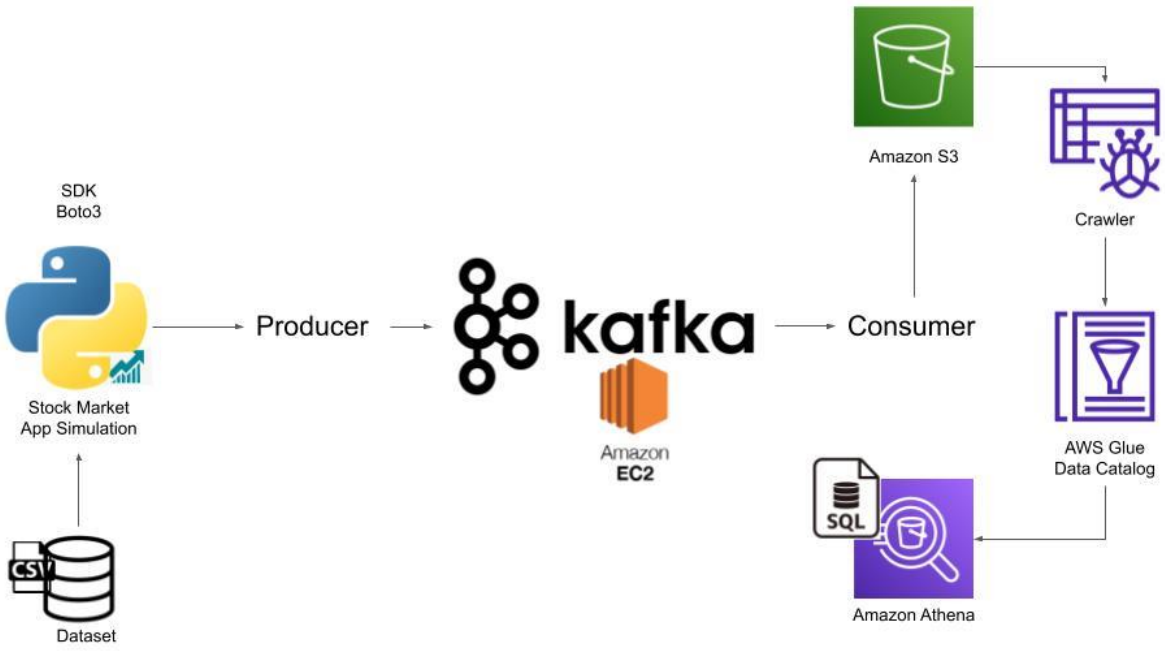
Stock Market Real-Time Data Analysis Using Kafka

Architecture



The Architecture of this project is as follows:

- The stock market is collect in the form of CSV file
- The Data is loaded and analysed using pandas library
- **Kafka Producer** and **Kafka consumer** are configured and data is sent to s3 buckets
- The **AWS crawler** is used to get the schema of the data
- Database is created using **AWS Glue Catalog** and the same database is used in **Athena** for the analysis purpose

Stock Market Data:

```
data.head()
```

	Index	Date	Open	High	Low	Close	Adj Close	Volume	CloseUSD
0	HSI	1986-12-31	2568.300049	2568.300049	2568.300049	2568.300049	2568.300049	0.0	333.879006
1	HSI	1987-01-02	2540.100098	2540.100098	2540.100098	2540.100098	2540.100098	0.0	330.213013
2	HSI	1987-01-05	2552.399902	2552.399902	2552.399902	2552.399902	2552.399902	0.0	331.811987
3	HSI	1987-01-06	2583.899902	2583.899902	2583.899902	2583.899902	2583.899902	0.0	335.906987
4	HSI	1987-01-07	2607.100098	2607.100098	2607.100098	2607.100098	2607.100098	0.0	338.923013

Kafka Producer Code:

```
import pandas as pd
from kafka import KafkaProducer
from time import sleep
from json import dumps
import json

data= pd.read_csv("https://raw.githubusercontent.com/darshilparmar/stock-market-kafka-data-engineering-project/main/indexProcesse

#this is the way in which kafkaproducer is connected to kafkatopics and pushes the data to topic
producer = KafkaProducer(bootstrap_servers=['54.198.84.237:9092'], #change ip here
                         value_serializer=lambda x:
                         dumps(x).encode('utf-8'))

#in this way the data is sent to kafkatopic
while True:
    stock_dict = data.sample(1).to_dict(orient='records')
    producer.send('demo_test', value=stock_dict)
```

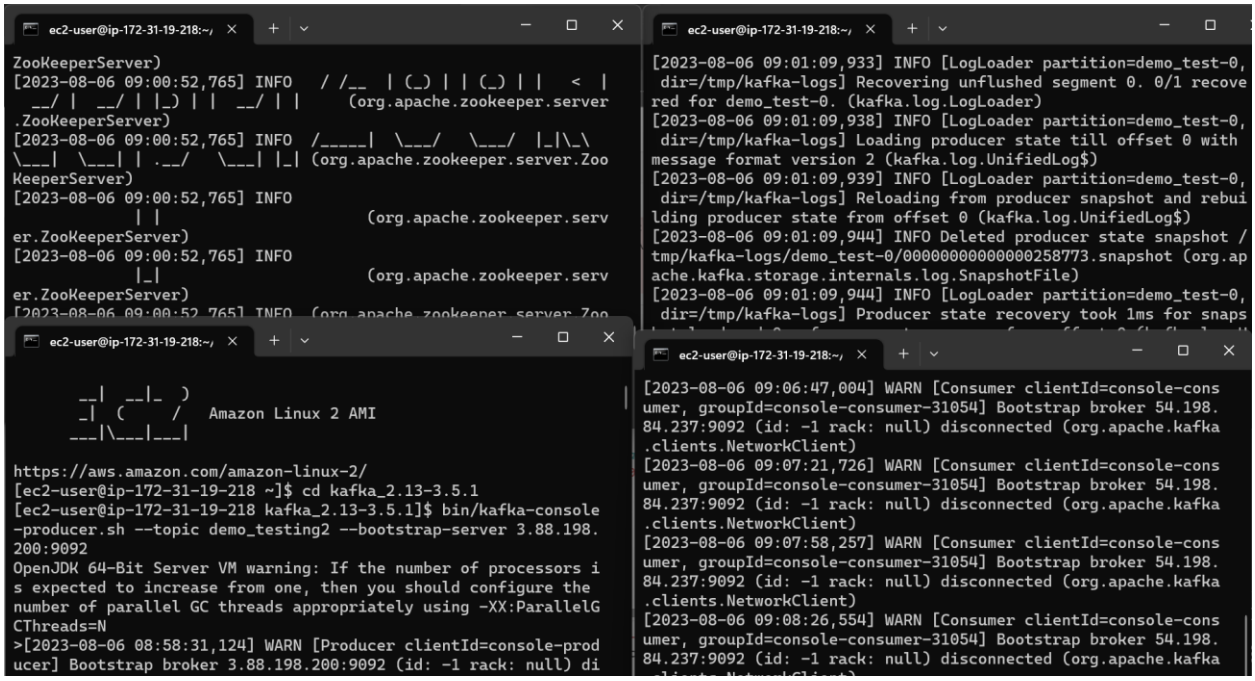
Kafka Consumer code:

```
import pandas
from kafka import KafkaConsumer
from time import sleep
from json import dumps,loads
import json
from s3fs import S3FileSystem



#this is the way kafkaconsumer is connected to kafkatopic
consumer = KafkaConsumer(
    'demo_test',
     bootstrap_servers=['54.198.84.237:9092'], #add your IP here
    value_deserializer=lambda x: loads(x.decode('utf-8')))

s3 = S3FileSystem()
#this loop will take the data from kafka topic and puts that data to s3 using dumps function
for count, i in enumerate(consumer):
    with s3.open("s3://stock-market-data-laxman/stock_market_{}.json".format(count), 'w') as file:
        json.dump(i.value, file)
```



Zookeeper ,Kafka,Kafka Producer and Kafka Consumer Running Servers