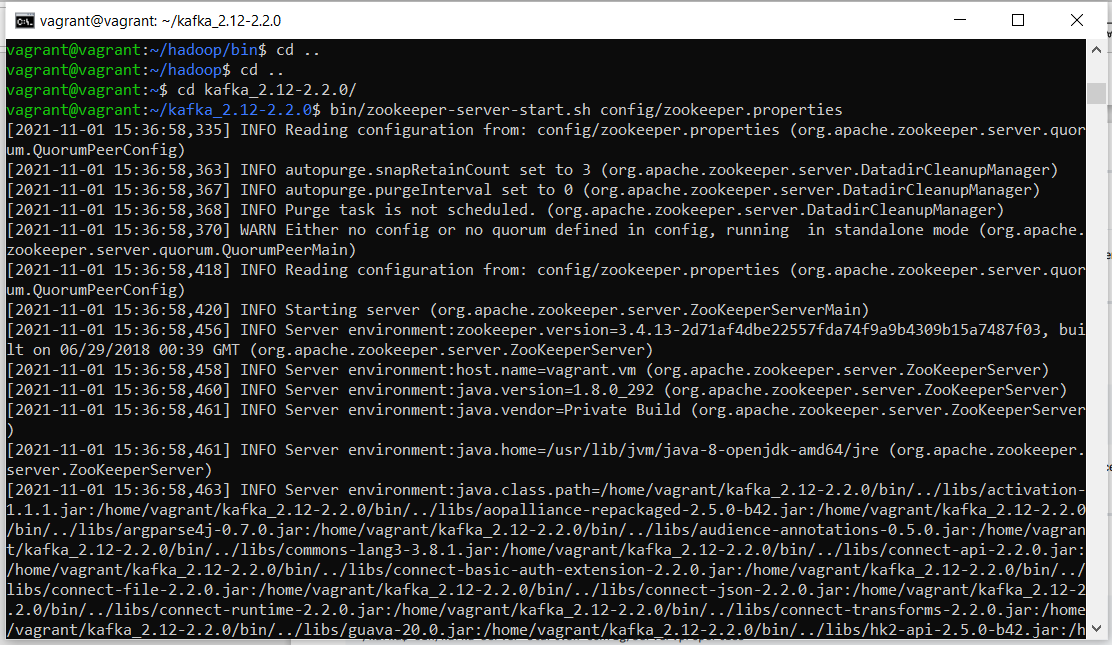
Data Ingestion and Loading:

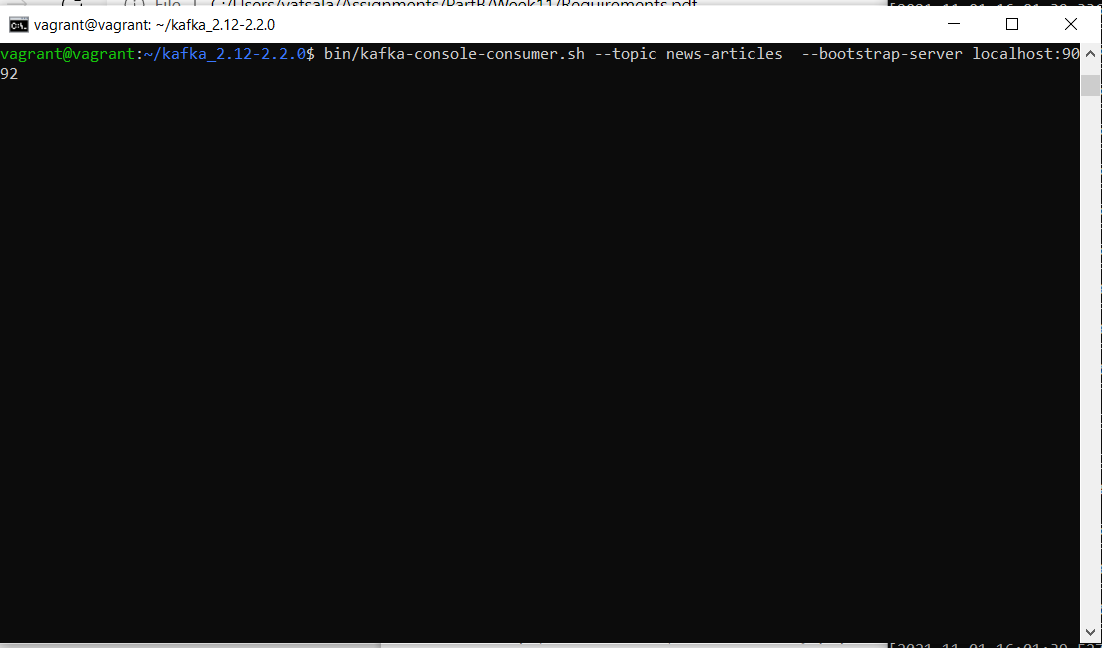
Here we are going to use kafka and zookeeper installed on vagrant. Data collection starts using RapidAPI from python and sends to kafka queue. From kafka consumer, we will read the data again using python and save the data to mongo db database, which installed locally on windows. This flow will help us handle large volume of data.

Please follow following steps for data ingestion and loading.

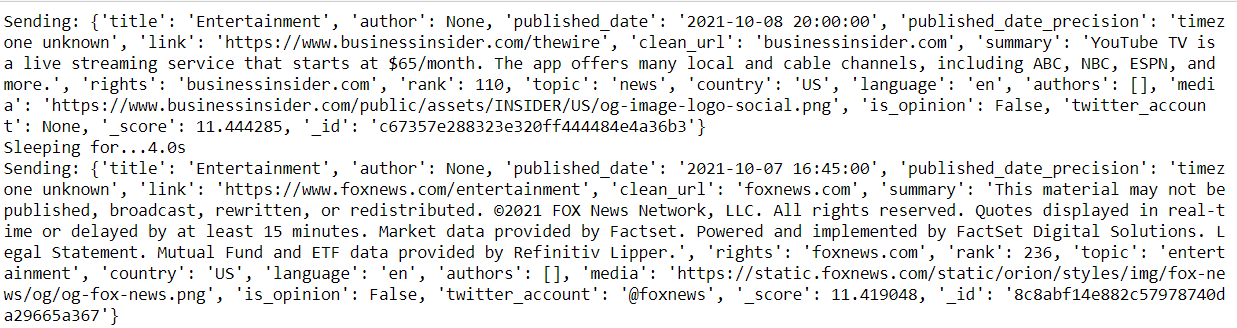
1. Install vagrant
2. Install zookeeper in vagrant
3. Install kafka producer in vagrant
4. Install kafka consumer in vagrant
5. Install mongo db (local on windows)
6. Start zookeeper in one of the vagrant terminals

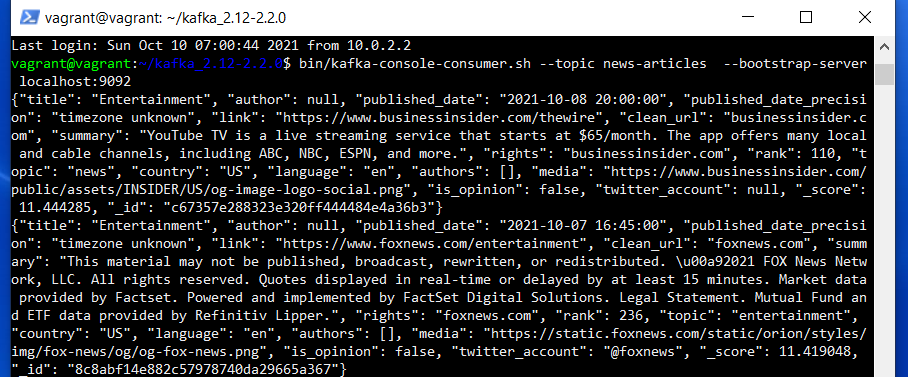


1. Start kafka server in another vagrant terminal
2. Start kafka consumer in another vagrant terminal



1. Run newsarticles\_producer.py at python prompt. Program will start sending the data collecting from rapidAPI to kafka



1. Consumer will start receiving the data
2. Start newsarticles\_consumer.py to collect the data from kafka queue and save it to mongo db.

