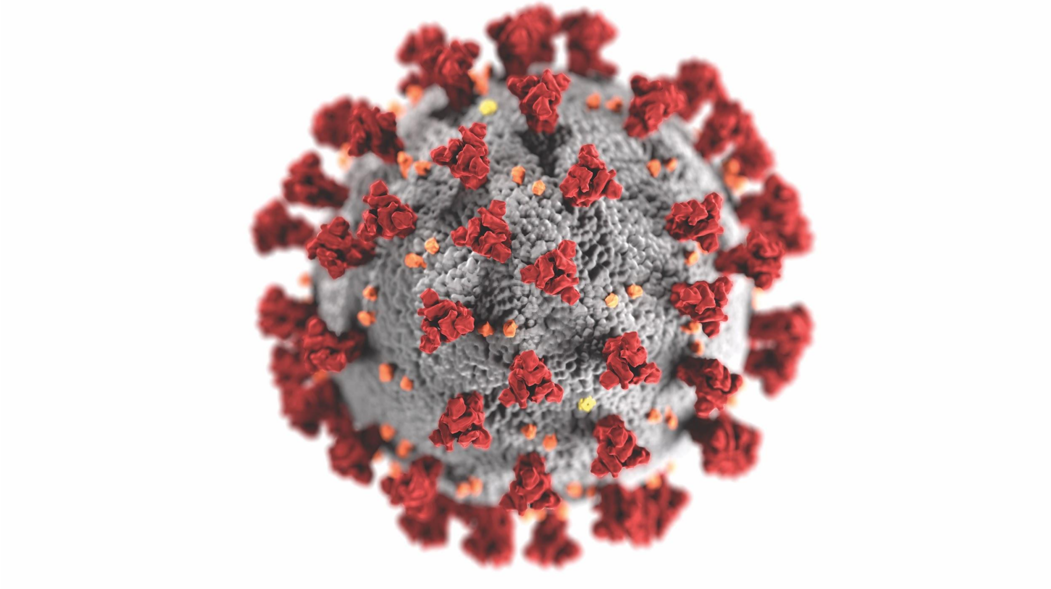
Covid-19 WebMap

Kleanthis Karavangelas
Liam Shamir
Dan Dunkers
Karanveer Singh
Dan Drzewicki

Agenda

1. Introduction
2. Strategy
3. Architecture
4. Stack
5. Demonstration
6. Future work
7. References



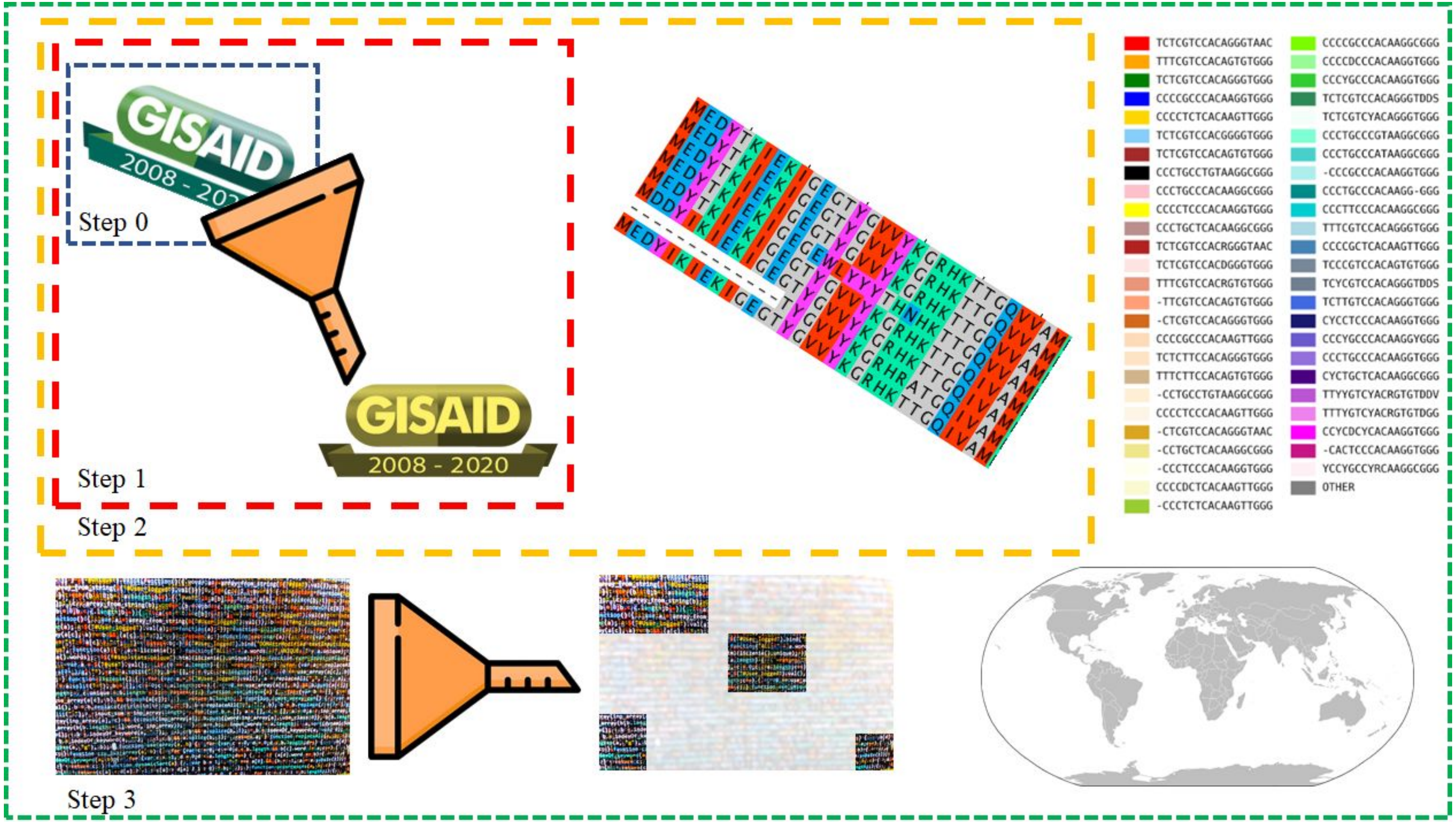
Background Info

- Covid-19 is caused by Sars-Cov-2 and responsible for coronavirus pandemic [1]
- Reported first in Wuhan, China. More than 6.4 million cases reported up until now [2]
- Classic methods of studying virus evolution:
    - Sequence alignment
    - Phylogenetic trees [3]
    - Reference sequences [4]

ISMs- Informative Subtype Markers [5]

- Framework for genetic subtyping
- Allows the subtyping of individual virus genomes
- Generates a signature for easy and efficient tracking of viral evolution through geography and time

ISM Pipeline



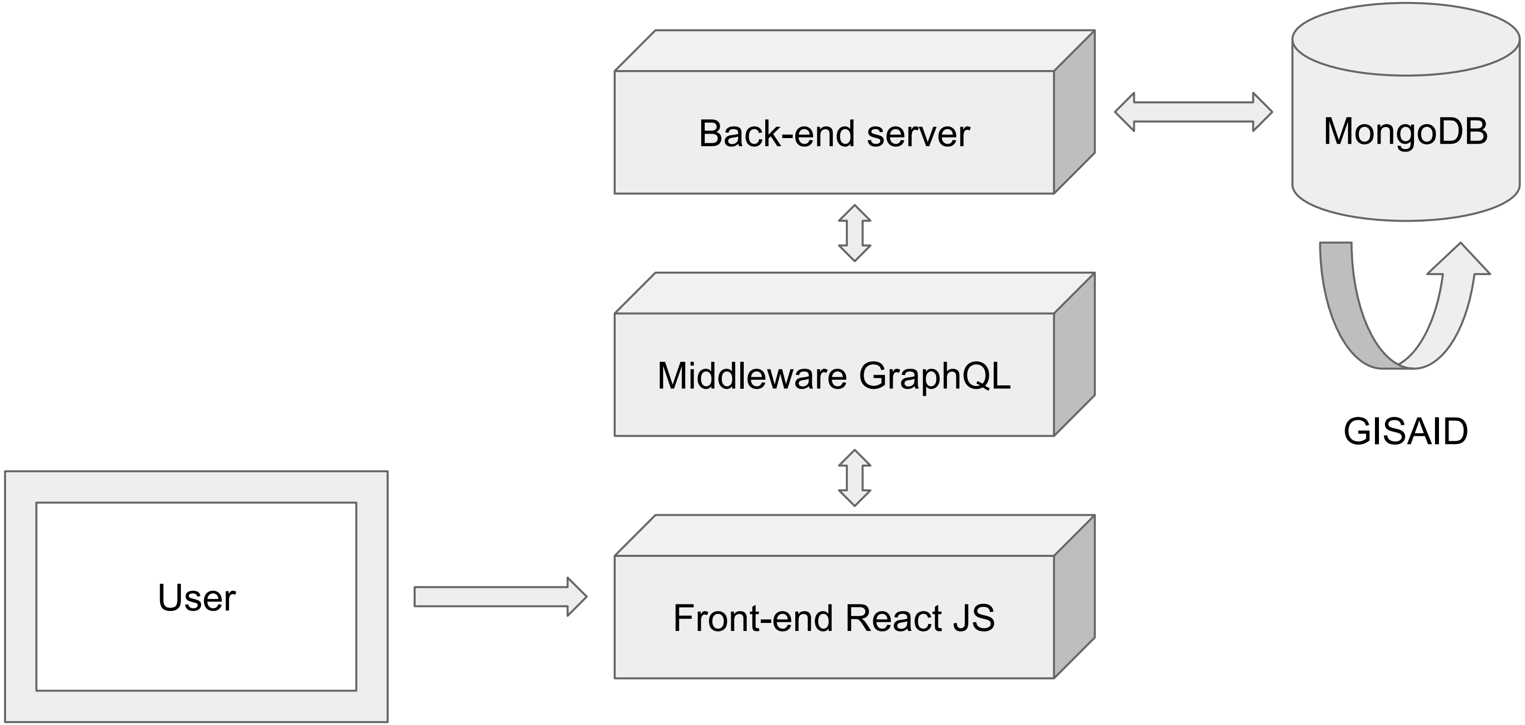
Related Work

- Efforts have been made to track the spread of the virus through a website
- Dong et al. have worked with Johns Hopkins to visualize and track reported cases of COVID-19 [6]
  - However, this study does not look at subtypes
- The study that calculates the ISMs by Zhao et al. displays only static images and graphs
  - Lacks ease of use and interactivity
  - Not as accessible as a website

Strategy/Framework

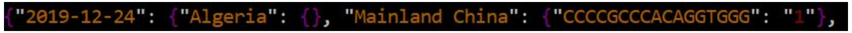
- Database → MongoDB
- Front-end → JavaScript (React)
- Back-end → JavaScript (nodejs)
- Middleware → GraphQL

System Architecture

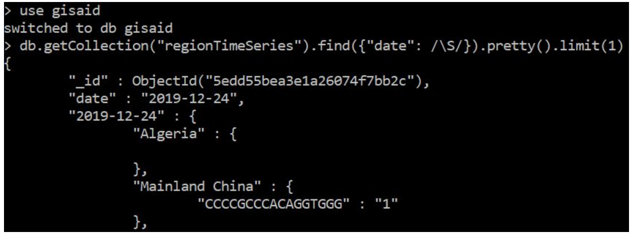


Backend

- Ultimately want to store json data for server-side access
  - region_time_series.json
  - region_pie_chart.json
- When a country is clicked on, graphql gets the data from the DB, processes it, and sends it to the front end
- Huge maintenance advantages
  - Scales with data size as more data becomes available
  - Loading new data is easy -> simply replace json file
  - Data is upserted so only new data needs to be saved
- Run python scripts whenever new json is available
  - Wraps ‘country’ key to pie chart data and ‘date’ key to time series data
  - Inserts into collection in db gisaid using pymongo



```
{"2019-12-24": {"Algeria": {}, "Mainland China": {"CCCCGCCCACAGGTGGG": "1"},
```



```
> use gisaid
switched to db gisaid
> db.getCollection("regionTimeSeries").find({"date": /\S/}).pretty().limit(1)
{
        "_id" : ObjectId("5edd55bea3e1a26074f7bb2c"),
        "date" : "2019-12-24",
        "2019-12-24" : {
                "Algeria" : {

                },
                "Mainland China" : {
                        "CCCCGCCCACAGGTGGG" : "1"
                },
```

Middleware

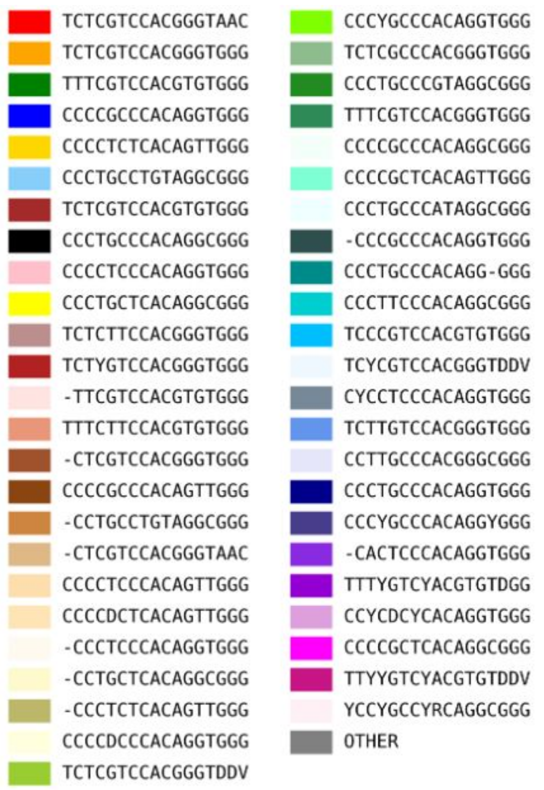
- Libraries
  - Express/Graphql Server
- Allows easy batching of queries on client side
- Provides own test bench in development without use of external programs
- Allows client to dictate to middleware necessary information.

Front-End

- Libraries
  - React
    - Allows for functional components
    - Keeping track of states
  - React simple maps
    - Interactive and customizable maps
  - Chart.js
    - Collection of charts and graphs compatible with React
  - Material UI
    - Collection of aesthetically pleasing React components
- Focus on interactivity and data visualization

Front-End

- Display an interactive graph
    - Time-series data and pie chart for each country
- Color-coding
    - Give each ISM a unique color
    - Graphs and charts with that ISM will use that color
    - Color-coding same as the one on EESI GitHub
    - Allows for easier visual comparisons
    - Countries are colored according to most common ISM



Demo

Future Work

- Hook up the middleware and front-end
  - So that processing can be done server side
- Improve performance
  - Optimize middleware JSON parsing
- Automate ISM pipeline and integrate with web server

References

1. Li, Qun et al. 2020. "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia". The New England Journal of Medicine.

2. Ezez. 2020. "COVID-19 situation reports". Who.int. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/.

3. Hadfield, James et al. 2018. "Nextstrain: real-time tracking of pathogen evolution". Bioinformatics 34 (23): 4121-4123. Oxford University Press (OUP). doi:10.1093/bioinformatics/bty407.

4. Wang, Changtai et al. 2020. "The establishment of reference sequence for SARS-CoV-2 and variation analysis". Journal of Medical Virology 92 (6): 667-674. Wiley. doi:10.1002/jmv.25762.

5. Zhao, Zhengqiao; Sokhansanj, Bahrad A. & Rosen, Gail L. 2020. "Characterizing geographical and temporal dynamics of novel coronavirus SARS-CoV-2 using informative subtype markers". Electrical and Computer Engineering, College of Engineering, Drexel University, Philadelphia, PA, USA.

6. Dong et al.; “An interactive web-based dashboard to track COVID-19 in real time”; The Lancet, Volume 20, Issue 5, May 01, 2020.