Coding Exercise

The Eclipse Foundation is composed of many open-source projects that are also hosted on GitHub. Part of our job is to make sure that these projects apply a secure software development lifecycle process and follow best practices wrt to protecting their repositories.

One aspect that recently got a lot of traction is the ability to enable "secret scanning" for a repository on GitHub where the contents of a repository are continuously checked by GitHub and partner companies if it might contain secrets that might have been accidentally pushed to the repository. The maintainers of these repositories are then warned and can take appropriate actions.

As an exercise, you should write a tool that is able to retrieve information for all repositories of a GitHub organisation (e.g. https://github.com/eclipse-cbi) whether secret_scanning is enabled for that repo or not. GitHub provides a REST Api to query various information about the repositories hosted, see more information at:
https://docs.github.com/en/rest?apiVersion=2022-11-28

Requirements:

- write a tool (using language / environment / framework of your choice) that is able to collect information for all repositories of a GitHub organisation if secret_scanning is enabled for them
- the GitHub organisation should be an input parameter to the tool
- the output could be a json file containing all repository names and the status of the secret_scanning feature for each of them

Hints:

- accessing the GitHub API can be done without any token, but the number of requests will be limited to 60 / hour
- you can generate a token to access the GitHub API from your GitHub account which allows you to do 5000 requests / hour
- the REST API supports conditional requests that are not counted towards the rate limit if data is unchanged

The idea of the exercise is to come up with a clean and working solution within 2h and be able to discuss various extensions to the tool:

- what would you do when using the tool to extract data for lots of organisations without hitting the rate limit
- how would you good runtime performance when extracting data for lots of organisations
- how would you schedule such a tool to monitor a set of organisations on regular basis