Step 1:

Starting kafka broker, producer and consumer services following the below steps:

a. Open a terminal

b. Expecting that the Kafka service is installed in kafka folder. If not, please follow the installation steps in the readme.md file.

c. Go to kafka folder using the below command:

cd ~/kafka

bin/zookeeper-server-start.sh config/zookeeper.properties

d. Open a new terminal

Go to kafka folder using the below command:

cd ~/kafka

bin/kafka-server-start.sh config/server.properties

e. Open a new terminal to run HBase service

sudo service hbase-master start;

sudo service hbase-regionserver start;

f. Open a new terminal to Create Movie Table in HBase

hbase shell

create 'movies', 'info', 'stats'

desc 'movies'

g. Open a new terminal to run Kafka Producer

cd <Jar File folder>

java -jar NetflixStreamProducer.jar

h. Open a new terminal to run Kafka Consumer

cd <Jar File folder>

spark-submit --class "consumer.NetflixConsumerSpark" --master local[*] NetflixStreamConsumerSpark.jar

h. Data Visualization

Install Python (Anaconda optional), Vs Code/any other Python IDE

Open NetflixMovieDataVizulation.ipynb from DataVisualization folder by using VS Code

Open an integrated terminal in VS code and install below packages

pip install happybase setuptools pandas matplotlib seaborn plotly nbformat

run NetflixMovieDataVizulation.ipynb

d. Copy the script we provided in our script folder in kafka folder and give executable permission using the below 2 commands:

cp <SCRIPT_FOLDER>/broker-start.sh ~/kafka/

chmod +x broker-start.sh

e. Now start all the services using the below command:

./broker-start.sh