The Impact of Covid-19 on Air Traffic:

Spatiotemporal/Time Series Forecasting with Deep Learning

Bo Yan, Yuan Hu, & Adelle Driker
(Group 6)

Table of Contents

Team Introductions

Professors Rose Yu & Ilkay Altintas de Callafon

- Advisors/Mentors

Bo Yan

- Record Keeper
- Software/ML/DL Engineer

Yuan Hu

- Budget Manager
- Data Engineer/Solution Architect

Adelle Driker

- Project Coordinator/Manager
- Data/Business Analyst

Table of Contents

Recap - Problem Definition

- Many global industries have been affected by the COVID 19 pandemic, the airline industry being one of the most heavily hit
    - E.g. London's Heathrow Airport reported a 97% decrease in passenger numbers between May 2019 and May 2020
- Creates uncertainty for both passengers and airline companies, especially due to the multiple waves of virus mutations
    - How should airlines plan future flights? When should passengers schedule their travels?

In other words, given a country's COVID situation, how should an airline/passengers plan ahead?

Recap - Data Definition

OpenSky Flight Data (Jan 2019 - Present)

- As-is, new files released on a monthly basis, multiple entries with missing data
- CallSign*, Number, ICAO24, Reg, TypeCode, Origin, Dest, First/Last Seen DT, Lat/Long/Alt of Origin & Dest

Johns Hopkins COVID19 Data (Jan 2020 - Present)

- Updates for historical inaccuracies, new files released daily
- Province/State, Country/Region**, Lat, Long, Dates

Airline Code and Country Mapping

- Sourced from IATA and ICAO, mostly complete
- Airline Name, IATA Designator, 3-Digit Code, ICAO Designator*, Country**
- Will be used to link together the OpenSky Flight and COVID19 datasets

Table of Contents

Target Stakeholders

Who would benefit the most from using this product?

Airline Companies

- Predict future demand
    - Utilize more efficient aircraft
- Optimize future prices

Table of Contents

Project Architecture

Key Features

- AWS based platform
- Orchestrated ETL
- noSQL database
- Multi-platform, open source analytics and interactive visualization web application

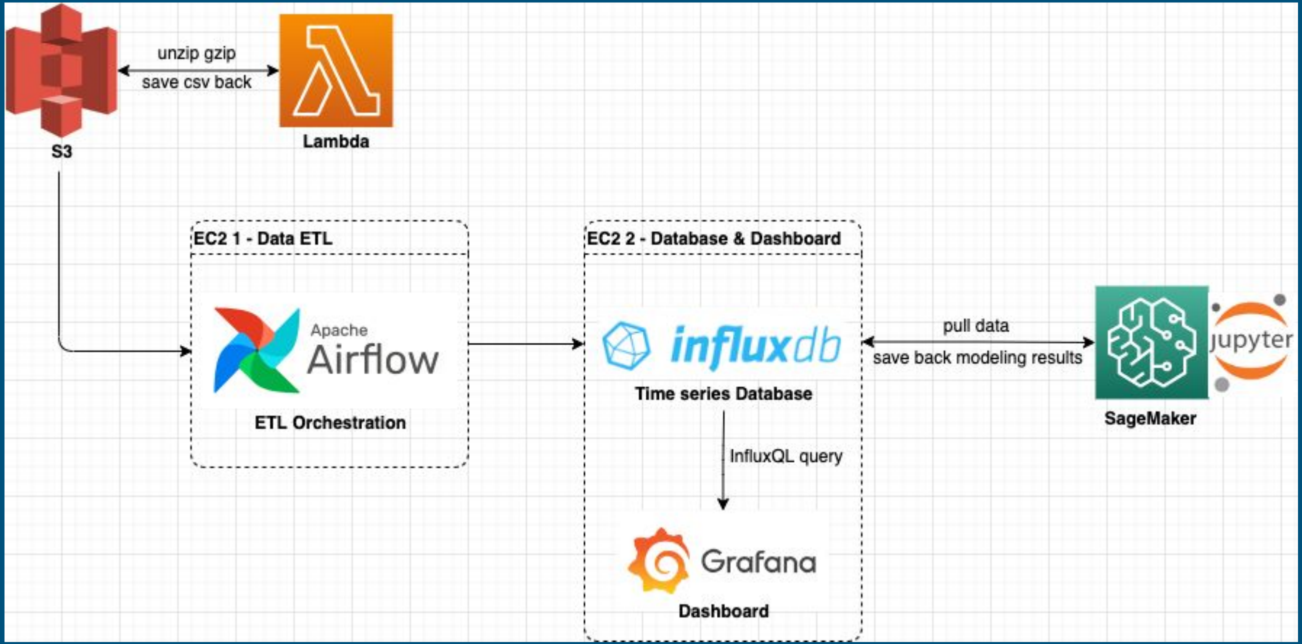


Table of Contents

Scalability - Architecture

Event-triggered S3 lambda function

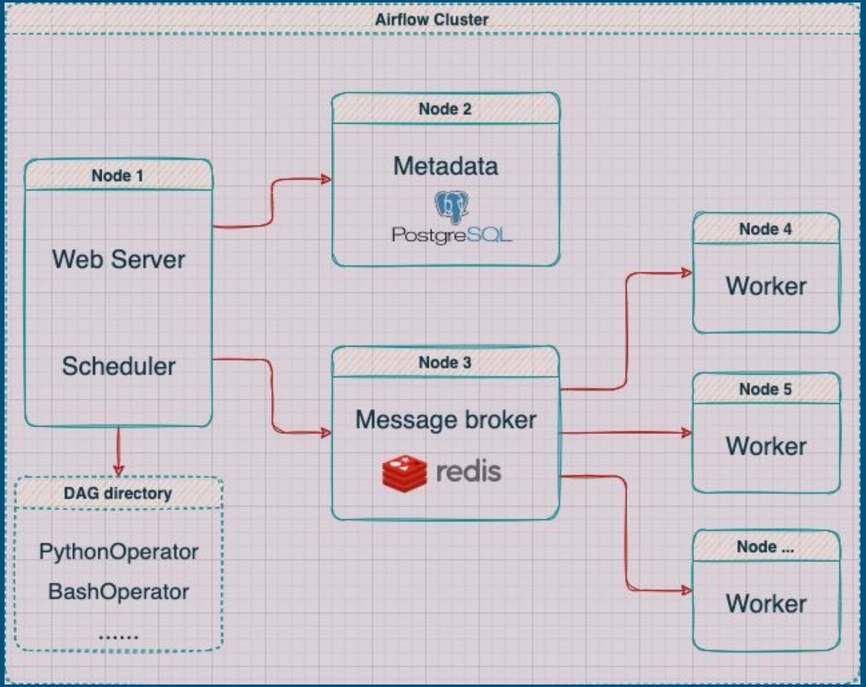
S3: Elastic scalability, flexible data structure

Lambda: Severless, Concurrent executions

Orchestrated ETL by Airflow

Celery executor

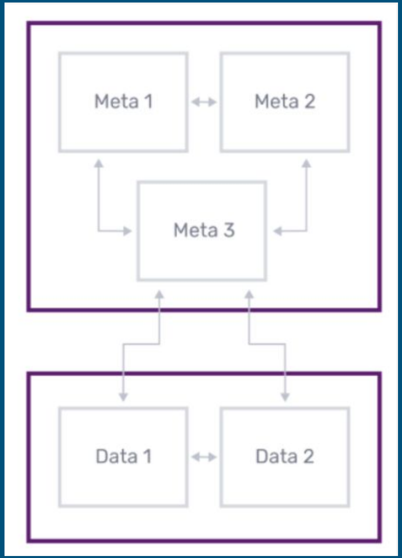
(how the number of workers scaled out)



Scalability - Architecture

Scalable time-series database(InfluxDB)

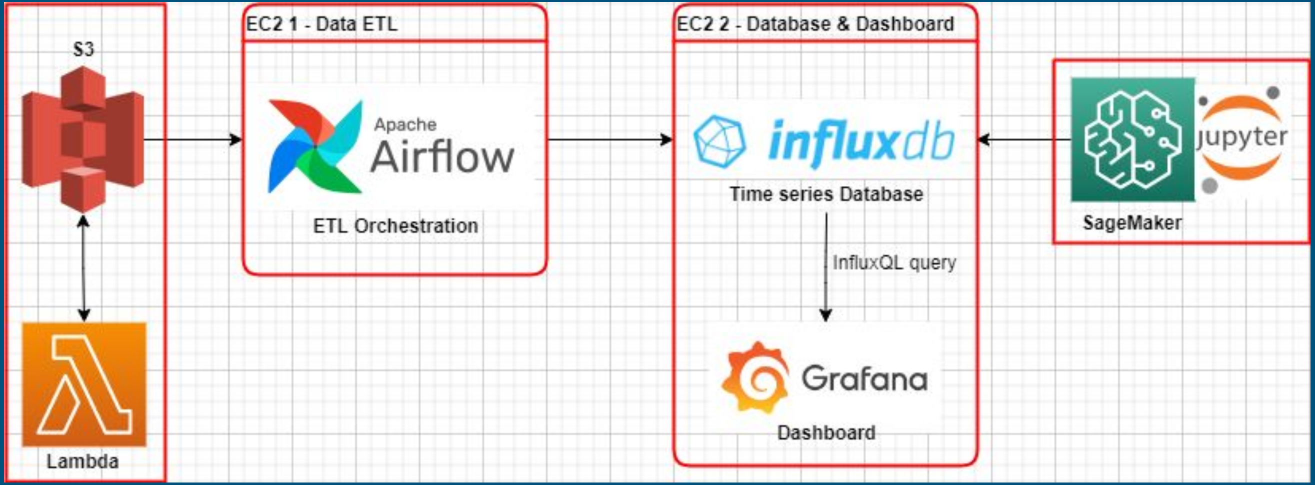
Cluster with meta and data nodes



Auto Scaling feature by SageMaker

Auto scaling brings more instance/removes unnecessary instance based on workload

Resilience to failure



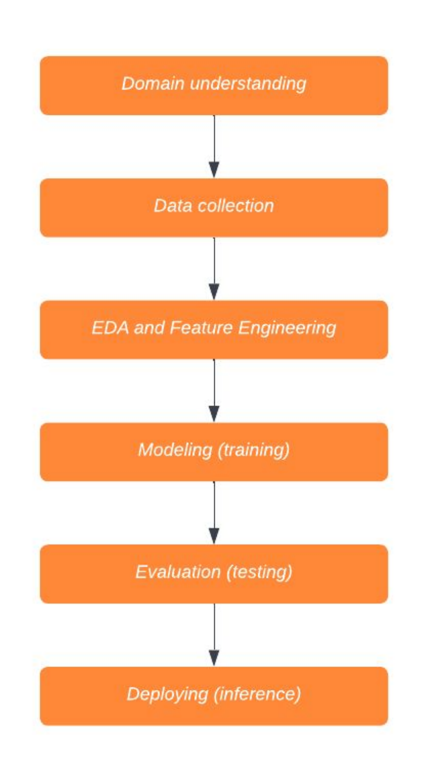
- Less effort on maintaining entire infrastructure
- Easy to reproduce development environment

Scalability - Modeling and Evaluation

Why scalability is important in machine learning/deep learning?

- Training a model can take a long time.
- If a model is too big, it can't fit into the working memory of the training device.
- Vertical scaling is expensive (data lives on a single node, adding more resources to a single node and adding additional CPU, RAM, and DISK to cope with an increasing workload).

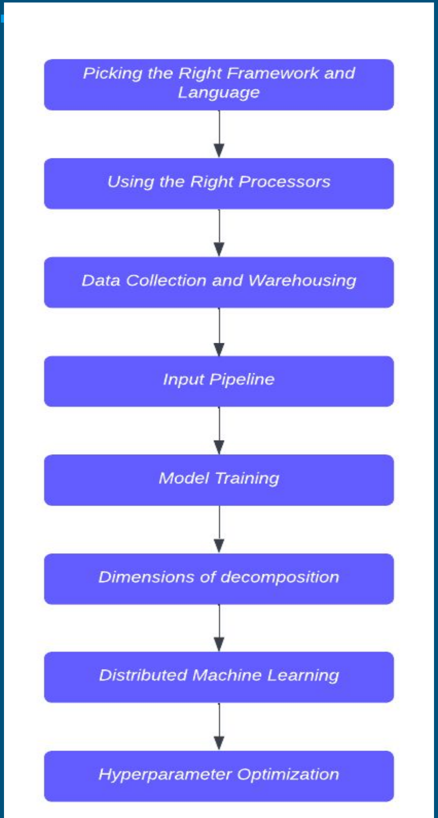
Scalability - Modeling and Evaluation



Opportunities to scale and challenges:

- Data handling: the way data fed to the training algorithm
- Try a bunch of training algorithms and architectures to figure out what fits our use-case the best
- Go back and forth between modeling and evaluation a few times (after tweaking the models) before getting the desired performance for a model

Scalability - Modeling and Evaluation



Solutions:

- Framework: Pytorch and Keras
- Language: Python
- Processor: CPUs (scalar processors)
- Input pipeline: Sagemaker, auto scaling
- Optimizer: Loss function
- Track versions and history of the models

Robustness

System Robustness:

- Amazon s3
  - Easily manage data at any scale with robust access controls
  - Backup and restore critical data with robust application feature
- Amazon EC2
  - Load balancers: distributes network or application traffic

Table of Contents

Data Product Preview - Grafana Dashboard

Table of Contents

Next Steps

1. Airflow DAGs still need further implementation
2. More features/variety of visualizations on dashboard
3. Incorporate country-level forecasts into the dashboard
4. Continue to optimize/tune models
5. Conduct end-to-end testing to ensure the final product is production-ready

Thank You

Q&A