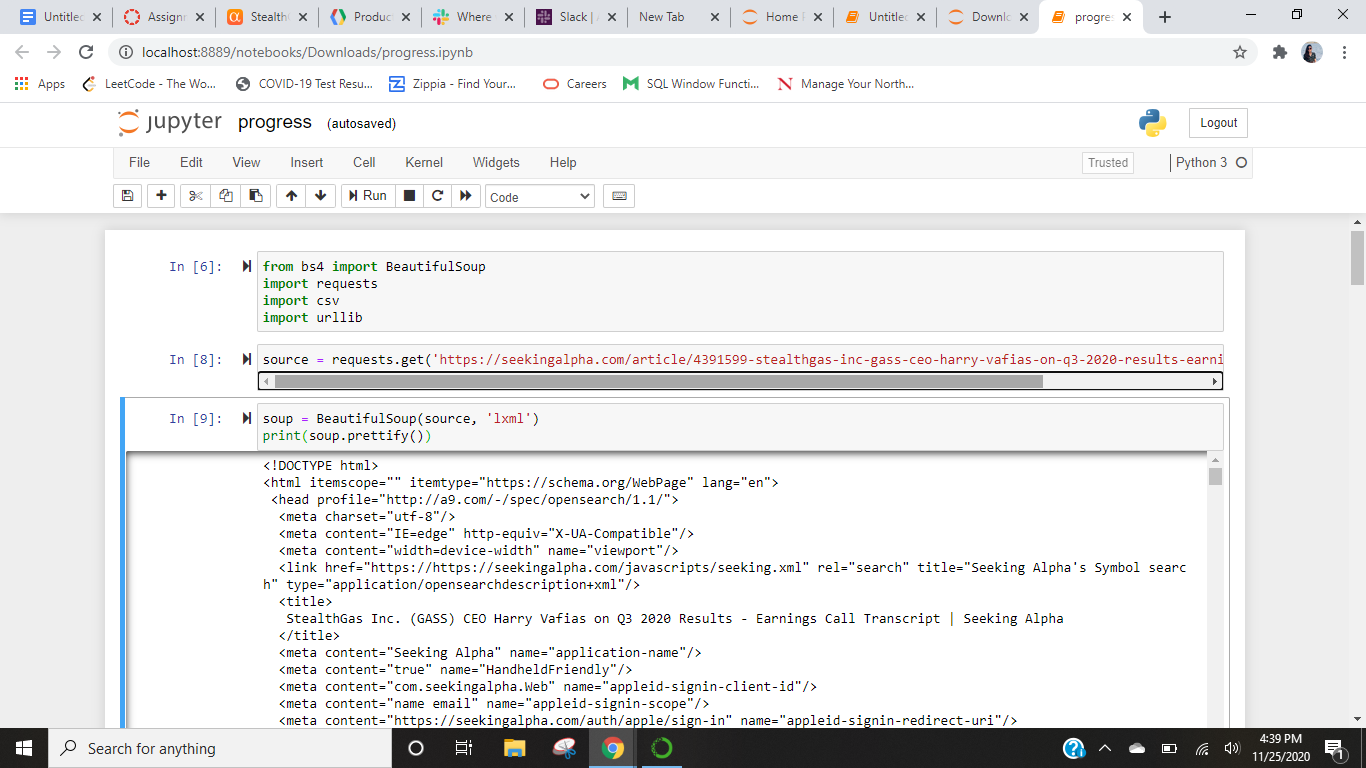
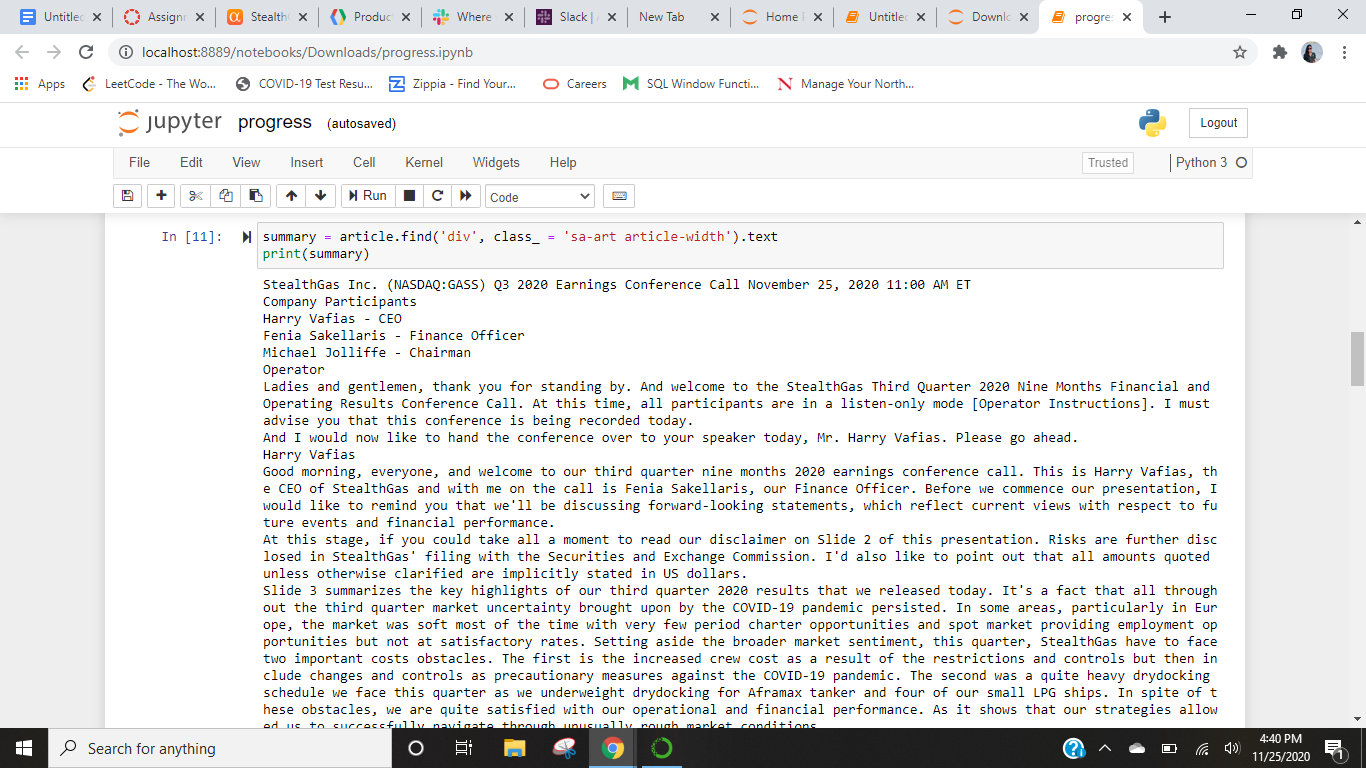
Productionalizing the Pipeline

# Scraping:

Raw data loaded from the scrapping site:



As we are going to use the transcript data which is present within the paragraph. So we scrapped only that data.



# Named entity recognition:

* It is also called entity identification or entity extraction ‒ is an AI technique that automatically identifies named entities in a text and classifies them into predefined categories.
* Entities can be names of people, organizations, locations, times, quantities, monetary values, percentages, and more.
* It also helps to easily analyze huge amounts of unstructured data.

**How Does Named Entity Recognition Work?**

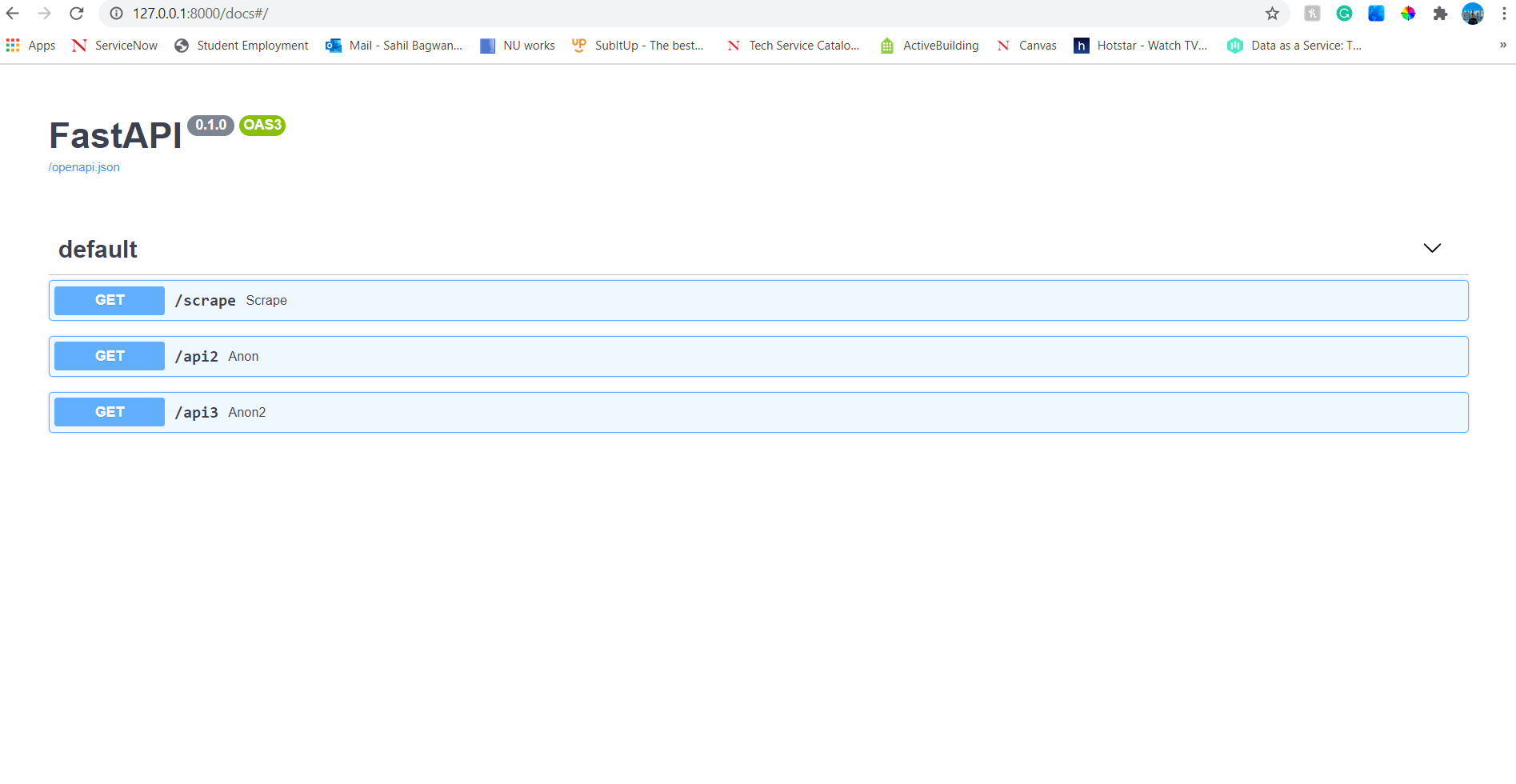
* This is done through machine learning and Natural Language Processing(NLP).
* NLP studies the structure and rules of language and creates intelligent systems capable of deriving meaning from text and speech, while machine learning helps machines learn and improve over time.

# Implement masking, and anonymization functions:

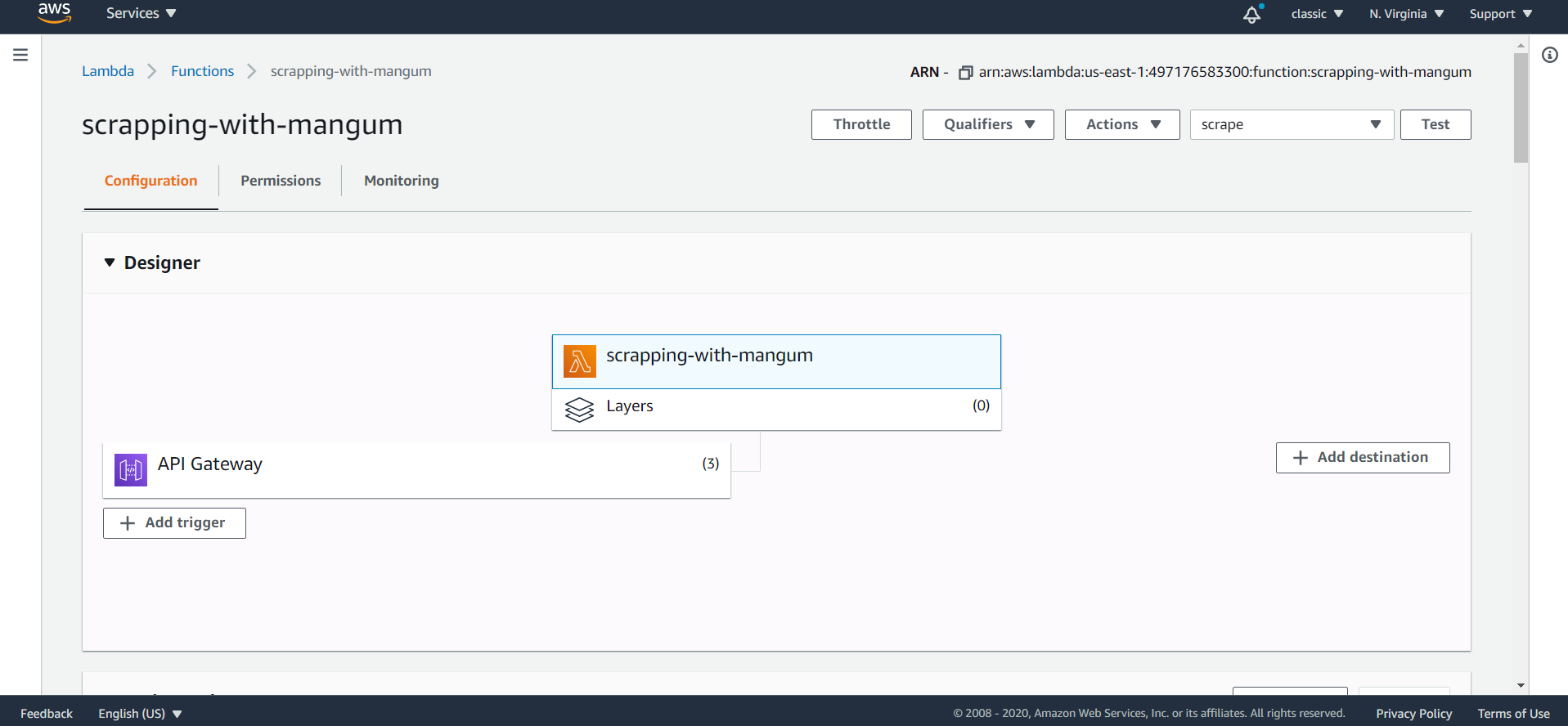
* Data masking is the process of replacing sensitive data copied from production databases to test non-production databases with realistic, but scrubbed, data based on masking rules.
* Data masking is ideal for situations when confidential or regulated data needs to be shared with non-production users.
* Data masking enables organizations to generate realistic and fully functional data with similar characteristics as the original data to replace confidential information.

# Steps :

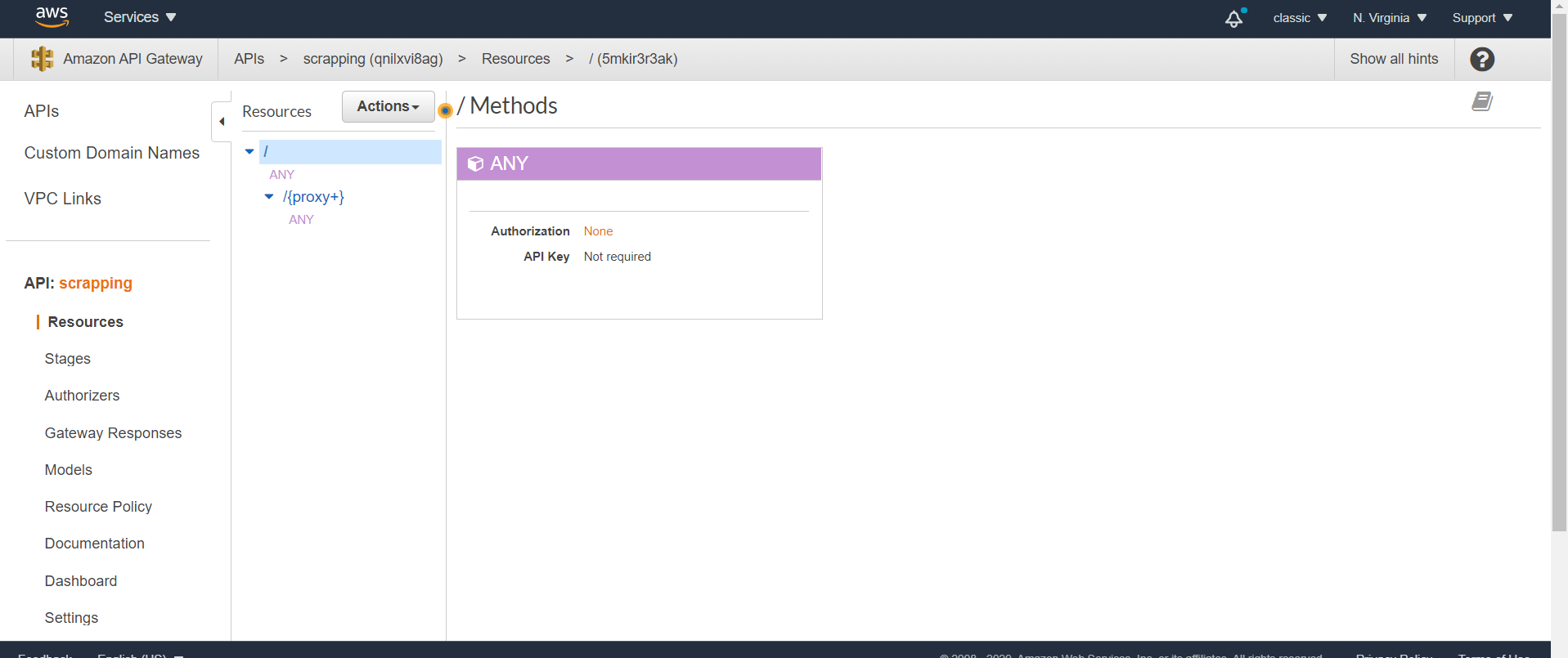
1. Designed fast API



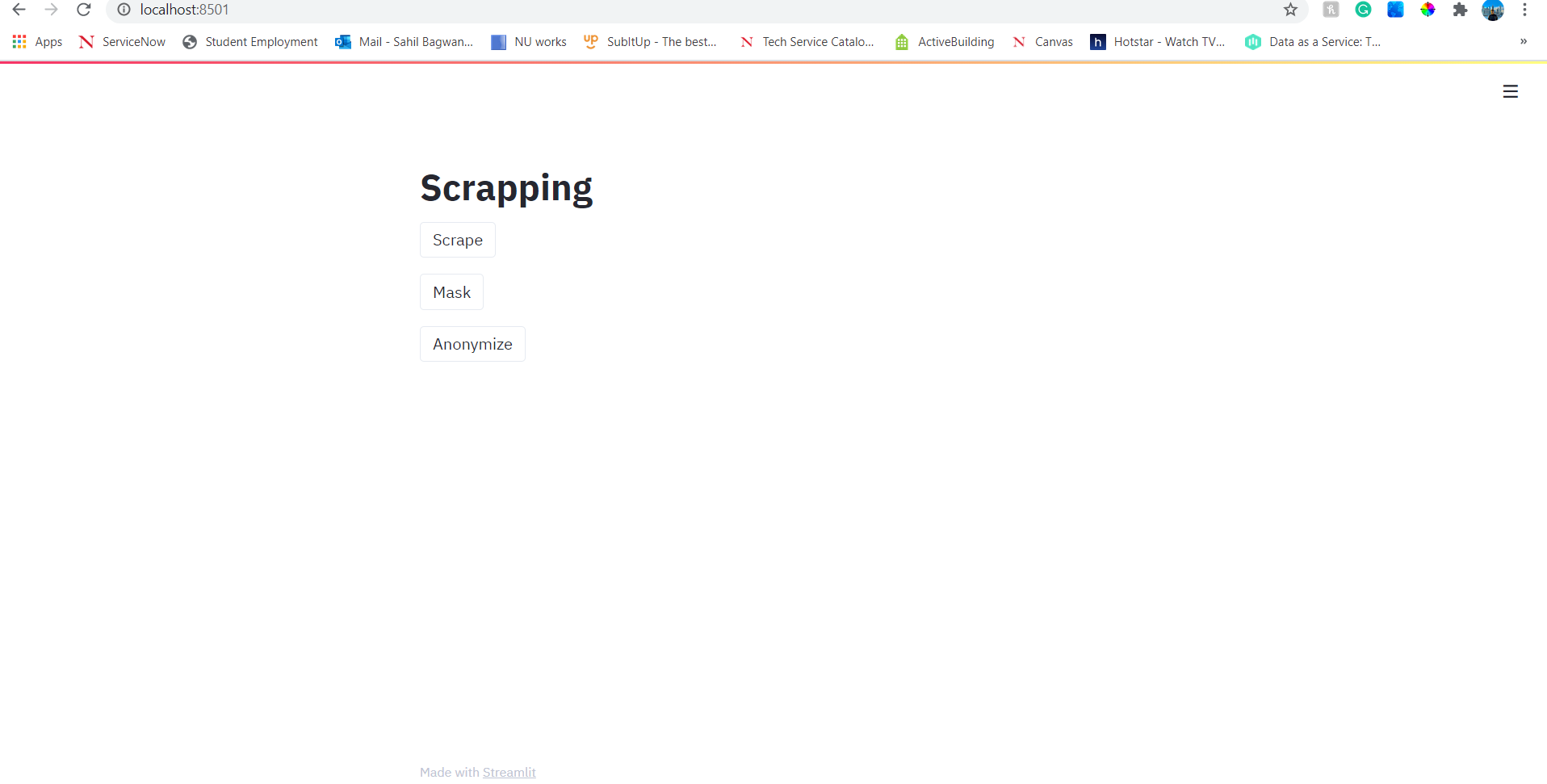
1. Wrapping the Fast API with AWS mangum



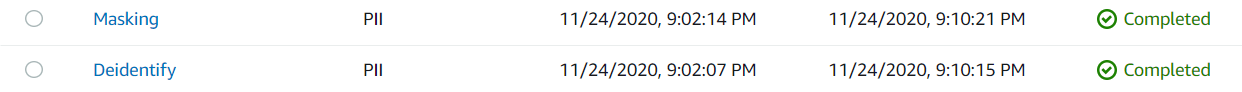
3. API gateway integration with AWS Lambda:



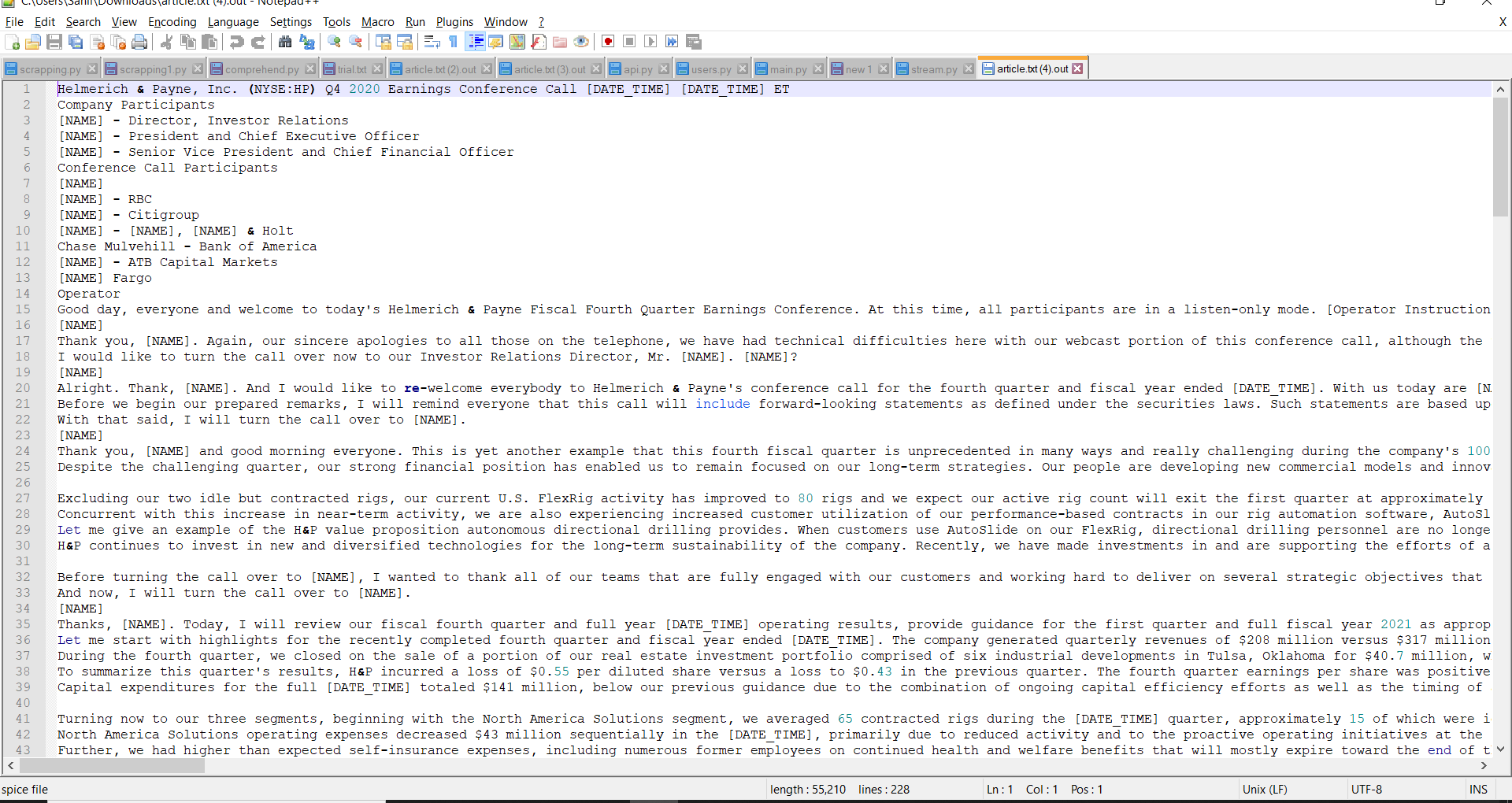
4. Deploying the designed APIs’ in Streamlit:



5. Running the Name Entity Recognition and Masking jobs in AWS Comprehend:



6. Output file of Named Entity Recognition:



7. Output file of Masking:

