sed -i 's|#advertised.listeners=PLAINTEXT://your.host.name:9092|advertised.listeners=PLAINTEXT://15.223.5.123:9092|' config/server.properties

ssh -i real-time-analytics-ec2-key-pair.pem ec2-user@15.156.199.54

 **Start Zookeeper:**

bash

Copy code

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

 **Verify Zookeeper is Running:**

bash

Copy code

echo ruok | nc localhost 2181

output should be imok

ps aux | grep zookeeper

netstat -an | grep 2181

 **Start Kafka:**

bash

Copy code

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



# Stop Zookeeper:

bash

Copy code

bin/zookeeper-server-stop.sh

# Stop Kafka:

bash

Copy code

bin/kafka-server-stop.sh

bin/kafka-console-consumer.sh --bootstrap-server 15.223.5.123:9092 --topic test --from-beginning

# Step-by-Step Guide to Create a Kafka Topic and Verify Setup

## Step 1: Create a Kafka Topic

Create a Topic Named test:

bash

Copy code

bin/kafka-topics.sh --create --topic test --bootstrap-server 172.31.17.38:9092 --partitions 1 --replication-factor 1

bin/kafka-topics.sh --create --topic user-activity --bootstrap-server 172.31.17.38:9092 --partitions 3 --replication-factor 1

## Step 2: List Kafka Topics

Verify the Topic Creation:

bash

Copy code

bin/kafka-topics.sh --list --bootstrap-server 172.31.17.38:9092

## Step 3: Describe the Kafka Topic

Get Details About the Topic:

bash

Copy code

bin/kafka-topics.sh --describe --topic test --bootstrap-server 172.31.17.38:9092

## Step 4: Produce and Consume Messages

Start a Kafka Producer

Start the Kafka Producer to Send Messages:

bash

Copy code

bin/kafka-console-producer.sh --broker-list 172.31.17.38:9092 --topic test

Type Messages:

Type a few messages and press Enter after each. For example:

bash

Copy code

Hello, Kafka!

This is a test message.

Start a Kafka Consumer

Open Another Terminal and Start the Kafka Consumer to Receive Messages:

bash

Copy code

bin/kafka-console-consumer.sh --bootstrap-server 172.31.17.38:9092 --topic test --from-beginning

Verify Messages:

You should see the messages you typed in the producer console appearing in the consumer console.

## Example Commands in Sequence

Create a Topic:

bash

Copy code

bin/kafka-topics.sh --create --topic test --bootstrap-server 15.223.5.123:9092 --partitions 1 --replication-factor 1

List Topics:

bash

Copy code

bin/kafka-topics.sh --list --bootstrap-server 15.223.5.123:9092

Describe Topic:

bash

Copy code

bin/kafka-topics.sh --describe --topic test --bootstrap-server 15.223.5.123:9092

Start Producer:

bash

Copy code

bin/kafka-console-producer.sh --broker-list 15.223.5.123:9092 --topic test

Type messages like:

bash

Copy code

Hello, Kafka!

This is a test message.

Start Consumer:

bash

Copy code

bin/kafka-console-consumer.sh --bootstrap-server 15.223.5.123:9092 --topic test --from-beginning

modifying spark config in the existing emr cluster:  
aws emr modify-instance-groups --cluster-id j-UCFR5FL9F7O3 --instance-groups InstanceGroupId= ig-2HZUZZK7HG0OE,Configurations=[{"Classification":"spark-defaults","Properties":{"spark.executor.memory":"4G","spark.executor.cores":"2","spark.driver.memory":"4G","spark.driver.cores":"2","spark.kafka.bootstrap.servers":"172.31.17.38:9092"}}]

I created the spark-config.json then used that instead of using above command.

ssh into the master node:  
  
ssh -i real-time-analytics-ec2-key-pair.pem hadoop@ec2-35-183-19-4.ca-central-1.compute.amazonaws.com

scp -i real-time-analytics-ec2-key-pair.pem -r data\_ingestion hadoop@ec2-35-183-19-4.ca-central-1.compute.amazonaws.com:/home/hadoop/

scp -i real-time-analytics-ec2-key-pair.pem -r data\_processing hadoop@ec2-35-183-19-4.ca-central-1.compute.amazonaws.com:/home/hadoop/