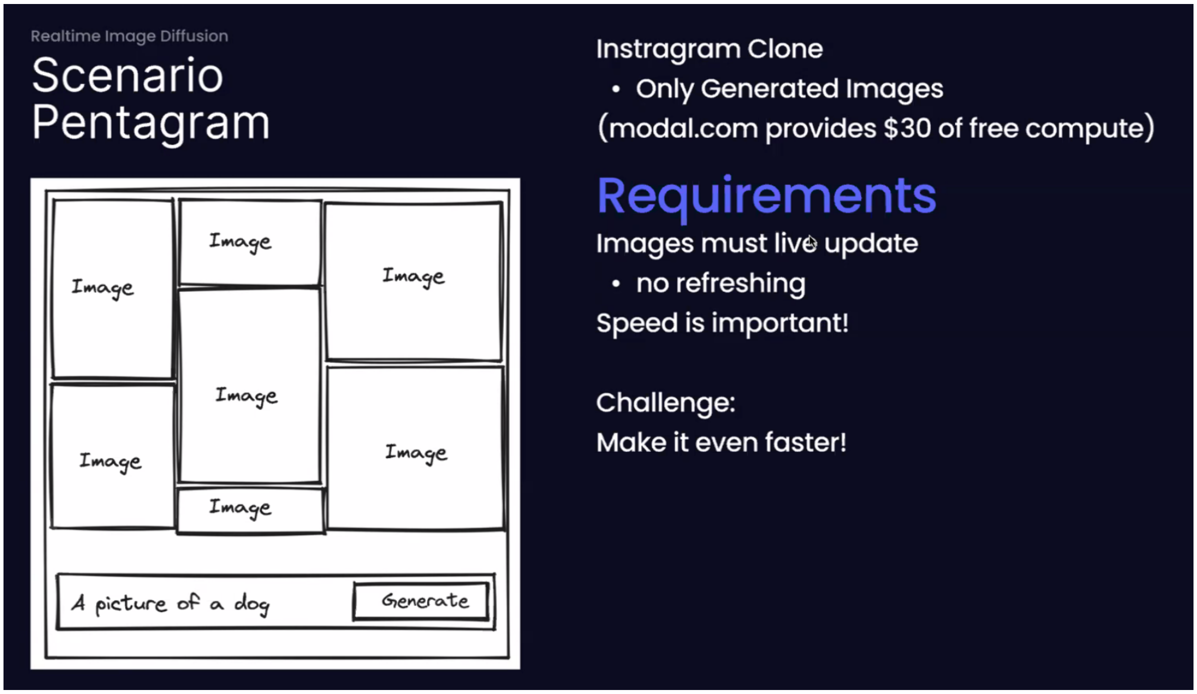
pentagram - instagram clone

12/16/24 monday

- ▼ slides



Realtime Image Diffusion

Scenario Pentagram

Instragram Clone

- Only Generated Images

(modal.com provides $30 of free compute)

Requirements

Images must live update

- no refreshing

Speed is important!

Challenge:

Make it even faster!

- host my own modal
- write my own api around my own diffusion modal
- host it on gpu on modal
- (server-less infrastructure): odds are my app is not always running constantly
  - i can spin my server down to 0, and when a user comes, i can turn the server on
  - i get a lower bill, but the user gets more latency - the user has to wait longer
  - lets ramp up and down amount of servers you have based on amount of users you have

- hassan built an app that got 1M users, what did the infra look like to handle that many users
    - used railway provider (10 servers that were always running)
    - if he could do it again, he would run it serverless on modal
- don’t have to train or fine tune a model
- just purely model inference
- save files in s3 and not a database
- blob storage: https://www.cloudflare.com/developer-platform/products/r2/
-