

KIND

Objective:

Populate a small database on AWS with Python & analyse the data:

1. Using a free tier AWS account, host and launch a small postgresql database in the cloud to store your project data.
2. Using python frameworks such as requests and beautiful soup, build a small app that can scrape news articles from a news website (CNN, ABC, etc) on a variety of topics. Take care with this python code and try to design the small app in a way that is easy to maintain and is error-tolerant. The code for the app should be hosted on github. Please do not push any secrets to the repo.
3. Scrape >200 news articles with your app and store this data in a new table created in your new AWS database, take care with the schema of the table.
4. In a jupyter/colab notebook query the news data in your database. You can use sqlalchemy and psycopg2 libraries to connect to your database. Run sql aggregate queries on your news data and display the results in the notebook. Interpret and analyse any results that look interesting to you.
5. Send us your github repo, credentials for your small database and your python notebook so that we can inspect your python code, database schema & SQL queries.

What we are looking for:

Much of the work we do here involves understanding a problem and working with raw data to create a solution, so this task is designed to let you show how you approach and solve problems.

Specifically, the skills we are most interested in are

- **Clean, efficient and maintainable python coding**
- **Ability to correctly wrangle and mutate raw data**
- **Approach to solving problems**
- **Ability to work with a tech stack to solve a problem (SQL->python->aws->linux)**

- **Understanding of data types & data structures**
- **Communication skills**

Setting up:

You will need an AWS account and a free tier EC2 server running Linux. If you don't have an AWS account you can create one for free via their website.

A suggested approach would be to:

1. Install a new postgres database into a T3.micro EC2 instance
2. Design and code a small scraping python app in vs code or similar
3. Use the app to scrape 200+ news articles and inspect the data. More data fields (meta data, sentiment, etc) is better and more impressive.
4. Design the database table schema
5. Create the new database table and insert the news data
6. Analyse the news data in a jupyter notebook
7. Summarise and send details back to us so we can read your code, sql & db schema.