BalkanID

Internship Recruitment Task

Submission Instructions

You need to submit your assessment by **24th April 2023, 4:30 am IST.** No further extension will be given.

Join the GitHub classroom first using the link below:
https://classroom.github.com/a/U95dUAR4

This will create a repository for you. Kindly make all commits and code pushes to this repository only. **No other repository submissions** will be considered.

After completing your task, fill your submission details on this Google Form:
https://forms.gle/d6e8rgKQMU7GUscx5

Problem Statement

You are tasked with creating a program that fetches data from the **Github API** using OAuth authentication and stores it in a Postgres database. The data should be normalized and deduplicated before being stored in the database. The program should then use the Github API to retrieve an access token, which it can use to fetch data from the API.

The program should be able to fetch repository data dynamically along with its owner information. The program should be able to fetch both public and private repositories of a user.

The program should also be able to handle errors with proper error handling and retries in case of network failures or other issues. It should log all errors and retries, along with the relevant data, to aid in debugging.

The program should **normalize the data** before storing it in the database. Normalize the tables as you feel necessary. Before storing the data, the program should check for duplicates using

the Repo ID and Owner ID fields. If a duplicate is found, the program should update the existing record instead of creating a new one.

Finally, the program should map the data to the CSV format and write it to a file. The CSV fields should be:

- Owner ID
- Owner Name
- Owner Email (Empty if null)
- Repo ID
- Repo Name
- Status (Public or Private)
- Stars Count

A user **should be able to download this CSV** by hitting an endpoint.

Guidelines

- You can choose to write your program in **Python or Go only**. It is crucial to ensure that your code is **well organized and easy to understand**, and that you provide clear instructions on how to run your program.

- All plagiarized submissions will be disqualified. Please **ensure that you use a VCS Platform like Github** and commit and push all your contributions on time. Kindly share the same.

- The Github repository must have a file called "**README.md**" which contains information about how to install and run your project, along with a clear understanding of your project, including relevant diagrams, if any. If you have deployed your application, you may include a link in the README.

- The task judging metrics will be based on the following points:
    - **Correctness:** Does the code correctly implement all the features desired in the problem statement?
    - **Code Quality:** Does the code follow best practices to ensure that it is clean, readable, maintainable, and comments are provided wherever necessary?
    - **Efficiency:** Is the code efficient (i.e, it performs operations in a sufficiently performant manner with less use of resources)?
    - **Secure:** Does the code or the public repository have any security issues/vulnerabilities?

Bonus Points:

- You can use Docker and containerize your application code to run.
- Your application can also fetch data from Organizations that you are a part of.