

CCC Analytics Data Engineering Take Home Assignment.

We would like to see how you are building a simple data pipeline. We want you to use following resources during this assignment in addition to Python.

News API

News API provides A simple REST API that can be used to retrieve breaking headlines and search for articles. A free News API account is required to obtain an API key.

Below is a list of News API endpoints.

Endpoint	Description
/v2/everything	search every article published by over 75,000 different sources large and small in the last 3 years.
/v2/top-headlines	Returns breaking news headlines for countries, categories, and singular publishers.
/v2/sources	Returns information (including name, description, and category) about the most notable sources available for searching.

You can find more information here: <https://newsapi.org/docs>

AWS S3

AWS S3 is a managed cloud storage service run by Amazon Web Service. You can create a free tier AWS account for a limited time that can be used for this assignment.

Use [boto3](#) Python SDK to interact with S3 from your python application.

Assignment:

Build a Python application to:

- First, retrieve all English language news sources. Use the language filter in sources end- point.
- Second, create a flattened version of top headlines for each of English language news source you have retrieved above and save them into a CSV file. CSV file should have *[timestamp]_headlines.csv* naming convention. *timestamp* is current system timestamp.
- Finally, upload all CSV files a S3 bucket.
[s3_bucket]/[news_source_name]/ [timestamp]_headlines.csv

Please share your application/code with us as a compressed file or as a git repository (in Github, Gitlab, or Bitbucket). No matter how far you go into completing this assignment. We would like to see up to the level you complete within 5 to7 days.