

Intro Data Engineer Exercise

This exercise is written in increasing order of difficulty, with each stage more complex than the last.

The expectation is not that you complete everything – the exercise is to generate discussion about what you learned and obstacles you overcame when solving problems. Depending on your familiarity with python, some of these steps may be easier than others.

As a Data Engineer, a lot of what we do is delivering data to the appropriate endpoint in the most efficient way possible. This exercise is intended to offer a flavour of that.

We do not ban the use of Chat GPT or Github Copilot to complete this assignment – however, we would strongly encourage you to complete the tasks in order – and not leap to the end. The conversation in your interview will also be centered around how you found the task, and why certain steps were taken. Generative AI is a great tool for coding, in the same way that StackOverflow is – but the conversation will be centered around your general approach to problem solving.

This exercise uses some of the Pokemon api. The documentation for it can be found here:

<https://pokeapi.co/docs/v2>

1. Use the requests library in python to call 'https://pokeapi.co/api/v2/pokemon/clefairy' and print the JSON response to console
2. Convert the response to a dictionary, and select only the id, name, height and weight of the pokemon. Create a dictionary containing only this information and print this to the console
3. Instead of having the code as one script, write a function to take the name of a pokemon as an argument, and return the appropriate dictionary for that pokemon as from step 2. You might at this point want to make the printed output of your program a little prettier.
4. Add the functionality to your code to accept arguments from the command line. The input to the command line should look something like 'python poke.py pikachu' and the returned value should be the dictionary for that pokemon.
5. We don't want to call the api more times than is necessary. Add functionality to your program to write the information we retrieve from the api to a table within a SQLite3 database (only the id, name, height and weight). When asked for a pokemon, your code should first check if the data is in the database, and read that instead of the api if it is.

Documentation for the SQLite3 library for python is here:

<https://docs.python.org/3/library/sqlite3.html>

Other things to consider:

- At the moment this code will probably throw an error if asked for a pokemon that doesn't exist, you might consider handling this error appropriately
- Your code may consider "Pikachu" and "pikachu" to be two separate pokemon. Can you handle that?

Sharing your work:

We recommend using your GitHub profile or creating a GitHub profile for this task as this is the easiest and most secure way for us to view your work. Please send a link to the recruiter at Oakland up to 24 hours before the interview.