# Problem Statement formation

COVID19 is on surge. Checking daily number of cases and trend helps to predict hospital capacity.

# Criteria for success

Ability to analysis and project daily COVID19 cases. Build continuous data pipeline to maintain data daily.

# Data sources

COVID19 Data - <https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_daily_reports>

JHU provides daily covid19 data in USA

# Scope of solution space

Download CSV files. Write Python script to retrieve data from the CSV files, clean data and load to database.

# Proposed architecture for the solution and rationale

ETL for smaller data sets.

# Choice of technology

Python – To load CSV files

MS SQL Server – To store data

Python - Data validation and cleanup