ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

KHOA CÔNG NGHỆ THÔNG TIN



**MÔN** **HỌC:** Hệ thống thông tin phục vụ trí tuệ kinh doanh

**CHỦ** **ĐỀ**: Đồ án thực hành - Xây dựng và khai thác KDL

**NHÓM**: TTKD-10

**GVHD**: Hồ Thị Hoàng Vy, Nguyễn Thị Như Anh

Người thực hiện

18120469 – Nguyễn Hoài Nam

18120510 – Cao Xuân Hồng Phúc

18120518 – Phạm Thị Bích Phượng

18120538 – Võ Nguyễn Hồng Sơn

TP Hồ Chí Minh, ngày 18 tháng 9 năm 2021

NHẬN XÉT CỦA GIÁO VIÊN

Tp. Hồ Chí Minh, ngày … tháng … năm 2019

Giáo viên lý thuyết

NHẬN XÉT CỦA GIẢNG VIÊN

Tp. Hồ Chí Minh, ngày … tháng … năm 2020

Giảng viên hướng dẫn

# 

**MỤC LỤC**

[I. THÔNG TIN NHÓM 4](#_Toc86949577)

[II. BẢNG PHÂN CÔNG CÔNG VIỆC VÀ ĐÁNH GIÁ 4](#_Toc86949578)

[III. NỘI DUNG 5](#_Toc86949579)

[1. Mô tả ý nghĩa các thuộc tính của các nguồn dữ liệu 5](#_Toc86949580)

[**a.** **Accidents** 5](#_Toc86949581)

[**b.** **Vehicles** 7](#_Toc86949582)

[**c.** **Casualties** 9](#_Toc86949583)

[**d.** **Postcodes** 11](#_Toc86949584)

[**e.** **PCD\_OA\_LSOA\_MSOA\_LAD\_AUG21\_UK\_LU** 12](#_Toc86949585)

[2. Mô tả quá trình ETL từ Source vào Stage 12](#_Toc86949586)

[IV. LINK GITHUB 15](#_Toc86949587)

# THÔNG TIN NHÓM

|  |  |  |
| --- | --- | --- |
| MSSV | HỌ TÊN | EMAIL |
| 18120469 | Nguyễn Hoài Nam | 18120469@student.hcmus.edu.vn |
| 18120510 | Cao Xuân Hồng Phúc | 18120510@student.hcmus.edu.vn |
| 18120518 | Phạm Thị Bích Phượng | 18120518@student.hcmus.edu.vn |
| 18120538 (\*) | Võ Nguyễn Hồng Sơn | 18120538@student.hcmus.edu.vn |

**(\*): Nhóm trưởng**

# BẢNG PHÂN CÔNG CÔNG VIỆC VÀ ĐÁNH GIÁ

|  |  |  |  |
| --- | --- | --- | --- |
| **Người thực hiện** | **Công việc thực hiện** | **Mức độ hoàn thành** | **Đánh giá của nhóm** |
| 18120469 – Nam | * Mô tả ý nghĩa thuộc tính của bảng PCD\_OA\_LSOA\_MSOA\_LAD\_AUG21\_UK\_LU | 100% | 10/10 |
| 18120510 – Phúc | * Mô tả ý nghĩa thuộc tính của bảng Postcodes | 100% | 10/10 |
| 18120518 – Phượng | * Mô tả ý nghĩa thuộc tính của bảng Vehicles , Casualities | 100% | 10/10 |
| 18120538 – Sơn | * Mô tả ý nghĩa thuộc tính của bảng Accidents | 100% | 10/10 |

# NỘI DUNG

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

### **Accidents**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | Accident Index | string | Mã vụ tai nạn |
| 2 | Police Force | int | Mã lực lượng cảnh sát |
| 3 | Accident Severity | int | Mức độ nghiêm trọng: tử vong, nghiêm trọng , nhẹ |
| 4 | Day of Week | int | Các ngày thứ trong tuần: Sunday, Monday,… |
| 5 | Local Authority (District) | int | Tên khu vực địa phương xảy ra vụ việc |
| 6 | Local Authority (Highway Authority - ONS code) | string | Tên đường chính xảy ra tai nạn |
| 7 | 1st Road Class | int | Đường cấp 1 |
| 8 | Road Type | int | Loại đường:  EX: One way street ( Đường 1 chiều) |
| 9 | Junction Detail | int | Chi tiết giao lộ:   * Bùng binh * Ngả tư |
| 10 | Junction Control | int | Kiểm soát giao lộ:   * Người có quyền * Tín hiệu giao thông |
| 11 | 2nd Road Class | int | Đường cấp 2 |
| 12 | Pedestrian Crossing-Human Control | int | Người kiểm soát phân luồng đưa người qua đường |
| 13 | Pedestrian Crossing-Physical Facilities | int | Cơ sở vật chất dành cho người đi bộ |
| 14 | Light Conditions | int | Các điều kiện ánh sáng  Ví dụ: bóng tối – đèn không sáng |
| 15 | Weather Conditions | int | Các điều kiện thời tiết  Ví dụ: Gió mạnh - mưa |
| 16 | Road Surface Conditions | int | Các điều kiện mặt đường  Ví dụ: Khô, ẩm |
| 17 | Special Conditions at Site | int | Các điều kiện đặc biệt ở hiện trường  Ví dụ: Mặt đường bị lỗi |
| 18 | Carriageway Hazards | int | Các mối nguy hiểm trên đường đi:  Ví dụ: Chó trên đường, tai nạn trước đó |
| 19 | Urban or Rural Area | int | Thành thị hay nông thôn |
| 20 | Did Police Officer Attend Scene of Accident | int | Cảnh sát có tham dự hiện trường vụ tai nạn không |
| 21 | LSOA\_of\_Accident\_Location | string | Mã địa điểm tai nạn |

### **Vehicles**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | Accident\_Index | String | Số thứ tự chỉ mục các vụ tai nạn |
| 2 | Vehicle\_Reference | Int | Thứ tự của phương tiện trong vụ tai nạn |
| 3 | Vehicle\_Type | Int | Các giá trị được biểu diễn dạng số theo bảng UK Accients – Codebook. Mỗi giá trị tương ứng với một loại xe khác nhau. Ví dụ 1: Pedal cycle, 18 :Tram,…. |
| 4 | Vehicle\_Manoeuvre | Int | Trước lúc tai nạn, phương tiện đã di chuyển như thế nào. 1: Nearside, 7:Offside. |
| 5 | Vehicle\_Location-Restricted\_Lane | Int | Vị trí của xe có vi phạm các làn đường bị cấm đi hay không. |
| 6 | Junction\_Location | Int | Vị trí xe trên giao lộ. Mỗi giá trị từ 0 đến 8 trong bảng Codebook thể hiện vị trí của xe. -1 là dữ liệu không có hoặc vượt quá khoảng giá trị trên. |
| 7 | Skidding\_and\_Overturning | Int | Phương tiện có bị mất lái trượt hay lật hay ko. 1: trượt đi, 2: trượt và lật , 4: bị rơi một phần của xe và lật. |
| 8 | Hit\_Object\_in\_Carriageway | Int | Xe đã tông vào vật thể gì trên tuyến đường đi.  Các giá trị 1-12 thể hiện các vật thể mà xe tông phải. Ngoại trừ -1: dữ liệu null hoặc ngoai khoảng giá trị. 0: không. |
| 9 | Vehicle\_Leaving\_Carriageway | Int | Xe văng khỏi đường đi như thế nào. Các giá trị được ghi trong code book. Ví dụ 1:Nearside(về phía bên trái, gần với giải phân cách), 7: Offside( về phía bên phải, gần với vạch kẻ đường ở giữa).  Lưu ý: UK đi xe bên tay trái. |
| 10 | Hit\_Object\_off\_Carriageway | Int | 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 | Int | Điêm đầu tiên của phương tiện bị ảnh hưởng. 0: did not impact, 1:Front, 2:Back, … |
| 12 | Was\_Vehicle\_Left\_Hand\_Drive? | Int | Tài xế đã bị mất lái. 1:No, 2:Yes. |
| 13 | Journey\_Purpose\_of\_Driver | Int | Mục đích di chuyển của tài xế.  Các giá trị biểu diên bằng số tương đương với. 1: Journey as part of work, 2: Commuting to/from work,… |
| 14 | Sex\_of\_Driver | Int | Giới tinh của tài xế. Giới tính được biểu diễn bằng số tương đương với các giá trị sau: 1: Male, 2:Female, 3: not known, -1: Data missing or out of range. |
| 15 | Age\_of\_Driver | Int | Tuổi hiện tại của tài xế. |
| 16 | Age\_Band\_of\_Driver | Int | Thang đo độ tuổi của tài xế. Giá trị từ 1-11 thể hiện thang đo độ của tuổi của tài xế trong bảng codebook. Ví dụ tuổi 56: 9. |
| 17 | Engine\_Capacity\_(CC) | Int | Dung tích của động cơ. Đơn vị đo là CC |
| 18 | Propulsion\_Code | Int | Động cơ sẽ chạy bằng loại nhiên liệu nào. Ví dụ 1:Petrol, 2: Heavy oil, M:undefined. |
| 19 | Age\_of\_Vehicle | Int | Tuổi hiện tại của phương tiện. |
| 20 | Driver\_IMD\_Decile | Int | 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 | Int | Loại khu vực mà tài xế sinh sống. 1: Urban Area, 2: Small Town, 3:Rural. |

### **Casualties**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | Accident\_Index | String | Số thứ tự chỉ mục các vụ tai nạn |
| 2 | Vehicle\_Reference | Int | Thứ tự của phương tiện của nạn nhân trong vụ tai nạn |
| 3 | Casualty\_Reference | Int | Số nạn nhân liên quan đến vụ tai nạn. |
| 4 | Casualty\_Class | Int | Nạn nhân thuộc nhóm người nào. 1: Driver or rider, 2: passenger, 3: Pedestrian. |
| 5 | Sex\_of\_Casualty | Int | Giới tính của nạn nhân. 1:Male, 2:Female, -1: Data missing or out of range |
| 6 | Age\_of\_Casualty | Int | Tuổi của nạn nhân. |
| 7 | Age\_Band\_of\_Casualty | Int | Thang đo độ tuổi của nạn nhân. Giá trị từ 1-11 thể hiện thang đo độ của nhóm tuổi của nạn nhân trong bảng codebook. Ví dụ tuổi 56: 9. |
| 8 | Casualty\_Severity | Int | Mức độ nghiêm trọng của xảy ra với nan nhân. Các giá trị biểu diễn kiểu số. 1: Fatal ,2:Serious, 3: Slight |
| 9 | Pedestrian\_Location | Int | Vị trí của đi bộ của nạn nhân. Ví dụ  0: not a Pedestrian, 1: Crossing on pedestrian crossing facility, 2: Crossing in zig-zag approach lines,…. |
| 10 | Pedestrian\_Movement | Int | Nạn nhân đi bộ di chuyển như thế nào. Ví dụ 0: not a Pedestrian,2: Crossing from nearside - masked by parked or stationary vehicle, 1: Crossing from driver’s nearside….. |
| 11 | Car\_Passenger | Int | Nếu nạn nhân đi car. Thì nạn nhân ngồi ở vị trí nào trên xe. Ví dụ các giá trị 0: not car passenger, 1: Front sear passenger, 2: Rear seat passenger, -1: Data missing or out of range. |
| 12 | Bus\_or\_Coach\_Passenger | Int | Có phải nạn nhân khi đi xe bus hoặc xe khách không. Ví dụ 0: not a bus or coach passenger( nạn nhân là người đi bộ), 1: Boarding, 2: Alighting(xuống xe),…. |
| 13 | Pedestrian\_Road\_Maintenance\_Worker | Int | Có phải nạn nhân là công nhân bảo trì đường cho người đi bộ. Mỗi giá trị số được biểu diễn có ý nghia riêng. 0: no/ not applicable, 1: Yes, 2: not known, -1:Data missing or out of range. |
| 14 | Casualty\_Type | Int | Loại phương tiện mà nạn nhân sử dụng. Mỗi giá trị số được biểu diễn có ý nghia riêng. 0: Pedestrian, 1: Cyclist, 2: Motorcycle 50cc and under rider or passenger,…. |
| 15 | Casualty\_Home\_Area\_Type | Int | Loại khu vực mà nạn nhân sinh sống. 1: Urban Area, 2: Small Town, 3:Rural. |

### **Postcodes**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | postcode | string | Mã bưu điện |
| 2 | easting | int | Hướng bắc - khoảng cách về phía bắc của vĩ độ. (đơn vị: mét) |
| 3 | northing | int | Hướng đông - Khoảng cách về phía đông của kinh độ. (đơn vị: mét) |
| 4 | latitude | float | Vĩ độ ( đơn vị: độ) |
| 5 | longitude | float | Kinh độ(đơn vị: độ) |
| 6 | city | string | Thành phố |
| 7 | county | string | Hạt - đơn vị hành chính cao nhất ở Anh |
| 8 | country\_code | string | Mã quốc gia - 3 chữ cái |
| 9 | country\_name | string | Tên quốc gia |
| 10 | iso3166-2 | string | Mã quốc gia ISO |
| 11 | region\_code | string | Mã vùng |
| 12 | region\_name | string | Tên vùng |

### **PCD\_OA\_LSOA\_MSOA\_LAD\_AUG21\_UK\_LU**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | pcd7 | string | Mã bưu điện 7 ký tự |
| 2 | pcd8 | string | Mã bưu điện 8 ký tự |
| 3 | pcds | string | Mã bưu điện có số ký tự tùy chỉnh |
| 4 | dointr | int | Ngày bắt đầu |
| 5 | Doterm | int | Ngày kết thúc |
| 6 | usertype | bool | Loại người dùng  0 = small user; 1 = large user |
| 7 | oa11cd | string | Vùng đầu ra 2011 |
| 8 | lsoa11cd | string | Vùng đầu ra lớp dưới 2011 |
| 9 | msoa11cd | string | Vùng đầu ra Lớp giữa 2011 |
| 10 | ladcd | string | Mã chính quyền đại phương |
| 11 | lsoa11nm | string | Địa chỉ Vùng đầu ra lớp dưới 2011 |
| 12 | msoa11nm | string | Địa chỉ Vùng đầu ra Lớp giữa 2011 |
| 13 | ladnm |  | Địa chỉ chính quyền đại phương |
| 14 | ladnmw |  |  |

## Mô tả quá trình ETL từ Source vào Stage

1. **Load Accidents, Casualties và Vehicles vào Stage**

Graphical user interface

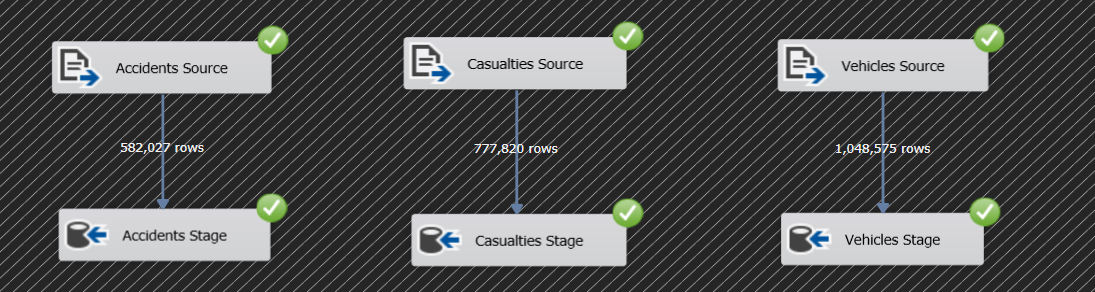
Description automatically generated

***Hình 1. Tạo Data Flow Task để thực hiện load 3 bảng Accidents, Casualties and Vehicles vào Stage***

***A picture containing timeline

Description automatically generated***

***Hình 2. Cấu hình các Sources và Destinations phù hợp.***



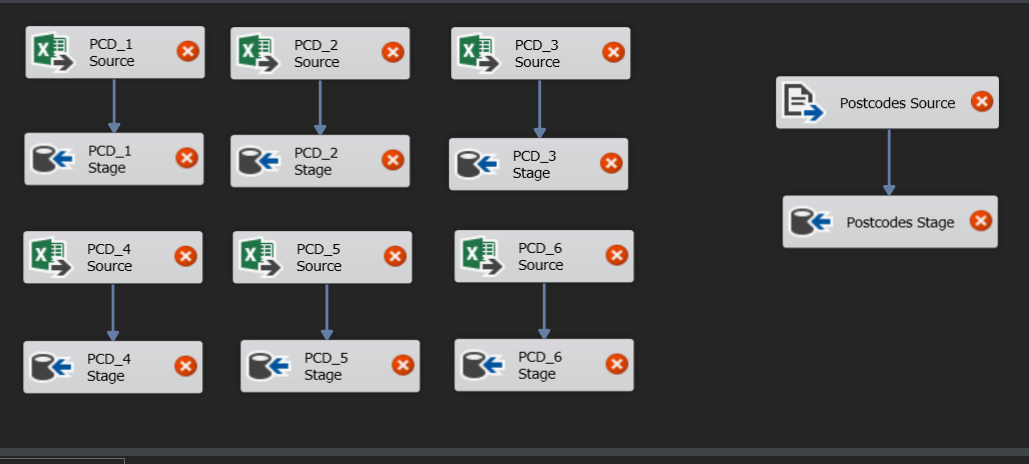
***Hình 3. Run thành công việc load từ Source vào Stage***

1. **Load Postcodes và PCD\_OA\_LSOA\_MSOA\_LAD\_AUG21\_UK\_LU vào Stage**

**Graphical user interface, text, application

Description automatically generated**

***Hình 4. Tạo Data Flow Task để thực hiện load PCD\_LSOA và Postcodes vào Stage***

****

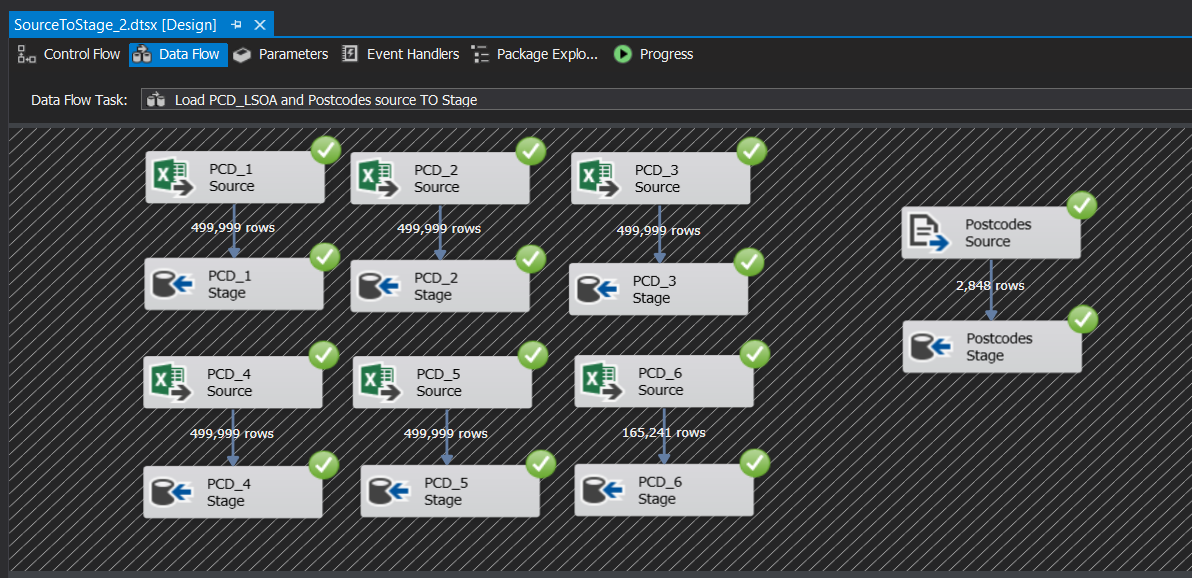
***Hình 5. Cấu hình các Sources và Destinations phù hợp.***

* Vì file PCD\_OA\_LSOA\_MSOA\_LAD\_AUG21\_UK\_LU.csv vượt quá 2 triệu dòng nên nhóm em đã SPLIT thành 6 file excel nhỏ để nạp vào Stage

Graphical user interface, text

Description automatically generated

* Sau khi cấu hình ta RUN và thành công như hình bên dưới:



# LINK GITHUB

<https://github.com/phucjeya/TTKD-10_DATH>