April 01, 2022

Project 2 Proposal – Team 1

Aaron Schomer, Bandon Cho, Jaceguai de Magalhaes, Jack Nguyen, Newyork Heryo

COVID-19 Analysis

The MVP will have the following features:

* Scala/Spark app CLI for user interactions.
* Users can login into the system using encrypted password.
* Users will have activities limited by their level
* Users will be able to run the queries using the CLI interface.
* System will log activities
* Users’ management through app.
* Use tableau to visualize results and analyze trends.

Queries:

Will be related to the development of the disease for different countries including the US and their states. The list bellow will update incrementally as specifics queries are defined.

1 – Select percentage of population confirmed, dead, and recovered.

2 – Cumulative average confirmed, deaths, and recovers.

3 - When was the peak of mortality rate of the pandemic?

4 - What are the top 10 countries with greatest amount of deaths?

5 - What are the correlation between deaths and population?

6 - What is the average recovered rate by countries?

7 – Confirmed by day, country, and states (when applicable)

8 – Total confirmed, death, and recover

9 – Confirmed spread speed for determined time and country

10 – Death spread speed for determined time and country

Stretches:

* Make a linear regression.
* Use Tableau visualizing Spatial Data.
* Use AWS or Azure.
* Find relations with other kind of Data, i.e., GDP.

General Schema

AWS/Azure

Query Results

COVID -19

Dataset

COVID - 19

Sanitized Dataset

CLI APP

-Manage Users

-Execute Queries

Tableau

(Visualize Results)

SPARK/HIVE/HDFS

(Users and Support Tables)