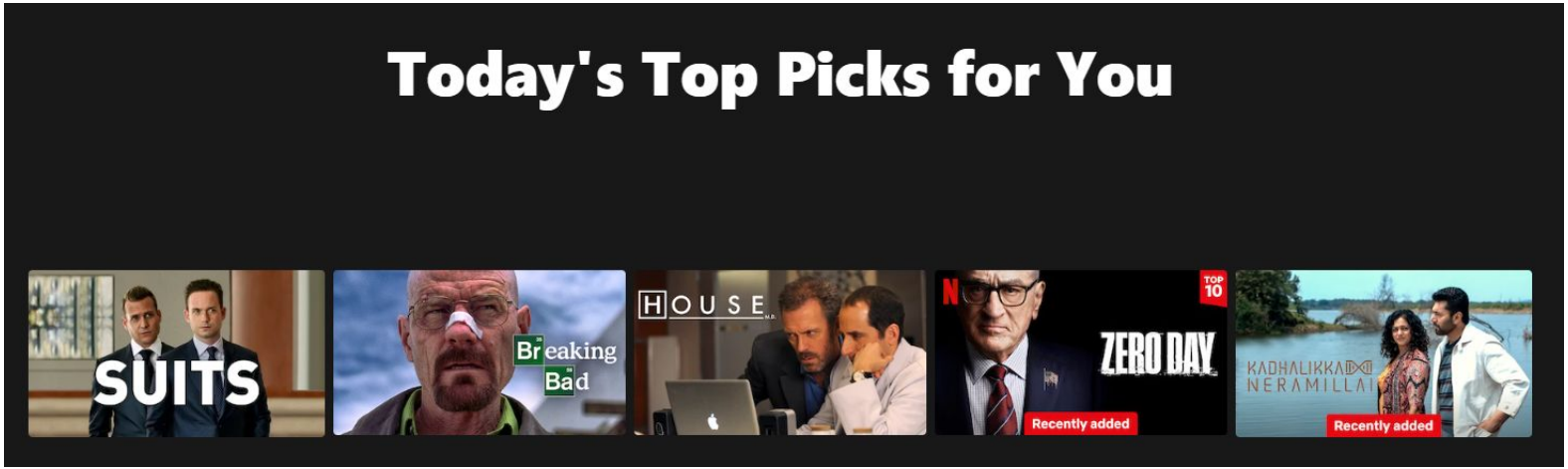
AI Society

Week 6: AI Student

[February 28, 2025]

Big Data Management And Deployments

Scenario – You are part of MLOps team at Flixx, a streaming service, are assigned with a task to send recommendations to millions of users every friday at 6PM.



Questions

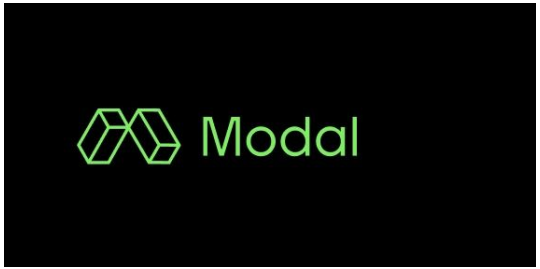
1. How to manage large datasets?
2. How to schedule deployments?
3. Do it one by one or all at once?

How to manage?

But…

https://treinetic.com/cloud-service-providers/

We will be using Modal



Why Modal?

1. Easy to set up.
2. Straightforward APIs
3. Free credits!

How to schedule deployments?

Cron

https://crontab.guru/

Offline vs Online Processing

Offline - Non real time. Like analyzing last month's sales data or sending recommendations to users every week.

Online - Real time. Like processing credit card transactions or responding to user clicks on a website.

Setting up Modal

Install just like a Python package

```
python3 -m pip install modal
```

Login using the following command

```
python3 -m modal setup
```

And we are ready to go…

Create Functional Blocks for Deployment

```
# App to deploy the code to
app = modal.App.lookup(name = 'recommender', create_if_missing = True)
# Storage volume to store and save data
volume = modal.Volume.from_name('recommender', create_if_missing = True)
# OS Image to run the code on and to install packages
image = modal.Image.debian_slim(python_version = '3.9').pip_install(
  'pandas',
  'torch',
  'sentence-transformers'
)
```

Upload data to Storage Volume

```
# Upload data to the volume
def upload_data():
    with volume.batch_upload() as batch:
        batch.put_file("movies_embeds.pkl", "/data/movies_embeds.pkl")
        batch.put_file("infer_dataset.csv", "/data/infer_dataset.csv")
```

Linking our function to the App

Running the App

```
if __name__ == "__main__":
    upload_data()
    modal.runner.deploy_app(app)
```