Lab 4 Report

Dylon Rajah

December 20, 2022

Project's Public GitHub Repository can be found here: https://github.com/d-rajah/Lab4

1 Program Summary

For Lab 4, I decided to implement option 2 - Integrate trending topics from Twitter with Kafka. The program runs on 3 instances on the cloud. One instance acts as a Manager and the other two are workers. The manager will have access to the Twitter API. However, **due to invalid Twitter credentials, the program will gather tweets OF trending topics**. I will elaborate more on this in section 2.1

2 Edited Files

2.1 Producer

The manager connects to Kaftka using the Kaftka Producer import. First it will create a Kaftka broker address with the IP and port provided of the worker servers, either IP will work. Then we connect to the Twitter API using the Bearer token provided by the developer section of Twitter. **To get the trending topics of Twitter, I need to apply for Elevated access. This is free and allows me more freedom when using the API. However, there is a 2 day processing time where they approve my request. After 2 days, I still haven't been approved and with the deadline coming up, I decided to slightly change the purpose. Instead, on December 17th, I looked up the trending topics on Twitter using this link: https://trends24.in/united-states/. From there I added the top 10 topics to an array and the program will search Twitter for 100 tweets containing any of those keywords. From there, the program continues as expected and the consumers will read whatever the producers publish.**

3 Consumers

In my cluster I had 2 consumers. Both consumers were able to read from the publisher sequentially - I did not implement it to retrieve at the same time, however I dont think this would be terribly difficult. The consumer will connect to the producer with the IP and Port number passed into it and will read the data that the producer is publishing. The topic is changed to "Tweets" and there is a text file that the output is being written to.
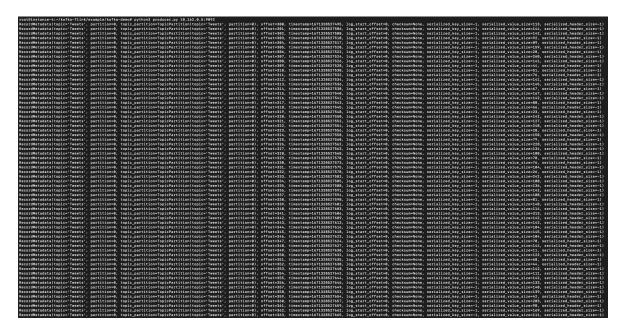
4 Demo

Figure 1: Producer Running

Figure 2: Consumer Retreiving Tweets

5 Difficulties Encountered

The hardest part of this lab was setting up Kaftka. I would say that itself was terribly difficult, just
it had a large learning curve. After I had it up and running it made a lot of sense. I have been using

ROS all semester when I help Dr.Lyons with research so I'm familiar with the publisher/subscriber model. Getting access to Twitter was also new but Im sure with the right credentials it would be a lot simpler.