**NỘI DUNG BÁO CÁO**

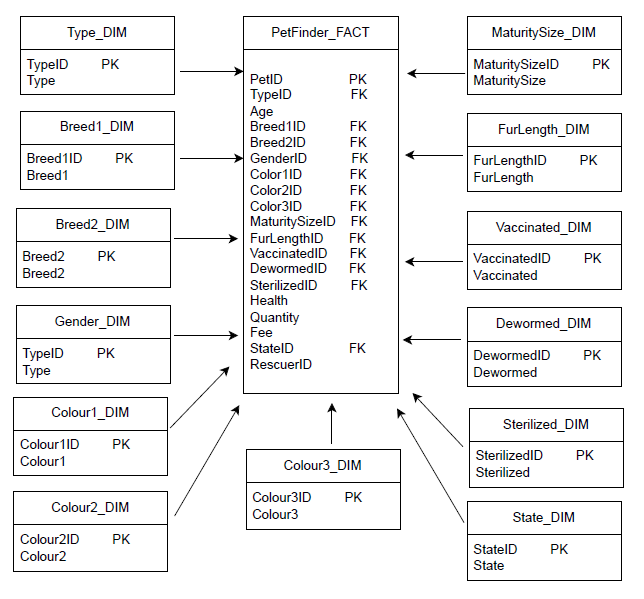
**1.Thiết kế ERD và xây dựng cơ sở dữ liệu………. trang 2 – trang 4**

**2. Xây dựng ETL cho Dimension Table…………..trang 5 – trang 22**

**3. Xây dựng ETL cho Fact Table………………….trang 23 – trang 37**

**1.Thiết kế ERD và xây dựng cơ sở dữ liệu:**

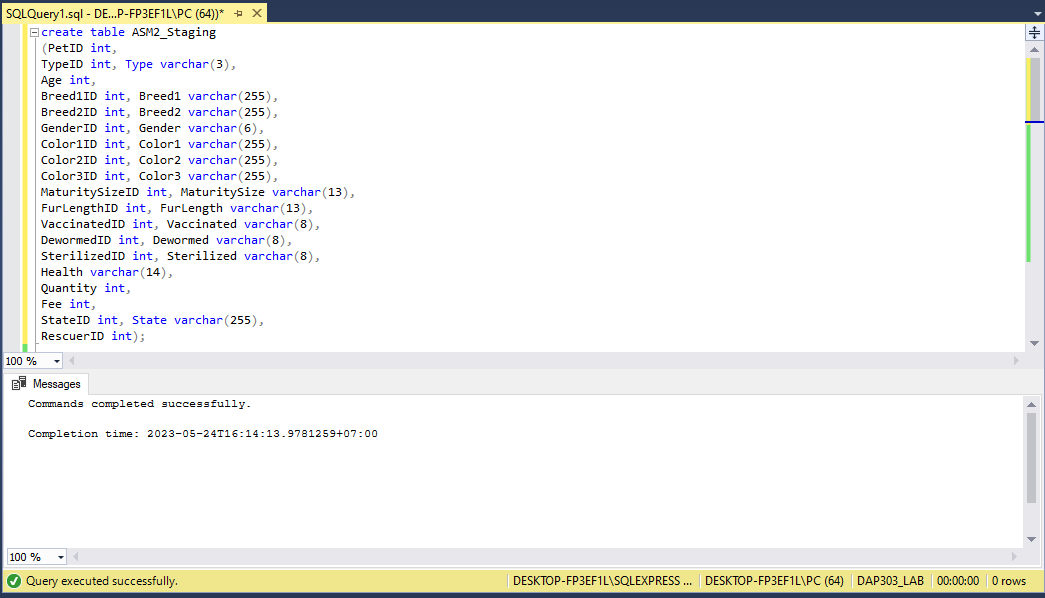
**ERD:**



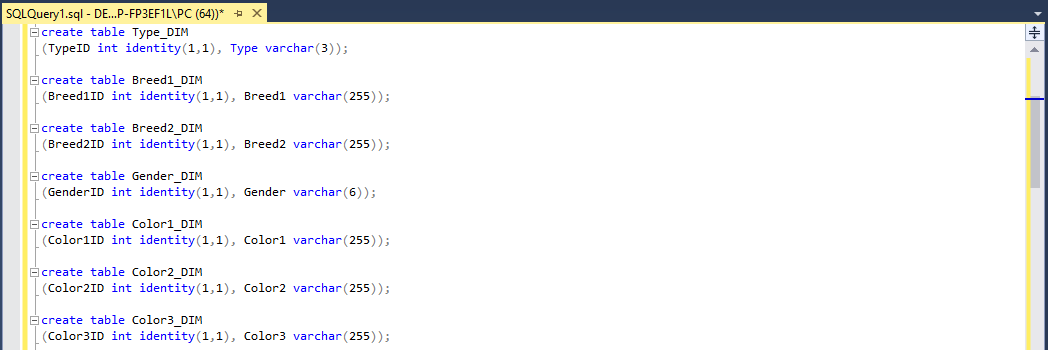
**Xây dựng cơ sở dữ liệu:**

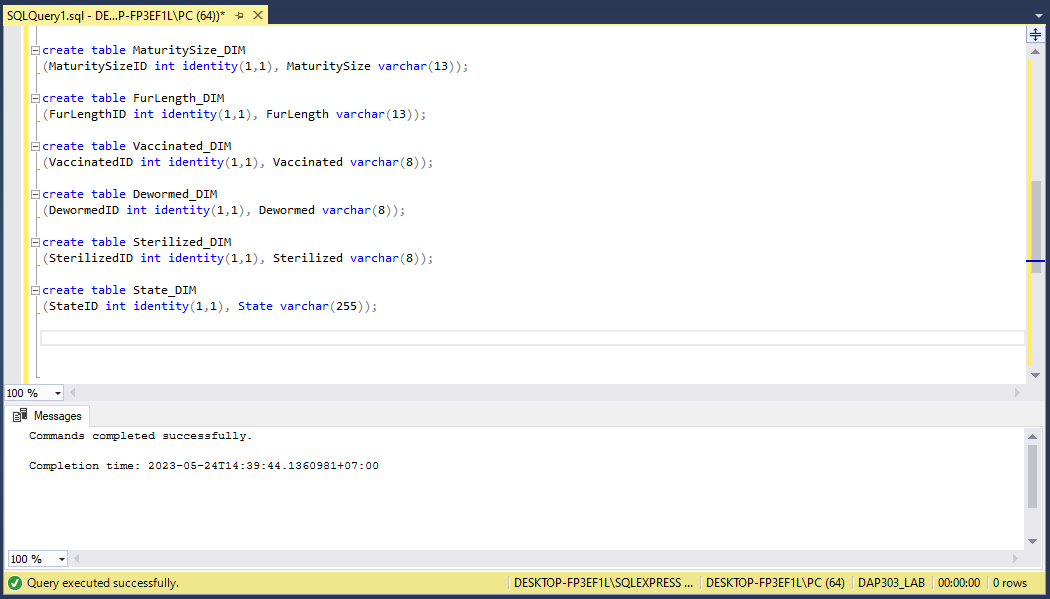
***+ Tạo các bảng trong SQL:***

Bảng “ASM2\_Staging” đưa dữ liệu từ csv vào SQL:

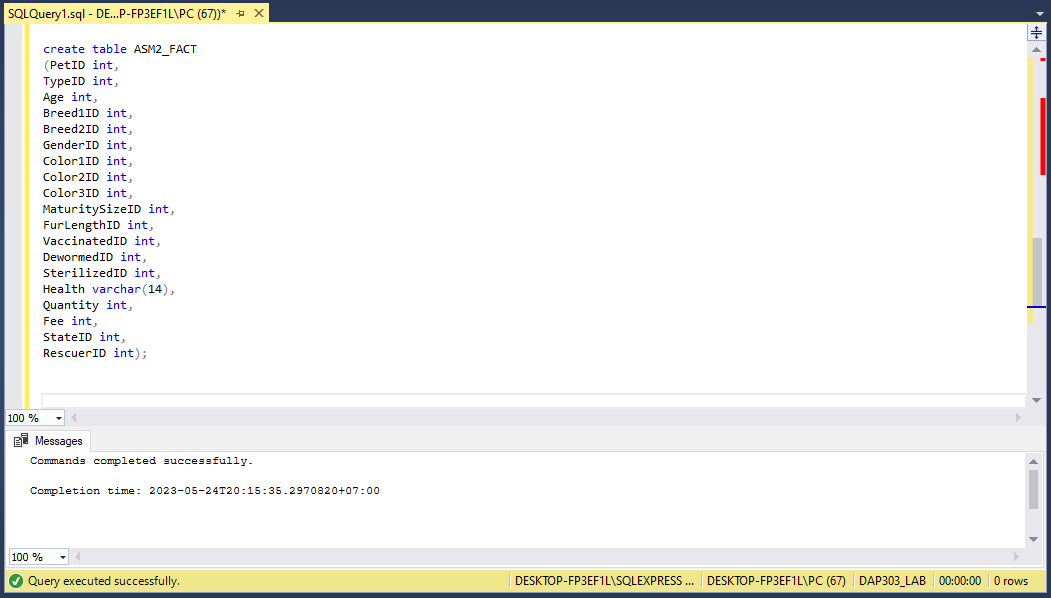
****

Tạo các bảng DIM:



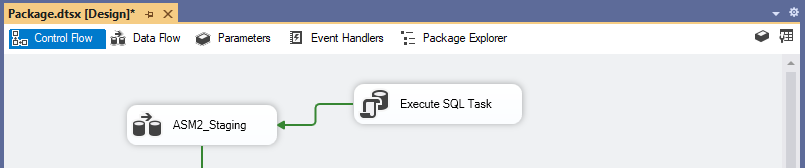


Tạo bảng FACT:



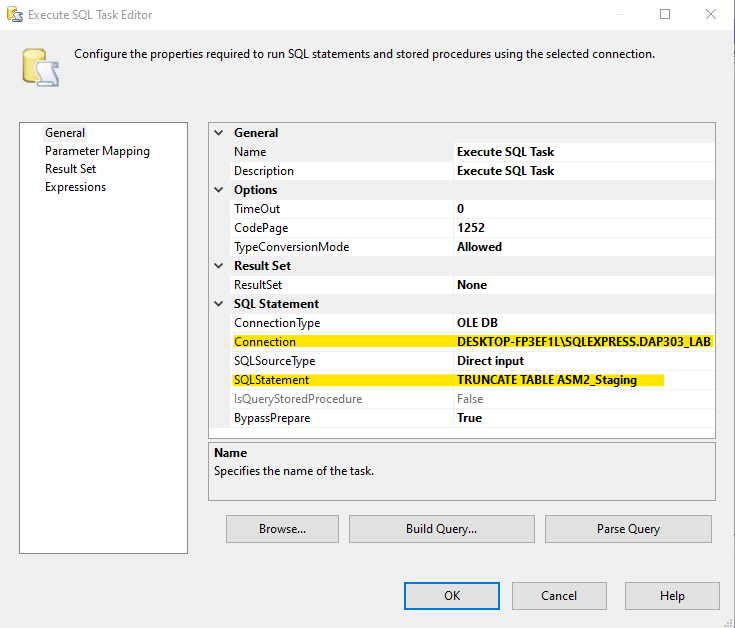
**2. Xây dựng ETL cho Dimension Table:**

* Tạo Data Flow “ASM2\_Staging” chuyển dữ liệu từ .CSV sang bảng “ASM2\_Staging” trên SQL:

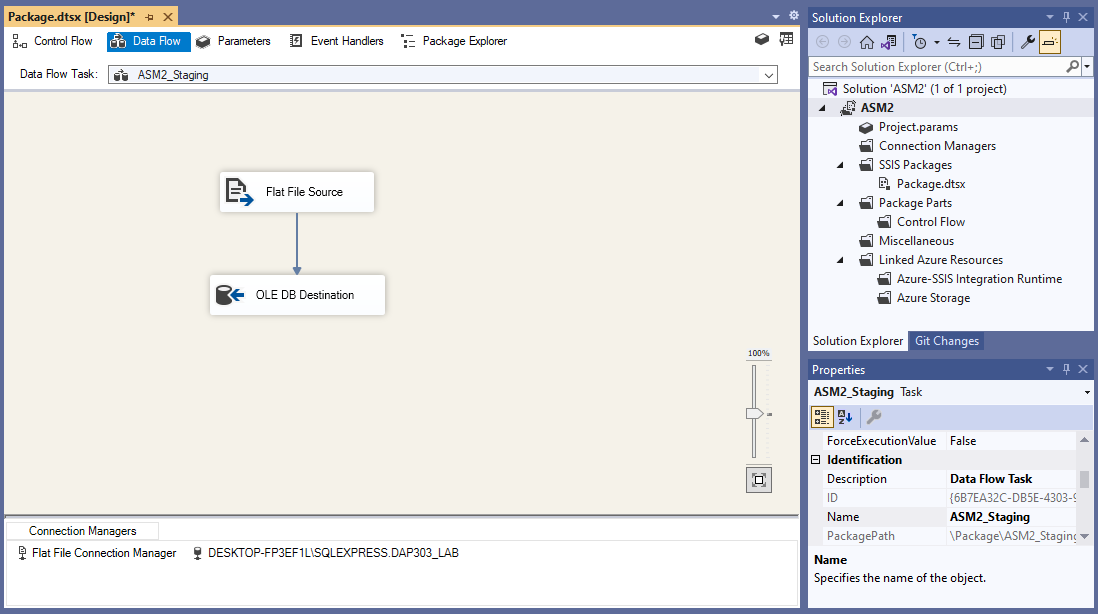


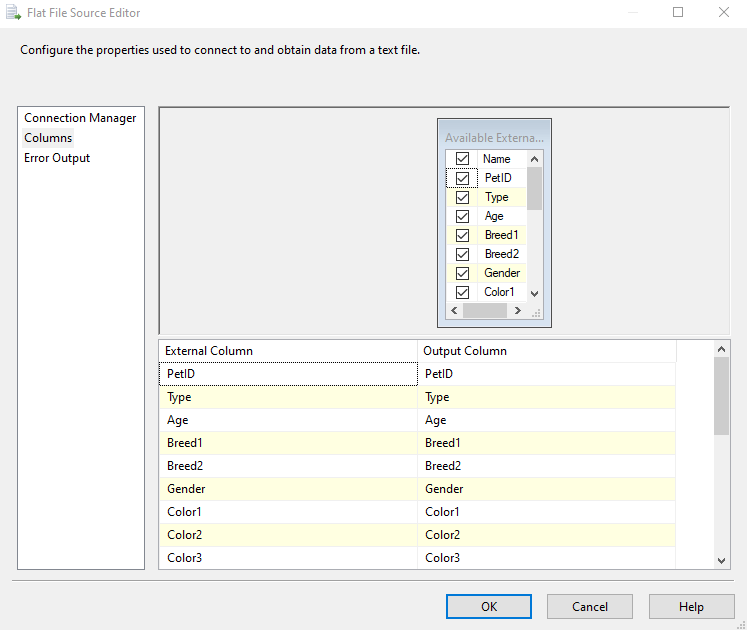
“ASM2\_Staging” sẽ có “Execute SQL Task”: sẽ xóa toàn bộ dữ liệu cũ của bảng để đưa dữ liệu mới vào (trong các trường hợp chạy lại Start của Flow này trên SSIS)

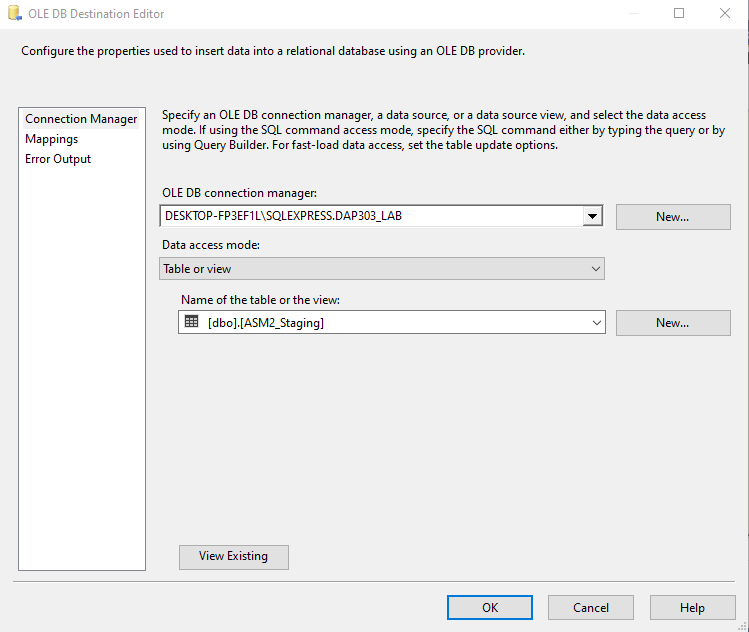
Chi tiết trong “Execute SQL Task”:



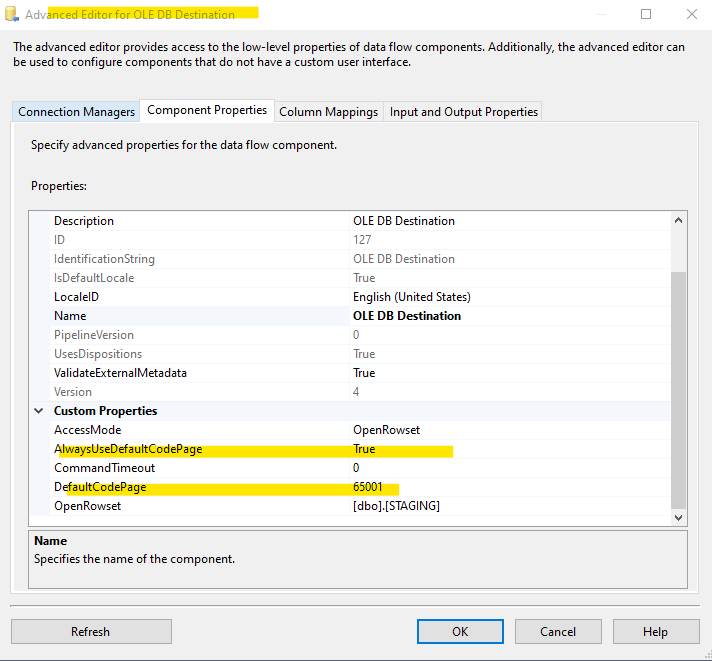
Chi tiết trong “ASM2\_Staging”:



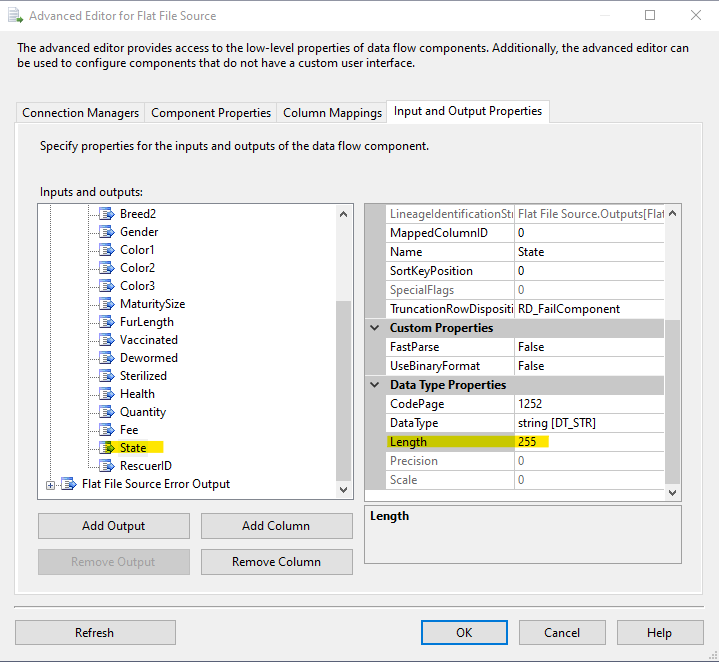


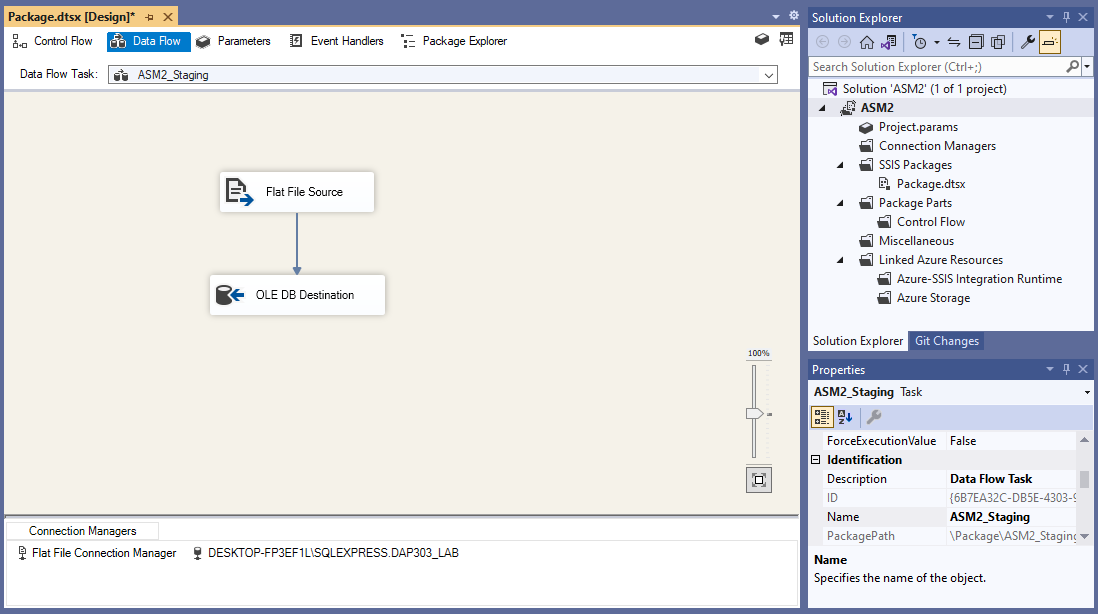


Sửa trong Destination trước khi Mapping:

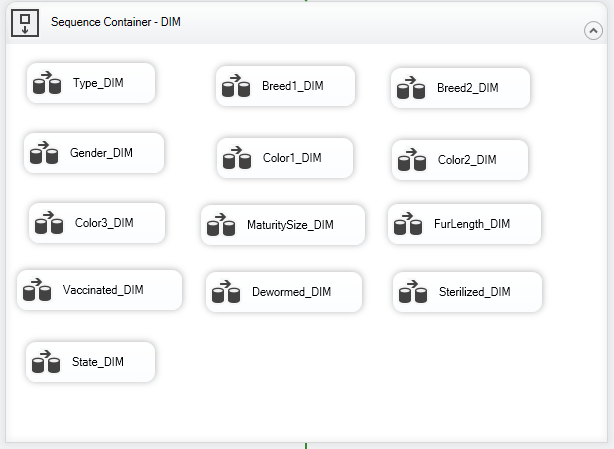


Sửa Length các cột có định độ dài khi tạo bảng ở SQL như Type, Breed1,…





* Tạo Data Flow cho các DIM lấy dữ liệu từ “ASM2\_Staging”:

Tạo từng Data Flow cho từng bảng DIM (theo ERD là 13 bảng), sau đó đưa 13 bảng này vào chung 1 Sequence Container

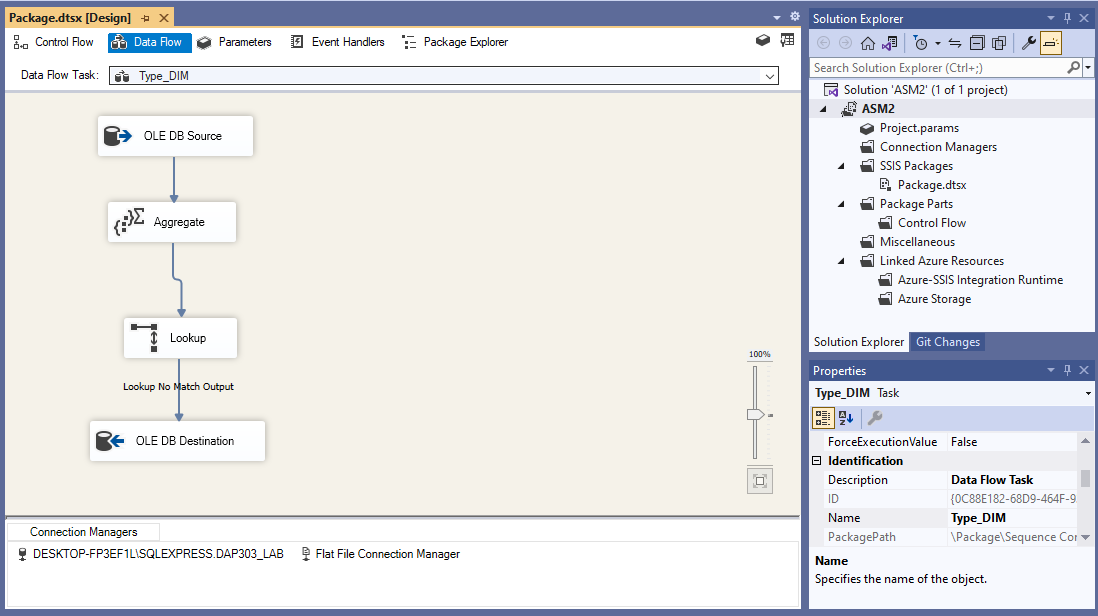
Các Flow của từng DIM đều thực hiện giống nhau:

Nguồn dữ liệu của các Data Flow này (OLE DB Source) từ “ASM2\_Staging” và kết quả (OLE DB Destination) sẽ được lưu tại bảng DIM tương ứng đã tạo trên SQL.

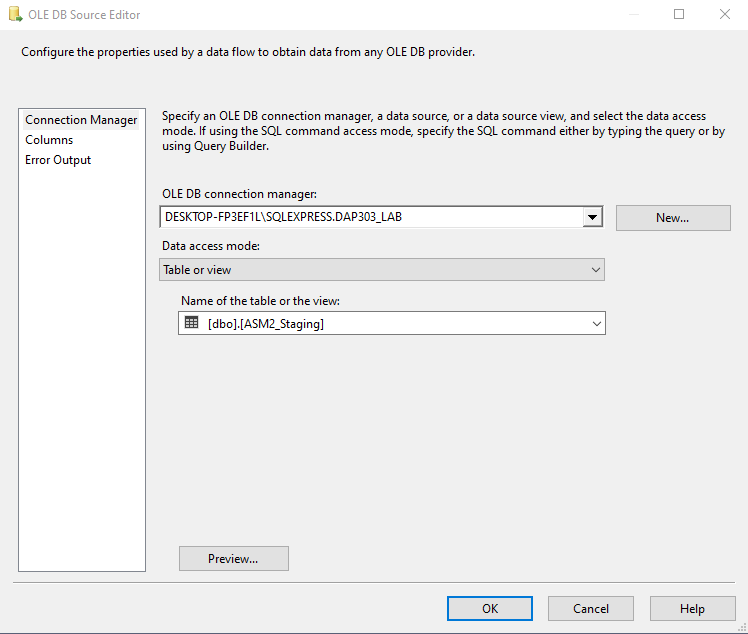
Aggregate: gộp các thông tin chung thành 1 thông tin duy nhất

Lookup: Tìm kiếm từng giá trị trong cột DIM rồi so sánh với kết quả đã có tại bảng DIM trên SQL và chỉ đưa vào bảng DIM trên SQL những giá trị chưa có

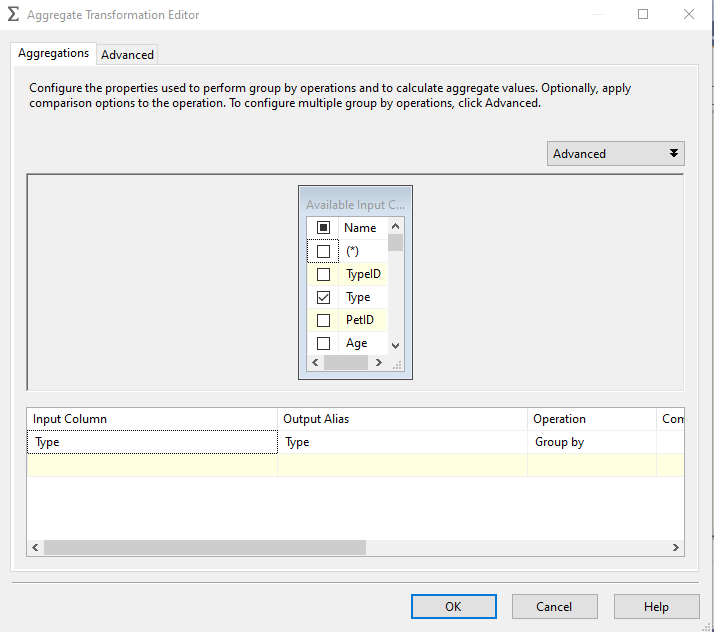
Các thông tin tạo 1 Flow cho 1 bảng DIM như sau:



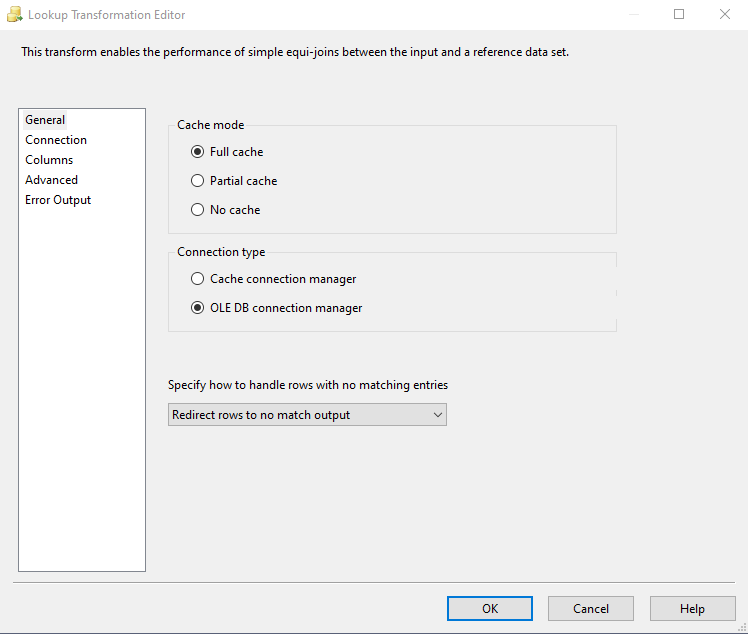
OLE DB Source:

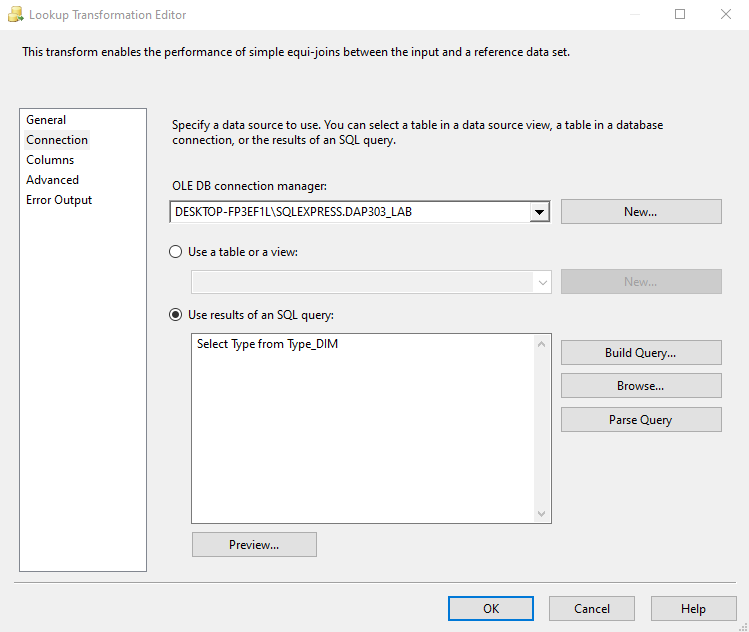


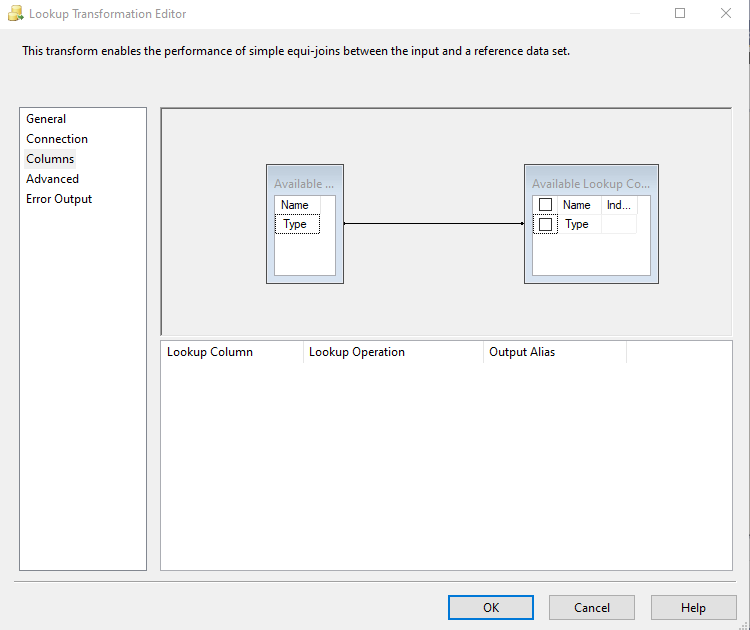
Aggregate:



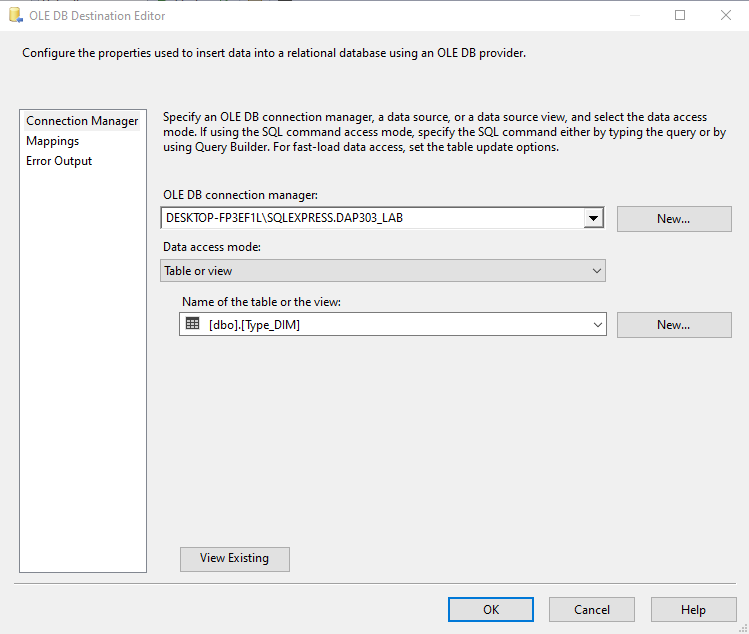
Lookup:

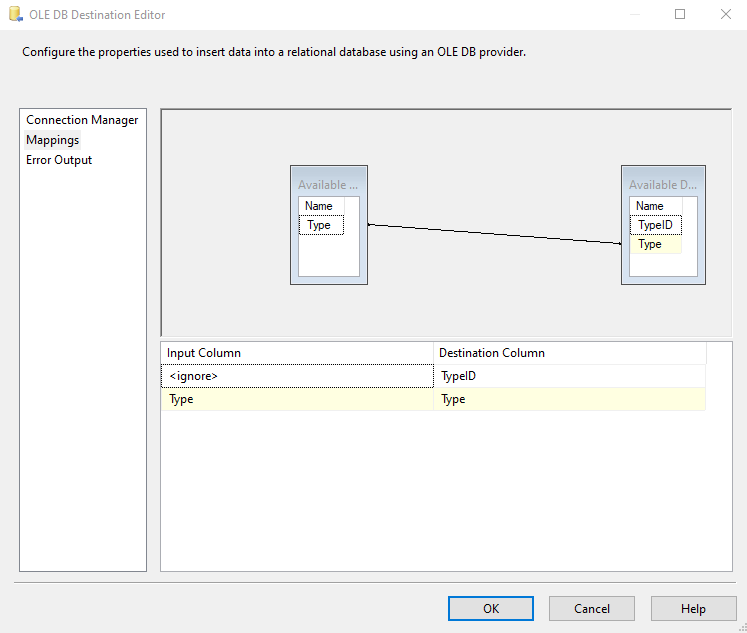




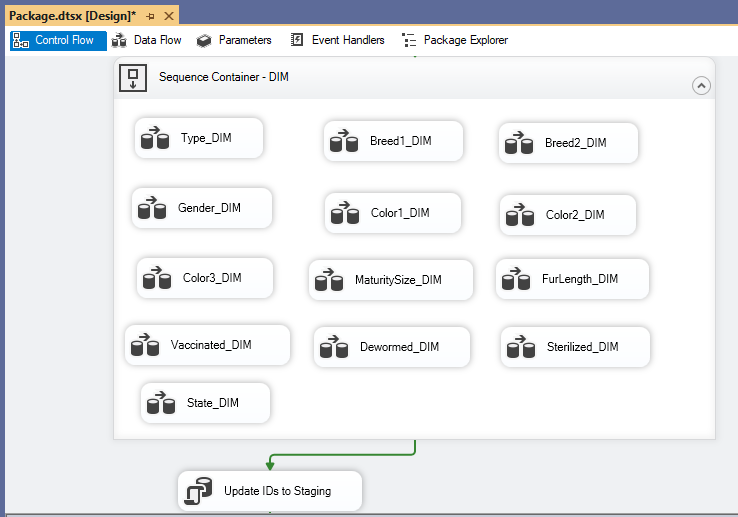


OLE DB Destination:

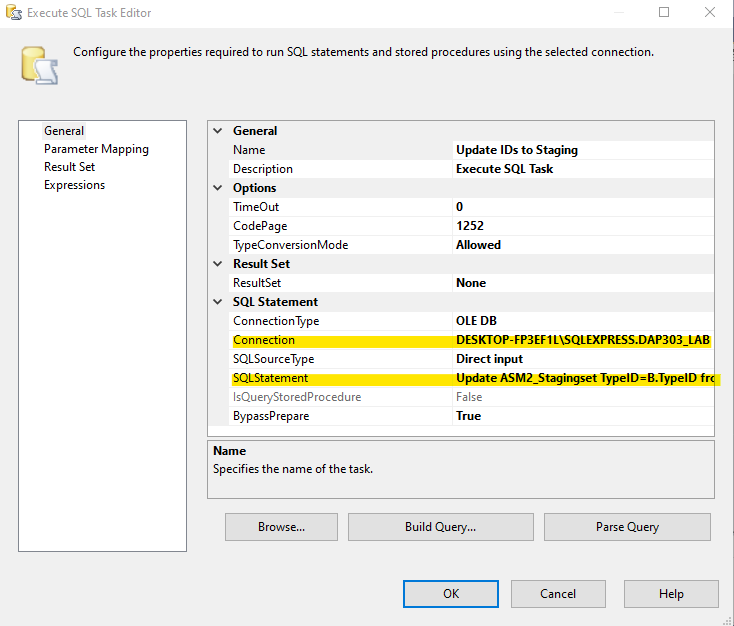




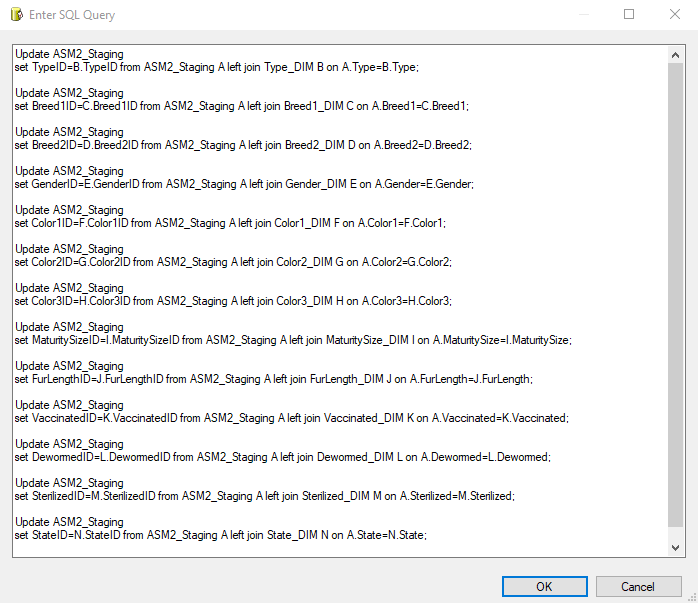
* Tạo Execute SQL Task là “Update IDs to Staging”: cập nhật các cột ID trong các bảng DIM vào bảng “ASM2\_Staging”:



Chi tiết trong Flow này như sau:

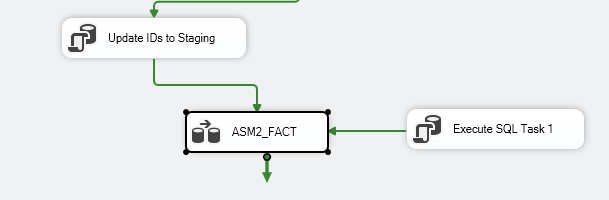


Chi tiết của SQLStatement như sau:

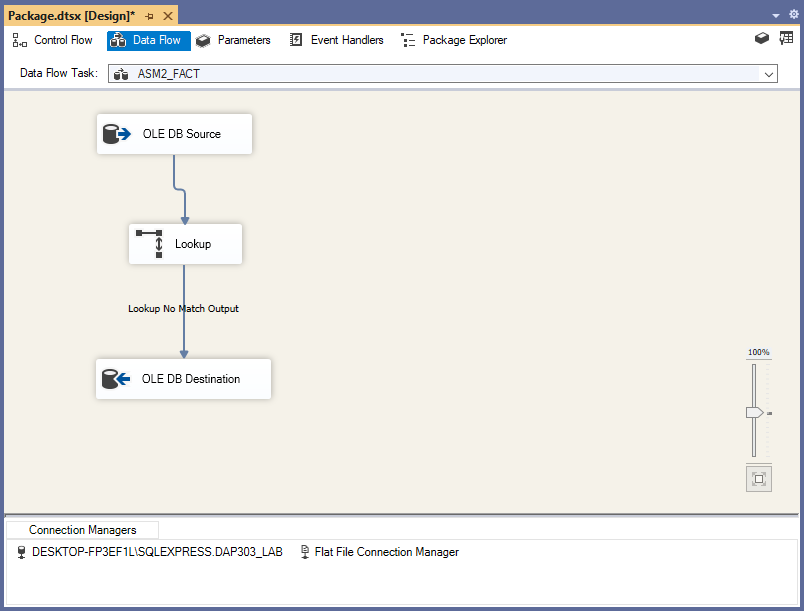


**3. Xây dựng ETL cho Fact Table:**

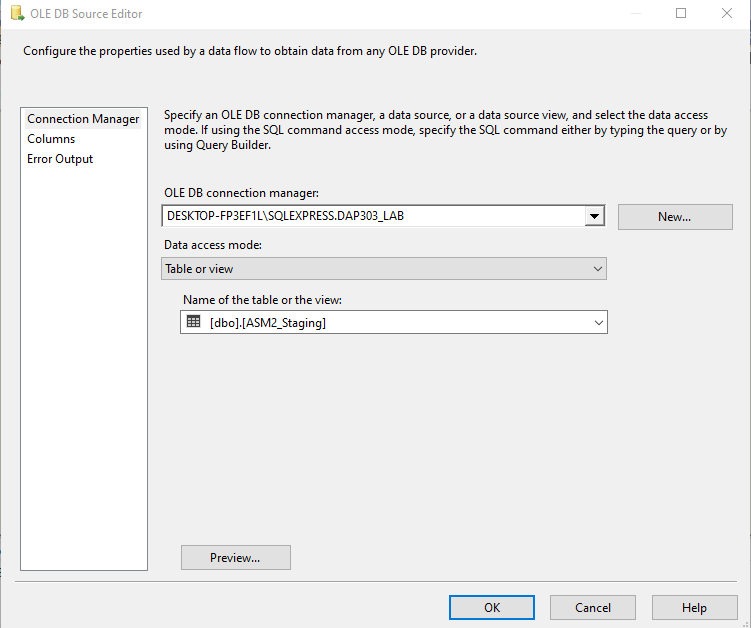
* Tạo Flow là bảng ASM2\_FACT:



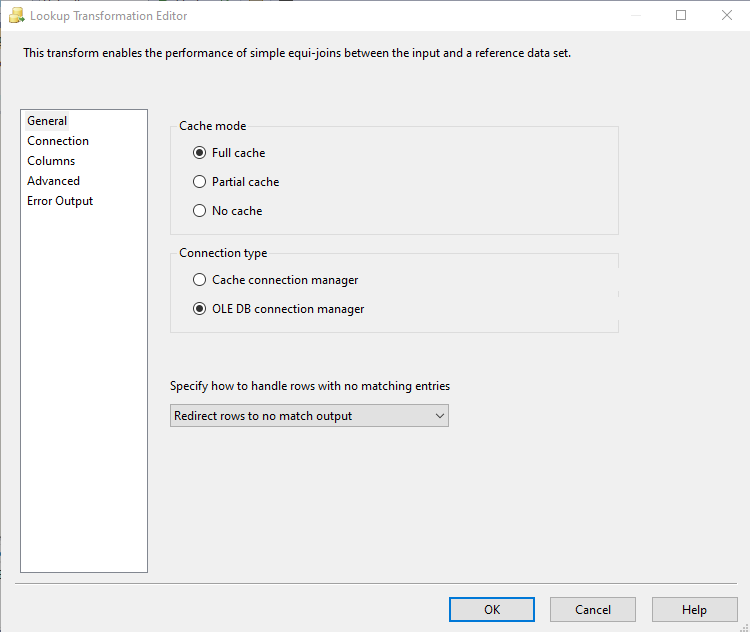
Chi tiết bên trong Flow ASM2\_FACT như sau:

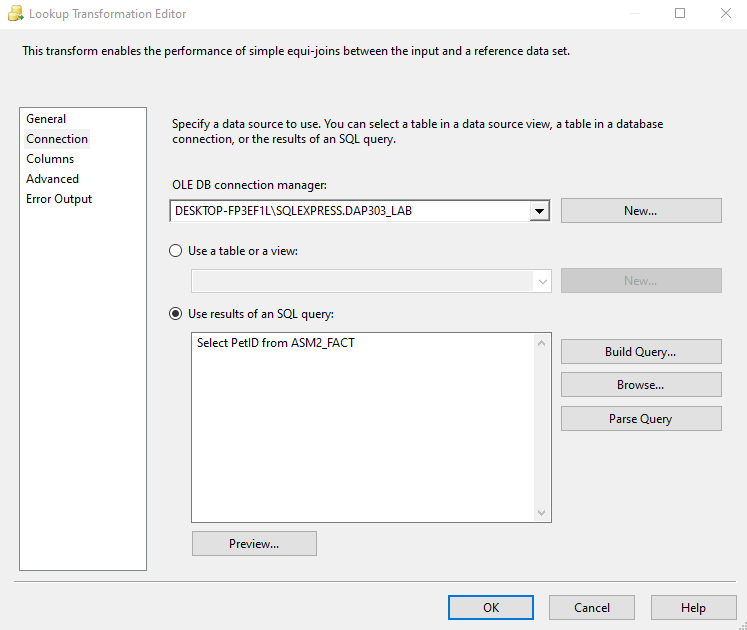


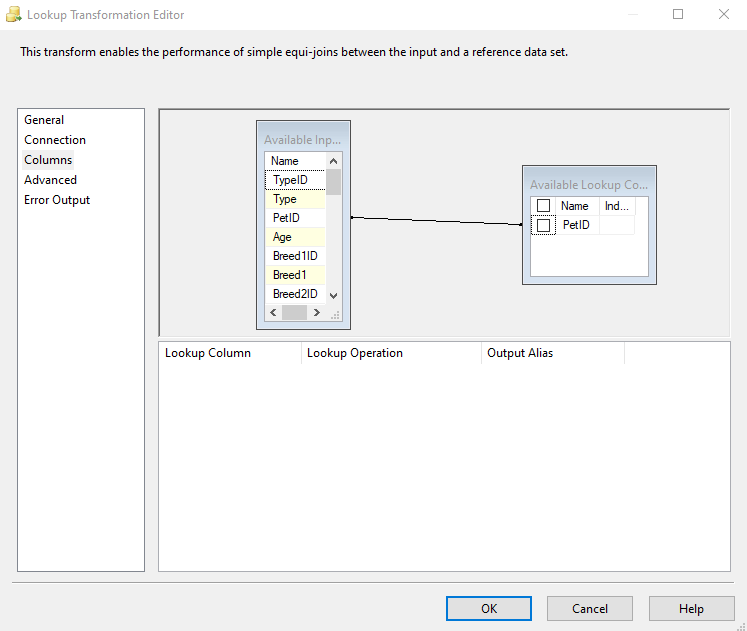
OLE DB Source:

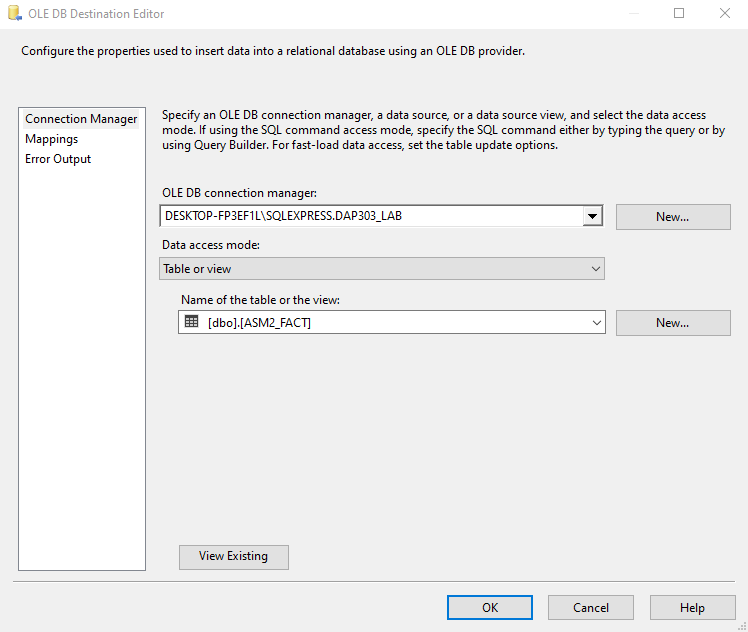


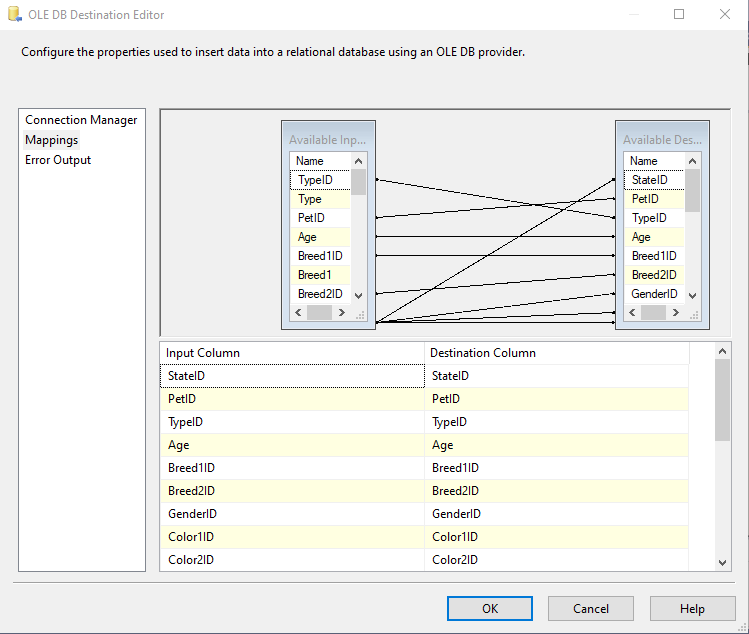
Lookup:



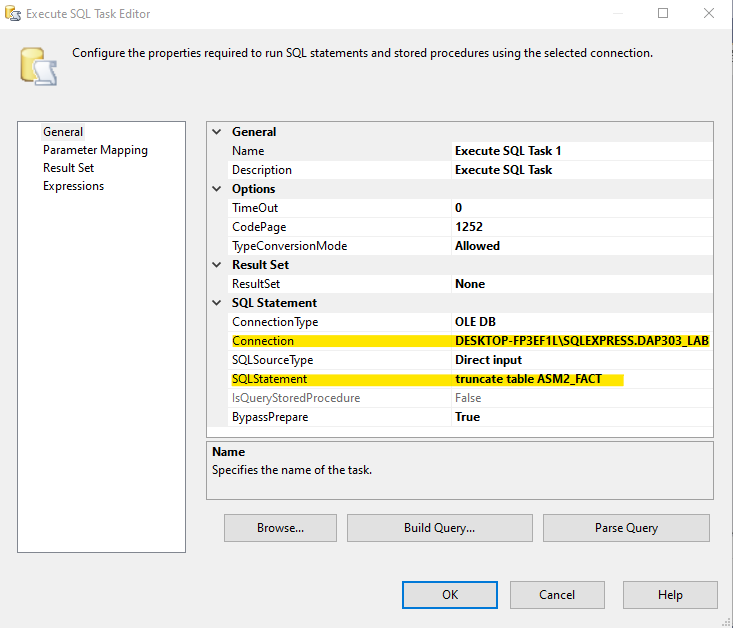




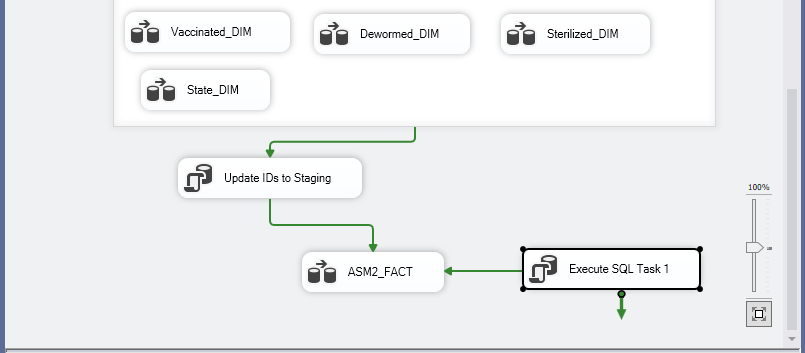
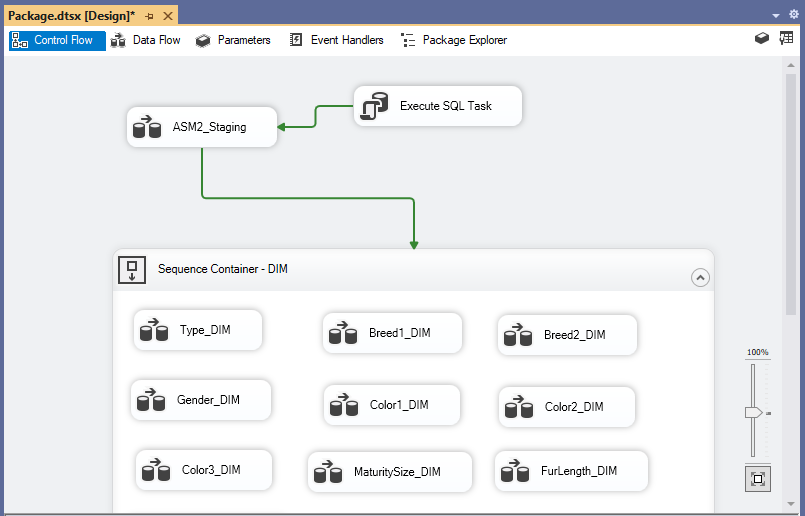




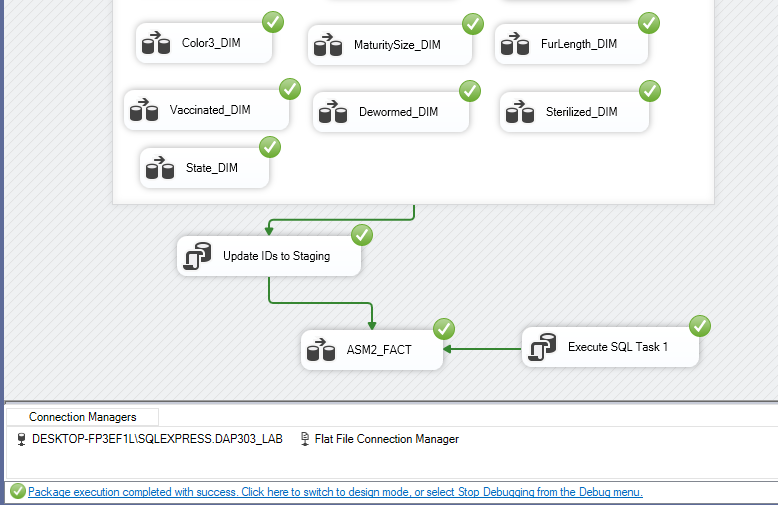
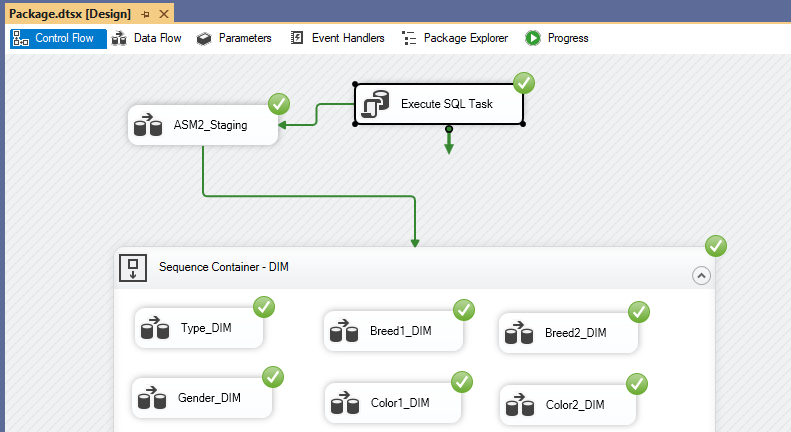
Data Flow của bảng ASM2\_FACT sẽ tạo 1 Execute SQL Task: “Execute SQL Task”, sẽ xóa toàn bộ dữ liệu của trong bảng FACT hiện tại và chỉ chứa dữ liệu mới



**Toàn bộ Data Flow sẽ như sau:**

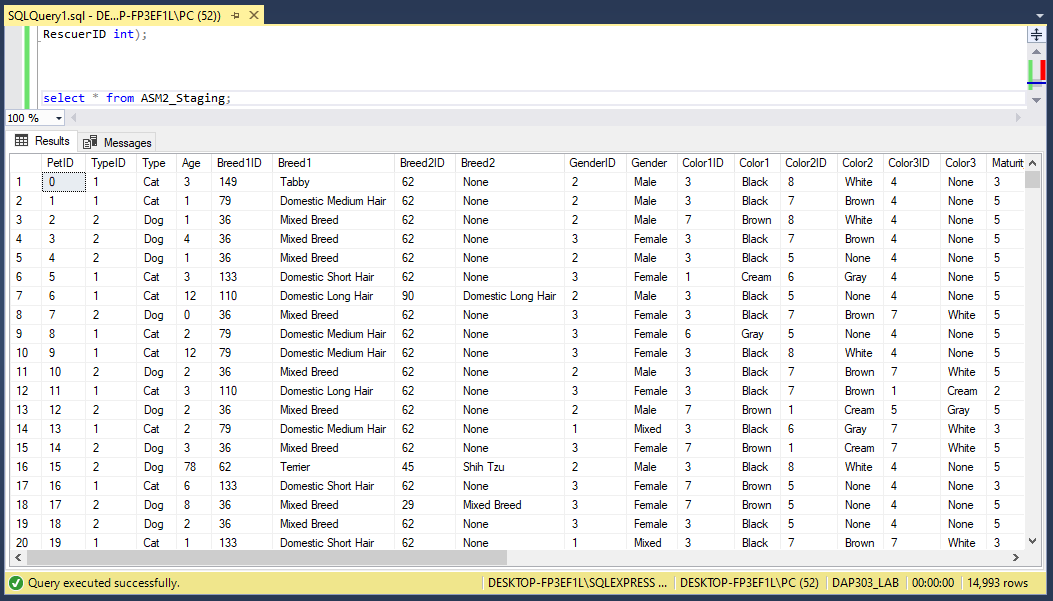


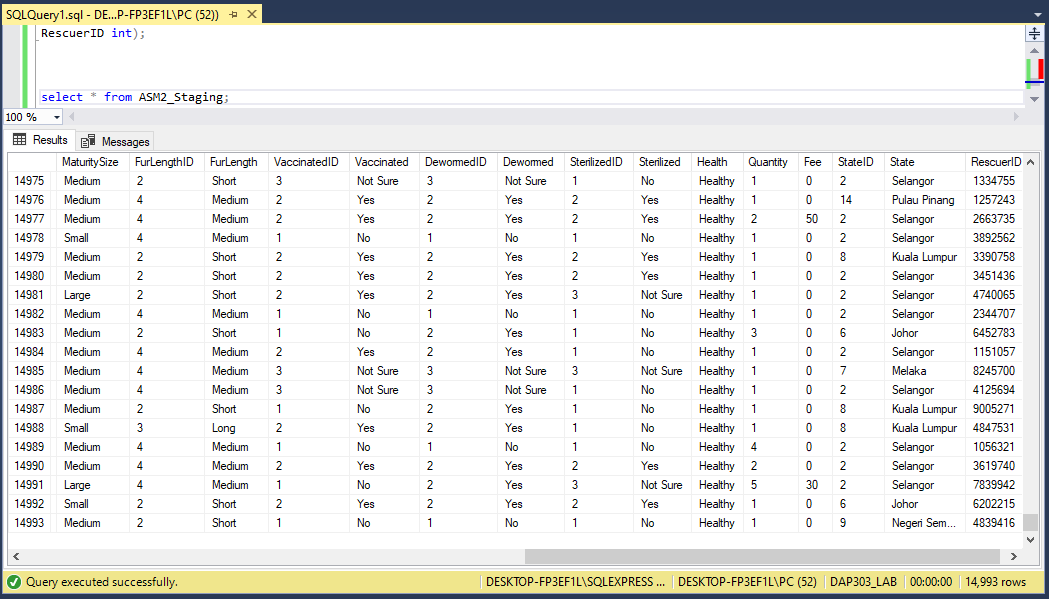
**Kết quả sau khi chạy ETL như sau:**



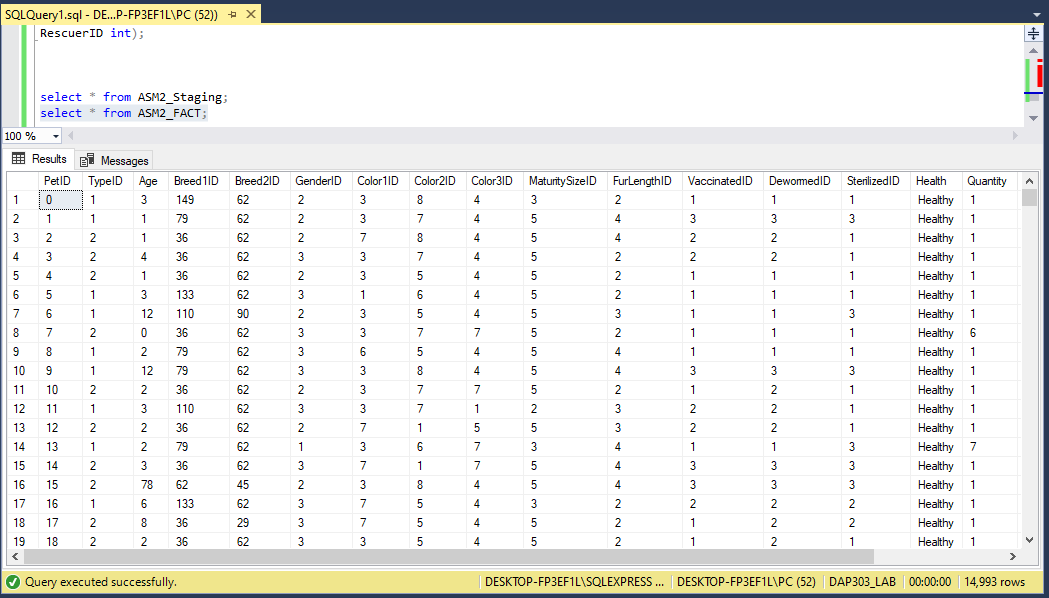
**Trên SQL xem thông tin các bảng như sau:**

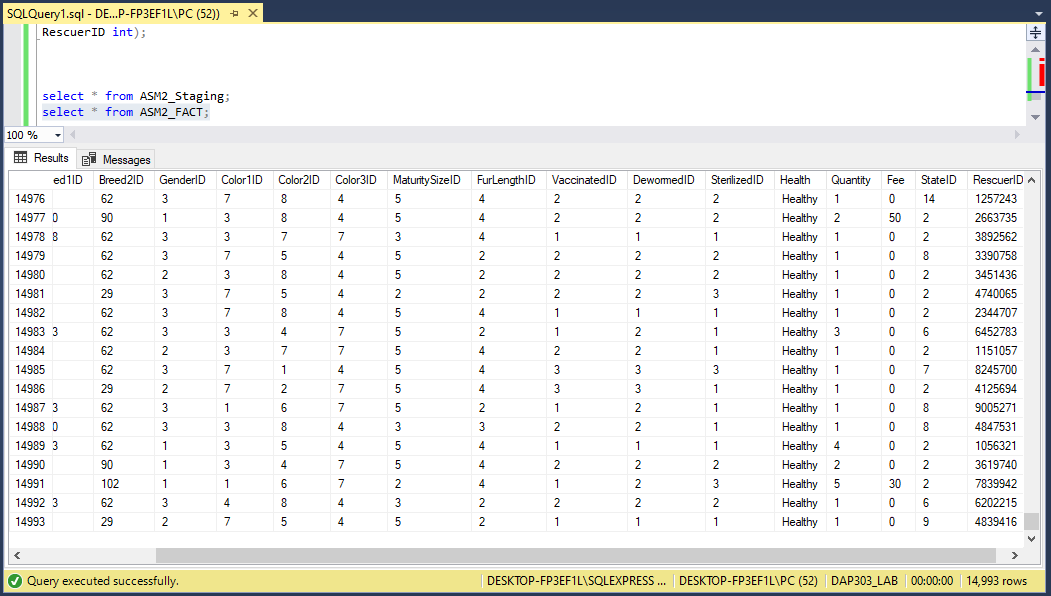
Bảng “ASM2\_Staging”:





Bảng “ASM2\_FACT”:





Các bảng DIM:

