In this homework we will build on top of the work from Homework 4

1. I will assume that you have successfully built the 2 apps required in HW 4. If not please reach out to me or the TAs for code on which to build.
2. You should now create two VMs running your web server, and ensure they are in different zones (in the same region) and place them behind a load balancer. I recommend using a network load balancer as it is easier to use. Make sure you configure health checks and other properties appropriately.
3. Modify your web server code to return the name of the zone the server is running in as a response header.
4. Modify the given client to extract and print the new response header your are returning in step #3.
5. Kill the web server in one of your VMs and report how quickly the load balancer notices its demise and routes the requests to the remaining healthy instance by monitoring any errors seen by your client (it should be in the order of seconds and definitely less than a minute). If not, stop your client to save on resource usage and use curl to debug your setup.
6. Once the load balancer has rerouted traffic to your one healthy VM, restart the web server on the other one and report how quickly the load balancer notices its presence and starts routing requests to it by monitoring the responses your client is getting (it should be in the order of seconds and definitely less than a minute). If not, stop your client to save on resource usage and use curl to debug your setup.

What to turn in:

- The python code for your modified server and client as a github link
- A pdf file describing all the necessary steps to configure and run your apps, your failover timing measurements, and the ratio of requests served by each backend VM.