Backend Engineer | Programming Task

Overview: This programming challenge consists of two tasks. First task requires you to crawl the data from a credit web source and parse the crawled data. In the second part, you will have to develop a backend engine that will expose a web API. These APIs can be used by client to search the data based on a specific criteria.

Please note that you are required to use Python version 3.x.x throughout this task.
Choice of other supporting frameworks and libraries is up to you.

Submission: You are required to submit this task using github. Please create a profile on github If you don't already have one. Upload all the code, graphs and results to a github repository and share the link to repository in the email.

You will have **one week** to submit this assignment.

Scoring: Primary scoring will be done on the correctness of your code and output. However, We will be considering several other factors for evaluation e.g.

- Code readability
- Performance
- Documentation
- Error handling
- Test cases
- Evaluation/accuracy scores.

You are encouraged to improvise on the above list, as this is only for the hint.

Task-01: The website you need to crawl is the Singapore Marine company listing. The website link is: https://www.sgmaritime.com/company-listings . This site contains company and contact information. This task has two parts: 1) You need to collect the basic raw data available in the link. Bonus points will be given if you collect more. The desired output format should be

[
{
"company name": "ASP SHIP MANAGEMENT GROUP", "url": "https://www.sgmaritime.com/companies/asp-ship-management-group ", "crawled_on": "2019-02-28",...},

{.....}

, ...]

Save this output to file named **company_index.json**.

2) In the second part you need to fetch detailed company profile based on the links you crawled in part 1. The detailed company profile should include:

- company name - company address - company description (business summary)

- company email - company phone number - company website

- contact person -product and services - categories

The output format should be like

[

{

"company name": "ASP SHIP MANAGEMENT GROUP", "company_url": "https://www.sgmaritime.com/companies/asp-ship-management-group", "company street address": "...", "country": "Singapore" (default), "company description":.. "category", "company phone number": [(84.75) 860 265,(84.75)860 346], "business":"Agriculture", "company website":"www.aspships.com ", "company email":"", contacts: [{'job_title':'Sales','name':Capt Robert Walker, 'email:'xxxxx'}, {'job_title':Commissioner,'name':XXX,'email:'xxxxx'}

},

{.....}

]

Save the output in **company_profiles.json**.

Task 2: This task involves building a backend web API. You are free to choose any web Framework e.g. *Flask, Django, Bottle, Pyramid* etc. Please use the crawled datafile **company_profiles.json.** For the first step of this task you should load this Data in the database of your choice *MySQL, PostgreSQL, MongoDb, SQLite* etc. Then using the web

framework, provide an API for the consumer to search company information. The consumer should also be able to apply filters. The API endpoint should support following operations:

1) Fetch list of companies: The endpoint should be able to fetch list of all companies without applying any filters. *Endpoint:* /companies *Output:* {

"status_code": 200,
"message": "successful",

"data":[{

"id": 1, "company_url": "https://www.sgmaritime.com/companies/asp-ship-management-group", "company_name": "...", "company_email": "", "company_website":"", "company street address": "", "country": "Indonesia", "company description": , "company phone number": [(84.75) 860 265,(84.75)860 346], "industry":"",.. },

...] }

2) Fetch specific company: This endpoint can fetch a specific company given its *name* *Endpoint:* companies?company_name=ben%20tre%20aquaproduct (Fetch company where name='ASP SHIP MANAGEMENT GROUP ')

Output
: {

"status_code": 200,
"message": "successful",

"data":[{

"id": 1, "company_url": "https://www.sgmaritime.com/companies/asp-ship-management-group", "company_name": "...", "company_email": "", "company_website":"", "company street address": "", "country": "Indonesia", "company description": , "company phone number": [(84.75) 860 265,(84.75)860 346], "industry":"",.. },

, }] }

Please provide answer to following questions:

1. Which database engine you choose and why?

2. Which web framework you choose and why?

3. Briefly describe the architecture of your application?