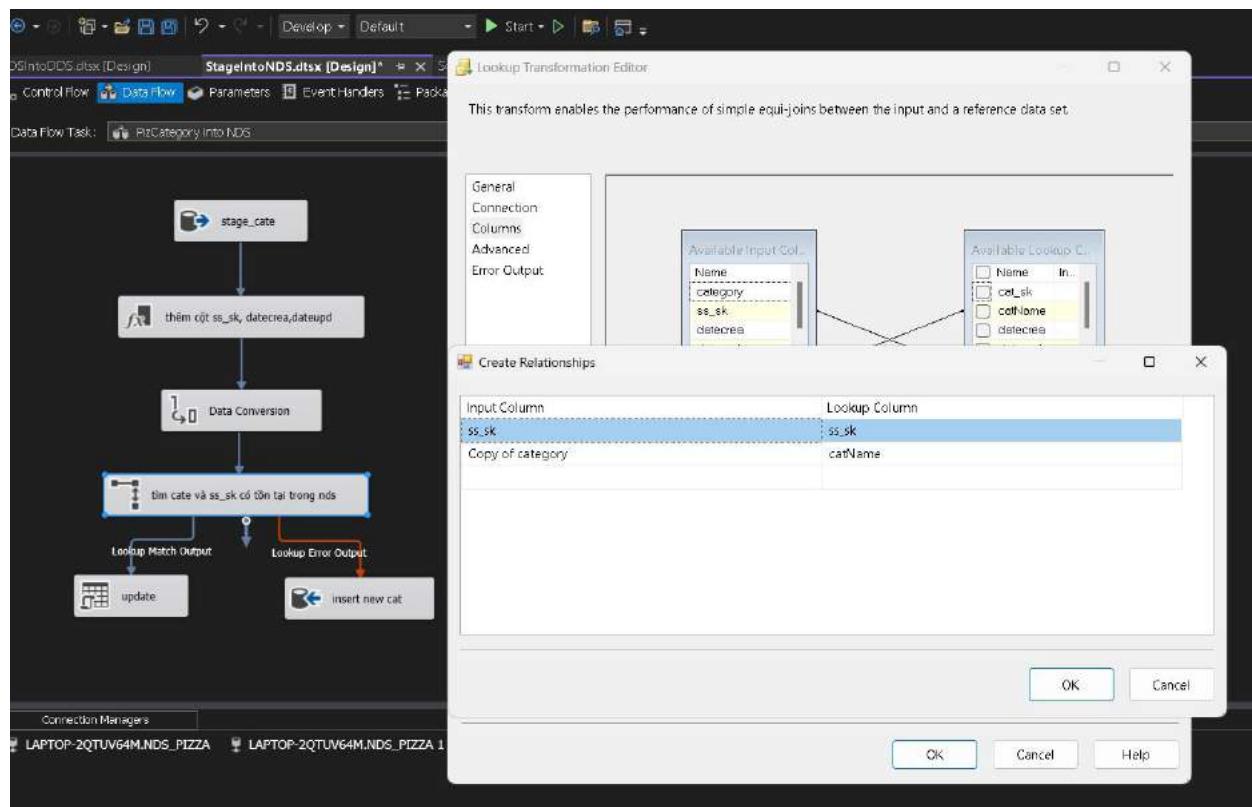
Pizzacategoey_stage lưu vào pizzacate_nds , tương tự như ở supermarket



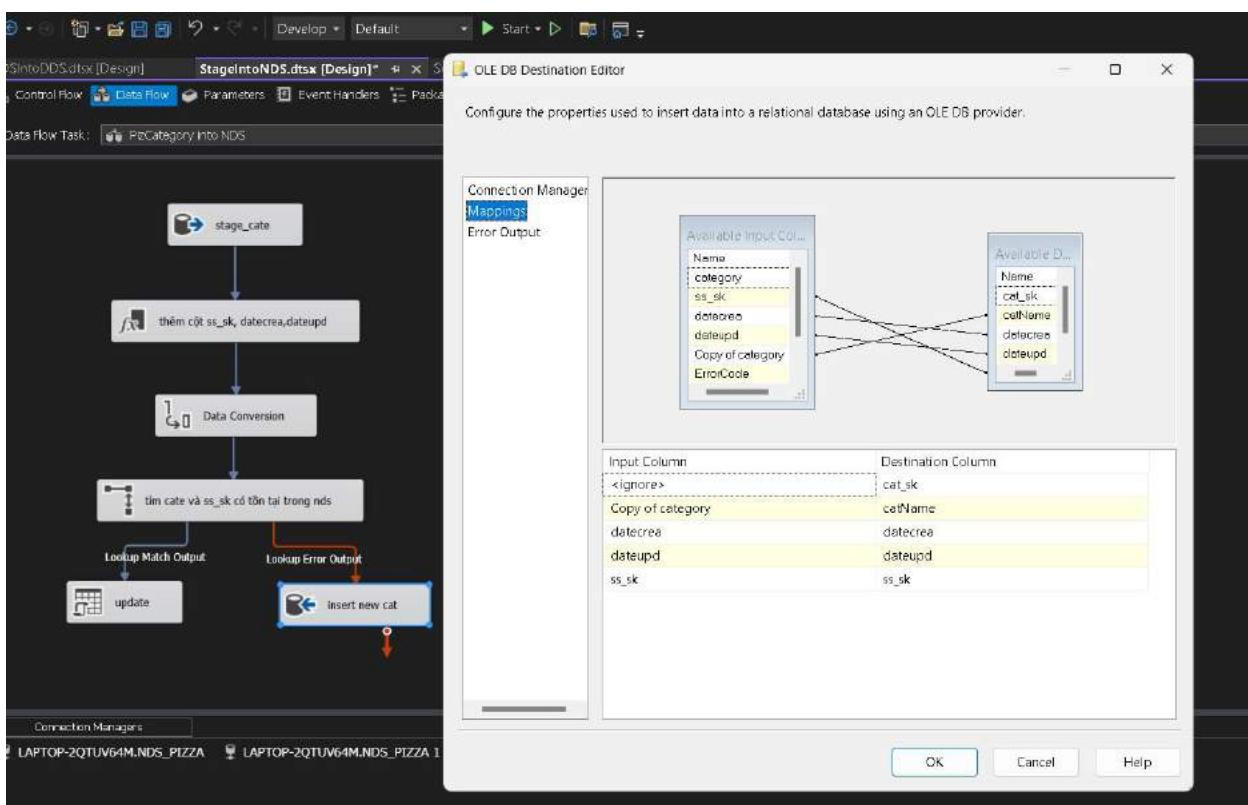
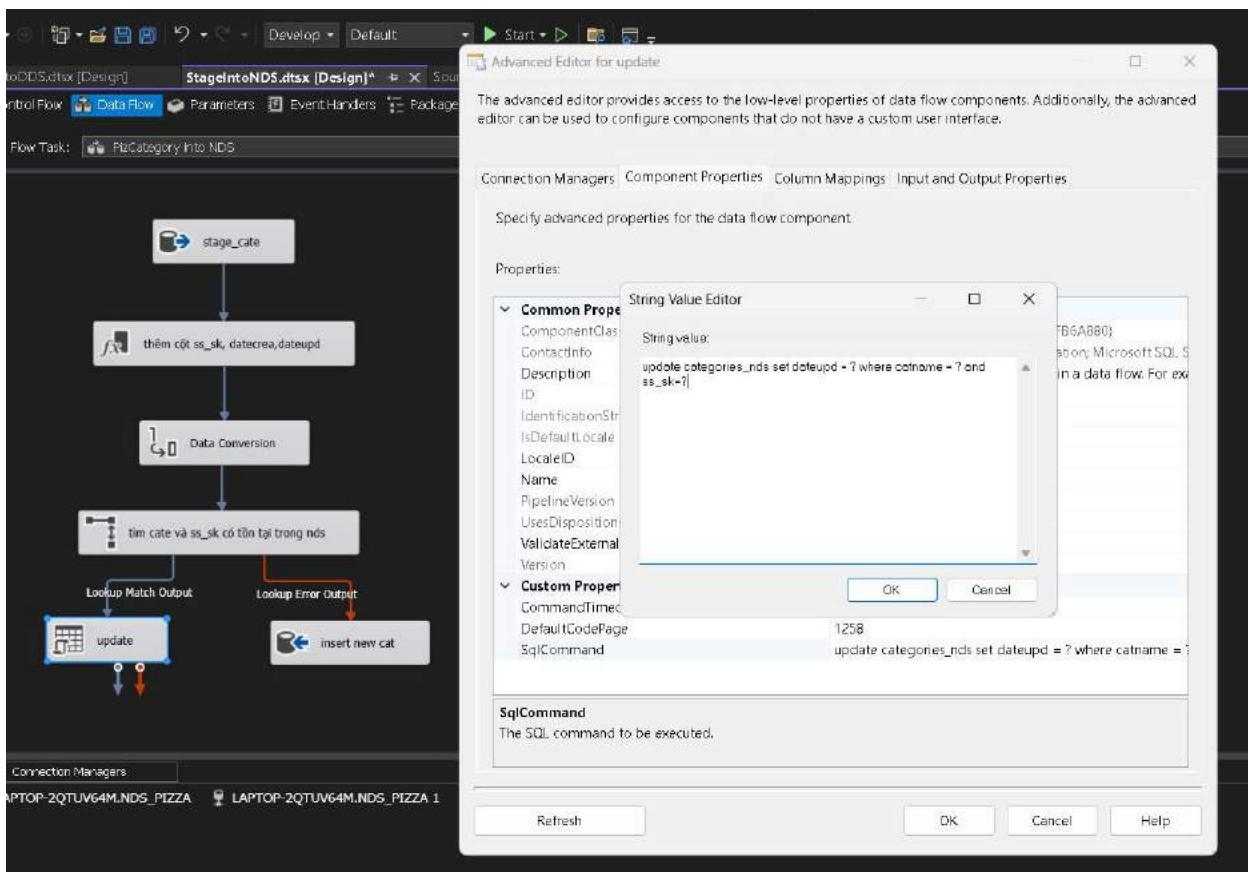
Thêm cột sourceid, datecrea, dateupd



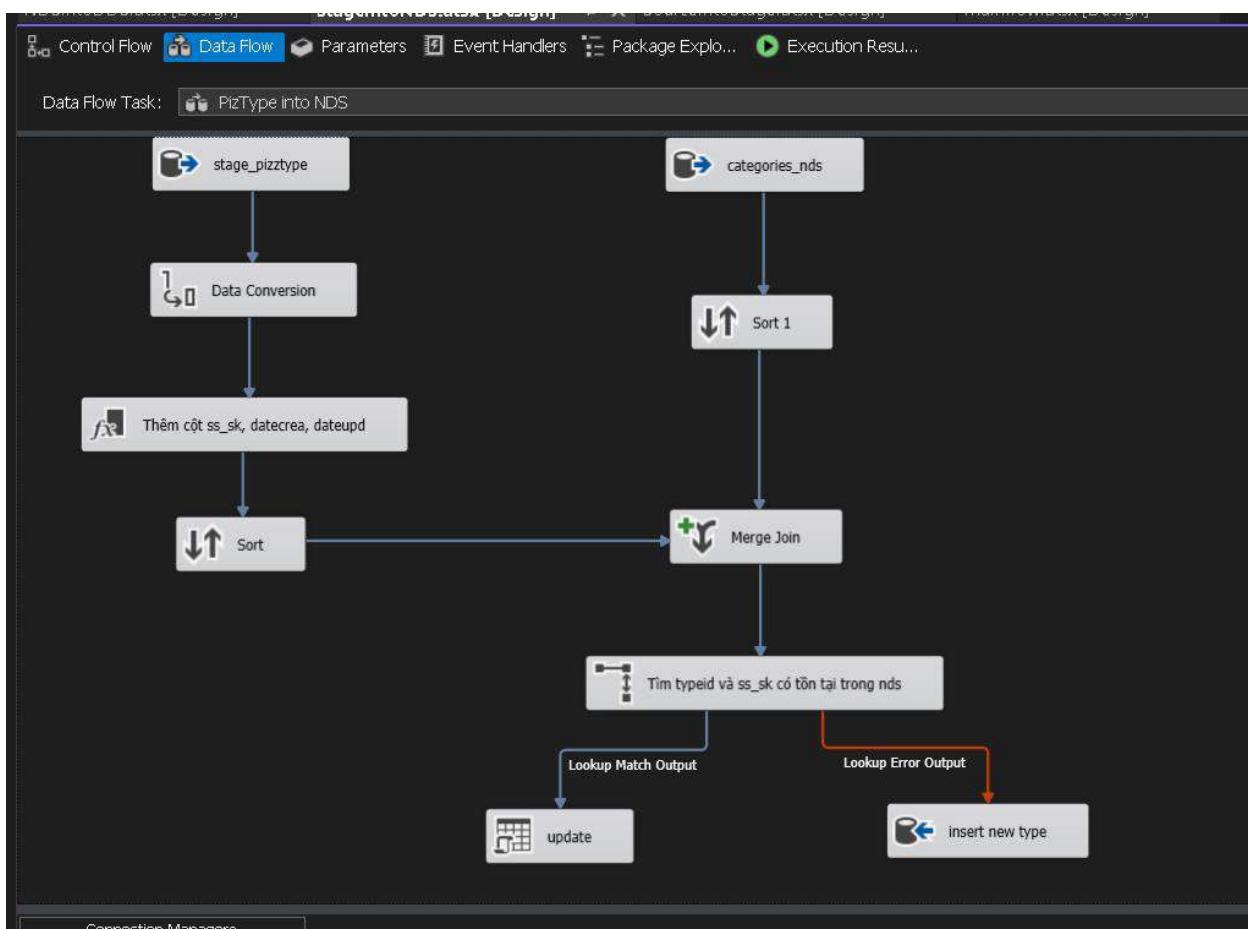
Tìm sourceid và category có tồn tại trong pizzacategori_nds không



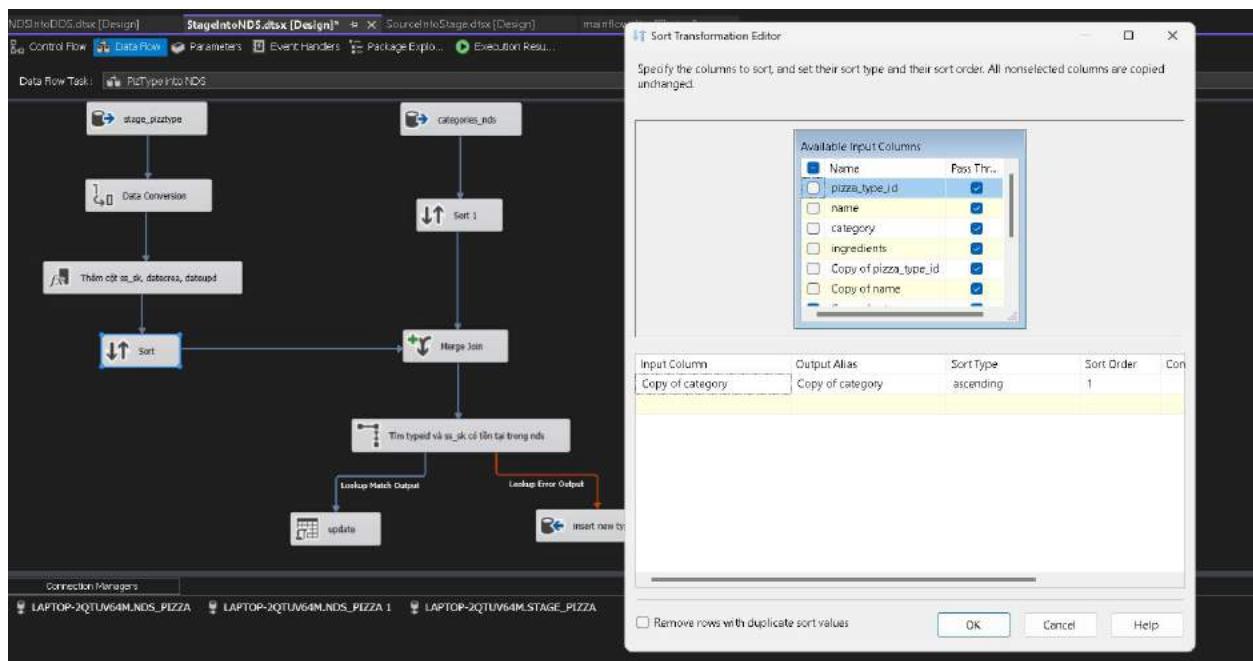
Nếu tồn tại thì update , còn không thì insert category mới

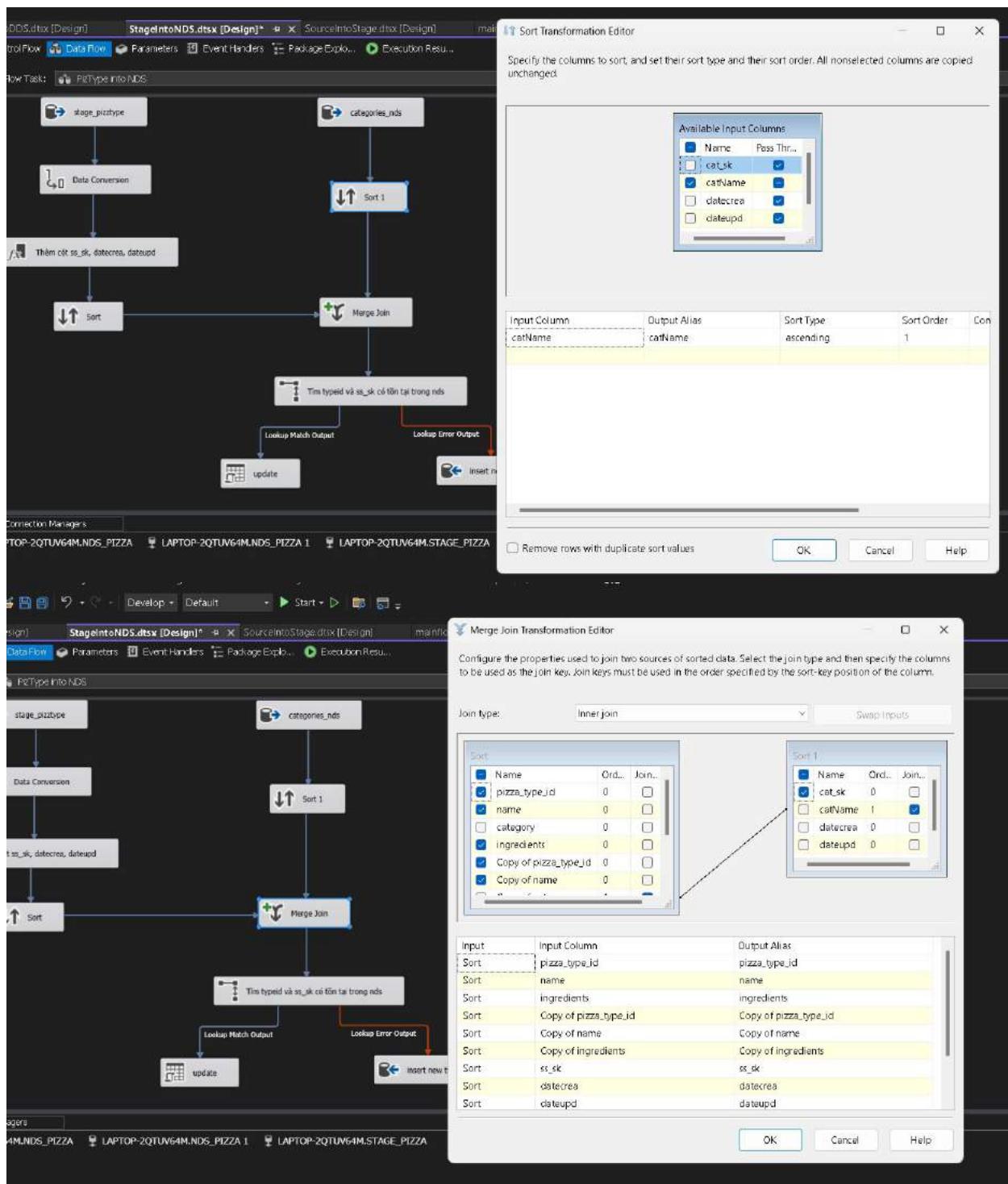


Tương tự PizzaType_stage vào pizzaType_nds



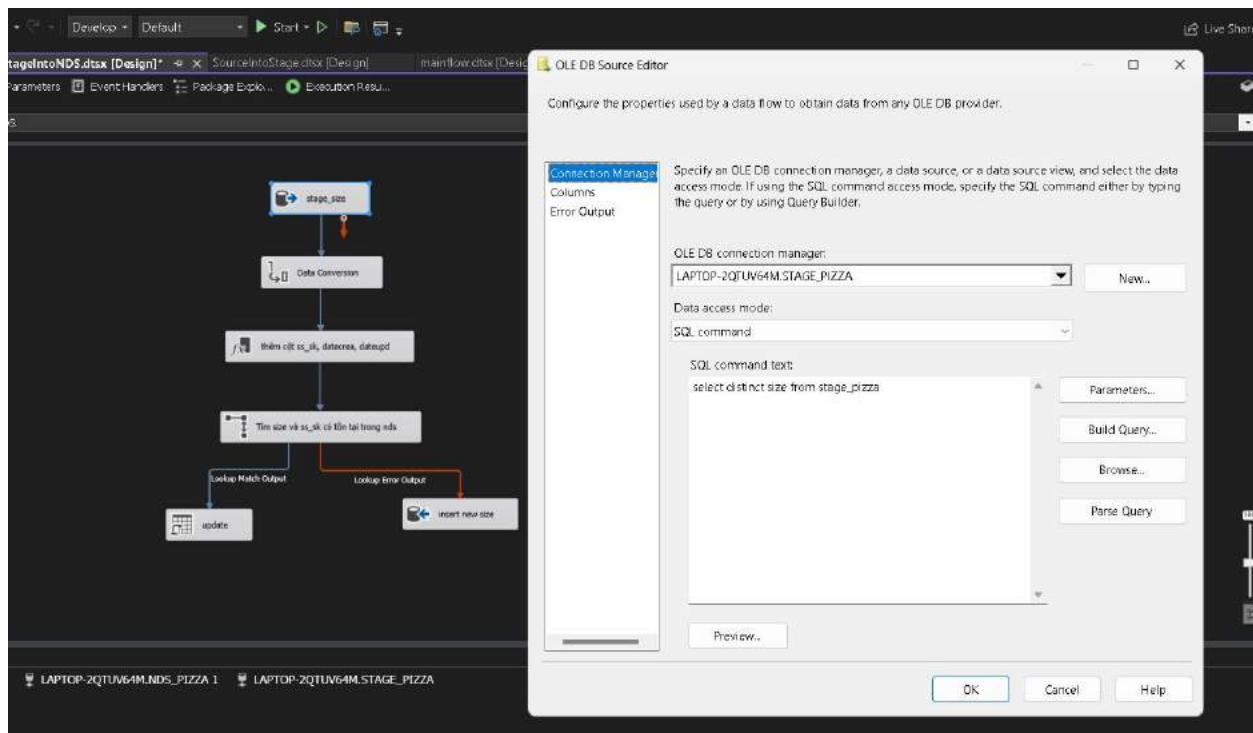
Pizzatype_stage merge join với pizzacategories_nds theo category và lấy surrogate key của pizzacategories_nds



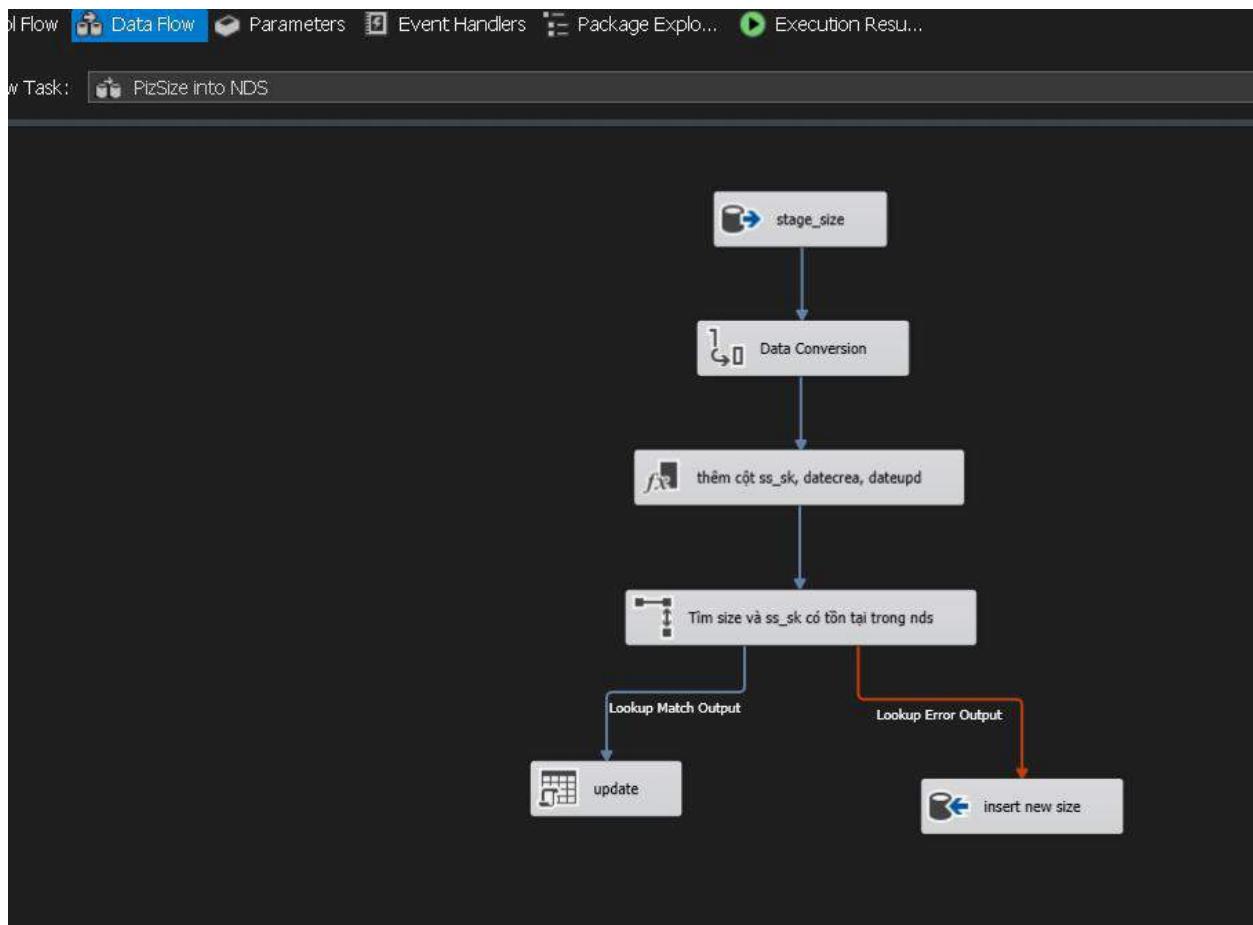


Và tương tự tìm sourceid và pizzatype_id có tồn tại trong pizzatype_nds không.

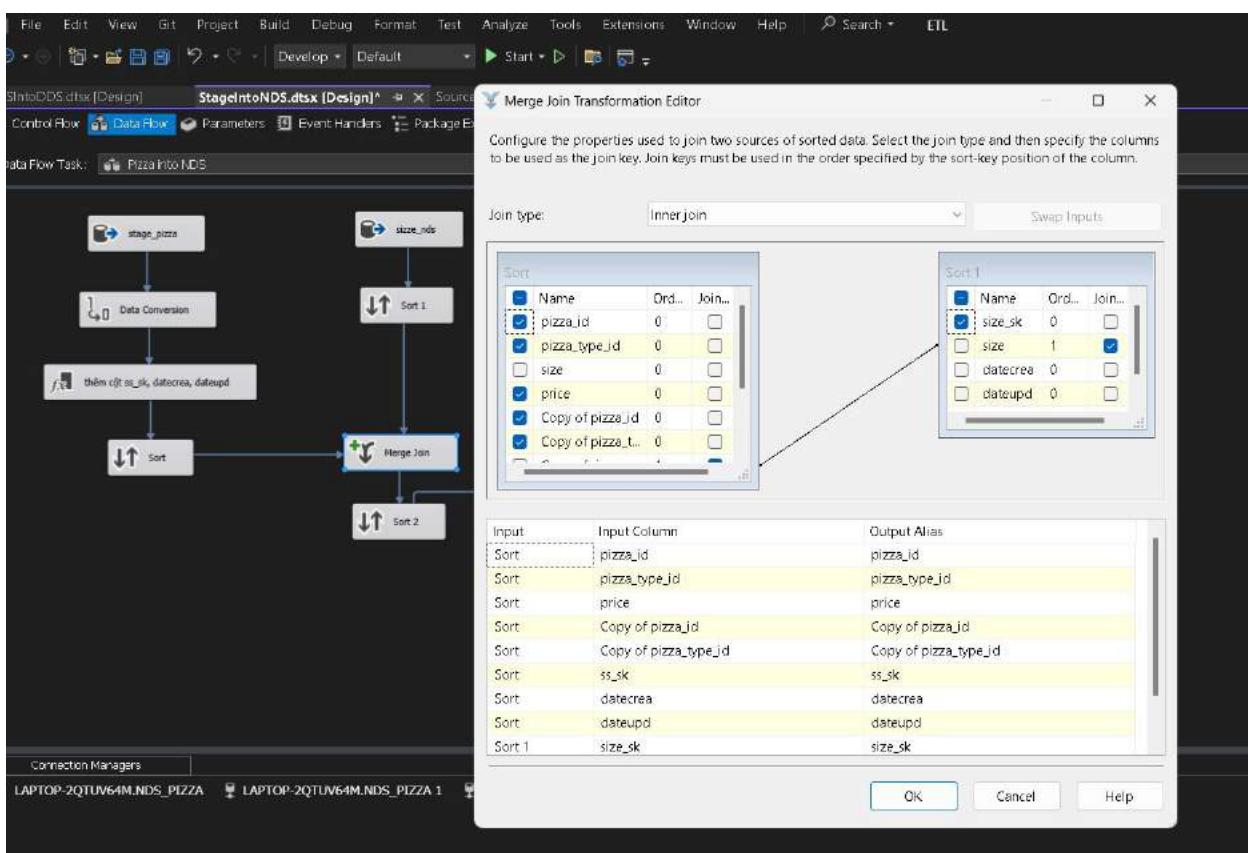
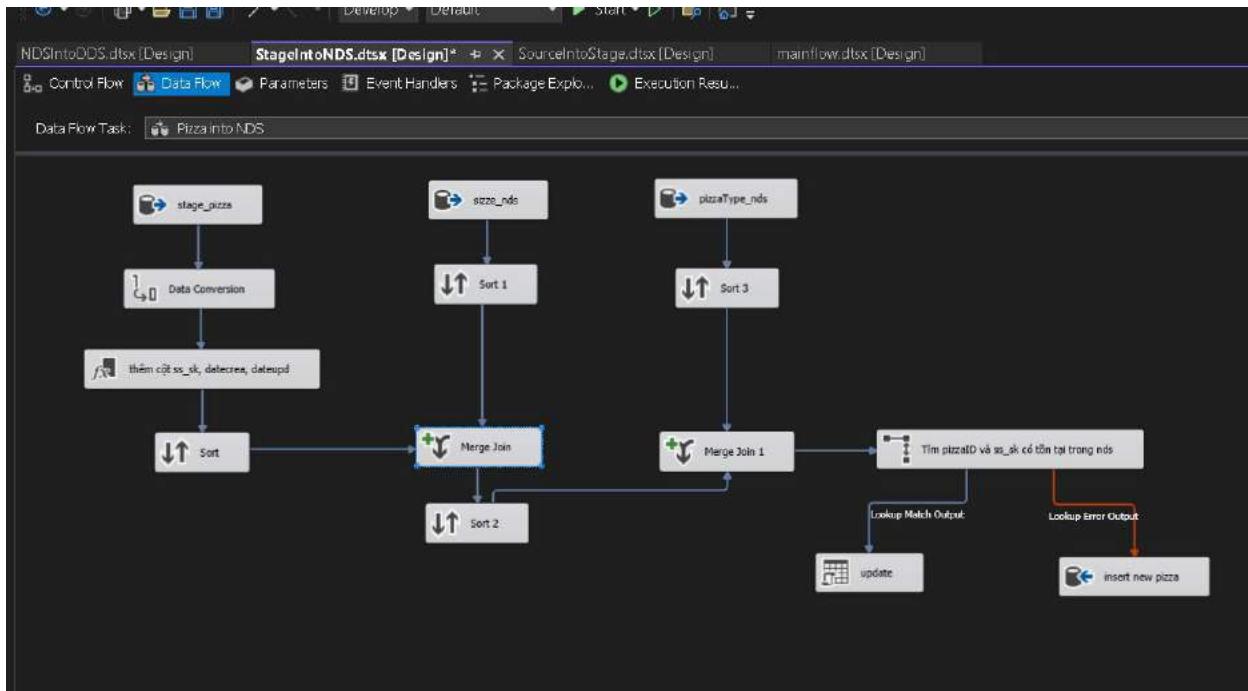
Tiếp theo pizzasize_stage vào pizzasize_nds



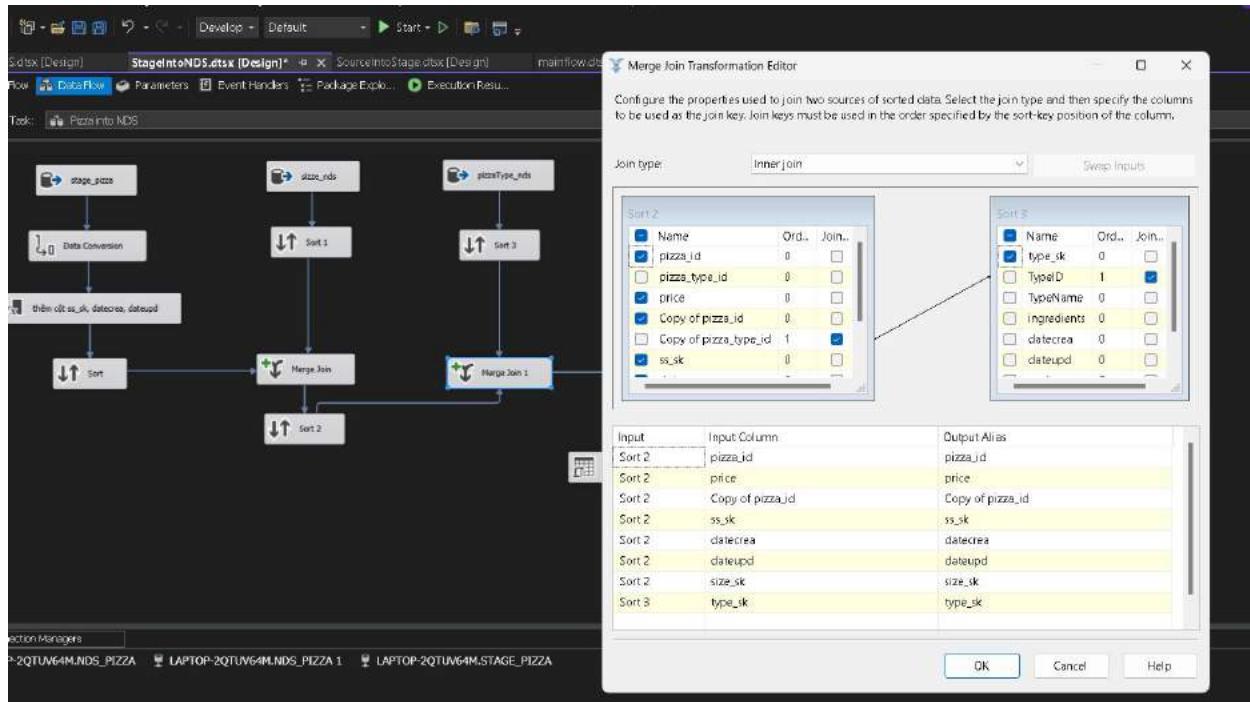
Thêm cột sourceid, datecrea,dateupd tương tự như trên



Tiếp theo là pizza_stage vào pizza_nds, đầu tiên merge join với size_nds theo size lấy SK của size_nds

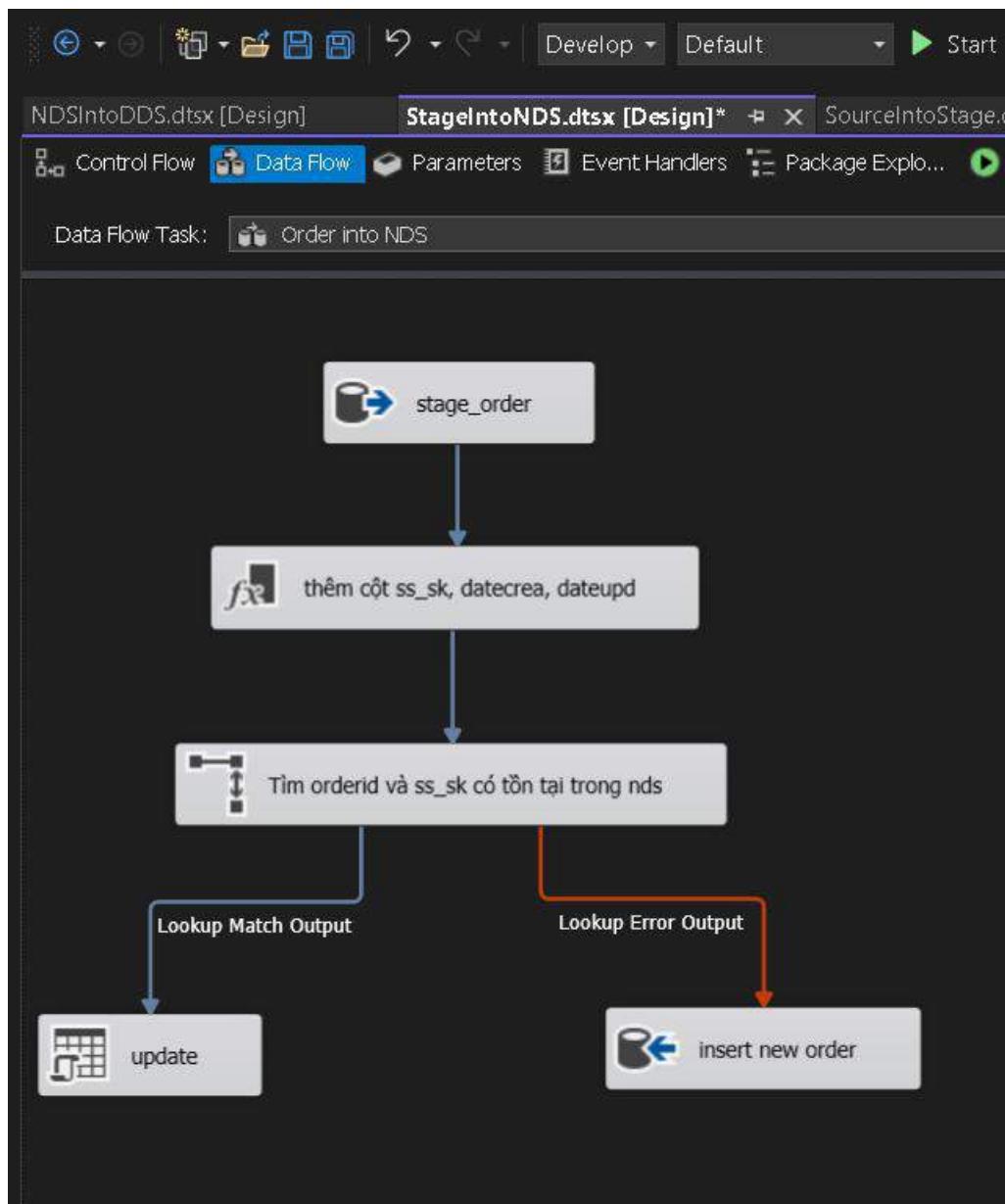


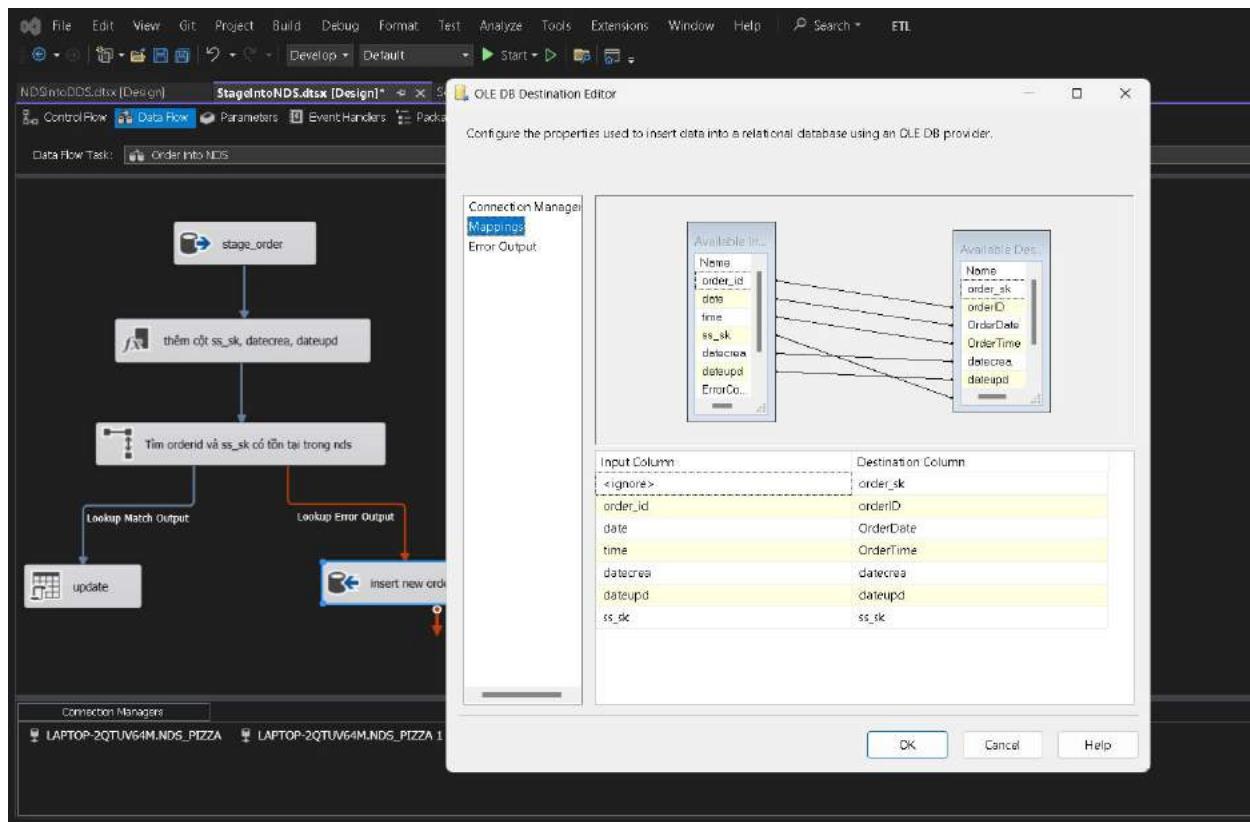
Pizza_stage merge join với pizztype_nds theo pizzatypeID lấy sk của pizztype_nds



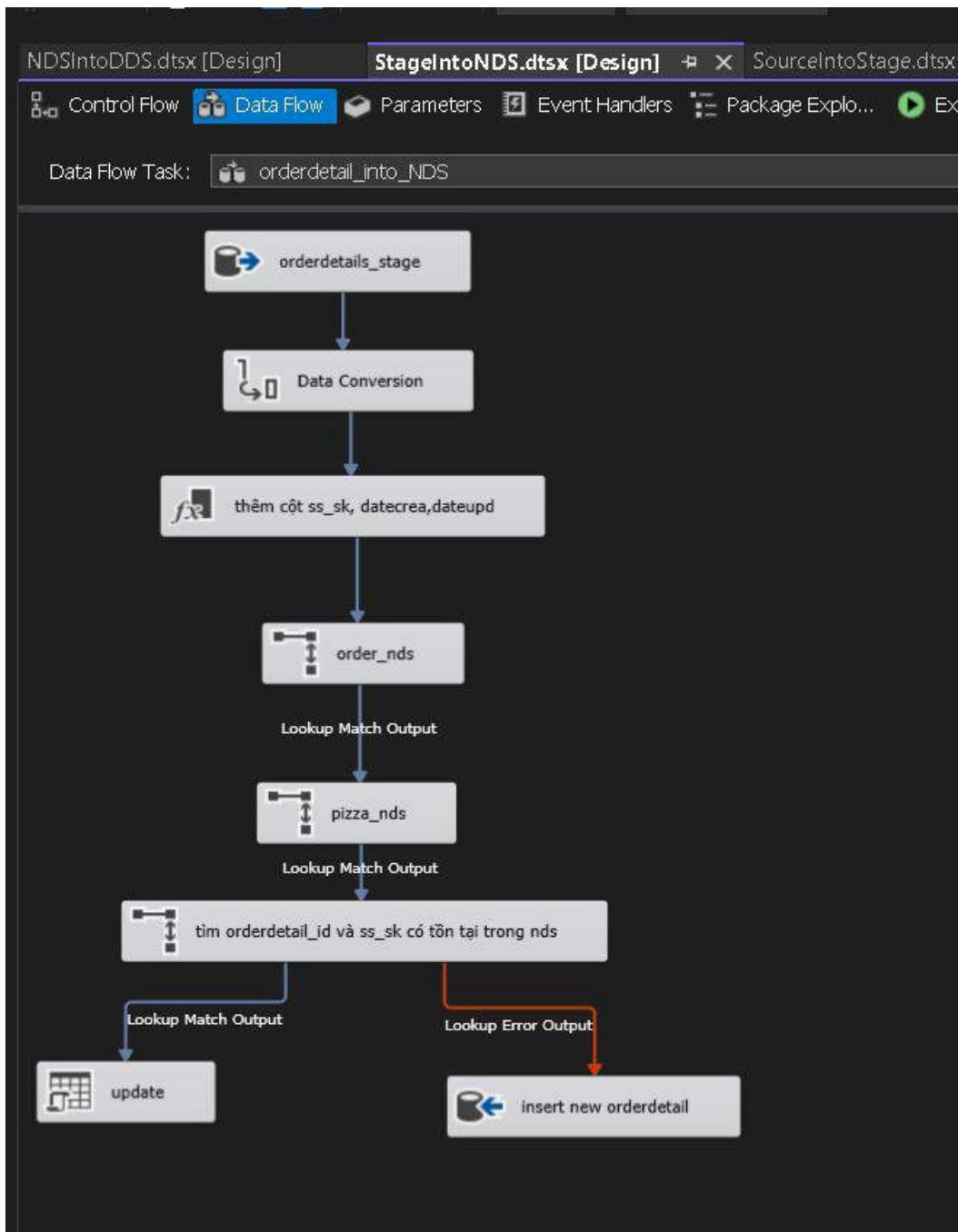
Và tương tự tìm sourceid và pizzaid có tồn tại trong pizza_nds không

Tiếp theo order_stage vào order_nds

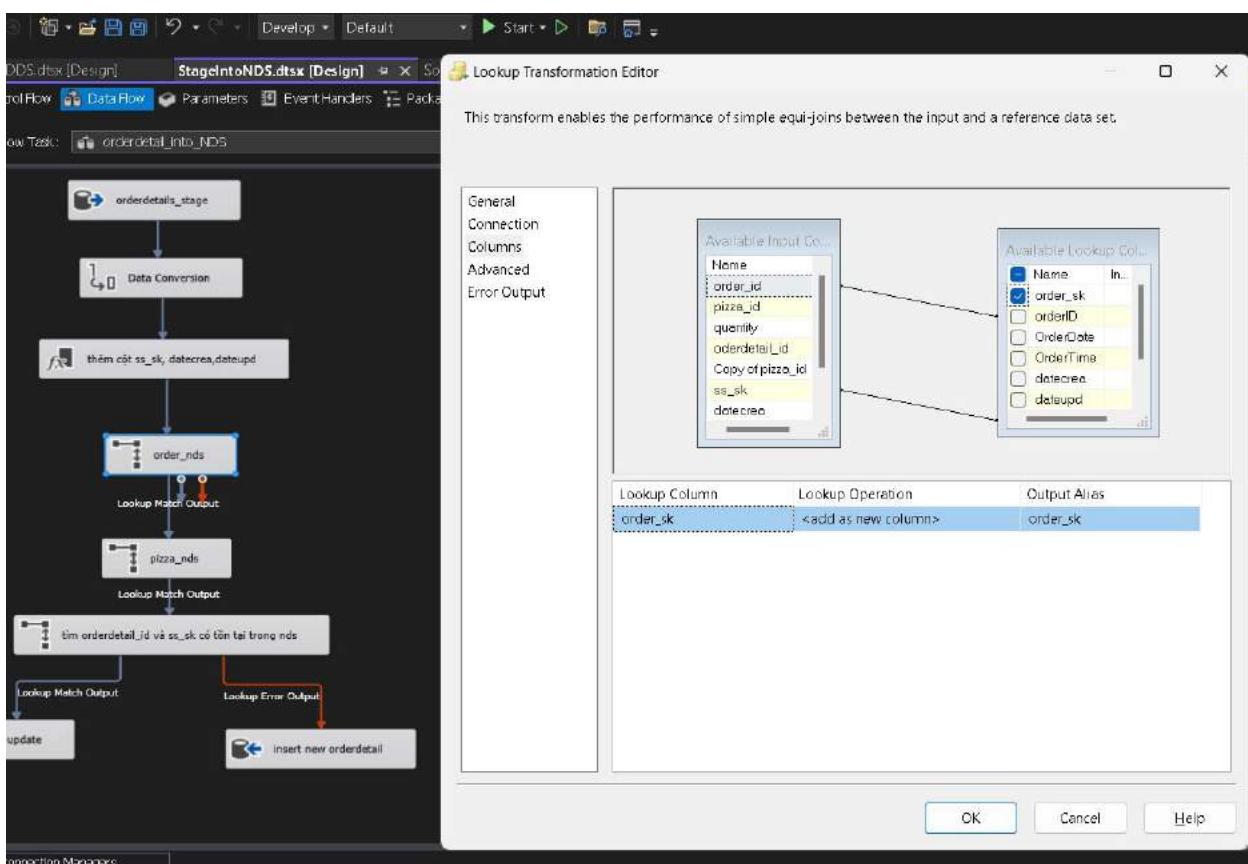
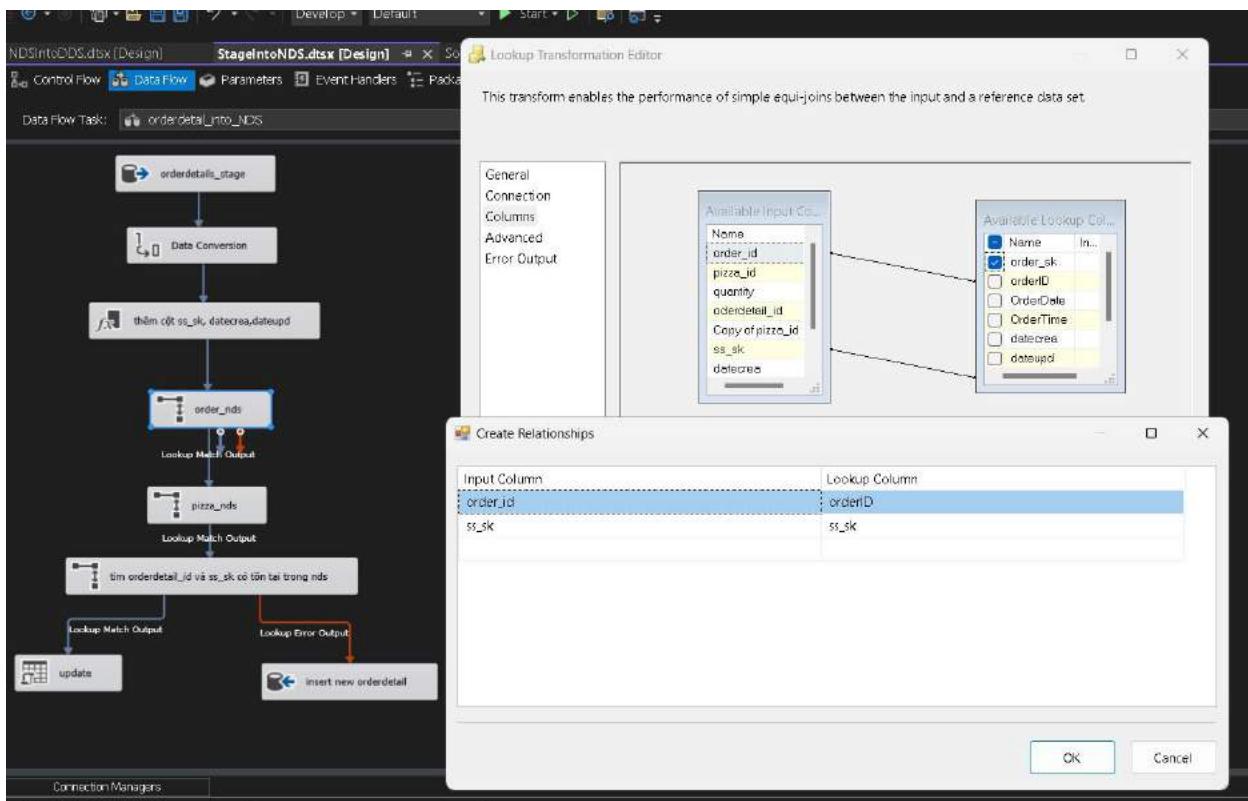




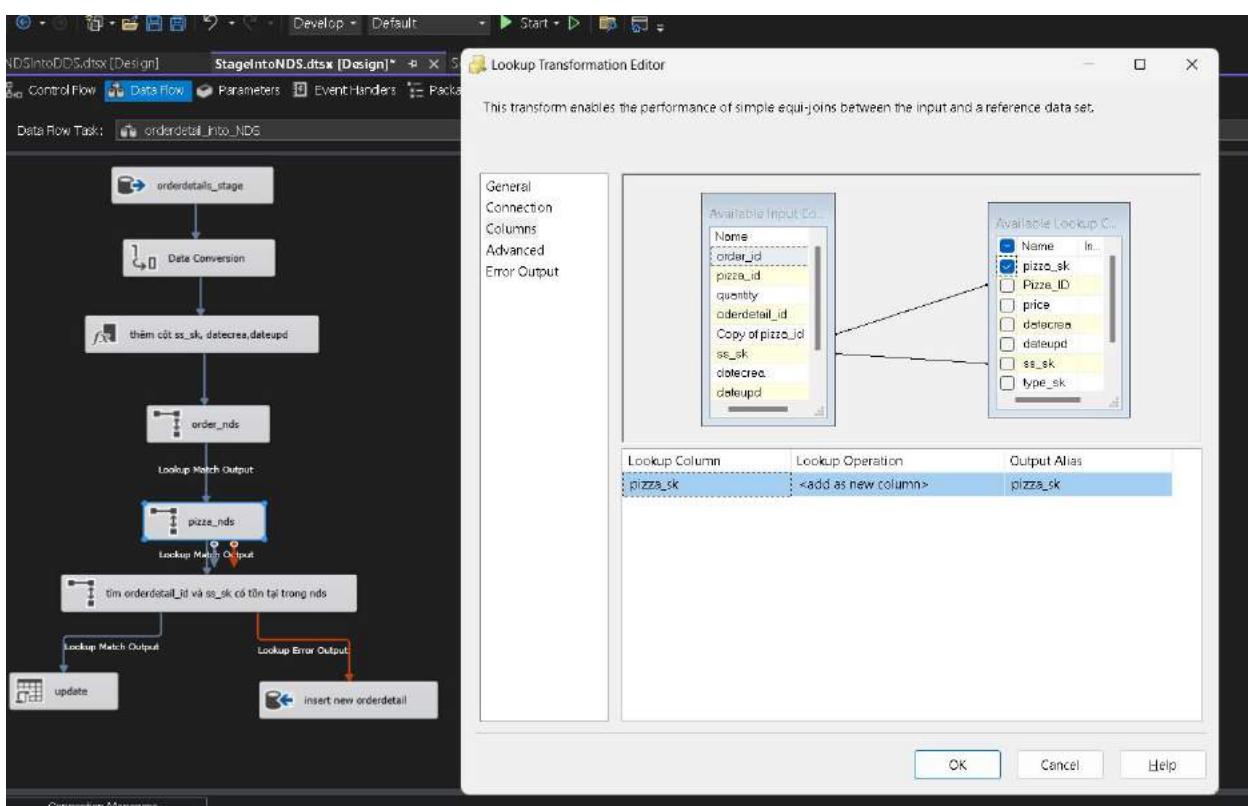
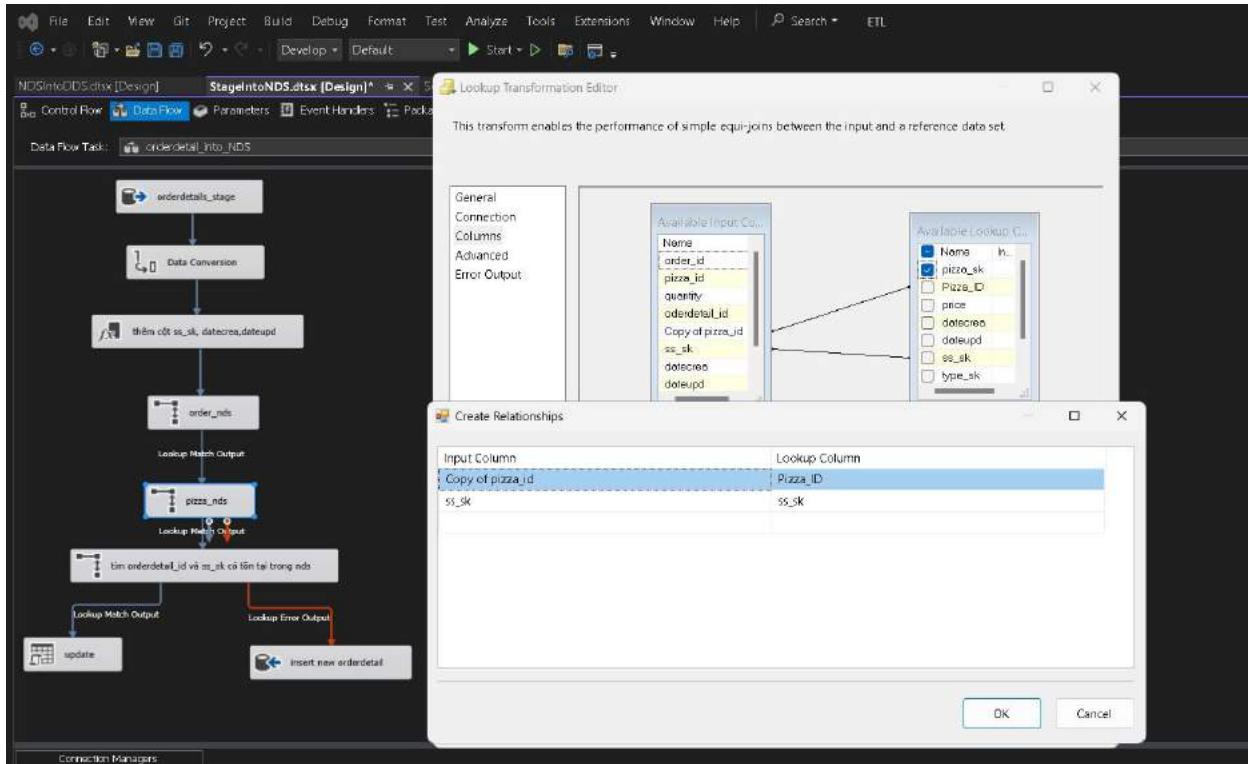
Tiếp theo orderdetails_stage vào orderdetail_nds, thì ngoài merge join thì mình có thể dùng look up để merge join



Đầu tiên tìm những dòng trong orderdetails_stage có orderid và sourceid có tồn tại trong order_nds và sau đó lấy order_sk của order_nds ra



Tương tự look up các dòng có pizzaid và sourceid tồn tại trong pizza_nds và lấy SK của pizza_nds ra



Và sau đó tìm orderdetail_id có tồn tại trong orderdetail_nds không

4. NDS vào DDS

Thiết kế DDS:

1. Thống kê số lượng order theo ngày, tháng , năm
2. Thống kê doanh thu theo từng loại pizza và theo từng tháng trong năm
3. Thống kê số lượng pizza đc order theo từng loại pizza và size

Sự kiện:

- Khi khách hàng order pizza

Bối cảnh sự kiện:

- Ai: Khách hàng
- Ở đâu: cửa hàng
- Cái gì: pizza
- Khi nào: ngày order

Đo lường: quantity, total

Fact table:

- Các giá trị có sẵn: quantity
- Các giá trị cần tính toán: total = quantity * price
- Độ mịn: 1 dòng trong bảng fact tương ứng với 1 hóa đơn có 1 pizza

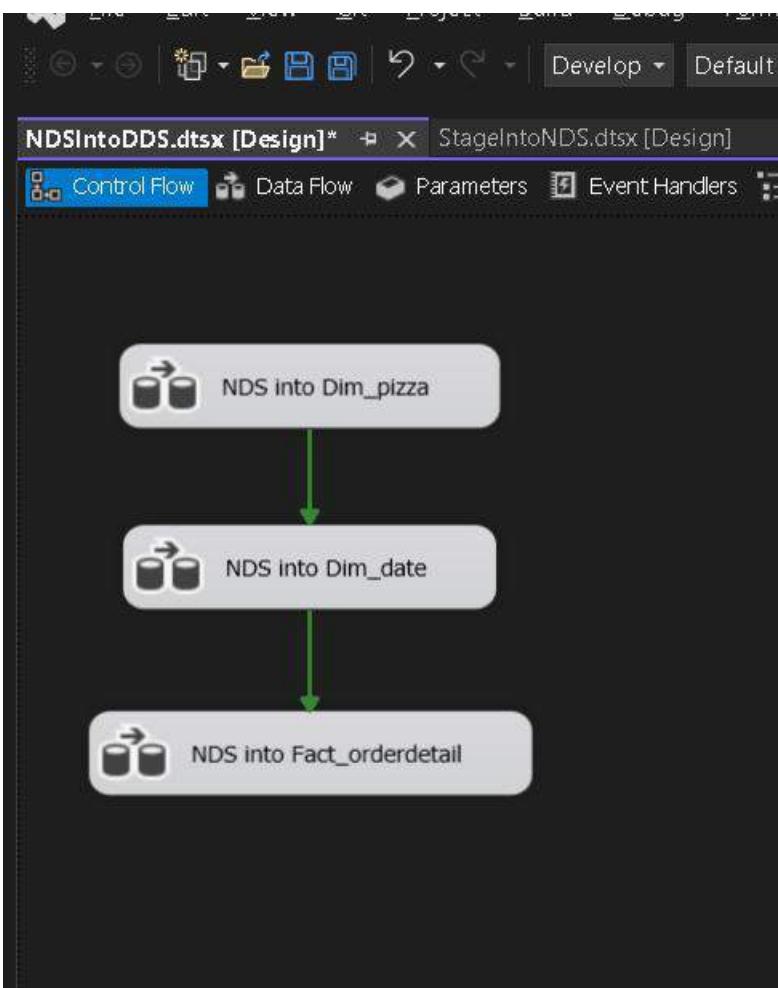
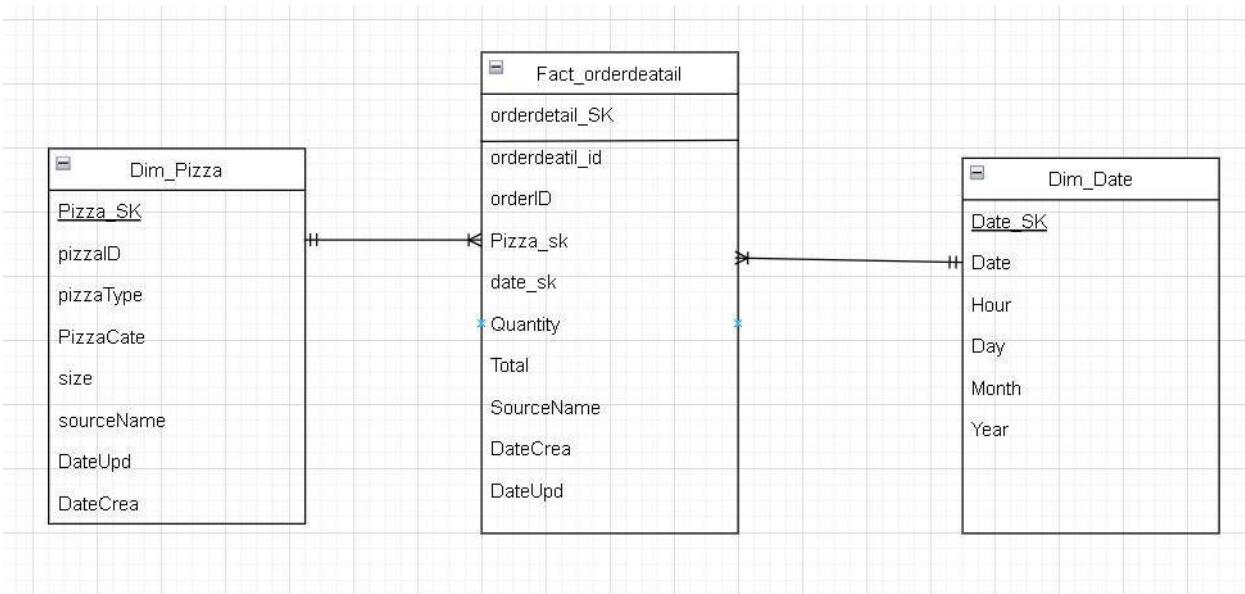
Dimention table:

- Dim_pizza: không cần lưu giá trị cũ -> SCD1: Ghi chép giá trị cũ

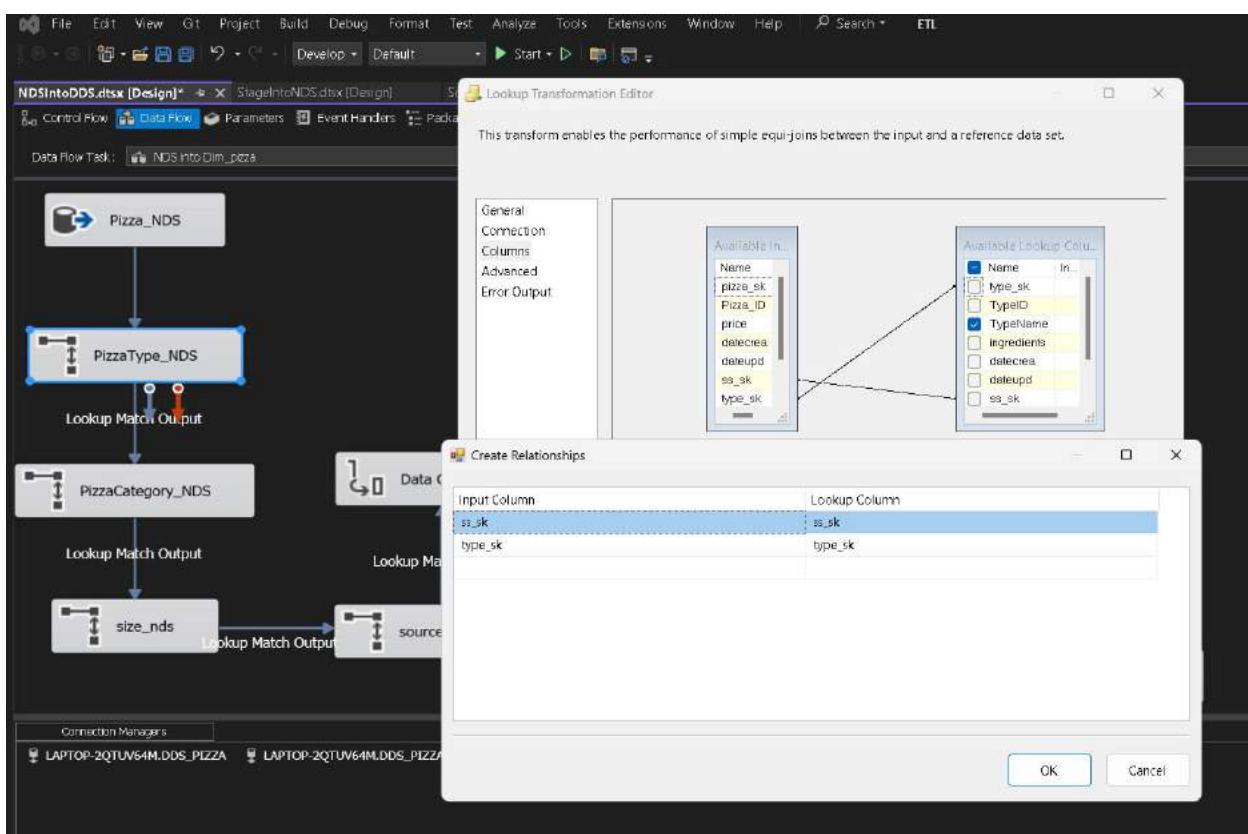
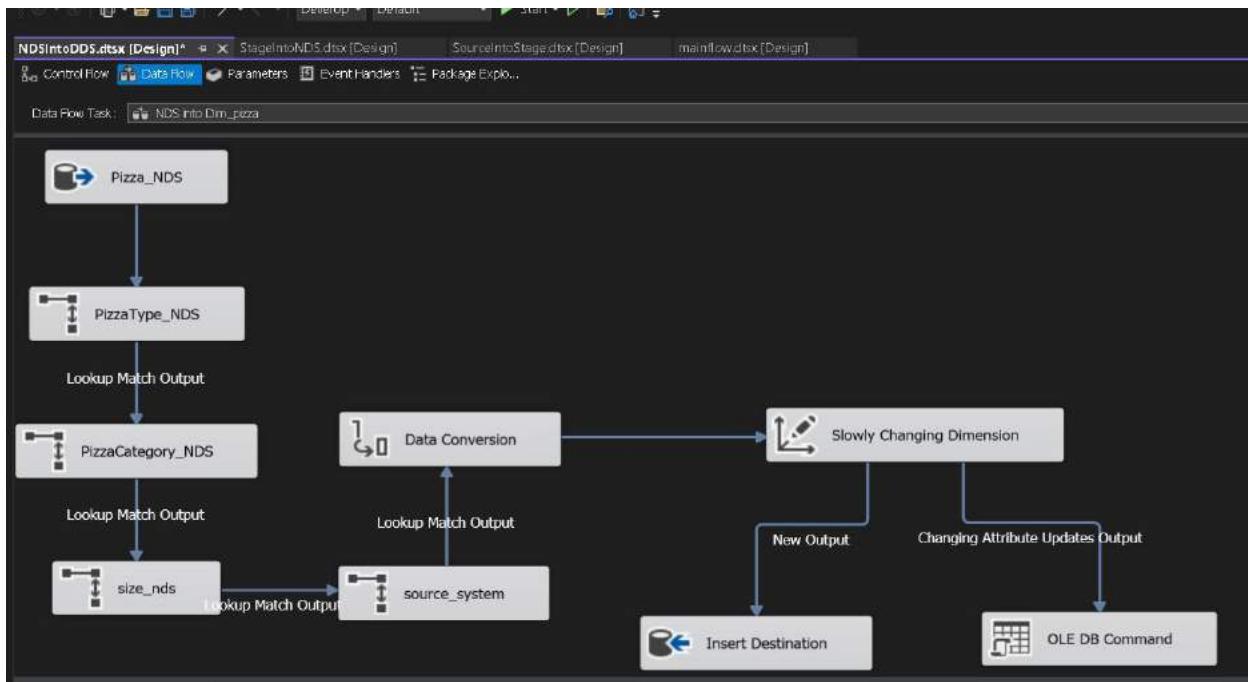
Phân chiêu:

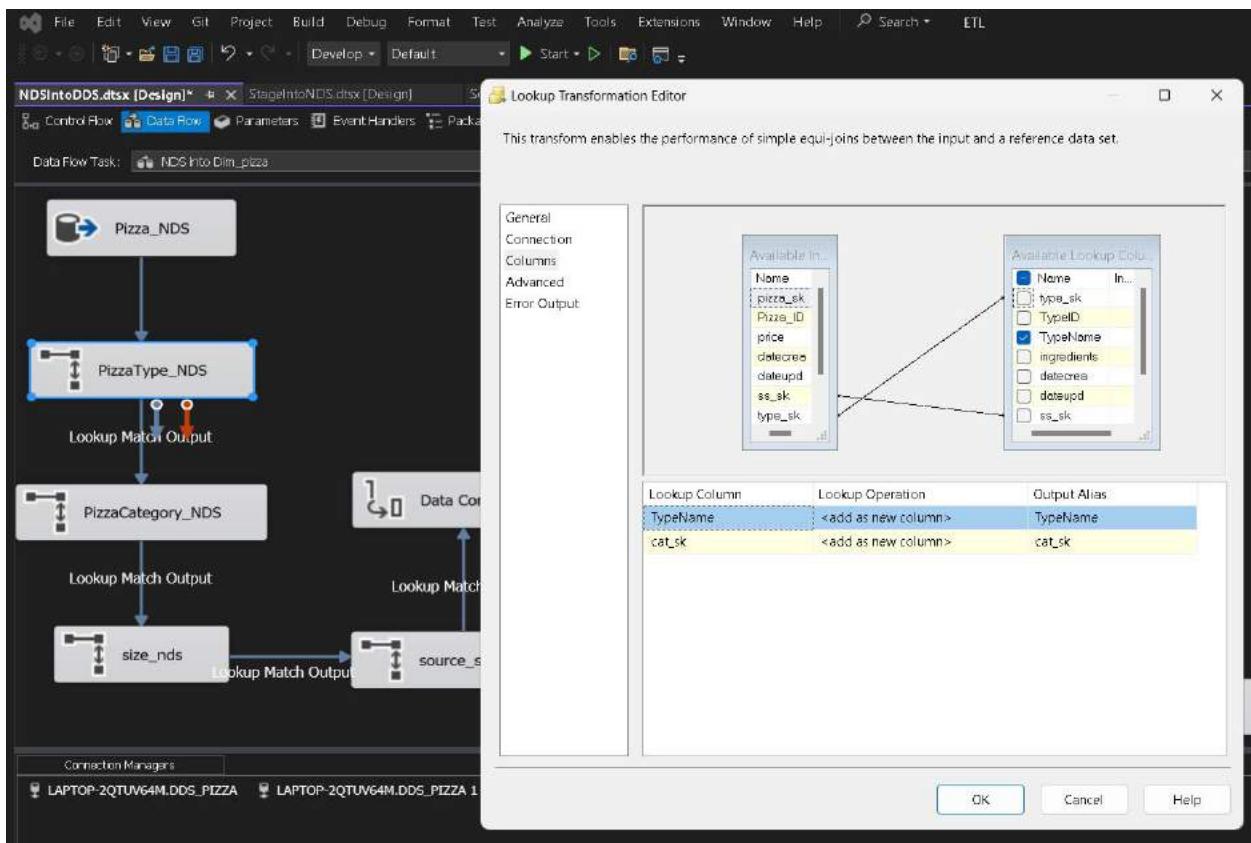
- Dim_date: năm -> tháng -> ngày -> giờ
- Dim_pizza: category -> type -> pizza

DDS: https://drive.google.com/file/d/1fCmhkA_GWHqdv_c1wfQqYgNCiW-2m-Jx/view?usp=sharing

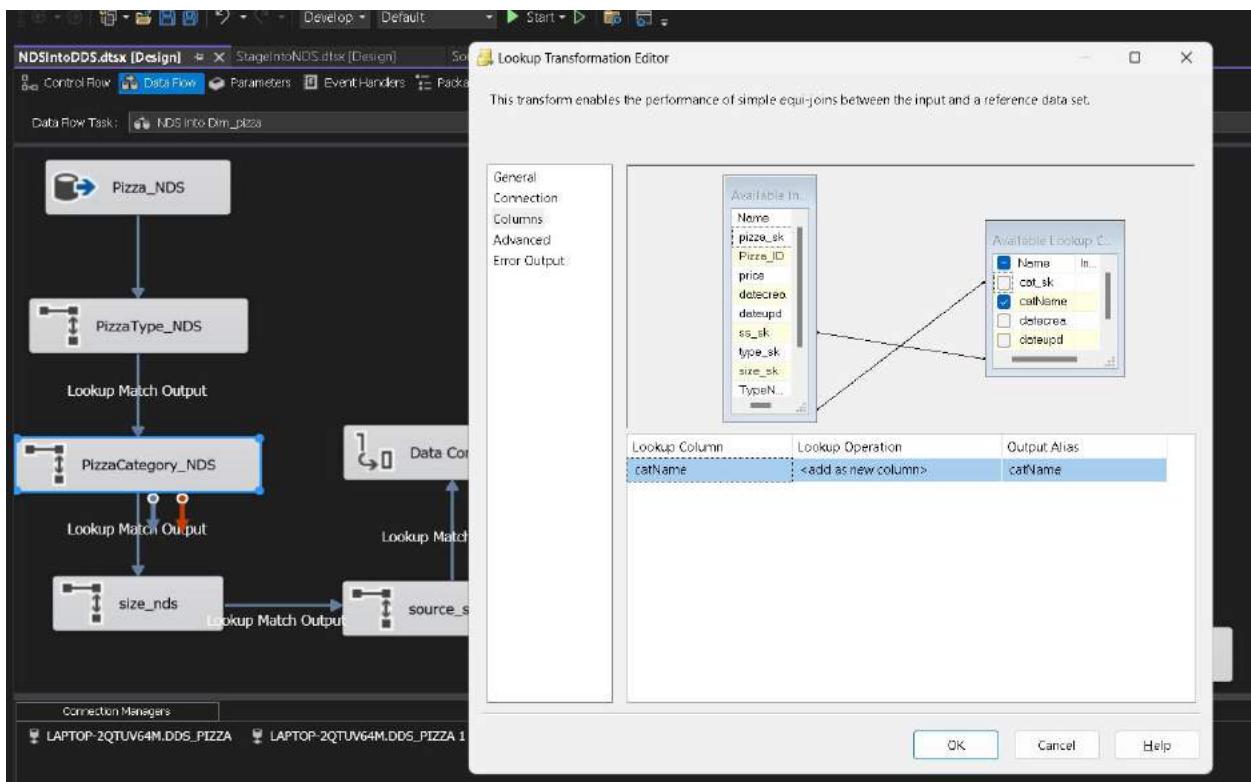


Đầu tiên pizza_nds vào Dim_pizza, lookup type_SK và sourceid có tồn tại trong pizzatype_nds và lấy typename, cat_sk của pizzatype_nds ra

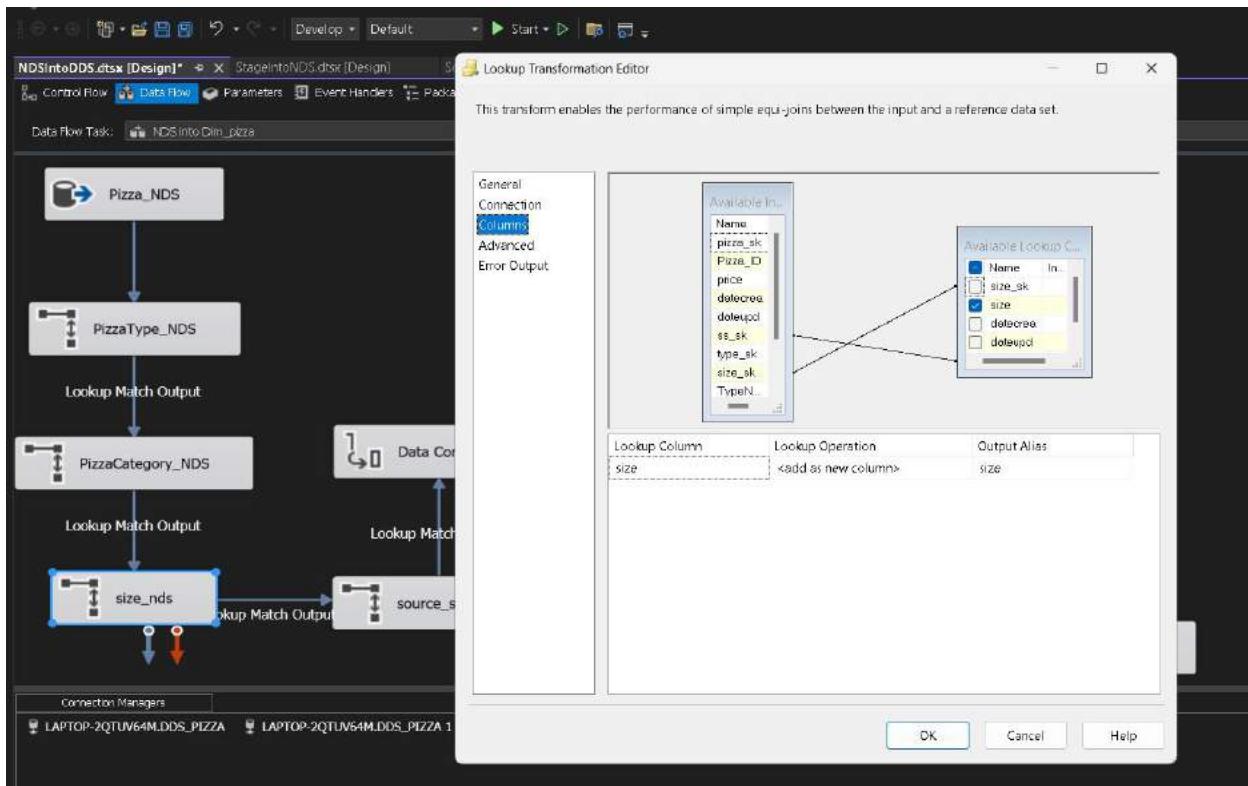




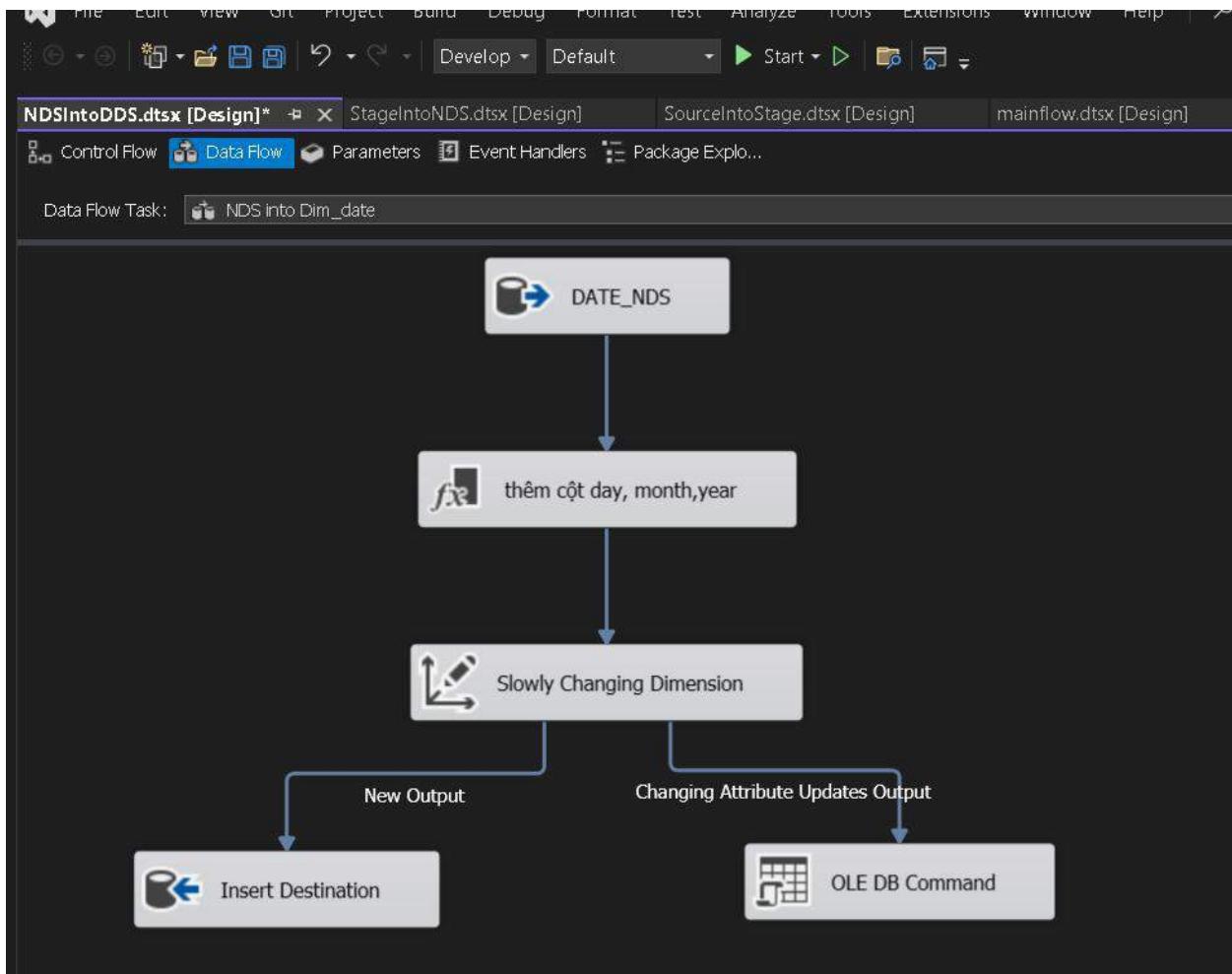
Sau khi có cat_sk thì tiếp tục loop up có cat_sk và sourceid có tồn tại trong cate_nds và lấy catname ra



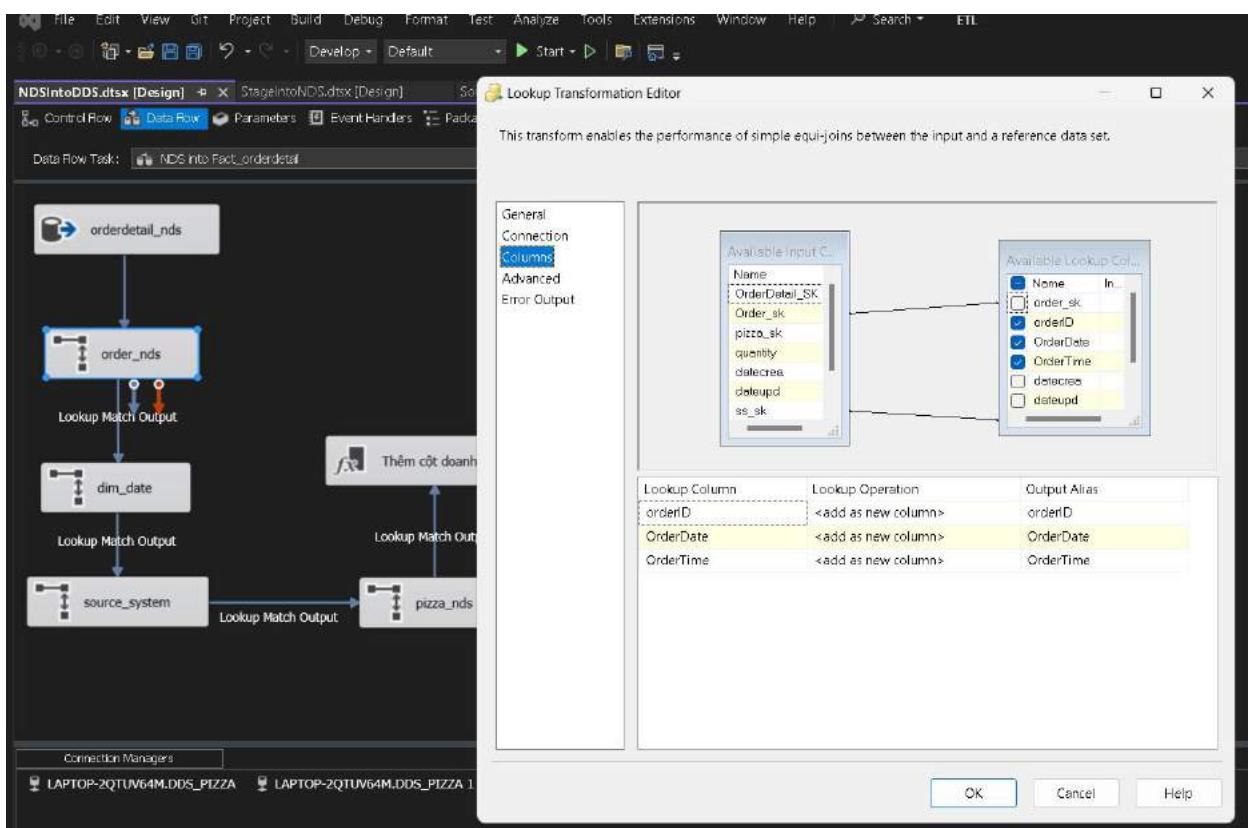
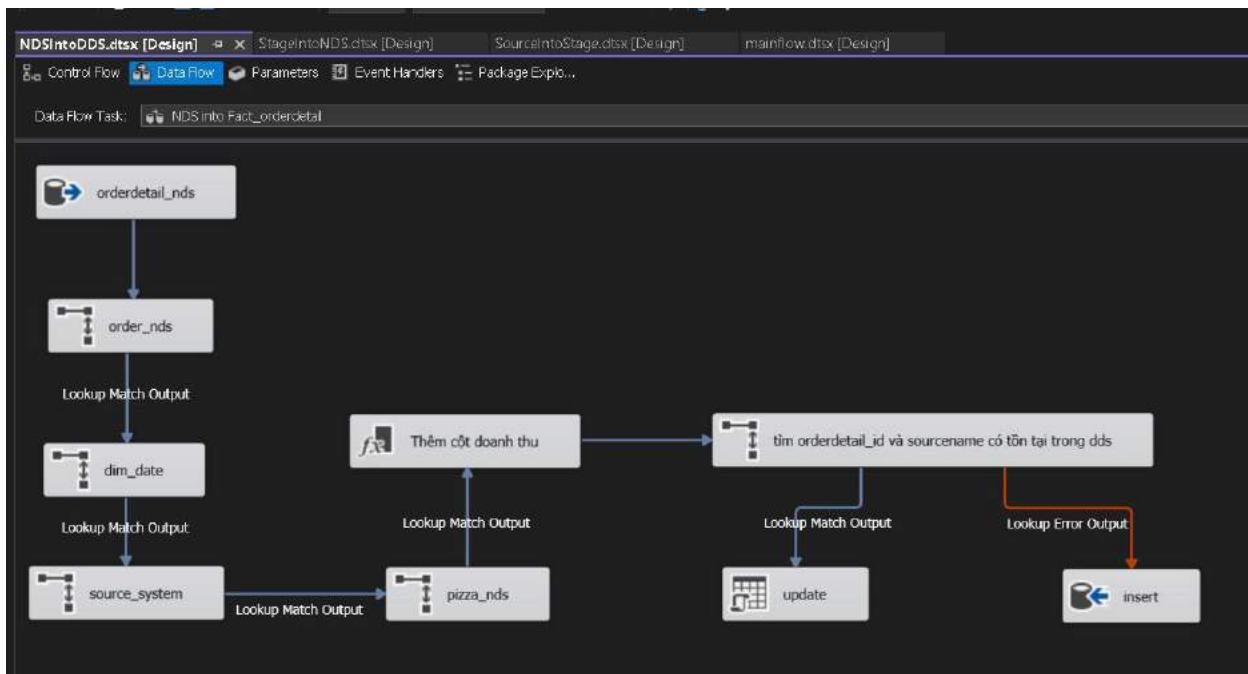
Tương tự loop up trong bảng size_nds



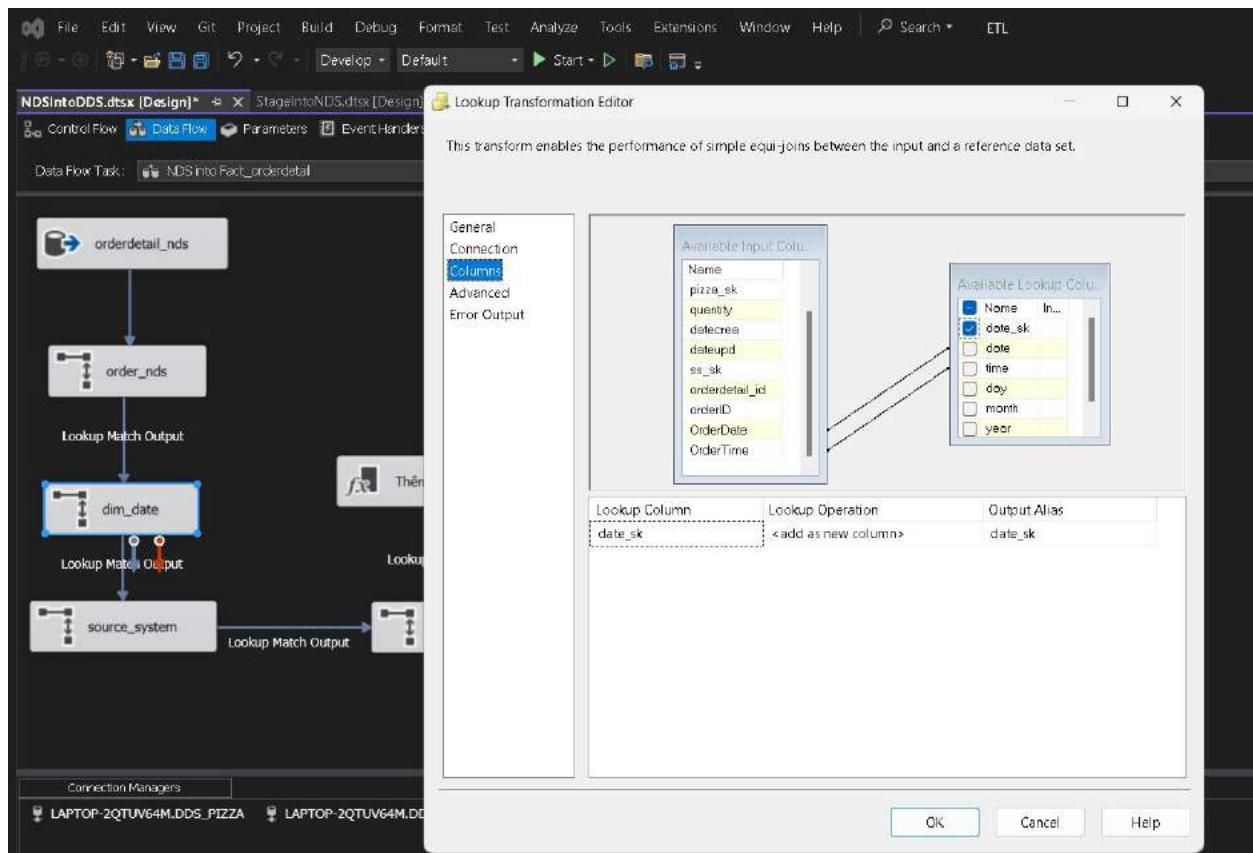
Tiếp theo date_nds vào dim_date tương tựj như supermarket sales



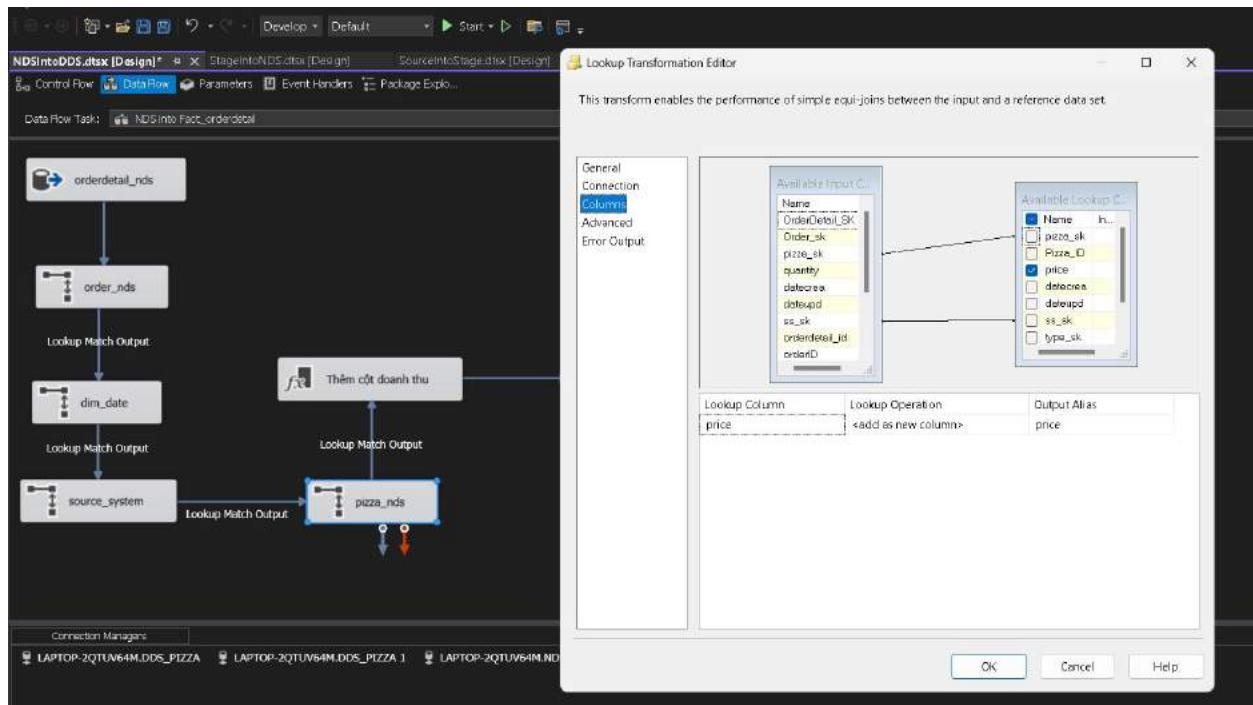
Tiếp theo xử lý bảng fact, orderdetails-nds loop up với order_nds theo order_sk và lấy orderid, orderdate, ordertime ra



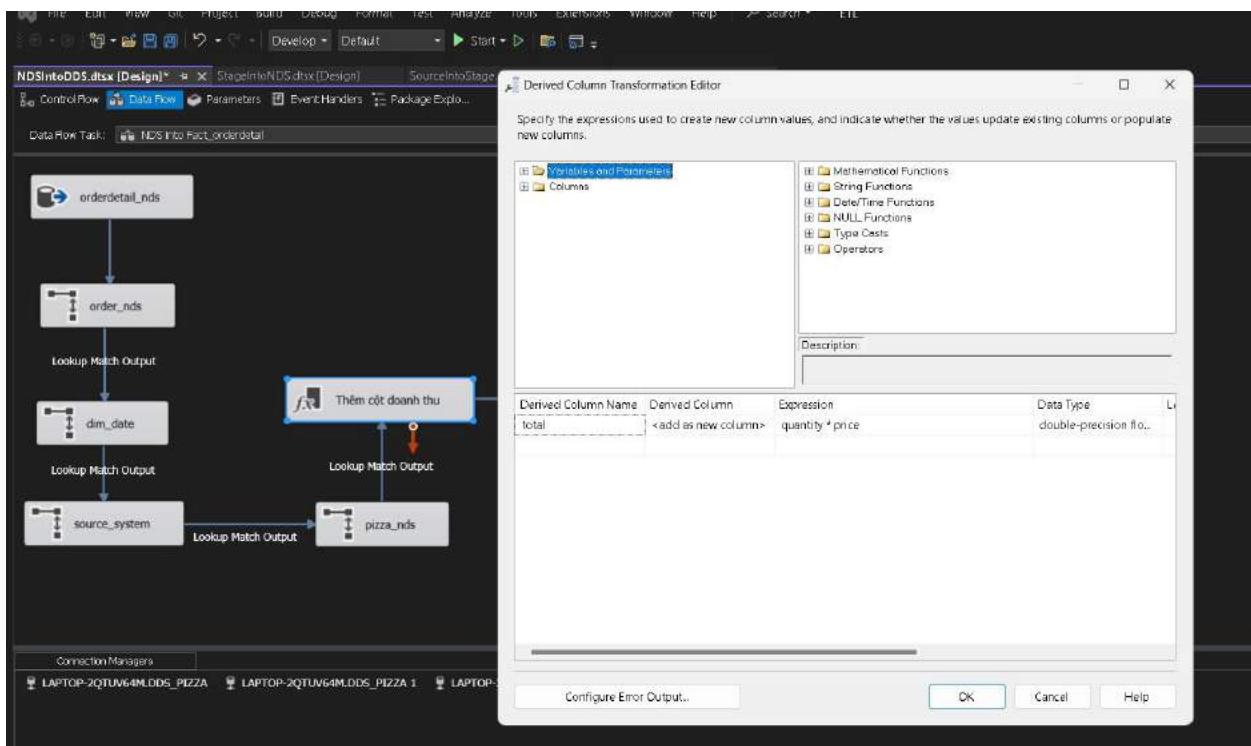
Lookup với bảng dim_date theo orderdate, ordertime và lấy datesk ra



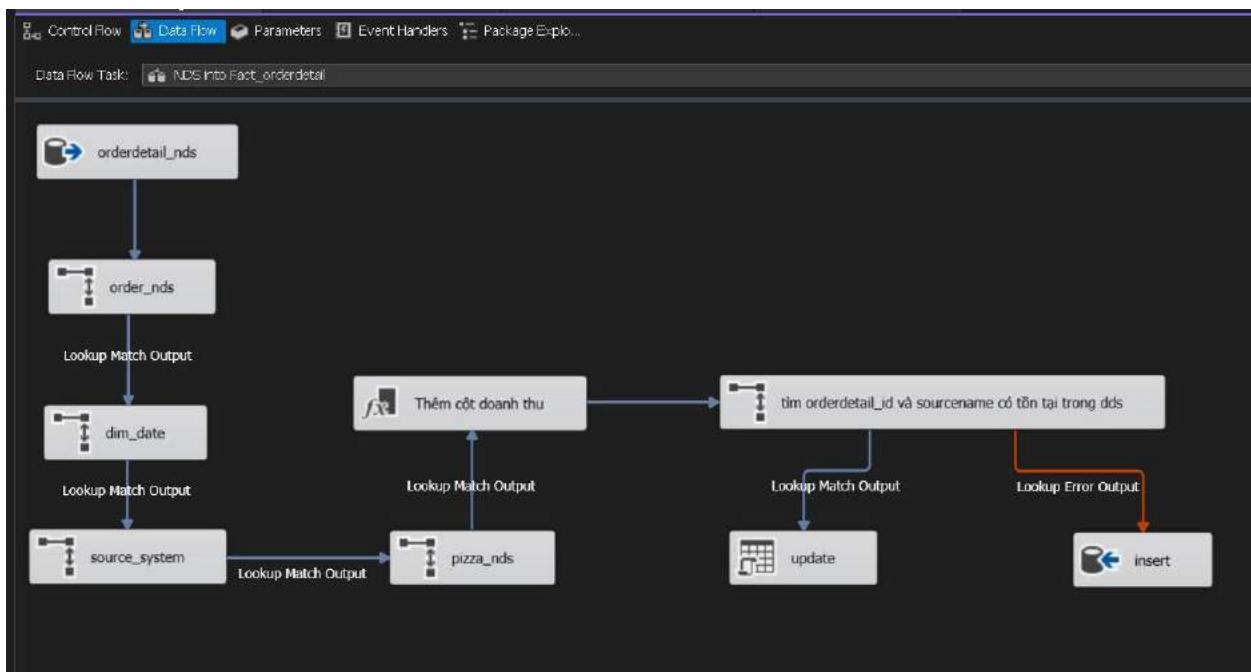
Loopup với pizza_nds theo pizza_sk để lấy price để tính doanh thu



Thêm cột total doanh thu

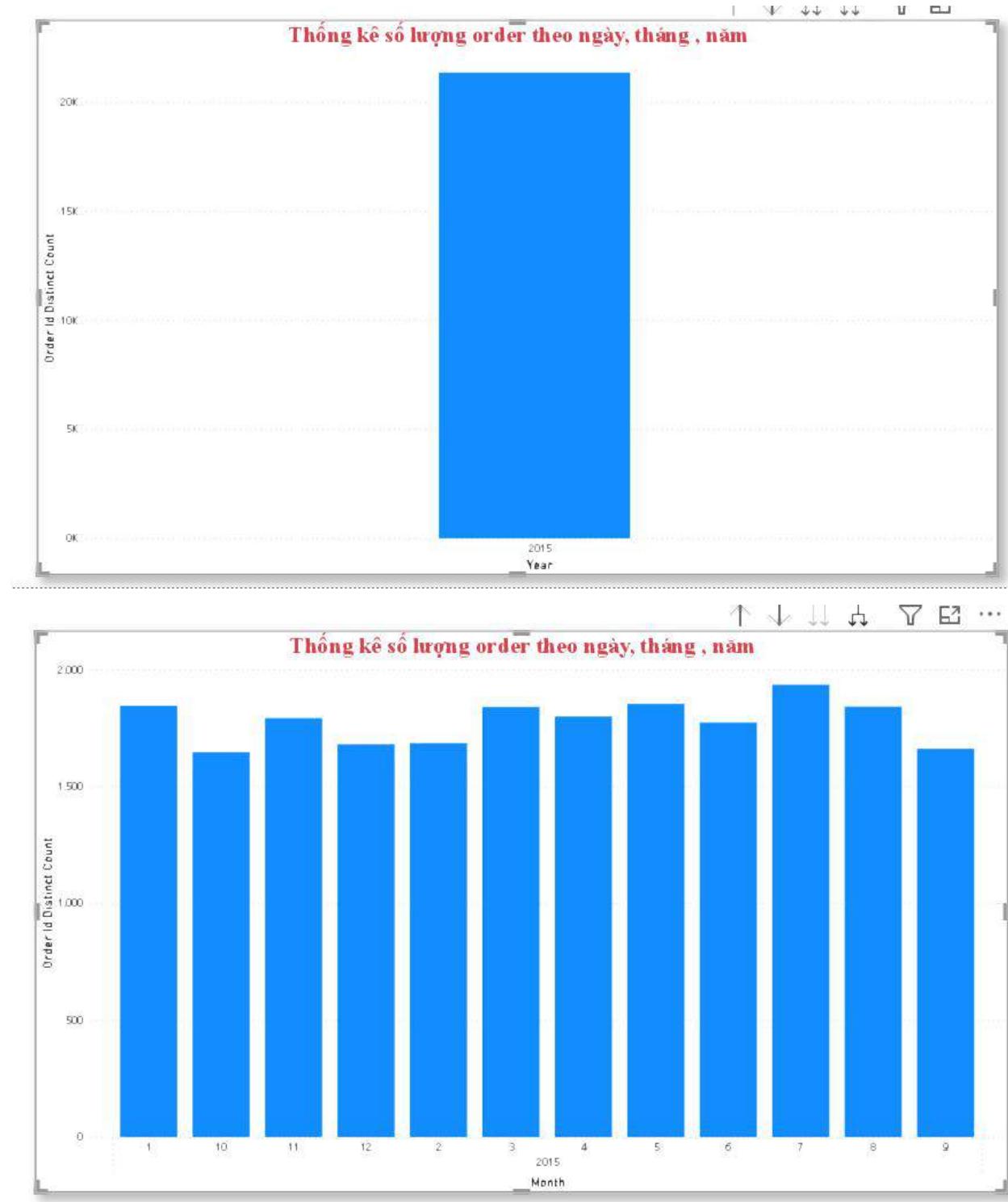


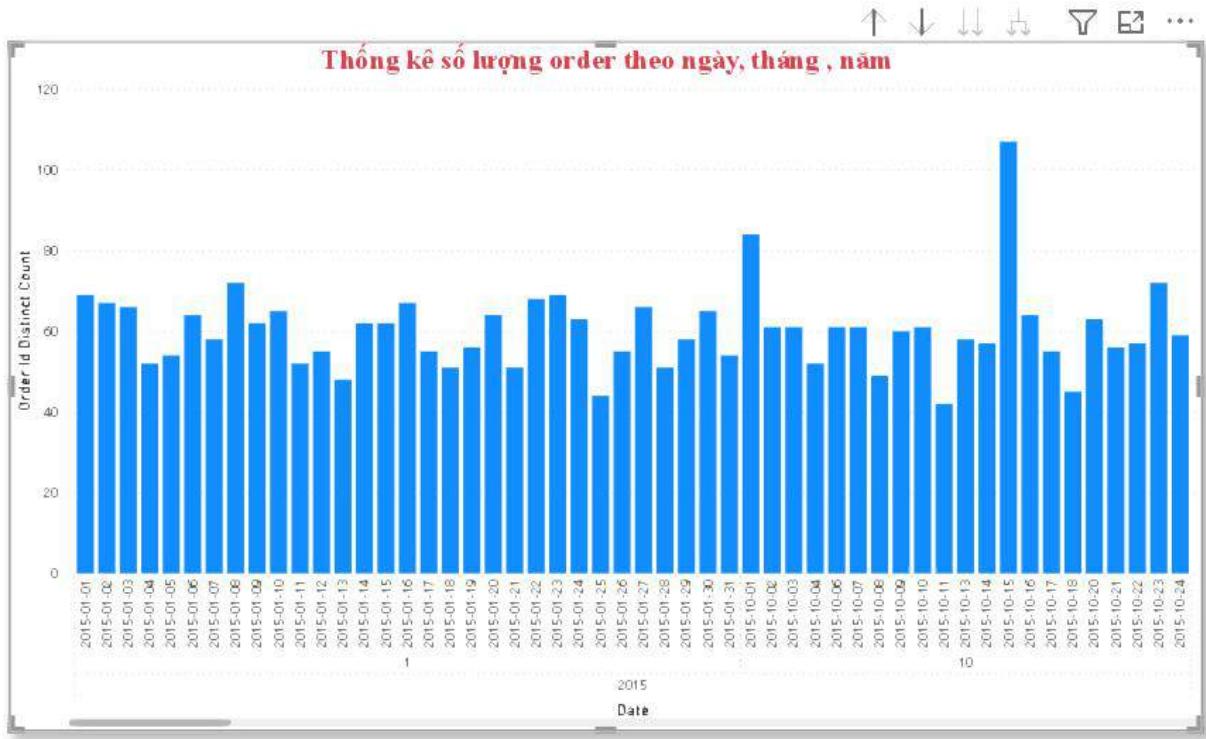
Và tương tự kiểm tra orderdetail_id có tồn tại trong bảng fact không



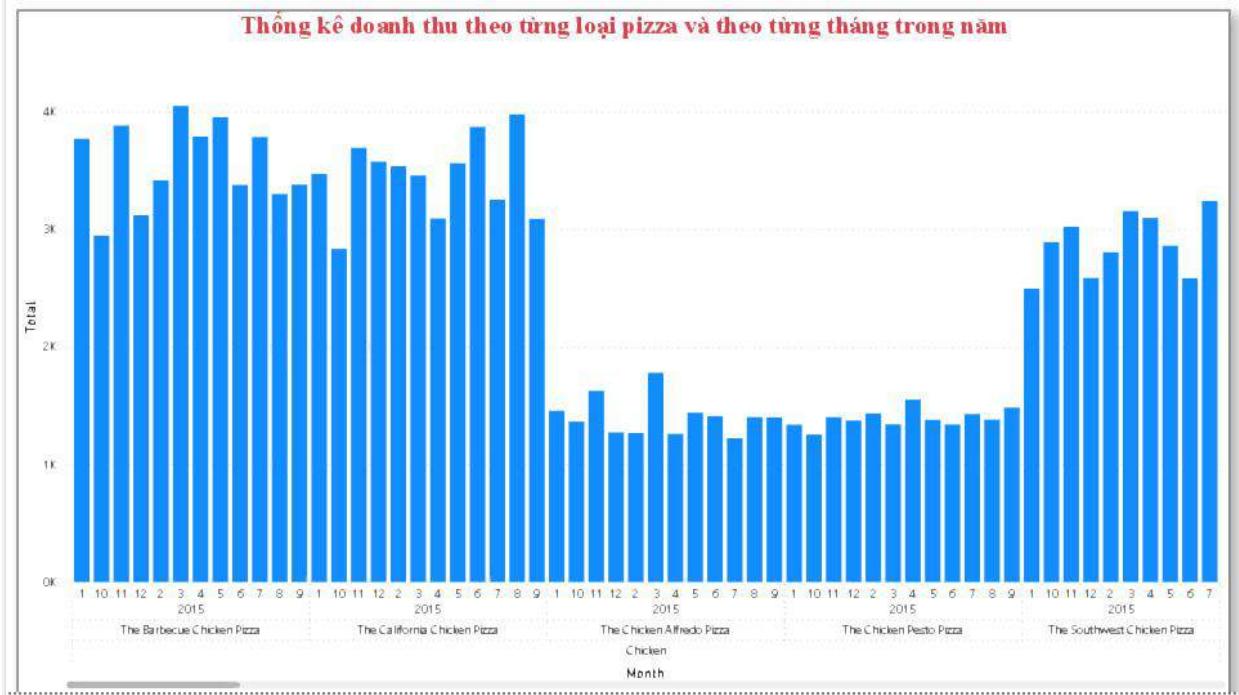
B) Report

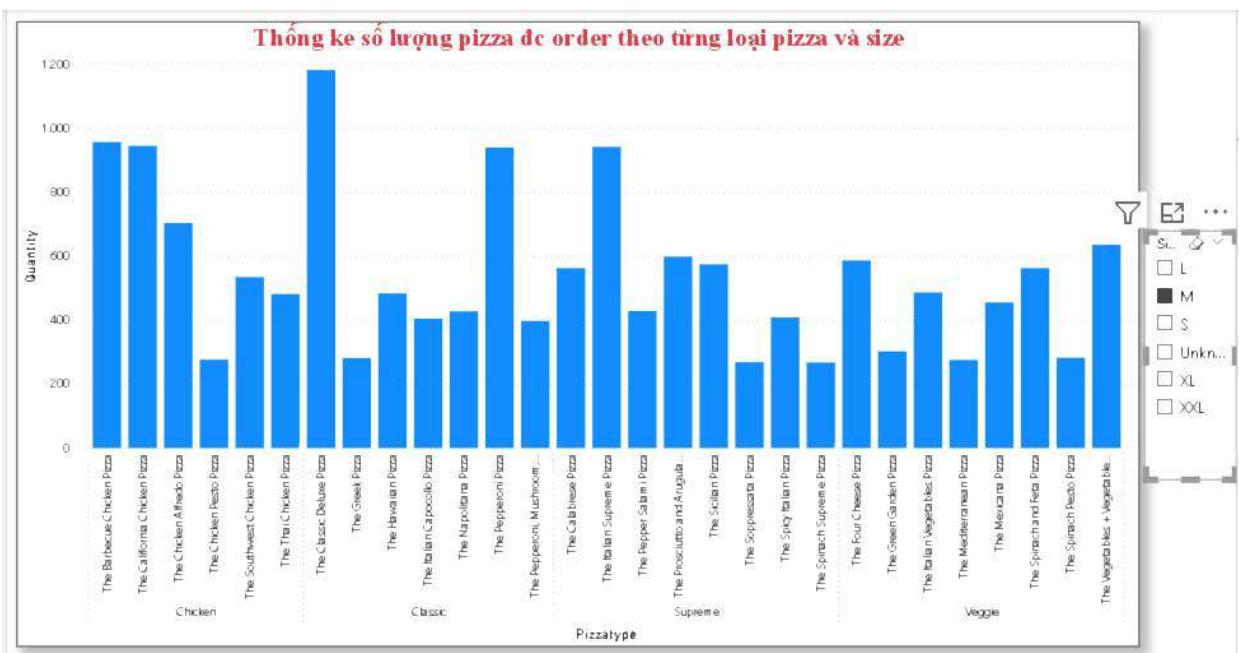
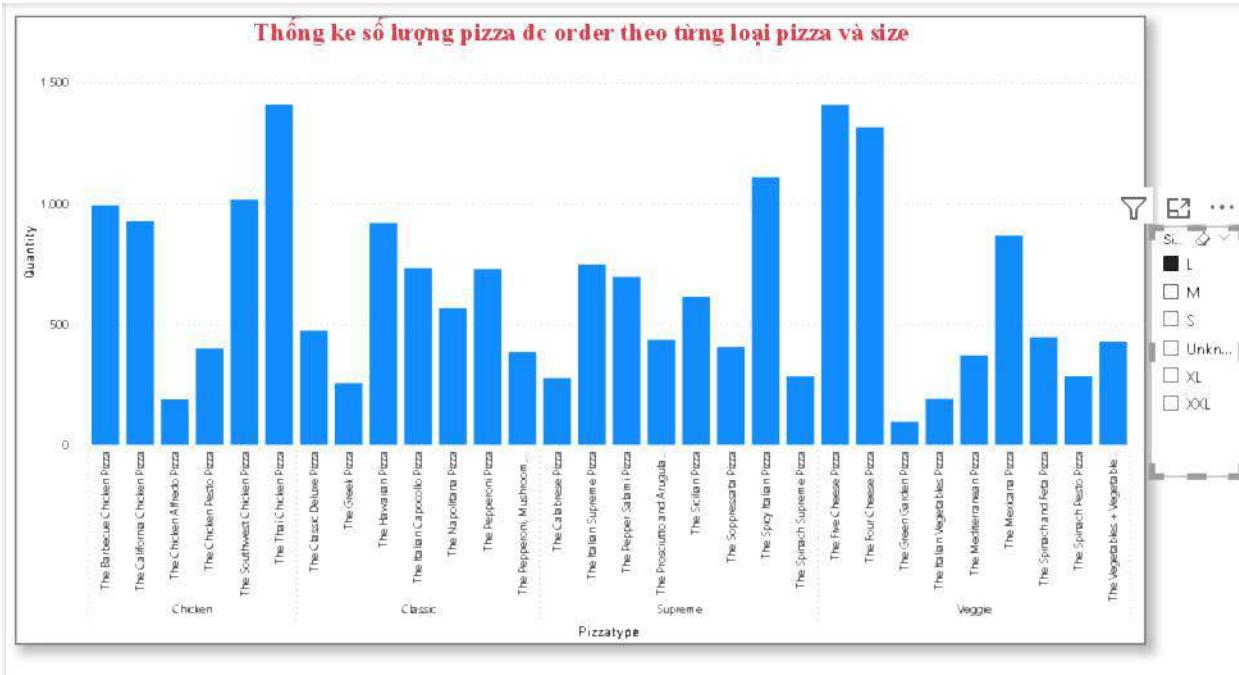
1. Thống kê số lượng order theo ngày, tháng , năm

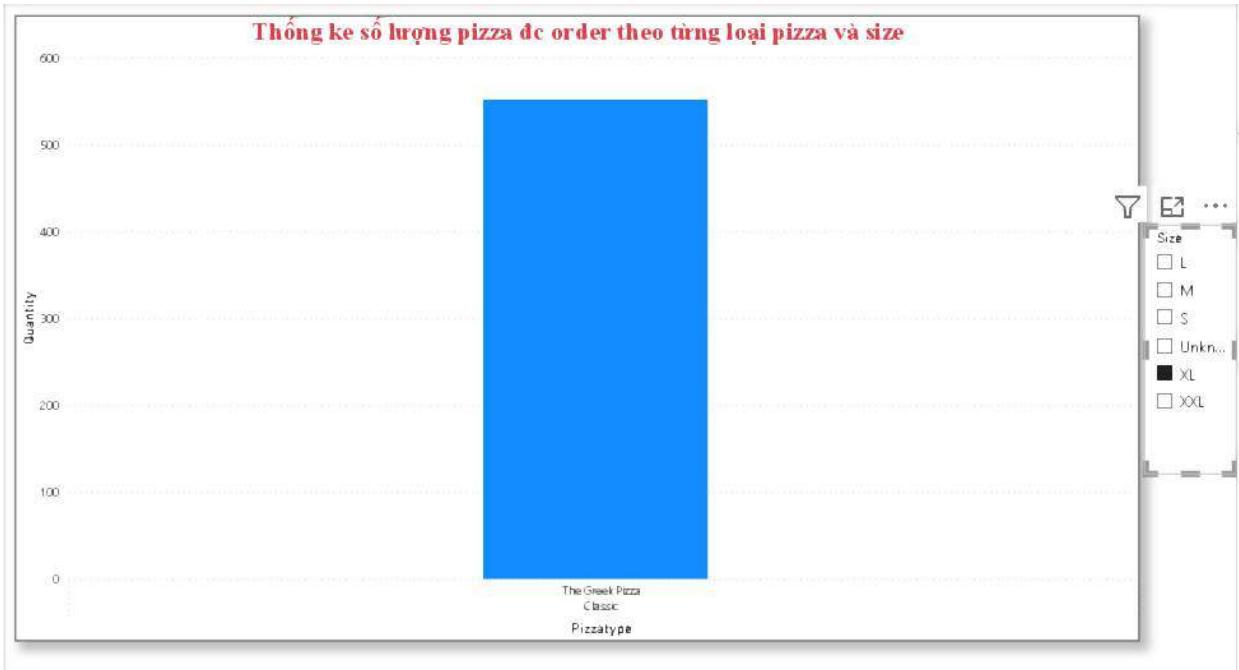




2. Thống kê doanh thu theo từng loại pizza và theo từng tháng trong năm







C) Olap, mdx

Phân chiêu bảng Dim_pizza

Phân chiêu bảng dim_date

The screenshot shows the Analysis Services Dimension Designer interface. The top menu bar has tabs for 'DIM DATE.dim [Design]', 'DIM PIZZA.dim [Design]', and 'DDS PIZZA.cube [Design]'. Below the menu are toolbars for 'Dimension Struct...', 'Attribute Relationships', 'Translations', and 'Browser'. The main area is divided into two sections: 'Attributes' on the left and 'Hierarchies' on the right. The 'Attributes' section lists attributes for the 'DIM DATE' dimension, including Date, Date Sk, Day, Month, Time, and Year. The 'Hierarchies' section shows a hierarchy structure with 'Year' at the top level, followed by 'Month', 'Date', and 'Time' levels, with '<new level>' below. A tooltip on the 'Hierarchy' button says: 'To create a new hierarchy, drag an attribute here.'

1. Total Sales by Pizza Category and Year

```
SELECT {[Measures].[Total]} ON 0,  
{[DIM DATE].[Year].[Year].Members * [DIM PIZZA].[Pizzacate].Members} ON 1  
FROM [DDS PIZZA]
```

The screenshot shows the Analysis Services Query Editor. The 'Results' tab is selected, displaying a table with five rows. The columns are labeled '2015', 'All', and 'Total'. The data shows total sales for each pizza category in 2015.

		Total
2015	All	817860.050000001
2015	Chicken	195919.5
2015	Classic	220053.1
2015	Supreme	208196.999999998
2015	Veggie	193690.450000003

2. Quantity Sold by Pizza Type for a Specific Month

```

SELECT {[Measures].[Quantity]} ON 0,
    ORDER (
        ORDER(
            [DIM DATE].[Month].[Month].Members * [DIM PIZZA].[Pizzatype].Members,
            [Measures].[Quantity],
            BDESC
        ),
        VAL([DIM DATE].[Month].CurrentMember.MemberValue),
        BASC
    )ON 1
FROM [DDS PIZZA];

```

		Quantity
1	All	4232
1	The Pepperoni Pizza	239
1	The Barbecue Chicken Pizza	211
1	The California Chicken Pizza	202
1	The Thai Chicken Pizza	199
1	The Sicilian Pizza	191
1	The Classic Deluxe Pizza	190
1	The Hawaiian Pizza	185
1	The Italian Supreme Pizza	167
1	The Four Cheese Pizza	159
1	The Spicy Italian Pizza	153
1	The Big Meat Pizza	150
1	The Vegetables + Vegetables Pizza	143
1	The Five Cheese Pizza	138
1	The Southwest Chicken Pizza	138
1	The Napolitana Pizza	136
		124

3. Top 5 Pizzas by Sales in a Specific Year

+ :: SQL

```
SELECT {[Measures].[Quantity],[Total]} ON 0,
    ORDER(
        TOPCOUNT([DIM PIZZA].[Pizzatype].[Pizzatype].MEMBERS,
            5,
            [Measures].[Total]
        ) * [DIM PIZZA].[Size].MEMBERS,
        [Measures].[Total],
        DESC
    ) ON 1
FROM [DDS PIZZA]
WHERE [DIM DATE].[Year].[2015];
```

108 %

Messages Results

		Quantity	Total
The Barbecue Chicken Pizza	All	2432	42768
The Barbecue Chicken Pizza	L	992	20584
The Barbecue Chicken Pizza	M	956	16013
The Barbecue Chicken Pizza	S	484	6171
The California Chicken Pizza	All	2370	41409.5
The California Chicken Pizza	L	927	19235.25
The California Chicken Pizza	M	944	15812
The California Chicken Pizza	S	499	6362.25
The Classic Deluxe Pizza	All	2453	38180.5
The Classic Deluxe Pizza	M	1181	18896
The Classic Deluxe Pizza	L	473	9696.5
The Classic Deluxe Pizza	S	799	9588
The Spicy Italian Pizza	All	1924	34831.25
The Spicy Italian Pizza	L	1109	23011.75
The Spicy Italian Pizza	M	408	6732
The Spicy Italian Pizza	S	407	5087.5

```
+ :: SQL Copy Caption
SELECT {[Measures].[Quantity],[Total]} ON 0,
    ORDER(
        TOPCOUNT([DIM PIZZA].[Pizzatype].[Pizzatype].MEMBERS,
            5,
            [Measures].[Total]
        ) * [DIM PIZZA].[Size].MEMBERS,
        [Measures].[Total],
        BDESC
    ) ON 1
FROM [DDS PIZZA]
WHERE [DIM DATE].[Year].[2015];
```

ANOTHER SORT

		Quantity	Total
The Thai Chicken Pizza	All	2371	43434.25
The Barbecue Chicken Pizza	All	2432	42768
The California Chicken Pizza	All	2370	41409.5
The Classic Deluxe Pizza	All	2453	38180.5
The Spicy Italian Pizza	All	1924	34831.25
The Thai Chicken Pizza	L	1410	29257.5
The Spicy Italian Pizza	L	1109	23011.75
The Barbecue Chicken Pizza	L	992	20584
The California Chicken Pizza	L	927	19235.25
The Classic Deluxe Pizza	M	1181	18896
The Barbecue Chicken Pizza	M	956	16013
The California Chicken Pizza	M	944	15812
The Classic Deluxe Pizza	L	473	9696.5
The Classic Deluxe Pizza	S	799	9588
The Thai Chicken Pizza	M	481	8056.75
The Spicy Italian Pizza	M	408	6732

4. Thống kê tổng số hóa đơn order theo ngày, tháng ,năm

```

//Thống kê tổng số hóa đơn order theo ngày, tháng, năm
WITH MEMBER [Measures].[OrderCount] AS [Measures].[FACT ORDERDETAILS Count]
SELECT {[Measures].[OrderCount]} ON 0,
{[DIM DATE].[Year].[Year] * [DIM DATE].[Month].[Month] * [DIM DATE].[Date].[Date]} ON 1
FROM [DDS PIZZA];

//Thống kê doanh thu theo từng loại pizza và theo từng tháng trong năm
SELECT {[Measures].[Total]} ON 0,
ORDER (
    ORDER(
        [DIM DATE].[Month].[Month].Members * [DIM PIZZA].[Pizzatype].Members,
        [Measures].[Quantity],
        BDESC
    ),
    VAL([DIM DATE].[Month].CurrentMember.MemberValue),
    BASC
)ON 1
FROM [DDS PIZZA];

```

100 %

			OrderCount
2015	1	2015-01-01	161
2015	1	2015-01-02	160
2015	1	2015-01-03	154
2015	1	2015-01-04	106
2015	1	2015-01-05	121
2015	1	2015-01-06	144
2015	1	2015-01-07	133
2015	1	2015-01-08	171
2015	1	2015-01-09	123
2015	1	2015-01-10	145
2015	1	2015-01-11	114
2015	1	2015-01-12	118
2015	1	2015-01-13	117
2015	1	2015-01-14	144
2015	1	2015-01-15	123
2015	1	2015-01-16	155

5. Thống kê doanh thu theo từng loại pizza và theo từng tháng trong năm

```

//Thống kê doanh thu theo từng loại pizza và theo từng tháng trong năm
SELECT {[Measures].[Total]} ON 0,
    ORDER (
        ORDER(
            [DIM DATE].[Month].[Month].Members * [DIM PIZZA].[Pizzatype].Members,
            [Measures].[Quantity],
            BDESC
        ),
        VAL([DIM DATE].[Month].CurrentMember.MemberValue),
        BASC
    )ON 1
FROM [DDS PIZZA];

//Thống kê số lượng pizza được order theo từng loại pizza và size
SELECT {[Measures].[Quantity],[Total]} ON 0,
    ORDER(
        ORDER(
            [DIM PIZZA].[Pizzatype].Members,
            [Measures].[Quantity],
            BDESC
        ),
        VAL([DIM PIZZA].[Size].CurrentMember.MemberValue),
        BASC
    )ON 1
FROM [DDS PIZZA];

```

100 %

		Total
Unknown	All	(null)
Unknown	The Barbecue Chicken Pizza	(null)
Unknown	The Big Meat Pizza	(null)
Unknown	The Brie Carre Pizza	(null)
Unknown	The Calabrese Pizza	(null)
Unknown	The California Chicken Pizza	(null)
Unknown	The Chicken Alfredo Pizza	(null)
Unknown	The Chicken Pesto Pizza	(null)
Unknown	The Classic Deluxe Pizza	(null)
Unknown	The Five Cheese Pizza	(null)
Unknown	The Four Cheese Pizza	(null)
Unknown	The Greek Pizza	(null)
Unknown	The Green Garden Pizza	(null)
Unknown	The Hawaiian Pizza	(null)
Unknown	The Italian Capocollo Pizza	(null)
Unknown	The Italian Supreme Pizza	(null)

6. Thống kê số lượng pizza được order theo từng loại pizza và size

```

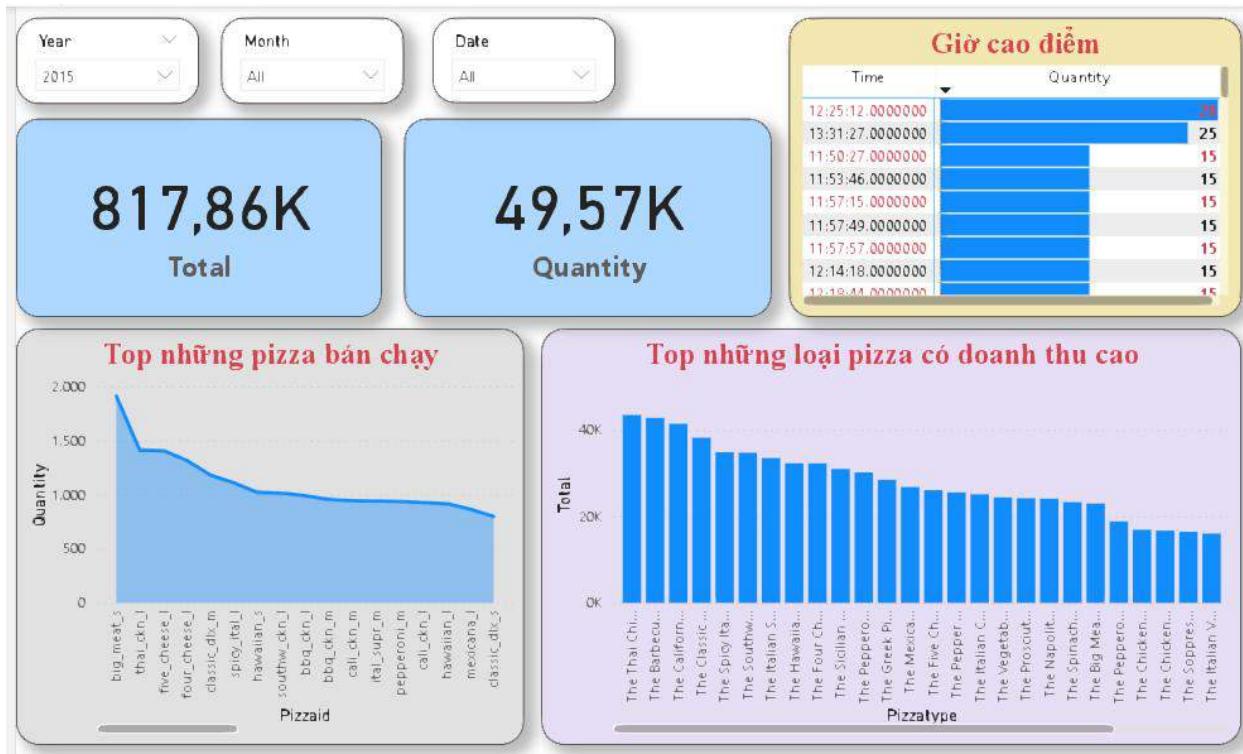
//Thống kê số lượng pizza được order theo từng loại pizza và size
SELECT {[Measures].[Quantity],[Total]} ON 0,
       ORDER(
           TOPCOUNT([DIM PIZZA].[Pizzatype].[Pizzatype].MEMBERS,
           5,
           [Measures].[Total]
           ) * [DIM PIZZA].[Size].MEMBERS,
           [Measures].[Total],
           BDESC
       ) ON 1
FROM [DDS PIZZA];

```

100 %

		Quantity	Total
The Thai Chicken Pizza	All	2371	43434.25
The Barbecue Chicken Pizza	All	2432	42768
The California Chicken Pizza	All	2370	41409.5
The Classic Deluxe Pizza	All	2453	38180.5
The Spicy Italian Pizza	All	1924	34831.25
The Thai Chicken Pizza	L	1410	29257.5
The Spicy Italian Pizza	L	1109	23011.75
The Barbecue Chicken Pizza	L	992	20584
The California Chicken Pizza	L	927	19235.25
The Classic Deluxe Pizza	M	1181	18896
The Barbecue Chicken Pizza	M	956	16013
The California Chicken Pizza	M	944	15812
The Classic Deluxe Pizza	L	473	9696.5
The Classic Deluxe Pizza	S	799	9588
The Thai Chicken Pizza	M	481	8056.75
The Spicy Italian Pizza	M	408	6732
		100	80000.00

D) Dashboard



E) Mining

	Chicken	Classic	Supreme	Veggie	S	M	L	Total > 5
0	1	0	0	0	1	0		0
0	1	0	0	0	1	0		0
0	0	0	1	0	1	0		0
1	0	0	0	0	0	1		0
0	0	1	0	0	0	1		0
0	0	0	1	0	0	0	1	0
0	0	1	0	0	0	0	1	0
0	0	1	0	0	1	0		0
0	0	1	0	0	0	1		0
1	0	0	0	1	0	0		0
0	1	0	0	1	0	0		0
0	0	1	0	1	0	0		0
0	0	1	0	1	0	0		0
0	0	0	1	0	0	1		1
0	1	0	0	0	0	0	1	1
0	0	0	1	1	0	0		1
0	0	0	1	1	0	0		1
0	0	1	0	0	0	1		1
0	0	1	0	0	0	0	1	1
0	1	0	0	1	0	0		1
0	0	1	0	1	0	0		1
0	0	0	1	1	0	0		1
1	0	0	0	0	0	0	1	0
0	0	0	1	0	0	0	1	0

Tương tự trước hết ta dùng one hot encoding để chuyển thông tin bán hàng thành dạng nhị phân trong đó một số thông tin đã được lược bỏ để tìm ra sự ảnh hưởng của loại bánh lên số lượng bánh mua

Bảng tần suất các thuộc tính

Feature Frequency Lookup Table

Feature Index	Feature	Frequency
1	Chicken	22.24%
2	Classic	29.99%
3	Supreme	24.22%
4	Veggie	23.55%
5	S	29.08%
6	M	31.64%
7	L	38.10%
8	Total > 5	12.06%

Bảng 2 way lift

2-Way Lift							
LHS	RHS	Occurance	LHS Freq	RHS Freq	Trans	Predicted Trans	Lift
Chicken	Total > 5	1306	0.222439325	0.120608803	48620	1304.384204	1.001239
Classic	Total > 5	1719	0.299856026	0.120608803	48620	1758.355738	0.977618
Supreme	Total > 5	1449	0.242225422	0.120608803	48620	1420.409872	1.020128
Veggie	Total > 5	1390	0.235479227	0.120608803	48620	1380.850185	1.006626
S	Total > 5	1674	0.290765117	0.120608803	48620	1705.046647	0.981791
M	Total > 5	1852	0.316433566	0.120608803	48620	1855.566434	0.998078
L	Total > 5	0	0.38103661	0.120608803	48620	2234.398684	0

Có thể thấy khách hàng mua bánh size L 100% khả năng sẽ mua ít hơn 5 bánh

Vẫn có thể mở rộng ra cho nhiều biến hơn

3-Way Lift							
LHS		RHS					
Feature 1 Index	Feature 2 Index						
4	5						
Feature 1	Feature 2	Feature 3	Occurrences Freq	LHS Freq	RHS Freq	Lift	
Veggie	S	Total > 5	0.006622789	0.05418	0.12061	1.013587	

Nhưng khả năng cao loại bánh không có ảnh hưởng đến số lượng bánh mua của khách hàng