

TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM  
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO ĐỒ ÁN CUỐI KỲ

**ĐỀ TÀI: XÂY DỰNG KHO DỮ LIỆU VỀ TAI NẠN XE HƠI  
TẠI VƯƠNG QUỐC ANH**

Môn học:

KHO DỮ LIỆU

Mã lớp học phần:

DAWH430784\_22\_2\_01

GVHD:

ThS. Nguyễn Văn Thành

Nhóm sinh viên thực hiện (Nhóm 17):

1. Huỳnh Nguyễn Như Nguyên – 20133019
2. Nguyễn Thị Diệu Hiền – 20133040
3. Nguyễn Thắng - 20133

*TP. Hồ Chí Minh, tháng 5 năm 2023*

## BẢNG THÔNG TIN CHI TIẾT NHÓM

STT	Họ và tên	MSSV
1	Huỳnh Nguyễn Như Nguyên	20133019
2	Nguyễn Thị Diệu Hiền	20133040
3	Nguyễn Quốc Thắng	20133120

## BẢNG PHÂN CÔNG CHI TIẾT

Công việc	Người thực hiện		
	Huỳnh Nguyễn Như Nguyên	Nguyễn Thị Diệu Hiền	Nguyễn Quốc Thắng
Chọn dataset	33%	33%	33%
Nghiên cứu tính cần thiết để xây dựng kho dữ liệu	33%	33%	33%
Khảo sát nghiên cứu, phân tích báo cáo nghiệp vụ	33%	33%	33%
Xây dựng lược đồ hình sao	33%	33%	33%
Xây dựng lược đồ hình bông tuyết	33%	33%	33%
Quá trình SSIS – Source to Stage	50%	50%	
Quá trình SSIS – Stage to NDS		60%	40%
Quá trình SSIS – NDS to Data Warehouse	30%	40%	30%
Quá trình SSAS – Xây dựng Cube, Phân chiềú	30%	70%	
Truy vấn với ngôn ngữ DMX	33%	33%	33%
Trực quan với Power BI	33%	33%	33%
Viết báo cáo	50%	50%	

## MỤC LỤC

LỜI CẢM ƠN.....	1
LÝ DO CHỌN ĐỀ TÀI.....	2
<b>A. TẬP DỮ LIỆU.....</b>	<b>3</b>
1. Giới thiệu bộ dữ liệu .....	3
2. Mô tả dữ liệu:.....	4
2.1. UK Car Accidents – Bảng Casualties .....	4
2.2. UK Car Accidents – Bảng Vehicles .....	7
2.3. UK Car Accidents – Bảng Accidents .....	14
2.4. Bảng Postcodes (UK Postcodes) .....	15
2.5. Bảng PCD_OA_LSOA_MSOA_LAD_AUG21_UK_LU .....	16
3. Công cụ hỗ trợ.....	17
<b>B. DATA WAREHOUSE .....</b>	<b>18</b>
1. Lược đồ ngôi sao (Star schema).....	18
2. Lược đồ chòm sao (Constellation Schema).....	19
3. Các bảng Dim.....	20
3.1. DimGender.....	20
3.2. DimAgeBand.....	20
3.3. DimCasualtySeverity.....	20
3.4. DimCasualtyType .....	21
3.5. DimAccidentSeverity .....	21
3.6. DimLocalAuthorityDistrict.....	21
3.7. DimUrbanOrRuralArea.....	22
3.8. DimRoadType .....	22
3.9. DimVehicleType .....	22
3.12. DimJourneyPurpose .....	23
3.11. Dim AgeGroup .....	23
3.10. DimBuiltUpRoad .....	23
3.13. DimTimeOfDay.....	24

3.14. DimTownCity .....	24
3.15. DimCounty.....	24
3.16. DimRegion .....	25
3.17. DimCountry .....	25
4. Các bảng Fact.....	25
4.1. FactAccidents .....	25
4.2. FactVehicles.....	26
4.3. FactCasualties.....	26
<b>C. QUÁ TRÌNH SSIS .....</b>	<b>27</b>
<b>I. ETL Source to Stage .....</b>	<b>27</b>
1. Bảng đổ dữ liệu.....	27
2. Quy trình ETL .....	28
2.1. Control Flow của quy trình .....	28
2.2. Bảng Accidents (Accidents Data Flow) .....	29
2.3. Bảng Vehicles, Casualties (Vehicles, Casualties Data Flow) .....	29
2.4. Các bảng Constant (Constant Data Flows).....	31
2.5. Bảng Postcodes (Postcodes Data Flow) .....	32
2.6. Bảng LSOA (LSOA Data Flow) .....	32
<b>II. ETL Stage to NDS .....</b>	<b>33</b>
1. Lược đồ NDS và mô tả.....	33
2. Qui trình ETL .....	35
2.1. Các bảng master .....	35
2.2. Các bảng địa lý .....	39
2.3. Các bảng Transactions .....	46
<b>III. ETL NDS to Data Warehouse .....</b>	<b>53</b>
1. Lược đồ DDS và mô tả.....	53
3.1. DimGender.....	53
3.2. DimAgeBand.....	54
3.3. DimCasualtySeverity.....	54
3.4. DimCasualtyType .....	54

3.5. DimAccidentSeverity .....	55
3.6. DimLocalAuthorityDistrict.....	55
3.7. DimUrbanOrRuralArea.....	55
3.8. DimRoadType .....	56
3.9. DimVehicleType .....	56
3.10. DimJourneyPurpose .....	56
3.11. Dim AgeGroup .....	57
3.12. DimBuiltUpRoad .....	57
3.13. DimTimeOfDay.....	57
3.14. DimTownCity .....	58
3.15. DimCounty.....	58
3.16. DimRegion .....	58
3.17. DimCountry .....	59
2. Quy trình ETL .....	59
2.1 Bảng DimCountry(DimCountry Data Flow) .....	59
2.2 Bảng DimRegion(DimRegion Data Flow) .....	61
2.3 Bảng DimCounty(DimCounty Data Flow) .....	63
2.4 Các bảng chiều còn lại(Other Dimensions Data Flow) .....	64
2.5 Các bảng Fact(Fact Data Flow) .....	66
2.6. Kết quả thực thi .....	68
<b>D. QUÁ TRÌNH SSAS.....</b>	<b>71</b>
1. Tạo Data Source: .....	71
2. Tạo Data Source View: .....	72
3. Tạo Cube .....	73
4. Phân chiều.....	75
4.1. Chiều Date .....	75
4.2. Chiều Địa lý .....	76
4.3. Các chiều còn lại .....	77
5. Sử dụng DMX để truy vấn .....	78
<b>E. REPORT SỬ DỤNG POWER BI .....</b>	<b>84</b>

## LỜI CẢM ƠN

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

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

Xuất phát từ mục đích học tập, tìm hiểu sâu hơn các kiến thức về kho dữ liệu, cũng như tìm hiểu kỹ về quy trình nghiệp vụ của việc lên ý tưởng, xây dựng một kho dữ liệu. Nhóm chúng em đã thực hiện đồ án “Xây dựng kho dữ liệu về tai nạn ô tô tại Vương Quốc Anh”. Trong quá trình thực hiện đồ án, dựa trên kiến thức được Thầy cung cấp qua các buổi học lý thuyết cũng như thực hành trên lớp, kết hợp với việc tự tìm hiểu những công cụ và kiến thức mới, nhóm đã cố gắng thực hiện đồ án một cách tốt nhất.

Với vốn kiến thức được tiếp thu trong quá trình học không chỉ là nền tảng cho quá trình nghiên cứu đồ án mà còn là hành trang quý báu để em bước vào đời một cách vững chắc và tự tin.

Tuy nhiên trong quá trình nghiên cứu đề tài, mặc dù đã cố gắng hoàn thành tốt công việc được phân công trong phạm vi và khả năng cho phép nhưng chắc chắn không tránh khỏi những thiếu sót. Rất mong nhận được sự góp ý của Thầy để chương trình được hoàn chỉnh hơn, đồng thời bổ sung vốn kinh nghiệm cho nhóm trên con đường sắp tới.

Nhóm em xin chân thành cảm ơn!

## LÝ DO CHỌN ĐỀ TÀI

Nhóm chúng em đã chọn đề tài "Xây dựng kho dữ liệu về tai nạn ô tô tại Vương Quốc Anh" vì nhiều lý do quan trọng, không chỉ về mặt xã hội mà còn cung cấp thông tin cần thiết để cải thiện an toàn giao thông và giảm thiểu tai nạn ô tô.

Trước hết, Vương Quốc Anh đã ghi nhận một số lượng đáng kể các vụ tai nạn giao thông hàng năm, gây thiệt hại về người và tài sản. Xây dựng một kho dữ liệu chính xác và chi tiết về tai nạn ô tô sẽ giúp chúng ta hiểu rõ hơn về các yếu tố gây tai nạn, như điều kiện đường, thời tiết, tốc độ, và hành vi của người lái xe. Điều này cung cấp một cơ sở chắc chắn để đưa ra các biện pháp và chính sách hiệu quả hơn nhằm giảm thiểu nguy cơ xảy ra tai nạn giao thông.

Ngoài ra, việc xây dựng một kho dữ liệu về tai nạn ô tô tại Vương Quốc Anh sẽ giúp chúng ta phân tích và đánh giá hiệu quả của các biện pháp an toàn giao thông đã được triển khai. Chúng ta có thể xem xét các chiến lược đã được sử dụng để giảm tai nạn, như việc cải thiện hệ thống đường, đèn giao thông, tăng cường kiểm tra an toàn xe cộ và các biện pháp giáo dục người lái xe. Dựa trên dữ liệu thu thập được, chúng ta có thể đưa ra đánh giá và khuyến nghị các biện pháp cải tiến để nâng cao hiệu quả và an toàn giao thông.

Hơn nữa, việc xây dựng kho dữ liệu về tai nạn ô tô tại Vương Quốc Anh có thể đóng góp vào nỗ lực toàn cầu để nghiên cứu về an toàn giao thông. Dữ liệu thu thập được có thể được so sánh và phân tích cùng với các kho dữ liệu tương tự từ các quốc gia khác, giúp tìm ra các xu hướng và thực hành tốt nhất về an toàn giao thông trên toàn cầu. Qua đó, chúng ta có thể học hỏi từ kinh nghiệm của những nước khác và áp dụng những giải pháp đã được chứng minh hiệu quả vào thực tế của Vương Quốc Anh.

Tóm lại, xây dựng kho dữ liệu về tai nạn ô tô tại Vương Quốc Anh là một đề tài đáng quan tâm và có ý nghĩa vô cùng quan trọng. Việc thu thập và phân tích dữ liệu này sẽ giúp cải thiện an toàn giao thông, đưa ra các biện pháp và chính sách hiệu quả hơn, dự đoán và phòng ngừa tai nạn, và đóng góp vào nghiên cứu toàn cầu về an toàn giao thông.

## A. TẬP DỮ LIỆU

### 1. Giới thiệu bộ dữ liệu

- **Dữ liệu UK Car Accidents 2005 – 2015** (Nhóm chỉ lấy dữ liệu từ năm 2011 đến 2014 để thực hiện bài cuối kỳ)
  - Nguồn dữ liệu: <https://www.kaggle.com/datasets/silicon99/dft-accident-data>
  - Tổng quan: Tập dữ liệu này chứa thông tin về các vụ tai nạn xe hơi ghi nhận tại Vương quốc Anh trong khoảng 2005 - 2015. Nó cung cấp thông tin về vị trí, thời gian, nguyên nhân và hậu quả của các vụ tai nạn. Tập dữ liệu này thường được sử dụng cho mục đích nghiên cứu về an toàn giao thông và phân tích dữ liệu để đưa ra các biện pháp cải thiện an toàn đường bộ.
  - Tập dữ liệu UK Car Accidents 2005 – 2015 bao gồm 3 file csv:
    - + Accidents: Chứa thông tin cơ bản về các vụ tai nạn
    - + Casualties: Chứa thông tin về các phương tiện tham gia vào các vụ tai nạn
    - + Vehicles: Chứa thông tin về những người bị thương trong các vụ tai nạn
- **Dữ liệu LSOA-Postcode mapping:**
  - Nguồn dữ liệu: <https://geoportal.statistics.gov.uk/datasets/postcode-to-output-area-to-lower-layer-super-output-area-to-middle-layer-super-output-area-to-local-authority-district-august-2021-lookup-in-the-uk-1/about>
  - Tổng quan: "LSOA-Postcode mapping" là một tập dữ liệu sự ánh xạ giữa mã bưu điện và mã LSOA (Lower Layer Super Output Area) ở Vương quốc Anh. Mã bưu điện là một mã số được sử dụng để xác định khu vực địa lý nhất định, trong khi mã LSOA là một đơn vị địa lý nhỏ hơn được sử dụng cho mục đích thống kê và nghiên cứu.
- **Dữ liệu UK – Postcodes:**
  - Nguồn dữ liệu: <https://github.com/academe/UK-Postcodes/blob/master/postcodes.csv>
  - Tổng quan: Tập dữ liệu "UK - Postcodes" chứa thông tin về mã bưu điện, khu vực địa lý tương ứng, vị trí địa lý (như tọa độ địa lý), các thành phố hoặc khu vực quản lý, và các thuộc tính khác liên quan đến mã bưu điện.

## 2. Mô tả dữ liệu:

### 2.1. UK Car Accidents – Bảng Casualties

STT	Thuộc tính	Mô tả
1	Accident_Index	Số thứ tự chỉ mục các vụ tai nạn
2	Vehicle_Reference	Thứ tự của phương tiện của nạn nhân trong vụ tai nạn
3	Casualty_Reference	Số nạn nhân liên quan đến vụ tai nạn
4	Casualty_Class	Nạn nhân thuộc nhóm người nào. 1: Driver or rider 2: passenger 3: Pedestrian.
5	Sex_of_Casualty	Giới tính của nạn nhân 1: Male 2: Female -1: Data missing or out of range
6	Age_of_Casualty	Tuổi của nạn nhân
7	Age_Band_of_Casualty	Các nhóm tuổi thương vong - Giá trị 1: nhóm tuổi từ 0 đến 5 tuổi - Giá trị 2: nhóm tuổi từ 6 đến 10 tuổi - Giá trị 3: nhóm tuổi từ 11 đến 15 tuổi - Giá trị 4: nhóm tuổi từ 16 đến 20 tuổi - Giá trị 5: nhóm tuổi từ 21 đến 25 tuổi - Giá trị 6: nhóm tuổi từ 26 đến 35 tuổi - Giá trị 7: nhóm tuổi từ 36 đến 45 tuổi

		<ul style="list-style-type: none"> <li>- Giá trị 8: nhóm tuổi từ 46 đến 55 tuổi</li> <li>- Giá trị 9: nhóm tuổi từ 56 đến 65 tuổi</li> <li>- Giá trị 10: nhóm tuổi từ 65 đến 75 tuổi</li> <li>- Giá trị 11: nhóm tuổi từ 75 tuổi trở lên</li> <li>- Giá trị -1: Mất dữ liệu hoặc nằm ngoài phạm vi</li> </ul>
8	Casualty_Severity	<p>Mức độ nghiêm trọng của xảy ra với nạn nhân. Các giá trị biểu diễn kiểu số.</p> <p>1: Fatal 2: Serious 3: Slight</p>
9	Pedestrian_Location	<p>Vị trí đi bộ của nạn nhân</p> <p>0: not a Pedestrian 1: Crossing on pedestrian crossing facility 2: Crossing in zig-zag approach lines, ...</p>
10	Pedestrian_Movement	<p>Nạn nhân đi bộ di chuyển như thế nào.</p> <p>0: not a Pedestrian, 2: Crossing from nearside – masked by parked or stationary vehicle 1: Crossing from driver's nearside, ...</p>
11	Car_Passenger	<p>Nếu nạn nhân đi car. Thì nạn nhân ngồi ở vị trí nào trên xe.</p> <p>0: not car passenger 1: Front seat passenger</p>

		2: Near seat passenger,  -1: Data missing or out of range.
12	Bus_or_Coach_Passenger	Có phải nạn nhân khi đi xe bus hoặc xe khách không.  0: not a bus or coach passenger (nạn nhân là người đi bộ)  1: Boarding  2: Alighting (xuống xe), ...
13	Pedestrian_Road_Maintenance_Worker	Có phải nạn nhân là công nhân bảo trì đường cho người đi bộ  0: no/ not applicable,  1: Yes,  2: Not known,  -1: Data missing or out of range
14	Casualty_Type	Loại phương tiện mà nạn nhân sử dụng.  0: Pedestrian  1: Cyclist  2: Motorcycle 50cc and under rider or passenger,....
15	Casualty_Home_Area_Type	Loại khu vực mà nạn nhân sinh sống.  1: Urban Area  2: Small Town  3: Rural

## 2.2. UK Car Accidents – Bảng Vehicles

STT	Thuộc tính	Mô tả
1	Accident_Index	Số thứ tự chỉ mục các vụ tai nạn
2	Vehicle_Reference	Thứ tự của phương tiện trong vụ tai nạn
3	Vehicle_Type	<p>Chỉ ra loại phương tiện sử dụng, và được chia thành các loại sau đây.</p> <ul style="list-style-type: none"> <li>- Giá trị 1: Pedal cycle - xe đạp</li> <li>- Giá trị 2: Motorcycle 50cc and under – Xe mô tô 50cc trở xuống</li> <li>- Giá trị 3: Motorcycle 125cc and under – Xe mô tô 125cc trở xuống</li> <li>- Giá trị 4: Motorcycle over 125cc and up to 500cc - Xe mô tô trên 125cc đến 500cc</li> <li>- Giá trị 5: Motorcycle over 500cc - Xe máy trên 500cc</li> <li>- Giá trị 8: Taxi/Private hire car - Taxi / thuê xe riêng</li> <li>- Giá trị 9: Car - Xe hơi</li> <li>- Giá trị 10: Minibus (8 - 16 passenger seats) - Xe buýt nhỏ (8-16 chỗ ngồi)</li> <li>- Giá trị 11: Bus or coach (17 or more pass seats) - Xe buýt hoặc xe khách (17 chỗ ngồi trở lên)</li> <li>- Giá trị 16: Ridden horse - Ngựa hoang</li> </ul>

		<ul style="list-style-type: none"> <li>- Giá trị 17: Agricultural vehicle - Xe công nông</li> <li>- Giá trị 18: Tram - Xe điện</li> <li>- Giá trị 19: Van / Goods 3.5 tonnes mgw or under - Van / Hàng hóa 3,5 tấn mgw trở xuống</li> <li>- Giá trị 20: Goods over 3.5t. and under 7.5t - Hàng trên 3,5t. và dưới 7,5t</li> <li>- Giá trị 21: Goods 7.5 tonnes mgw and over - Hàng hóa 7,5 tấn mgw trở lên</li> <li>- Giá trị 22: Mobility scooter - Xe tay ga di động</li> <li>- Giá trị 23: Electric motorcycle - Xe máy điện</li> <li>- Giá trị 90: Other vehicle - Xe khác</li> <li>- Giá trị 97: Motorcycle - unknown cc - Xe máy - phân khối không xác định</li> <li>- Giá trị 98: Goods vehicle – unknown weight - Xe chở hàng - không rõ trọng lượng</li> </ul>
4	Vehicle_Manoeuvre	<p>Trước lúc tai nạn, phương tiện đã di chuyển như thế nào.</p> <p>1: Nearside</p> <p>7: Offside.</p>
5	Vehicle_Location_Restricted_Lane	Vị trí của xe có vi phạm các làn đường bị cấm đi hay không

6	Junction_Location	<p>Chỉ ra vị trí giao lộ của phương tiện tai nạn, và được chia thành các loại sau đây</p> <ul style="list-style-type: none"> <li>- Giá trị 0: Not at or within 20 metres of junction - Không ở hoặc trong vòng 20 mét của đường giao nhau</li> <li>- Giá trị 1: Approaching junction or waiting/parked at junction approach - Tiếp cận đường giao nhau hoặc chờ / đỗ ở đường giao nhau</li> <li>- Giá trị 2: Cleared junction or waiting/parked at junction exit - Đường giao nhau đã thông thoáng hoặc chờ / đỗ ở lối ra đường giao nhau</li> <li>- Giá trị 3: Leaving roundabout - Rời khỏi bùng binh</li> <li>- Giá trị 4: Entering roundabout - Đi vào bùng binh</li> <li>- Giá trị 5: Leaving main road - Ra khỏi đường chính</li> <li>- Giá trị 6: Entering main road - Vào đường chính</li> <li>- Giá trị 7: Entering from slip road - Đi vào từ đường trượt</li> <li>- Giá trị 8: Mid Junction - on roundabout or on main road - Giao lộ Giữa – trên bùng binh hoặc trên đường chính</li> </ul>
---	-------------------	---

		<ul style="list-style-type: none"> <li>- Giá trị -1: Data missing or out of range</li> <li>- dữ liệu bị mất hoặc nằm ngoài phạm vi</li> </ul>
7	Skidding_and_Overturning	<p>Chỉ ra phương tiện bị trượt hoặc lật, và được chia thành các loại sau đây.</p> <p>0: None</p> <p>1: Skidded</p> <p>2: Skidded and overturned</p> <p>3: Jackknifed</p> <p>4: Jackknifed and overturned</p> <p>5: Overturned</p> <p>-1: Data missing or out of range</p>
8	Hit_Object_in_Carriageway	<p>Xe đã tông vào vật thể gì trên tuyến đường đi.</p> <p>Các giá trị 1-12 thể hiện các vật thể mà xe tông phải.</p> <p>-1: dữ liệu null hoặc ngoài khoảng giá trị.</p> <p>0: không.</p>
9	Vehicle_Leaving_Carriageway	<p>Xe văng khỏi đường đi như thế nào. Các giá trị được ghi trong code book.</p> <p>Ví dụ</p> <p>1: Nearside (về phía bên trái, gần với giải phân cách)</p> <p>7: Offside (về phía bên phải, gần với vạch kẻ đường ở giữa)</p>

		Lưu ý: UK đi xe bên tay trái.
10	Hit_Object_off_Carriageway	Xe đã tông vào vật thể gì bên ngoài tuyến đường. Các giá trị biểu diễn bằng số. Mỗi số tương ứng với một vật thể.
11	1st_Point_of_Impact	Điểm đầu tiên của phương tiện bị ảnh hưởng.  0: did not impact  1: Front  2: Back, ...
12	Was_Vehicle_Left_Hand_Drive?	Chỉ ra người điều khiển phương tiện có lái xe bằng tay trái hay không.  1: No  2: Yes  -1: Data missing or out of range - dữ liệu bị thiếu hoặc nằm ngoài phạm vi
13	Journey_Purpose_of_Driver	Chỉ ra mục đích của hành trình của người điều khiển phương tiện. và được chia thành các loại sau đây.  1: Journey as part of work  2: Commuting to/from work – Đi làm đến / đi từ nơi làm việc  3: Taking pupil to/from school - Đưa học sinh đến / từ trường  4: Pupil riding to/from school – Đưa đón học sinh  5: Other - khác

		<p>6: Not known - chưa biết</p> <p>15: Other/Not known (2005-10) - khác/ chưa biết</p> <p>-1: Data missing or out of range - dữ liệu bị thiếu hoặc ngoài phạm vi</p>
14	Sex_of_Driver	<p>Giới tính của tài xế. Giới tính được biểu diễn bằng số tương đương với các giá trị sau:</p> <p>1: Male</p> <p>2: Female</p> <p>3: not known</p> <p>-1: Data missing or out of range</p>
15	Age_of_Driver	Tuổi hiện tại của tài xế
16	Age_Band_of_Driver	<p>Thuộc tính chỉ ra nhóm tuổi của nạn nhân, được lưu trữ theo quy luật code từ 1-11 (code – độ tuổi).</p> <p>1: 0 - 5</p> <p>2: 6 - 10</p> <p>3: 11 – 15</p> <p>4: 16 – 20</p> <p>5: 21 – 25</p> <p>6: 26 – 35</p> <p>7: 36 – 45</p> <p>8: 46 – 55</p> <p>9: 56 – 65</p>

		10: 66 – 75  1: Over 75  -1: Data missing or out range
17	Engine_Capacity_(CC)	Dung tích của động cơ. Đơn vị đo là CC
18	Propulsion_Code	Động cơ sẽ chạy bằng loại nhiên liệu nào.  1: Petrol  2: Heavy oil  M: undefined
19	Age_of_Vehicle	Tuổi hiện tại của phương tiện
20	Driver_IMD_Decile	Thể hiện mức độ khó khăn thiếu thốn của khu vực mà tài xế sinh sống. Các giá trị được biểu diễn từ -1-10 thể hiện ở các mức độ.  1: nhiều nhất là 10%  9: ít nhất 10-20%
21	Driver_Home_Area_Type	Loại khu vực mà tài xế sinh sống.  1: Urban Area,  2: Small Town,  3: Rural.

### 2.3. UK Car Accidents – Bảng Accidents

STT	Thuộc tính	Mô tả
1	Accident_Index	Số thứ tự chỉ mục các vụ tai nạn
2	Police_Force	Mã lực lượng cảnh sát
3	Accident_Severity	Đây là thuộc tính chỉ ra mức độ nghiêm trọng của 1 vụ tai nạn, gồm có 3 mức độ là (fatal, serious, slight).
4	Day_of_Week	Ngày trong tuần: Sunday, Monday, ...
5	Local_Authority_(District)	Tên khu vực địa phương xảy ra vụ việc
6	Local_Authority_(Highway)	Tên đường chính xảy ra tai nạn
7	1st_Road_Class	Đường cấp 1
8	Road_Type	Loại đường.  Ví dụ: One way street (Đường 1 chiều)
9	Junction_Detail	Chi tiết giao lộ:  - Bùng binh  - Ngả tư
10	Junction_Control	Kiểm soát giao lộ:  - Người có quyền  - Tín hiệu giao thông
11	2nd_Road_Class	Đường cấp 2
12	Pedestrian_Crossing_Human_Control	Người kiểm soát phân luồng đưa người qua đường
13	Pedestrian_Crossing_Physical_Facilities	Cơ sở vật chất dành cho người đi bộ

14	Light_Conditions	Các điều kiện ánh sáng Ví dụ: Bóng tối – đèn không sáng
15	Weather_Conditions	Các điều kiện thời tiết Ví dụ: Gió mạnh - mưa
16	Road_Surface_Conditions	Các điều kiện mặt đường Ví dụ: Khô, ẩm, ...
17	Special_Conditions_at_Site	Các điều kiện đặc biệt ở hiện trường Ví dụ: Mặt đường bị lỗi
18	Carriageway_Hazards	Các mối nguy hiểm trên đường đi. Ví dụ: Chó trên đường, tai nạn trước đó, ...
19	Urban_or_Rural_Area	Thành thị hay nông thôn
20	Did_Police_Officer_Attend_Scene_of_Accident	Cảnh sát có tham dự hiện trường vụ tai nạn không
21	Time	Thời gian xảy ra vụ tai nạn
22	LSOA_of_Accident_Location	Mã địa điểm tai nạn

#### 2.4. Bảng Postcodes (UK Postcodes)

STT	Thuộc tính	Mô tả
1	city	Thị trấn(town) hoặc thành phố (city)
2	county	Quận(the county) hoặc khu vực đô thị (city conurbation) hoặc khu vực thành phố (municipal region).
3	country_code	Mã quốc gia gồm 3 chữ cái

4	country_name	Tên của quốc gia
5	region_code	Mã vùng (không chính thức )
6	region_name	Tên vùng
7	postcode	Mã bưu điện
8	easting	Giá trị hướng đông - tối đa 7 chữ số
9	northing	Giá trị hướng bắc - tối đa 7 chữ số
10	latitude	Góc vĩ độ tính bằng độ
11	longitude	Góc kinh độ tính bằng độ
12	iso3166-2	Mã quốc gia ISO

#### 2.5. Bảng PCD\_OA\_LSOA\_MSOA\_LAD\_AUG21\_UK\_LU

STT	Thuộc tính	Mô tả
1	pcd7	Mã postcode 7 ký tự
2	pcd8	Mã postcode 8 ký tự
3	pcds	Khoảng cách giữa quận và đơn vị hành
4	dointr	Ngày postcode có hiệu lực
5	doterm	Ngày postcode hết hiệu lực
6	usertype	Loại người dùng
7	oa11cd	Mã Output Area dân số trung bình khu vực vào khoảng 310 người
8	lsoa11cd	Mã Lower layer super output area dân số vào khoảng 1500

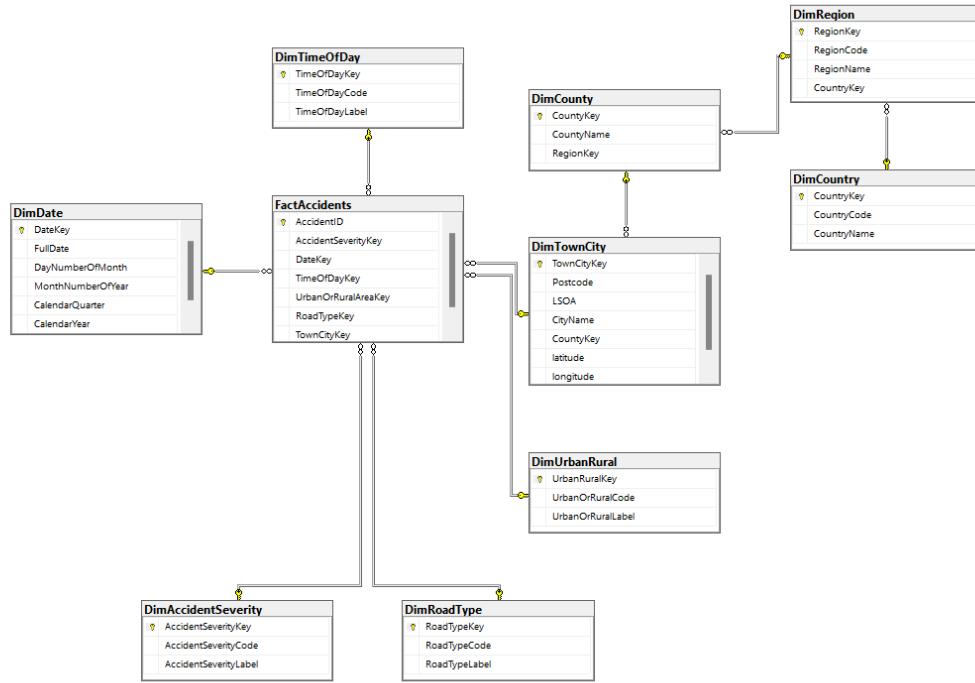
9	msoa11cd	Mã Middle layer super output area dân số vào khoảng 7500
10	ladcd	Mã vùng đất
11	lsoa11nm	Tên của Lower layer super output area
12	msoa11nm	Tên của Middle layer super output area
13	adnm	Tên vùng đất

### 3. Công cụ hỗ trợ

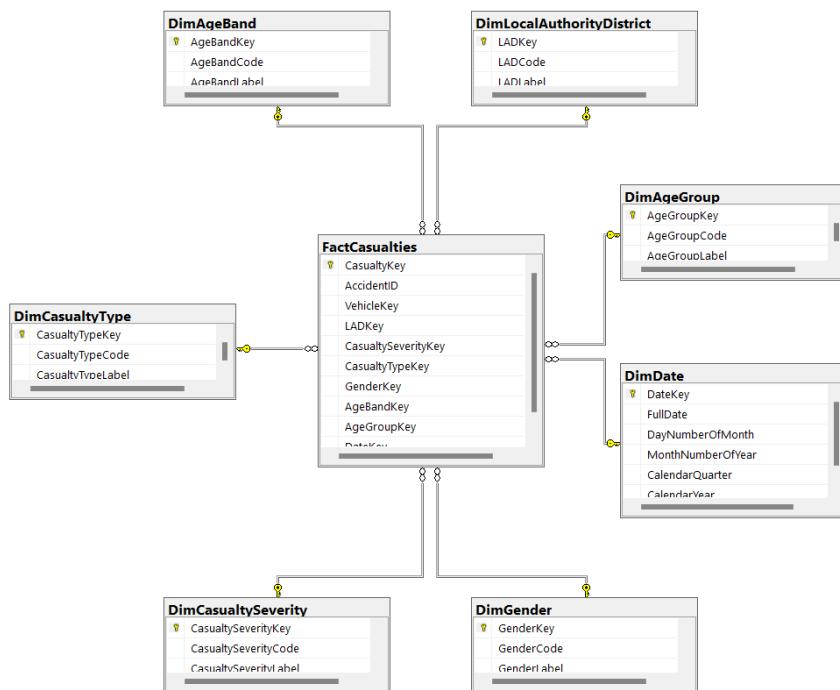
- Visual Studio 2019/ 2022 tích hợp SSIS (SQL Server Integration Services) và SSAS (SQL Server Analysis Services).
- SQL Server Management Studio, SQL Server 2019
- Power BI.

## B. DATA WAREHOUSE

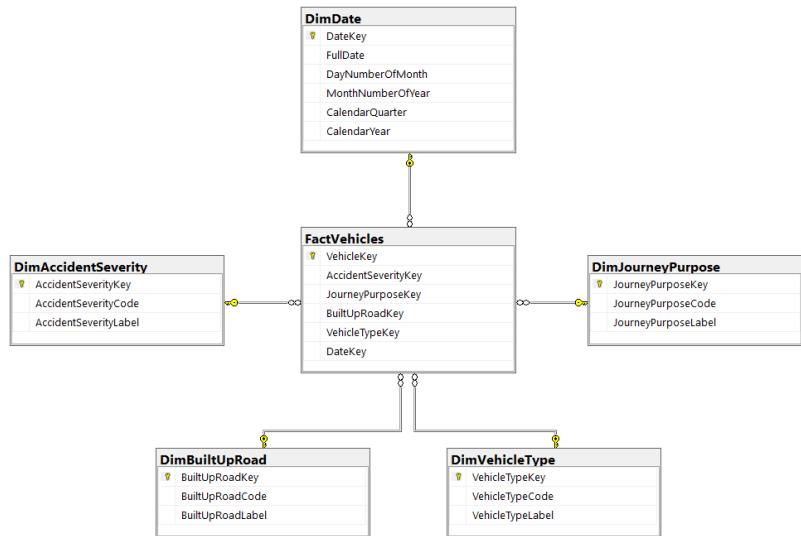
### 1. Lược đồ ngôi sao (Star schema)



Hình 1 - Lược đồ ngôi sao bảng Fact Accidents

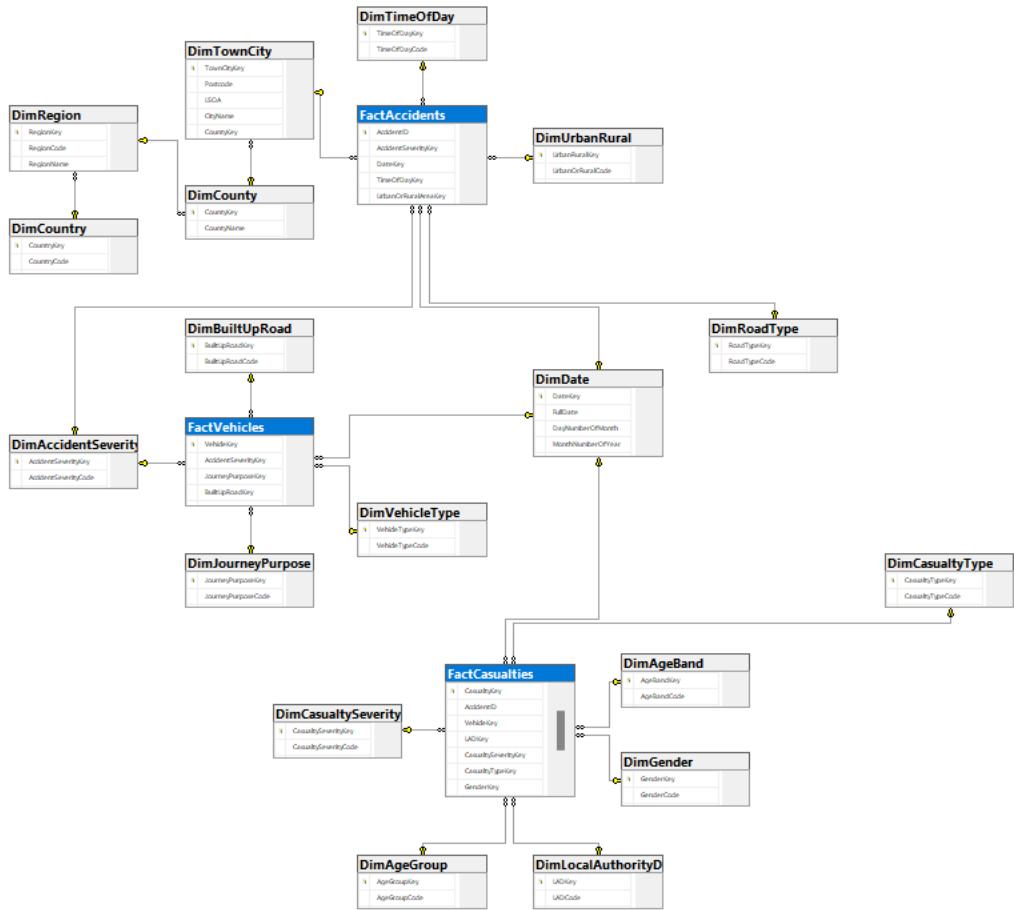


Hình 2 - Lược đồ ngôi sao bảng Fact Casualties



Hình 3 - Lược đồ ngôi sao bảng Fact Vehicles

## 2. Lược đồ chòm sao (Constellation Schema)



### 3. Các bảng Dim

#### 3.1. DimGender

Thuộc tính	Mô tả
GenderKey int identity(1,1),	Mã tự tăng của bảng giới tính
Code int,	Mã giới tính
Label varchar(100),	Nhãn giới tính
CONSTRAINT PK_G PRIMARY KEY(GenderKey)	Set GenderKey là PK

#### 3.2. DimAgeBand

Thuộc tính	Mô tả
AgeBandKey int identity(1,1),	Mã tự tăng bảng nhóm độ tuổi (trong khoảng từ x → y)
Code int,	Mã nhóm tuổi
Label varchar(100),	Nhãn
CONSTRAINT PK_AB PRIMARY KEY(AgeBandKey)	Set AgeBandKey là Pk

#### 3.3. DimCasualtySeverity

Thuộc tính	Mô tả
CasualtySeverityKey int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã độ nghiêm trọng của nạn nhân
CONSTRAINT PK_CS PRIMARY KEY(CasualtySeverityKey)	Set CasualtySeverityKey là PK

### *3.4. DimCasualtyType*

Thuộc tính	Mô tả
CasualtyTypeKey int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã loại tai nạn
Label varchar(100),	Nhãn
CONSTRAINT PK_CT PRIMARY KEY(CasualtyTypeKey)	Set CasualtyTypeKey là PK

### *3.5. DimAccidentSeverity*

Thuộc tính	Mô tả
AccidentSeverityIndex int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã độ nghiêm trọng của vụ tai nạn
Label varchar(20),	Nhãn
CONSTRAINT PK_AS PRIMARY KEY(AccidentSeverityKey)	Set AccidentSeverityKey là PK

### *3.6. DimLocalAuthorityDistrict*

Thuộc tính	Mô tả
LADKey int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã địa phương
Label varchar(50),	Nhãn
CONSTRAINT PK_LAD PRIMARY KEY(LADKey)	Set LADKey là PK

### *3.7. DimUrbanOrRuralArea*

Thuộc tính	Mô tả
UrbanOrRuralKey int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã nơi xảy ra tai nạn (Thành thị hay nông thôn)
Label varchar(10),	Nhãn
CONSTRAINT PK_UORA PRIMARY KEY(UrbanOrRuralKey)	Set UrbanOrRuralKey là PK

### *3.8. DimRoadType*

Thuộc tính	Mô tả
RoadTypeKey int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã loại đường
Label varchar(50),	Nhãn
CONSTRAINT PK_RT PRIMARY KEY(RoadTypeKey)	Set RoadTypeKey là PK

### *3.9. DimVehicleType*

Thuộc tính	Mô tả
VehicleTypeIndex int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã loại phương tiện
Label varchar(100),	Nhãn
CONSTRAINT PK_VT PRIMARY KEY(VehicleTypeKey)	Set VehicleTypeKey là PK

### *3.12. DimJourneyPurpose*

Thuộc tính	Mô tả
JourneyPurposeKey int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã chuyển đi
Label varchar(100),	Nhãn
CONSTRAINT PK_JP PRIMARY KEY(JourneyPurposeKey)	Set JourneyPurposeKey là PK

### *3.11. Dim AgeGroup*

Thuộc tính	Mô tả
AgeGroupKey int identity(1,1),	Mã tự tăng của bảng nhóm tuổi
Code int,	Mã nhóm tuổi
Label varchar(100),	Nhãn
CONSTRAINT PK_AG PRIMARY KEY(AgeGroupKey)	Set AgeGroupKey là PK

### *3.10. DimBuiltUpRoad*

Thuộc tính	Mô tả
BuiltUpRoadKey int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã đường nơi xảy ra tai nạn (Có phải khu vực đông dân cư không)
Label varchar(100),	Nhãn
CONSTRAINT PK_BUR PRIMARY KEY(BuiltUpRoadKey)	Set BuiltUpRoadKey là PK

### *3.13. DimTimeOfDay*

Thuộc tính	Mô tả
TimeOfDayIndex int identity(1,1)	Mã tự tăng (SK)
Code int,	Mã của thời gian trong ngày
Label varchar(100),	Nhãn

### *3.14. DimTownCity*

Thuộc tính	Mô tả
TownCityIndex int identity(1,1),	Mã tự tăng
Postcode varchar(10),	mã bưu điện
lsoa varchar(10),	Mã địa lý
CityName varchar(50),	Tên thành phố
County int,	Mã hạt/ tỉnh
CONSTRAINT PK_TC PRIMARY KEY(TownCityKey)	Set TownCityKey là PK

### *3.15. DimCounty*

Thuộc tính	Mô tả
CountyIndex int identity(1,1),	Mã tự tăng của bảng Hạt/ tỉnh
CountyName varchar(50),	Tên hạt/ tỉnh (đơn vị hành chính lớn nhất ở Anh)
Region int,	Mã vùng/ miền
CONSTRAINT PK_C PRIMARY KEY(CountyIndex)	set CountyKey là PK

### *3.16. DimRegion*

Thuộc tính	Mô tả
RegionIndex int identity(1,1),	Mã tự tăng của bảng vùng/miền
RegionCode varchar(10),	Mã vùng/ miền
RegionName varchar(50),	Tên vùng
Country int,	Mã quốc gia của vùng
CONSTRAINT PK_R PRIMARY KEY(RegionIndex)	Set RegionKey là PK

### *3.17. DimCountry*

Thuộc tính	Mô tả
CountryIndex int identity(1,1),	Mã tự tăng của quốc gia
CountryCode varchar(10),	Mã quốc gia theo data
CountryName varchar(20),	Tên quốc gia
CONSTRAINT PK_CY PRIMARY KEY(CountryIndex)	Set CountryKey là PK

## 4. Các bảng Fact

### *4.1. FactAccidents*

Thuộc tính	Mô tả
AccidentID int ,	Mã tai nạn (SK)
DateKey,	Mã Thời gian
AccidentSeverityKey int,	Mã Mức độ nghiêm trọng của vụ tai nạn
TownCityKey int,	Mã thành phố
TimeOfDay int	Mã khung thời gian trong ngày
RoadTypeKey int,	Mã loại đường
UrbanOrRuralAreaKey int,	Mã thành thị hay nông thôn
CONSTRAINT PK_A PRIMARY KEY(AccidentID)	Set AccidentID là PK

#### 4.2. FactVehicles

Thuộc tính	Mô tả
VehicleKey int,	Mã phương tiện (SK)
VehicleTypeKey int,	Mã loại phương tiện
JourneyPurposeKey int,	Mã Mục đích chuyến
AccidentSeverityKey int,	Mã mức độ nghiêm trọng
BuiltUpRoadKey int	Mã biển thị tốc độ trên 50 hay dưới 50
DateKey	Mã thời gian
CONSTRAINT PK_V PRIMARY KEY(VehicleKey)	Set VehicleKey là PK

#### 4.3. FactCasualties

Thuộc tính	Mô tả
CasualtyKey int,	Mã nạn nhân (SK)
AccidentID int,	ID của vụ tai nạn
VehicleKey int,	Của phương tiện nào
LADKey int,	Mã quận
GenderKey int,	Giới tính của nạn nhân bị tai nạn
AgeBandKey int,	Độ tuổi của nạn nhân bị tai nạn
AgeGroupKey int,	Mã nhóm tuổi của nạn nhân (Children: 0-15, Young adult: 16-17, Adult: 18-59, 60 and over)
CasualtySeverityKey int,	Mã mức độ nghiêm trọng của nạn nhân
CasualtyTypeKey int,	Mã loại tai nạn
DateKey	Mã thời gian
CONSTRAINT PK_CST PRIMARY KEY(CasualtyKey)	Set CasualtyKey là PK

## C. QUÁ TRÌNH SSIS

### I. ETL Source to Stage

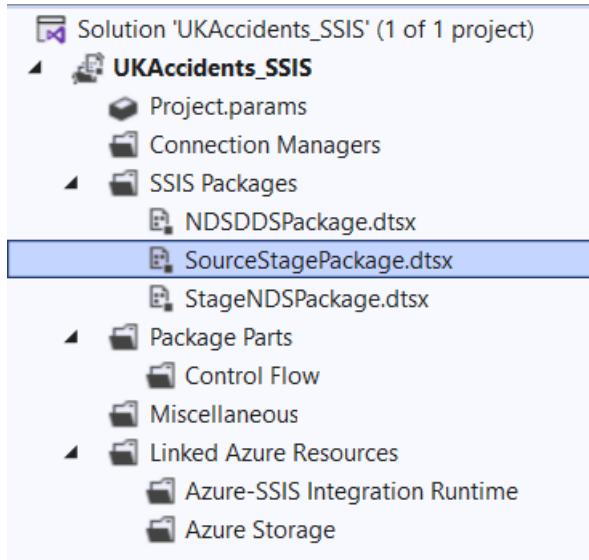
Source to Stage là quá trình đưa dữ liệu từ các file dữ liệu gốc (Source) vào Staging Area.

#### 1. Bảng đổ dữ liệu

Source ID	Source	Stage tables	Thuộc tính (ghi * nếu không có thay đổi)
1	Accidents.csv	Accidents	*
1	Vehicles.csv	Vehicles	*
1	Casualties.csv	Casualties	*
1	UK Accidents - Codebook.xlsx	Gender	code, label
		AgeBand	code, label
		CasualtySeverity	code, label
		CasualtyType	code, label
		AccidentSeverity	code, label
		LocalAuthorityDistrict	code, label
		UrbanOrRuralArea	code, label
		RoadType	code, label
		VehicleType	code, label
		JourneyPurpose	code, label
		AgeGroup	code, label
		BuiltUpRoad	code, label
		TimeOfDay	code, label
2	Postcodes.csv	Postcodes	postcode, city, county, country_code, country_name, region_code, region_name
2	PCD_OA_LSOA_MSOA_LA D_AUG21_UK_LU.csv	LSOA_PCD	lsoa11cd (mã LSOA), pcd7 (postcode 7 ký tự)

## 2. Quy trình ETL

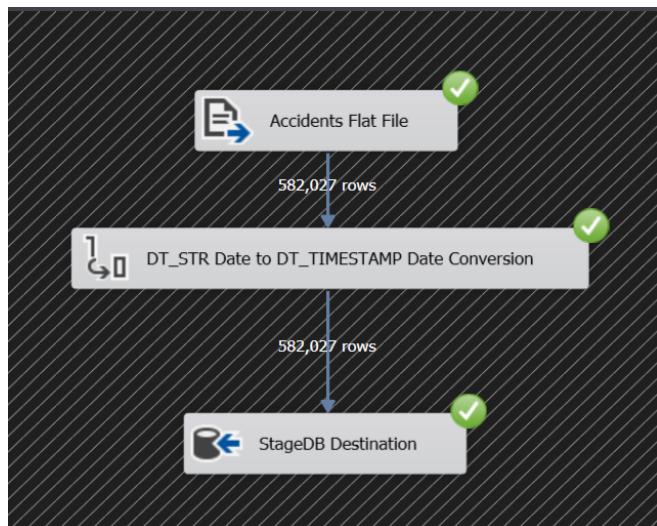
Package của quy trình có tên là SourceStagePackage. Ở giai đoạn này, thực hiện đổ dữ liệu từ Source vào Staging Area.



### 2.1. Control Flow của quy trình



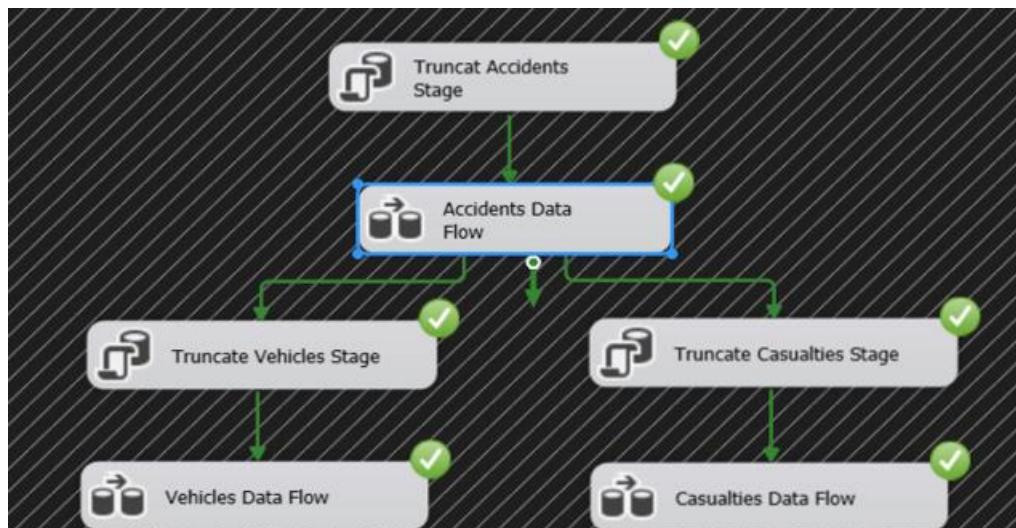
## 2.2. Bảng Accidents (Accidents Data Flow)



Nhóm em sử dụng component **Data Conversion** để thực hiện chuyển đổi kiểu dữ liệu của thuộc tính **Date** trong bảng **Accidents** từ **DT\_STR** sang **DT\_TIMESTAMP**.

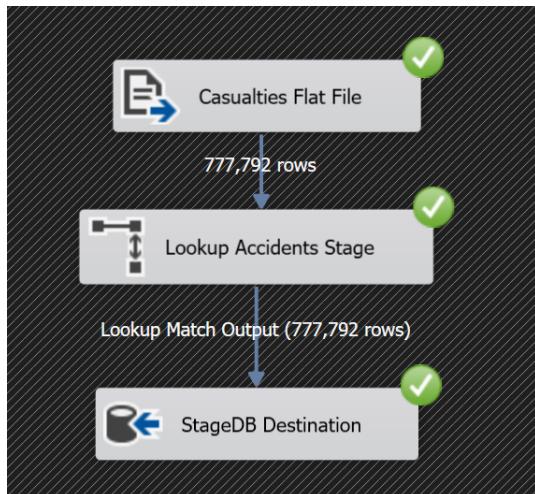
## 2.3. Bảng Vehicles, Casualties (Vehicles, Casualties Data Flow)

Bảng Vehicles và Casualties là 2 bảng con của Accidents nên sẽ được đổ sau khi bảng Accidents hoàn thành đổ dữ liệu.

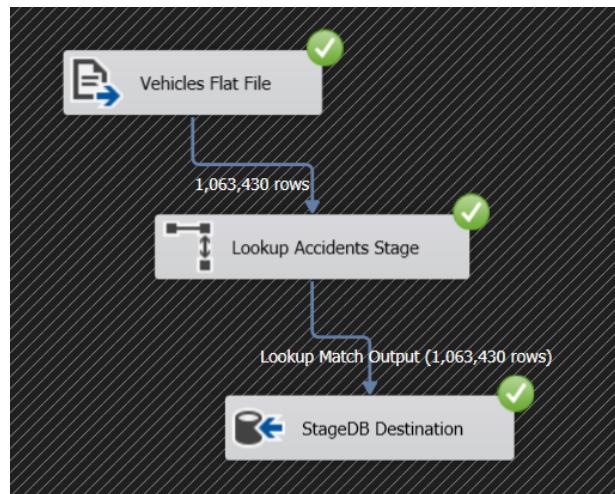


Tương tự như Accidents, 2 bảng này cũng được truncate trước (truncate trong Stage database) trước khi được đổ dữ liệu.

- Bên trong Data Flow của Vehicles (Casualties tương tự):



*Data Flow của Vehicles*



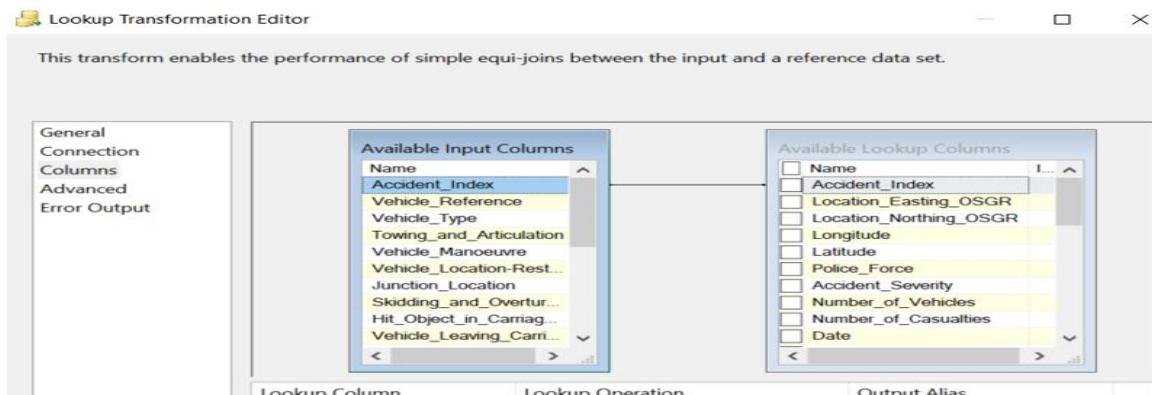
*Data Flow của Casualties*

Ở đây nhóm chúng em sử dụng component **Lookup**.

Lý do sử dụng component này là:

- Bảng Vehicles và Casualties không có nhãn thời gian riêng, phải phụ thuộc vào bảng Accidents qua Accident\_index.
- Để kiểm tra đảm bảo Accidents đã tồn tại trong Stage trước khi đổ Vehicles và Casualties, tránh để dư thừa dữ liệu và tốn chi phí thời gian. Ví dụ như chỉ lấy mốc từ 2011-2012 nhưng đổ dữ liệu Casualties từ 2011-2014, dư thừa dữ liệu 2 năm.

Ở đây Vehicles.Accident\_index ánh xạ đến Accidents.Accident\_index:



## 2.4. Các bảng Constant (Constant Data Flows)

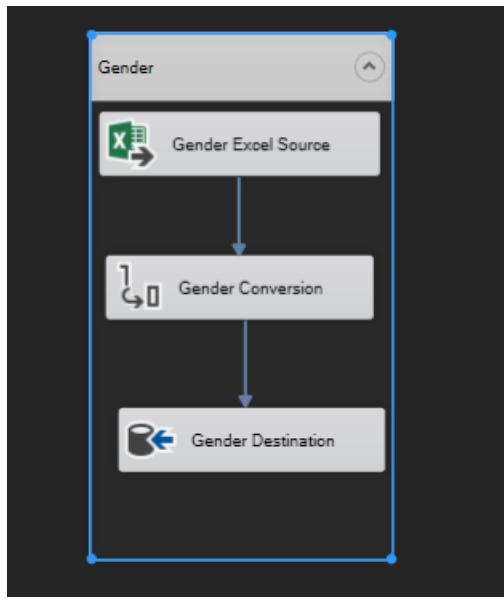
Các bảng Constant trong Stage sẽ trở thành các Dimension trong Data Warehouse sau này.



Data Flow của các bảng:



Ví dụ luồng Gender



Dữ liệu được lấy từ tập tin **UK Accidents - Codebook.xls** nên bên trong mỗi luồng sẽ sử dụng source là **Excel Source**.

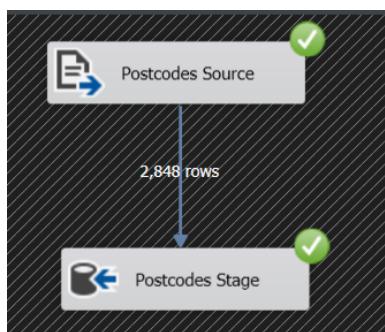
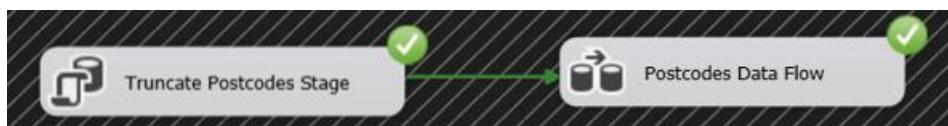
Tiếp đến sử dụng component **Data Conversion** để chuyển đổi kiểu dữ liệu.

Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Page
code	Copy of code	single-byte signed integer				
label	Copy of label	string [DT_STR]	100			1252 (ANSI - Latin I)

Cuối cùng là component **OLE DB Destination** để định nghĩa đích đến cho dữ liệu.

### 2.5. Bảng Postcodes (Postcodes Data Flow)

Tương tự như bảng Vehicles nhưng không cần Lookup tồn tại trong Data Flow.



### 2.6. Bảng LSOA (LSOA Data Flow)



Sử dụng component **Flat File Source** và chỉ lọc ra những thuộc tính cần thiết là **pcd7** và **lsoa11cd**.

Sử dụng component **Derived Column** để xử lý loại bỏ dấu nháy kép trong dữ liệu của thuộc tính pcd7 và lsoa11cd

Derived Column Name	Derived Column	Expression	Data Type	Length	Precision	Scale	Code Page
lsoa11cd	Replace 'lsoa11cd'	REPLACE(lsoa11cd,"\"", "")	string [DT_STR]	50			1252 (ANSI - Latin I)
pcd7	Replace 'pcd7'	REPLACE(pcd7,"\"", "")	string [DT_STR]	50			1252 (ANSI - Latin I)

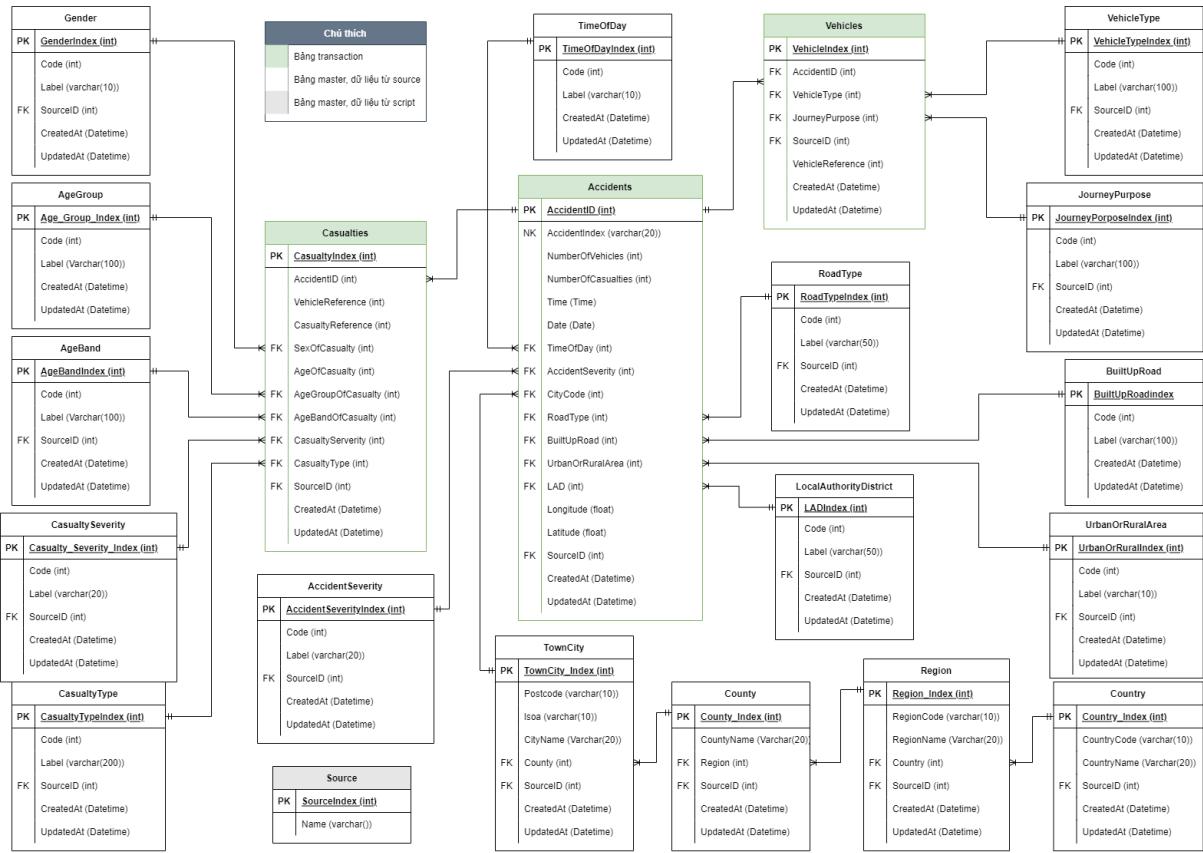
Sau đó đổ dữ liệu vào **OLE DB Destination**.

## II. ETL Stage to NDS

### 1. Lược đồ NDS và mô tả

Bảng danh sách các bảng trong NDS

Source ID	Stage table(s)	NDS table	Type
1	Gender	Gender	Master
1	Age_Band	AgeBand	Master
1	Casualty_Severity	CasualtySeverity	Master
1	Casualty_Type	CasualtyType	Master
1	Accident_Severity	AccidentSeverity	Master
1	Local_Authority_District	LocalAuthorityDistrict	Master
1	Urban_or_Rural_Area	UrbanOrRuralArea	Master
1	Road_Type	RoadType	Master
1	Vehicle_Type	VehicleType	Master
1	Journey_Purpose	JourneyPurpose	Master
1	Accidents	Accidents	Transaction
1	Vehicles	Vehicles	Transaction
1	Casualties	Casualties	Transaction
2	Postcodes + LSOA_PCD	TownCity	Master
2	Postcodes + LSOA_PCD	County	Master
2	Postcodes + LSOA_PCD	Region	Master
2	Postcodes + LSOA_PCD	Country	Master
-	-	Source	Master
1	Age_Group	AgeGroup	Master
1	Built_up_Road	BuiltUpRoad	Master
1	Time_of_Day	TimeOfDay	Master

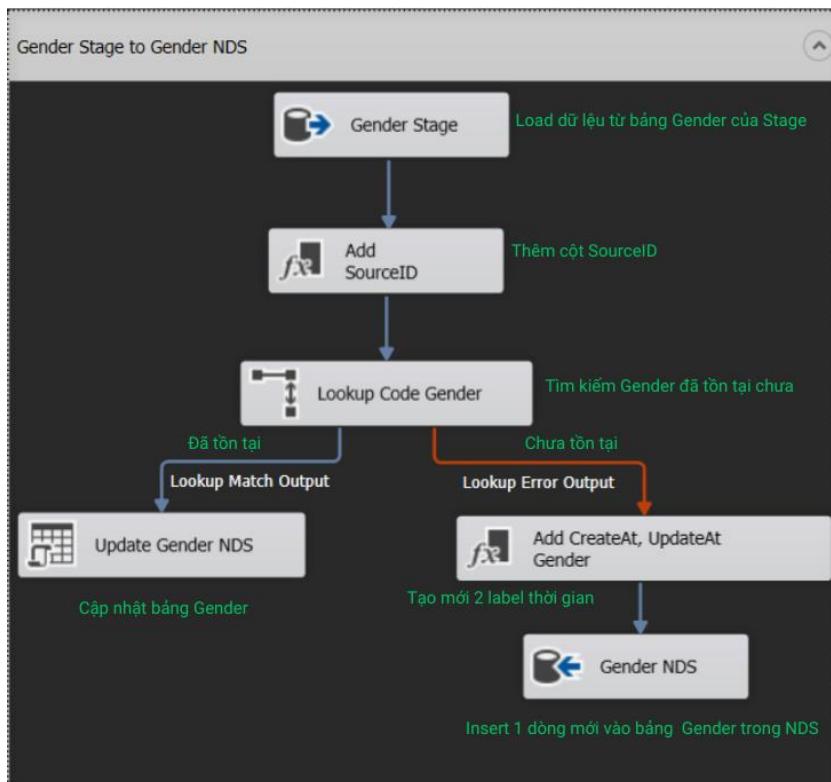


## 2. Qui trình ETL

### 2.1. Các bảng master

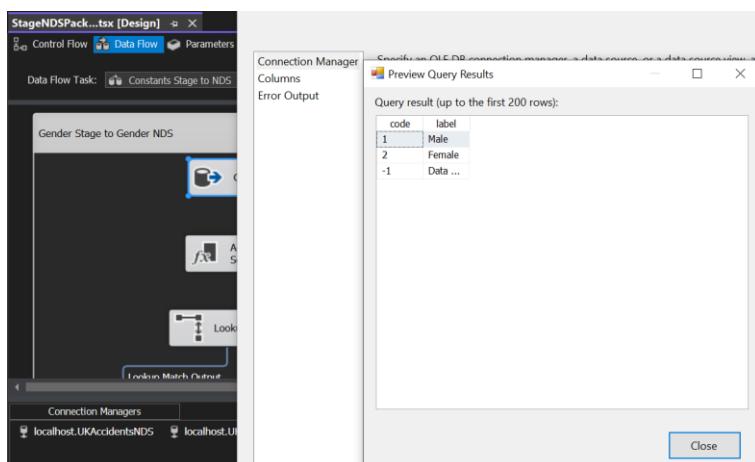
Bao gồm các bảng Gender, AgeGroup, AgeBand, CasualtySeverity, CasualtyType, AccidentSeverity, LocalAuthorityDistrict, UrbanOrRuralArea, BuiltUpRoad, RoadType, JourneyPurpose, VehicleType

Ví dụ minh họa nạp bảng Gender

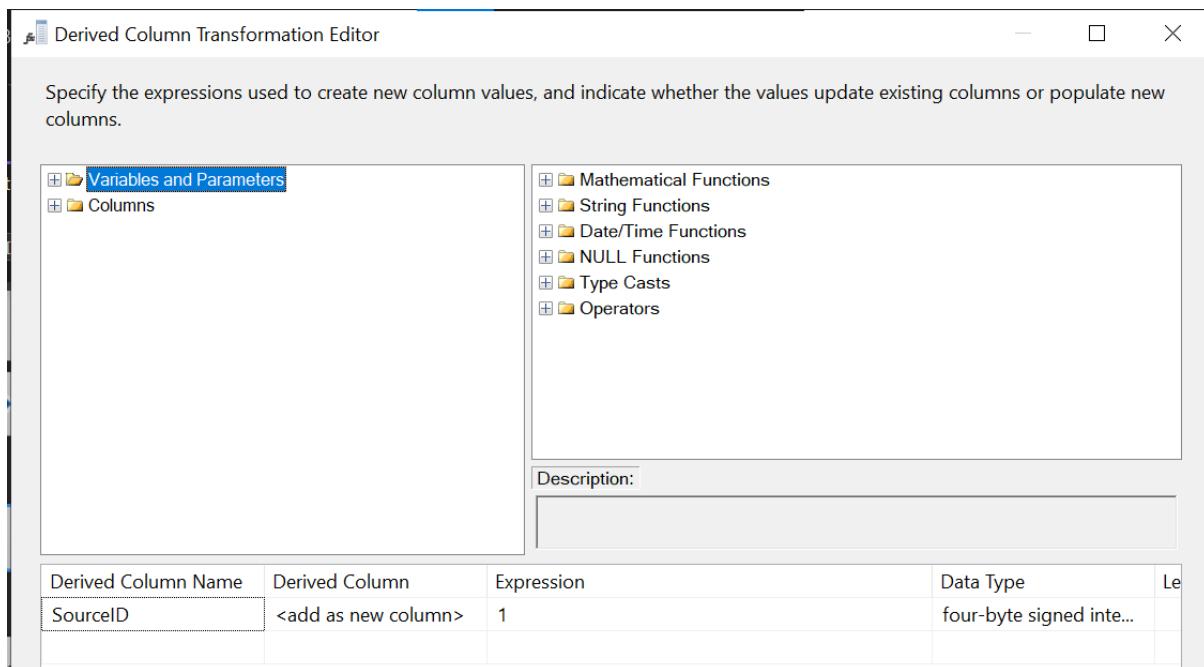


#### 2.1.1. Bảng Stage

Thiết lập bảng Gender Stage

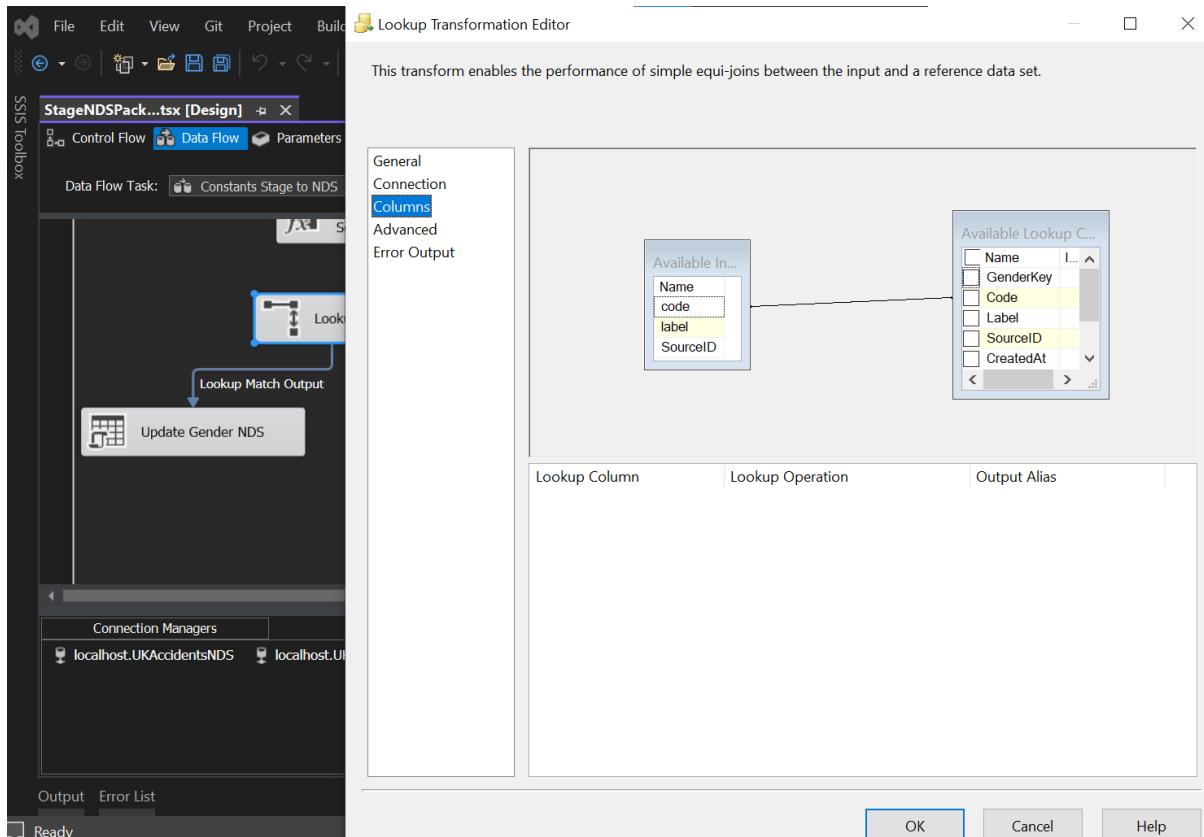


## 2.1.2. Thêm Source ID



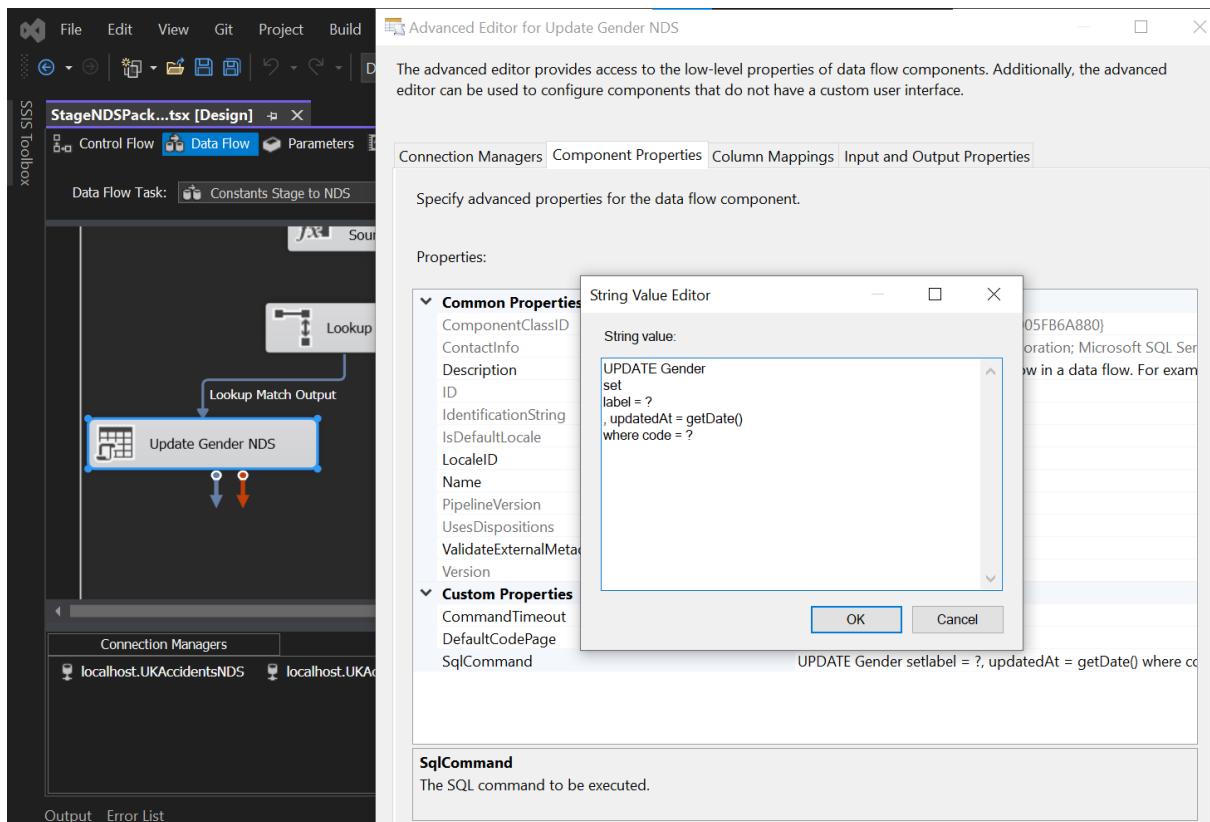
Giá trị SourceID = 1 là lấy dữ liệu trong file Accident\_CodeBook

## 2.1.3. Look up Code Gender

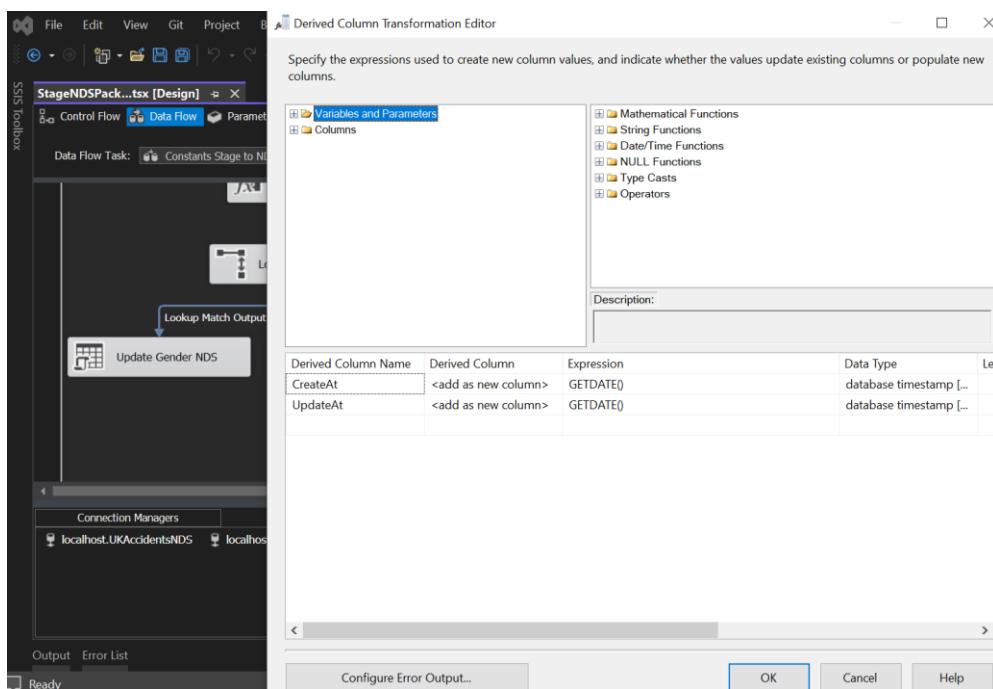


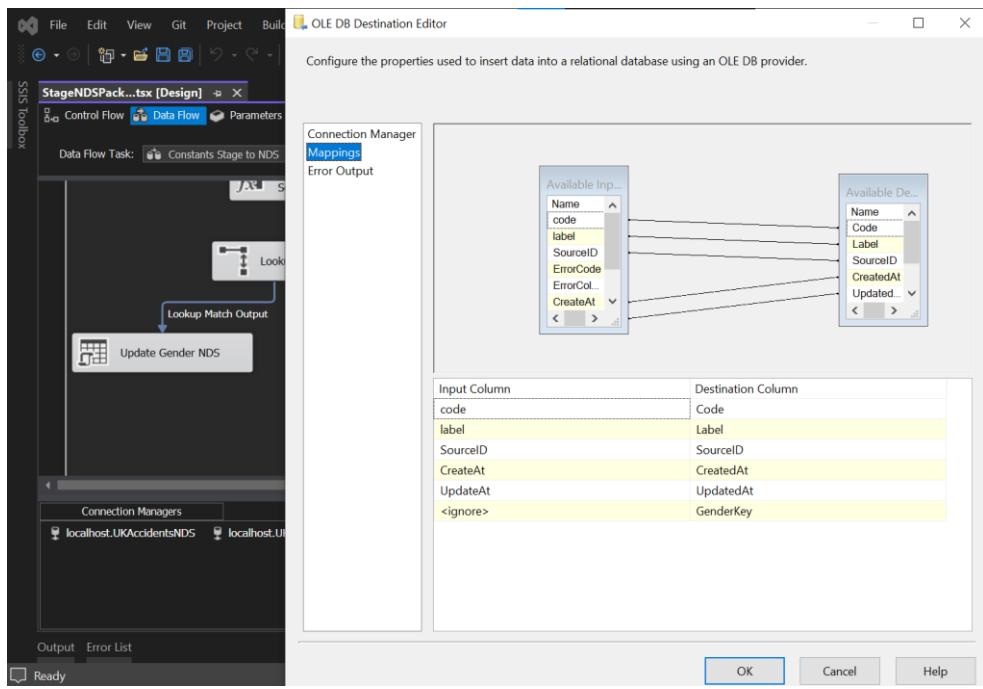
## 2.1.4. Xử lý kết quả lookup

Trường hợp Code Gender đã tồn tại thì update dữ liệu



Trường hợp chưa tồn tại thì thêm nhãn thời gian và insert dòng dữ liệu mới



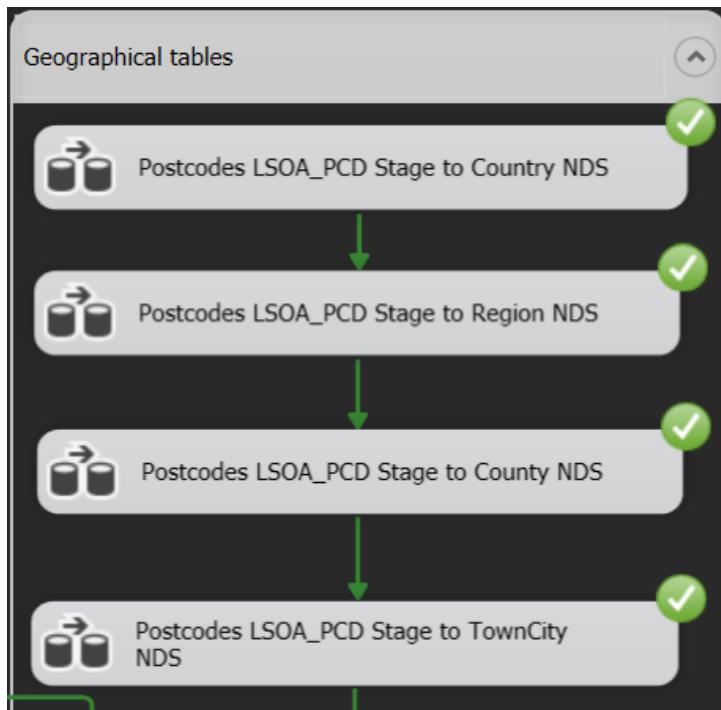


Các bảng còn lại tương tự Gender

#### 2.1.5. Kết quả thực thi



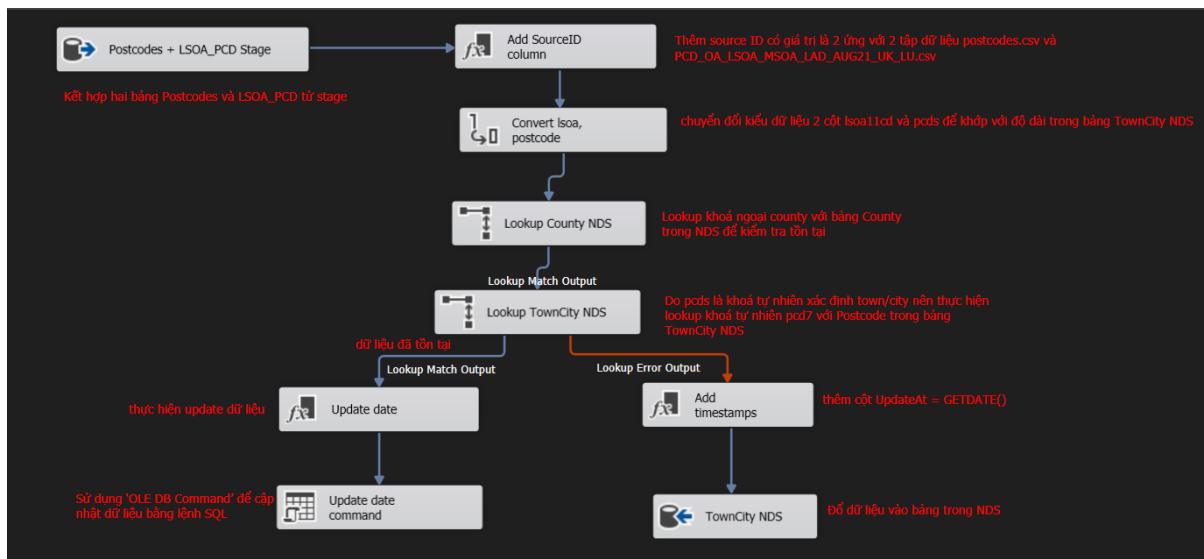
## 2.2. Các bảng địa lý



Thực hiện nạp từ bảng cha (cấp cao) đến bảng con (cấp thấp) vì bảng con sẽ tham chiếu khoá ngoại đến bảng cha:

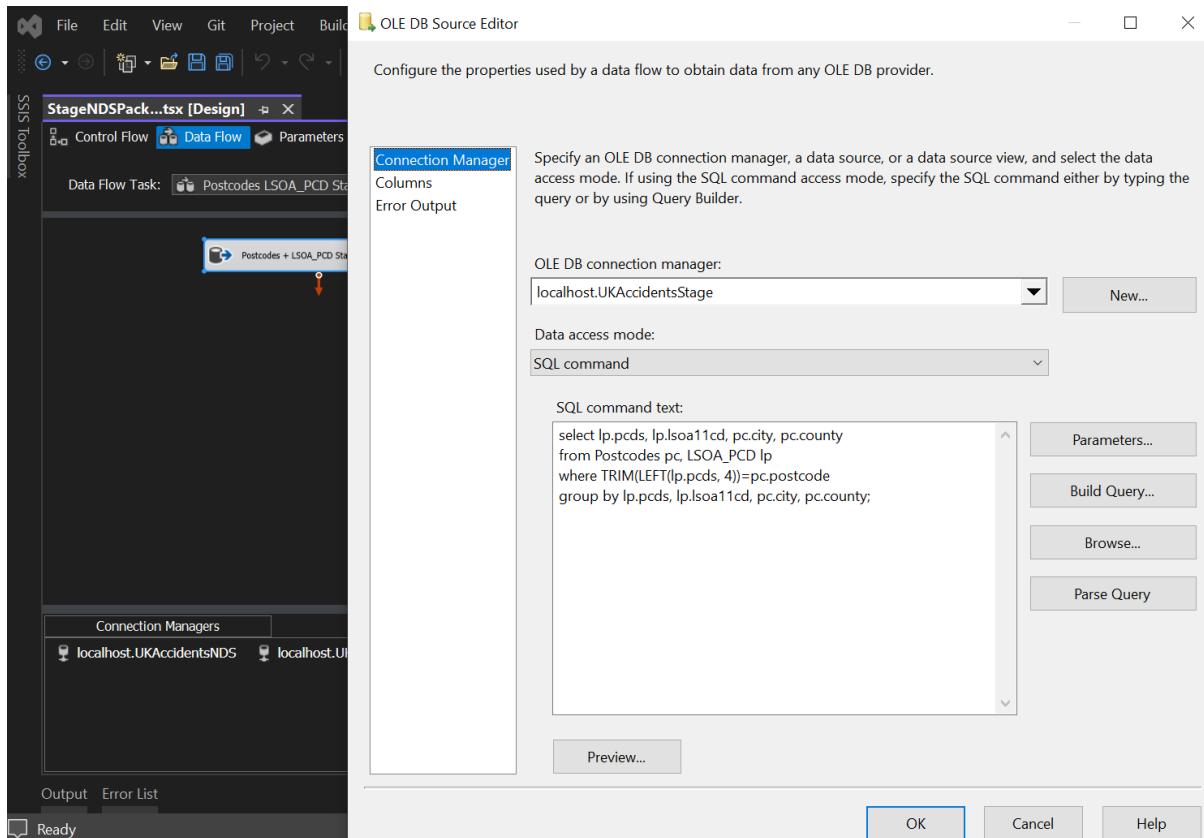
Country → Region → County → Town/City

Ví dụ minh họa nạp bảng TownCity



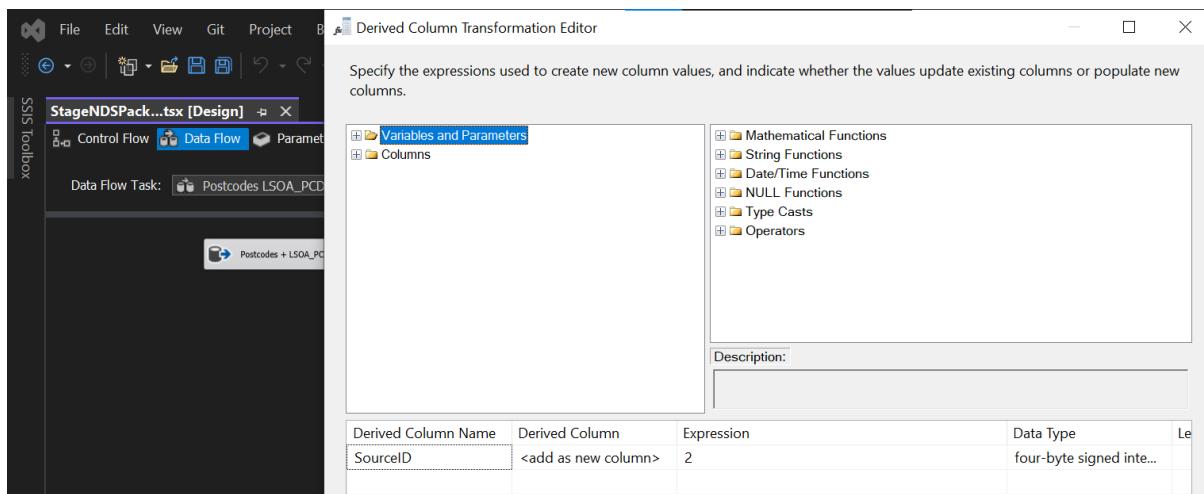
## 2.2.1. Bảng Stage

Kết hợp hai bảng Postcodes và LSOA\_PCD từ stage:



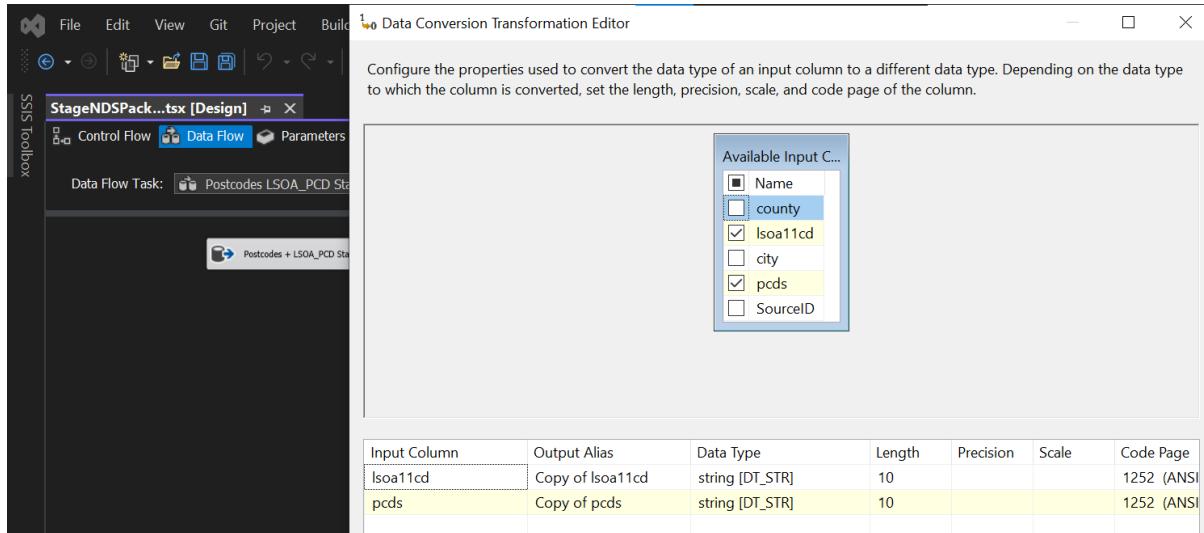
## 2.2.2. Thêm SourceID

Thêm source ID có giá trị là 2 ứng với 2 tập dữ liệu postcodes.csv và PCD\_OA\_LSOA\_MSKOA\_LAD\_AUG21\_UK\_LU.csv.



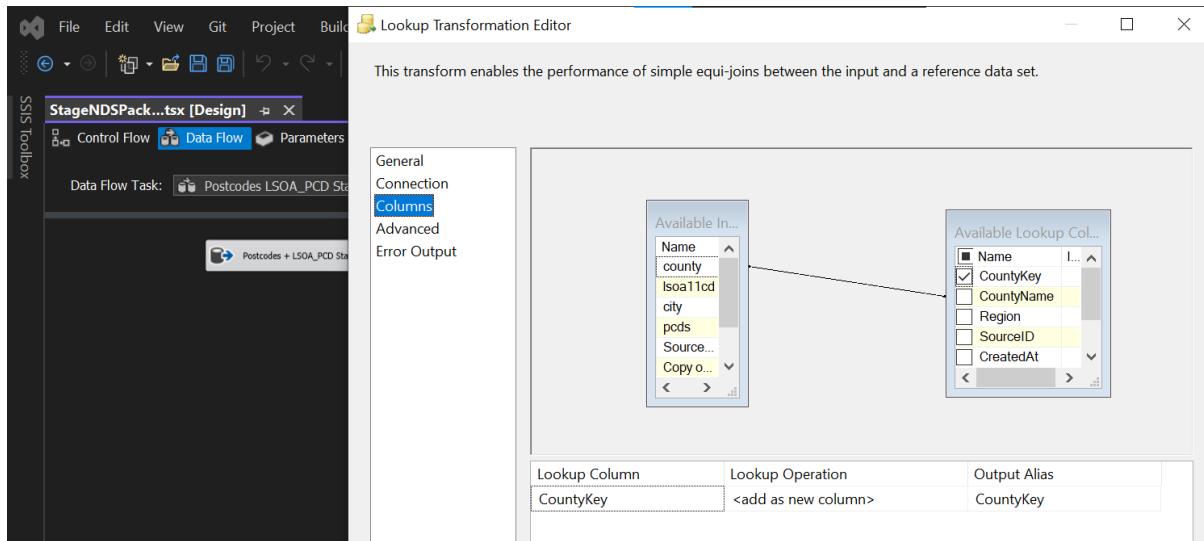
### 2.2.3. Chuyển đổi dữ liệu

Sử dụng component 'Data Conversion' để chuyển đổi kiểu dữ liệu 2 cột lsoa11cd và pcds để khớp với độ dài trong bảng TownCity NDS:



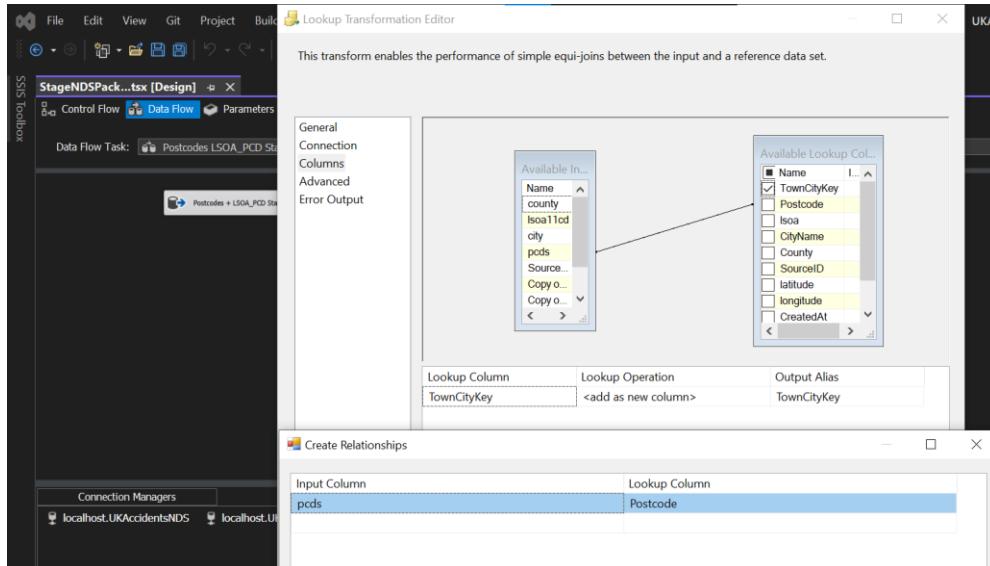
### 2.2.4. Lookup khoá ngoại

Lookup khoá ngoại county với bảng County trong NDS để kiểm tra tồn tại



## 2.2.5. Lookup khoá tự nhiên

Do pcds là khoá tự nhiên xác định town/city nên thực hiện lookup khoá tự nhiên pcds với Postcode trong bảng TownCity NDS.



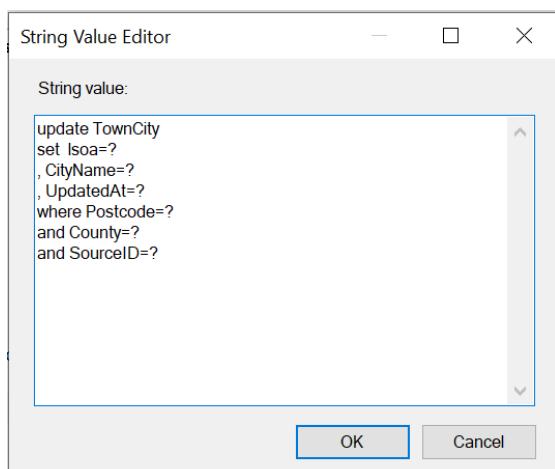
## 2.2.6. Upsert

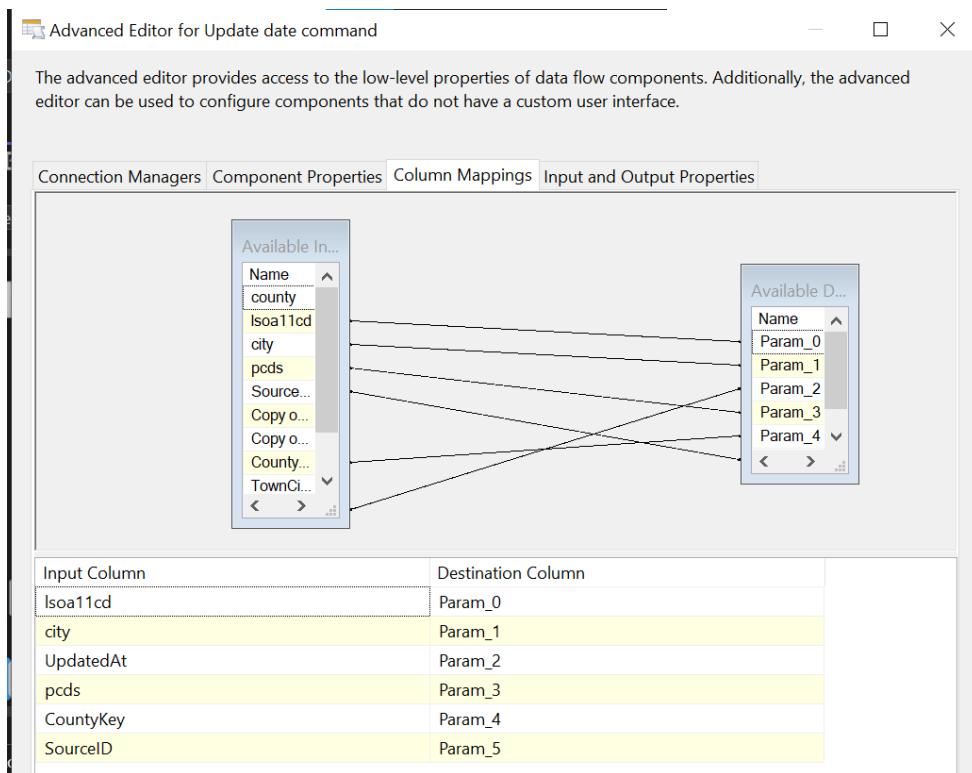
Nhận kết quả lookup phía trên để thực hiện insert hoặc update dữ liệu

Nếu dữ liệu đã tồn tại, thực hiện update dữ liệu:

- Sử dụng 'Derived Column' thêm cột UpdatedAt = GETDATE()
- Sử dụng 'OLE DB Command' để cập nhật dữ liệu bằng lệnh SQL

Derived Column Name	Derived Column	Expression	Data Type	Length
UpdatedAt	<add as new column>	GETDATE()	database timestamp	[...]

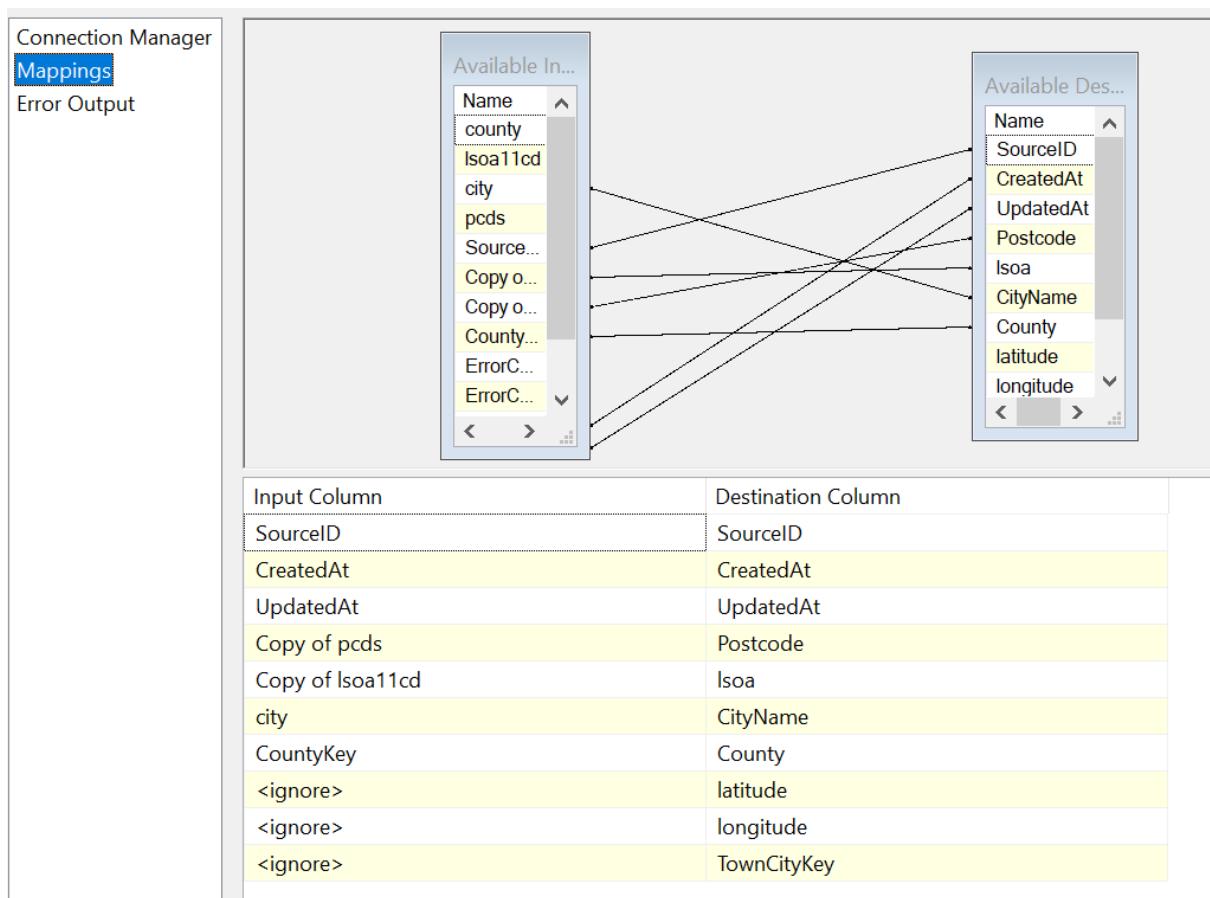




Nếu dữ liệu chưa có trong NDS, thực hiện insert dữ liệu:

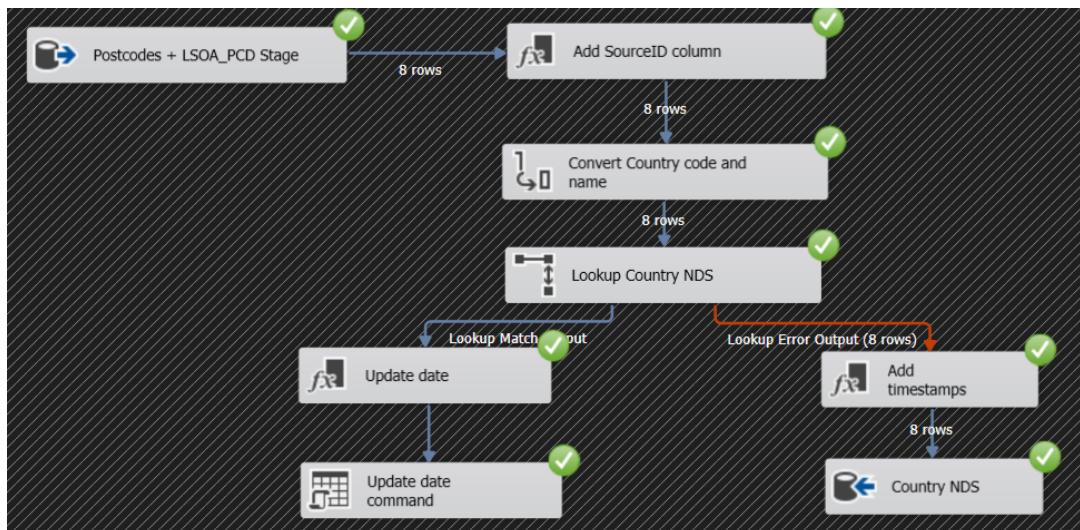
- Sử dụng 'Derived Column' thêm cột UpdateAt = GETDATE()
- Sử dụng 'OLE DB Destination' để đổ dữ liệu vào bảng trong NDS

Derived Column Name	Derived Column	Expression	Data Type	Length
CreatedAt	<add as new column>	GETDATE()	database timestamp	[...]
UpdatedAt	<add as new column>	GETDATE()	database timestamp	[...]

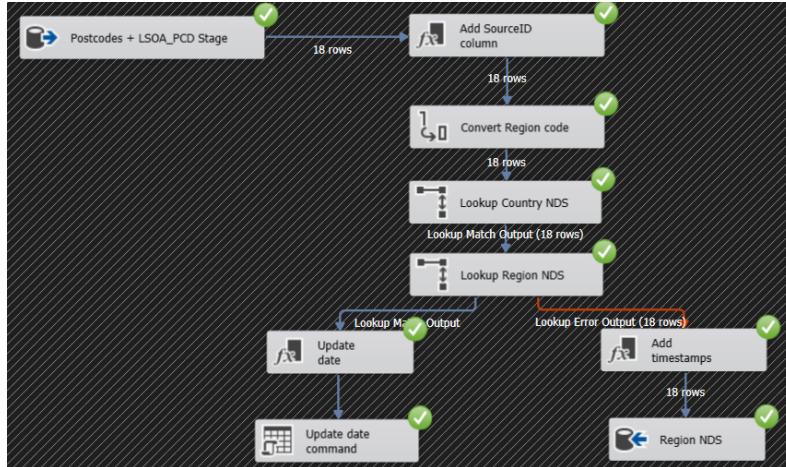


## 2.2.7. Kết quả thực thi

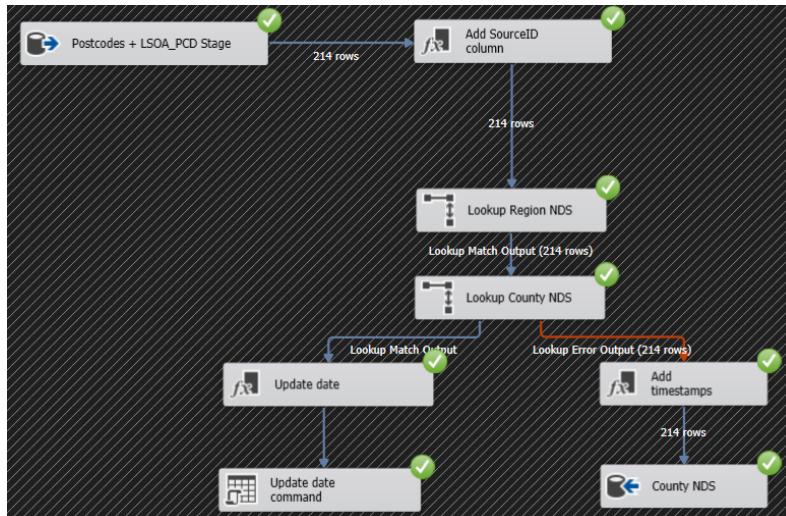
Bảng Country



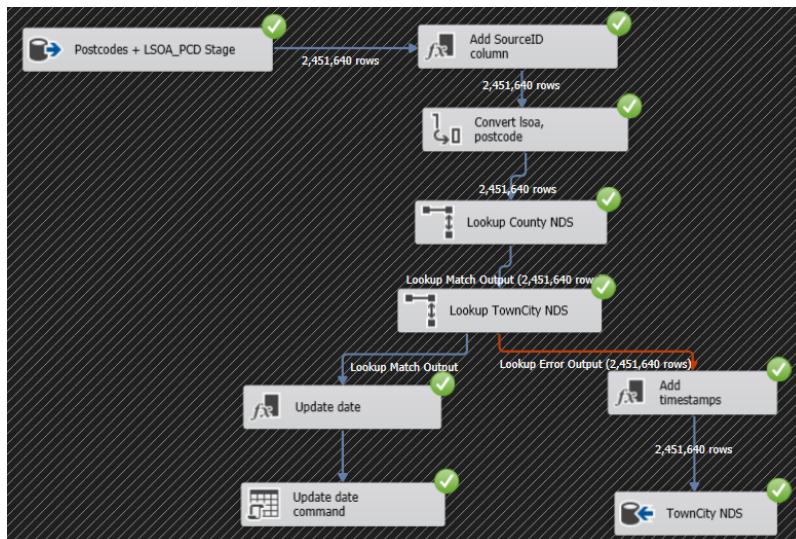
## Bảng Region



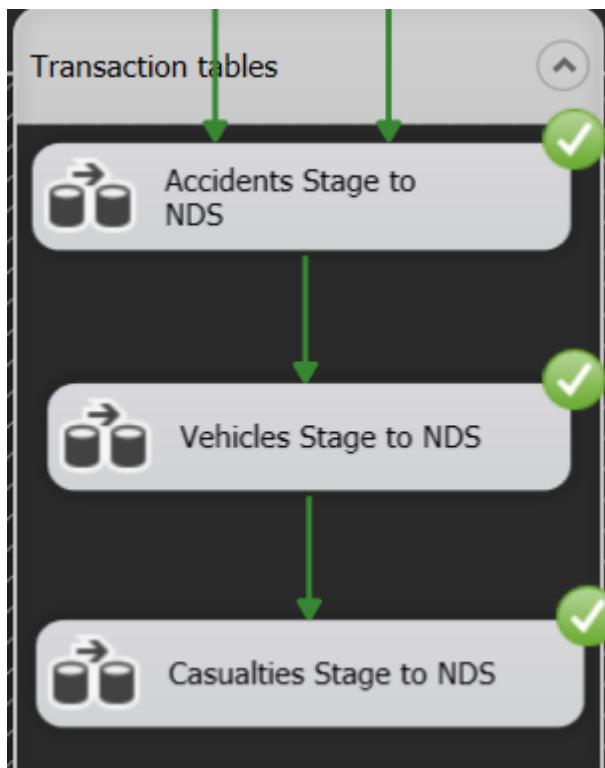
## Bảng County



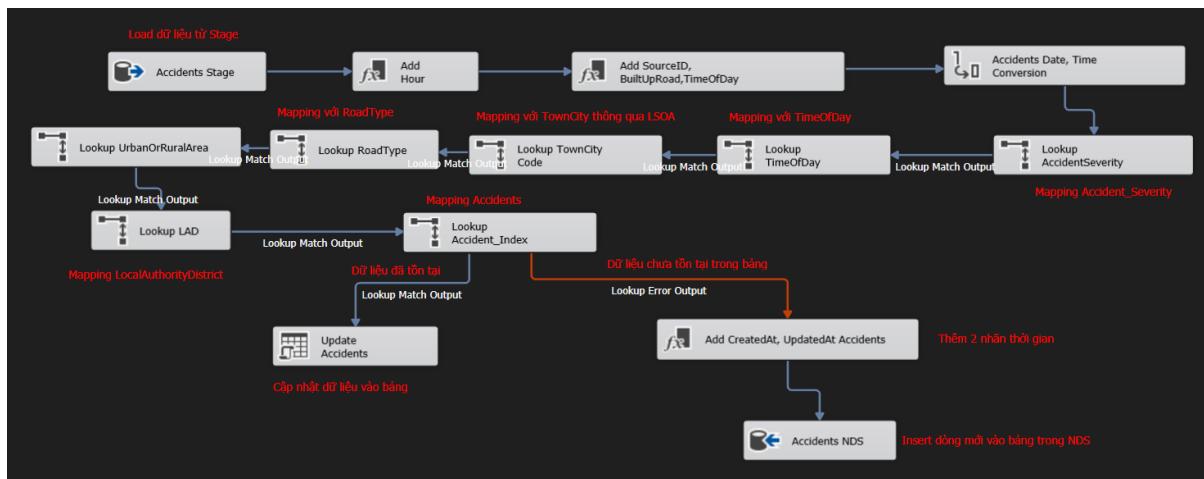
## Bảng TownCity



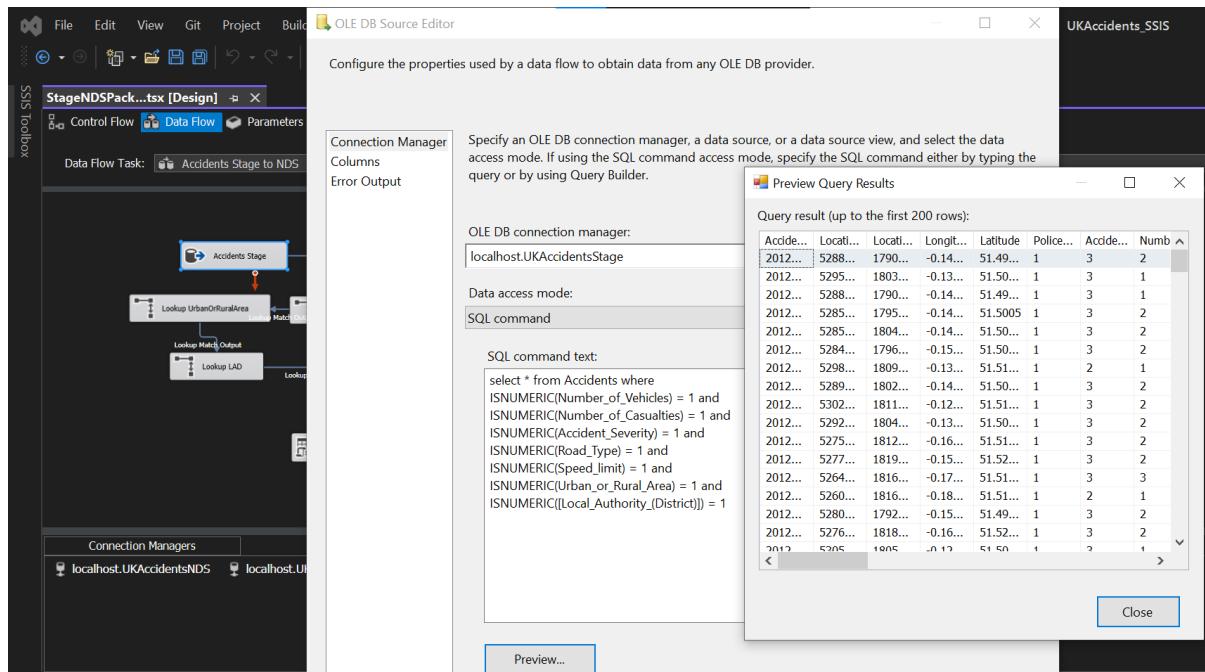
### 2.3. Các bảng Transactions



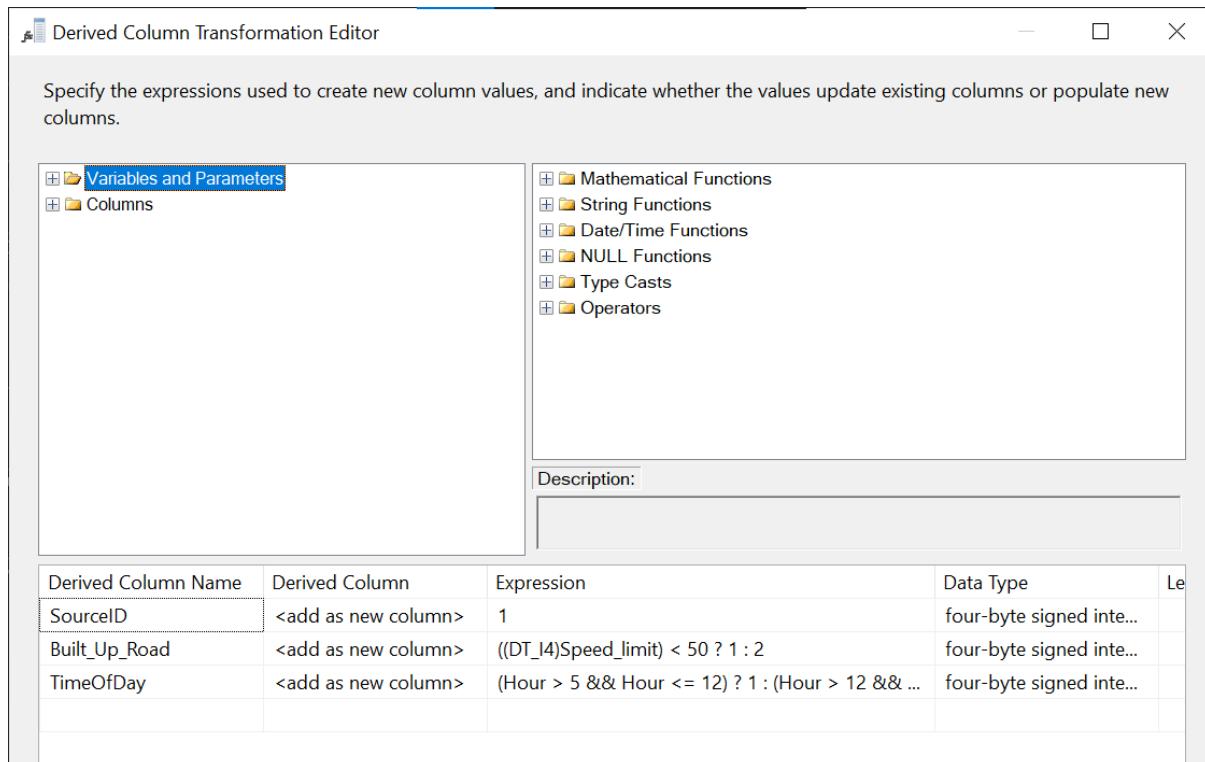
#### Ví dụ minh họa bảng Accidents



### 2.3.1. Accident Stage



### 2.3.2. Add SourceID, BuiltUpRoad, TimeOfDay



### 2.3.3. Accidents Data Conversion

Data Conversion Transformation Editor

Configure the properties used to convert the data type of an input column to a different data type. Depending on the data type to which the column is converted, set the length, precision, scale, and code page of the column.

Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Page
Time	Copy of Time	database time [DT_DBT...]				
Date	Copy of Date	database date [DT_DBD...]				

### 2.3.4. Lookup Accident\_Severity

Lookup Transformation Editor

This transform enables the performance of simple equi-joins between the input and a reference data set.

General	Connection	Columns	Advanced	Error Output

Available Input Columns

Name
Accident_Index
Accident_Severity
Number_of_Vehicles
Number_of_Casualties
Date
Time
Local_Authority_(Dis...)
Road_Type
Speed_limit
Urban_or_Rural_Area

Available Lookup Columns

Name
AccidentSeverityKey
Code
Label
SourceID
CreatedAt

Lookup Column      Lookup Operation      Output Alias

Lookup Column	Lookup Operation	Output Alias
AccidentSeverityKey	<add as new column>	AccidentSeverityKey

Create Relationships

Input Column	Lookup Column
Accident_Severity	Code

### 2.3.5. Lookup City Code

This screenshot shows the 'Lookup Transformation Editor' window. On the left, a sidebar lists 'General', 'Connection', 'Columns', 'Advanced', and 'Error Output'. The main area has two 'Available' panes: 'Available Input Columns' (containing 'Accident\_Index', 'Accident\_Severity', 'Number\_of\_Vehicles', 'Number\_of\_Casualties', 'Date', 'Time', 'Local\_Authority\_(Dis...)', 'Road\_Type', 'Speed\_limit', 'Urban\_or\_Rural\_Area') and 'Available Lookup Col...' (containing 'Name', 'TownCityKey', 'Postcode', 'Isoa', 'CityName', 'County', 'SourceID', 'latitude', 'longitude', 'CreatedAt'). A connection line connects 'TownCityKey' in the lookup pane to 'TownCityKey' in the 'Lookup Column' section of the bottom table. The table also includes 'Lookup Operation' (<add as new column>) and 'Output Alias' (TownCityKey). Below this is a 'Create Relationships' dialog showing a single relationship between 'LSOA\_of\_Accident\_Location' (Input Column) and 'Isoa' (Lookup Column).

### 2.3.6. Lookup Road\_type

This screenshot shows the 'Lookup Transformation Editor' window. The sidebar and overall layout are identical to the previous screenshot. The 'Available Input Columns' pane includes 'Road\_Type'. A connection line connects 'RoadTypeKey' in the lookup pane to 'RoadTypeKey' in the 'Lookup Column' section of the bottom table. The table includes 'Lookup Operation' (<add as new column>) and 'Output Alias' (RoadTypeKey). Below this is a 'Create Relationships' dialog showing a single relationship between 'Road\_Type' (Input Column) and 'Code' (Lookup Column).

### 2.3.7. Lookup Urban or Rural Area

Lookup Transformation Editor

This transform enables the performance of simple equi-joins between the input and a reference data set.

General  
Connection  
Columns  
Advanced  
Error Output

Available Input Columns

Name
Accident_Index
Accident_Severity
Number_of_Vehicles
Number_of_Casualties
Date
Time
Local_Authority_(District)
Road_Type
Speed_limit
Urban_or_Rural_Area

Available Lookup Columns

Name
<input checked="" type="checkbox"/> UrbanOrRuralKey
<input type="checkbox"/> Code
<input type="checkbox"/> Label
<input type="checkbox"/> SourceID
<input type="checkbox"/> CreatedAt

Lookup Column      Lookup Operation      Output Alias

UrbanOrRuralKey	<add as new column>	UrbanOrRuralKey
-----------------	---------------------	-----------------

Create Relationships

Input Column	Lookup Column
Urban_or_Rural_Area	Code

### 2.3.8. Lookup LAD

Lookup Transformation Editor

This transform enables the performance of simple equi-joins between the input and a reference data set.

General  
Connection  
Columns  
Advanced  
Error Output

Available Input Columns

Name
Accident_Index
Accident_Severity
Number_of_Vehicles
Number_of_Casualties
Date
Time
Local_Authority_(District)
Road_Type
Speed_limit
Urban_or_Rural_Area

Available Lookup...

Name
<input checked="" type="checkbox"/> LADKey
<input type="checkbox"/> Code
<input type="checkbox"/> Label
<input type="checkbox"/> Source...
<input type="checkbox"/> Creat...

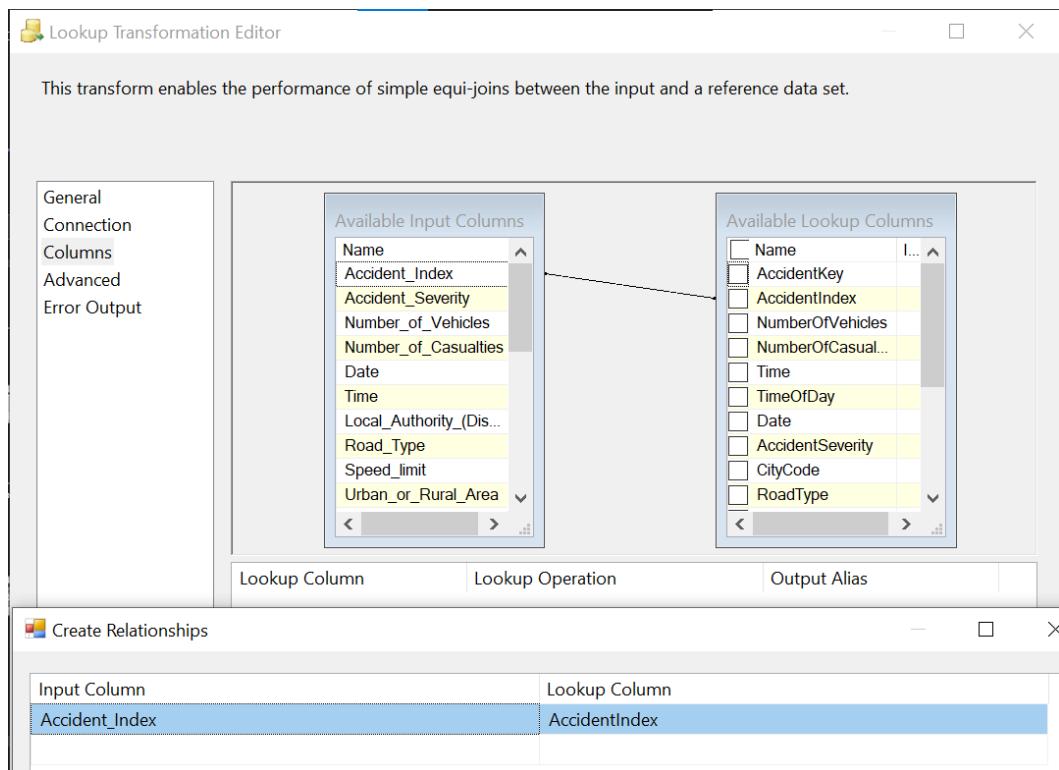
Lookup Column      Lookup Operation      Output Alias

LADKey	<add as new column>	LADKey
--------	---------------------	--------

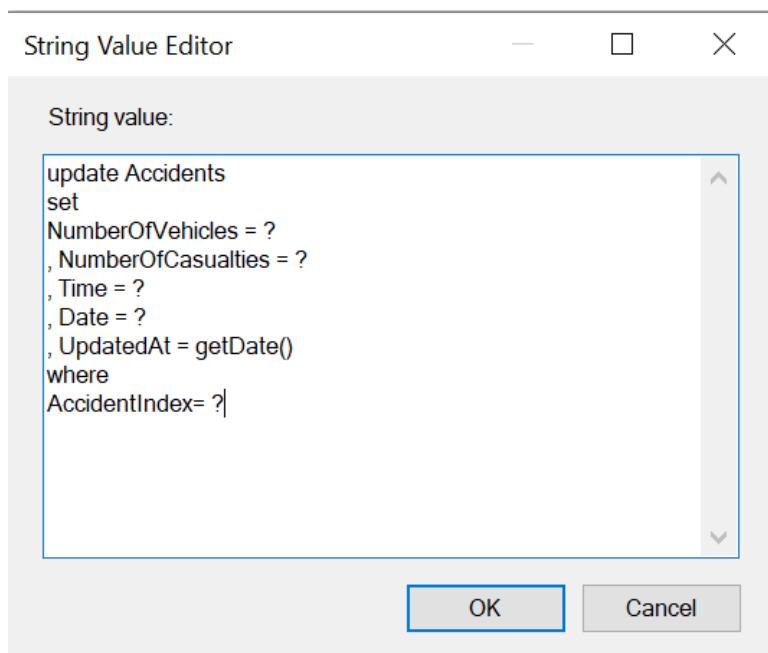
Create Relationships

Input Column	Lookup Column
Local_Authority_(District)	Code

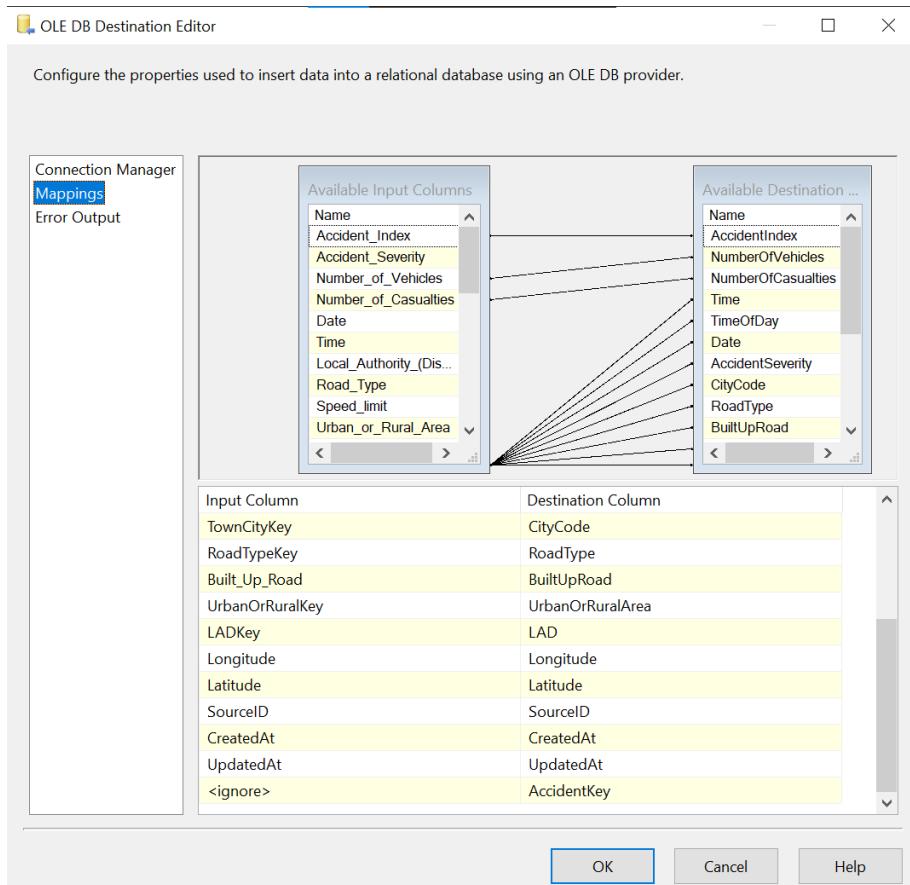
### 2.3.9. Lookup Accident Index



Trường hợp dữ liệu đã tồn tại thì update



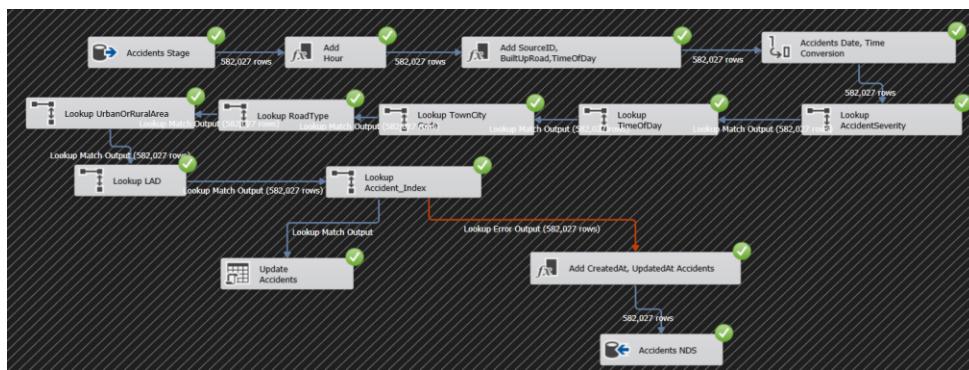
Trường hợp dữ liệu chưa tồn tại thì tiến hành insert dòng dữ liệu mới



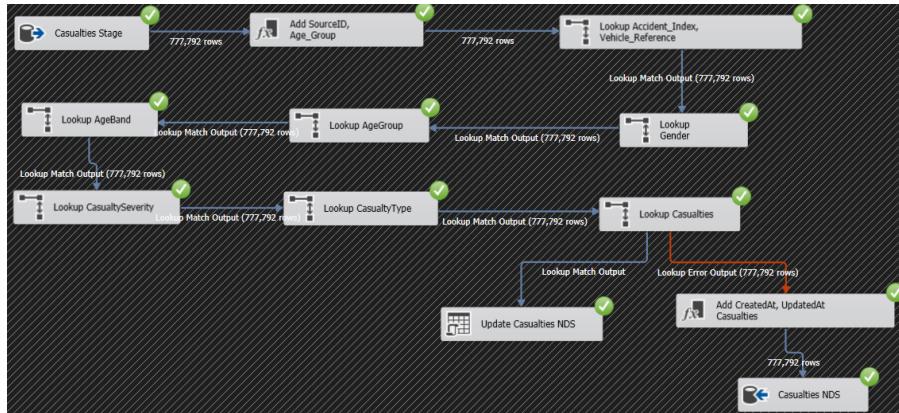
Các bảng còn lại tương tự

### 2.3.10. Kết quả thực thi

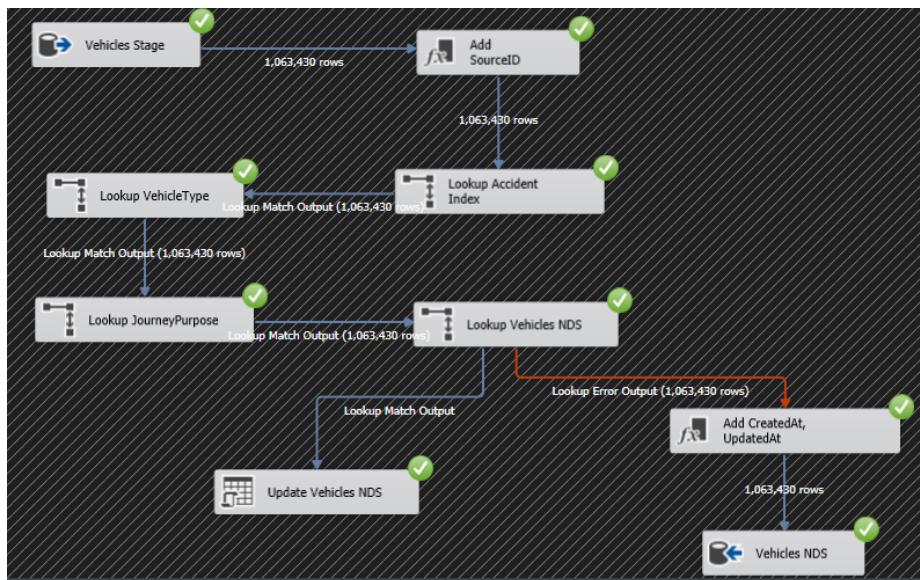
#### Bảng Accidents



## Bảng Casualties



## Bảng Vehicles



## III. ETL NDS to Data Warehouse

### 1. Lược đồ DDS và mô tả

#### 3.1. DimGender

Thuộc tính	Mô tả
GenderKey int identity(1,1),	Mã tự tăng của bảng giới tính
Code int,	Mã giới tính
Label varchar(100),	Nhãn giới tính
CONSTRAINT PK_G PRIMARY KEY(GenderKey)	Set GenderKey là PK

### *3.2. DimAgeBand*

Thuộc tính	Mô tả
AgeBandKey int identity(1,1),	Mã tự tăng bảng nhóm độ tuổi (trong khoảng từ x → y)
Code int,	Mã nhóm tuổi
Label varchar(100),	Nhãn
CONSTRAINT PK_AB PRIMARY KEY(AgeBandKey)	Set AgeBandKey là Pk

### *3.3. DimCasualtySeverity*

Thuộc tính	Mô tả
CasualtySeverityKey int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã độ nghiêm trọng của nạn nhân
CONSTRAINT PK_CS PRIMARY KEY(CasualtySeverityKey)	Set CasualtySeverityKey là PK

### *3.4. DimCasualtyType*

Thuộc tính	Mô tả
CasualtyTypeKey int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã loại tai nạn
Label varchar(100),	Nhãn
CONSTRAINT PK_CT PRIMARY KEY(CasualtyTypeKey)	Set CasualtyTypeKey là PK

### *3.5. DimAccidentSeverity*

Thuộc tính	Mô tả
AccidentSeverityIndex int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã độ nghiêm trọng của vụ tai nạn
Label varchar(20),	Nhãn
CONSTRAINT PK_AS PRIMARY KEY(AccidentSeverityKey)	Set AccidentSeverityKey là PK

### *3.6. DimLocalAuthorityDistrict*

Thuộc tính	Mô tả
LADKey int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã địa phương
Label varchar(50),	Nhãn
CONSTRAINT PK_LAD PRIMARY KEY(LADKey)	Set LADKey là PK

### *3.7. DimUrbanOrRuralArea*

Thuộc tính	Mô tả
UrbanOrRuralKey int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã nơi xảy ra tai nạn (Thành thị hay nông thôn)
Label varchar(10),	Nhãn
CONSTRAINT PK_UORA PRIMARY KEY(UrbanOrRuralKey)	Set UrbanOrRuralKey là PK

### *3.8. DimRoadType*

Thuộc tính	Mô tả
RoadTypeKey int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã loại đường
Label varchar(50),	Nhãn
CONSTRAINT PK_RT PRIMARY KEY(RoadTypeKey)	Set RoadTypeKey là PK

### *3.9. DimVehicleType*

Thuộc tính	Mô tả
VehicleTypeIndex int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã loại phương tiện
Label varchar(100),	Nhãn
CONSTRAINT PK_VT PRIMARY KEY(VehicleTypeKey)	Set VehicleTypeKey là PK

### *3.10. DimJourneyPurpose*

Thuộc tính	Mô tả
JourneyPurposeKey int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã chuyển đi
Label varchar(100),	Nhãn
CONSTRAINT PK_JP PRIMARY KEY(JourneyPurposeKey)	Set JourneyPurposeKey là PK

### *3.11. Dim AgeGroup*

Thuộc tính	Mô tả
AgeGroupKey int identity(1,1),	Mã tự tăng của bảng nhóm tuổi
Code int,	Mã nhóm tuổi
Label varchar(100),	Nhãn
CONSTRAINT PK_AG PRIMARY KEY(AgeGroupKey)	Set AgeGroupKey là PK

### *3.12. DimBuiltUpRoad*

Thuộc tính	Mô tả
BuiltUpRoadKey int identity(1,1),	Mã tự tăng (SK)
Code int,	Mã đường nơi xảy ra tai nạn (Có phải khu vực đông dân cư không)
Label varchar(100),	Nhãn
CONSTRAINT PK_BUR PRIMARY KEY(BuiltUpRoadKey)	Set BuiltUpRoadKey là PK

### *3.13. DimTimeOfDay*

Thuộc tính	Mô tả
TimeOfDayIndex int identity(1,1)	Mã tự tăng (SK)
Code int,	Mã của thời gian trong ngày
Label varchar(100),	Nhãn

### *3.14. DimTownCity*

Thuộc tính	Mô tả
TownCityIndex int identity(1,1),	Mã tự tăng
Postcode varchar(10),	mã bưu điện
lsoa varchar(10),	Mã địa lý
CityName varchar(50),	Tên thành phố
County int,	Mã hạt/ tinh
CONSTRAINT PK_TC PRIMARY KEY(TownCityKey)	Set TownCityKey là PK

### *3.15. DimCounty*

Thuộc tính	Mô tả
CountyIndex int identity(1,1),	Mã tự tăng của bảng Hạt/ tinh
CountyName varchar(50),	Tên hạt/ tinh (đơn vị hành chính lớn nhất ở Anh)
Region int,	Mã vùng/ miền
CONSTRAINT PK_C PRIMARY KEY(CountyIndex)	set CountyKey là PK

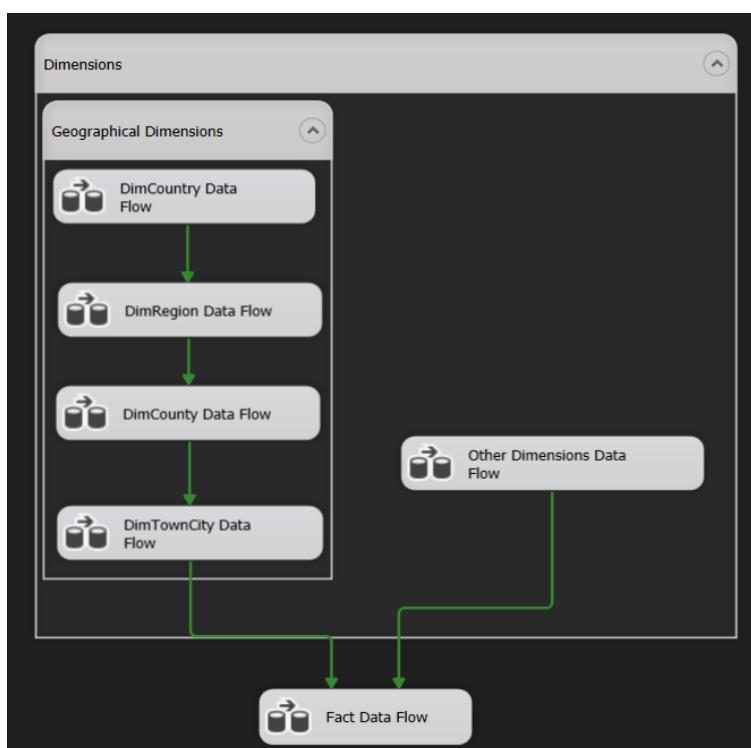
### *3.16. DimRegion*

Thuộc tính	Mô tả
RegionIndex int identity(1,1),	Mã tự tăng của bảng vùng/miền
RegionCode varchar(10),	Mã vùng/ miền
RegionName varchar(50),	Tên vùng
Country int,	Mã quốc gia của vùng
CONSTRAINT PK_R PRIMARY KEY(RegionIndex)	Set RegionKey là PK

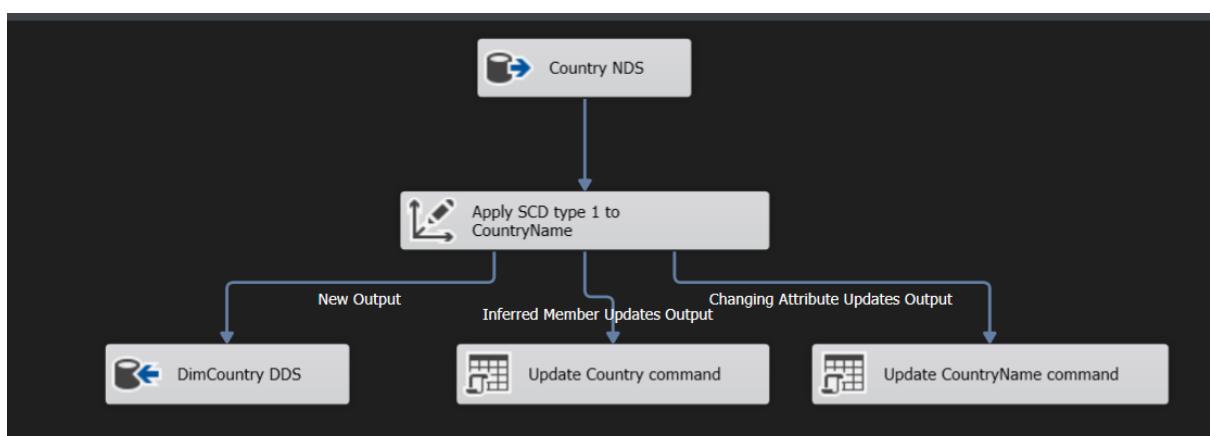
### 3.17. DimCountry

Thuộc tính	Mô tả
CountryIndex int identity(1,1),	Mã tự tăng của quốc gia
CountryCode varchar(10),	Mã quốc gia theo data
CountryName varchar(20),	Tên quốc gia
CONSTRAINT PK_CY PRIMARY KEY(CountryIndex)	Set CountryKey là PK

## 2. Quy trình ETL



### 2.1 Bảng DimCountry(DimCountry Data Flow)



## Load dữ liệu Country từ NDS

Connection Manager  
Columns  
Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.

OLE DB connection manager: localhost.UKAccidentsNDS New...

Data access mode: Table or view

Name of the table or the view: [dbo].[Country]

**Preview Query Results**

Query result (up to the first 200 rows):

Count...	Count...	Count...	Sourc...	Create...	Updat...
1	IMN	Isle of...	2	5/15/...	5/15/...
2	IOM	IMN	2	5/15/...	5/15/...
3	SCT	Scotland...	2	5/15/...	5/15/...
4	ENG	England	2	5/15/...	5/15/...
5	WLS	Wales	2	5/15/...	5/15/...
6	NIR	North...	2	5/15/...	5/15/...
7	GGR	Guernsey	2	5/15/...	5/15/...
8	JEY	Jersey	2	5/15/...	5/15/...

## Áp dụng SCD loại 1 lên thuộc tính CountryName

**Slowly Changing Dimension Wizard**

**Select a Dimension Table and Keys**  
Select a dimension table to load and map columns in the transformation input to

Connection manager: localhost.UKAccidentsDDS

Table or view: [dbo].[DimCountry]

**Input Columns**   **Dimension Columns**   **Key Type**

CountryCode	CountryCode	Not a key column
CountryKey	CountryKey	Business key
CountryName	CountryName	Not a key column

Help   < Back   Next >   Finish >>   Cancel

**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.

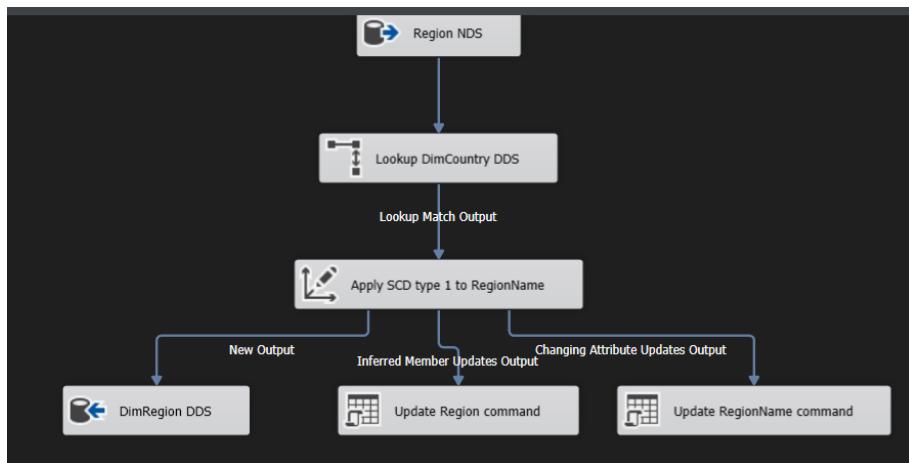
**Historical Attribute**  
Select this type when changes in column values are saved in new records. Previous values are saved in

Dimension Columns   Change Type

CountryName	Changing a...
-------------	---------------

Help   < Back   Next >   Finish >>   Cancel

## 2.2 Bảng DimRegion(DimRegion Data Flow)



Load dữ liệu Region từ NDS

Connection Manager  
Columns  
Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.

OLE DB connection manager:  
localhost.UKAccidentsNDS

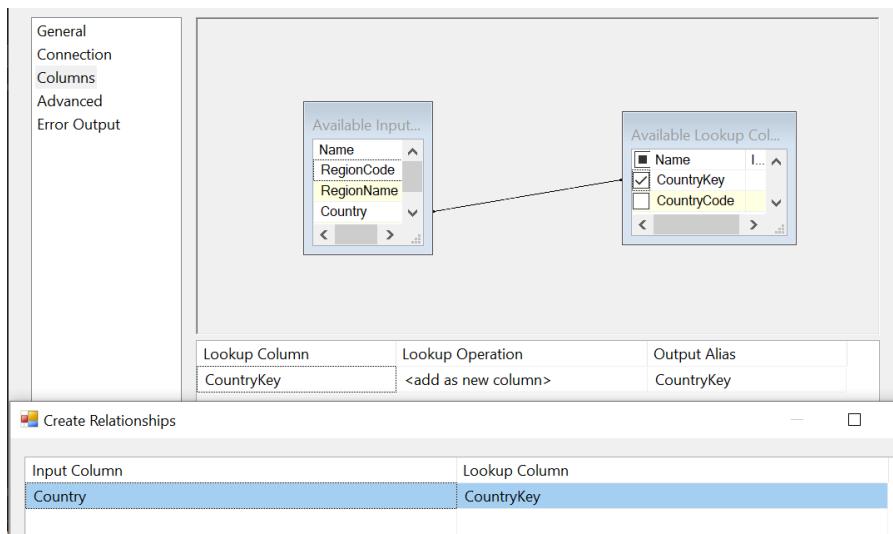
Data access mode:  
Table or view

Name of the table or the view:  
[dbo].[Region]

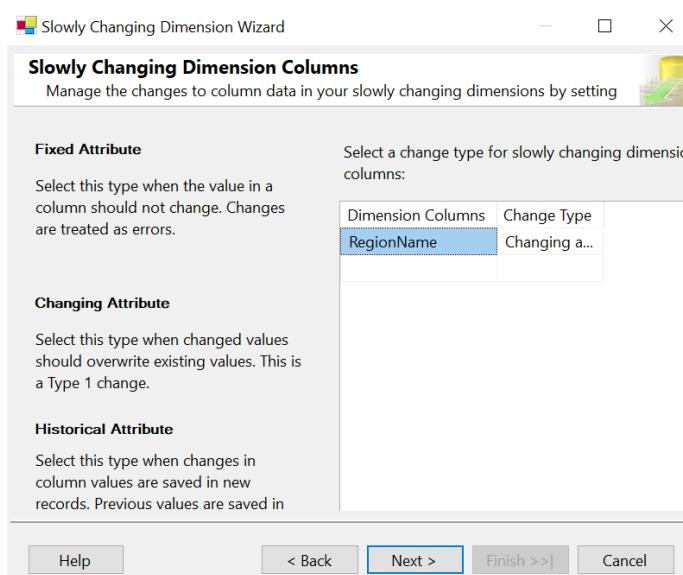
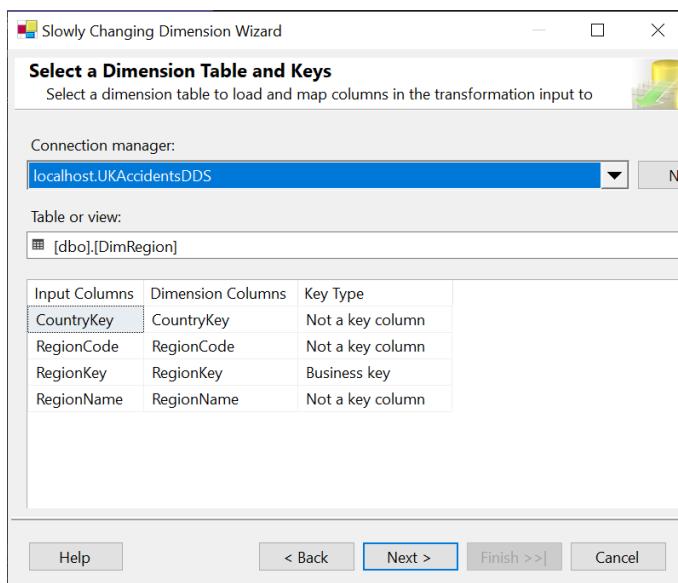
Preview Query Results

RegionID	RegionName	Country	Source	CreateDate	UpdateDate
1	SOUTH	South...	4	2	5/15/...
2	NORTH	North...	4	2	5/15/...
3	WEST	West ...	4	2	5/15/...
4	SOUTH	South...	4	2	5/15/...
5	EAST	#N/A	4	2	5/15/...
6	NI	North...	6	2	5/15/...
7	SCOT	Scotla...	3	2	5/15/...
8	IMN	Isle of...	1	2	5/15/...
9	IM	IMN,I...	2	2	5/15/...
10	NORTH	North...	4	2	5/15/...
11	EAST	East ...	4	2	5/15/...

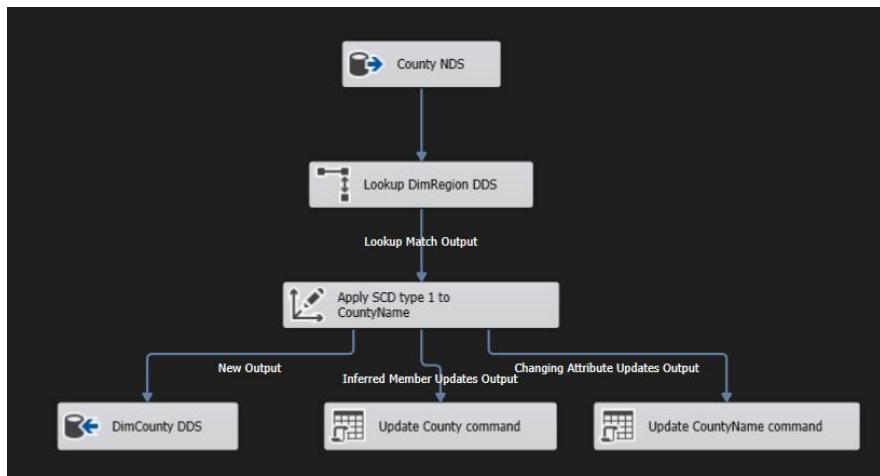
Lookup Country



## Áp dụng SCD loại 1 lên thuộc tính RegionName



## 2.3 Bảng DimCounty(DimCounty Data Flow)



Load dữ liệu County từ NDS

Connection Manager  
Columns  
Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.

OLE DB connection manager: localhost.UKAccidentsNDS

Data access mode: Table or view

Name of the table or view: [dbo].[County]

Preview Query Results

Query result (up to the first 200 rows):

Count...	Count...	Region	Sourc...	Create...	Updat...
1	Liverp...	2	2	5/15/...	5/15/...
2	North...	7	2	5/15/...	5/15/...
3	Swindl...	1	2	5/15/...	5/15/...
4	Merth...	13	2	5/15/...	5/15/...
5	Leices...	11	2	5/15/...	5/15/...
6	City o...	14	2	5/15/...	5/15/...
7	Stoke...	3	2	5/15/...	5/15/...
8	Sunde...	10	2	5/15/...	5/15/...
9	Bury	2	2	5/15/...	5/15/...
10	St Hel...	2	2	5/15/...	5/15/...
11	Magh...	6	2	5/15/...	5/15/...
12	Orkne...	7	2	5/15/...	5/15/...
13	Blaen...	13	2	5/15/...	5/15/...
14	Worce...	3	2	5/15/...	5/15/...
15	Essex	5	2	5/15/...	5/15/...
16	Salford	2	2	5/15/...	5/15/...

Preview... Close

Áp dụng SCD loại 1 lên thuộc tính CountyName

Slowly Changing Dimension Wizard

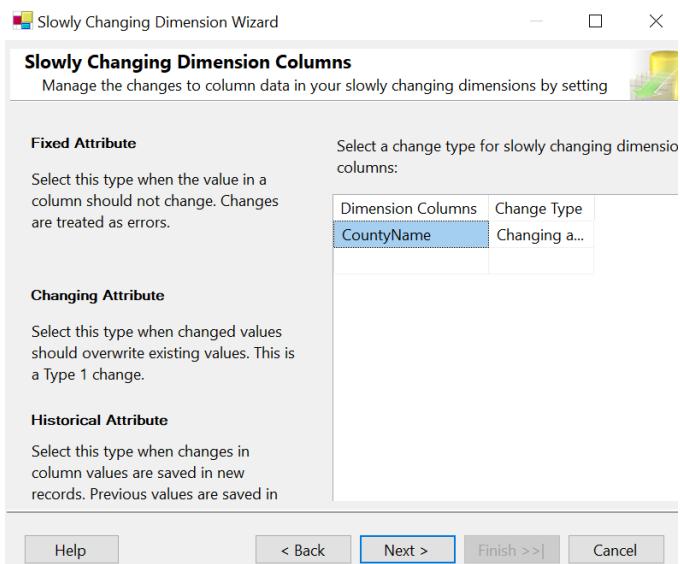
Select a Dimension Table and Keys

Select a dimension table to load and map columns in the transformation input to

Connection manager: localhost.UKAccidentsDDS

Table or view: [dbo].[DimCounty]

Input Columns	Dimension Columns	Key Type
CountyKey	CountyKey	Business key
CountyName	CountyName	Not a key column
RegionKey	RegionKey	Not a key column

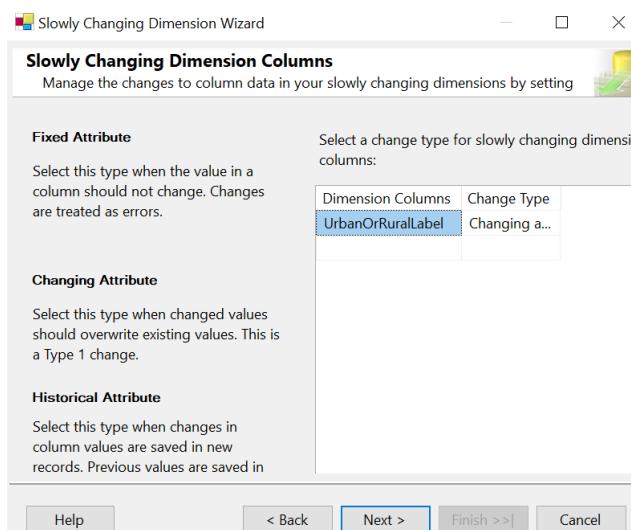
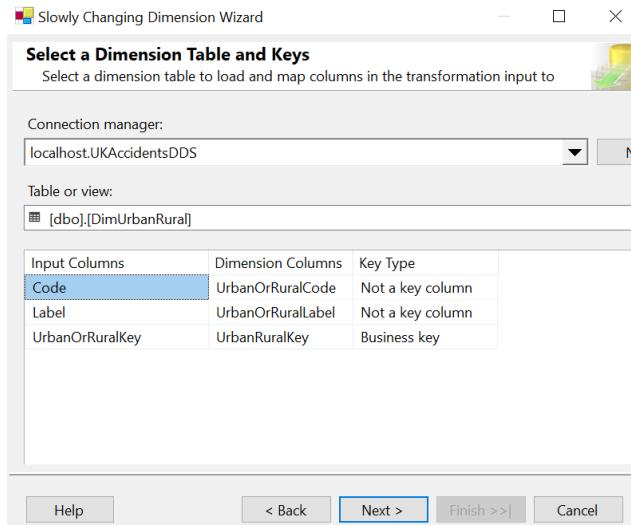


## 2.4 Các bảng chiều còn lại(Other Dimensions Data Flow)

Ví dụ: bảng DimUrbanOrRuralArea(các bảng còn lại tương tự)

Load dữ liệu UrbanOrRural từ NDS

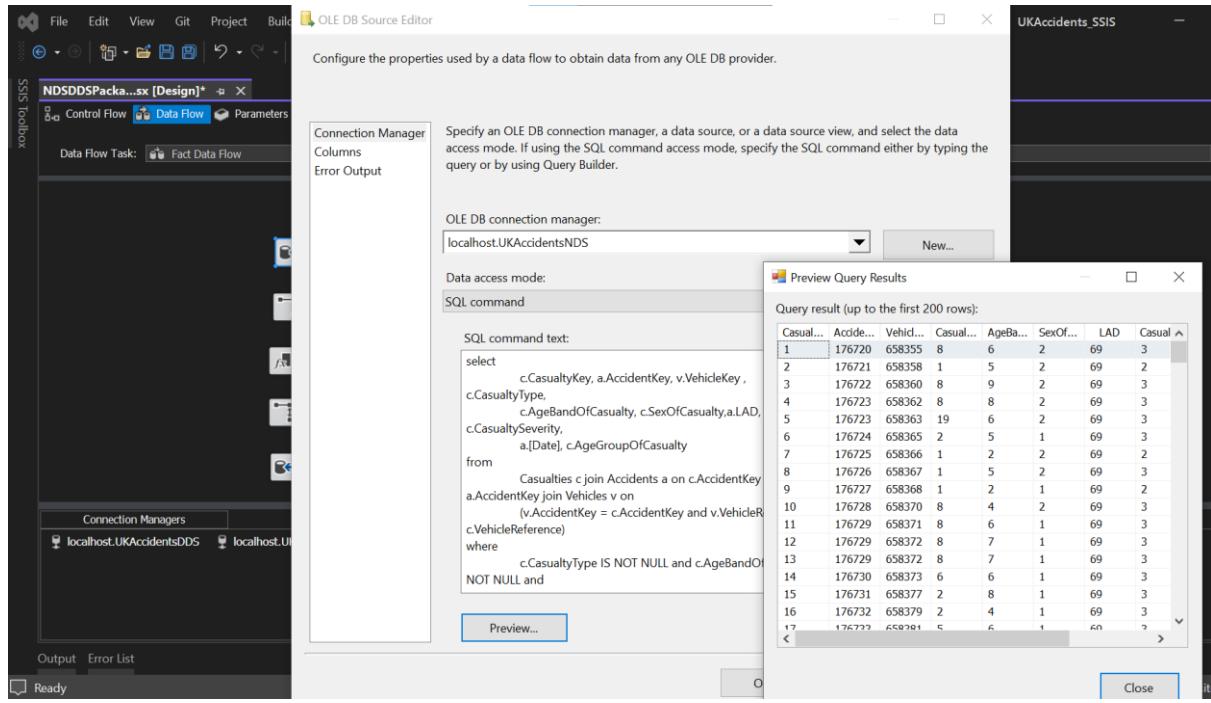
Áp dụng SCD loại 1 lên thuộc tính UrbanOrRuralLabel



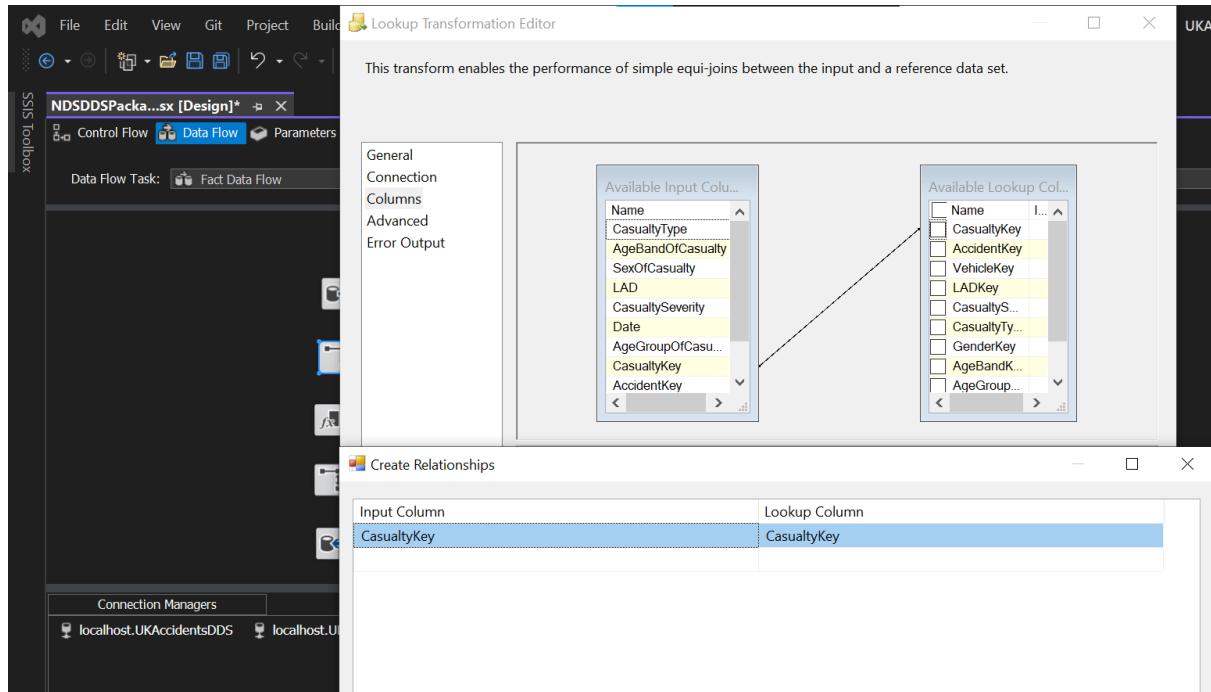
## 2.5 Các bảng Fact(Fact Data Flow)

Ví dụ: bảng FactCasualties(bảng FactAccident và FactVehicles tương tự)

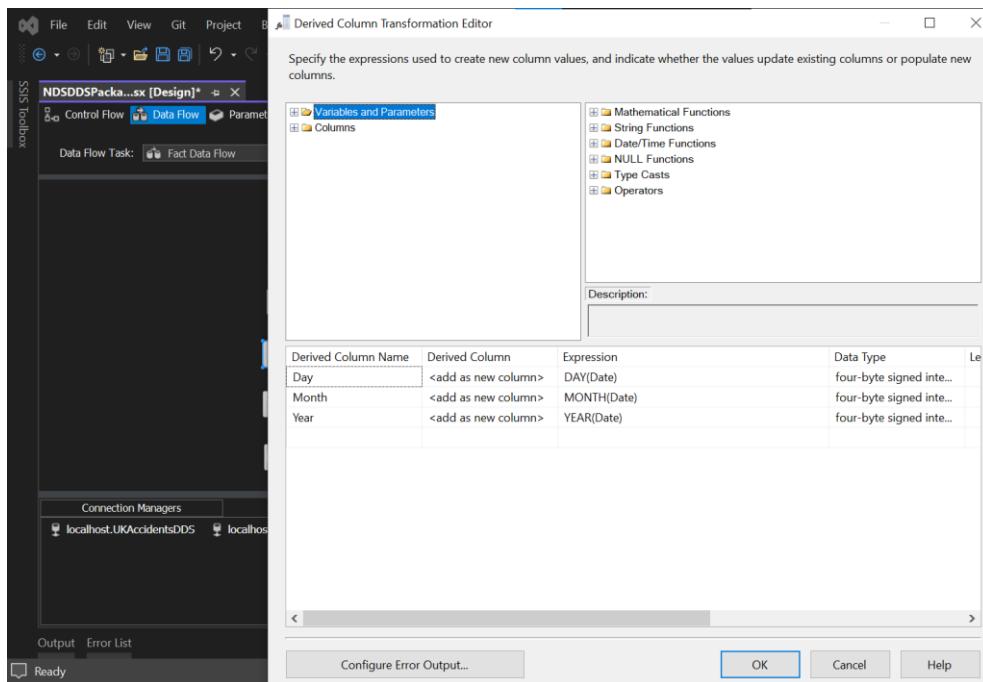
Load dữ liệu từ NDS



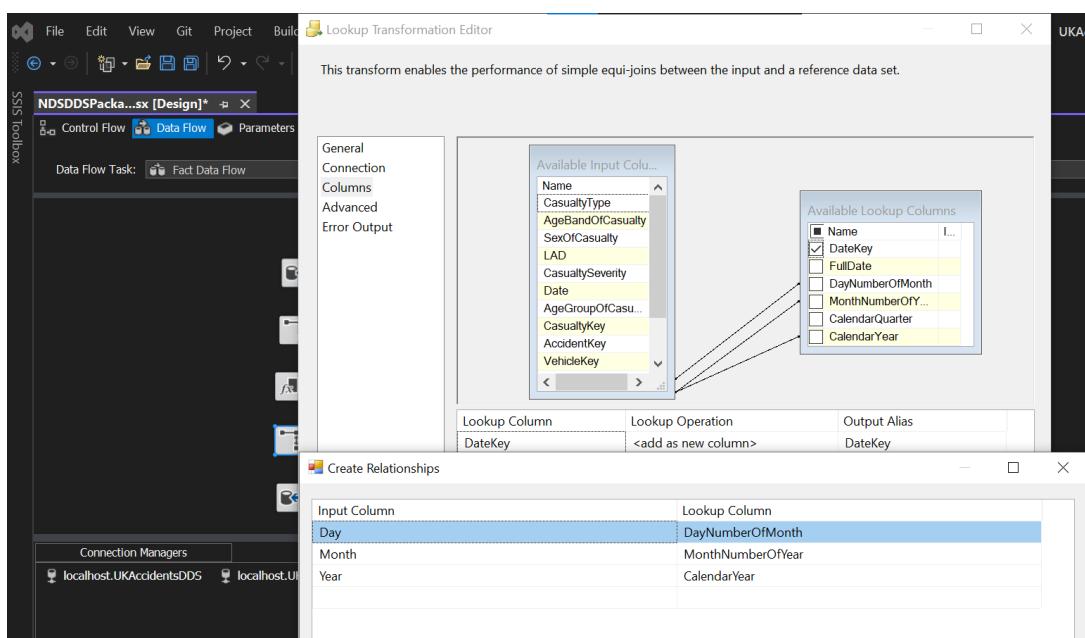
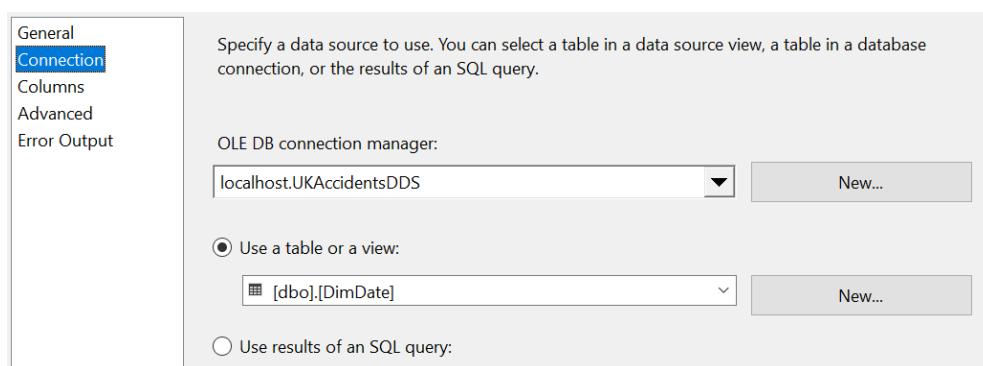
Lookup bảng FactCasualties tìm kiếm các dữ liệu đã tồn tại hay chưa



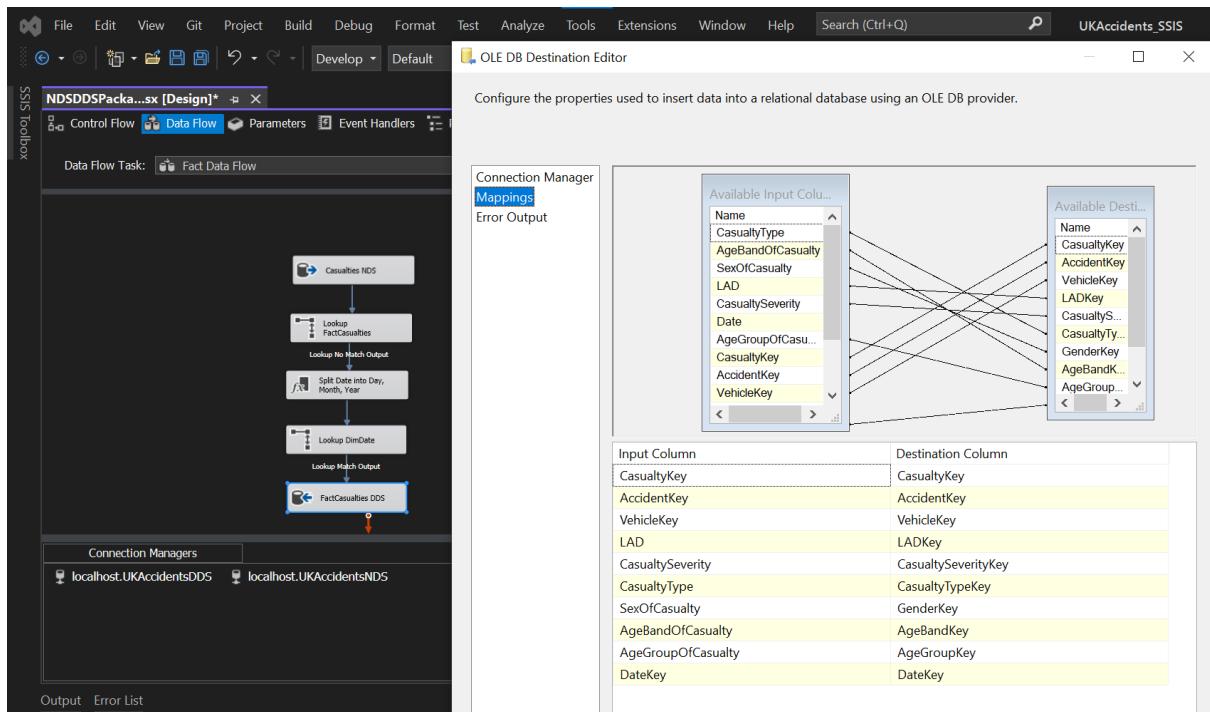
Tách Day, Month, Year từ cột Date bằng hàm DAY(), MONTH(), YEAR()



## Lookup Day, Month, Year

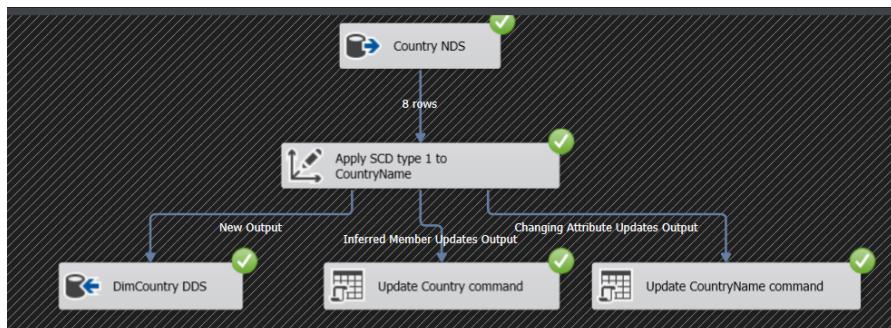


## Đổ dữ liệu vào bảng FactCasualties

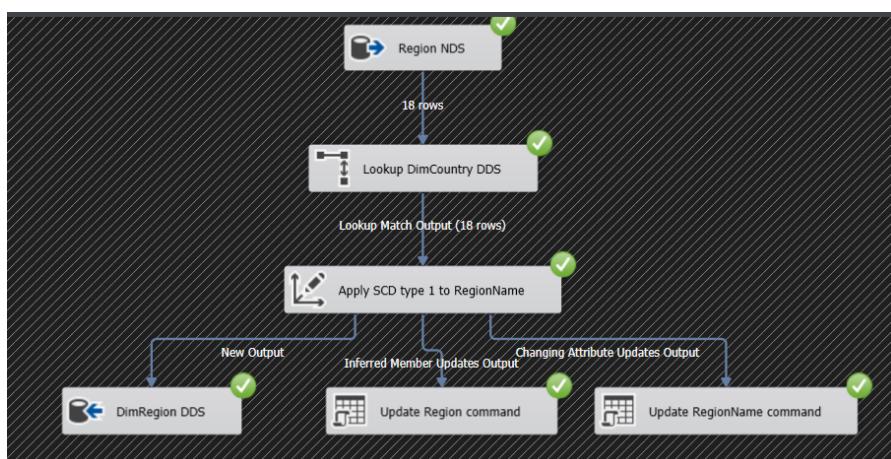


## 2.6. Kết quả thực thi

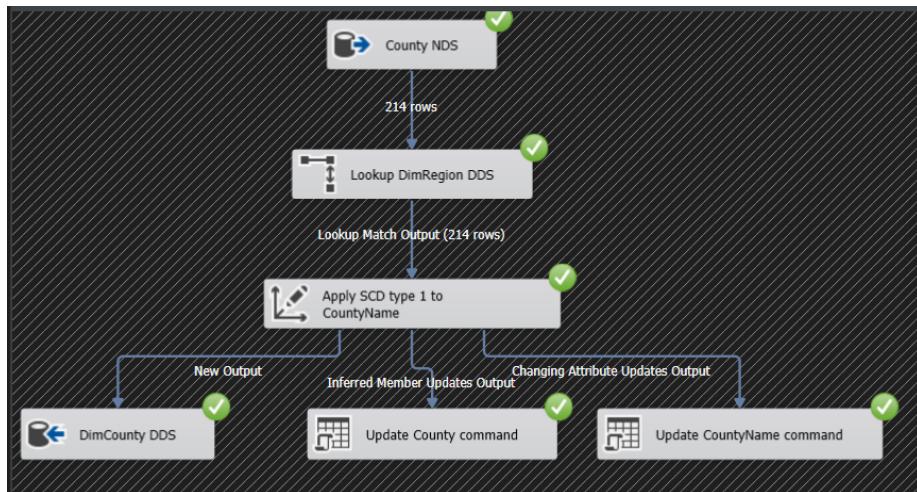
### Bảng DimCountry



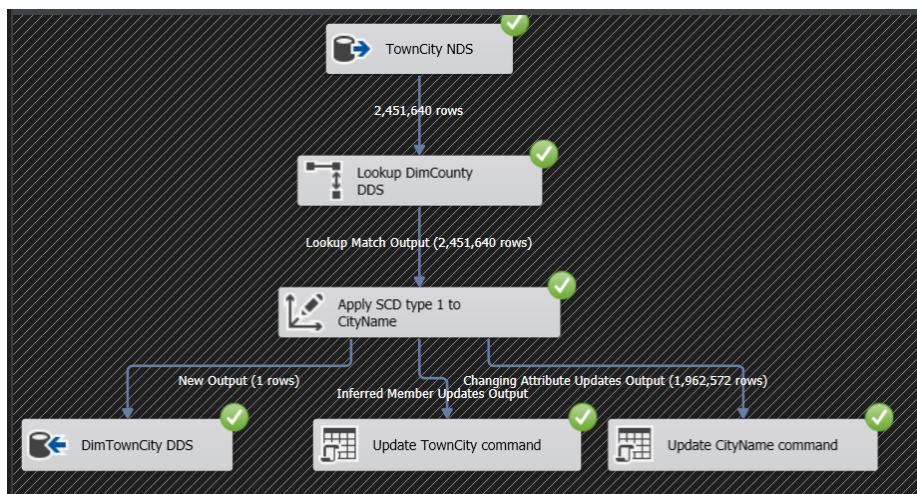
### Bảng DimRegion



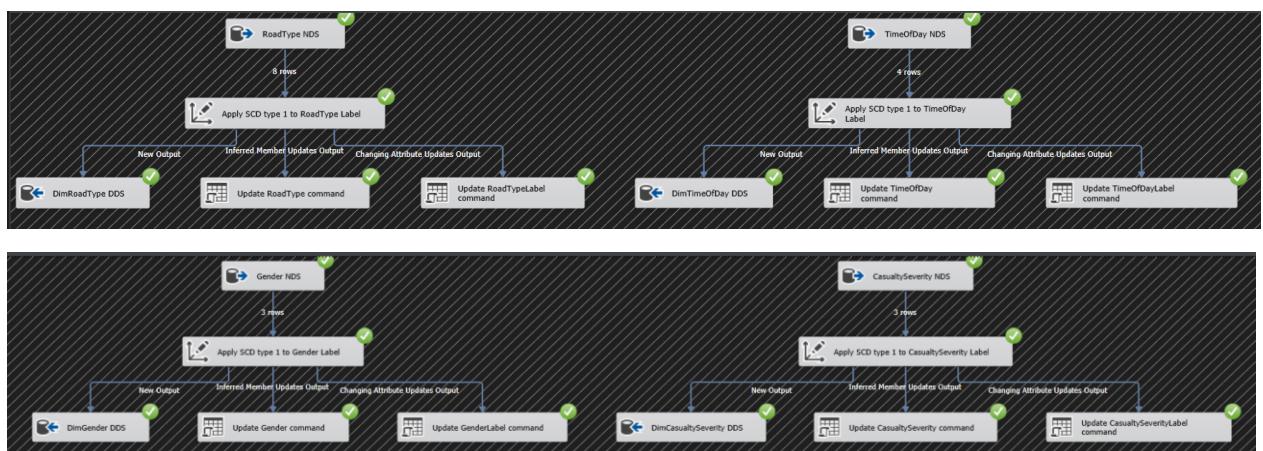
## Bảng DimCounty



## Bảng DimTownCity

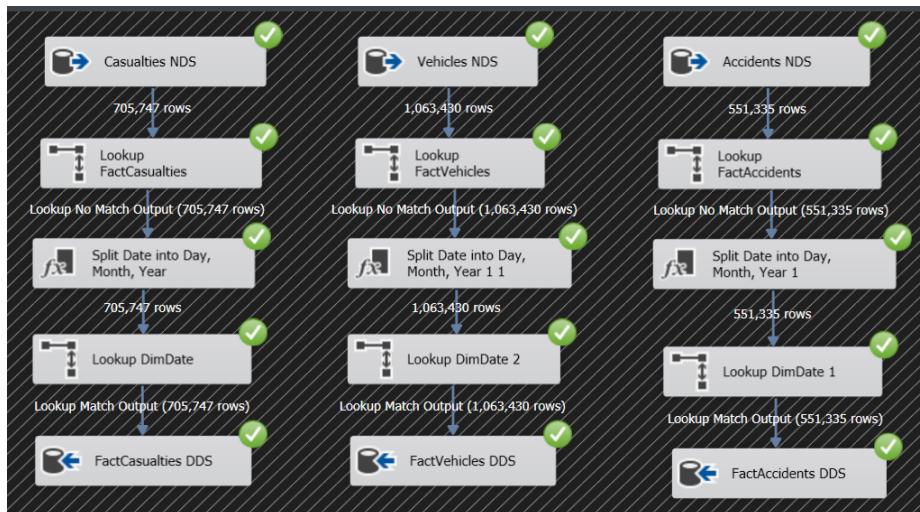


## Các bảng Dim còn lại





## Các bảng Fact

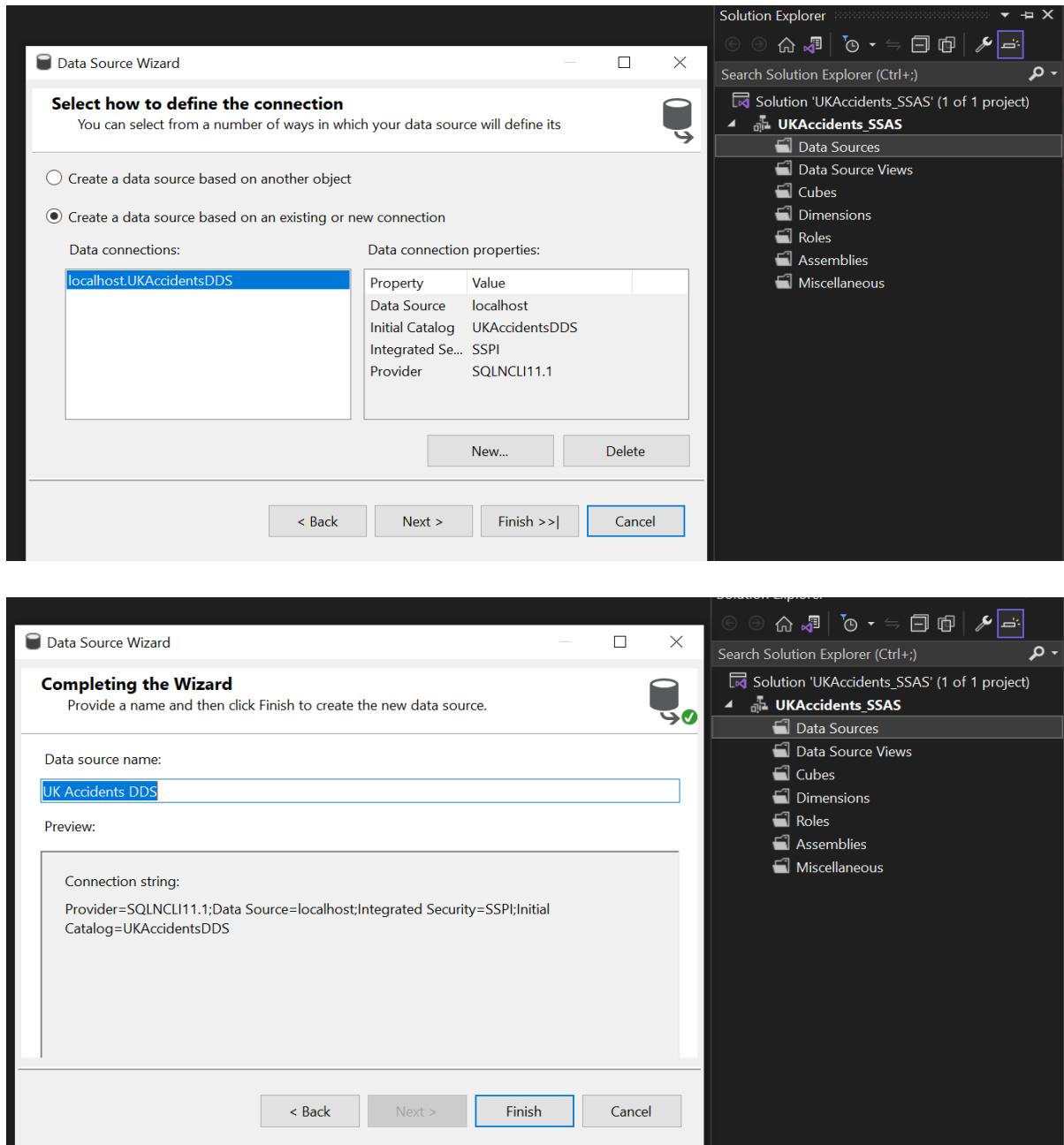


## D. QUÁ TRÌNH SSAS

### 1. Tạo Data Source:

Chuột phải vào DataSource và chọn New Data Source.Chọn Next để tiếp tục.

Connect đến Data có sẵn hoặc tạo connect mới. Connect đến DDS/ Data Warehouse đã tạo trước đó.

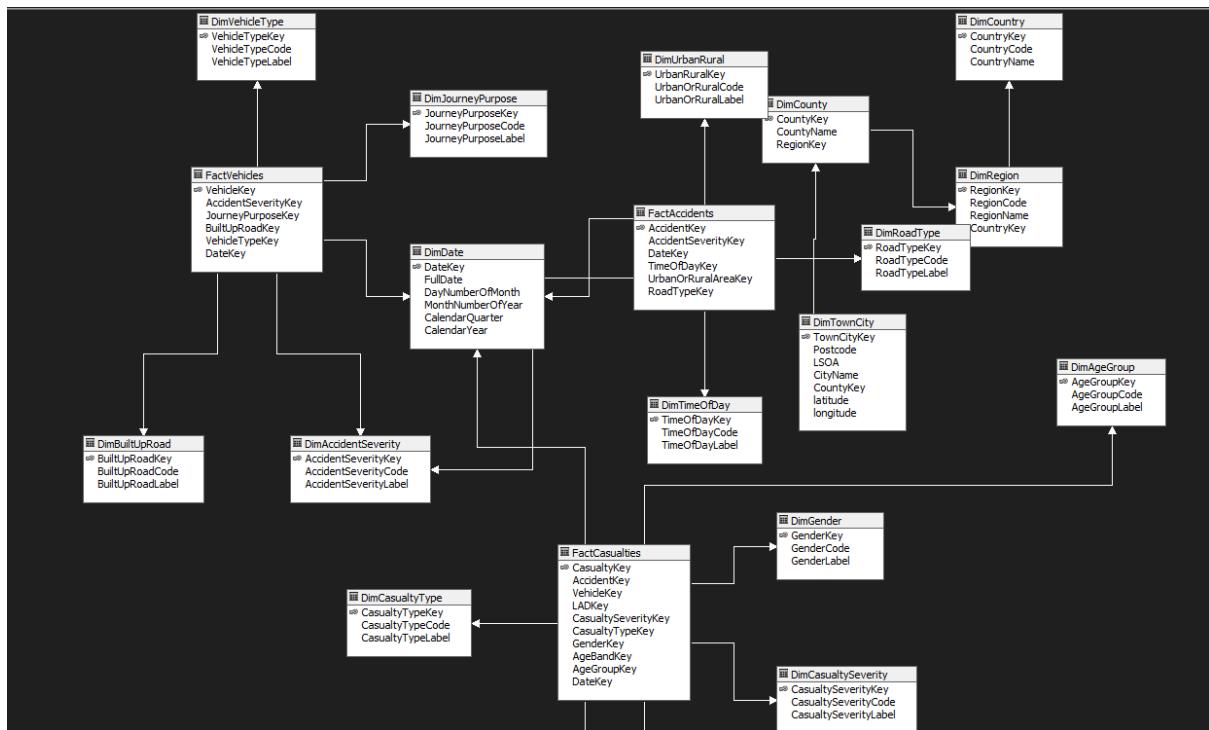
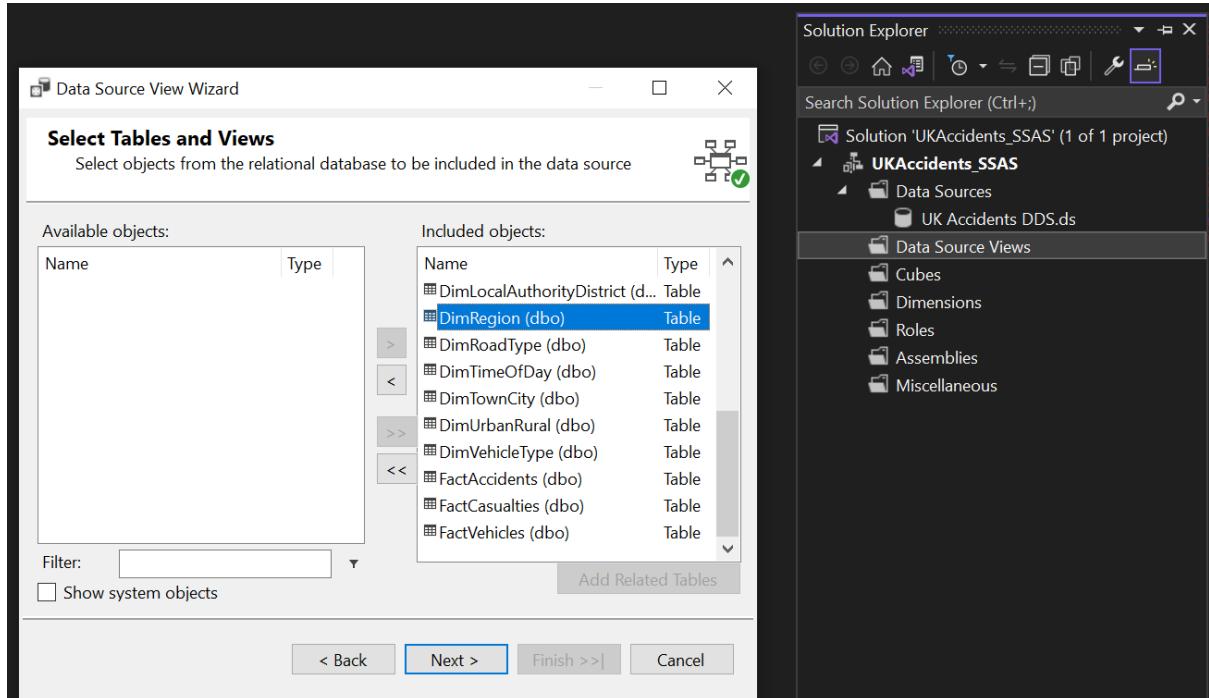


Chọn Finish để hoàn tất.

## 2. Tạo Data Source View:

Chọn chuột phải vào Data Source Views và New Data Source View để tiến hành tạo data mới.

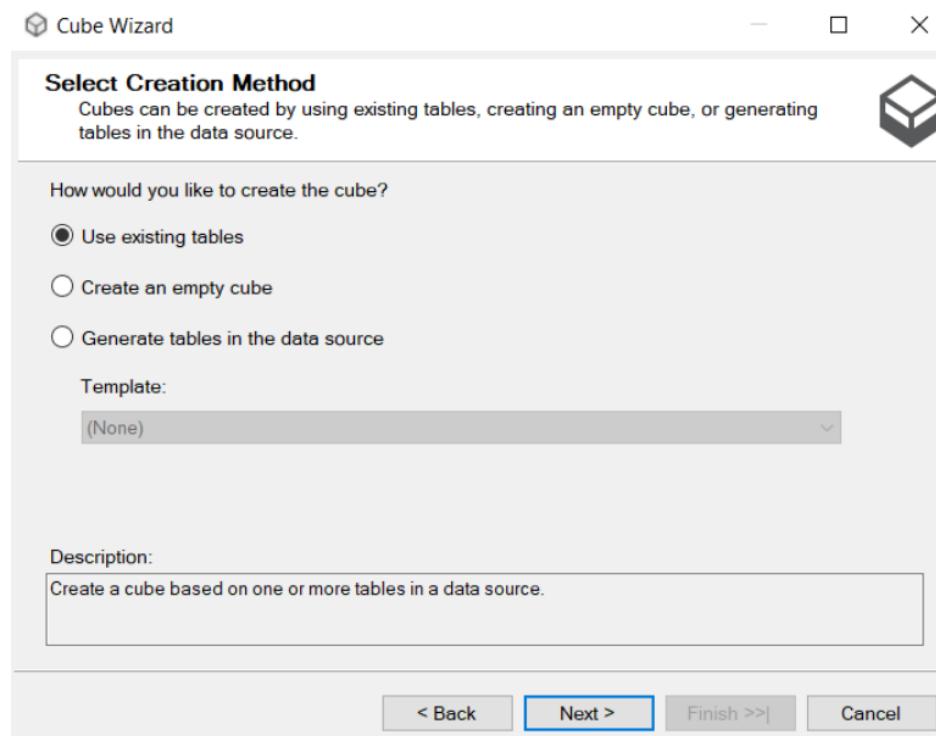
Tại **Select Tables and Views** chọn ra các bảng cần sử dụng và next



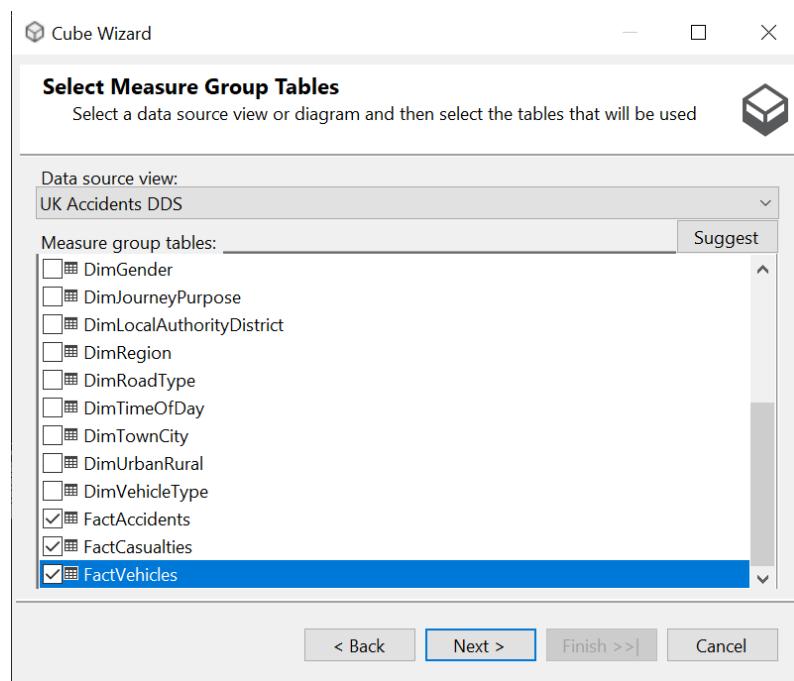
### 3. Tạo Cube

Chuột phải vào Cubes và chọn New Cube.

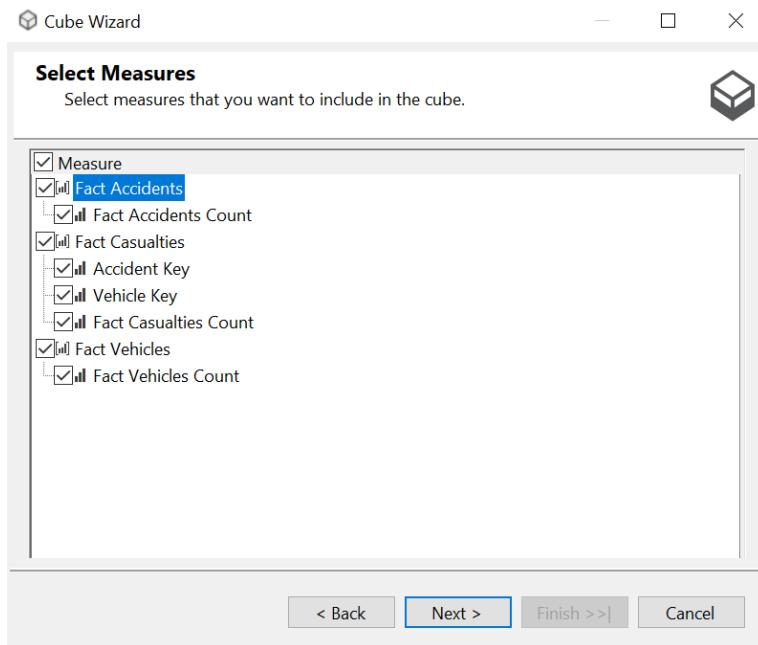
Chọn Using Existing Tables



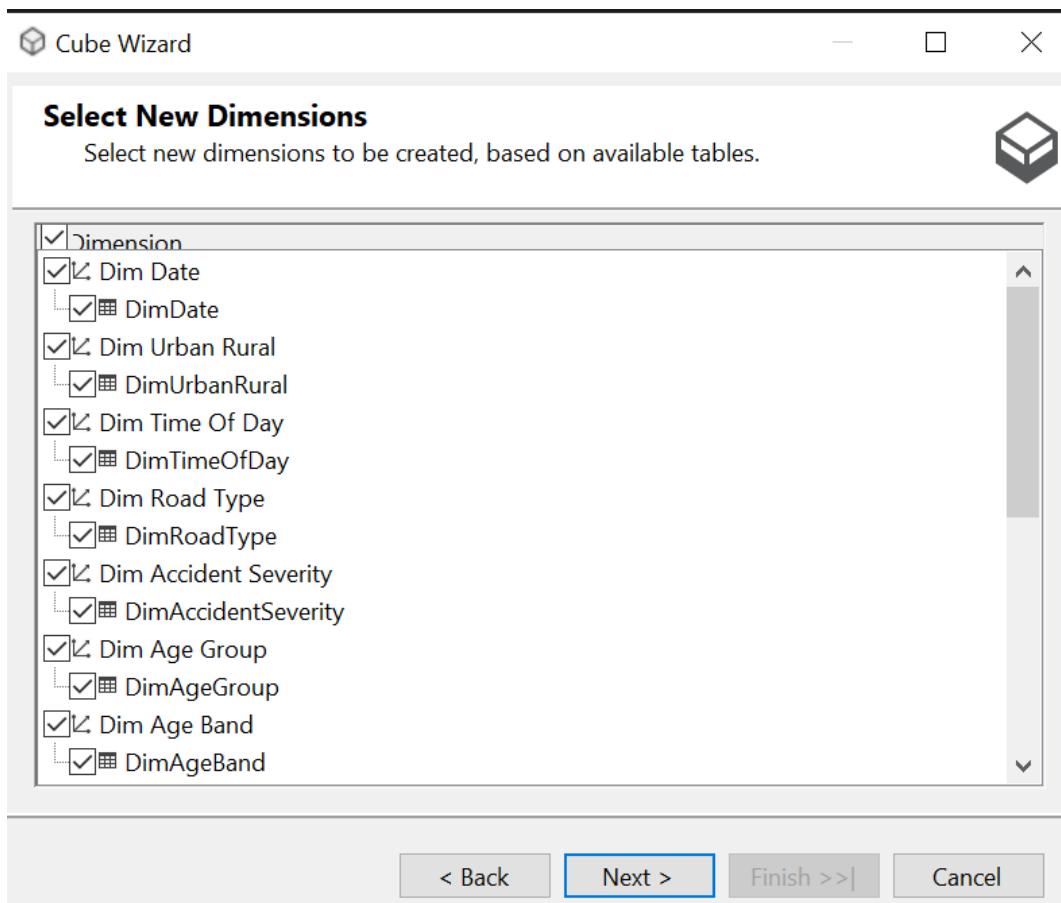
Chọn các bảng measure(fact)

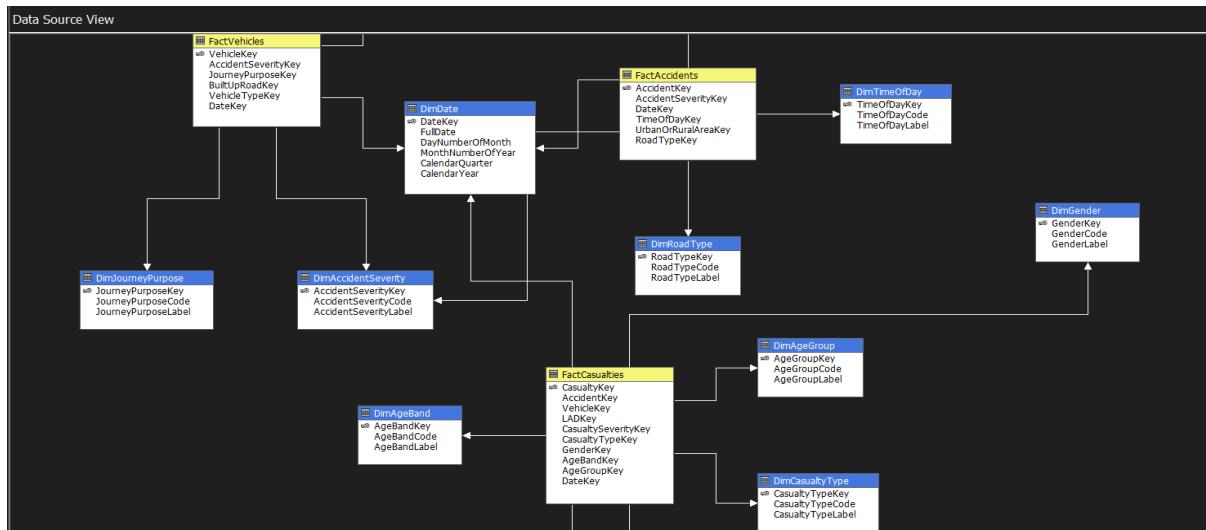


Chọn thuộc tính cần thiết trong các bảng fact



Chọn các bảng Dimensions

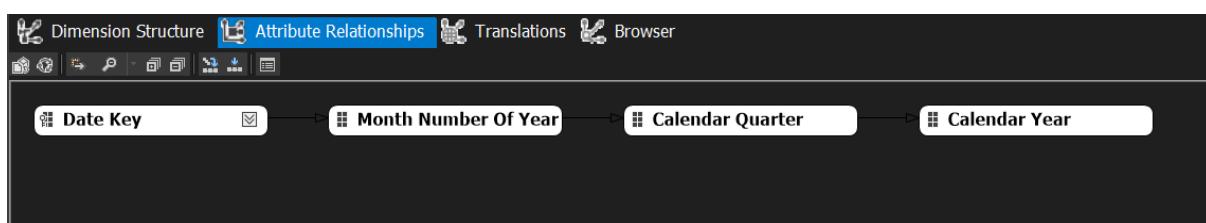




## 4. Phân chiều

### 4.1. Chiều Date

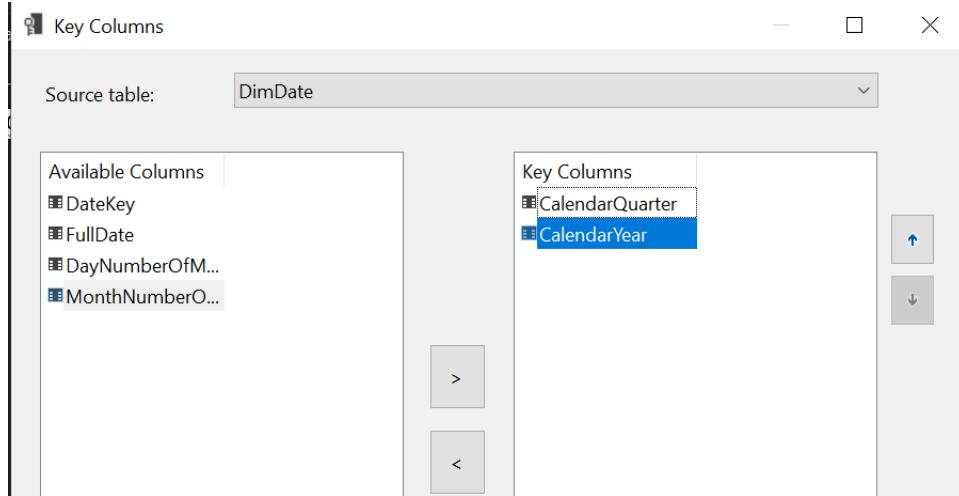
This screenshot shows the 'Dimension Structure' tab for the 'Dim Date.dim' dimension. The 'Attributes' pane lists attributes such as DateKey, FullDate, DayNumberofMonth, MonthNumberofYear, and CalendarQuarter. The 'Hierarchies' pane displays a hierarchy structure for 'Calendar' with levels: Calendar Year, Calendar Quarter, Month Number Of Year, and Date Key. A tooltip explains how to create a new hierarchy by dragging an attribute here. The 'Data Source View' pane shows the 'DimDate' table with its columns: DateKey, FullDate, DayNumberofMonth, MonthNumberofYear, CalendarQuarter, and CalendarYear.



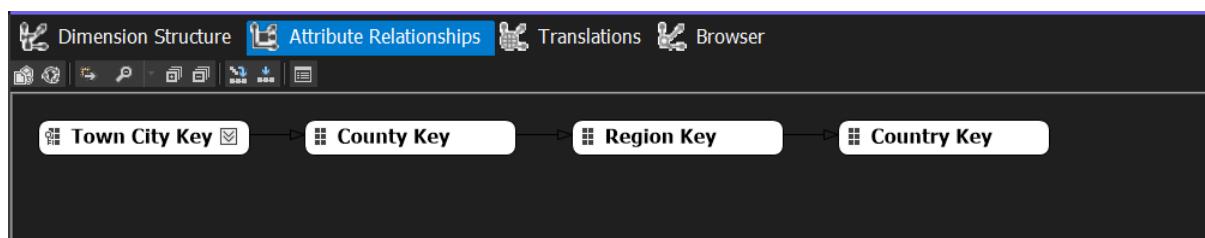
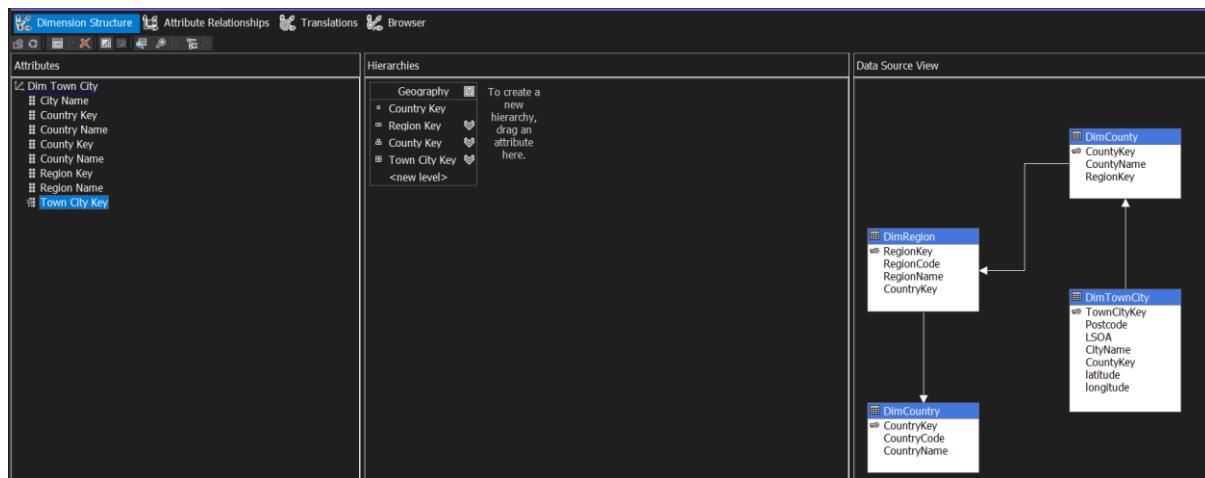
Điều chỉnh khóa chính cho bảng MonthNumberofYear

This screenshot shows the 'Key Columns' dialog for the 'DimDate' table. The 'Source table:' dropdown is set to 'DimDate'. The 'Available Columns' pane lists DateKey, FullDate, DayNumberofM..., and CalendarQuarter. The 'Key Columns' pane contains MonthNumberofYear and CalendarYear. Navigation arrows allow for moving columns between the panes.

## Điều chỉnh khóa chính cho bảng CalendarQuarter



## 4.2. Chiều Địa lý



### 4.3. Các chiều còn lại

Ví dụ: chiều TimeOfDay (các bảng còn lại tương tự)

The screenshot shows the 'Dimension Structure' tab selected in the top navigation bar. The main area is divided into three sections: 'Attributes', 'Hierarchies', and 'Data Source View'. The 'Attributes' section contains a table with one row for 'Dim Time Of Day' containing three columns: 'Time Of Day Code', 'Time Of Day Key', and 'Time Of Day Label'. The 'Hierarchies' section has a placeholder message: 'To create a new hierarchy, drag an attribute here.' The 'Data Source View' section shows a table with one row for 'DimTimeOfDay' containing three columns: 'TimeOfDayKey', 'TimeOfDayCode', and 'TimeOfDayLabel'. The 'TimeOfDayKey' column is highlighted with a blue background.

The screenshot shows the 'Attribute Relationships' tab selected in the top navigation bar. The main area is divided into two sections: 'Attributes' and 'Attribute Relationships'. The 'Attributes' section contains a table with one row for 'Time Of Day Key' containing three columns: 'Time Of Day Code', 'Time Of Day Key', and 'Time Of Day Label'. The 'Time Of Day Key' column is highlighted with a blue background. The 'Attribute Relationships' section shows a table with one row for 'Time Of Day Key' containing two columns: 'Time Of Day Code' and 'Time Of Day Label'. The 'Time Of Day Code' column is highlighted with a blue background.

## 5. Sử dụng DMX để truy vấn

- Report the number of casualties by Severity (Fatal, Serious, Slight) in the Local Authority Districts over years. / Thống kê số lượng nạn nhân theo Mức Độ Nghiêm Trọng (Fatal, Serious, Slight) ở các Địa phương (Local\_Authority\_(District)) trong tất cả các năm.

```
-- 5.1
with MEMBER MEASURES.[fatal or Serious] as sum
(
    {[Dim Casualty Severity].[Casualty Severity Key].&[1],[Dim Casualty Severity].[Casualty Severity Key].&[2]},
    [Measures].[Fact Casualties Count]
)
MEMBER MEASURES.[fatal] as sum
(
    [Dim Casualty Severity].[Casualty Severity Key].&[1],
    [Measures].[Fact Casualties Count]
)
MEMBER MEASURES.[Serious] as sum
(
    [Dim Casualty Severity].[Casualty Severity Key].&[2],
    [Measures].[Fact Casualties Count]
)
MEMBER MEASURES.[Slight] as sum
(
    [Dim Casualty Severity].[Casualty Severity Key].&[3],
    [Measures].[Fact Casualties Count]
)
MEMBER MEASURES.[All] as sum
(
    [Dim Casualty Severity].[Casualty Severity Key],
    [Measures].[Fact Casualties Count]
)
select {MEASURES.[fatal],MEASURES.[Serious],MEASURES.[fatal or Serious],MEASURES.[Slight],MEASURES.[All]} on 0,[Dim Local Authority District].[LAD Label].[LAD Label] on 1
from [UK Accidents DDS];
```

Kết quả:

	fatal	Serious	fatal or Serious	Slight	All
Aberdeen City	6	87	93	217	310
Aberdeenshire	25	177	202	367	569
Adur	5	122	127	608	735
Allerdale	30	154	184	1230	1414
Alnwick	(null)	(null)	(null)	(null)	(null)
Amber Valley	6	173	179	1172	1351
Angus	6	37	43	139	182
Argyll and Bute	15	114	129	472	601
Arun	18	259	277	1248	1525
Ashfield	14	204	218	1314	1532
Ashford	24	229	253	1749	2002
Aylesbury Vale	25	274	299	1801	2100
Babergh	16	124	140	1263	1403
Barking and Dagenham	15	157	172	2064	2236
Barnet	28	422	450	4452	4902
Barnsley	23	313	336	2577	2913
Barrow-in-Furness	6	59	65	582	647
Basildon	9	236	245	1805	2050
Basingstoke and Deane	20	330	350	1391	1741
Bassetlaw	28	224	252	1293	1545
Bath and North East Somerset	18	134	152	1388	1540
Bedford	16	227	243	1846	2089
Berwick-upon-Tweed	(null)	(null)	(null)	(null)	(null)
Bexley	11	147	158	1871	2029

2. Report the number of casualties by Severity (Fatal, Serious, Slight) in the Local Authority Districts by Quarter in years. / Thống kê số lượng nạn nhân theo Mức Độ Nghiêm Trọng ở các Địa Phương (Local\_Authority\_(District)) theo các Quý trong từng năm.

```

SELECT NON EMPTY CROSSJOIN([Dim Date].[Calendar].[Calendar Year],[Dim Date].[Calendar Quarter].[Calendar Quarter]) on columns,
NON EMPTY CROSSJOIN([Dim Local Authority District].[LAD Label].[LAD Label],[Dim Casualty Severity].[Casualty Severity Label].[Casualty Severity Label]) on rows
FROM [UK Accidents DDS]
WHERE [Measures].[Fact Casualties Count];

```

## Kết quả:

		2011	2011	2011	2011	2012	2012	2012	2012	2013	2013	2013	2013	2014	2014	2014	2014
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Barking and Dagenham	Fatal	(null)	2	1	1	(null)	1	1	1	3	(null)	3	(null)	1	1	1	(null)
Barking and Dagenham	Serious	9	8	10	15	12	7	16	7	4	11	9	11	6	8	11	13
Barking and Dagenham	Slight	125	138	131	141	134	111	125	132	95	103	120	120	122	144	151	172
Barnet	Fatal	3	3	1	1	1	1	2	3	2	4	(null)	2	(null)	(null)	1	4
Barnet	Serious	29	31	23	38	22	25	25	27	22	32	24	34	19	23	21	27
Barnet	Slight	291	313	290	270	290	289	242	267	260	227	276	296	285	306	275	275
Barnsley	Fatal	2	3	1	2	2	1	1	1	(null)	5	1	2	(null)	1	1	(null)
Barnsley	Serious	20	14	22	16	24	18	27	21	19	20	16	16	17	16	27	20
Barnsley	Slight	152	164	193	179	128	153	135	184	118	143	119	170	149	196	154	240
Barrow-in-Furness	Fatal	(null)	1	(null)	(null)	1	(null)	(null)	1	(null)	(null)	(null)	(null)	(null)	2	(null)	1
Barrow-in-Furness	Serious	3	4	5	(null)	5	(null)	3	1	6	3	4	5	7	3	4	6
Barrow-in-Furness	Slight	36	28	44	33	20	39	46	39	31	28	33	31	44	53	34	43
Basildon	Fatal	(null)	(null)	1	(null)	1	2	(null)	(null)	(null)	1	1	(null)	2	(null)	(null)	1
Basildon	Serious	11	14	12	11	16	12	13	20	11	11	21	13	21	15	16	19
Basildon	Slight	109	94	115	83	114	121	114	112	72	118	114	111	129	154	130	115
Basingstoke and Deane	Fatal	2	1	1	3	3	2	1	1	2	2	(null)	(null)	(null)	1	1	(null)
Basingstoke and Deane	Serious	18	22	24	14	9	27	23	15	12	24	23	20	18	22	33	26
Basingstoke and Deane	Slight	91	85	102	75	76	85	123	81	79	74	84	106	82	90	78	80
Bassetlaw	Fatal	2	4	1	2	2	3	2	1	1	1	1	3	2	(null)	3	(null)
Bassetlaw	Serious	17	21	10	8	12	19	22	18	10	13	20	14	13	8	10	9
Bassetlaw	Slight	96	90	76	86	86	73	100	70	78	76	83	84	72	69	79	75
Bath and North East Somerset	Fatal	(null)	1	(null)	1	(null)	(null)	1	3	2	2	(null)	2	2	1	2	1
Bath and North East Somerset	Serious	2	3	8	3	8	7	8	6	9	7	17	13	6	9	21	7

3. Report the number of accidents by Severity and Time of Day (Morning: 5am-12pm, Afternoon: 12pm-5pm, Evening: 5pm-9pm, Night: 9pm-5am) over years. /  
 Thông kê số lượng TNGT theo Mức Độ Nghiêm Trọng và Thời Điểm Trong Ngày (Morning: 5am-12pm, Afternoon: 12pm-5pm, Evening: 5pm-9pm, Night: 9pm-5am) trong các năm.

```
SELECT NON EMPTY CROSSJOIN(CROSSJOIN({[Dim Gender].[Gender Label].[Gender Label]-[Dim Gender].[Gender Label].&[Data missing or out of range]},[Dim Casualty Type].[Casualty Type Label].[Casualty Type Label]),{[Dim Age Band].[Age Band Label].[Age Band Label],[Dim Age Band].[Age Band Label]}) on rows,
NON EMPTY [Dim Date].[Calendar Year].[Calendar Year] on columns
FROM [UK Accidents DDS]
WHERE [Measures].[Fact Casualties Count];
```

## Kết quả:

		2011	2012	2013	2014
Afternoon	Fatal	500	491	500	501
Afternoon	Serious	7182	7050	7018	7248
Afternoon	Slight	44709	42678	42158	44792
Evening	Fatal	336	272	311	307
Evening	Serious	3794	3871	3755	3994
Evening	Slight	21941	20688	21167	22163
Morning	Fatal	501	457	446	469
Morning	Serious	6199	6062	5988	6559
Morning	Slight	41530	39988	40497	42751
Night	Fatal	376	325	330	357
Night	Serious	2441	2394	2423	2448
Night	Slight	10759	10148	10134	10327

4. Report the number of accidents by Severity, Urban or Rural Area and Road Type over years. / Thống kê số lượng TNGT theo Mức Độ Nghiêm Trọng, Vùng (Urban\_or\_Rural\_Area), và Kiểu Đường (Road Type) trong các năm.

```

SELECT NON EMPTY CROSSJOIN([Dim Road Type].[Road Type Label].[Road Type Label],[Dim Accident Severity].[Accident Severity Label].[Accident Severity Label]) on rows,
NON EMPTY CROSSJOIN([Dim Date].[Calendar Year].[Calendar Year],[Dim Urban Rural].[Urban Or Rural Label].[Urban Or Rural Label]) on columns
FROM [UK Accidents DDS]
WHERE [Measures].[Fact Accidents Count];

```

Kết quả:

		2011	2011	2012	2012	2013	2013	2014	2014
		Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
Dual carriageway	Fatal	230	101	188	88	232	82	202	101
Dual carriageway	Serious	1285	1166	1203	1091	1216	1112	1393	1210
Dual carriageway	Slight	8437	8546	8116	8176	8237	8394	8517	8911
One way street	Fatal	3	14	5	13	(null)	10	3	16
One way street	Serious	41	283	39	271	31	317	37	356
One way street	Slight	201	2017	154	1778	184	2054	150	2229
Roundabout	Fatal	13	11	13	17	12	12	9	9
Roundabout	Serious	375	559	367	524	371	551	396	639
Roundabout	Slight	3323	5510	3240	5323	3090	5225	3359	5768
Single carriageway	Fatal	881	442	765	447	818	404	825	452
Single carriageway	Serious	6524	9189	6352	9329	6403	8990	6747	9253
Single carriageway	Slight	28315	60665	27291	57627	26995	58099	27237	62064
Slip road	Fatal	8	4	5	3	13	1	8	2
Slip road	Serious	92	40	104	45	78	56	102	45
Slip road	Slight	836	486	850	464	764	490	825	545
Unknown	Fatal	5	1	1	(null)	3	(null)	4	3
Unknown	Serious	22	40	18	34	25	34	29	42
Unknown	Slight	198	405	145	338	154	270	144	284

5. Report the number of accidents by Journey Purpose and Vehicle Type / Tổng hợp số lượng tai nạn theo Mục Đích Hành Trình (Journey Purpose) và Loại Phương Tiện (Vehicle\_Type).

```

SELECT NON EMPTY CROSSJOIN([Dim Road Type].[Road Type Label].[Road Type Label],[Dim Accident Severity].[Accident Severity Label].[Accident Severity Label]) on rows,
NON EMPTY CROSSJOIN([Dim Date].[Calendar Year].[Calendar Year],[Dim Urban Rural].[Urban Or Rural Label].[Urban Or Rural Label]) on columns
FROM [UK Accidents DDS]
WHERE [Measures].[Fact Accidents Count];

```

**Kết quả:**

	Commuting to/from work	Data missing or out of range	Journey as part of work	Not known	Other	Pupil riding to/from school
Agricultural vehicle	36	(null)	1759	443	18	2
Bus or coach (17 or more pass seats)	117	(null)	22038	3072	54	18
Car	74345	3	72826	585059	16757	1186
Data missing or out of range	(null)	(null)	2	108	1	(null)
Electric motorcycle	2	(null)	1	15	(null)	1
Goods 7.5 tonnes mgw and over	184	1	17109	2517	31	6
Goods over 3.5t. and under 7.5t	210	(null)	5149	1992	32	(null)
Goods vehicle - unknown weight	7	(null)	200	177	2	(null)
Minibus (8 - 16 passenger seats)	109	(null)	1510	881	18	4
Mobility scooter	3	(null)	16	387	48	2
Motorcycle - unknown cc	21	(null)	9	247	8	1
Motorcycle 125cc and under	5053	(null)	3374	21975	529	203
Motorcycle 50cc and under	1364	(null)	1078	8254	303	366
Motorcycle over 125cc and up to 500cc	1418	(null)	866	7054	160	31
Motorcycle over 500cc	4419	(null)	2319	22311	639	29
Other vehicle	262	(null)	3003	4133	39	10
Pedal cycle	12140	1	5701	60101	1599	1884
Ridden horse	(null)	(null)	19	482	21	(null)
Taxi/Private hire car	441	(null)	14959	5069	79	10
Tram	(null)	(null)	76	4	(null)	(null)
Van / Goods 3.5 tonnes mgw or under	4117	(null)	25382	22020	551	20

6. Report the number of casualties by Severity, Casualty Type and Age Group over years. / Thống kê số lượng nạn nhân theo Mức Độ Nghiêm Trọng, Loại Nạn Nhân (Casualty Type) và Độ Tuổi trong các năm.

```

SELECT NON EMPTY CROSSJOIN([Dim Casualty Type].[Casualty Type Label].[Casualty Type Label], {[Dim Age Group].[Age Group Label].[Age Group Label],[Dim Age Group].[Age Group Label]}) on rows,
NON EMPTY CROSSJOIN([Dim Date].[Calendar Year].[Calendar Year],[Dim Casualty Severity].[Casualty Severity Label].[Casualty Severity Label]) on columns
FROM [UK Accidents DDS]
WHERE [Measures].[Fact Casualties Count];

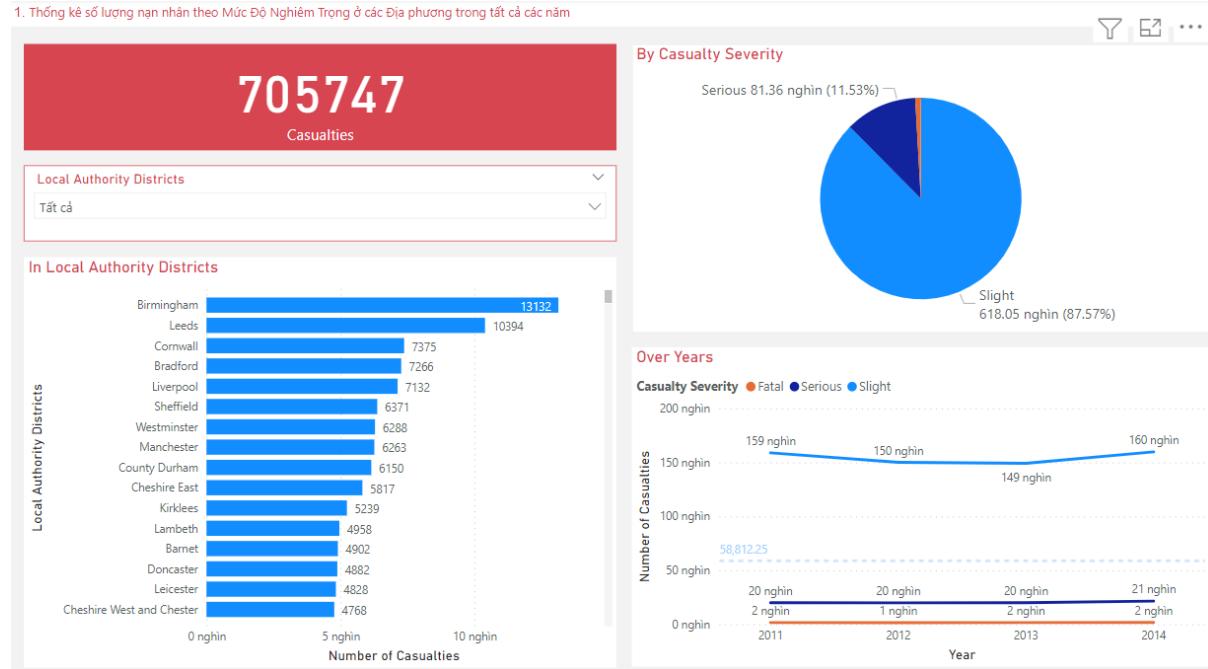
```

### Kết quả:

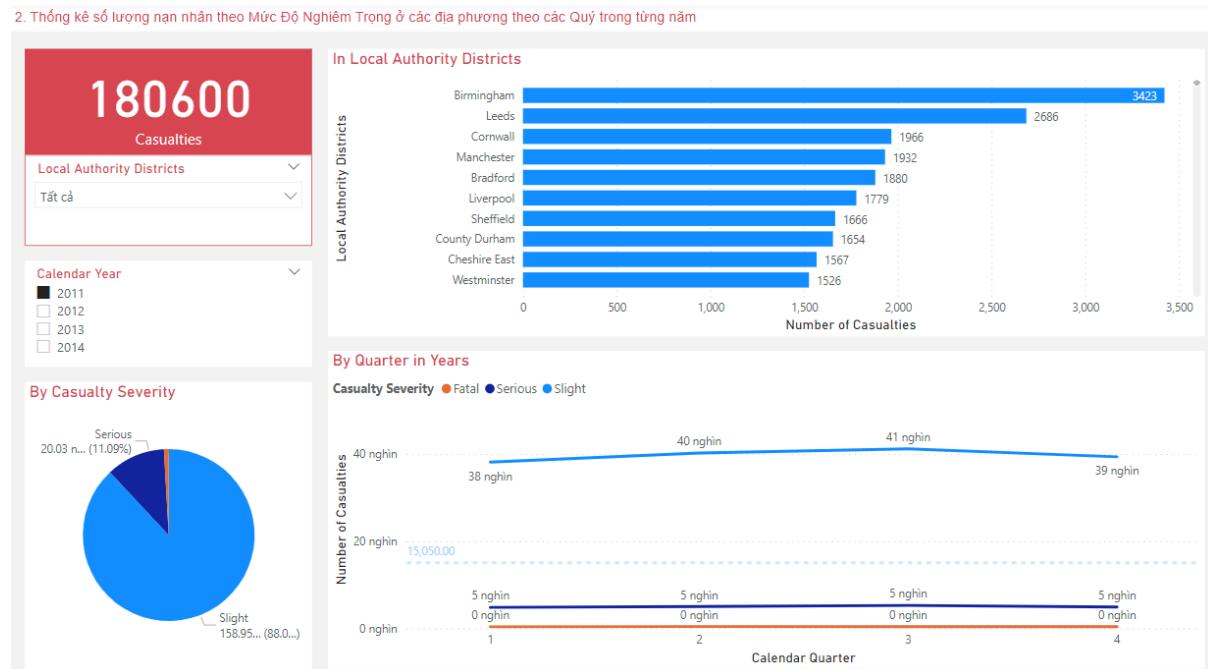
		2011	2011	2011	2012	2012	2012	2013	2013	2013	2014	2014
		Fatal	Serious	Slight	Fatal	Serious	Slight	Fatal	Serious	Slight	Fatal	Serious
Motorcycle over 125cc and up to 500cc rider or passenger	Young adult	2	24	77	4	27	59	1	29	47	1	13
Motorcycle over 125cc and up to 500cc rider or passenger	All	31	537	1427	27	543	1430	26	513	1362	20	506
Motorcycle over 500cc rider or passenger	60 and over	27	173	298	19	170	243	20	202	274	22	221
Motorcycle over 500cc rider or passenger	Adult	201	2114	4201	177	1831	3623	206	1983	3870	213	2146
Motorcycle over 500cc rider or passenger	Children	(null)	3	14	(null)	1	14	1	3	11	(null)	4
Motorcycle over 500cc rider or passenger	Young adult	1	22	48	3	10	34	1	6	29	(null)	13
Motorcycle over 500cc rider or passenger	All	229	2312	4561	199	2012	3914	228	2194	4184	235	2384
Other vehicle occupant	60 and over	9	41	164	4	45	134	6	38	109	3	17
Other vehicle occupant	Adult	8	77	567	1	73	525	4	71	346	5	69
Other vehicle occupant	Children	(null)	7	37	(null)	8	27	(null)	3	21	(null)	12
Other vehicle occupant	Young adult	(null)	3	23	(null)	8	27	(null)	7	13	(null)	7
Other vehicle occupant	All	17	128	791	5	134	713	10	119	489	8	105
Pedestrian	60 and over	160	979	2317	160	1054	2261	136	1026	2375	170	1168
Pedestrian	Adult	195	2124	8981	182	2314	8748	201	2194	9263	196	2118
Pedestrian	Children	27	1273	5141	17	1247	4508	21	1185	4384	25	1169
Pedestrian	Young adult	13	302	1294	7	314	1185	11	241	1235	14	231
Pedestrian	All	395	4678	17733	366	4929	16702	369	4646	17257	405	4686
Taxi/Private hire car occupant	60 and over	(null)	21	192	1	23	177	1	28	211	1	27
Taxi/Private hire car occupant	Adult	6	116	2016	7	112	2112	6	114	1961	2	96
Taxi/Private hire car occupant	Children	(null)	4	104	(null)	3	79	1	5	84	(null)	5

## E. REPORT SỬ DỤNG POWER BI

1. Thống kê số lượng nạn nhân theo Mức Độ Nghiêm Trọng (Fatal, Serious, Slight) ở các Địa phương (Local\_Authority\_(District)) trong tất cả các năm.

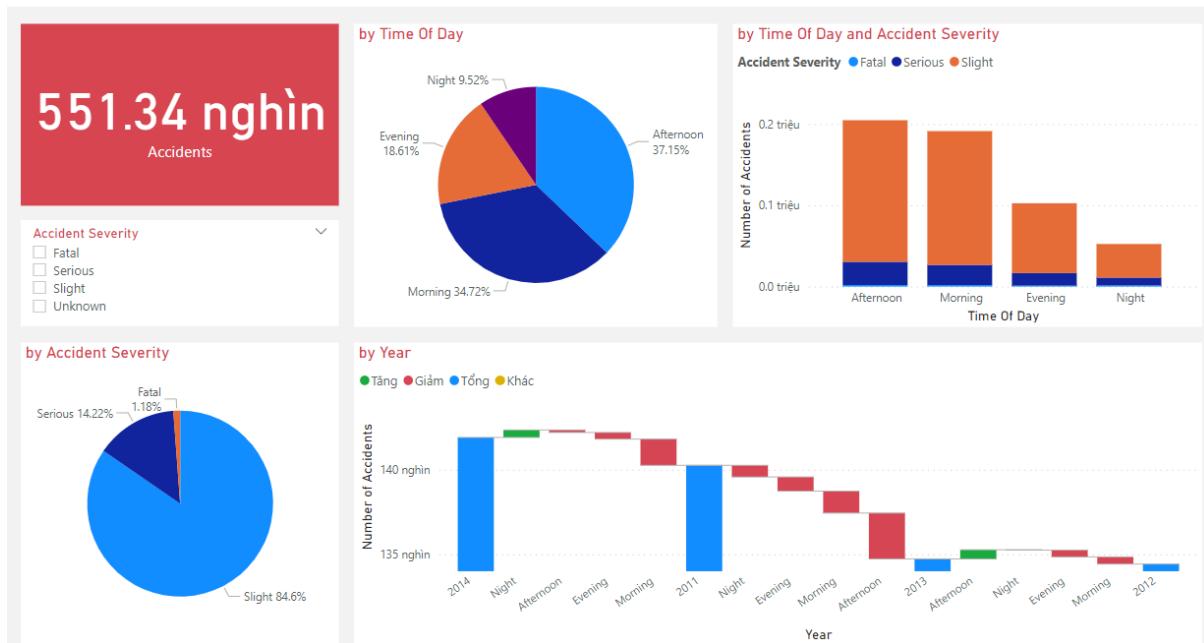


2. Thống kê số lượng nạn nhân theo Mức Độ Nghiêm Trọng ở các Địa Phương (Local\_Authority\_(District)) theo các Quý trong từng năm.



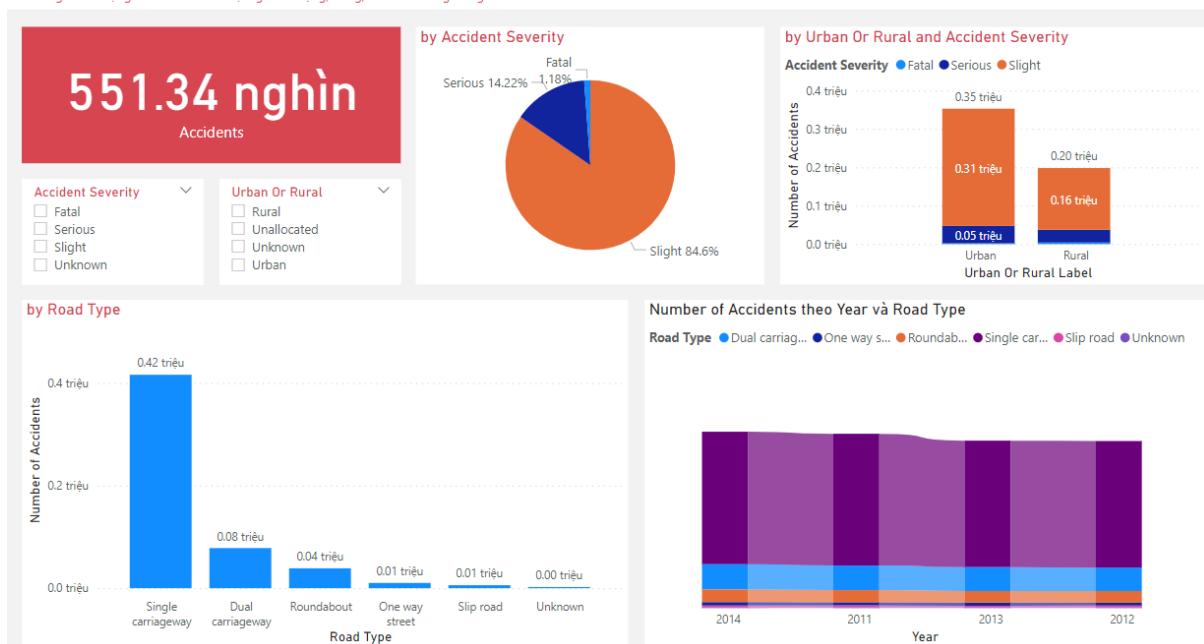
3. Thống kê số lượng TNGT theo Mức Độ Nghiêm Trọng và Thời Điểm Trong Ngày (Morning: 5am-12pm, Afternoon: 12pm-5pm, Evening: 5pm-9pm, Night: 9pm-5am) trong các năm.

4. Thống kê số lượng TNGT theo Mức Độ Nghiêm Trọng và Thời Điểm Trong Ngày (Morning: 5am-12pm, Afternoon: 12pm-5pm, Evening: 5pm-9pm, Night: 9pm-5am) trong các năm



4. Thống kê số lượng TNGT theo Mức Độ Nghiêm Trọng, Vùng (Urban\_or\_Rural\_Area), và Kiểu Đường (Road Type) trong các năm.

5. Thống kê số lượng TNGT theo Mức Độ Nghiêm Trọng, Vùng, và Kiểu Đường trong các năm



## 5. Tổng hợp số lượng tử vong theo Giới tính, Loại nạn nhân và Nhóm tuổi theo các năm

3. Thống kê số lượng người tử vong theo Giới Tính, Loại Nạn Nhân và Nhóm Tuổi theo các năm



## 6. Thống kê số lượng nạn nhân theo Mức Độ Nghiêm Trọng, Loại Nạn Nhân (Casualty Type) và Độ Tuổi trong các năm.

6. Thống kê số lượng nạn nhân theo Mức Độ Nghiêm Trọng, Loại Nạn Nhân (Casualty Type) và Độ Tuổi trong các năm, Độ Tuổi

