# Auto Log

Insight Data Engineering Fellowship, Silicon Valley

Gene Der Su

# Motivation

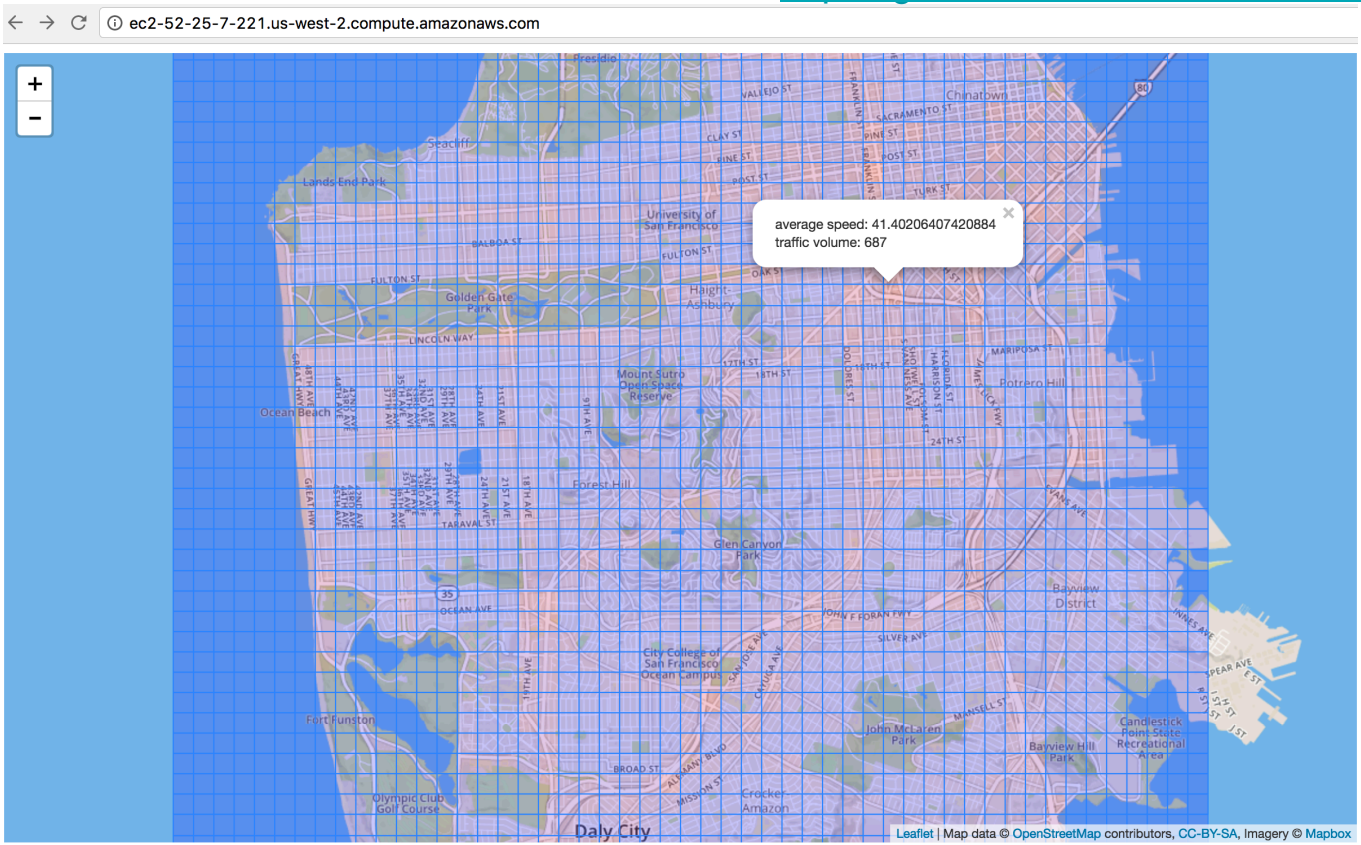
- Traffic in the Bay Area is a headache

- This framework can be used in to avoid traffic dense areas and for companies to re-route their customers/ vehicles

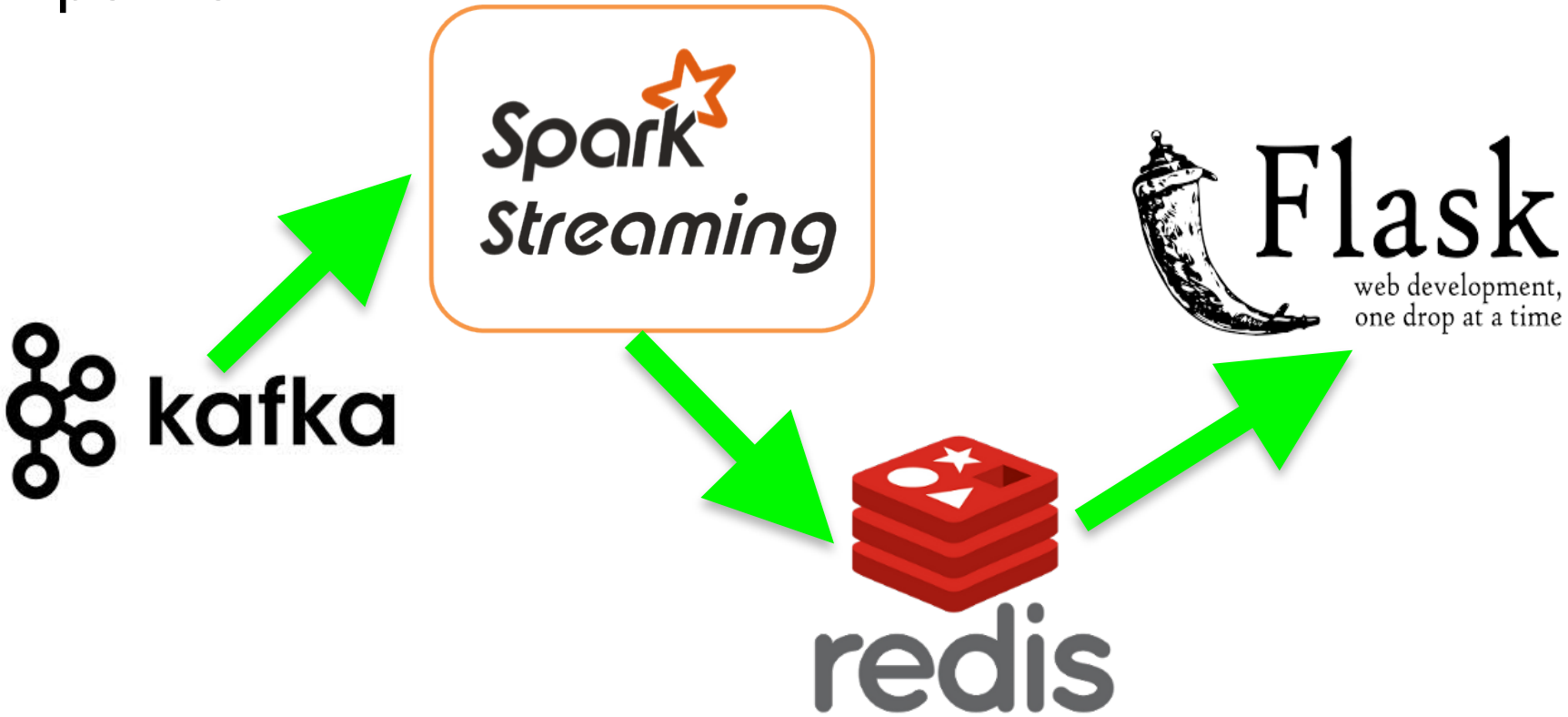
# Product

http://ec2-52-34-86-155.us-west-2.compute.amazonaws.com/

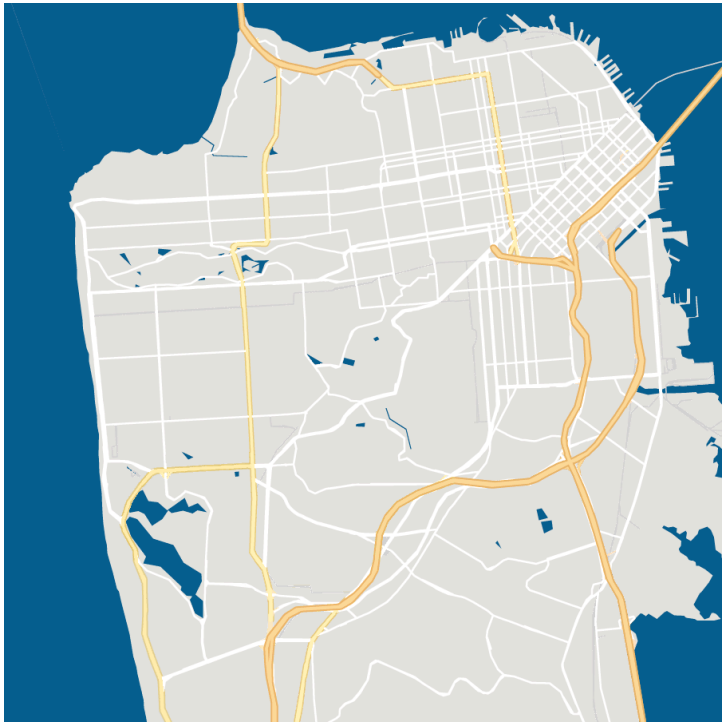
https://github.com/GeneDer/Auto-Log#product



# Pipeline


Spark
Streaming
kafka
Flask
web development,
one drop at a time
redis

# Data



```
37.7574275756;-122.495782221;0;4;10.8982281987
37.7881020756;-122.438275338;1;4;11.2431736567
37.7627444889;-122.510030195;2;4;9.9391131626
37.6981235422;-122.393128143;3;4;16.2580655369
37.80950606;-122.477414351;4;4;8.80097933628
```

- Simulated on the simplified San Francisco map
- (latitude, longitude, car id, simulation iteration, randomly generated speed)

# Queries

- The map is divided into 50 by 50 grids
- For each grid, the car density and the average is calculated and updated with the past 20 seconds data
- Since each grid is independent, all of them can be run in parallel on multiple nodes

# Challenges

- Set up simulator that can mimic traffics in real world
- Compute and update the data in real time

# Other considerations

- Simulated data can behave different than real world
- The simulator can generate **~280MB/minute** on one t2.micro with1 million cars
- The pipeline uses 10 m4.large. It costs around $790.6 per month or **$26.35 per day** in a 30 day month

# About me

Machine Learning Engineer at GoFind.ai

Masters in Computer Science from UC Davis

Bachelor in Applied Math from UC Merced

Love badminton and cycling

# Miscellaneous

- The grid is 790 ft. by 790 ft. square (around 2 city blocks covered by each side)
- Average processing time: 13s @ 2668records/sec
- Data can be filtered by timestamp so data transfer delay won't be a big issue
- Mapping of city block to an id can be done for showing specific traffic volume on each street
- Data skew needs to be tested. stress test will be done
- When will this break?