INCUBIG

BACKEND DEVELOPER ASSIGNMENT

How to Solve:

At Incubig, you will be expected to follow the following framework for any problem you will encounter. Hence, you are expected to follow the following steps for understated problems

A. Identify steps (system design) to solve the (pseudo code on whiteboard/paper)
B. Identify potential bottlenecks (computationally expensive pieces of code) and way to mitigate those (if possible)
C. Web search for optimization (list your finding)
D. Actual code

Problem I:

Access a zipped file from cloud, manipulate the contents and store the manipulated version back on cloud.

Details and steps:

1. **Download:** Given a file (zipped) on cloud (AWS S3/Gcloud Bucket), download it in memory and unzip it (Calling origin as Bucket A)
2. **Unzipping:** It's an concatenated XML file (i.e.. XML blocks joined together each having <Doctype> tag
3. **Extraction:** Extract individual XMLs
4. **Manipulation:** Extract contents of 5 tags (dummy tag names – alpha, beta, gamma, delta, theta), please note that these tags might be nested inside some other tags
5. **Storage:** Save the result from Step 3 in another Bucket (Bucket B) and results from Step 4 in Bucket C

Sample Zipped File: Request if needed

Problem II :

Given a list of phrases (List A), build a classifier which takes a new phrase as input (eg. Data science) and finds most similar item in List A

Sample List: [Click here](#)

Problem III (Optional):

Given the documents given created in Problem I, Step 3 of above problems, create a LSI (latent semantic index)

P.S. – It's an open problem (feel free to use any of the pre build python libraries)