Objective:
The objective of this project is to automatiacally scrape and process multi-scale spatiotemporal COVID-19 data across the world.

Abstraction:
. Corona virus has been recognized as a global threat across the world and economic disruption were acquire throughout the world.
. With the help of covid tests, most countries publish number of cases frequently.

- Covid-19 data has no standardized format, it is updated by different countries in a unique channel, format and at different intervals.

- The goal of this application is to develop a toolset using cloud based web scrapping methods to store covid-19 cases automatically (in automated manner)/

- The system will publish real time dynamic covid-19 data to the public access.

System requirement:

- This stoftware requires some minimum hardware and software requirements to execute and run efficiently.

- This system requires a processor of i3 and above with 2.60GHZ of processing speed along with a 1GB of ram and 160Gb of hard Drive space. A cache memory of 512 KB is required, also with computer accessories like

keyboard and monitor to operate.

- When it comes to software specifications Microsoft visual studio .net is used as the front-end and SQL server is used as the backend, which is running on top of windows 7 or above operating system.

Level - 1

. Datasource from the cloud is classified as structured and un-structured data using a datasource classifier.

. In case of structured data the data is downloaded and then merged into the database.

In case of unstructured data, the data is cleaned and converted by the system to the required format. And the n cleaned data is passed to a data crawler which scrapes the covid dataset from the cleaned data and then merges it to the database.

. Once we have the datas in the database the system will publish real time dynamic covid-19 data to the public access

Module : Data collection,

Structured:

1.) When the data source is structured and it is available in the required data format it is straight forward to store the dataset in the system.

2.) The app database table contains columns such as date, location and number of cases, etc…

Un-structured:

1.) When the datasource in not in the required format and structure.   It can be manually cleaned and uploaded using the form.

Semi-structured:

- In case a datasource is structured but it's format is different from the required format, the datasource available in one format (eg. CSV, SQL) can be converted into another (eg.Excel) format for uploading the data.