

ARMO cloud software engineering candidate exercise

Language: Python/Go

Level: experienced

Estimated time: 1d

Name: the docker image crawler

Problem definition

We find ourselves many times in the need of information and statistics about docker container images out in the “wild”. In order to get to this information, we need to go over a huge number of docker container images and look inside them.

To accomplish this, we need to do some sort of enumeration of all publicly available docker images. Then, using this enumeration we should pass the names of images to a worker who will do the actual data mining in a single docker image

Exercise

Write a command line python application

Inputs

- host name of a docker registry to “crawl”
- name of worker python app to run on each docker image name (workers)

Output

- Call the worker app on each image name you find in the docker registry

Remarks

- Don't run more worker apps in parallel than the number of CPU cores you have because it is ineffective (let's say you have 4 cores, the 5th worker should not start while all 4 are running)
- Only enumerate docker images who are tagged. Docker API enables you the get of list of image tags.
- The python application does not need to be production grade code, but it should run a simple case of your choice
- It should show you concept of designing system and code
- You are free to use python or Go libraries