Covid 19 Analysis

1. Miêu tả:

* Hệ thống cập nhật dữ liệu về các ca mắc mới, khỏi bệnh hay tử vong do covid 19 qua từng ngày theo từng khu vực.
* Hệ thống sẽ tự động trích xuất dữ liệu về số ca biến đổi theo ngày và thêm vào data warehouse. Kết quả cuối cùng là một data warehouse với dữ liệu đã được làm sạch và sẵn sàng cho phân tích.
* Data warehouse bao gồm dữ liệu về codvi19 theo khu vực, dân số khu vực và thu nhập bình quân đầu người của khu vực để phục vụ cho việc phân tích dữ liệu.

1. Hình thức nộp:

* Bản nộp sẽ bao gồm:
  + Tài liệu hướng dẫn set up trên AWS.
  + File python cho phép trích xuất, làm sạch dữ liệu.
  + Username và password giúp mentor connect tới redshift.

1. Bản demo được show ra:

* Các biểu đồ trên QuickSight trả lời cho các câu hỏi:
  + Xu hướng mở rộng hay thu hẹp của covid 19 theo một vùng cụ thể.
  + Tương quan giữa số ca mắc của các khu vực.
  + Tương quan giữa số ca mắc với dân số.
  + Tương quan giữa số ca khỏi bệnh và tử vong với thu nhập bình quân đầu người.
  + …..

1. Công nghệ:

* AWS: Lambda, S3, Athena, Glue, Redshift, QuickSight
* Python

1. Tables:

* DIM\_DATE
* DIM\_COUNTRY
* FACT\_COVID19
* FACT\_POPULATION
* FACT\_GDP\_PER\_CAPITA

Diagram

Description automatically generated

1. Kiến trúc

Diagram

Description automatically generated