Create EC2 instance. After that follow below steps.

Create AWS Credentials folder to access its resources further in code

In SSH; go to home if not already there (cd ~)

1. Create new folder:

mkdir .aws

2. Create a new file

vim .aws/credentials

3. paste entire AWS CLI content – as it is – save and quit

[you need to modify the credentials file on every launch since AWS CLI is not constant]

To Install Jupyter and run on EC2

SSH into the instance and type below commands one by one.

1. sudo su (root user)

2. yum update (to perform updates, of any required)

3. Create and Activate python virtual env

python3 -m venv venv (To create virtual environment)  
source venv/bin/activate (To activate virtual environment)

4. Install the required packages – make note that there were no errors [warnings can be ignored]

pip install pyyaml ipython jupyter ipyparallel pandas boto3 -U

5. Enable IPython Cluster

ipcluster nbextension enable

6. Start an ipcluster with 4 engines

ipcluster start -n 4

[DO NOT stop/interrupt the process – otherwise you will be unable to finish this assignment]

7. Let the previous SSH terminal be as it is and start a new session (SSH again to the instance)

source venv/bin/activate

jupyter notebook --port=8888 --no-browser --ip=0.0.0.0 --allow-root

Commands to Download Kafka

wget https://downloads.apache.org/kafka/3.7.0/kafka\_2.13-3.7.0.tgz

tar -xvf kafka\_2.13-3.7.0.tgz

Java is Prerequisite to run Kafka.

java -version (to check java version)

sudo yum install java-1.8.0-openjdk (to install java)

Change to Kafka directory before starting server

cd kafka\_2.13-3.7.0

To Start Zoo-keeper:

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

After running Zoo-keeper, open new terminal and SSH into it to start the Kafka Server

Start Kafka-server:

export KAFKA\_HEAP\_OPTS="-Xmx256M -Xms128M"

cd kafka\_2.13-3.7.0

By default, host pointed to the private address. So update server.properties to run it on public IP

To do this, Do "sudo nano config/server.properties" - change ADVERTISED\_LISTENERS to public ip of the EC2 instance. After that run below command to run the Kafka Server with updated config file.

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

After running Kafka-Server, again open new terminal and SSH.

Create the topic:

cd kafka\_2.13-3.7.0

bin/kafka-topics.sh --create --topic demotest --bootstrap-server 52.91.127.14:9092 --replication-factor 1 --partitions 1 (Here we are creating topic name called demotest. You can give desire name)

Start Producer:

bin/kafka-console-producer.sh --topic demotest --bootstrap-server 52.91.127.14:9092

After running Kafka-Server, again open new terminal and SSH.

Start Consumer:

cd kafka\_2.13-3.7.0

bin/kafka-console-consumer.sh --topic demotest --bootstrap-server 52.91.127.14:9092

write anything in producer and that text you can find in consumer. Great your producer and consumer are ready.