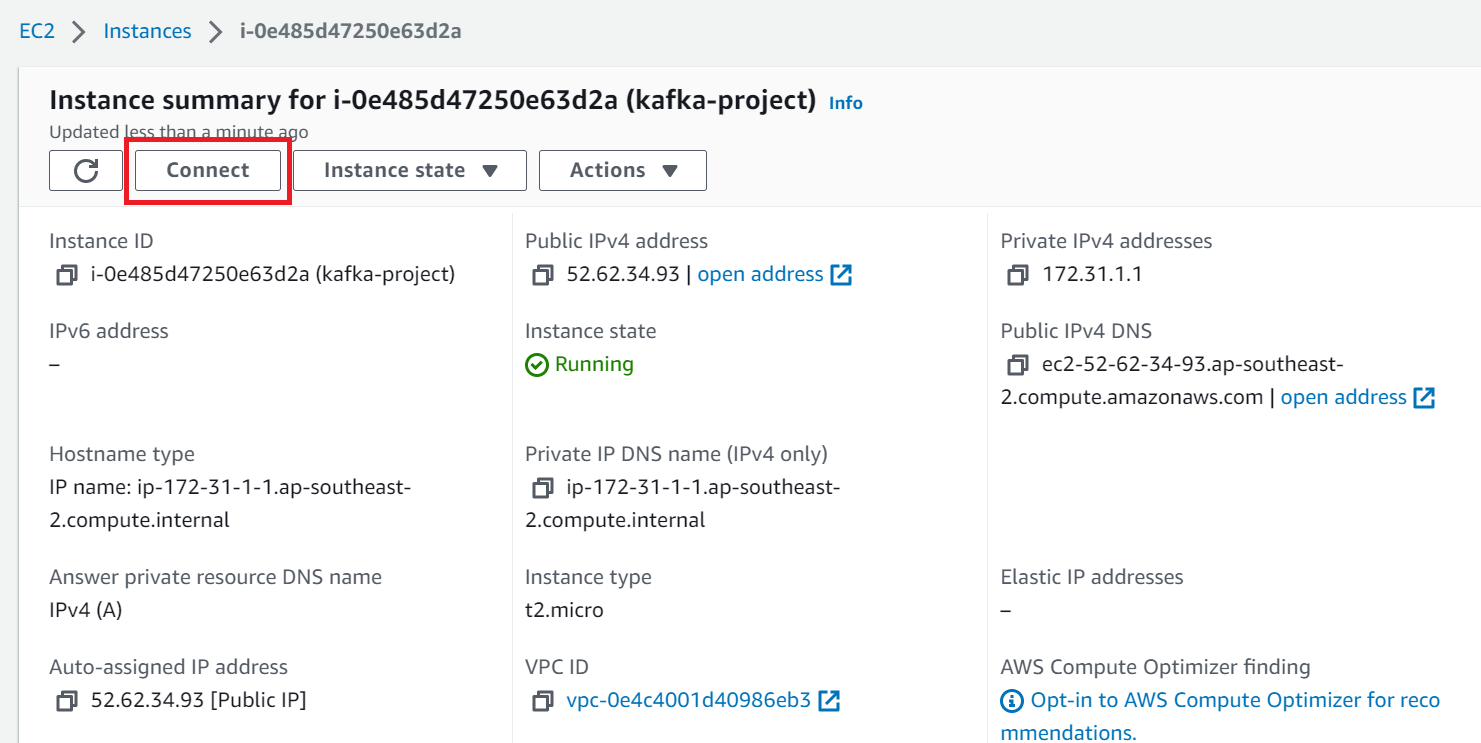
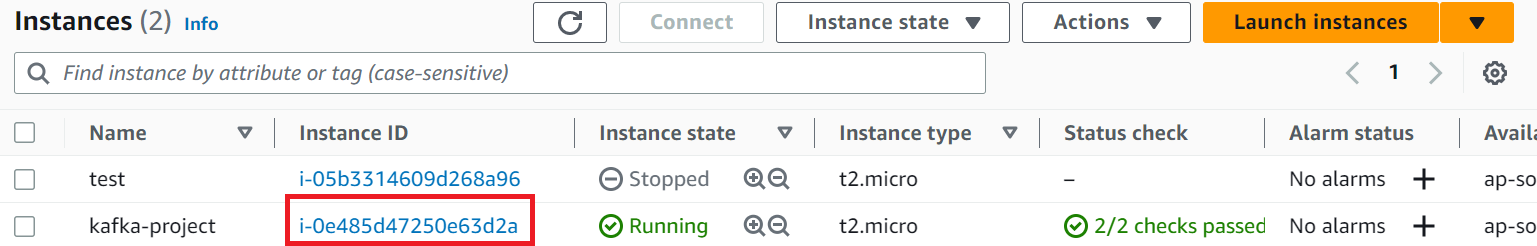
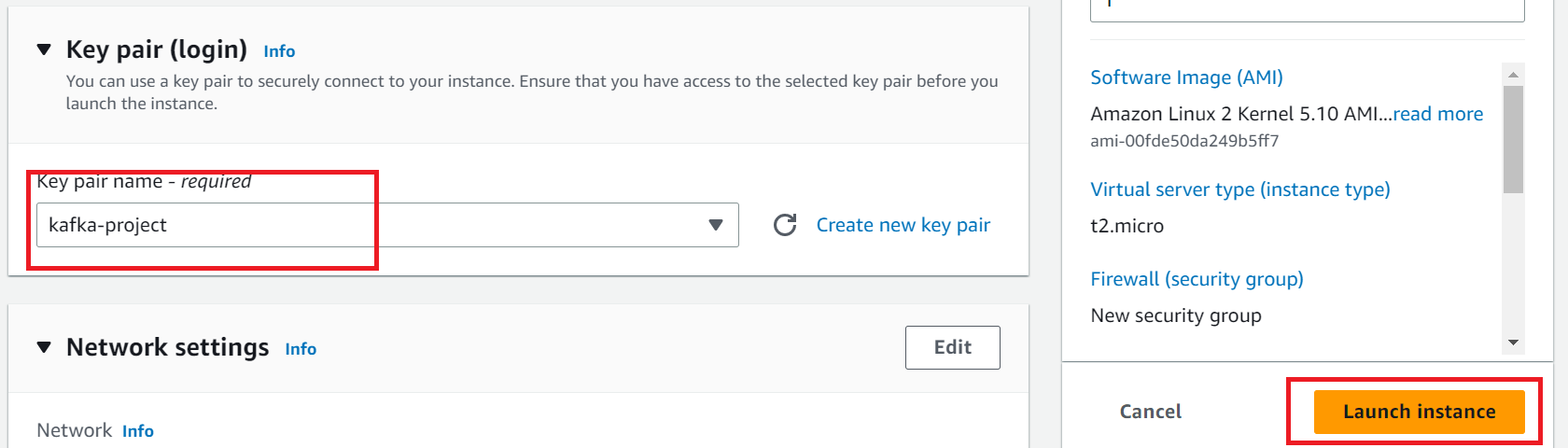
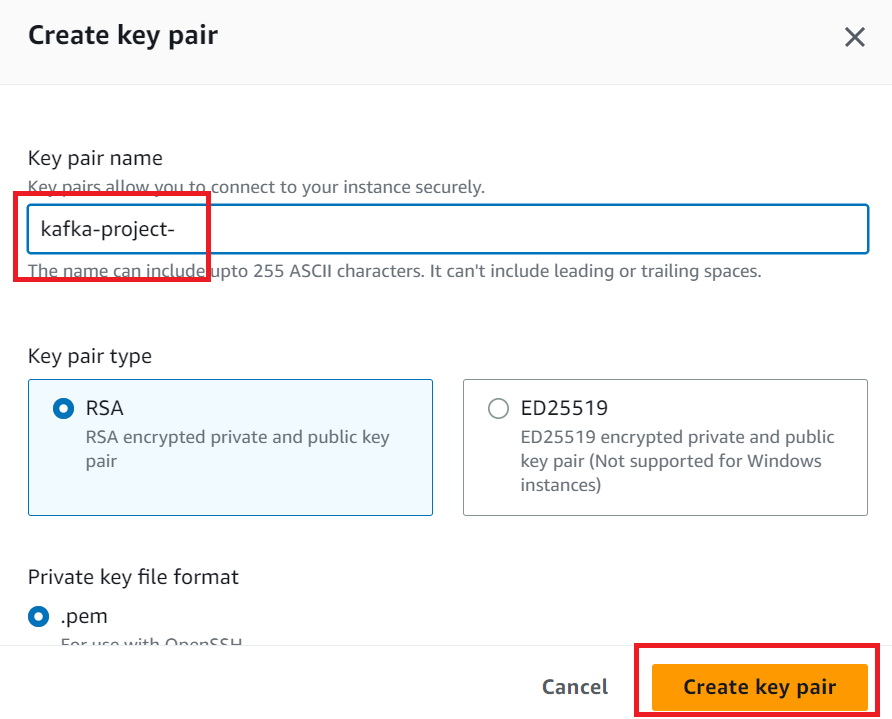
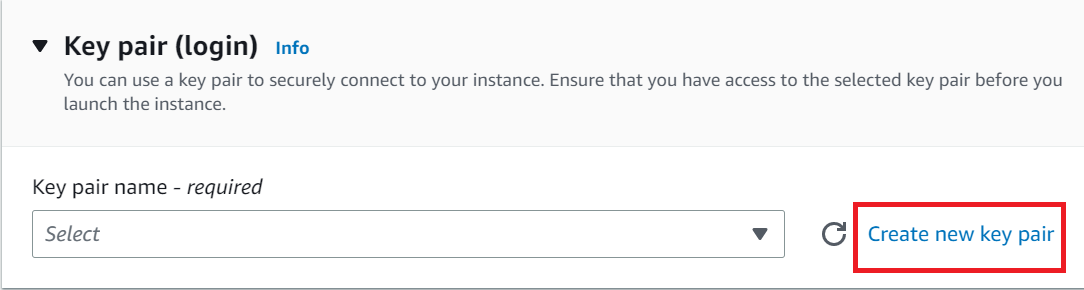
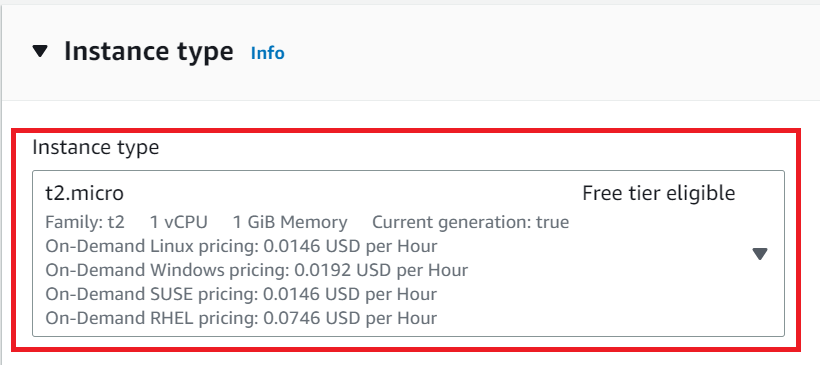
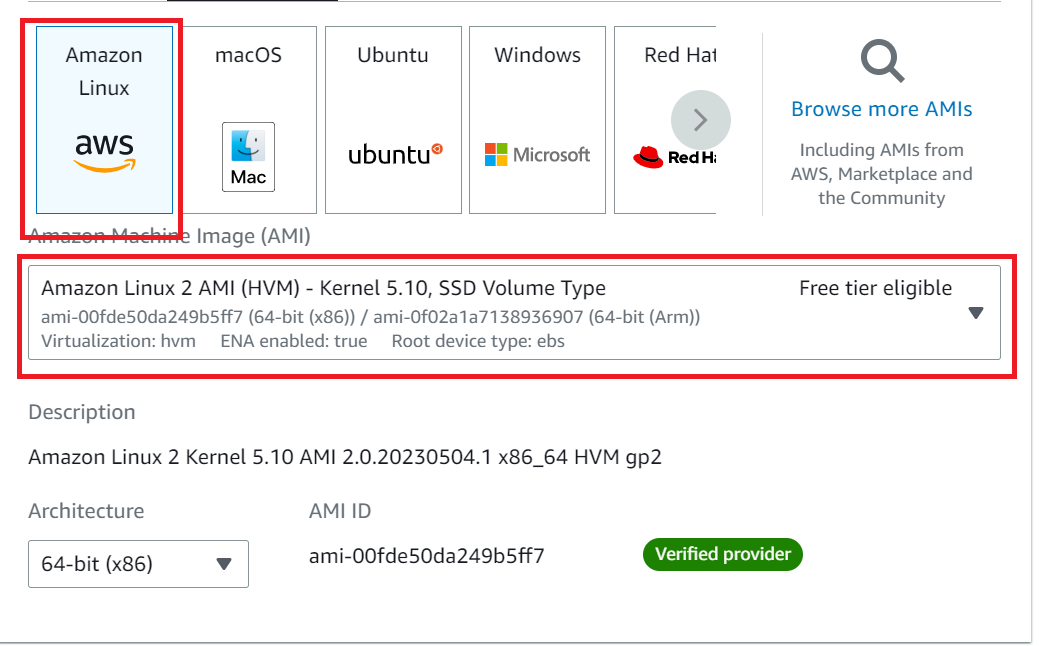
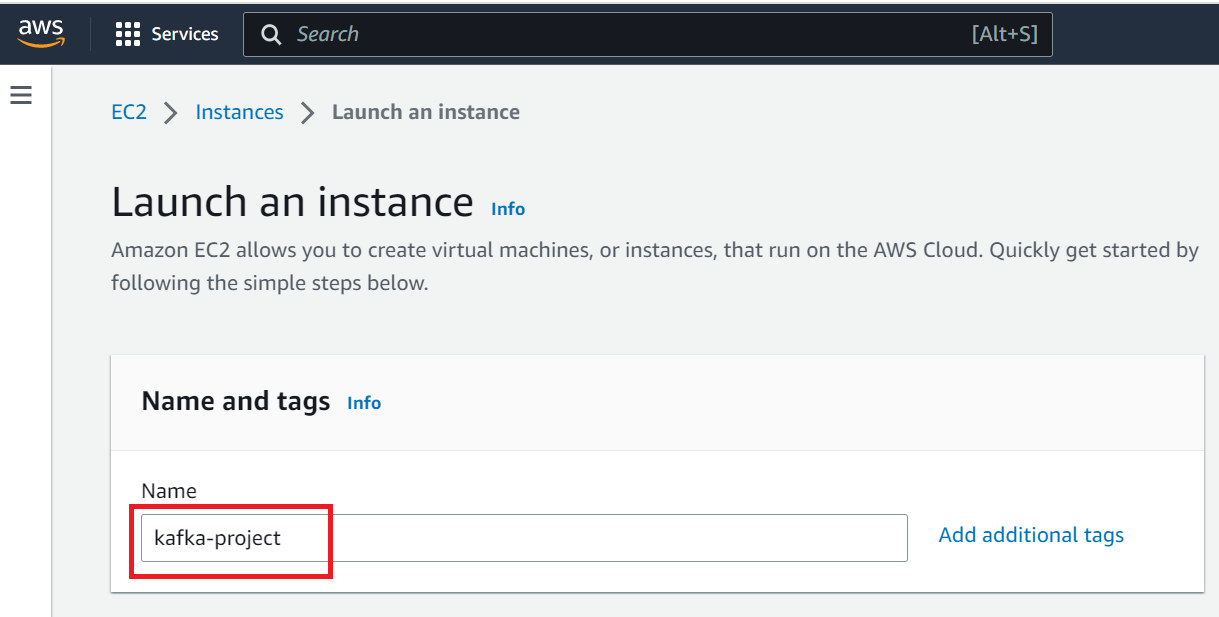
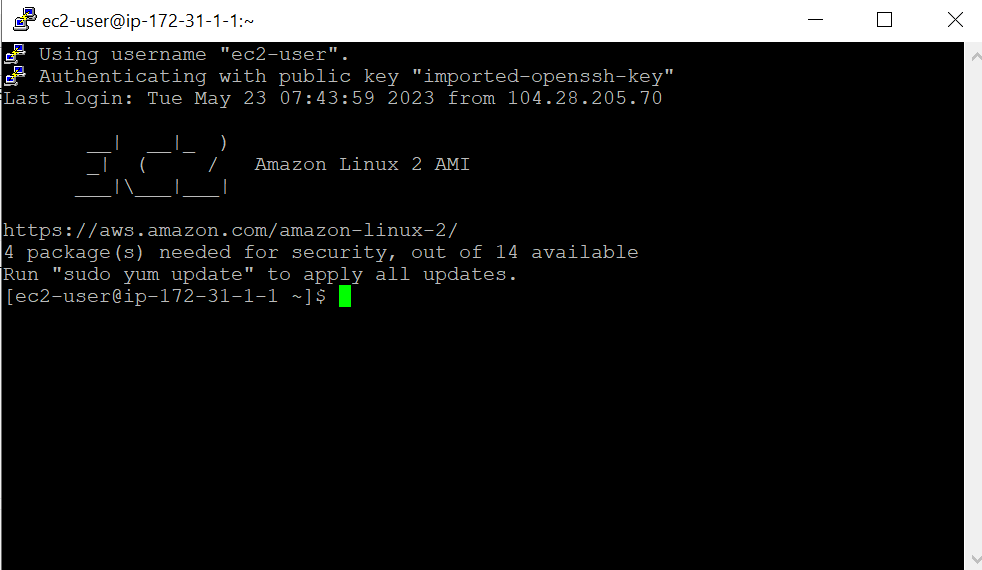
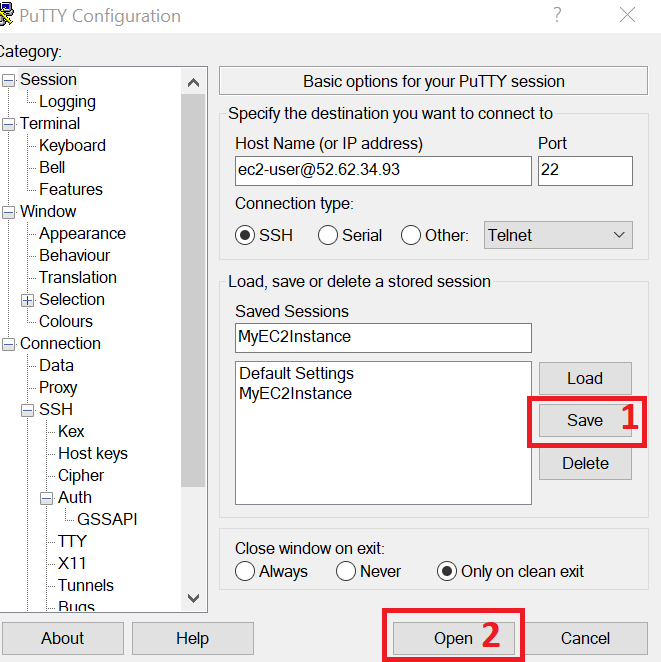
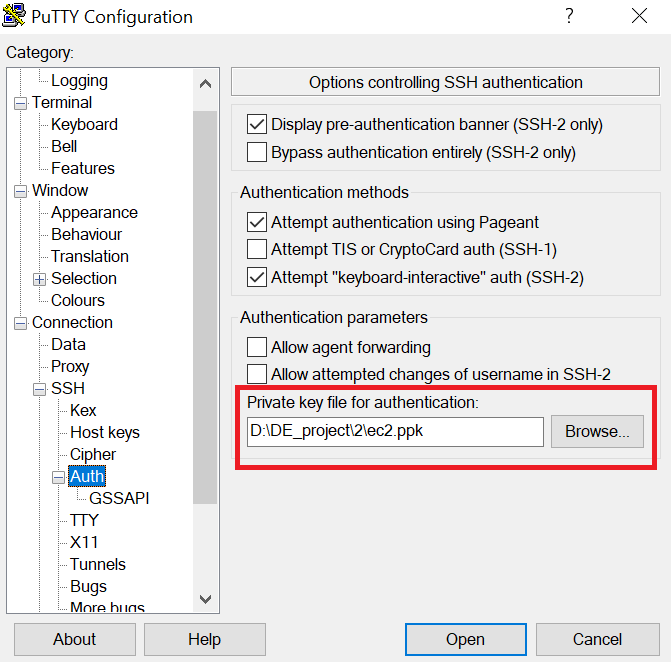
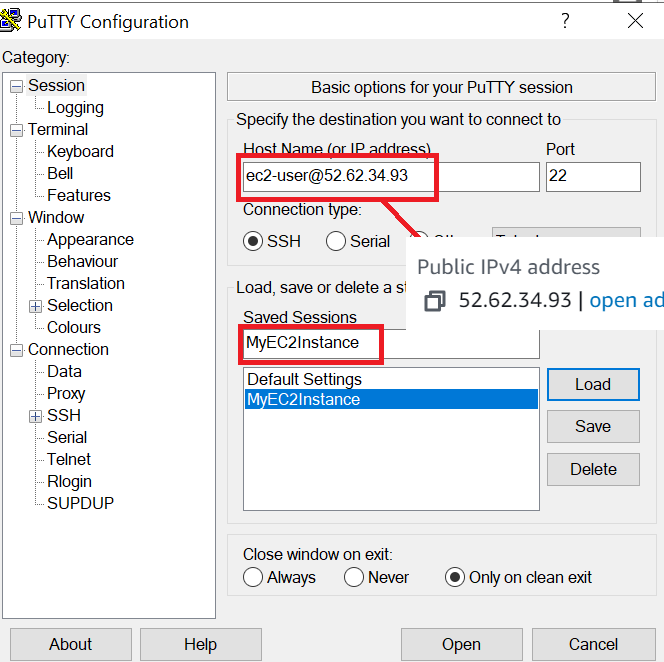
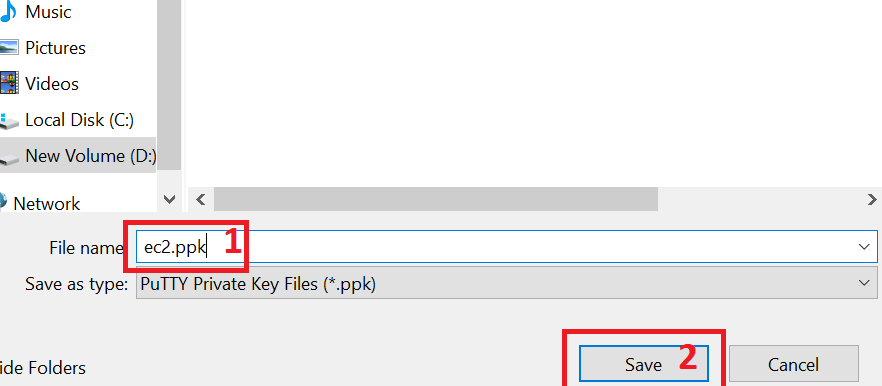
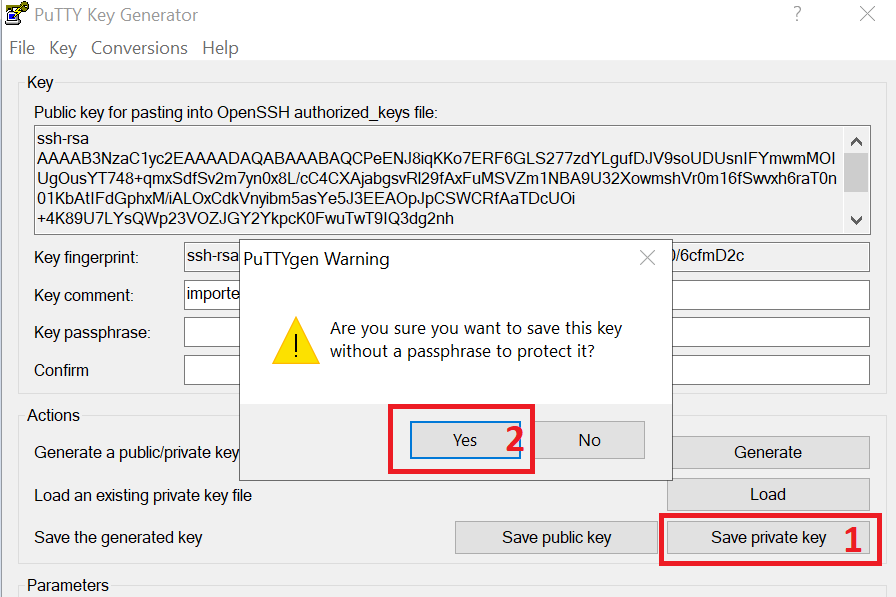
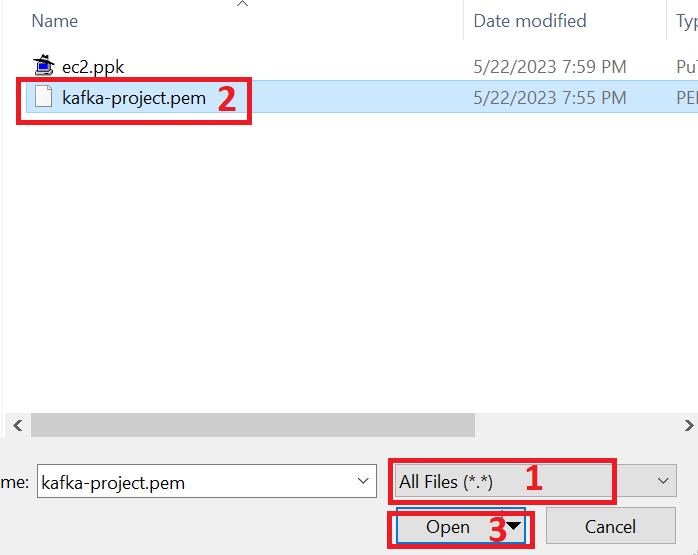
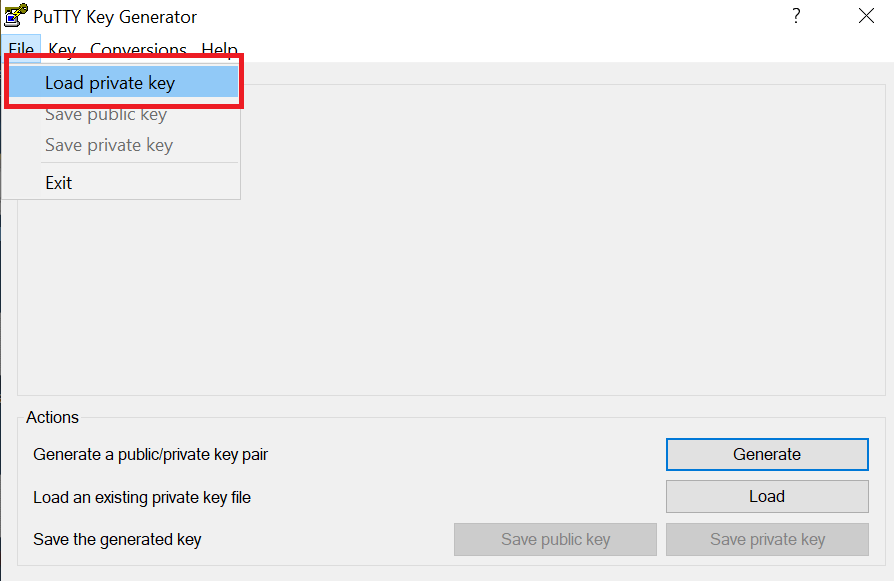
Launch instance EC2



Connect to your EC2 instance

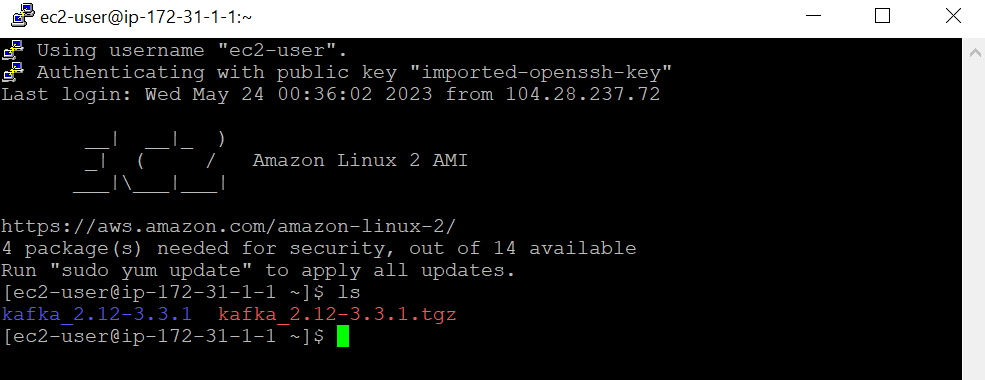


Install Kafka:

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

wget https://downloads.apache.org/kafka/3.3.1/kafka\_2.12-3.3.1.tgz

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



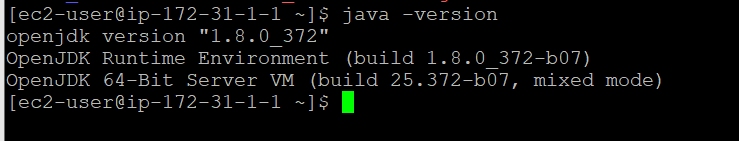
Install Java:

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

java -version

sudo yum install java-1.8.0-openjdk

java -version

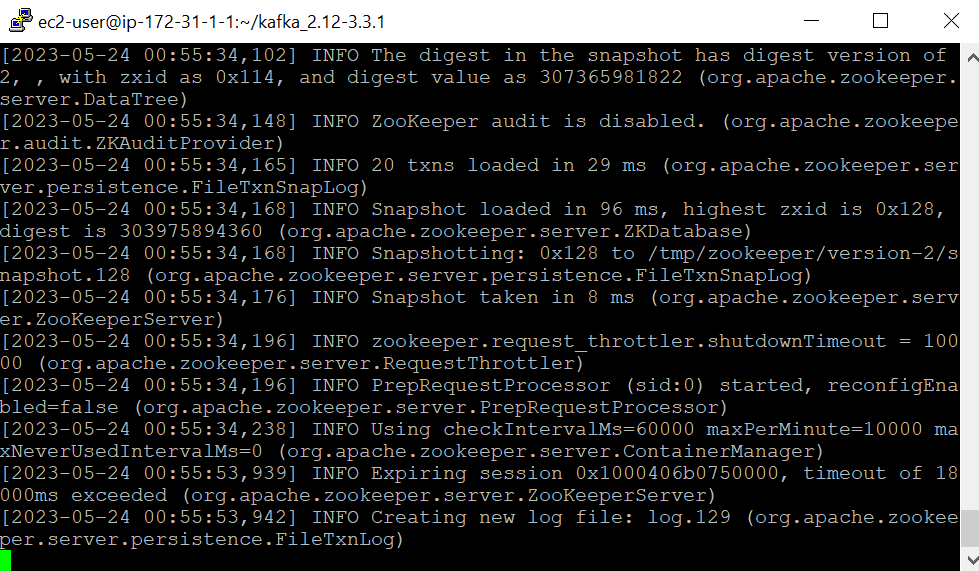


Start Zoo-keeper:

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

cd kafka\_2.12-3.3.1

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



Open another window to start kafka

But first ssh to to your ec2 machine as done above

Start Kafka-server:

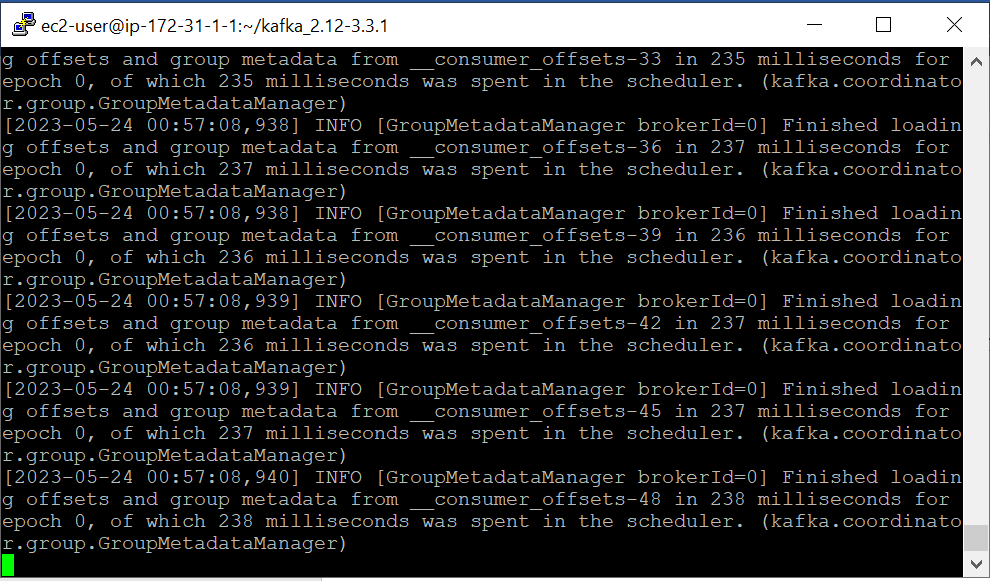
----------------------------------------

Duplicate the session & enter in a new console --

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

cd kafka\_2.12-3.3.1

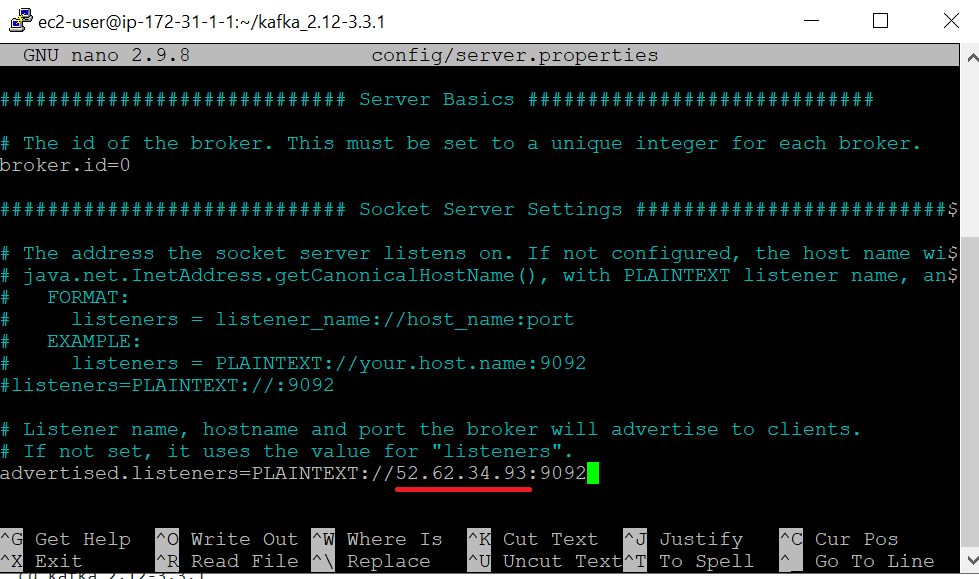
bin/kafka-server-start.sh config/server.properties



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



Ctrl X – Y – Enter (Save)

Then Restart Zoo-keeper and Kafka-server

Create the topic:

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

Duplicate the session & enter in a new console --

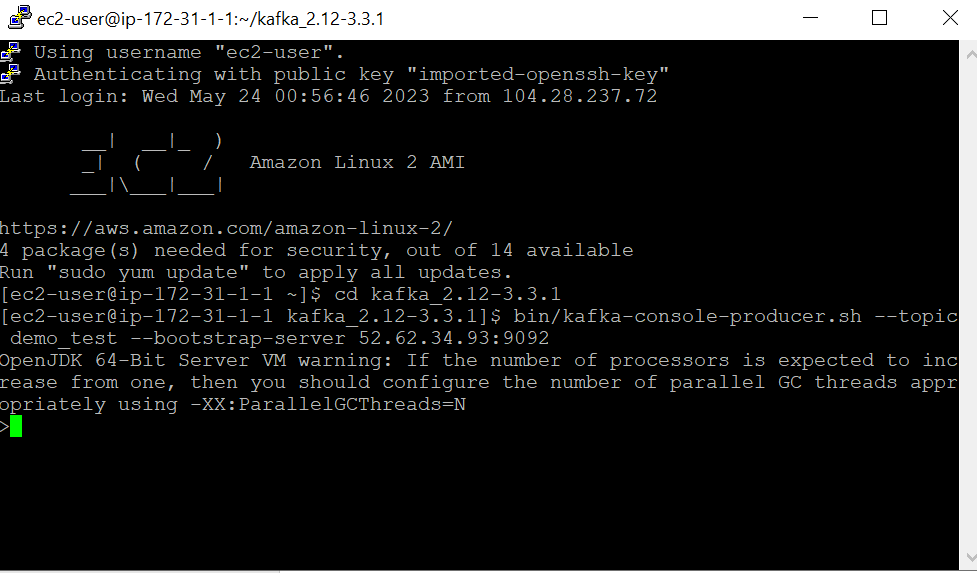
cd kafka\_2.12-3.3.1

bin/kafka-topics.sh --create --topic demo\_test --bootstrap-server 52.62.34.93:9092 --replication-factor 1 --partitions 1

Start Producer:

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

bin/kafka-console-producer.sh --topic demo\_test --bootstrap-server 52.62.34.93:9092



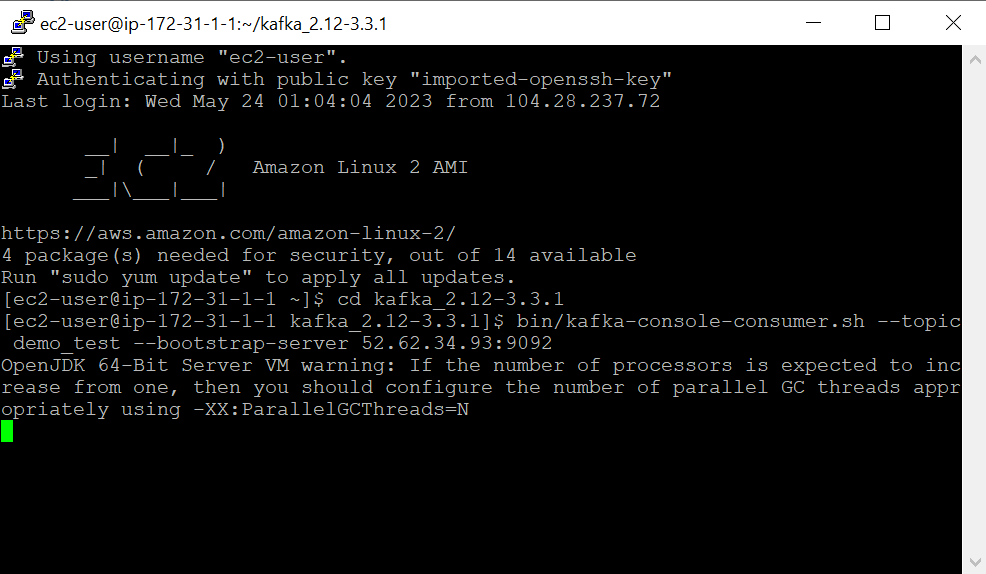
Start Consumer:

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

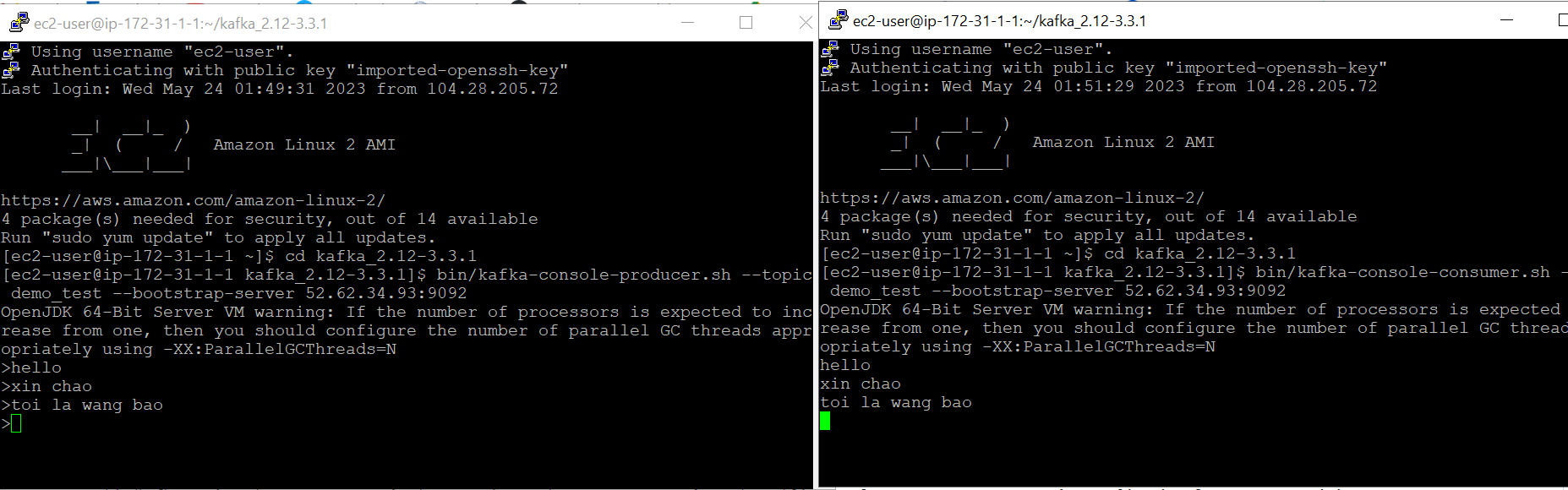
Duplicate the session & enter in a new console --

cd kafka\_2.12-3.3.1

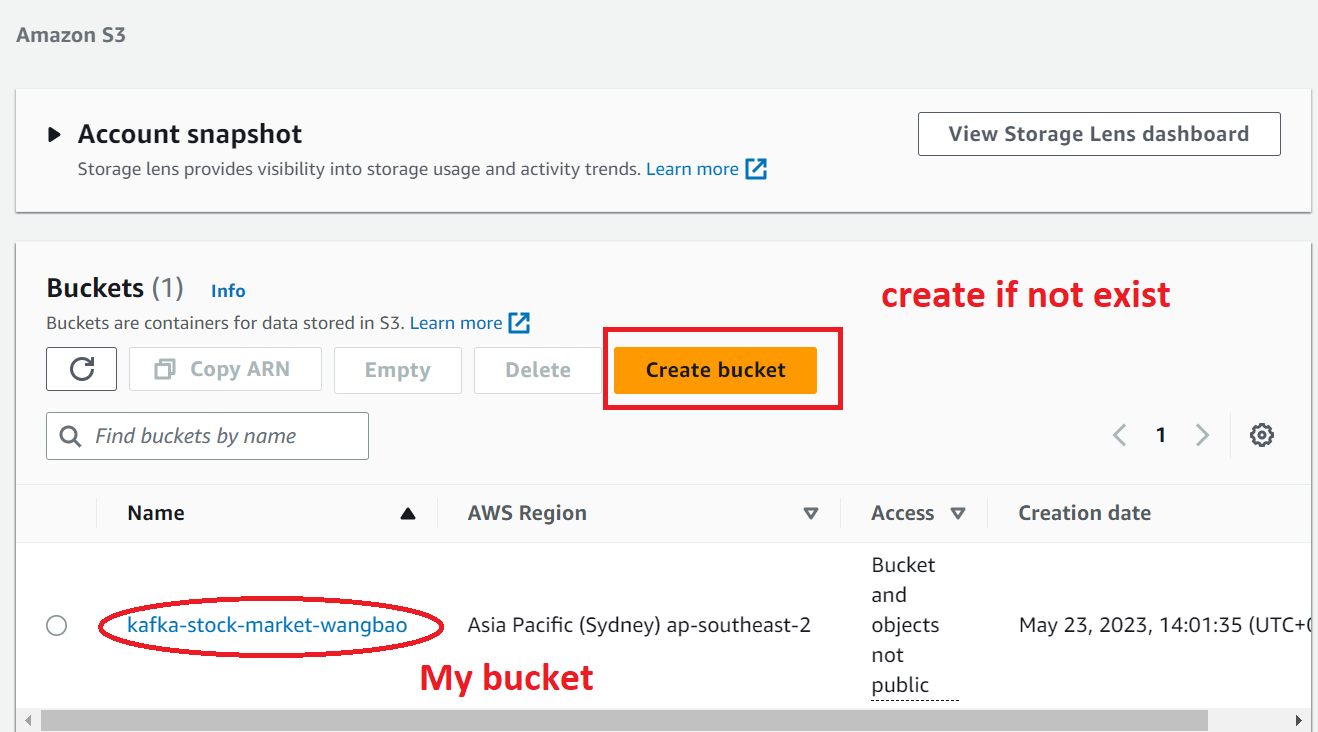
bin/kafka-console-consumer.sh --topic demo\_test --bootstrap-server 52.62.34.93:9092



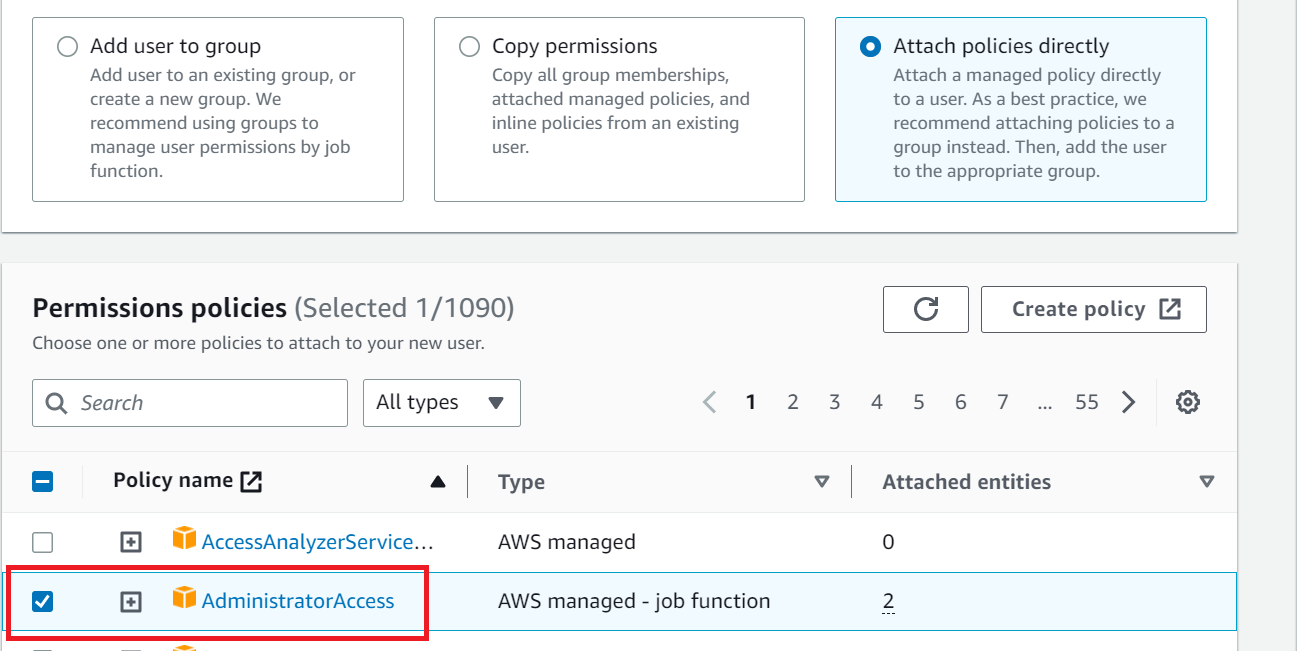
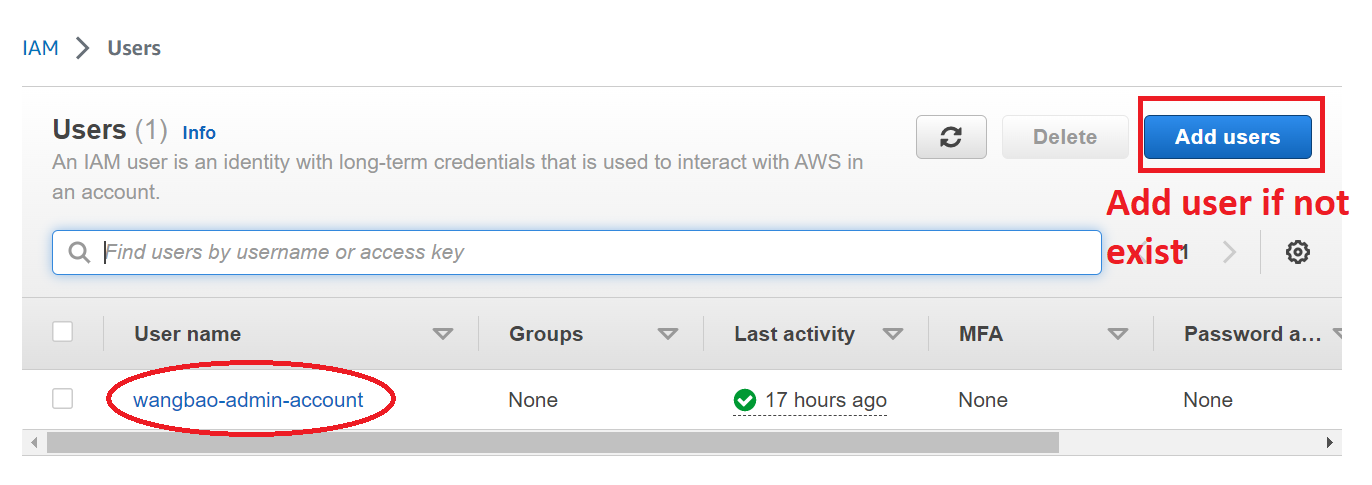
Listen



S3

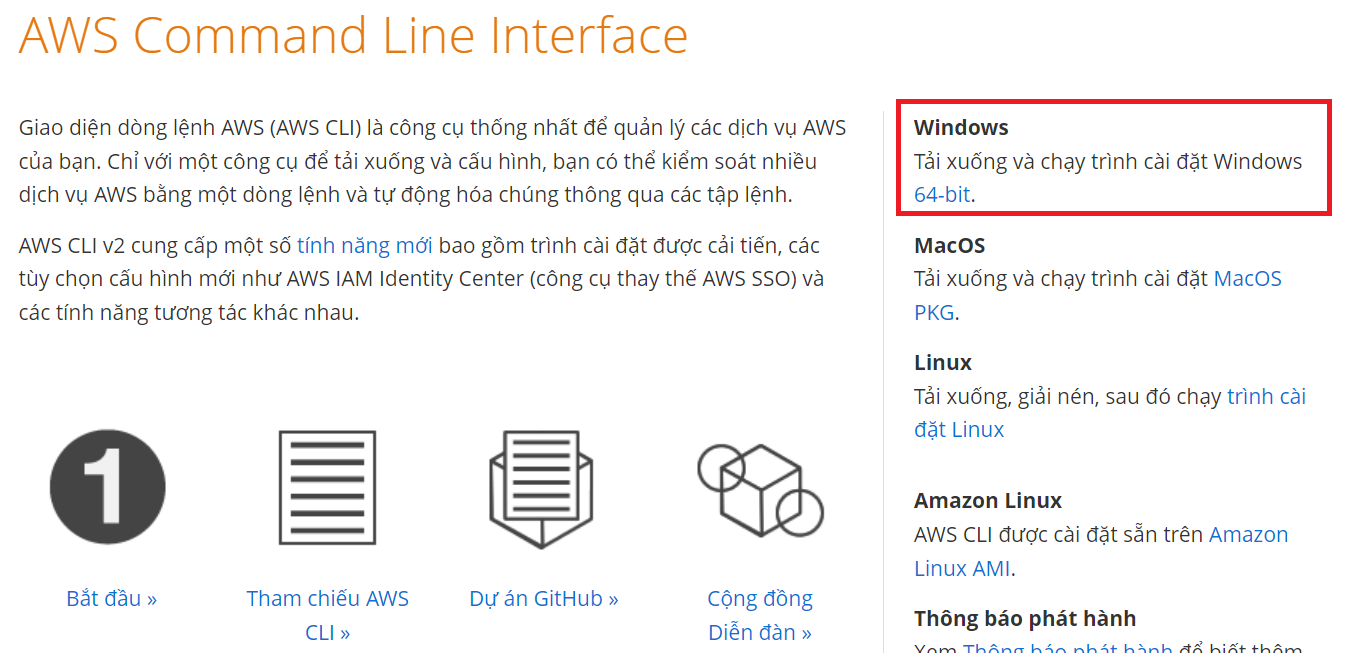


IAM

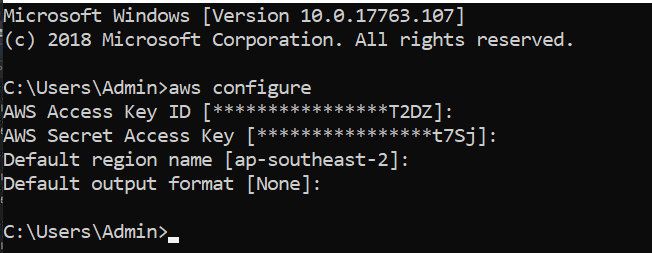


Create Access key -> Dowload csv file (access key)

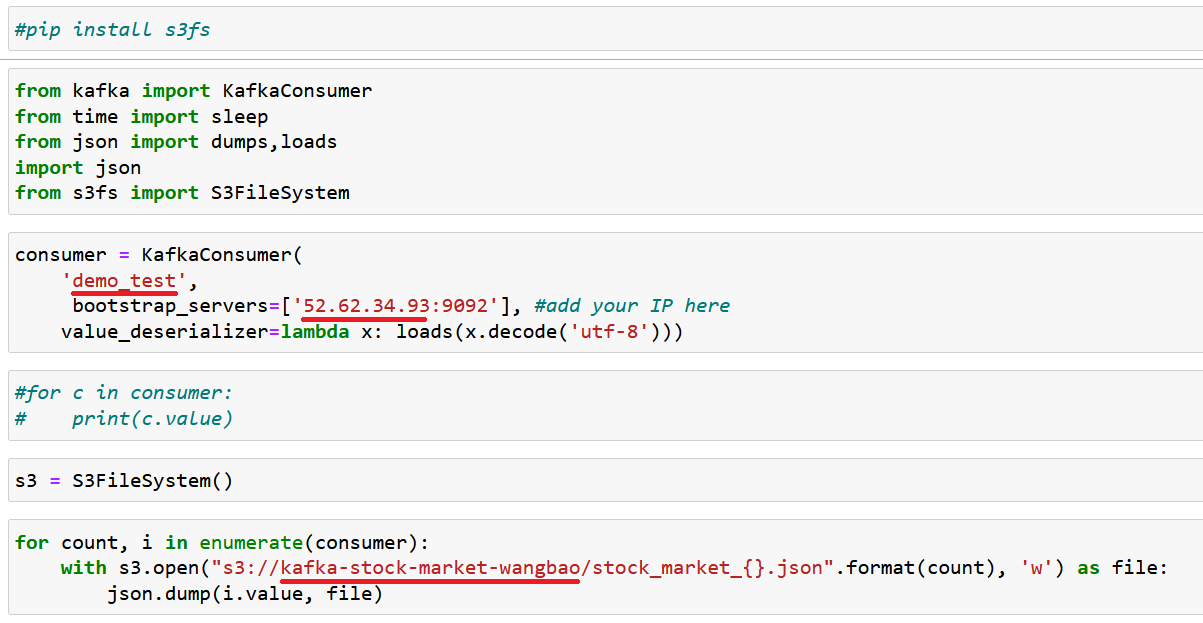
Dowload AWS CLI



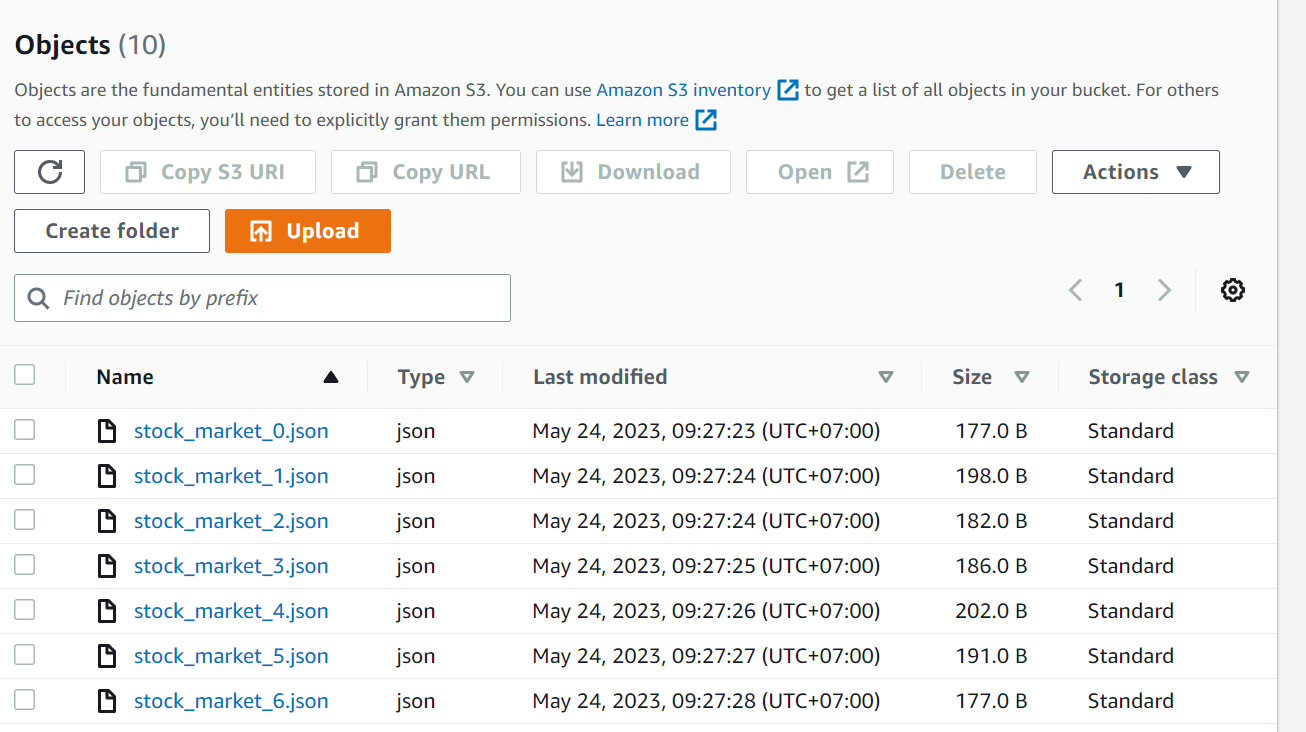
Set up



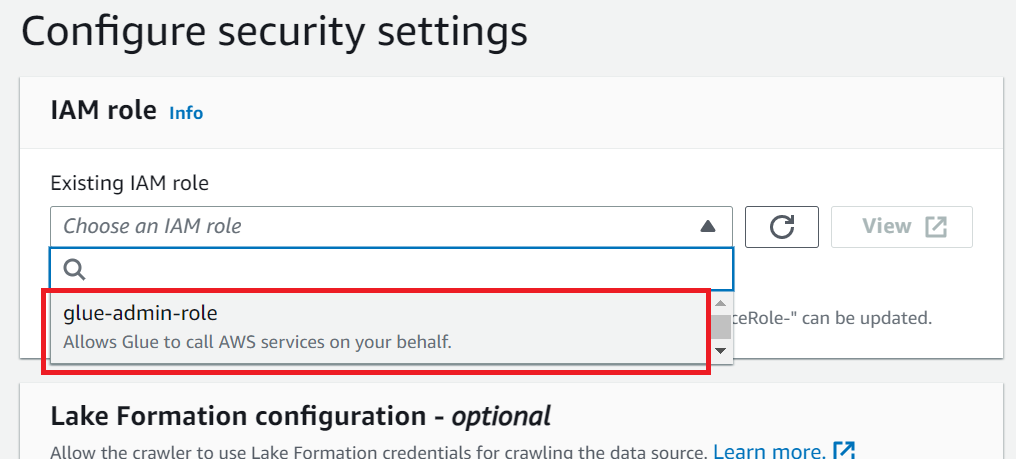
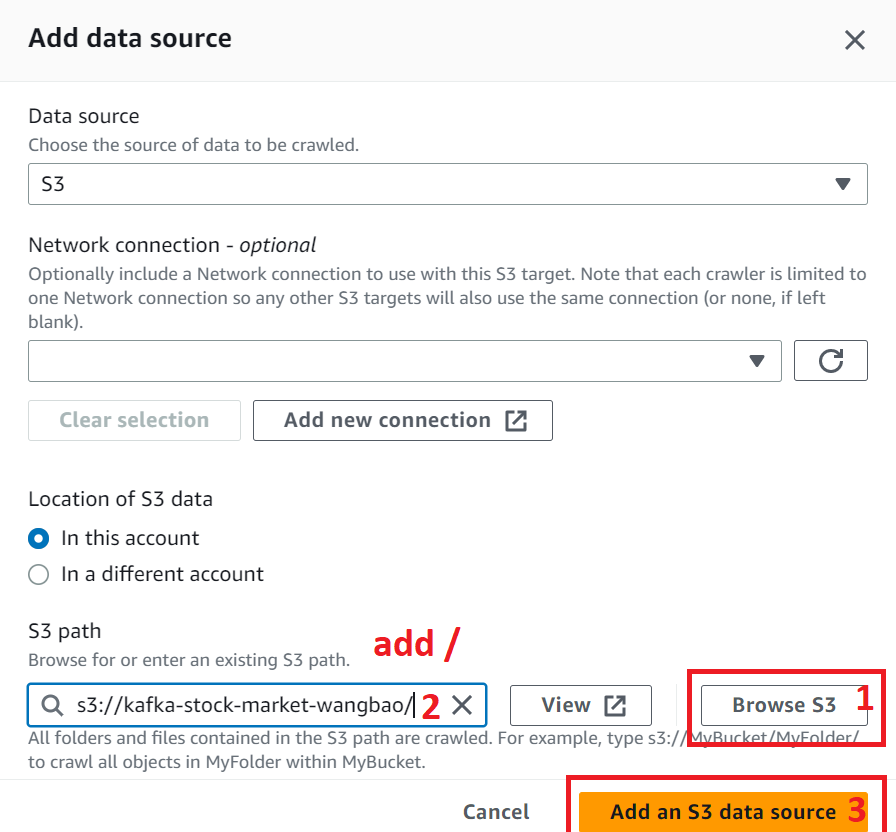
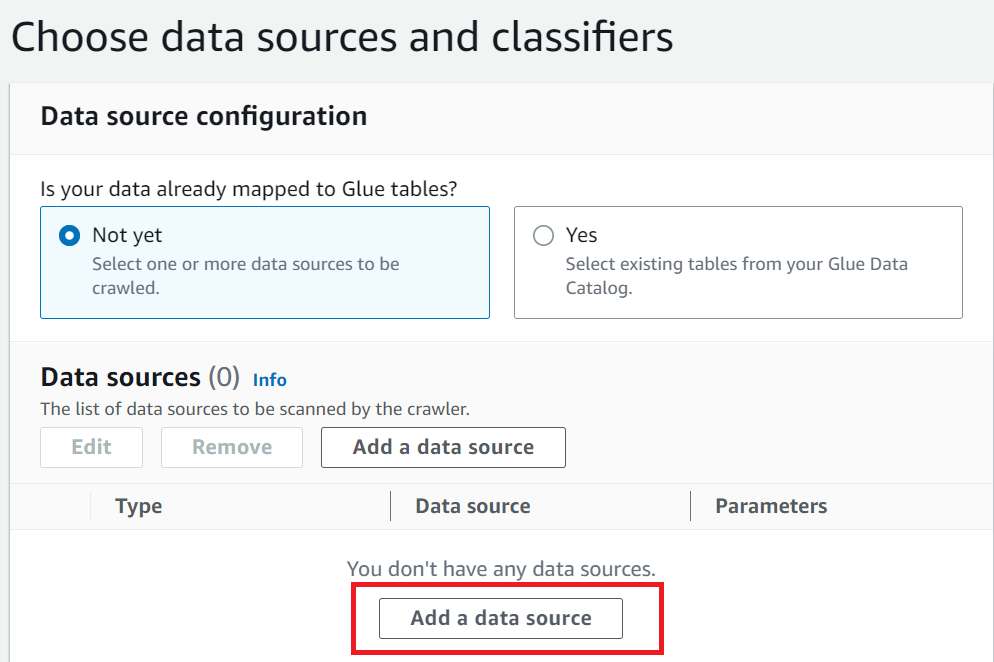
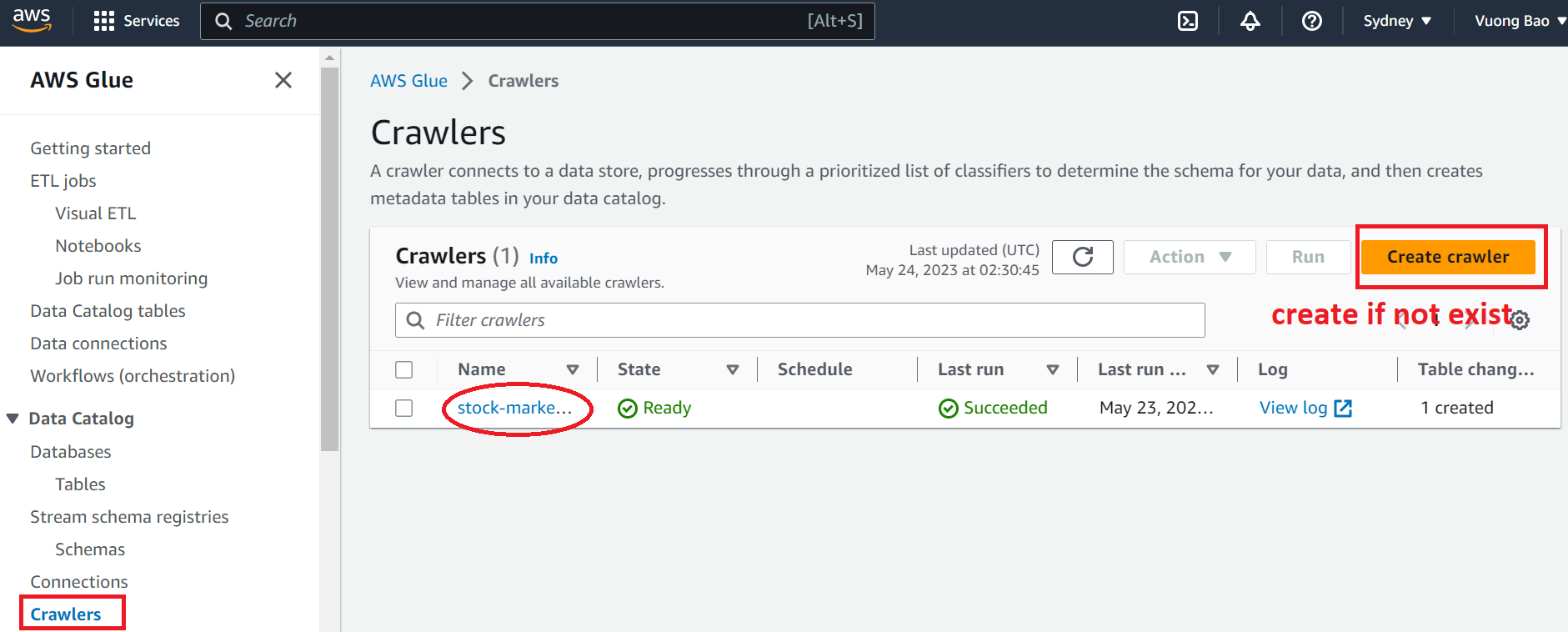
Jupyter

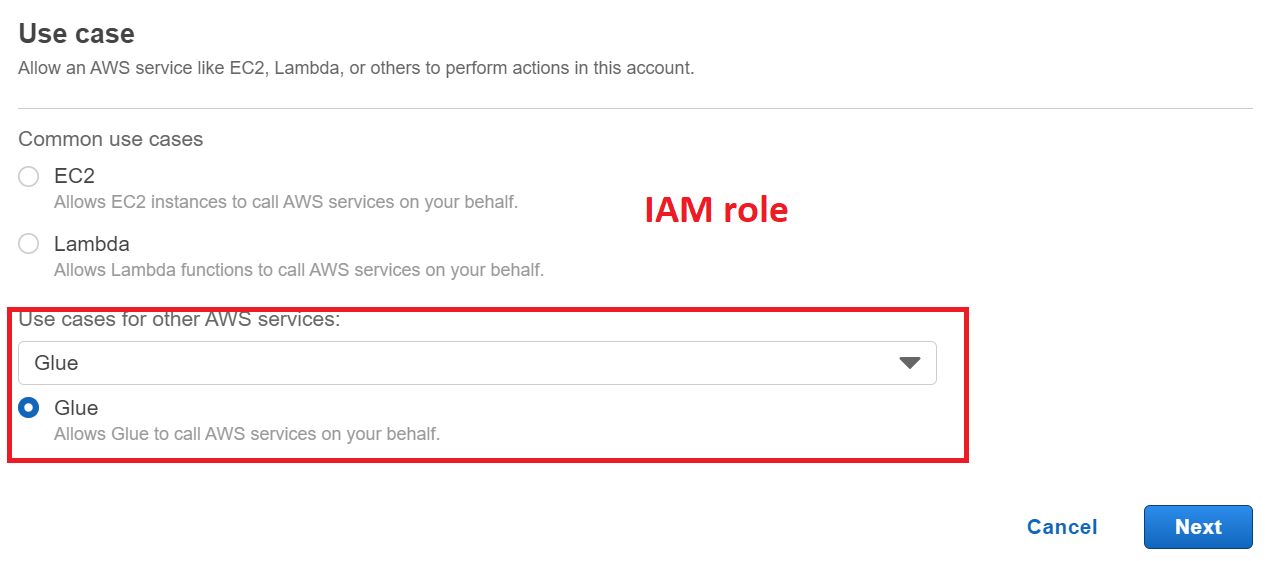


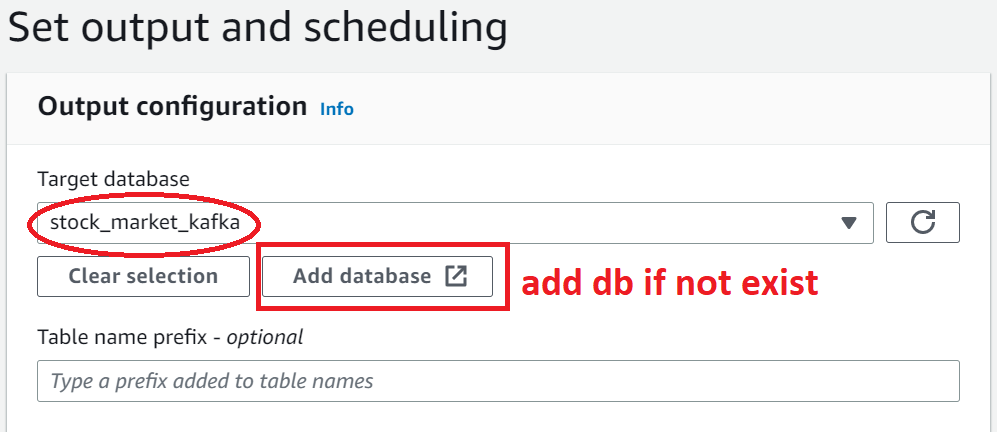
S3



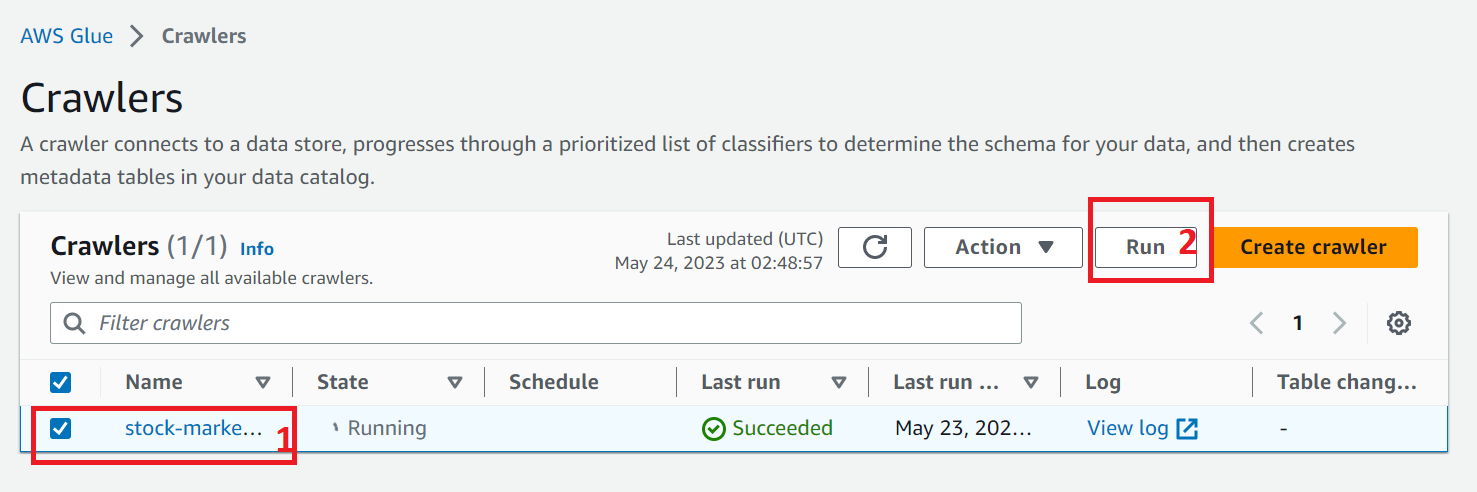
Glue



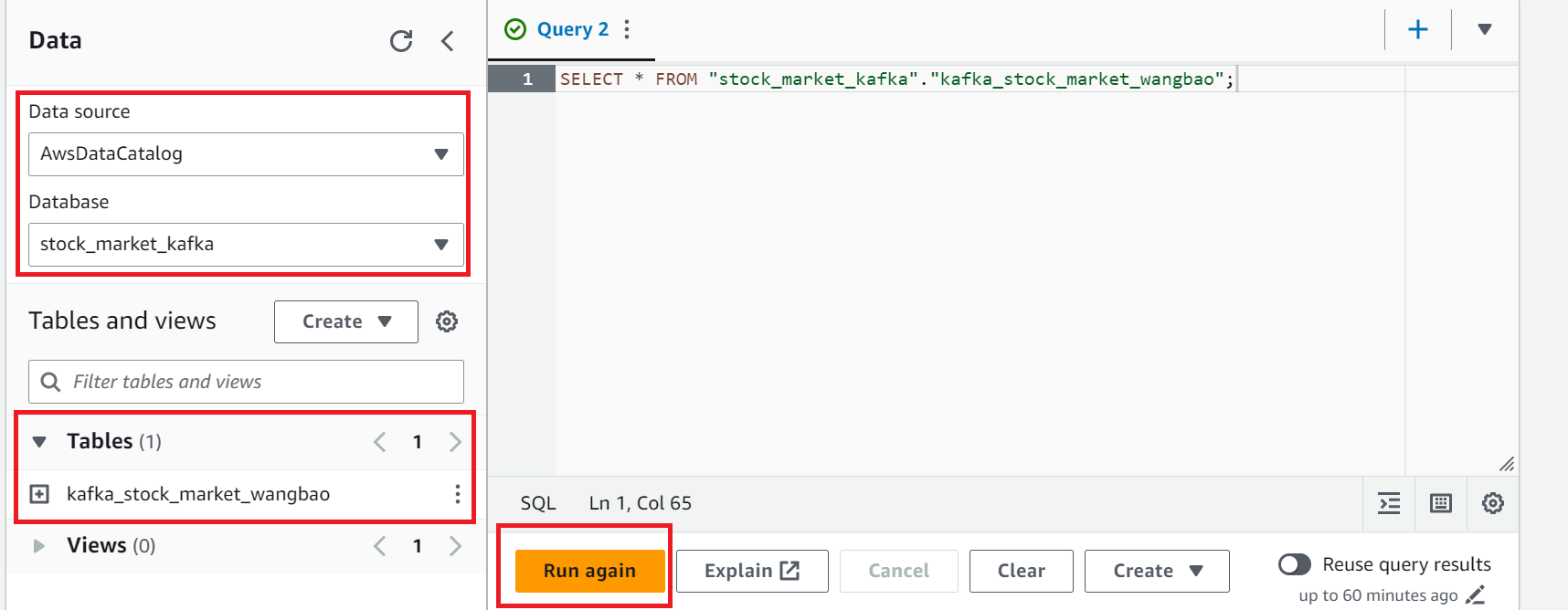


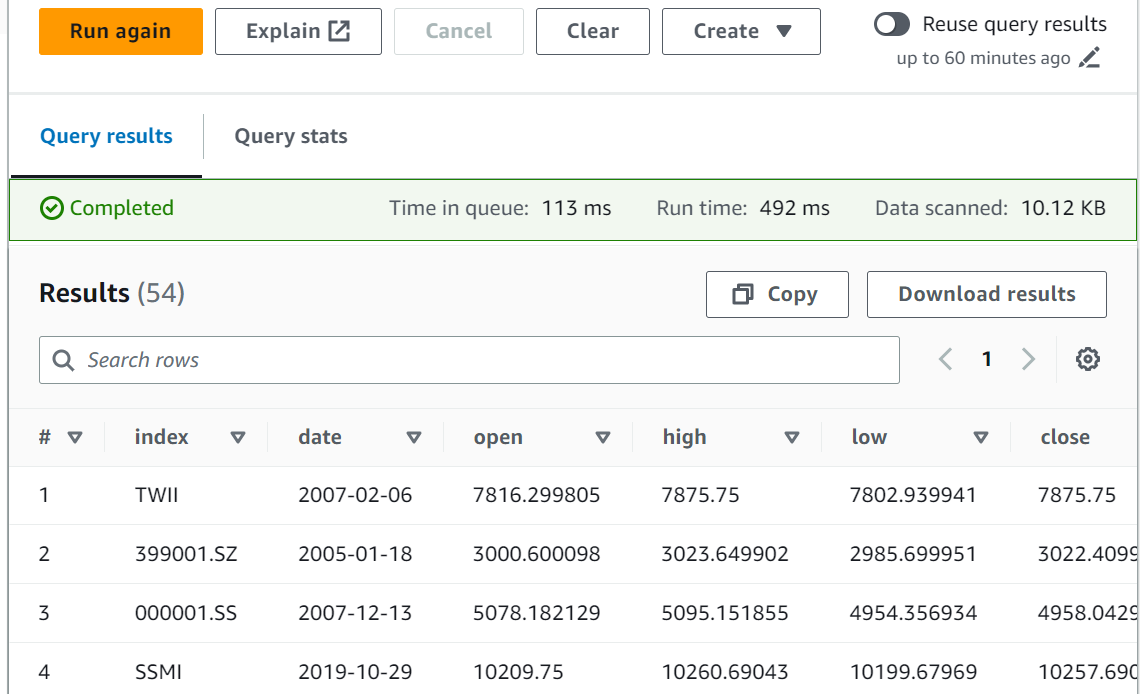


Next -> Create Crawler



Athena





Nếu không thể query. Thực hiện bước sau

