



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM

Khoa Công nghệ Thông tin

ĐỒ ÁN THỰC HÀNH

BUILDING AND MINING DATA WAREHOUSE

Student: 21127702 - Bùi Nguyễn Tin
19127407 - Nguyễn Huy Hoàng
20127189 - Nguyễn Quốc Huy
20127204 - Nguyễn Phụng Khanh

Course: CSC12107 - 21HTTT2

Teacher: Hồ Thị Hoàng Vy
Tiết Gia Hồng
Nguyễn Ngọc Minh Châu
Lê Nguyễn Hoài Nam

Thành phố Hồ Chí Minh – 2024

MỤC LỤC

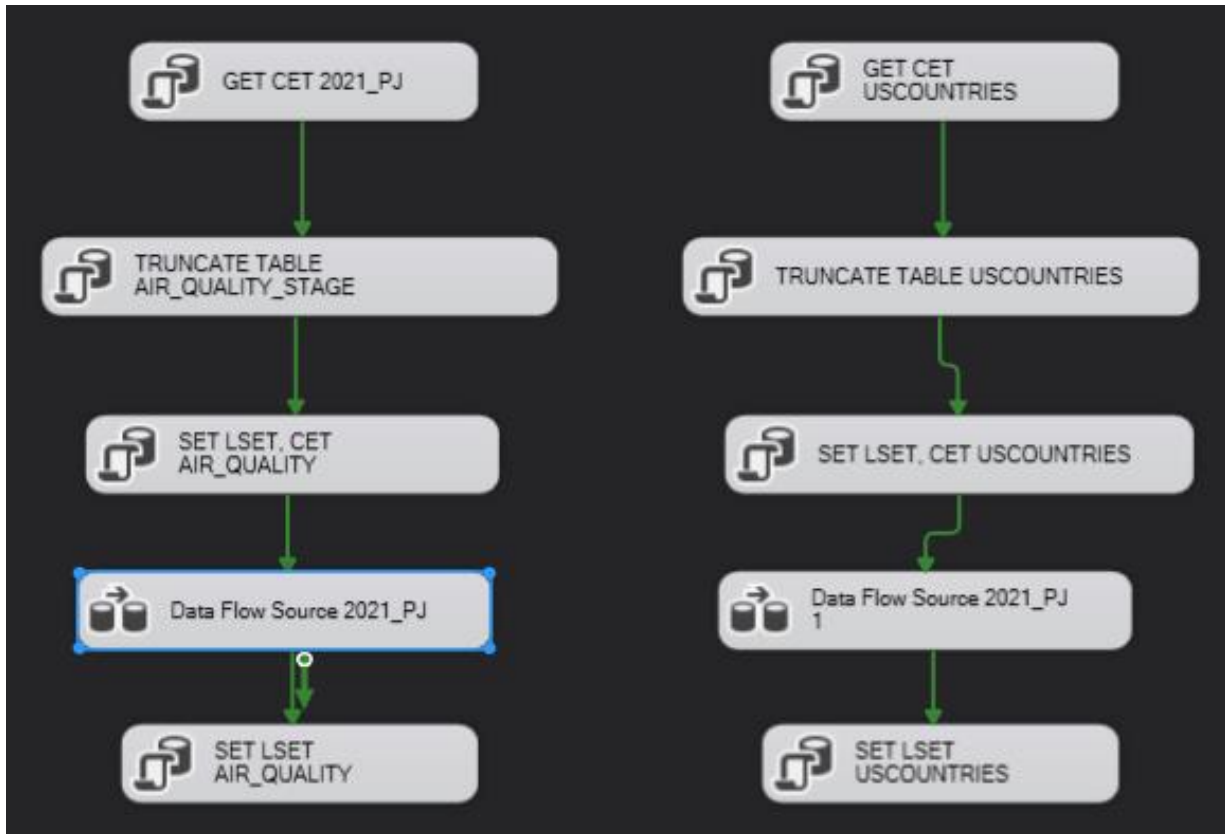
MỤC LỤC	2
Thông tin thành viên và phân công	3
NỘI DUNG.....	4
1. Source đổ data vào Stage	4
2. Stage sang NDS	6
NGUỒN THAM KHẢO	10

Thông tin thành viên và phân công

STT	Họ và Tên	MSSV	Công việc	Mức độ hoàn thành
1	Nguyễn Quốc Huy	20127189	<ul style="list-style-type: none">● Source sang Stage● Stage sang NDS● Report	100%
2	Nguyễn Phương Khanh	20127204	<ul style="list-style-type: none">● Source sang Stage● Report	100%
3	Nguyễn Huy Hoàng	19127407	<ul style="list-style-type: none">● Stage sang NDS● Report	100%
4	Bùi Nguyễn Tin	21127702	<ul style="list-style-type: none">● Source sang Stage● Stage sang NDS● Report	100%

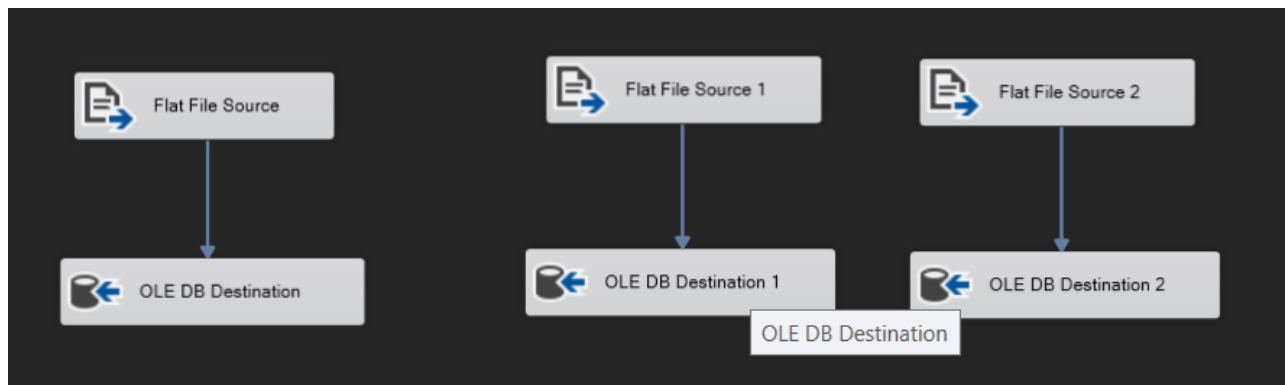
NỘI DUNG

1. Source dữ liệu vào Stage

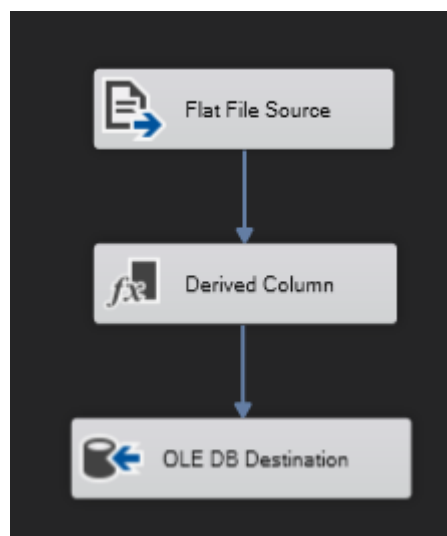


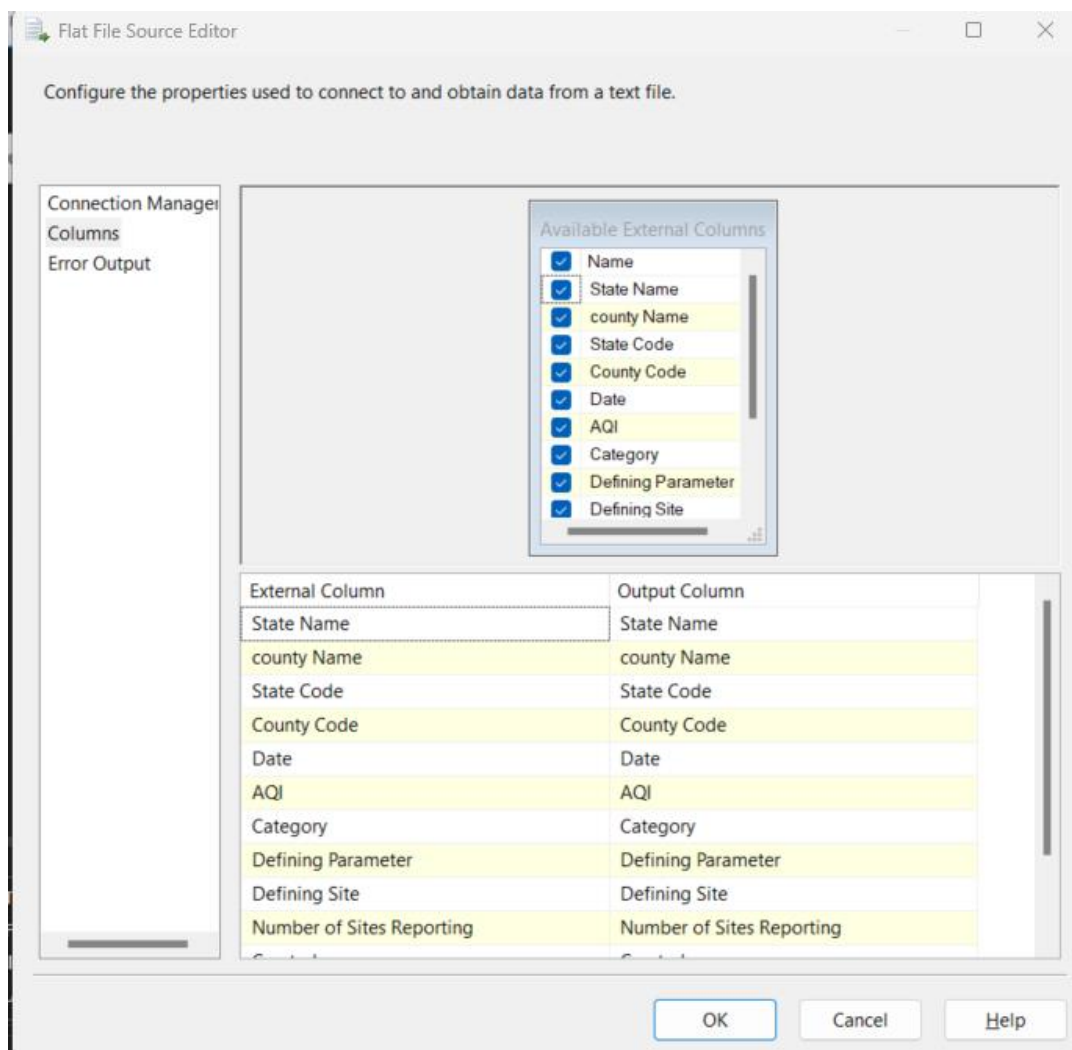
Đây là sơ đồ quy trình xử lý dữ liệu, nó mô tả các bước thực hiện như sau:

1. GET CET 2021_PJ: Bước này lấy dữ liệu từ nguồn dữ liệu có tên [CET 2021_PJ].
2. TRUNCATE TABLE AIR_QUALITY_STAGE: Bước này xóa toàn bộ dữ liệu trong bảng [AIR_QUALITY_STAGE].
3. SET LSET, GET AIR_QUALITY: Bước này thiết lập và lấy dữ liệu [AIR_QUALITY].
4. TRUNCATE TABLE USCOUNTRIES: Bước này xóa toàn bộ dữ liệu trong bảng [USCOUNTRIES].
5. SET LSET, GET USCOUNTRIES: Bước này thiết lập và lấy dữ liệu [USCOUNTRIES].
6. Data Flow Source 2021_PJ và Data Flow Source 2021_PJ 1: Đây là hai nguồn dữ liệu được sử dụng trong quy trình này.

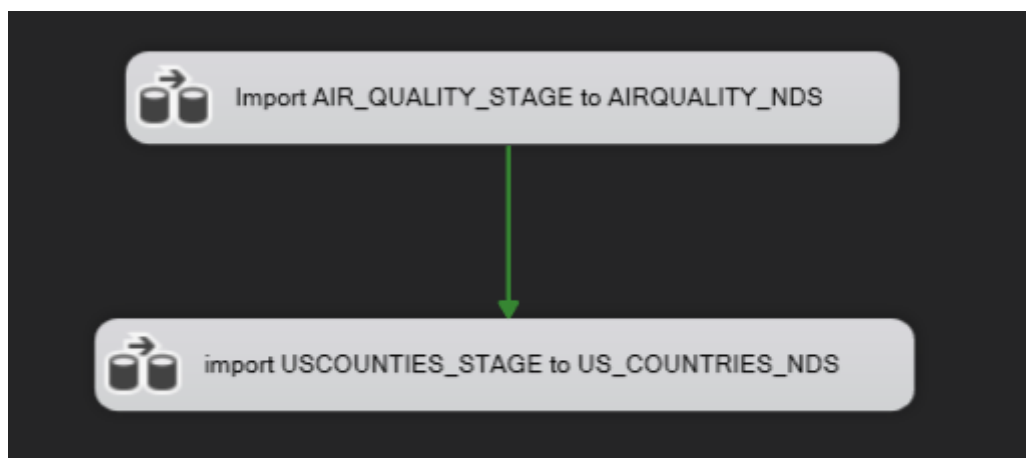


Tương tự cho DataFlow của county từ Source sang Stage:

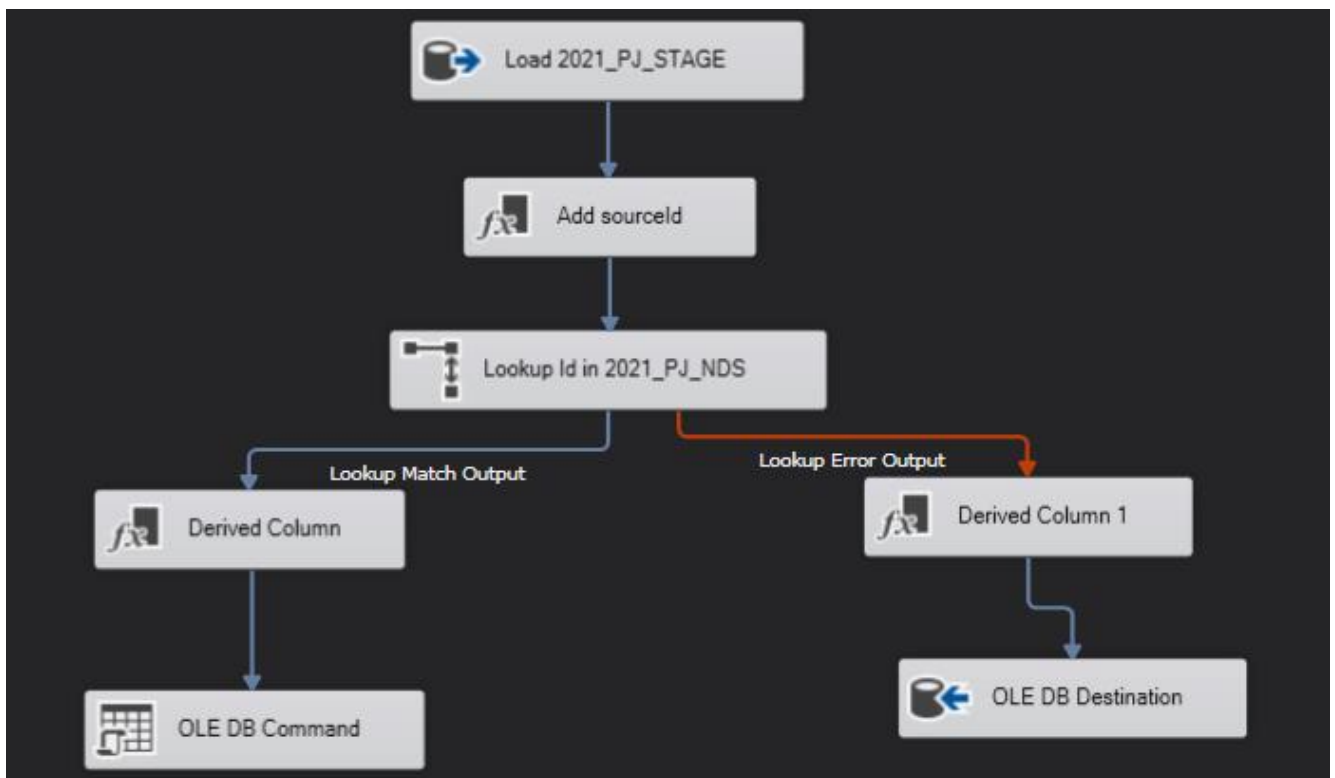




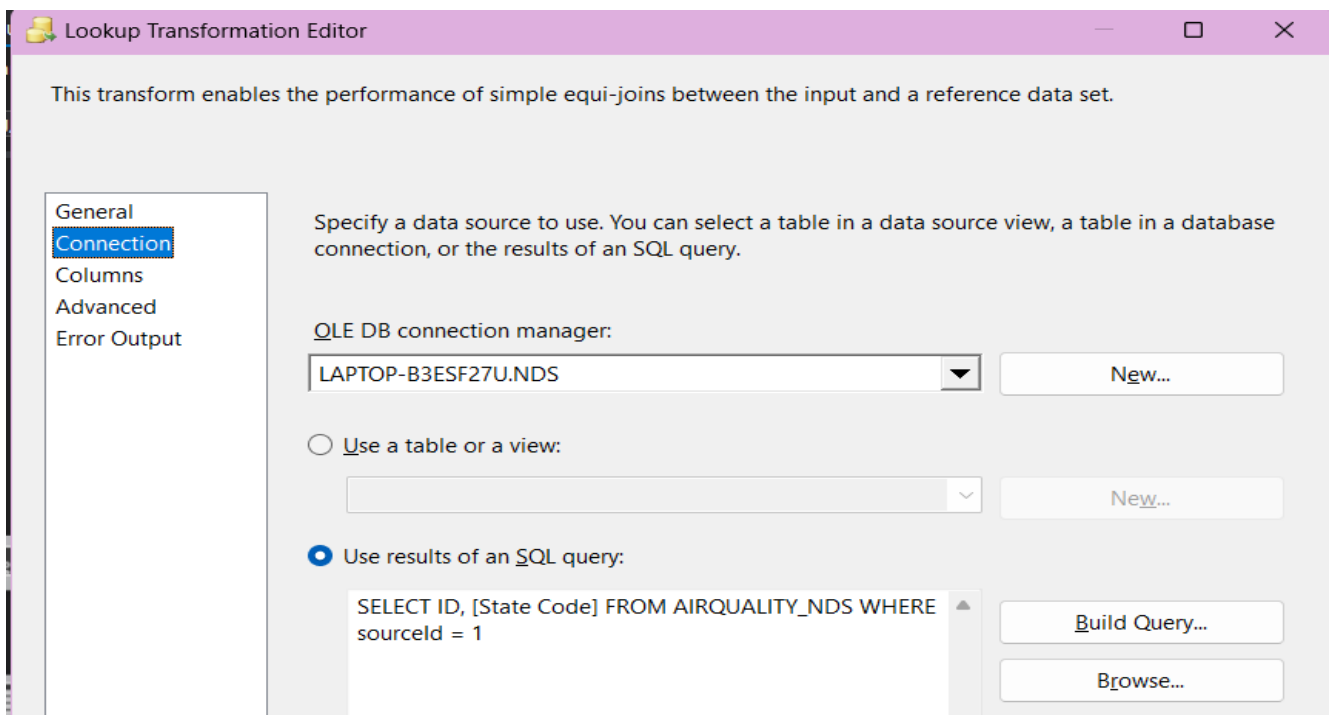
2. Stage sang NDS



Quá trình ETL từ Stage sang NDS của bảng [AIRQUALITY_NDS]:



Sau khi add SourceId vào, tiến hành Lookup bằng cách Select ID và State Code trong bảng [AIRQUALITY_NDS]:



Theo hướng Lookup thành công thì update lại cột [Last updated]:

Advanced Editor for OLE DB Command

The advanced editor provides access to the low-level properties of data flow components. Additionally, the advanced editor can be used to configure components that do not have a custom user interface.

Connection Managers | **Component Properties** | Column Mappings | Input and Output Properties

Specify advanced properties for the data flow component.

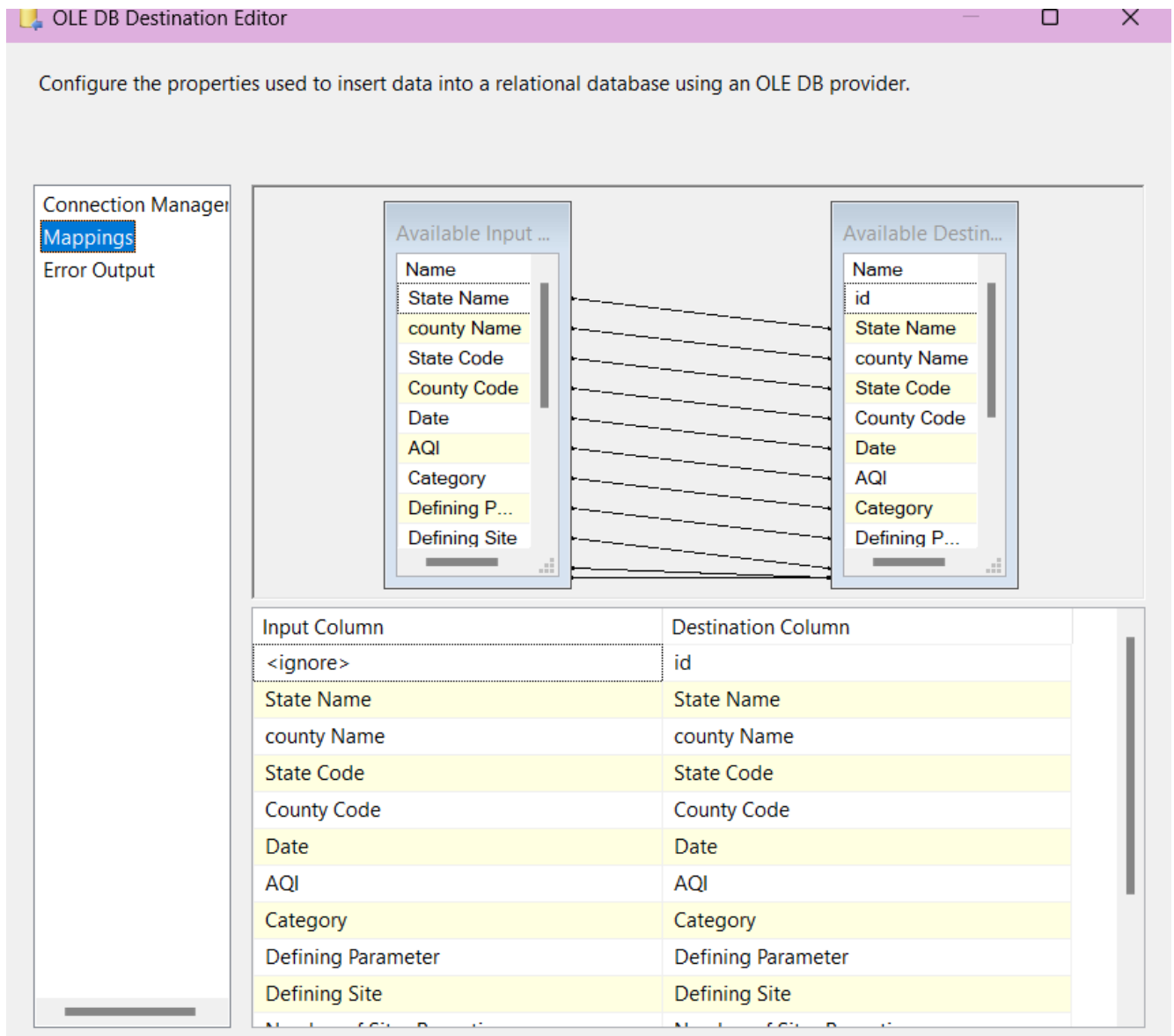
Properties:

Common Properties	
ComponentClassID	{4AA3F603-2558-44FD-9CAD-F8005FB6A880}
ContactInfo	OLE DB Command;Microsoft Corporation; Microsoft SQL S
Description	Runs an SQL statement for each row in a data flow. For exa
ID	212
IdentificationString	OLE DB Command
IsDefaultLocale	True
LocaleID	English (United States)
Name	OLE DB Command
PipelineVersion	0
UsesDispositions	True
ValidateExternalMetadata	True
Version	2
Custom Properties	
CommandTimeout	0
DefaultCodePage	1252
SqlCommand	UPDATE AIRQUALITY_NDS SET [Last Updated] = ? WHERE sou

Name
Specifies the name of the component.

Refresh OK Cancel Help

Theo hướng lookup không thành công thì tiến hành thêm mới vào từ Stage vào bảng [AIRQUALITY_NDS]:



NGUỒN THAM KHẢO

[1]: BI_supermarket_sales project - Duc An Nguyen

https://github.com/ducannng/BI_supermarket_sales

[2]: ETL Project From Excel Data Source to Star Schema with SSIS - Chuc Nguyen Van

[ETL Project From Excel Data Source to Star Schema with SSIS](#)

[3]: How to build ETL Project with SSIS Step by Step: Derived columns with conditional expression - Chuc Nguyen Van

[How to build ETL Project with SSIS Step by Step: Derived columns with conditional expression \(?\)](#)