Real-time data analytics Using Kafka

Practical

- AWS account → Free tier

Services

- EC2 (Virtual Machine)
- Pycharm
- Google Colab
- S3 → data Storage
- Glue → Crawling data from S3
- Athena → SQL

Step 1 Setting up VM → EC2

- Amazon Linux 2 AMI

→ Security → Security group → Add 1 inbound rule

All traffic → IPv4

Step2 → Pycharm (local machine) → Terminal

↓ SSH command.

EC2

Step3 In your virtual machine install

1. JVM (17) ✓
2. Kafka ←

Download latest version of kafka

```
curl -O https://dlcdn.apache.org/kafka/4.0.0/kafka_2.13-4.0.0.tgz
```

→ Extract Kafka

```
tar -xzf kafka_2.13-4.0.0.tgz
cd kafka_2.13-4.0.0
```

→ Install Java

```
sudo yum install -y java-17-amazon-corretto
```

→ cd kafka

→ `KAFKA_CLUSTER_ID="$(bin/kafka-storage.sh random-uuid)"`

→

```
bin/kafka-storage.sh format --standalone -t $KAFKA_CLUSTER_ID -c
config/server.properties
```

⇒ sudo nano config/server.properties

→ advertised.listeners=PLAINTEXT://localhost:9092,CONTROLLER://localhost:9093

Replace with EC2 ip address

107.20.130.244

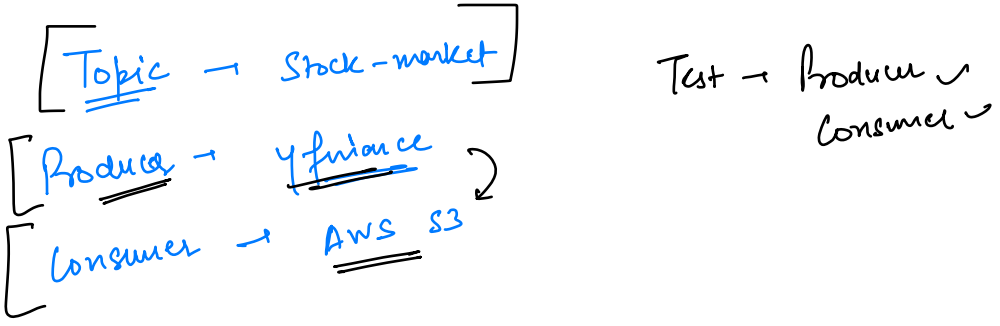
advertised.listeners=PLAINTEXT://107.20.130.244:9092,CONTROLLER://107.20.130.244:9093

→ Adjust the memory

export KAFKA_HEAP_OPTS="-Xms512M -Xmx512M"

→ Start the Kafka server

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



Topic creation

```
bin/kafka-topics.sh --create --topic stock-market --bootstrap-server 107.20.130.244:9092
```

Producer

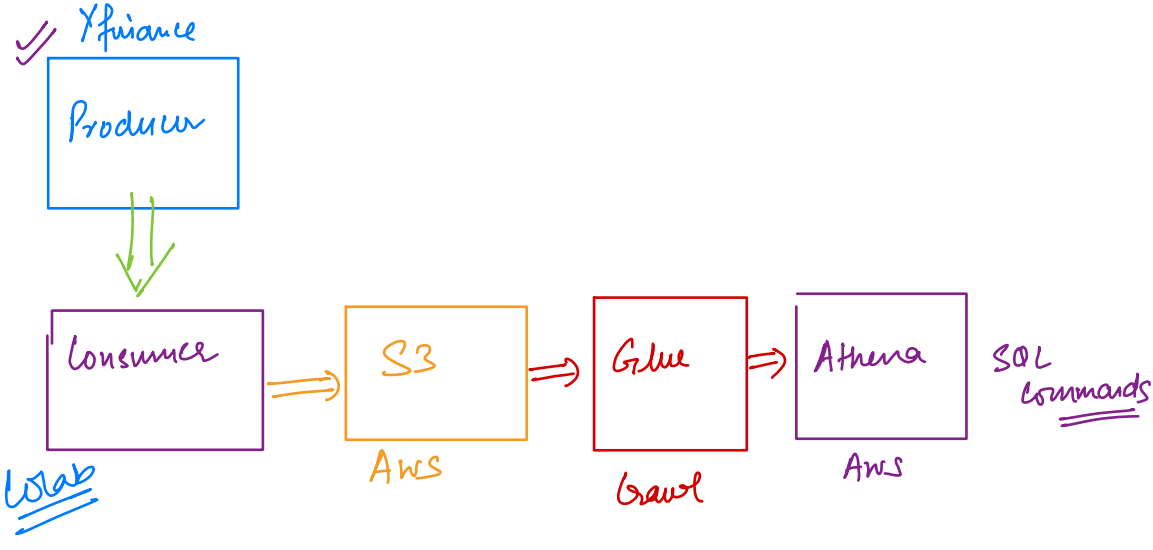
```
bin/kafka-console-producer.sh --topic stock-market --bootstrap-server 107.20.130.244:9092
```

new-terminal → Consumer

```
bin/kafka-console-consumer.sh --topic stock-market --from-beginning --bootstrap-server 107.20.130.244:9092
```

→ Producer Colab

→ Consumer Colab.



→ Security credentials

[Access ID
&
Secret Access key]