Project Proposal

Developing Kafka Deployment and Publisher/Consumer for REST Based Services

Tanu Bordia, Vishesh Ruparelia
IMT2016002, IMT2016006

Course: NC 864/ SDN & NFV

1 Objective & Scope:

The objective of this project is to deploy Kafka for publisher/consumer REST based services. Kafka topic will load the exposed data from HackerNews using standard REST API(s). The above will be implemented using the Python. The data will then be published and consumed by a program running locally, a remote application and ElasticSearchDB.

2 Approach:

Our approach involves pulling data from official HackerNews REST API servers and then publishing it to a topic in the Kafka cluster. Every entry under a topic is called a record which can be identified using a sequential id number. This number will be called offset of that record. The published data in turn will be consumed by three consumers:

1. A process running on the localhost which logs the latest record.
    - A daemonized python script which reads the records from the message queue and logs it to a file.
2. A web server displaying latest posts related to a particular topic in the HackerNews feed.
    - Django/Flask based server which reads messages from the queue, filters it based on some parameters and displays it on the web.
3. ElasticSearch instance storing the records which are accessed by Kibana for visualization.

3 Timeline:

Below is a tentative timeline for delivery of particular aspects of the project:

Timeline	
Date	**Target**
20/10/2019	Learning best practices of working with Kafka
25/10/2019	Create publisher using data read through HackerNews API
30/10/2019	Developing consumer applications and setting them up to run with Kafka
7/11/2019	Setup ELK stack
14/11/2019	Debugging and project report

4 References:

- Apache Kafka: `https://kafka.apache.org/intro`
- Github
- REST API: `https://restfulapi.net/`
- Kafka for Python: `https://pypi.org/project/kafka-python/`
- HackerNews API: `https://hackernews.api-docs.io`