**BỘ MÔN HỆ THỐNG THÔNG TIN – KHOA CÔNG NGHỆ THÔNG TIN**

Nhóm sinh viên thực hiện: Nhóm 09

1712468 – Võ Công Huân

1712748 – Nguyễn Văn Tâm

1712566 – Phạm Thị Kiều Loan

1712935 – Phommala Sisouvanh

GV phụ trách: Cô Hồ Thị Hoàng Vy  
 Cô Tiết Gia Hồng

ĐỒ ÁN MÔN HỌC: HỆ THỐNG THÔNG TIN PHỤC VỤ TRÍ TUỆ KINH DOANH

HỌC KỲ I – NĂM HỌC 2022 - 2023

**HTTT PHỤC VỤ TRÍ TUỆ KINH DOANH**

**ĐẠI HỌC KHOA HỌC TỰ NHIÊN THÀNH PHỐ HỒ CHÍ MINH, ĐẠI HỌC QUỐC GIA TP HCM**



**BẢNG THÔNG TIN CHI TIẾT NHÓM**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Mã nhóm:** | CQ2019-BI#9 | | | |
| **Tên nhóm:** | NHÓM 09 | | | |
| **Số lượng:** | **4** | | | |
| **MSSV** | **Họ tên** | **Email** | **Điện thoại** | **Hình ảnh** |
| 1712468 | Võ Công Huân | conghuanbttbqn@gmail.com | 0399776804 |  |
| 1712748 | Nguyễn Văn Tâm | ntam597.van@gmail.com | 0373222833 |  |
| 1712935 | Phommala Sisouvanh | tonsisouvanh7@gmail.com | 0375002567 |  |
| 1712566 | Phạm Thị Kiều Loan | phamthikieuloan99@gmail.com | 0378136224 |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Bảng phân công & đánh giá hoàn thành công việc** | | | |
| **Công việc thực hiện** | **Người thực hiện** | **Mức độ hoàn thành** | **Đánh giá của nhóm** |
| Tìm hiểu và mô tả ý nghĩa của các thuộc tính trong file dữ liệu COVID19 |  |  |  |
| Thực hiện Data Profiling để kiểm soát các kiểu dữ liệu cần lưu trữ |  |  |  |
| Tạo project ETL and đẩy lên github |  |  |  |
| Tạo file report, mô tả lại thuộc tính của các bảng, push lên github |  |  |  |
| Tạo package để xử lý nạp dữ liệu từ file xuống database |  |  |  |
| Thiết kế mô hình NDS |  |  |  |
| Thiết kế mô hình DDS |  |  |  |
| Viết script để tạo cơ sở dữ liệu Stage, Meta data. |  |  |  |
| Viết script để tạo cơ sở dữ liệu NDS |  |  |  |
| Viết script để tạo cơ sở dữ liệu DDS |  |  |  |
| Tạo package để xử lý nạp dữ liệu từ Source vào Stage |  |  |  |
| Tạo package để xử lý nạp dữ liệu từ Stage vào NDS |  |  |  |
| Tạo package để xử lý nạp dữ liệu từ NDS vào DDS |  |  |  |
| Unit test quá trình ETL |  |  |  |
| Tạo project OLAP và push lên github |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**YÊU CẦU ĐỒ ÁN- BÀI TẬP**

|  |  |
| --- | --- |
| **Loại bài tập** | **🞎 Lý thuyết 🗹 Thực hành 🗹 Đồ án 🞎 Bài tập** |
| **Ngày bắt đầu** |  |
| **Ngày kết thúc** | **29/12/2022** |

# **Yêu cầu của Đồ án/Bài tập**

* Xây dựng và phân tích dữ liệu về Covid-19 trong các năm 2020 – 2022
* Thiết kế kho dữ liệu (KDL), tổng hợp, nạp dữ liệu các nguồn vào KDL và thiết kế, xây dựng Cube.

# **Kết quả**

1. **Mô tả ý nghĩa các thuộc tính của các nguồn dữ liệu sau:**
2. **Bảng Case Report**

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Mô tả** | **Kiểu dữ liệu** |
| 1 | **Outcome** | Kết quả:   * Resolved - Điều trị * Fatal - Tử vong | Varchar |
| 2 | **Age** | Tuổi của bệnh nhân | Varchar |
| 3 | **Gender** | Giới tính | Varchar |
| 4 | **Reporting PHU** | Các PHU được báo cáo | varchar |
| 5 | **SpecimenDate** | Ngày lấy mẫu | date |
| 6 | **CaseReported Date** | Ngày trả báo cáo | date |
| 7 | **PHUCity** | Thành phố của các PHU được báo cáo. | varchar |
| 8 | **TestReported Date** | Ngày trả kết quả | date |
| 9 | **CaseAcquisition info** | Thông tin ca nhiễm:   * CC: dương tính xác định được nguồn lây (closed contact) * No known Epi-link: dương tính không rõ nguồn lây * OB: bùng phát (Outbreak) * Travel | Varchar |
| 10 | **AccurateEpisode Dt** | Ngày khởi phát | date |
| 11 | **PHU Address** | Địa chỉ của PHU | varchar |
| 12 | **PHU Website** | Địa chỉ website của PHU | varchar |
| 13 | **OutbreakRelated** | Có liên quan đến đợt bùng phát | varchar |
| 14 | **PHU Latitude** | Vĩ tuyến của PHU | real |
| 15 | **PHU Longitude** | Kinh tuyển của PHU | real |
| 16 | **PHU Postal Code** | Mã vùng của PHU | varchar |

1. **Bảng Compiled COVID-19 CASE DETAILS**

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Mô tả** | **Kiểu dữ liệu** |
| 1 | **row\_id** |  | int |
| 2 | **date\_reported** | Ngày báo cáo | datetime |
| 3 | **health\_region** | Tên PHU | varchar |
| 4 | **age\_group** | Nhóm tuổi, được phân loại (đã được mô tả ở bảng Vaccine By Age PHU ở phía dưới) | varchar |
| 5 | **gender** | Giới tính | varchar |
| 6 | **exposure** | Phơi nhiễm   * Outbreak - Bùng phát * Close Contact - Liên hệ chặt chẻ * Not Reported - Không được báo cáo * Travel-Related - Du lịch | varchar |
| 7 | **case\_status** | Trạng thái ca nhiễm   * Recovered - Phục hồi * Deceased - Tử vong * Active - Điều trị tích cực | varchar |

1. **Bảng Ongoing outbreaks PHU**

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Mô tả** | **Kiểu dữ liệu** |
| 1 | **date** | Ngày thống kê số đợt bùng phát dịch | date |
| 2 | **phu\_num** | ID của PHU | int |
| 3 | **outbreak\_group** | Cơ sở bùng phát dịch:   * 1 Congregate Care - Chăm sóc cộng đồng * 2 Congregate Living - Lưu trú cộng đồng * 3 Education - Giáo dục * 4 Workplace - Nơi làm việc * 5 Recreational - Cơ sở giải trí   6 Other/Unknown - Không xác định | varchar |
| 4 | **number\_ongoing\_outbreaks** | Số đợt bùng phát đang diễn ra | int |

1. **Bảng Public Health Units Group**

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Mô tả** | **Kiểu dữ liệu** |
| 1 | **PHU\_Group** | Nhóm PHU | varchar |
| 2 | **PHU\_City** | Thành phố của PHU | varchar |
| 3 | **PHU\_region** | Tên PHU | varchar |

1. **Bảng Public health unit**

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Mô tả** | **Kiểu dữ liệu** |
| 1 | **PHU\_ID** | Định danh của đơn vị chăm sóc y tế cộng đồng | int |
| 2 | **Reporting\_PHU** | Các PHU được báo cáo | varchar |
| 3 | **Reporting\_PHU\_Address** | Địa chỉ PHU được báo cáo | varchar |
| 4 | **Reporting\_PHU\_City** | Thành phố của các PHU được báo cáo. | varchar |
| 5 | **Reporting\_PHU\_Postal\_Code** | Mã bưu điện của PHU được báo cáo | varchar |
| 6 | **Reporting\_PHU\_Website** | Địa chỉ website của PHU được báo cáo | varchar |
| 7 | **Reporting\_PHU\_Latitude** | Vĩ tuyến PHU | float |
| 8 | **Reporting\_PHU\_Longitude** | Kinh tuyến PHU | float |

1. **Bảng Vaccine By Age PHU**

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Mô tả** | **Kiểu dữ liệu** |
| 1 | **Date** | Ngày báo cáo | date |
| 2 | **PHU ID** | Định danh của đơn vị chăm sóc y tế cộng đồng | int |
| 3 | **Agegroup** | Nhóm tuổi, được phân loại gồm:   * 5 to 11 years old * 12 to 17 year olds * 18 to 29 years old * 30 to 39 years old * 40 to 49 years old * 50 to 59 years old * 60 to 69 years old * 70 to 79 years old * 80 years and older * Adults\_18plus * Ontario\_12plus * Ontario\_5plus * Undisclosed\_or\_missing | varchar |
| 4 | **At least one dose\_cumulative** | Số người tiêm được ít nhất 1 mũi | int |
| 5 | **Second\_dose\_cumulative** | Số người tiêm được 1 mũi | int |
| 6 | **fully\_vaccinated\_cumulative** | Số người tiêm đủ vaccin. Tiêm đầy đủ nghĩa là:   * Tiêm 1 mũi Janssen (Johnson & Johnson) * Tiêm 2 mũi trong danh mục vaccin được Bộ y tế Canada phê duyệt * Tiêm 1 mũi trong danh mục được Bộ ý tế phê duyệt + 1   mũi trong danh mục không được phê duyệt   * Tiêm 3 mũi vaccin thuộc loại bất kỳ | int |
| 7 | **third\_dose\_cumulative** | Số người tiêm được 3 mũi (tập con của số người tiêm đủ) | int |

1. **Thiết kế kho dữ liệu (KDL), tổng hợp, nạp dữ liệu các nguồn vào KDL và thiết kế, xây dựng Cube:**