**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**NGÀNH CÔNG NGHỆ THÔNG TIN**

****

**PROJECT SUBMISSION 2**

Phạm Ngọc Thùy Trang 18127022

Ngô Thị Thanh Thảo 18127053

Hoàng Nguyên Trúc 18127055

**Môn học: Hệ thống thông tin phục vụ trí tuệ kinh doanh**

**Thành phố Hồ Chí Minh**

**Ộ MÔN HỆ**

# Mục lục

[I. Mục lục 1](#_Toc88312186)

[II. Bảng thông tin chi tiết nhóm 2](#_Toc88312187)

[III. Phân công công việc 3](#_Toc88312188)

[IV. Mô tả ý nghĩa các thuộc tính của nguồn dữ liệu 4](#_Toc88312189)

[1. Postcode 4](#_Toc88312190)

[2. LSOA 5](#_Toc88312191)

[3. Bảng Casualties: 11](#_Toc88312192)

[4. Bảng Accidents: 15](#_Toc88312193)

[5. Bảng Vehicles: 19](#_Toc88312194)

[V. ETL từ Source sang Stage 22](#_Toc88312195)

[1. Xử lý dữ liệu khi từ Source sang Stage 22](#_Toc88312196)

[2. Các bước thực hiện khi đổ từ bảng nguồn vào stage 26](#_Toc88312197)

[3. Flow khi thực hiện trên Visual Studio Code 27](#_Toc88312198)

[VI. Stage sang NDS 30](#_Toc88312199)

[1. Thiết kế NDS dưới dạng mô hình ER 30](#_Toc88312200)

[2. Lược đồ quan hệ dưới dạng physical: 37](#_Toc88312201)

[3. Flow khi chuyển đổi dữ liệu từ Stage sang NDS 58](#_Toc88312202)

[VII. References 68](#_Toc88312203)

# Bảng thông tin chi tiết nhóm

|  |  |  |  |
| --- | --- | --- | --- |
| **Mã nhóm:** | Nhóm 2 | | |
| **Số lượng:** | 3 | | |
| **MSSV** | **Họ tên** | **Email** | **SĐT** |
| 18127022 | Phạm Ngọc Thùy Trang | 18127022@  student.hcmus.edu.vn |  |
| 18127053 | Ngô Thị Thanh Thảo | 18127053@  student.hcmus.edu.vn |  |
| 18127055 | Hoàng Nguyên Trúc | 18127055@  student.hcmus.edu.vn |  |

# Phân công công việc

|  |  |  |
| --- | --- | --- |
| **Họ tên thành viên** | **Nội dung công việc** | **Phần trăm hoàn thành (%)** |
| Phạm Ngọc Thùy Trang |  |  |
| Ngô Thị Thanh Thảo |  |  |
| Hoàng Nguyên Trúc |  |  |

# Mô tả ý nghĩa các thuộc tính của nguồn dữ liệu

## Postcode

**Các thuộc tính sử dụng:**

* **Postcode (Mã bưu chính):** là mã được quy định bởi liên hiệp bưu chính toàn cầu giúp định vị khi chuyển thư, bưu phẩm. Hiểu đơn giản, mã bưu chính là một “tọa độ” của địa chỉ trên “bản đồ bưu chính”. Mỗi quốc gia có một hệ thống ký hiệu mã bưu chính riêng. Mỗi mã bưu chính trong một quốc gia, ký hiệu cho một địa phận bưu điện duy nhất. Địa phận bưu điện có thể là một làng nhỏ, một thành phố nhỏ, một quận của một thành phố lớn.
* **Latitude:** là vĩ độ - một giá trị xác định vị trí của một điểm trên bề mặt Trái Đất ở phía bắc hay phía nam của xích đạo.
* **Longitude:** là kinh độ - một giá trị tọa độ địa lý theo hướng đông-tây, được sử dụng phổ biến nhất trong bản đồ học và hoa tiêu toàn cầu
* **City:** Thành phố
* **County (Hạt):** Là một khu vực nằm trong một bang và lớn hơn thành phố, có chính quyền riêng để xử lý các công việc địa phương. Ví dụ, Hạt El Paso nằm trong bang Colorado, có thành phố Colorado Springs và các cộng đồng dân cư khác.
* **Region\_code:** mã vùng (mã khu vực địa lý). Thông thường, một khu vực địa lý có thể bao gồm nhiều quốc gia khác nhau.
* **Region name:** Tên khu vực địa lý tương ứng với region\_code
* **Country\_code:** Mã quốc gia (một loại mã địa lý bằng chữ hoặc bằng số được dùng để đại diện cho một quốc gia hoặc một khu vực phụ thuộc)
* **Country\_name:** Tên quốc gia tương ứng với country\_code

## LSOA

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên thuộc tính** | **Ý nghĩa/Mô tả thuộc tính** | **Phạm vi giá trị của mã** | **Độ dài thuộc tính** |
| **pcd7** | Còn được gọi là unit postcode (mã bưu chính độc nhất) phiên bản dài 7 ký tự. Pdc7 chính là những mã bưu chính còn hoạt động cho đến thời điểm hiện tại nằm trong vương quốc Anh, Channel Islands và Isle of Man có nhận thư hàng tháng từ Royal Mail - công ty dịch vụ bưu chính và chuyển phát nhanh đa quốc gia của Anh, cũng như tất cả những mã bưu chính bị đóng và không được sử dụng lại bởi Royal Mail trong vương quốc Anh và bởi các cơ quan hành chính của Channel Island và Isle of Man | AB1Δ1AA-ZE999ZZ (trong đó Δ đại diện cho ký tự khoảng trắng) | 7 ký tự (bao gồm ký tự khoảng trắng) |
| **pcd8** | Như trên, chỉ khác ở điểm đây là mã bưu chính độc nhất phiên bản 8 ký tự | AB1ΔΔ1AAZE99Δ9ZZ (trong đó Δ là ký tự khoảng trắng) | 8 ký tự (bao gồm ký tự khoảng trắng) |
| **oac11cd** | Viết tắt của cụm từ Output Area và có tên gọi đầy đủ là 2011 Census Output Area (hoặc là SA - Small Area - những khu vực nhỏ trong vương quốc Anh). Trong đó, pseudo oac11cd code sẽ được dùng cho Channel Islands và Isle of Man. oac11cd code sẽ để trống đối với các mã bưu chính không nằm trong hệ thống lưới. Oac11d được dùng để định nghĩa một cấp độ địa lý có cấp cao hơn unit postcode. Một output area sẽ có thể có hoặc không có khu vực kế vị (cũng là một output area), nhưng sẽ phải có một khu vực cha - hay còn gọi là SOA (Lower Super Output Area) | E00XXXXXX, W00XXXXXX, S00XXXXXX, 95AA010001- 95ZZ519999, pseudo hoặc null (không có thông tin).  Cụ thể hơn:  E00000001 - E00165665 = England; W00000001 - W00009769 = Wales; S00000001 - S00042604 = Scotland; 95AA010001 - 95ZZ519999 = Northern Ireland; L99999999 (pseudo code) = Channel Islands; M99999999 (pseudo code) = Isle of Man | 9 - 10 ký tự (hoặc chỉ có 1 ký tự nào trong trường hợp không có thông tin) |
| **ladnmw** | Tên khu vực tương ứng với mã oac11cd, nếu không có thì để trống |  |  |
| **lsoa11cd** | Viết tắt của cụm từ Lower Super Output Area và có tên gọi đầy đủ là 2011 Census Lower Layer Super Output Area, hoặc là Data Zone (DZ). **lsoa11cd** đại diện cho mã 2011 LSOA của khu vực England và xứ Wales, đại diện cho mã SOA của khu vực Northern Ireland and DZ code cho Scotland. Trong đó, pseudo **lsoa11cd** code sẽ được dùng cho Channel Islands và Isle of Man. **lsoa11cd** code sẽ để trống đối với các mã bưu chính không có mã OA. Một LSOA có thể có nhiều hoặc không có các output area. Thông thường, một LSOA sẽ phải có một MSOA (middle super output area). **lsoa11cd** sẽ có tên gọi là DZONE1 dùng để chỉ những mã bưu chính nằm trong Scotland, và sẽ được gọi là SOA1 khi nó đại diện cho những mã bưu chính nằm trong England và Wales. | E01XXXXXX  S01XXXXXX  W01XXXXXX  Null (nếu không có thông tin hiển thị)  Cụ thể hơn:  E01000001- E01032482 = England;  W01000001- W01001896 = Wales;  S99999999 (pseudo) = Scotland;  N99999999 (pseudo) = Northern Ireland;  L99999999 (pseudo) = Channel Islands;  M99999999 (pseudo) = Isle of Man | 9 ký tự (hoặc chỉ 1 ký tự null nếu không có thông tin) |
| **lsoa11nm** | Tên khu vực tương ứng với mã **lsoa11cd**, nếu không có thì để trống |  |  |
| **msoa11cd** | Viết tắt của cụm từ Lower Super Output Area và có tên gọi đầy đủ là 2011 Middle Lower Layer Super Output Area đối với khu vực England và Wales hoặc có tên là Intermediate Zone(IZ) đối với khu vực Scotland. Trong đó, pseudo **msoa11cd** code sẽ được dùng cho Channel Islands và Isle of Man. **msoa11cd** code sẽ để trống đối với các mã bưu chính nằm trong hệ thống lưới. Một MSOA có thể có nhiều hoặc không có LSOA cũng như có thể có hoặc không có các workplace zone. Thông thường, một MSOA sẽ phải có một khu vực cha gọi là Local Authority - LA (Chính quyền địa phương). Tuy nhiên, nó có thể có hoặc không có LSOA cũng như workplace zone. | * **E02XXXXXX** * **S02XXXXXX** * **W02XXXXXX**   Cụ thể là:  E02000001- E02006781 = England; W02000001- W02000413 = Wales; S99999999 (pseudo) = Scotland; N99999999 (pseudo) = Northern Ireland; L99999999 (pseudo) = Channel Islands; M99999999 (pseudo) = Isle of Man | 9 ký tự (hoặc chỉ 1 ký tự null nếu không có thông tin) |
| **msoa11nm** | Tên khu vực tương ứng với mã **msoa11cd**, nếu không có thì để trống. |  |  |
| **ladcd** | Viết tắt của cụm từ Local Authority, **ladcd** chính là  Quận hiện tại mà mã bưu chính đã được chỉ định. LA có rất nhiều tên đầy đủ khác nhau, trong đó có Local Authority District (LAD), hoặc là unitary authority (UA), hoặc là metropolitan district (MD), hoặc London borough (LB), hoặc council area (CA) hoặc district council area (DCA). Trong đó, pseudo **ladcd** code sẽ được dùng cho Channel Islands và Isle of Man. **ladcd** code sẽ để trống đối với các mã bưu chính không nằm trong hệ thống lưới. Một LA có thể có nhiều hoặc không có các phường cũng như có thể có hoặc không có các khu vực cộng đồng. Tuy nhiên, một LA cần phải thuộc về một quốc gia nào đó. | E06XXXXXX  E07XXXXXX  E08XXXXXX  E09XXXXXX  N09XXXXXX  S12XXXXXX  W06XXXXXX  Cụ thể hơn:  E06000001 - E06000056 = England (UA); E07000001 - E07000239 = England (LAD); E08000001 - E08000036 = England (MD); E09000001 - E09000033 = England (LB); W06000001 - W06000024 = Wales (UA); S12000005 - S12000044 = Scotland (CA); 95A – 95Z = Northern Ireland (DCA); L99999999 (pseudo) = Channel Islands; M99999999 (pseudo) = Isle of Man |  |
| **ladnm** | Tên chính quyền địa phương tương ứng với mã ladcd, nếu không có thì để trống |  |  |

## Bảng Casualties:

**Các thuộc tính sử dụng:**

* **Accident\_Index:** mã số tai nạn
* **Sex\_of\_Casualty:** giới tính nạn nhân. Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 1 | Nam |
| 2 | Nữ |
| 3 | Không rõ |
| -1 | Dữ liệu bị thiếu hoặc nằm ngoài phạm vi dữ liệu có sẵn |

* **Age\_Band\_of\_Casualty:** nhóm độ tuổi thương vong (0-5, 6-10, 11-15,...hơn 75 tuổi). Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | 0-5 |
| 2 | 6-10 |
| 3 | 11-15 |
| ... | ... |

* **Casualty\_Severity:** mức độ nghiêm trọng gồm ba mức độ: gây tử vong,

nghiêm trọng, nhẹ. Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 1 | Gây tử vong |
| 2 | Nghiêm trọng |
| 3 | Nhẹ |

* **Casualty\_Type:** loại nạn nhân. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 0 | Người dân |
| 1 | Người đi xe đạp |
| 2 | Hành khách hoặc những người đi xe moto có phân khối <= 50cc |
| ... | ... |

* **Age\_of\_Casualty:** độ tuổi của nạn nhân bị thương vong.
* **Age\_Band\_of\_Casualty**: các nhóm độ tuổi của nạn nhân bị thương vong. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | 0-5 |
| 2 | 6-10 |
| 3 | 11-15 |

* **Casualty\_Home\_Area\_Type**: loại khu vực nạn nhân sinh sống. Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 1 | Khu vực thành thị |
| 2 | Thị trấn nhỏ |
| 3 | Nông thôn |
| -1 | Dữ liệu bị thiếu hoặc nằm ngoài phạm vi dữ liệu có sẵn |

* **Casualty\_Class**: lớp nạn nhân (gồm 3 loại: người lái xe hoặc tài xế, cư dân, hành khách). Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 1 | Tài xế hoặc người lái xe |
| 2 | Hành khách |
| 3 | Cư dân |

* **Car\_passenger:** Vị trí của hành khách bị xảy ra tai nạn nếu có đi ô tô. Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 0 | Không phải hành khách đi ô tô |
| 1 | Hành khách ngồi ghế trước |
| 2 | Hành khách ngồi ghế sau |
| -1 | Dữ liệu bị thiếu hoặc nằm ngoài phạm vi dữ liệu có sẵn |

* **Bus\_passenger:** Vị trí của hành khách bị xảy ra tai nạn nếu có đi xe buýt. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 0 | Không phải là hành khách đi xe buýt hoặc xe khách |
| 1 | Lên xe |
| 2 | Xuống xe |
| 3 | Hành khách đứng trên xe |
| 4 | Hành khách ngồi trên xe |
| -1 | Dữ liệu bị thiếu hoặc nằm ngoài phạm vi dữ liệu có sẵn |

* **Pedestrian\_Location**: Vị trí của người dân bị xảy ra tai nạn. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 0 | Không phải là khu dân cư |
| 1 | Đi qua dải đi bộ |
| 2 | Đi qua đường zig zag |
| ... | ... |

* **Pedestrian\_Movement:** Chuyển động của người dân. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 0 | Không phải là người đi |
| 1 | Băng qua từ phía gần người lái |
| 2 | Băng qua đường gần - bị che bởi xe đang đỗ hoặc đang dừng |
| ... | ... |

* **Vehicle\_Reference:** mã phương tiện

## Bảng Accidents:

**Các thuộc tính sử dụng:**

* **Accident\_Index:** mã số của tai nạn
* **Latitude:** là vĩ độ - một giá trị xác định vị trí của một điểm trên bề mặt Trái Đất ở phía bắc hay phía nam của xích đạo.
* **Longitude:** là kinh độ - một giá trị tọa độ địa lý theo hướng đông-tây, được sử dụng phổ biến nhất trong bản đồ học và hoa tiêu toàn cầu
* **Accident\_Severity:** mức độ nghiêm trọng. Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 1 | Gây tử vong |
| 2 | Nghiêm trọng |
| 3 | Nhẹ |

* **Local\_Authority(District):** địa phương xảy ra tai nạn

|  |  |
| --- | --- |
| Code | Label |
| 1 | Westminster |
| 2 | Camden |
| 3 | Islington |
| ... | ... |

* **Date:** Ngày/tháng/năm xảy ra tai nạn
* **Number\_of\_Vehicles**: Số lượng phương tiện tham gia
* **Number\_of\_Casualties**: Số lượng thương vong
* **Speed\_limit:** tốc độ tối đa cho phép
* **Local\_Authority\_(District)** **(tính theo quận):** Các mã LA (chính quyền địa phương) tại các quận. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Westminster |
| 2 | Camden |
| 3 | Islington |
| …. | …. |

* **Local\_Authority\_(Highway):** Các mã LA (chính quyền địa phương) tại các đường cao tốc. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| S12000033 | Aberdeen City |
| S12000034 | Aberdeenshire |
| S12000041 | Angus |
| ... | ... |

* **Day\_of\_Week**: thứ (buổi) xảy ra tai nạn. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Chủ nhật |
| 2 | Thứ hai |
| 3 | Thứ bai |
| ... | ... |
| 7 | Thứ bảy |

* **Time:** Thời điểm xảy ra tai nạn (giờ:phút)
* **Weather\_Conditions**: Điều kiện thời tiết. Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 1 | Thời tiết ổn và không có gió mạnh, gió giật |
| 2 | Trời mưa nhưng không có gió mạnh, gió giật |
| 3 | Tuyết rơi nhưng không có gió mạnh, gió giật |
| ... | …. |

* **Light\_Conditions:** Điều kiện ánh sáng. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Ánh sáng ban ngày |
| 2 | Trời tối, đèn sáng |
| 3 | Trời tối, đèn không sáng |
| ... | ... |

* + **Road\_Surface**: Bề mặt đường. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Khô ráo |
| 2 | Ẩm ướt |
| 3 | Tuyết phủ |
| ... | ... |

* + **Road\_Type:** Loại đường. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Vòng xoay |
| 2 | Đường một chiều |
| 3 | Xa lộ hai chiều |
| ... | ... |

* + **Urban\_or\_Rural\_Area:** Loại khu vực. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Đô thị |
| 2 | Nông thôn |
| 3 | Chưa phân bổ |

* **LSOA\_of\_Accident\_Location**: Mã vị trí nơi xảy ra tai nạn

## Bảng Vehicles:

**Các thuộc tính sử dụng:**

* + **Accident\_Index:** Mã số tai nạn
  + **Vehicle\_Reference:** Mã phương tiện
  + **Driver\_Home\_Area\_Type:** Loại khu vực nhà ở của tài xế. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Khu đô thị |
| 2 | Thị trấn nhỏ |
| 3 | Nông thôn |
| -1 | Không có dữ liệu |

* + **Sex\_of\_Driver:** Giới tính của tài xế. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Nam |
| 2 | Nữ |
| 3 | Không rõ |
| -1 | Dữ liệu bị thiếu hoặc nằm ngoài phạm vi dữ liệu có sẵn |

* + **Age\_Band\_of\_Driver:** Nhóm độ tuổi của tài xế (0-5, 6-10, 11-15,...hơn 75 tuổi). Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | 0-5 |
| 2 | 6-10 |
| 3 | 11-15 |
| ... | ... |

* **Journey\_Purpose\_of\_Driver:** Mục đích hành trình. Các giá trị được quy

định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Hành trình như một phần công việc |
| 2 | Đi làm/Tan làm |
| 3 | Đưa/đón học sinh đi học |
| ... | ... |

* **Was\_Vehicle\_Left\_Hand\_Drive:** Tay thuận của tài xế có là tay trái không

|  |  |
| --- | --- |
| Code | Label |
| 1 | Không |
| 2 | Có |
| -1 | Không có dữ liệu |

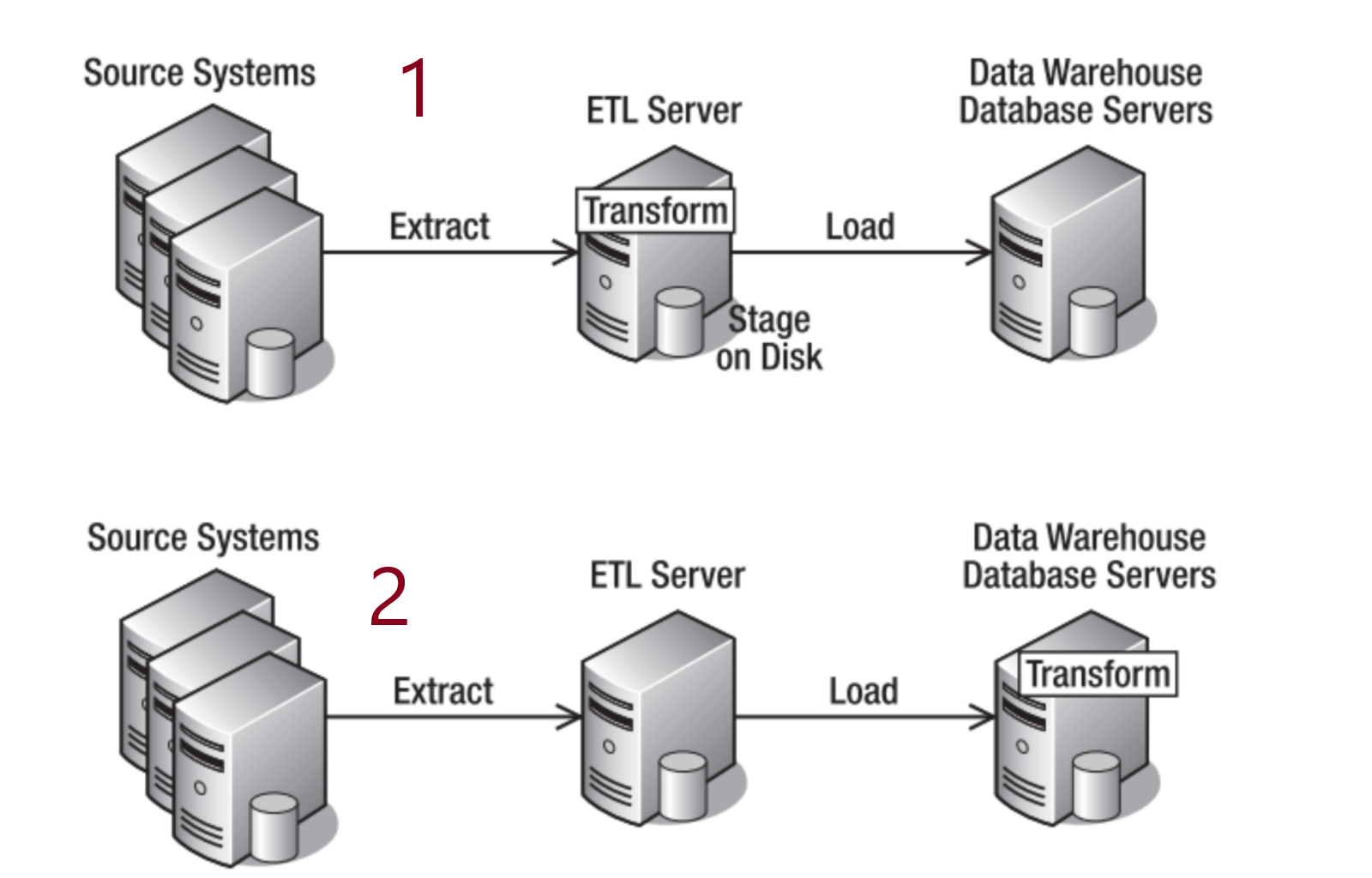
* **Vehicle\_Type:** Loại phương tiện. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Xe đạp |
| 2 | Xe mô tô dưới 50 phân khối |
| 3 | Xe mô tô dưới 125 phân khối |
| ... | ... |

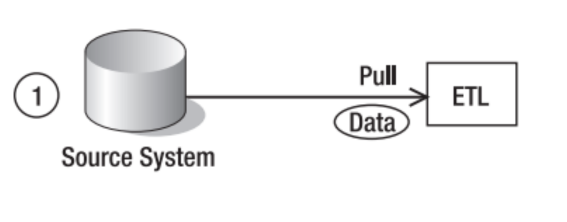
# ETL từ Source sang Stage

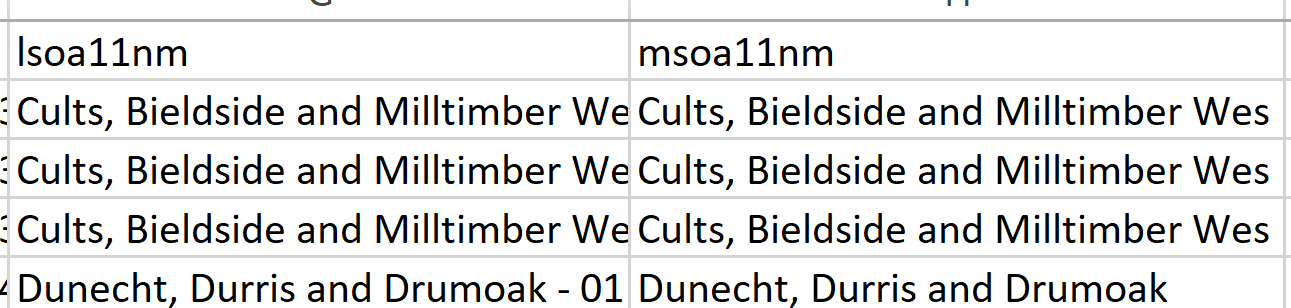
## Xử lý dữ liệu khi từ Source sang Stage

* Kiến trúc mà nhóm tiếp cận là kiến trúc số 1 vì dữ liệu sử dụng trên đồ án khá nhỏ gồm 5 file CSV có dung lượng nhỏ hơn … Vì vậy thực hiện transform (xử lý) ở memory nhanh hơn sau đó load vào Data Warehouse luôn. Độ lớn bộ nhớ của ETL server đủ để có thể hỗ trợ cho việc transform ở memory. Data WareHouse chỉ có nhiệm vụ nhận dữ liệu đã qua transform chứ không transform tại Data WareHouse như kiến trúc số 2 vì nhóm không có Data Warehouse đủ mạnh để vừa chứa dữ liệu vừa transform, có khả xử lý song song đồng thời



* Ở đây đối tượng chủ động để lấy dữ liệu từ source system là ETL Server viết những thủ tục, chương trình để lấy dữ liệu từ source system một cách đều đặn. Source System đóng vai trò bị động. Hình ảnh dưới đây minh họa thể hiện rõ đối tượng lấy dữ liệu từ source system:



* Theo đồ án, quá trình ETL dữ liệu từ Source sang State sử dụng 5 Source csv là:
* Accidents1114.csv
* Casualty1114.csv
* PCD\_OA\_LSOA\_MSOA\_LAD\_AUG21\_UK\_LU.csv
* postcodes.csv
* Vehicles1114.csv
* Với các nguồn như Accident, Casualty, postcodes, nhóm sử dụng trực tiếp lấy từ nguồn là flat file (file csv), đồng thời tạo thêm 2 trường đại diện cho nhãn đại diện cho thời gian (create\_date và update\_date) và 1 nhãn đại diện cho trạng thái “Dòng dữ liệu đã được xóa hay chưa” có tên là TrangThai (ở đây đang xét ngữ cảnh là dữ liệu bị xóa “soft deleted”) thủ công trên các file csv. Nhóm cũng đã chủ động bỏ những cột không cần thiết (không phục vụ quá trình phân tích và làm report) trong quá trình ETL.
* Với nguồn PCD\_OA\_LSOA\_MSOA\_LAD\_AUG21\_UK\_LU thì nhóm quyết định sử dụng cơ sở dữ liệu quan hệ để lưu dữ liệu nguồn và từ bây giờ, bất kỳ mọi thay đổi nào hoặc thêm mới dữ liệu sẽ được thực hiện dựa trên cơ sở dữ liệu quan hệ này thay vì file flat file như các bảng khác. Bên cạnh đó, việc tạo thêm 2 trường đại diện cho nhãn thời gian (createdate và updatedate) và TrangThai đại điện cho việc dòng dữ liệu đã bị xóa hay chưa. Nhóm cũng đã chủ động bỏ những cột không cần thiết (không phục vụ quá trình phân tích và làm report) trong quá trình ETL.
* Lý do việc nguồn PCD\_OA\_LSOA\_MSOA\_LAD\_AUG21\_UK\_LU thì nhóm quyết định sử dụng cơ sở dữ liệu quan hệ để lưu dữ liệu nguồn chứ không trực tiếp lấy từ nguồn là flat file (file csv) vì gặp một số khó khăn trong quá trình nhóm thử nghiệm trực tiếp lấy nguồn PCD\_OA\_LSOA\_MSOA\_LAD\_AUG21\_UK\_LU là flat file. Như hình minh họa ở dưới, khi xử lý dữ liệu, delimiter của SSIS là dấu phẩy thì lsoa11nm sẽ có giá trị là “Cults, Bieldside and Milltimber Wes” và msoa11nm là “Cults, Bieldside and Milltimber Wess”, nhưng delimiter của SSIS nhận diện sai trong phẩy trong csv vì dữ liệu cột lsoa11nm có dấu phẩy giống với dấu phẩy ngăn cách các cột dữ liệu trong csv dẫn đến việc mapping việc lỗi. Ví dụ như sau quá trình mapping thì lsoa11nm sẽ là “Cults” và msoa11nm là “Bieldside and Milltimber Wes” sai với kết quả mong muốn ban đầu. 

Nên nhóm quyết định với nguồn PCD\_OA\_LSOA\_MSOA\_LAD\_AUG21\_UK\_LU thì sử dụng cơ sở dữ liệu quan hệ để lưu dữ liệu nguồn .

* Đối với các nhãn đại diện cho thời gian là create\_date và update\_date, nhóm quyết định sử dụng kiểu dữ liệu là string thay cho timestamp hay datetime là vì khi thiết lập kết nối cho flat file cũng như thực hiện mapping column giữa dữ liệu nguồn và dữ liệu đích. Nhóm nhận thấy rằng luôn có lỗi xảy ra khi người dùng chọn kiểu dữ liệu là datetime hoặc timestamp khi đọc dữ liệu từ file csv lên, kể cả khi dùng script task để custom kiểu dữ liệu. SSIS không thể chuyển đổi được dữ liệu có được ở trường create\_date và update\_date thành kiểu dữ liệu ở dạng datetime. “Error 0xc02020a1: Data Flow Task 1: Data conversion failed. The data conversion for column "create\_date" returned status value 2 and status text "The value could not be converted because of a potential loss of data.". chính là lỗi mà nhóm gặp phải. Để giải quyết vấn đề này, ban đầu nhóm có thử sử dụng derived column và dùng expression để chuyển đổi kiểu dữ liệu cho các cột này, nhưng vẫn không thành công. Cuối cùng, nhóm quyết định dùng kiểu dữ liệu là chuỗi thay cho

kiểu dữ liệu datetime. Điều này áp dụng cho cả các bảng ở cơ sở dữ liệu Stage và Temp. Các data flow vẫn chạy đúng và ổn định.

* Nhóm quyết định tạo thêm 1 cơ sở dữ liệu tạm gọi là Temp bên cạnh cơ sở dữ liệu nguồn là Stage và Metadata. Lý do là vì nếu load trực tiếp dữ liệu từ flat file csv thì không thể dùng cách tiếp cận “Incremental Extract” như làm việc với nguồn là cơ sở dữ liệu do ta không thể sử dụng SQL Command trên file CSV. Vì lẽ đó mà nhóm cần đổ dữ liệu vào cơ sở dữ liệu tạm trước, sau đó từ cơ sở dữ liệu tạm này sẽ dùng SQL Command để thực hiện “Incremental Extract” vào cơ sở dữ liệu Stage. Những dữ liệu được lấy sẽ là những dữ liệu có thời gian cập nhật hoặc thời gian được tạo lớn hơn hoặc bằng thời điểm LSET và nhỏ hơn thời điểm CET, đồng thời dữ liệu cũng phải có trạng thái “False” (tức chưa bị xóa soft deleted).
* Tương ứng với 5 file csv ta sẽ có 5 bảng là:
* Postcodes
* Accidents2011\_2014
* Casualty2011\_2014
* Vehicles2011\_2014
* UK\_Area\_Information
* Chiến lược rút trích dữ liệu từ cơ sở dữ liệu: Incremental extract
  + Cơ bản về đặc điểm của chiến lược Incremental extract là chỉ rút trích những thay đổi so với lần rút trích gần đây nhất (Dòng mới thêm, dòng mới xoá, dòng được cập nhật) dựa vào thuộc tính nhãn thời gian và những dòng có trạng thái dữ liệu cho biết dữ liệu chưa bị xóa (soft deleted) hay chưa.
  + Trong đồ án này sử dụng hai nhãn thời gian là create\_date (ngày tạo) và update\_date (ngày cập nhật) và nhãn đại diện cho Trạng thái của dữ liệu. Giá trị ban đầu của create\_date là thời điểm tạo ra/thêm mới dòng dữ liệu, giá trị ban đầu của update\_date là NULL, giá trị ban đầu của trạng thái là false (tức là dữ liệu chưa có bị xóa - soft deleted). Mỗi khi có một dòng thay đổi thì cập nhật lại nhãn thời gian và trạng thái dữ liệu.
* Khi đổ dữ liệu vào nguồn tất cả các bảng có tên giống với tên đặt của csv ngoại trừ csv có tên là PCD\_OA\_LSOA\_MSOA\_LAD\_AUG21\_UK\_LU khi đổ vào nguồn đổi tên bảng thành UK\_Area\_Information
* Tạo một cơ sở dữ liệu metadata gồm bảng có các cột Name\_DataFlow (Tên bảng), LSET, CET. Trong đó:
* LSET: Thời điểm rút trích hiện tại
* CET: Thời điểm rút trích thành công gần đây nhất
* Tạo một cơ sở dữ liệu Stage có cấu trúc và các bảng giống với cơ sở dữ liệu Temp và cơ sở dữ liệu nguồn
* Tạo một cơ sở dữ liệu Temp có cấu trúc và các bảng giống với cơ sở dữ liệu nguồn

## Các bước thực hiện khi đổ từ bảng nguồn vào stage

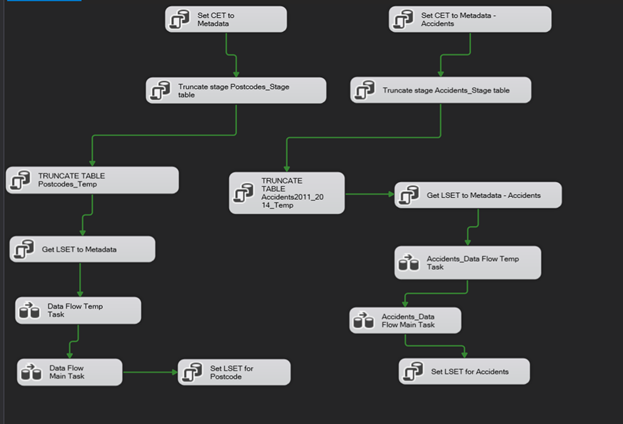
* Đối với PCD\_OA\_LSOA\_MSOA\_LAD\_AUG21\_UK\_LU:

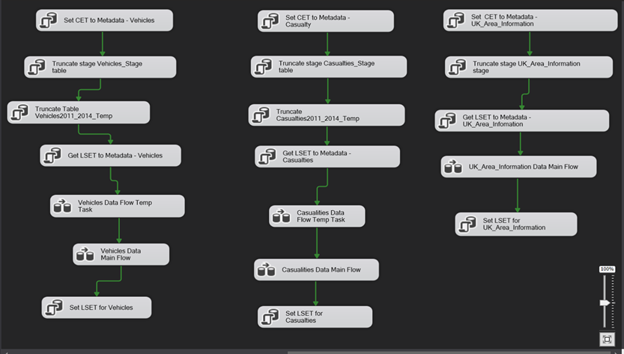
1. Lấy thông tin LSET được lưu trong metadata
2. Lấy CET bằng thời gian khởi động package
3. Truncate bảng Stage hiện tại: TRUNCATE TABLE <BangStage>
4. Rút trích dữ liệu từ bảng Source sang bảng Stage: select \* from BangSource where (createdate >= LSET and createdate < CET) or (updatedate >= LSET and updatedate < CET) AND (TrangThai = ‘FALSE’)
5. Cập nhật LSET = CET

* Đối với các nguồn còn lại:

1. Lấy thông tin LSET được lưu trong metadata
2. Lấy CET bằng thời gian khởi động package
3. Truncate bảng Stage hiện tại: TRUNCATE TABLE <BangStage>
4. Truncate bảng Temp hiện tại: TRUNCATE TABLE <BangTemp>
5. Rút trích dữ liệu từ flat file csv cụ thể vào trong bảng Temp
6. Rút trích dữ liệu từ bảng Temp sang bảng Stage: SELECT \* FROM <Tên bảng TEMP> WHERE (create\_date >= CAST(? AS VARCHAR)) AND (create\_date < CAST(? AS VARCHAR)) OR (update\_date >= CAST(? AS VARCHAR)) AND (update\_date < CAST(? AS VARCHAR)) AND (trangthai = 'FALSE').
7. Cập nhật LSET = CET.

## Flow khi thực hiện trên Visual Studio Code

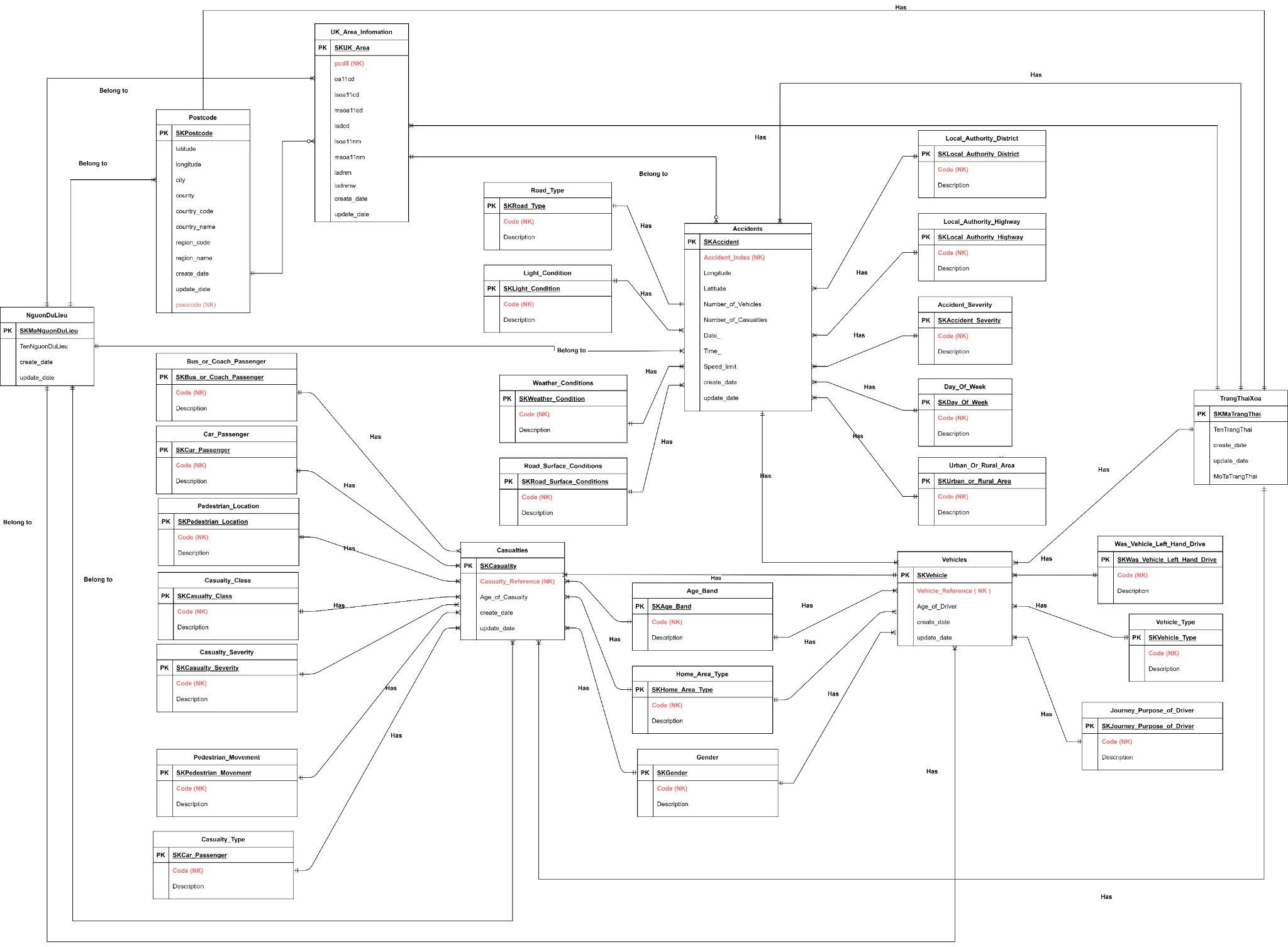




# Stage sang NDS

## Thiết kế NDS dưới dạng mô hình ER

Hình vẽ dưới thể hiện rõ hơn cơ sở dữ liệu sau khi chuẩn hóa thành NDS dưới dạng mô hình ER, ta sẽ dễ dàng thấy rõ sự thay đổi dữ liệu sau khi chuẩn hóa



**Mô tả dữ liệu:**

1. **Accidents** (Accident\_Index**,** Longitude, Latitude, Number\_of\_Vehicles, Number\_of\_Casualties, Date\_, Time\_, Speed\_limit, create\_date, update\_date, Accident\_Severity, Day\_of\_Week, Local\_Authority\_District, Local\_Authority\_Highway, Road\_Type, Light\_Conditions, Weather\_Conditions, Road\_Surface\_Conditions, Urban\_or\_Rural\_Area, SKMaTrangThai, SKMaNguonDuLieu, pcd8)

PHỤ THUỘC HÀM F = {Accident\_Index **->** Longitude, Latitude, Number\_of\_Vehicles, Number\_of\_Casualties, Date\_, Time\_, Speed\_limit, create\_date, update\_date, Accident\_Severity, Day\_of\_Week, Local\_Authority\_District, Local\_Authority\_Highway, Road\_Type, Light\_Conditions, Weather\_Conditions, Road\_Surface\_Conditions, Urban\_or\_Rural\_Area, SKMaTrangThai, SKMaNguonDuLieu, pcd8}

DẠNG CHUẨN: BCK

1. **Postcode** (postcode ,latitude, longitude, city, county, country\_code, country\_name ,region\_code, region\_name, create\_date, update\_date, SKMaTrangThai, SKMaNguonDuLieu)

PHỤ THUỘC HÀM F = {postcode -> latitude, longitude, city, county, country\_code, country\_name ,region\_code, region\_name, create\_date, update\_date, SKMaTrangThai, SKMaNguonDuLieu}

DẠNG CHUẨN: BCK

1. **UK\_Area\_Information** (pcd8 **,** oa11cd, lsoa11cd, msoa11cd, ladcd, lsoa11nm, msoa11nm, ladnm,ladnmw, create\_date, update\_date, SKMaTrangThai, SKMaNguonDuLieu, postcode)

PHỤ THUỘC HÀM F = {pcd8 **->** oa11cd, lsoa11cd, msoa11cd, ladcd, lsoa11nm, msoa11nm, ladnm,ladnmw, create\_date, update\_date, SKMaTrangThai, SKMaNguonDuLieu, postcode}

DẠNG CHUẨN: BCK

1. **Vehicles** (Vehicle\_Reference ,Accident\_Index **,** Age\_of\_Driver, create\_date, update\_date,Vehicle\_Type, Was\_Vehicle\_Left\_Hand\_Drive, Journey\_Purpose\_of\_Driver, Sex\_of\_Driver, Age\_Band\_of\_Driver, Driver\_Home\_Area\_Type, SKMaTrangThai, SKMaNguonDuLieu)

PHỤ THUỘC HÀM F = {Vehicle\_Reference ,Accident\_Index **->** Age\_of\_Driver, create\_date, update\_date,Vehicle\_Type, Was\_Vehicle\_Left\_Hand\_Drive, Journey\_Purpose\_of\_Driver, Sex\_of\_Driver, Age\_Band\_of\_Driver, Driver\_Home\_Area\_Type, SKMaTrangThai, SKMaNguonDuLieu }

DẠNG CHUẨN: BCK

1. **Casualties** ( Vehicle\_Reference, Accident\_Index, Casualty\_Reference, Age\_of\_Casualty, create\_date, update\_date, Casualty\_Class, Sex\_of\_Casualty, Age\_Band\_of\_Casualty, Casualty\_Severity, Pedestrian\_Location. Pedestrian\_Movement, Car\_Passenger, Bus\_or\_Coach\_Passenger, Casualty\_Type, Casualty\_Home\_Area\_Type, SKMaTrangThai, SKMaNguonDuLieu)

PHỤ THUỘC HÀM F = {Vehicle\_Reference, Accident\_Index, Casualty\_Reference -> Age\_of\_Casualty, create\_date, update\_date, Casualty\_Class, Sex\_of\_Casualty, Age\_Band\_of\_Casualty, Casualty\_Severity, Pedestrian\_Location. Pedestrian\_Movement, Car\_Passenger, Bus\_or\_Coach\_Passenger, Casualty\_Type, Casualty\_Home\_Area\_Type, SKMaTrangThai, SKMaNguonDuLieu }

DẠNG CHUẨN: BCK

1. **NguonDuLieu**(SKMaNguonDuLieu,TenNguonDuLieu, create\_date, update\_date)

PHỤ THUỘC HÀM F = {SKMaNguonDuLieu ->TenNguonDuLieu, create\_date, update\_date)}

DẠNG CHUẨN: BCK

1. **TrangThaiXoa**(SKMaTrangThai,TenTrangThai, create\_date, update\_date, MoTaTrangThai)

PHỤ THUỘC HÀM F = {SKMaTrangThai ->TenTrangThai, create\_date, update\_date, MoTaTrangThai }

DẠNG CHUẨN: BCK

1. **Bus\_or\_Coach\_Passenger**(Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Car\_Passenger**(Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Pedestrian\_Location** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Casualty\_Class** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Casualty\_Severity** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Pedestrian\_Movement** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Casualty\_Type** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Road\_Type** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Light\_Condition** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Weather\_Conditions** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Road\_Surface\_Conditions** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Age\_Band** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Home\_Area\_Type** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Gender** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Local\_Authority\_District** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Local\_Authority\_Highway** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Accident\_Severity** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Day\_Of\_Week** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Urban\_Or\_Rural\_Area** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Was\_Vehicle\_Left\_Hand\_Drive** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Vehicle\_Type** (Code, Description)

PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

1. **Journey\_Purpose\_of\_Driver** (Code, Description)

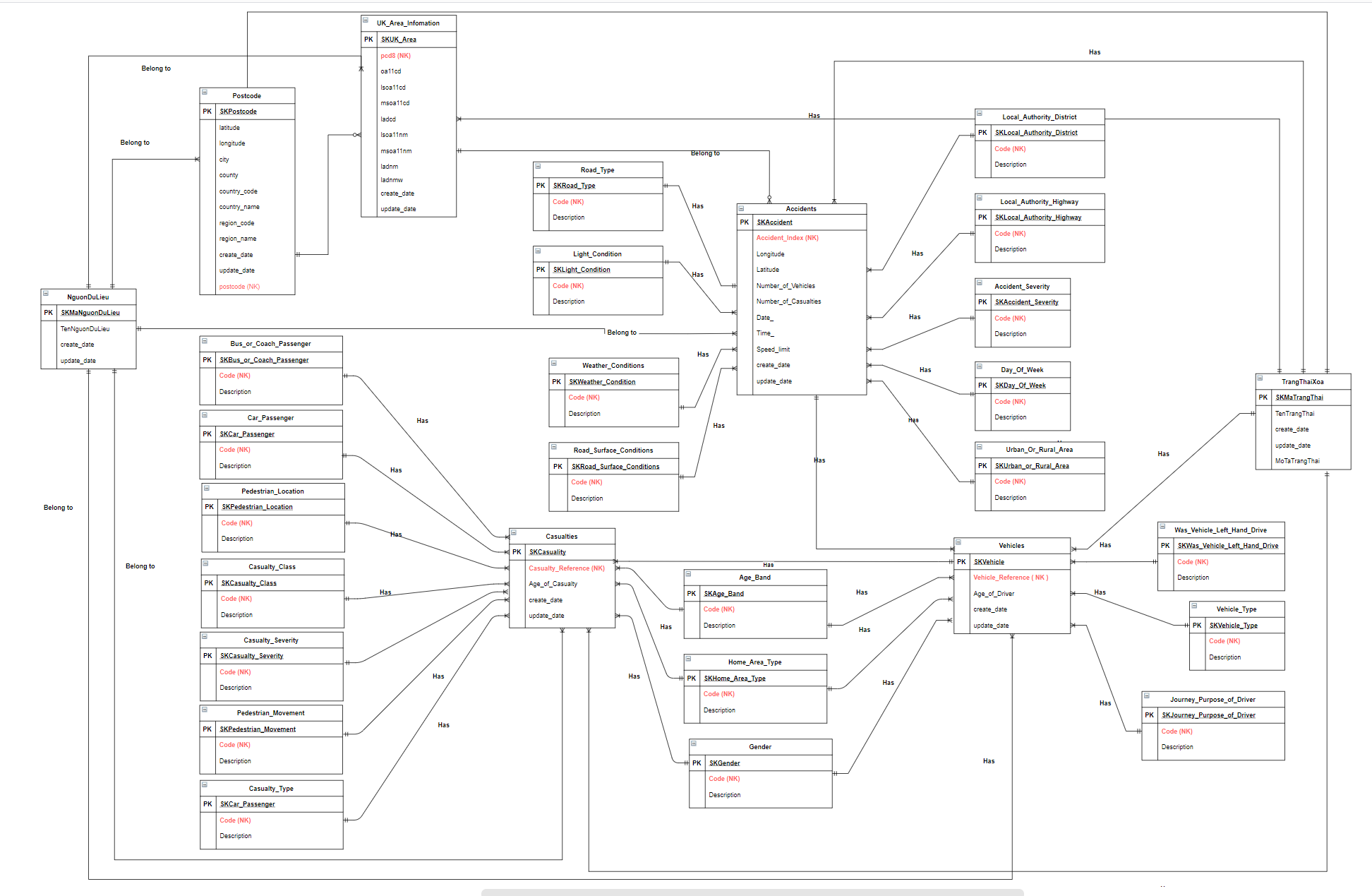
PHỤ THUỘC HÀM F = {Code -> Description}

DẠNG CHUẨN: BCK

* **Kết luận**: Như vậy, lược đồ tổng thể đạt dạng chuẩn BCK, đáp ứng cấu trúc trong NDS

## Lược đồ quan hệ dưới dạng physical:

Khi cài đặt ở mức vật lý, để tránh tình trạng sai sót và thất thoát dữ liệu trong tương lai khi các mã khoá như Accident\_Index, postcode, pcd8,…. bị thay đổi sẽ ảnh hưởng đến tham chiếu khóa ngoại. Nhóm chúng em quyết định tạo thêm một trường khóa tự tăng có tiền tố là SK ở mỗi bảng với kiểu dữ liệu là int. Khi đó, các bảng khác nếu có tham chiếu sẽ tham chiếu tới các giá trị có kiểu dữ liệu là int này, và ta cũng không cần phải lo lắng nếu trong tương lai các mã thực thể kia có thay đổi.



**Mô tả cấu trúc bảng trong NDS**

* Cấu trúc của bảng Postcode\_NDS trong NDS

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKPostcodes | Khoá tự tăng của bảng Postcode\_NDS |  |  |
| postcode | Mã postcode trong source system | Postcodes\_Stage.postcode |  |
| latitude | Vĩ độ của postcode | Postcodes\_Stage.latitude |  |
| longitude | Kinh độ của postcode | Postcodes\_Stage.longitude |  |
| city | Thành phố | Postcodes\_Stage.city |  |
| county | hat, khu vực nằm trong một bang và lớn hơn thành phố | Postcodes\_Stage.county |  |
| country\_code | Mã quốc gia | Postcodes\_Stage.country\_code |  |
| country\_name | Tên quốc gia tương ứng với country code | Postcodes\_Stage.country\_name |  |
| region\_code | Mã khu vực | Postcodes\_Stage.region\_code |  |
| region\_name | Tên khu vực tương ứng với region\_code | Postcodes\_Stage.region\_name |  |
| create\_date | Thời gian mà record được tạo trong Dimension Data Store |  |  |
| update\_date | Thời gian mà record được cập nhật trong Dimension Data Store |  |  |
| SKMaTrangThai | Trạng thái của dòng record | Tham chiếu khóa ngoại từ bảng TrangThaiXoa |  |
| SKMaNguonDuLieu | Nguồn chứa record | Tham chiếu khóa ngoại từ bảng NguonDuLieu |  |

* Cấu trúc của bảng UK\_Area\_Information\_NDS trong NDS

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKUK\_Area | Khóa tự tăng của bảng UK\_Area\_Information\_NDS |  |  |
| pcd8 | Mã khóa pcd8 trong source system | UK\_Area\_Information\_Stage.pcd8 | Lấy các ký tự từ đầu đến khoảng trắng gần nhất trong pcd8 |
| oa11cd | Output Area (2011 Census Output Area) | UK\_Area\_Information\_Stage.oa11cd |  |
| lsoa11cd | Lower Super Output Area | UK\_Area\_Information\_Stage.lsoa11cd |  |
| msoa11cd | Middle Lower Layer Super Output Area | UK\_Area\_Information\_Stage.msoa11cd |  |
| ladcd | Quận hiện tại mà mã bưu chính đã được chỉ định | UK\_Area\_Information\_Stage.ladcd |  |
| lsoa11nm | Tên khu vực tương ứng với mã **lsoa11cd** | UK\_Area\_Information\_Stage.lsoa11nm |  |
| msoa11nm | Tên khu vực tương ứng với mã **msoa11cd** | UK\_Area\_Information\_Stage.msoa11nm |  |
| ladnm | Tên chính quyền địa phương tương ứng với mã ladcd | UK\_Area\_Information\_Stage.ladnm |  |
| ladnmw | Tên khu vực tương ứng với mã oac11cd | UK\_Area\_Information\_Stage.ladnmw |  |
| create\_date | Thời gian mà record được tạo trong Dimension Data Store |  |  |
| update\_date | Thời gian mà record được cập nhật trong Dimension Data Store |  |  |
| SKMaTrangThai | Trạng thái của dòng record | Tham chiếu khóa ngoại từ bảng TrangThaiXoa |  |
| SKMaNguonDuLieu | Nguồn chứa record | Tham chiếu khóa ngoại từ bảng NguonDuLieu |  |
| postcode | UK\_Area\_Information\_NDS thuộc postcode nào | Tham chiếu khóa ngoại từ bảng Postcode\_NDS |  |

* Cấu trúc của bảng Casualties2011\_2014\_NDS trong NDS

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKCasuality | Khóa tự tăng của bảng Casualties2011\_2014\_NDS |  |  |
| Casualty\_Reference | Mã khóa của bảng Casualties2011\_2014\_NDS trong source system | Casualties2011\_2014\_Stage.Casualty\_Reference |  |
| SKVehicle | Mã khóa đồng thời là khóa ngoại của bảng Casualties2011\_2014\_NDS | Tham chiếu khóa ngoại từ bảng Vehicles2011\_2014\_NDS |  |
| Casualty\_Class | Lớp nạn nhân | Tham chiếu khóa ngoại từ bảng Casualty\_Class |  |
| Sex\_of\_Casualty | Giới tính nạn nhân | Tham chiếu khóa ngoại từ bảng Gender |  |
| Age\_of\_Casualty | Độ tuổi nạn nhân | Casualties2011\_2014\_Stage.Age\_of\_Casualty |  |
| Age\_Band\_of\_Casualty | Thuộc nhóm độ tuổi nào | Tham chiếu khoá ngoại từ bảng Age\_Band |  |
| Casualty\_Severity | Mức độ nghiêm trọng | Tham chiếu khóa ngoại từ bảng Casualty\_Severity |  |
| Pedestrian\_Location | Vị trí của người dân xảy ra tai nạn | Tham chiếu khóa ngoại từ bảng Pedestrian\_Location |  |
| Pedestrian\_Movement | Chuyển động của người dân | Tham chiếu khóa ngoại từ bảng Pedestrian\_Movement |  |
| Car\_Passenger | Vị trí của hành khách bị tại nạn nếu có đi ô tô | Tham chiếu khóa ngoại từ bảng Car\_Passenger |  |
| Bus\_or\_Coach\_Passenger | Vị trí của hành khách bị tai nạn nếu có đi xe buýt | Tham chiếu khóa ngoại từ bảng Bus\_or\_Coach\_Passenger |  |
| Casualty\_Type | Loại nạn nhân | Tham chiếu khóa ngoại từ bảng Casualty\_Type |  |
| Casualty\_Home\_Area\_Type | Loại khu vực nạn nhân sinh sống | Tham chiếu khóa ngoại từ bảng Home\_Area\_Type |  |
| create\_date | Thời gian mà record được tạo trong Dimension Data Store |  |  |
| update\_date | Thời gian mà record được cập nhật trong Dimension Data Store |  |  |
| SKMaTrangThai | Trạng thái của dòng record | Tham chiếu khóa ngoại từ bảng TrangThaiXoa |  |
| SKMaNguonDuLieu | Nguồn chứa record | Tham chiếu khóa ngoại từ bảng NguonDuLieu |  |

* Cấu trúc của bảng Vehicles2011\_2014\_NDS trong NDS

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKVehicle | Khoá tự tăng của bảng Vehicles2011\_2014\_NDS |  |  |
| Vehicle\_Reference | Mã khoá của Vehicle trong bảng Source System | Vehicles2011\_2014\_Stage.Vehicle\_Reference |  |
| SKAccident | Khoá chính đồng thời là khoá ngoại của bảng Vehicles2011\_2014\_NDS |  |  |
| Vehicle\_Type | Loại phương tiện | Tham chiếu khoá ngoại từ bảng Vehicle\_Type |  |
| Was\_Vehicle\_Left\_Hand\_Drive | Tay thuận của tài xế có là tay trái không | Tham chiếu khoá ngoại từ bảng Was\_Vehicle\_Left\_Hand\_Drive |  |
| Journey\_Purpose\_of\_Driver | Mục đích chuyến đi | Tham chiếu khoá ngoại từ bảng Journey\_Purpose\_of\_Driver |  |
| Sex\_of\_Driver | Giới tính của tài xế | Tham chiếu khoá ngoại từ bảng Gender |  |
| Age\_of\_Driver | Độ tuổi của tài xế | Vehicles2011\_2014\_Stage.Age\_of\_Driver |  |
| Age\_Band\_of\_Driver | Tài xế thuộc nhóm tuổi | Tham chiếu khoá ngoại từ bảng Age\_Band |  |
| Driver\_Home\_Area\_Type | Loại khu vực nhà ở của tài xế | Tham chiếu khoá ngoại từ bảng Home\_Area\_Type |  |
| create\_date | Thời gian mà record được tạo trong Dimension Data Store |  |  |
| update\_date | Thời gian mà record được cập nhật trong Dimension Data Store |  |  |
| SKMaTrangThai | Trạng thái của dòng record | Tham chiếu khóa ngoại từ bảng TrangThaiXoa |  |
| SKMaNguonDuLieu | Nguồn chứa record | Tham chiếu khóa ngoại từ bảng NguonDuLieu |  |

* Cấu trúc của bảng Accidents2011\_2014\_NDS trong NDS

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKAccident | Khóa tự tăng của bảng Accidents2011\_2014\_NDS |  |  |
| Accident\_Index | Mã khóa trong source system | Accidents2011\_2014\_Stage.Accident\_Index |  |
| SK\_UKArea | Thuộc UK\_Area\_Information | Tham chiếu khóa ngoại từ bảng UK\_Area\_Information\_NDS |  |
| Longitude | Kinh độ xảy ra tai nạn | Accidents2011\_2014\_Stage.Longitude |  |
| Latitude | Vĩ độ xảy ra tai nạn | Accidents2011\_2014\_Stage.Latitude |  |
| Accident\_Severity | Mức độ nghiêm trọng | Tham chiếu khóa ngoại từ bảng Accident\_Severity |  |
| Number\_of\_Vehicles | Số xe trong tai nạn | Accidents2011\_2014\_Stage.Number\_of\_Vehicles |  |
| Number\_of\_Casualties | Số người bị thương | Accidents2011\_2014\_Stage. |  |
| Date\_ | Ngày xảy ra | Accidents2011\_2014\_Stage.Date\_ |  |
| Day\_of\_Week | Thứ, buổi xảy ra tai nạn | Tham chiếu khóa ngoại từ bảng |  |
| Time\_ | Thời gian xảy ra tai nạn | Accidents2011\_2014\_Stage.Time\_ |  |
| Local\_Authority\_District | Mã LA (chính quyền địa phương) tại các quận | Tham chiếu khóa ngoại từ bảng |  |
| Local\_Authority\_Highway | Mã LA (chính quyền địa phương) tại các đường cao tốc | Tham chiếu khóa ngoại từ bảng |  |
| Road\_Type | Loại đường | Tham chiếu khóa ngoại từ bảng |  |
| Speed\_limit | Tốc độ tối đa cho phép | Accidents2011\_2014\_Stage.Speed\_limit |  |
| Light\_Conditions | Điều kiện ánh sáng | Tham chiếu khóa ngoại từ bảng Light\_Conditions |  |
| Weather\_Conditions | Điều kiện thời tiết | Tham chiếu khóa ngoại từ bảng Weather\_Conditions |  |
| Road\_Surface\_Conditions | Điều kiện bề mặt đường | Tham chiếu khóa ngoại từ bảng Road\_Surface\_Conditions |  |
| Urban\_or\_Rural\_Area | Bề mặt đường | Tham chiếu khóa ngoại từ bảng Urban\_or\_Rural\_Area |  |
| create\_date | Thời gian mà record được tạo trong Dimension Data Store |  |  |
| update\_date | Thời gian mà record được cập nhật trong Dimension Data Store |  |  |
| SKMaTrangThai | Trạng thái của dòng record | Tham chiếu khóa ngoại từ bảng TrangThaiXoa |  |
| SKMaNguonDuLieu | Nguồn chứa record | Tham chiếu khóa ngoại từ bảng NguonDuLieu |  |

* **Cấu trúc bảng NguonDuLieu trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKMaNguonDuLieu | Khóa tự tăng của bảng NguonDuLieu |  |  |
| TenNguonDuLieu | Tên của nguồn dữ liệu | QLBI\_NDS.TrangThaiXoa.TenNguonDuLieu |  |
| create\_date | Thời gian mà record được tạo trong Dimension Data Store |  |  |
| update\_date | Thời gian mà record được cập nhật trong Dimension Data Store |  |  |

* **Cấu trúc bảng TrangThaiXoa trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKMaTrangThai | Khóa tự tăng của bảng TrangThaiXoa |  |  |
| TenTrangThai | Tên của trạng thái | QLBI\_NDS.TrangThaiXoa.TenTrangThai |  |
| create\_date | Thời gian mà record được tạo trong Dimension Data Store |  |  |
| update\_date | Thời gian mà record được cập nhật trong Dimension Data Store |  |  |

* **Cấu trúc bảng Bus\_or\_Coach\_Passenger trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKBus\_or\_Coach\_Passenger | Khóa tự tăng của bảng Bus\_or\_Coach\_Passenger |  |  |
| Code | Mã khóa của Bus\_Coach\_Passenger trong source system | QLBI\_NDS.Bus\_or\_Coach\_Passenger.Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Bus\_or\_Coach\_Passenger.Description |  |

* **Cấu trúc bảng Car\_Passenger trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKCar\_Passenger | Khóa tự tăng của bảng Car\_Passenger |  |  |
| Code | Mã khóa của Car\_Passenger trong source system | QLBI\_NDS.Car\_Passenger.Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Car\_Passenger.Description |  |

* **Cấu trúc bảng Pedestrian\_Location trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKPedestrian\_Location | Khóa tự tăng của bảng Pedestrian\_Location |  |  |
| Code | Mã khóa của Pedestrian\_Location trong source system | QLBI\_NDS.Pedestrian\_Location .Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Pedestrian\_Location .Description |  |

* **Cấu trúc bảng Casualty\_Class trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKCasualty\_Class | Khóa tự tăng của bảng Casualty\_Class |  |  |
| Code | Mã khóa của Casualty\_Class  trong source system | QLBI\_NDS.Casualty\_Class .Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Casualty\_Class .Description |  |

* **Cấu trúc bảng Casualty\_Severity trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKCasualty\_Severity | Khóa tự tăng của bảng Casualty\_Severity |  |  |
| Code | Mã khóa của Casualty\_Severity trong source system | QLBI\_NDS.Casualty\_Severity .Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Casualty\_Severity .Description |  |

* **Cấu trúc bảng Pedestrian\_Movement trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKPedestrian\_Movement | Khóa tự tăng của bảng Pedestrian\_Movement |  |  |
| Code | Mã khóa của Pedestrian\_Movement trong source system | QLBI\_NDS.Pedestrian\_Movement.Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Pedestrian\_Movement .Description |  |

* **Cấu trúc bảng Casualty\_Type trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKCasualty\_Type | Khóa tự tăng của bảng Casualty\_Type |  |  |
| Code | Mã khóa của Casualty\_Type trong source system | QLBI\_NDS.Casualty\_Type.Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Casualty\_Type.Description |  |

* **Cấu trúc bảng Road\_Type trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKRoad\_Type | Khóa tự tăng của bảng Road\_Type |  |  |
| Code | Mã khóa của Road\_Type trong source system | QLBI\_NDS.Road\_Type.Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Road\_Type.Description |  |

* **Cấu trúc bảng Light\_Condition trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKLight\_Condition | Khóa tự tăng của bảng Light\_Condition |  |  |
| Code | Mã khóa của Light\_Condition trong source system | QLBI\_NDS.Light\_Condition.Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Light\_Condition.Description |  |

* **Cấu trúc bảng Weather\_Conditions trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKWeather\_Conditions | Khóa tự tăng của bảng Weather\_Conditions |  |  |
| Code | Mã khóa của Weather\_Conditions trong source system | QLBI\_NDS.Weather\_Conditions .Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Weather\_Conditions.Description |  |

* **Cấu trúc bảng Road\_Surface\_Conditions trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKRoad\_Surface\_Conditions | Khóa tự tăng của bảng Road\_Surface\_Conditions |  |  |
| Code | Mã khóa của Road\_Surface\_Conditions trong source system | QLBI\_NDS.Road\_Surface\_Conditions.Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Road\_Surface\_Conditions.Description |  |

* **Cấu trúc bảng Age\_Band trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKAge\_Band | Khóa tự tăng của bảng Age\_Band |  |  |
| Code | Mã khóa của Age\_Band trong source system | QLBI\_NDS.Age\_Band .Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Age\_Band.Description |  |

* **Cấu trúc bảng Home\_Area\_Type trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKHome\_Area\_Type | Khóa tự tăng của bảng Home\_Area\_Type |  |  |
| Code | Mã khóa của Home\_Area\_Type trong source system | QLBI\_NDS.Home\_Area\_Type.Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Home\_Area\_Type .Description |  |

* **Cấu trúc bảng Gender trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKGender | Khóa tự tăng của bảng Gender |  |  |
| Code | Mã khóa của Gender trong source system | QLBI\_NDS.Gender .Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Gender.Description |  |

* **Cấu trúc bảng Local\_Authority\_District trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKLocal\_Authority\_District | Khóa tự tăng của bảng Local\_Authority\_District |  |  |
| Code | Mã khóa của Local\_Authority\_District trong source system | QLBI\_NDS.Local\_Authority\_District.Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Local\_Authority\_District .Description |  |

* **Cấu trúc bảng Local\_Authority\_Highway trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKLocal\_Authority\_Highway | Khóa tự tăng của bảng Local\_Authority\_Highway |  |  |
| Code | Mã khóa của Local\_Authority\_Highway trong source system | QLBI\_NDS.Local\_Authority\_Highway .Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Local\_Authority\_Highway.Description |  |

* **Cấu trúc bảng Accident\_Severity trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKAccident\_Severity | Khóa tự tăng của bảng Accident\_Severity |  |  |
| Code | Mã khóa của Accident\_Severity trong source system | QLBI\_NDS.Accident\_Severity.Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Accident\_Severity.Description |  |

* **Cấu trúc bảng Day\_Of\_Week trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKDay\_Of\_Week | Khóa tự tăng của bảng Day\_Of\_Week |  |  |
| Code | Mã khóa của Day\_Of\_Week trong source system | QLBI\_NDS.Day\_Of\_Week .Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDSDay\_Of\_Week..Description |  |

* **Cấu trúc bảng Urban\_Or\_Rural\_Area trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKUrban\_Or\_Rural\_Area | Khóa tự tăng của bảng Urban\_Or\_Rural\_Area |  |  |
| Code | Mã khóa của Urban\_Or\_Rural\_Area trong source system | QLBI\_NDS.Urban\_Or\_Rural\_Area.Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Urban\_Or\_Rural\_Area.Description |  |

* **Cấu trúc bảng Was\_Vehicle\_Left\_Hand\_Drive trong NDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKWas\_Vehicle\_Left\_Hand\_Drive | Khóa tự tăng của bảng Was\_Vehicle\_Left\_Hand\_Drive |  |  |
| Code | Mã khóa của Was\_Vehicle\_Left\_Hand\_Drive trong source system | QLBI\_NDS.Was\_Vehicle\_Left\_Hand\_Drive.Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Was\_Vehicle\_Left\_Hand\_Drive.Description |  |

* **Cấu trúc bảng Vehicle\_Type trong NDS**

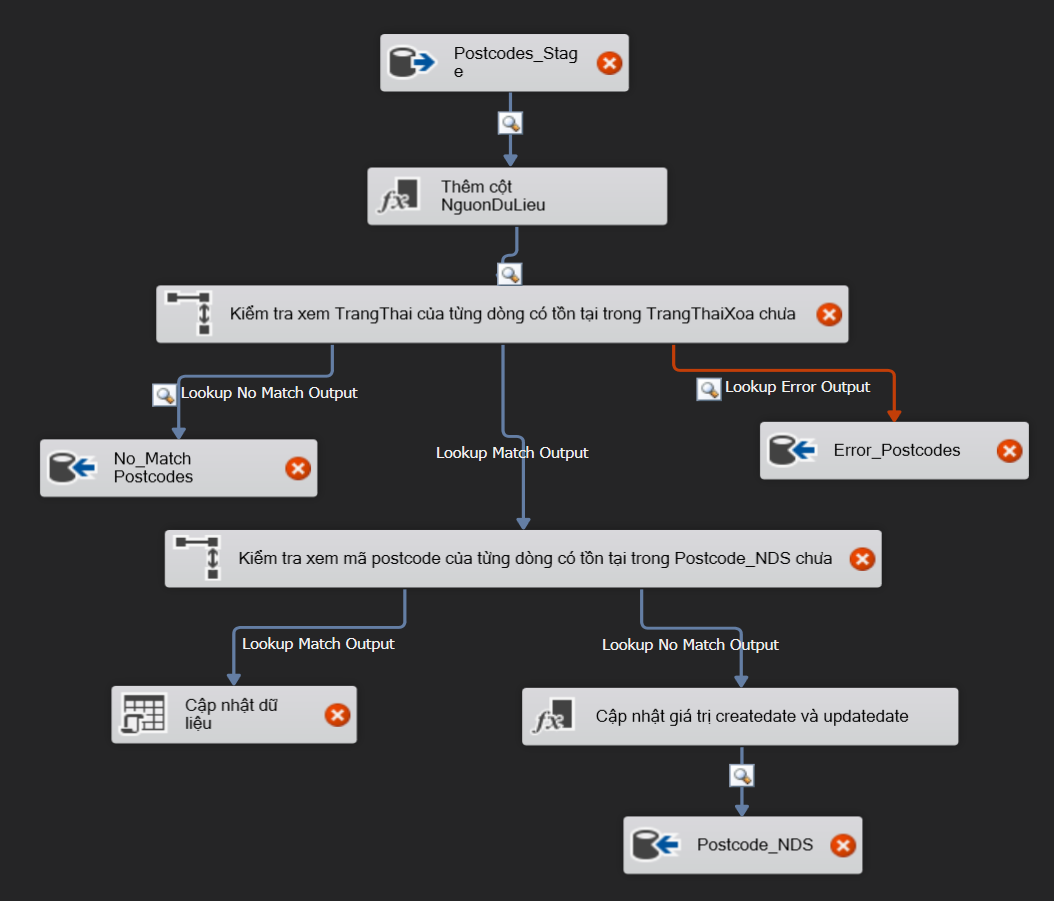
|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKVehicle\_Type | Khóa tự tăng của bảng Vehicle\_Type |  |  |
| Code | Mã khóa của Vehicle\_Type trong source system | QLBI\_NDS.Vehicle\_Type.Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Vehicle\_Type.Description |  |

* **Cấu trúc bảng Journey\_Purpose\_of\_Driver trong NDS**

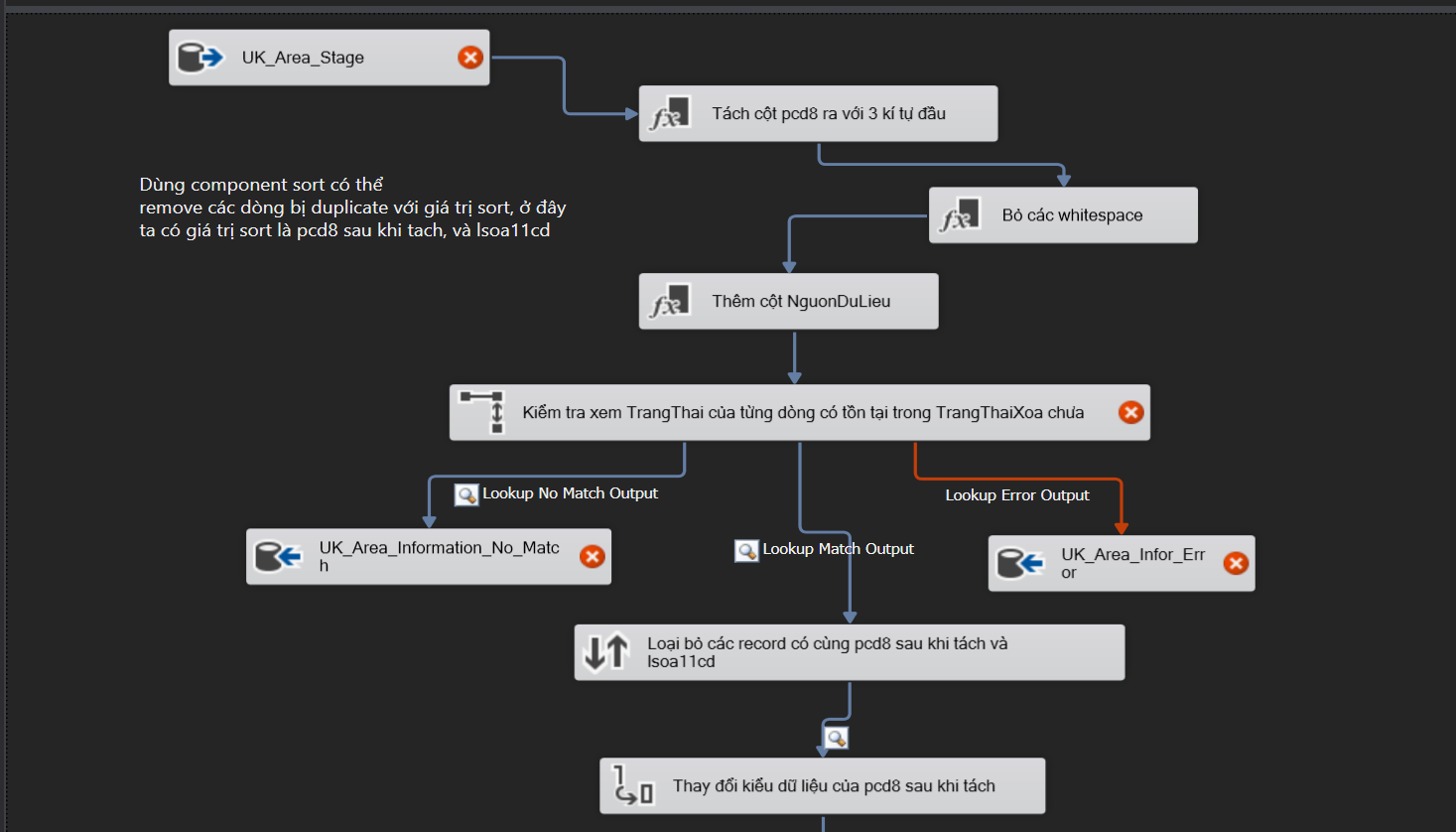
|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột** | **Mô tả** | **Nguồn** | **Transformation** |
| SKJourney\_Purpose\_of\_Driver | Khóa tự tăng của bảng Journey\_Purpose\_of\_Driver |  |  |
| Code | Mã khóa của Journey\_Purpose\_of\_Driver trong source system | QLBI\_NDS.Journey\_Purpose\_of\_Driver .Code |  |
| Description | Mô tả ý nghĩa của mã code | QLBI\_NDS.Journey\_Purpose\_of\_Driver.Description |  |

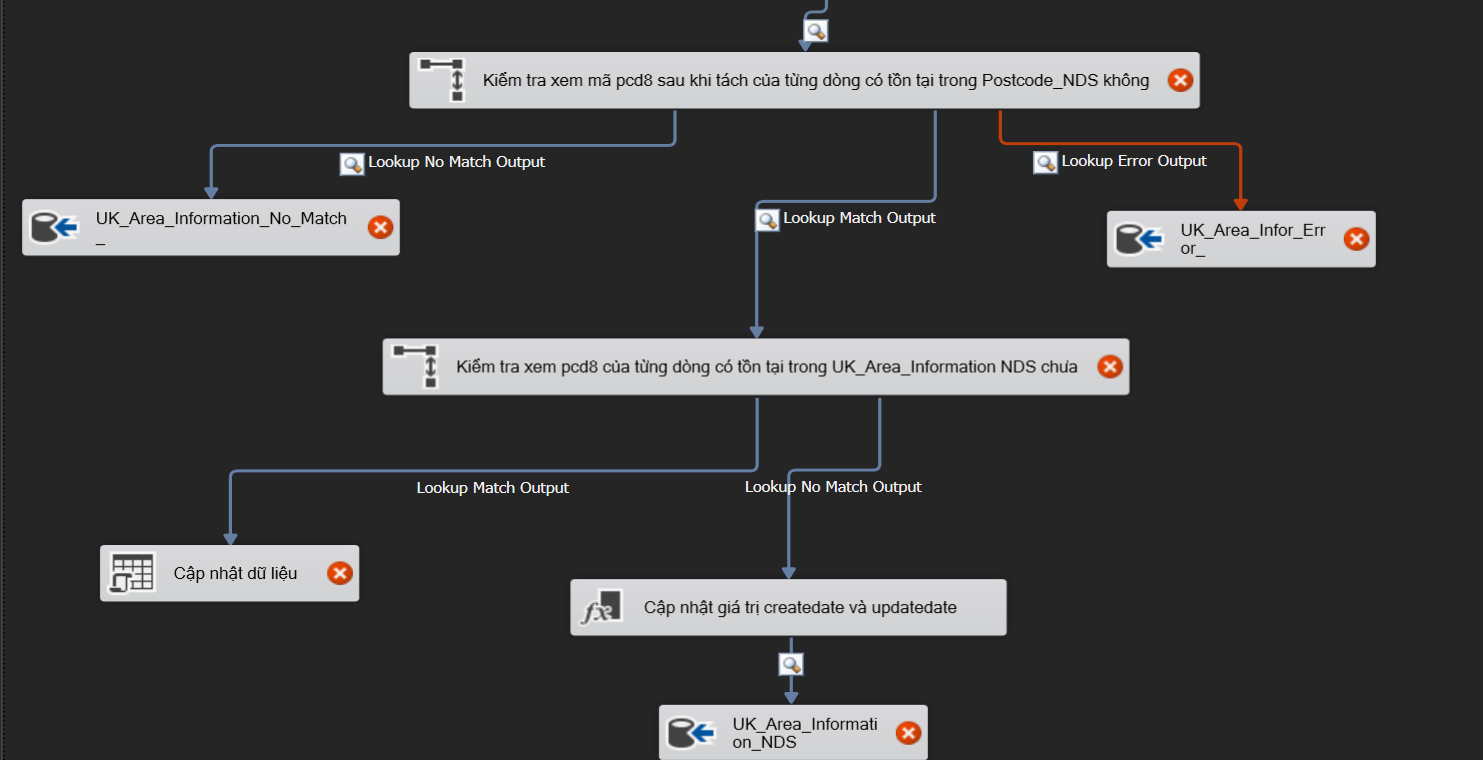
## Flow khi chuyển đổi dữ liệu từ Stage sang NDS

* **Flow chuyển đổi dữ liệu từ Postcode\_Stage sang NDS**

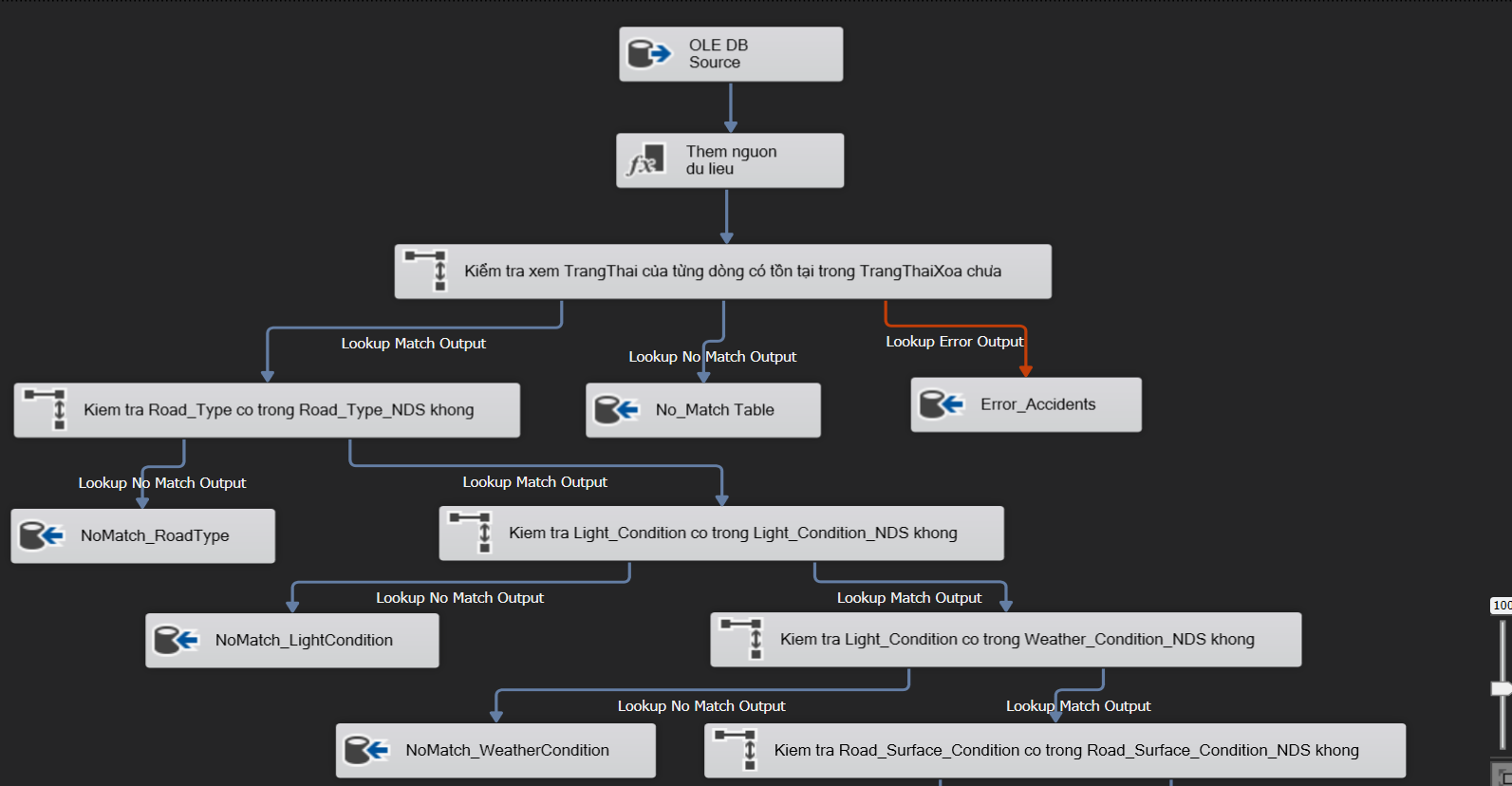
****

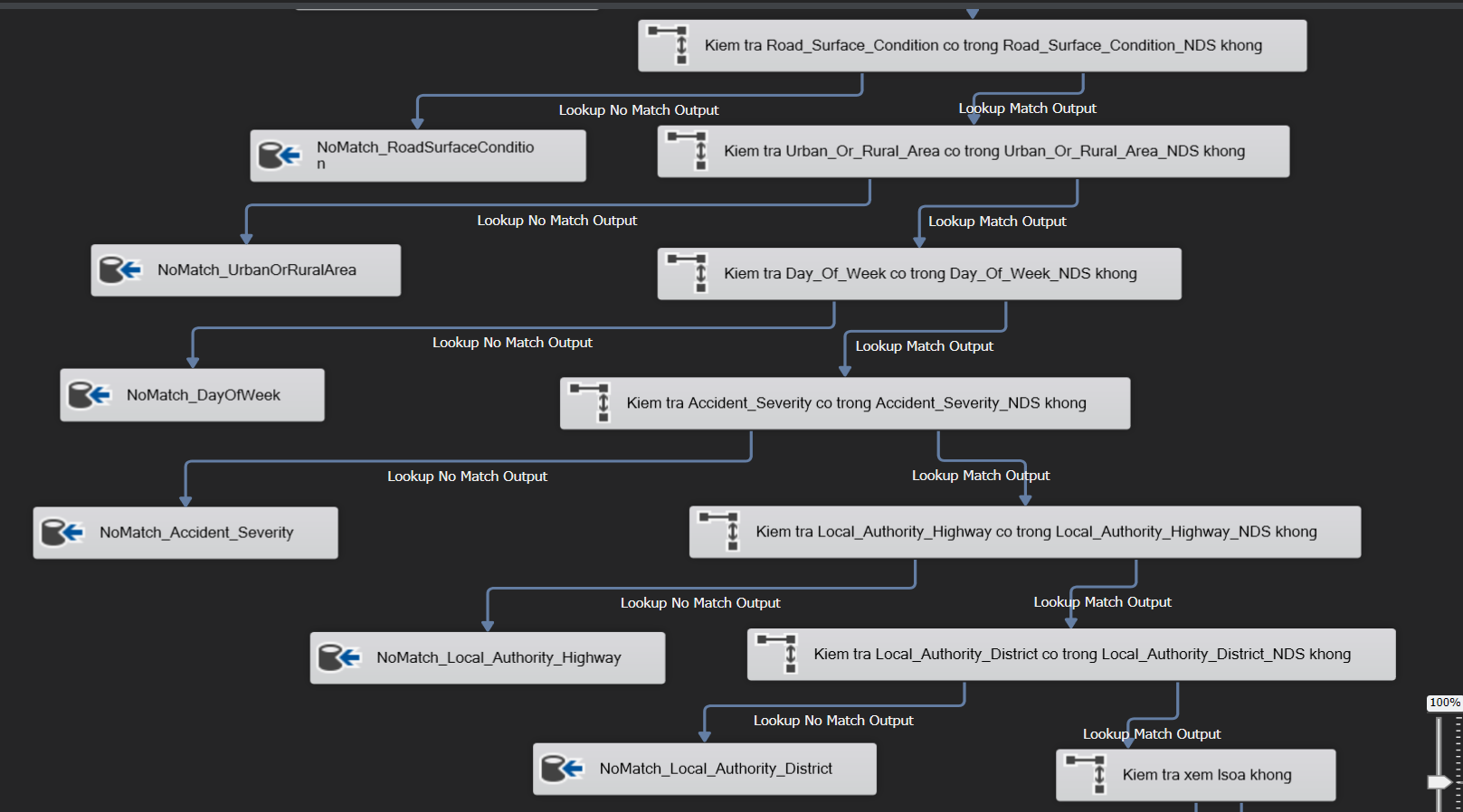
* **Bước 1:** Đổ dữ liệu Postcode\_Stage vào để xử lý.
* **Bước 2**: Thêm cột MaNguonDL trong data pipeline dữ liệu.
* **Bước 3**: Kiểm tra xem TrangThai đã từng tồn tại trong bảng TrangThaiXoa chưa? Nếu có thì trả về SKMaTrangThai. Không có thì đổ vào bảng No\_Match Postcodes. Nếu có lỗi trong quá trình chạy flow thì đổ vào bảng Error\_Postcodes kèm ghi chú lỗi xảy ra.
* **Bước 4:** Kiểm tra xem mã postcode đã từng tồn tại trong bảng Postcode\_NDS chưa? Nếu có thì trả về SKPostcodes và đi tiếp bước 4. Nếu chưa có thì cập nhật lại ngày tạo và ngày cập nhật bằng thời gian hiện tại và đổ dữ liệu vào bảng Postcode\_NDS.
* **Bước 5:** Cập nhật lại dữ liệu.
* **Flow chuyển đổi dữ liệu từ UK\_Area\_Information\_Stage sang NDS**

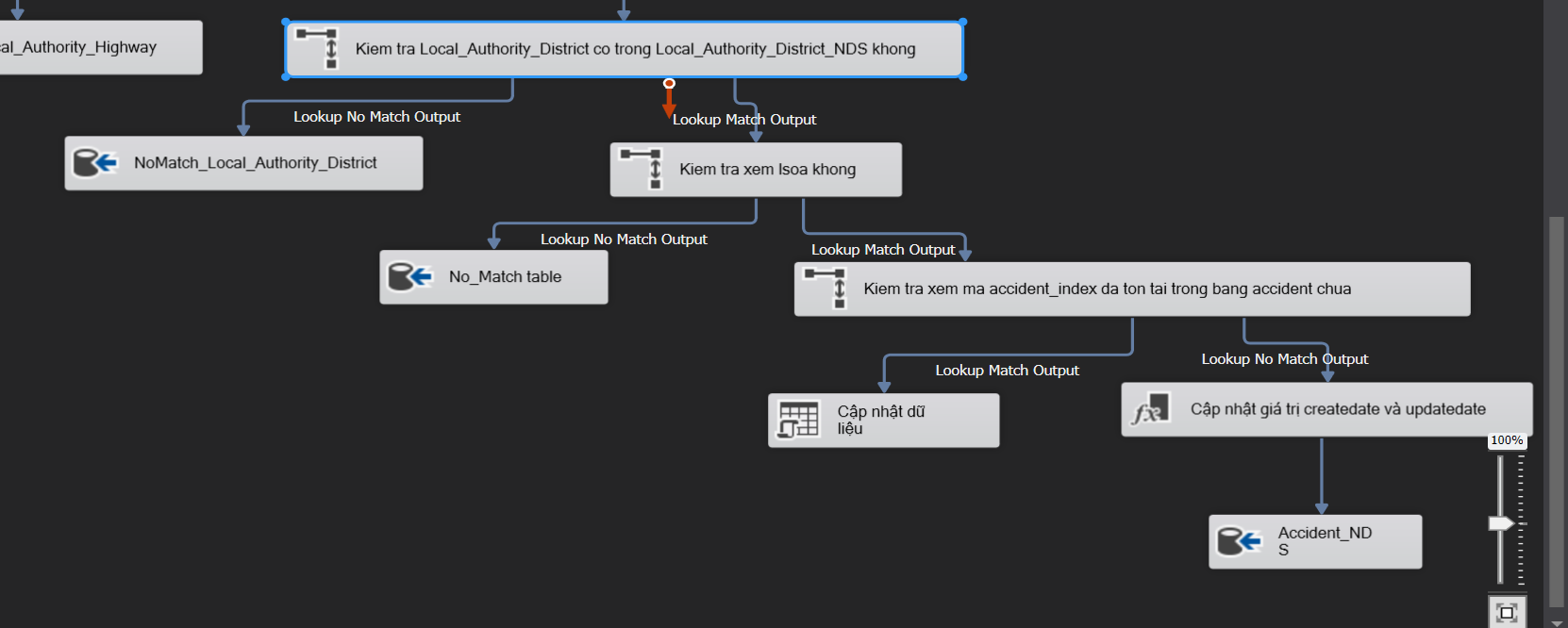
****

****

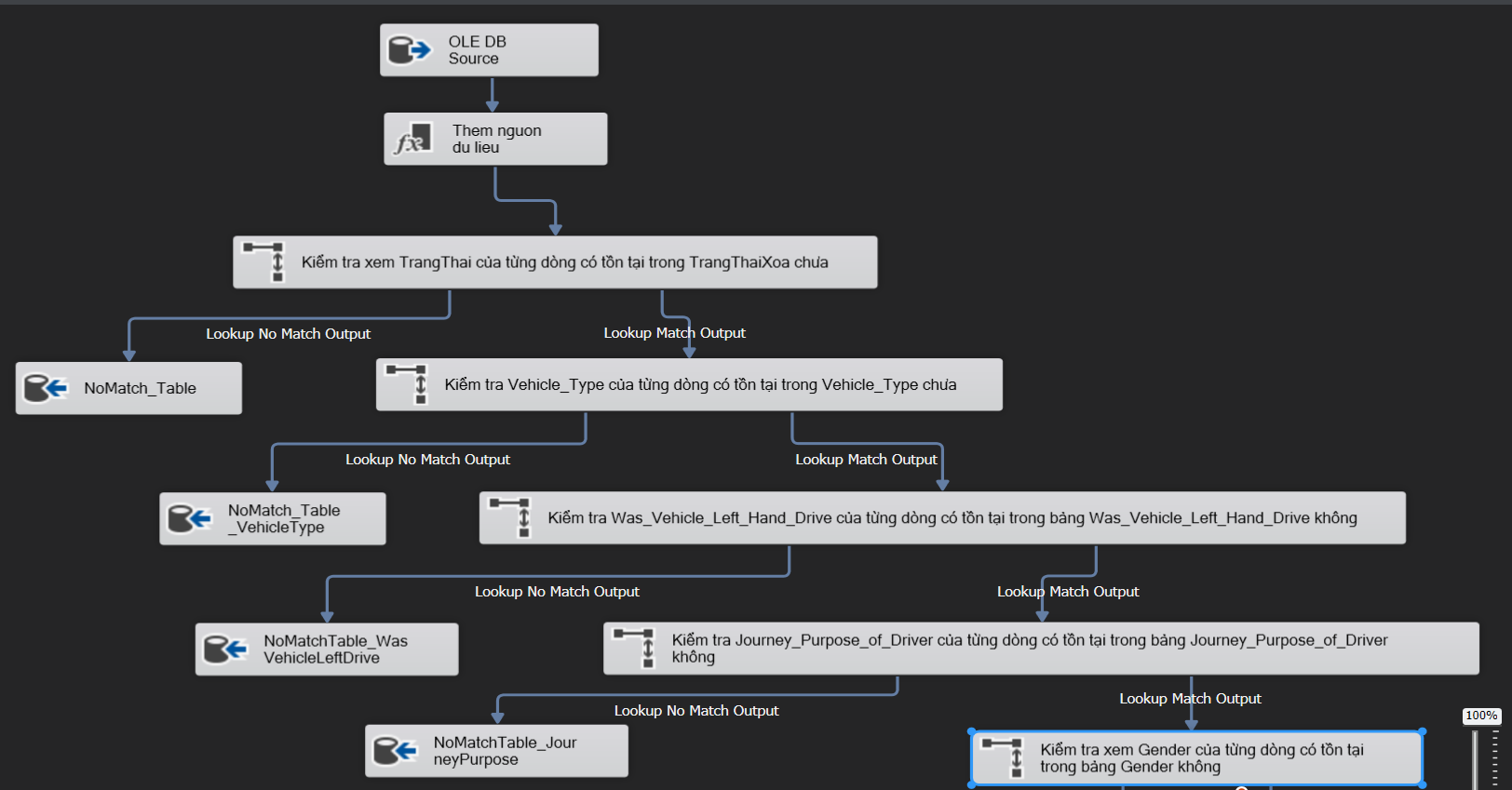
* **Bước 1:** Đổ dữ liệu UK\_Area\_Stage vào để xử lý.
* **Bước 2**: Tách những ký tự đầu trước khoảng trắng của pcd8 để làm khóa. Dùng expression là “LEN(SUBSTRING(pcd8,1,FINDSTRING(pcd8 + ""," ",1))) == 0 ? (SUBSTRING(pcd8,1,LEN(pcd8) - FINDSTRING(REVERSE(pcd8)," ",1) - 3)) : (SUBSTRING(pcd8,1,FINDSTRING(pcd8 + ""," ",1)))”
* **Bước 3:** Sau khi tách sẽ bỏ các khoảng trắng để loại bỏ trường hợp dữ liệu khi tách ra còn dư khoảng trắng. Dùng expression là TRIM(pcd8\_SAUKHITACH).
* **Bước 4**: Thêm cột MaNguonDL.
* **Bước 5**: Kiểm tra xem TrangThai đã từng tồn tại trong bảng TrangThaiXoa chưa? Nếu có thì trả về SKMaTrangThai. Không có thì đổ vô bảng UK\_Area\_Information\_No\_Match. Nếu có lỗi trong quá trình chạy flow thì đổ vô bảng UK\_Area\_Information\_Error kèm ghi chú lỗi xảy ra.
* **Bước 6:** Loại bỏ các record có pcd8 sau khi tách và lsoa11cd giống nhau.
* **Bước 7:**Thay đổi kiểu dữ liệu của pcd8 sau khi tách thành string [DT\_STR].
* **Bước 8:** Kiểm tra xem pcd8 sau khi tách của từng dòng có tồn tại trong bảng Postcode\_NDS chưa? Nếu có thì đi tới bước 9. Không có thì đổ vô bảng UK\_Area\_Information\_No\_Match. Nếu có lỗi trong quá trình chạy flow thì bỏ vô bảng UK\_Area\_Information\_Error kèm ghi chú lỗi xảy ra.
* **Bước 9:** Kiểm tra xem pcd8 của từng dòng đã tồn tại trong bảng UK\_Area\_Information\_Stage chưa? Nếu không có thì cập nhật lại giá trị createdate và updatedate rồi đổ dữ liệu vào UK\_Area\_Information\_NDS. Nếu có thì cập nhật lại dữ liệu
* **Flow chuyển đổi dữ liệu từ Accidents\_Stage sang NDS**

****

****

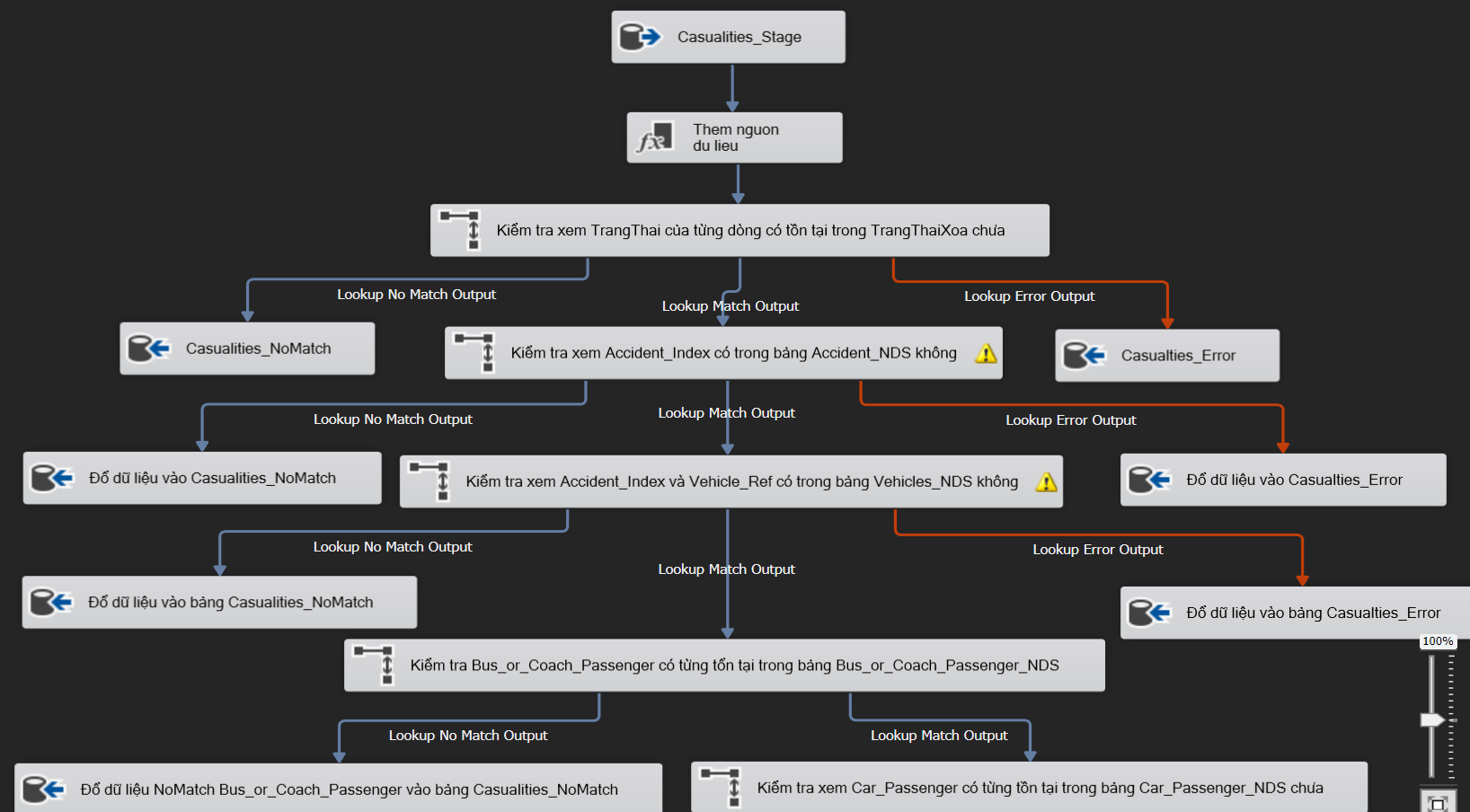
****

* **Bước 1:** Đổ dữ liệu Accidents2011\_2014\_Stage vào để xử lý.
* **Bước 2**: Thêm cột MaNguonDL trong data pipeline dữ liệu.
* **Bước 3**: Kiểm tra xem TrangThai đã từng tồn tại trong bảng TrangThaiXoa chưa? Nếu có thì trả về SKMaTrangThai. Nếu không có thì đổ vào bảng Accidents2011\_2014\_No\_Match. Nếu có lỗi trong quá trình chạy flow thì bỏ vô bảng Accidents2011\_2014\_Error kèm ghi chú lỗi xảy ra.
* **Bước 4:** Kiểm tra xem mã Road\_Type đã từng tồn tại trong bảng Road\_Type\_NDS chưa? Nếu có thì trả về SKRoad\_Type và đi tiếp bước 5. Nếu không có thì đổ vào bảng Accidents2011\_2014\_No\_Match
* **Bước 5**: Kiểm tra xem mã Light\_Conditions đã từng tồn tại trong bảng Light\_Conditions \_NDS chưa? Nếu có thì trả về SKLight\_Conditions và đi tiếp bước 6. Nếu không có thì đổ vào bảng Accidents2011\_2014\_No\_Match.
* **Bước 6**: Kiểm tra xem mã Road\_Surface\_Conditions đã từng tồn tại trong bảng Road\_Surface\_Conditions \_NDS chưa? Nếu có thì trả về SKRoad\_Surface\_Conditions và đi tiếp bước 7. Nếu không có thì đổ vào bảng Accidents2011\_2014\_No\_Match.
* **Bước 7**: Kiểm tra xem mã Urban\_or\_Rural\_Area đã từng tồn tại trong bảng Urban\_or\_Rural\_Area \_NDS chưa? Nếu có thì trả về SKUrban\_or\_Rural\_Area và đi tiếp bước 8. Nếu không có thì đổ vào bảng Accidents2011\_2014\_No\_Match.
* **Bước 8**: Kiểm tra xem mã Day\_of\_Week đã từng tồn tại trong bảng Day\_of\_Week\_NDS chưa? Nếu có thì trả về SKDay\_of\_Week và đi tiếp bước 9. Nếu không có thì đổ vào bảng Accidents2011\_2014\_No\_Match.
* **Bước 9**: Kiểm tra xem mã Accident\_Severity đã từng tồn tại trong bảng Accident\_Severity\_NDS chưa? Nếu có thì trả về SKAccident\_Severity và đi tiếp bước 10. Nếu không có thì đổ vào bảng Accidents2011\_2014\_No\_Match.
* **Bước 10**: Kiểm tra xem mã Local\_Authority\_Highway đã từng tồn tại trong bảng Local\_Authority\_Highway\_NDS chưa? Nếu có thì trả về SKLocal\_Authority\_Highway và đi tiếp bước 11. Nếu không có thì đổ vào bảng Accidents2011\_2014\_No\_Match.
* **Bước 11**: Kiểm tra xem mã Local\_Authority\_District đã từng tồn tại trong bảng Local\_Authority\_District\_NDS chưa? Nếu có thì trả về SKLocal\_Authority\_District và đi tiếp bước 12. Nếu không có thì đổ vào bảng Accidents2011\_2014\_No\_Match.
* **Bước 12**: Kiểm tra xem mã LSOA đã từng tồn tại trong bảng UK\_Area\_Information\_NDS chưa? Nếu có thì trả về SK\_UKArea và đi tiếp bước 13. Nếu không có thì đổ vào bảng Accidents2011\_2014\_No\_Match.
* **Bước 13**: Kiểm tra xem mã Accident\_Index đã từng tồn tại trong bảng Accidents2011\_2014\_NDS chưa? Nếu không có thì cập nhật lại giá trị createdate và updatedate rồi đổ dữ liệu vào Accidents2011\_2014\_NDS. Nếu có thì cập nhật lại dữ liệu.
* **Flow chuyển đổi dữ liệu từ Vehicles\_Stage sang NDS**

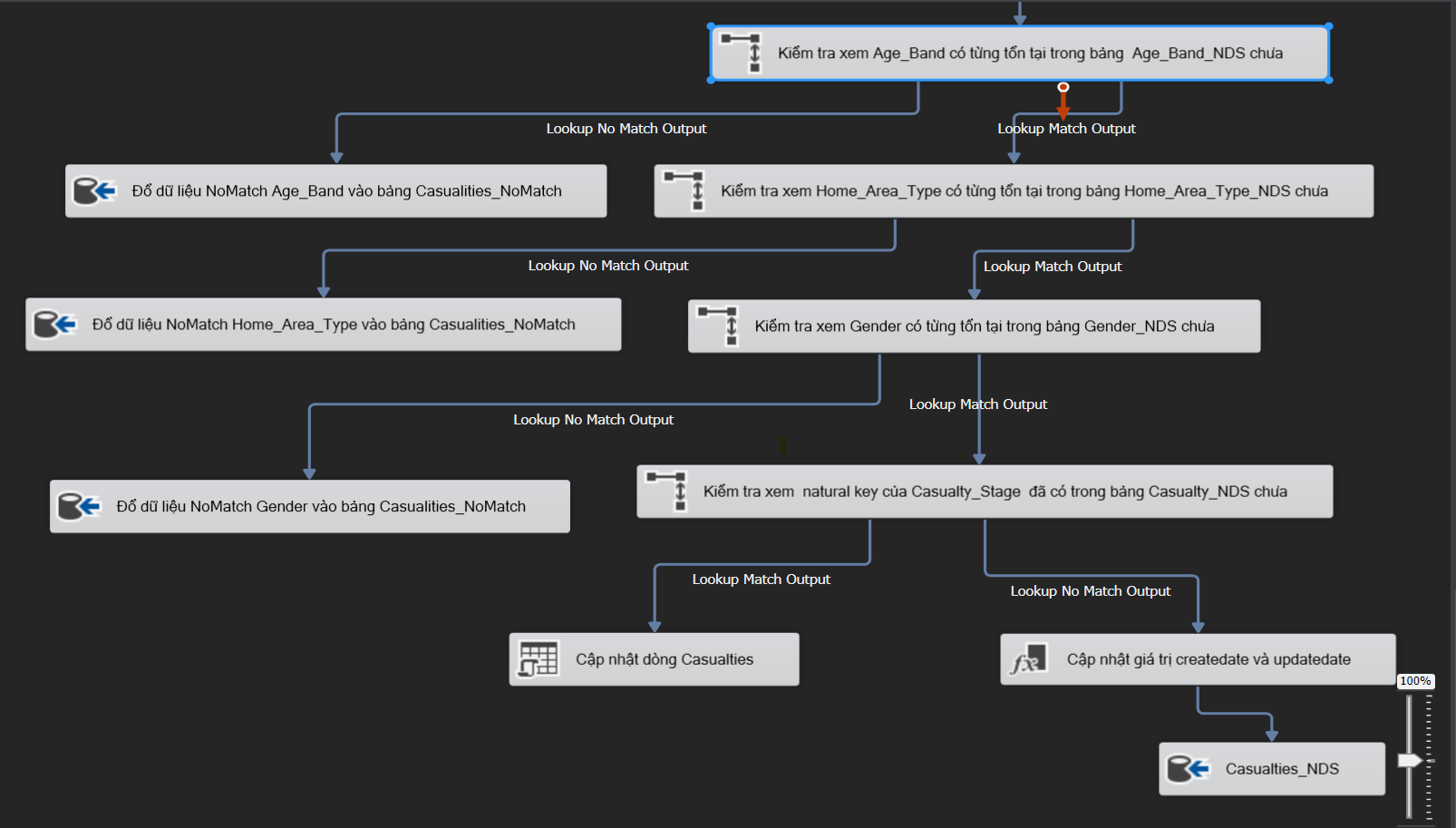
****

****

* **Bước 1:** Đổ dữ liệu Vehicles2011\_2014\_Stage vào để xử lý.
* **Bước 2**: Thêm cột MaNguonDL trong data pipeline dữ liệu.
* **Bước 3**: Kiểm tra xem TrangThai đã từng tồn tại trong bảng TrangThaiXoa chưa? Nếu có thì trả về SKMaTrangThai. Nếu không có thì đổ vào bảng Vehicles2011\_2014\_No\_Match. Nếu có lỗi trong quá trình chạy flow thì bỏ vô bảng Vehicles2011\_2014\_Error kèm ghi chú lỗi xảy ra.
* **Bước 4:** Kiểm tra xem mã Vehicle\_Type đã từng tồn tại trong bảng Vehicle\_Type\_NDS chưa? Nếu có thì trả về SKVehicle\_Type và đi tiếp bước 5. Nếu không có thì đổ vào bảng Vehicles2011\_2014\_No\_Match.
* **Bước 5 :** Kiểm tra xem mã Was\_Vehicle\_Left\_Hand\_Drive đã từng tồn tại trong bảng Was\_Vehicle\_Left\_Hand\_Drive\_NDS chưa? Nếu có thì trả về SKWas\_Vehicle\_Left\_Hand\_Drive và đi tiếp bước 6. Nếu không có thì đổ vào bảng Vehicles2011\_2014\_No\_Match.
* **Bước 6:** Kiểm tra xem mã Journey\_Purpose\_of\_Driver đã từng tồn tại trong bảng Journey\_Purpose\_of\_Driver \_NDS chưa? Nếu có thì trả về SKJourney\_Purpose\_of\_Driver và đi tiếp bước 7. Nếu không có thì đổ vào bảng Vehicles2011\_2014\_No\_Match.
* **Bước 7:** Kiểm tra xem mã Sex\_of\_Driver đã từng tồn tại trong bảng Gender\_NDS chưa? Nếu có thì trả về SKGender và đi tiếp bước 8. Nếu không có thì đổ vào bảng Vehicles2011\_2014\_No\_Match.
* **Bước 8:** Kiểm tra xem mã Age\_Band\_of\_Driver đã từng tồn tại trong bảng Age\_Band \_NDS chưa? Nếu có thì trả về SKAge\_Band và đi tiếp bước 9. Nếu không có thì đổ vào bảng Vehicles2011\_2014\_No\_Match.
* **Bước 9:** Kiểm tra xem mã Driver\_Home\_Area\_Type đã từng tồn tại trong bảng Home\_Area\_Type\_NDS chưa? Nếu có thì trả về SKHome\_Area\_Type và đi tiếp bước 10. Nếu không có thì đổ vào bảng Vehicles2011\_2014\_No\_Match.
* **Bước 10:** Kiểm tra xem mã Accident\_Index đã từng tồn tại trong bảng Accidents2011\_2014\_NDS chưa? Nếu có thì trả về SKAccident và đi tiếp bước 11. Nếu không có thì đổ vào bảng Vehicles2011\_2014\_No\_Match.
* **Bước 11**: Kiểm tra xem mã Vehicle\_Reference đã từng tồn tại trong bảng Vehicles2011\_2014\_NDS chưa? Nếu không có thì cập nhật lại giá trị createdate và updatedate rồi đổ dữ liệu vào Vehicles2011\_2014\_NDS. Nếu có thì cập nhật lại dữ liệu.
* **Flow chuyển đổi dữ liệu từ Casualties\_Stage sang NDS**







* **Bước 1:** Đổ dữ liệu Casualities2011\_2014\_Stage vào để xử lý.
* **Bước 2**: Thêm cột MaNguonDL trong data pipeline dữ liệu.
* **Bước 3**: Kiểm tra xem TrangThai đã từng tồn tại trong bảng TrangThaiXoa chưa? Nếu có thì trả về SKMaTrangThai. Nếu không có thì đổ vào bảng Casualities2011\_2014\_No\_Match. Nếu có lỗi trong quá trình chạy flow thì bỏ vô bảng Casualities2011\_2014\_Error kèm ghi chú lỗi xảy ra.
* **Bước 4:** Kiểm tra xem mã Accident\_Index đã từng tồn tại trong bảng Accidents2011\_2014\_NDS chưa? Nếu có thì trả về SKAccident và đi tiếp bước 5. Nếu không có thì đổ vào bảng Casualities2011\_2014\_No\_Match.
* **Bước 5 :** Kiểm tra xem mã Accident\_Index và Casualty\_Reference đã từng tồn tại trong bảng Vehicles2011\_2014\_NDS chưa? Nếu có thì trả về SKVehicles và đi tiếp bước 6. Nếu không có thì đổ vào bảng Casualities2011\_2014\_No\_Match.
* **Bước 6:** Kiểm tra xem mã Bus\_or\_Coach\_Passenger đã từng tồn tại trong bảng Bus\_or\_Coach\_Passenger\_NDS chưa? Nếu có thì trả về SKBus\_or\_Coach\_Passenger và đi tiếp bước 7. Nếu không có thì đổ vào bảng Casualities2011\_2014\_No\_Match.
* **Bước 7:** Kiểm tra xem mã Pedestrian\_Location đã từng tồn tại trong bảng Pedestrian\_Location\_NDS chưa? Nếu có thì trả về SKPedestrian\_Location và đi tiếp bước 8. Nếu không có thì đổ vào bảng Casualities2011\_2014\_No\_Match.
* **Bước 8:** Kiểm tra xem mã Casualty\_Class đã từng tồn tại trong bảng Casualty\_Class\_NDS chưa? Nếu có thì trả về SKCasualty\_Classvà đi tiếp bước 9. Nếu không có thì đổ vào bảng Casualities2011\_2014\_No\_Match.
* **Bước 9:** Kiểm tra xem mã Casualty\_Severity đã từng tồn tại trong bảng Casualty\_Severity\_NDS chưa? Nếu có thì trả về SKCasualty\_Severity và đi tiếp bước 10. Nếu không có thì đổ vào bảng Casualities2011\_2014\_No\_Match.
* **Bước 10:** Kiểm tra xem mã Pedestrian\_Movement đã từng tồn tại trong bảng Pedestrian\_Movement\_NDS chưa? Nếu có thì trả về SKPedestrian\_Movement và đi tiếp bước 11. Nếu không có thì đổ vào bảng Casualities2011\_2014\_No\_Match.
* **Bước 11:** Kiểm tra xem mã Casualty\_Type đã từng tồn tại trong bảng Casualty\_Type\_NDS chưa? Nếu có thì trả về SKCasualty\_Type và đi tiếp bước 12. Nếu không có thì đổ vào bảng Casualities2011\_2014\_No\_Match.
* **Bước 12:** Kiểm tra xem mã Age\_Band\_of\_Casualty đã từng tồn tại trong bảng Age\_Band\_NDS chưa? Nếu có thì trả về SKAge\_Band và đi tiếp bước 13. Nếu không có thì đổ vào bảng Casualities2011\_2014\_No\_Match.
* **Bước 13:** Kiểm tra xem mã Casualty\_Home\_Area\_Type đã từng tồn tại trong bảng Home\_Area\_Type\_NDS chưa? Nếu có thì trả về SKHome\_Area\_Type và đi tiếp bước 14. Nếu không có thì đổ vào bảng Casualities2011\_2014\_No\_Match.
* **Bước 14:** Kiểm tra xem mã Sex\_of\_Casualty đã từng tồn tại trong bảng Gender\_NDS chưa? Nếu có thì trả về SKGender và đi tiếp bước 15. Nếu không có thì đổ vào bảng Casualities2011\_2014\_No\_Match.
* **Bước 15**: Kiểm tra xem mã SKCasuality, Casualty\_Reference, SKVehicles đã từng tồn tại trong bảng Casualities2011\_2014\_NDS chưa? Nếu không có thì cập nhật lại giá trị createdate và updatedate rồi đổ dữ liệu vào Casualities2011\_2014\_NDS. Nếu có thì cập nhật lại dữ liệu.

# References

<http://doc.ukdataservice.ac.uk/doc/7961/mrdoc/pdf/7961_nspl_user_guide_2011.pdf>

[Lower Super Output Area - Area (findthatpostcode.uk)](https://findthatpostcode.uk/areatypes/lsoa11.html) (lsoa)

[Output area - Area (findthatpostcode.uk)](https://findthatpostcode.uk/areatypes/oa11.html) (oa)

[Middle Super Output Area - Area (findthatpostcode.uk)](https://findthatpostcode.uk/areatypes/msoa11.html) (msoa)

[Local Authority - Area (findthatpostcode.uk)](https://findthatpostcode.uk/areatypes/laua.html?p=45) (ladcd)

[Postcode to Output Area to Lower Layer Super Output Area to Middle Layer Super Output Area to Local Authority District (February 2018) Lookup in the UK - data.gov.uk](https://data.gov.uk/dataset/9b090605-9861-4bb4-9fa4-6845daa2de9b/postcode-to-output-area-to-lower-layer-super-output-area-to-middle-layer-super-output-area-to-local-authority-district-february-2018-lookup-in-the-uk)