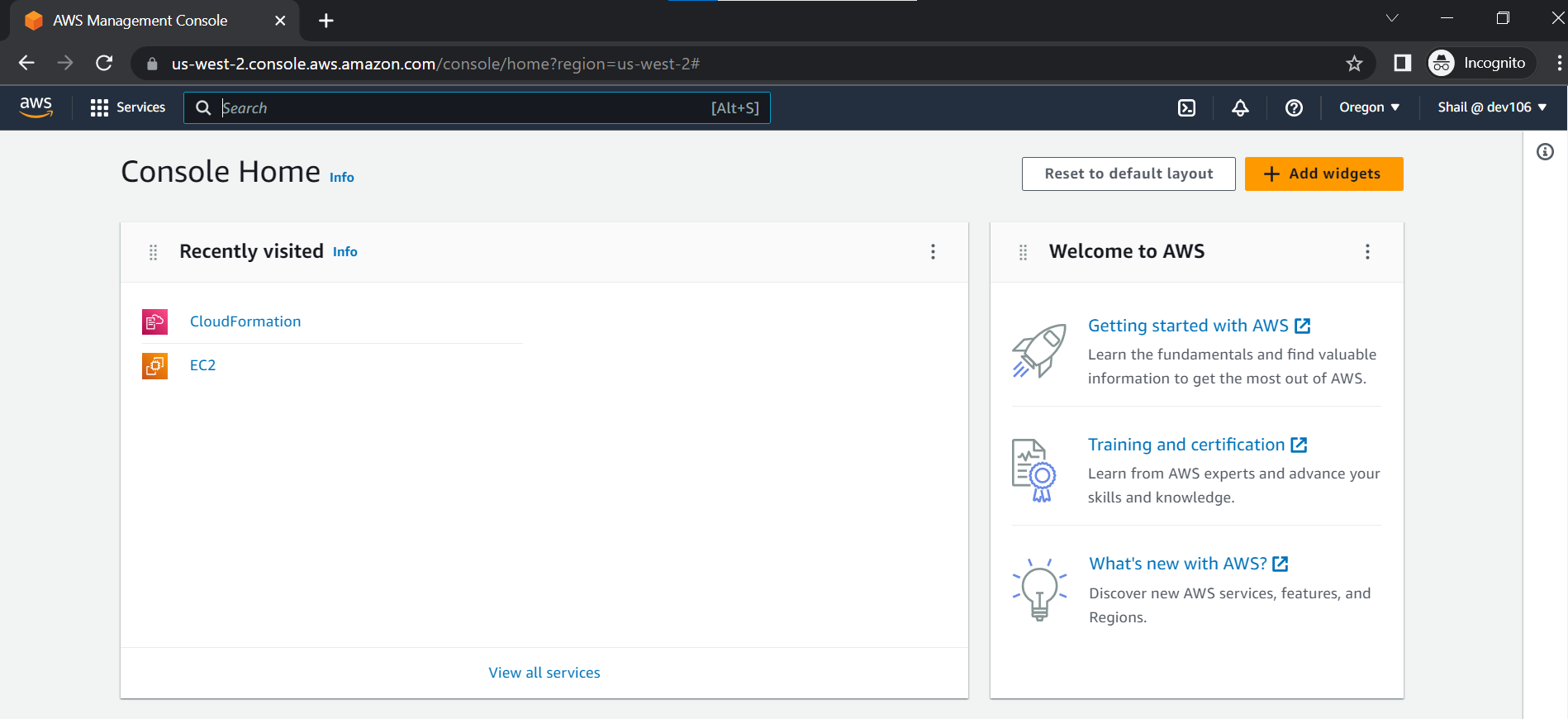
Username - Shail

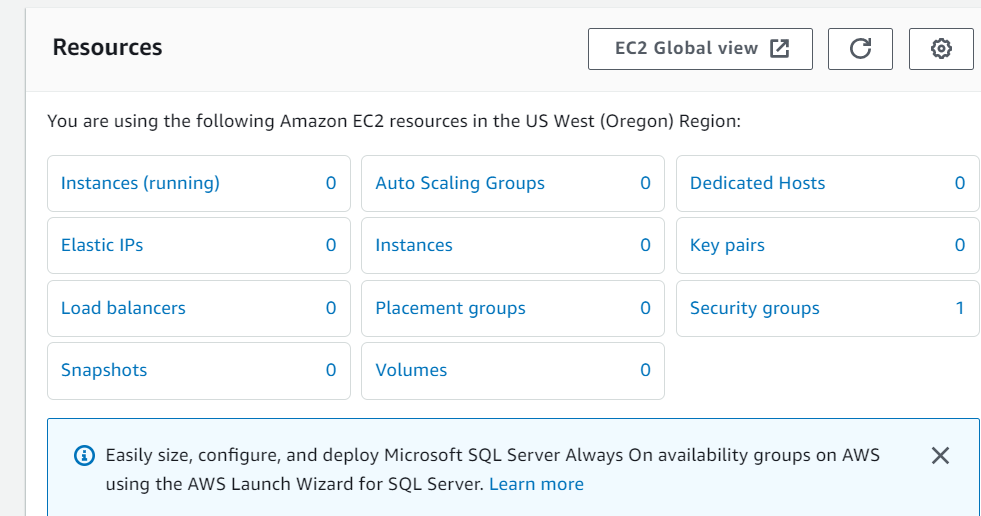
Password - Shailpatel86$

<https://dev106.signin.aws.amazon.com/console>

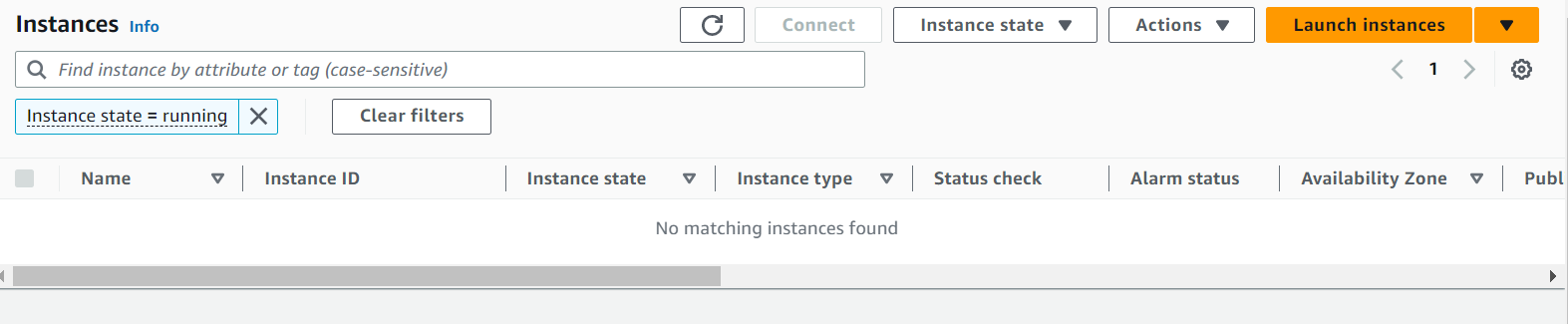
Go to the above link and log in with the given credentials



Click on EC2

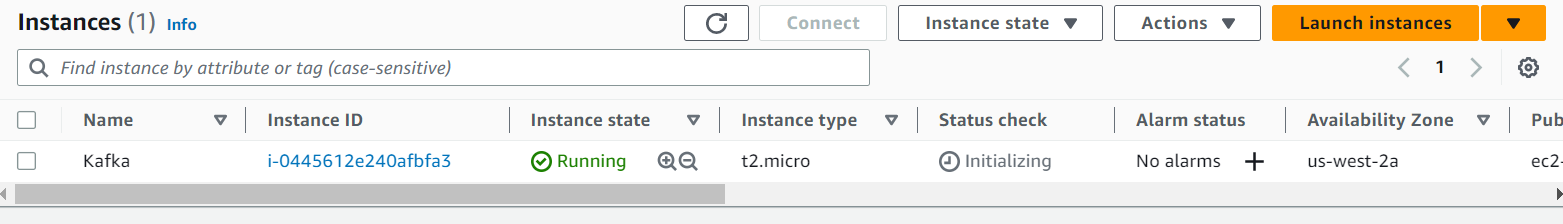


Click on Instances



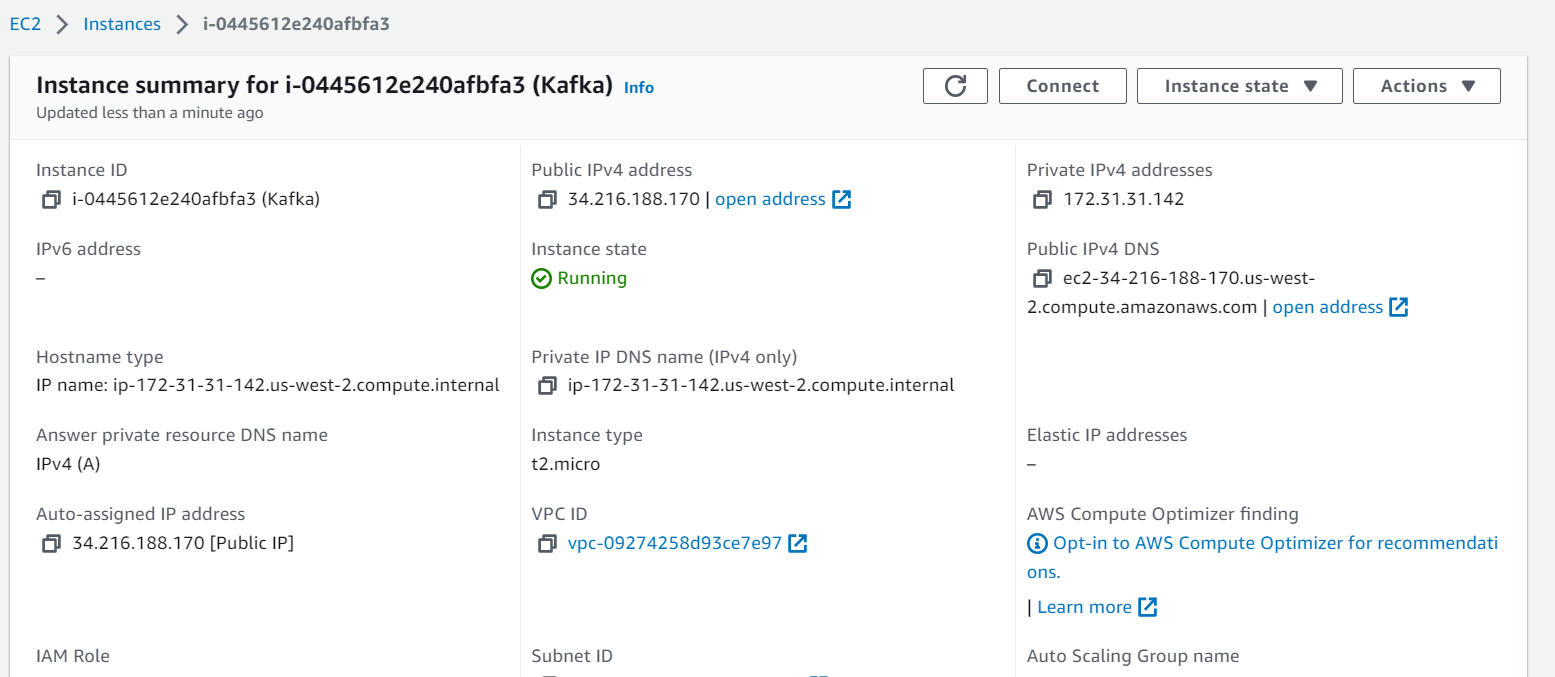
Click on launch instance

Give a name to your instance, create a keypair and keep the rest of the configurations to default and click on launch instance.



Click on the instance id

Click on connect



We will run the following commands on our instance to setup kafka and zookeeper

We get https://archive.apache.org/dist/kafka/3.3.1/kafka\_2.12-3.3.1.tgz

tar -xvf kafka\_2.12-3.3.1.tgz

-----------------------------------------------------------------------

java -version

sudo yum install java-1.8.0-openjdk

java -version

cd kafka\_2.12-3.3.1

# for ubuntu -

sudo apt-get update

sudo apt-get install -y openjdk-8-jdk

java -version

cd kafka\_2.12-3.3.1

------------------------------------------------------------------------

Start Zoo-keeper:

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

-----------------------------------------------------------------------

Start Kafka-server:

Duplicate the session & enter in a new console --

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

cd kafka\_2.12-3.3.1

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

It is pointing to private server , change server.properties so that it can run in public IP

To do this , you can follow any of the 2 approaches shared belwo --

Do a "sudo nano config/server.properties" - change ADVERTISED\_LISTENERS to public ip of the EC2 instance

--------------------------------------------------------------------------------------------------------------

Create the topic:

Duplicate the session & enter in a new console --

cd kafka\_2.12-3.3.1

bin/kafka-topics.sh --create --topic demo\_testing2 --bootstrap-server 54.215.179.39:9092 --replication-factor 1 --partitions 1

------------------------------------------------------------------------------------------------------------

Start Producer:

bin/kafka-console-producer.sh --topic demo\_testing2 --bootstrap-server 54.215.179.39:9092

----------------------------------------------------------------------------------------------------------------

Start Consumer:

Duplicate the session & enter in a new console --

cd kafka\_2.12-3.3.1

bin/kafka-console-consumer.sh --topic demo\_testing2 --bootstrap-server 54.215.179.39:9092

--------------------------------------------------------------------------------------------------------------

In Databricks create a ckuster and install this library on the cluster

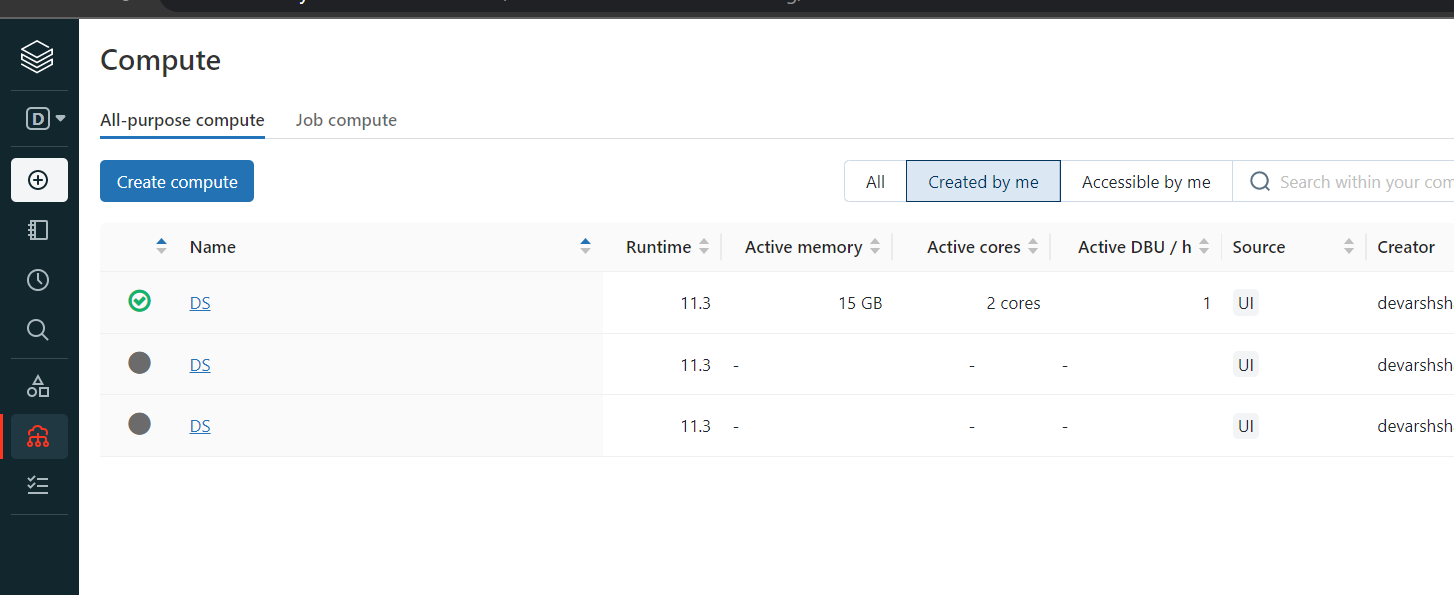
kafka-python

Attach the cluster to both the notebooks, change the IP to the current instance's IP and run the notebook.

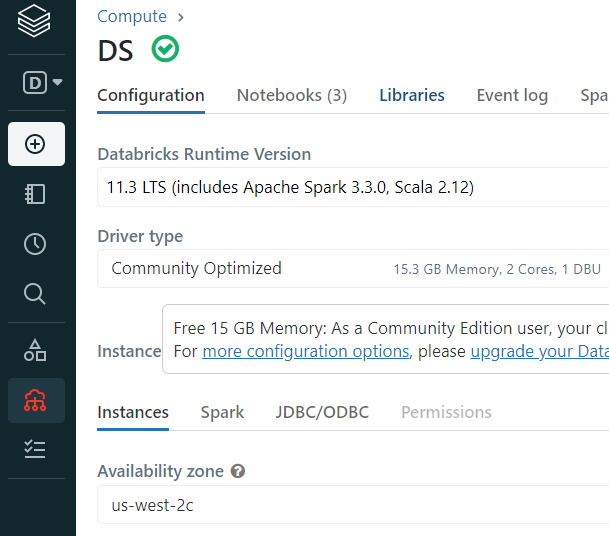
----------------------------------------------------------------------------------------------------------------

After the kafka cluster is setup on the instance, we will go to our Databricks notebooks and setup kafka consumer and kafka producer

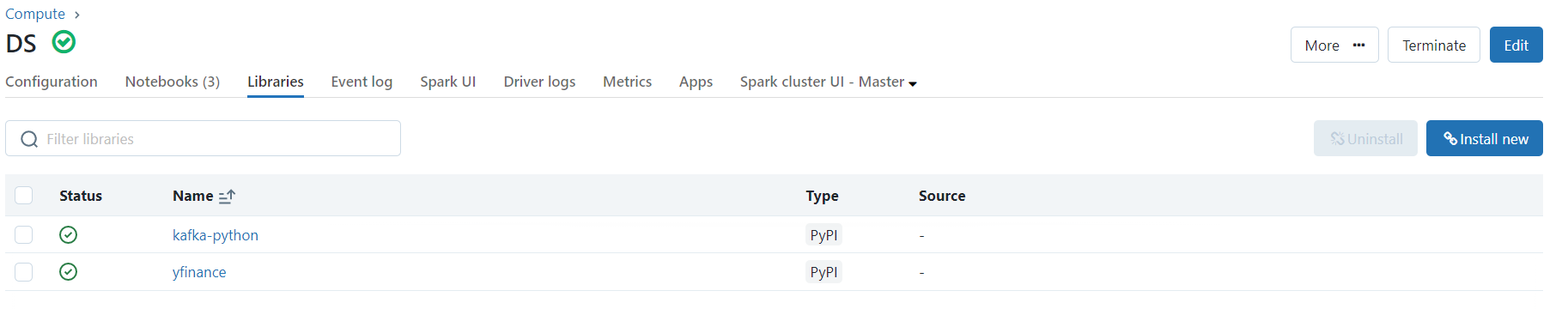
Go to compute and click on create compute to set up a cluster



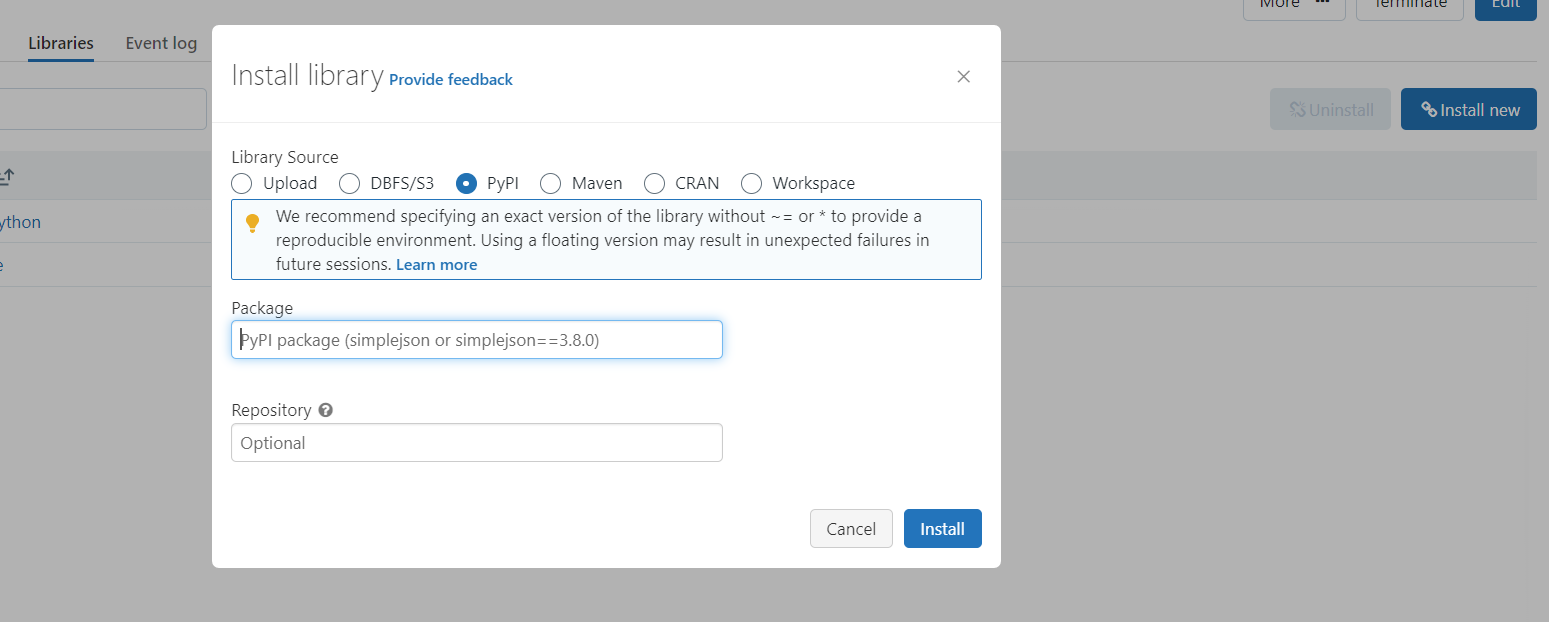
Click on the cluster name and go to libraries



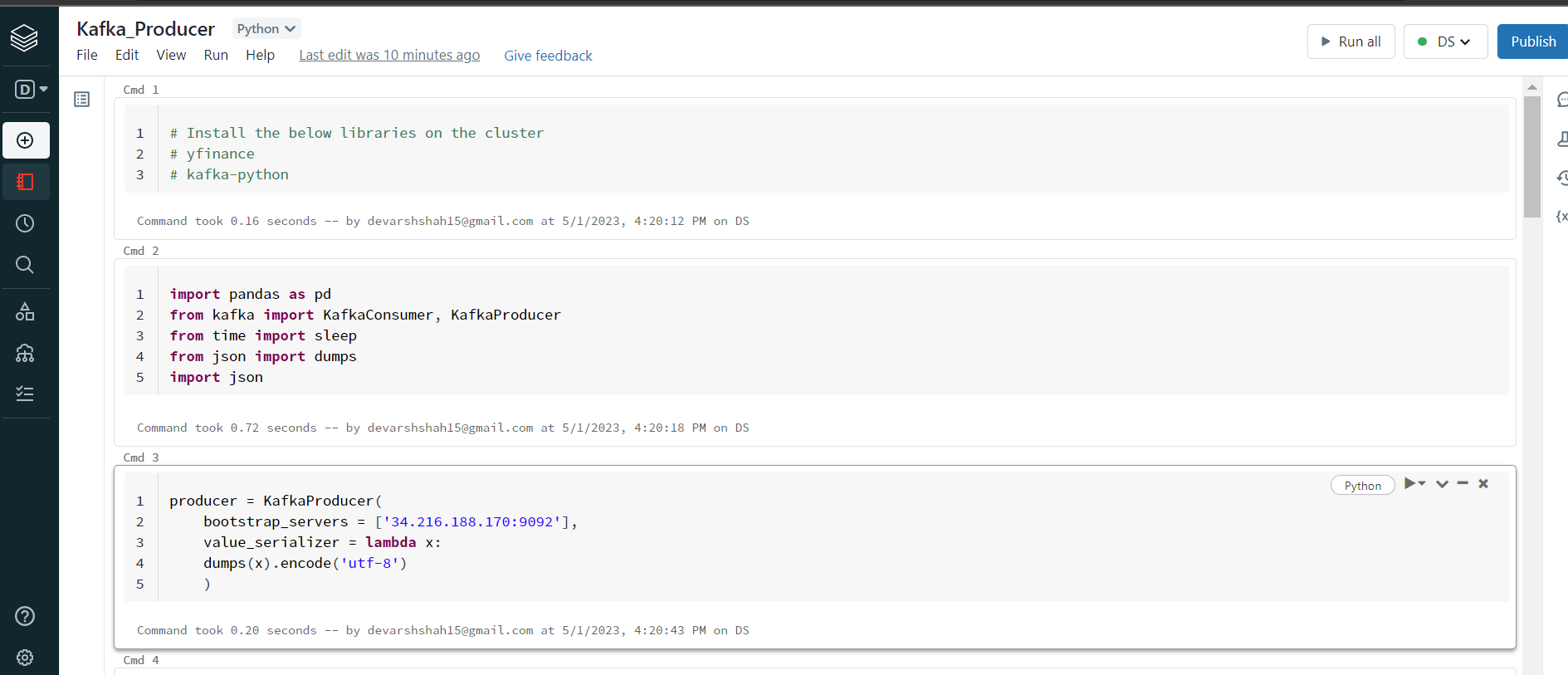
Inside it click on Install new



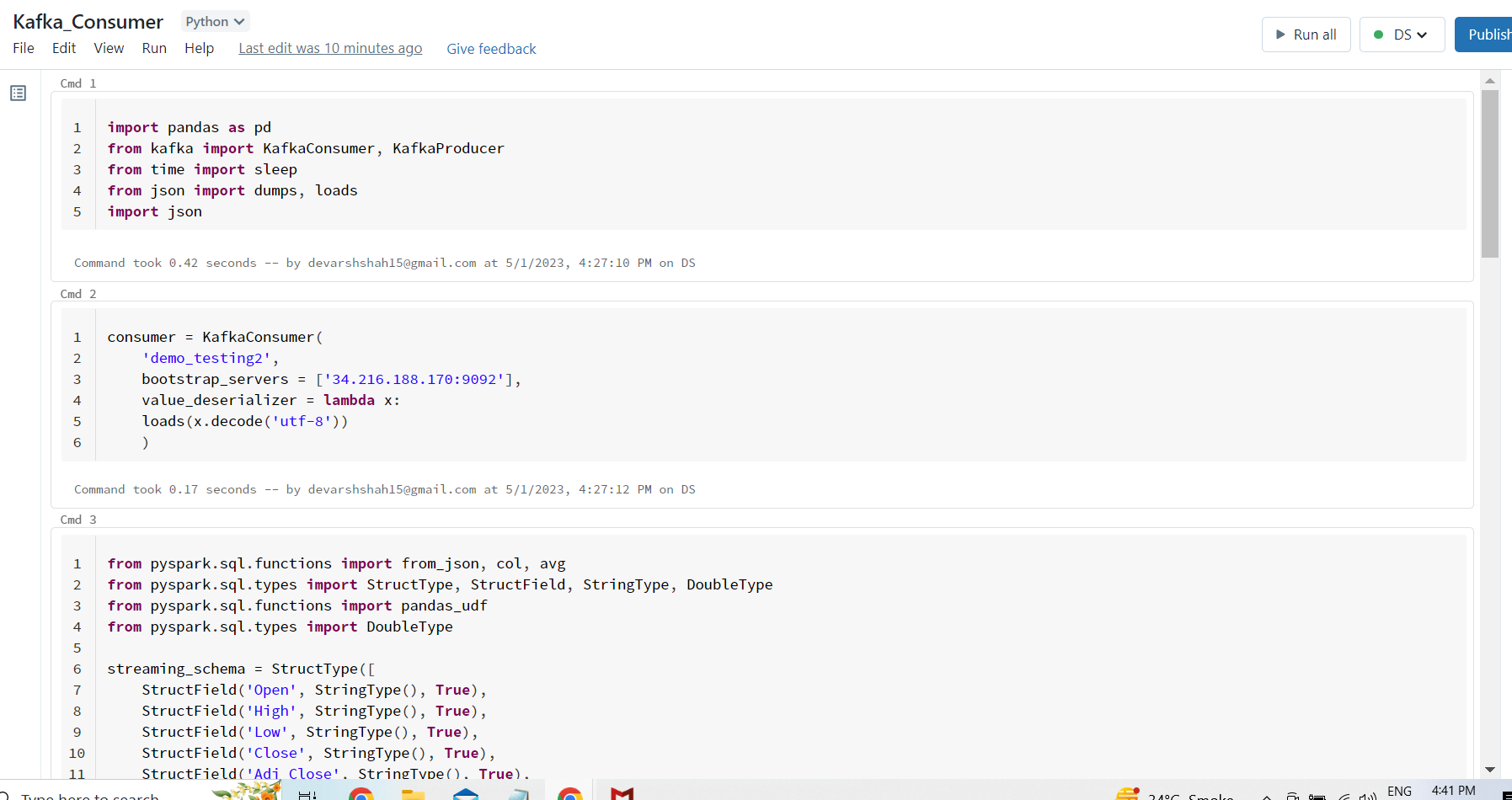
Select pypi and install two libraries one by one, kafka-python and yfinance



Go to kafka producer notebook and click on run all



Go to kafka consumer notebook and click run all



You will start seeing the values in the stream

