

# CREATING EMR CLUSTER AWS (INSTRUCTIONS)

## 1. AWS

Go to the official web site of Amazon Web Services <https://aws.amazon.com/> and click on **Sign in to the Console button** (see Figure 1)

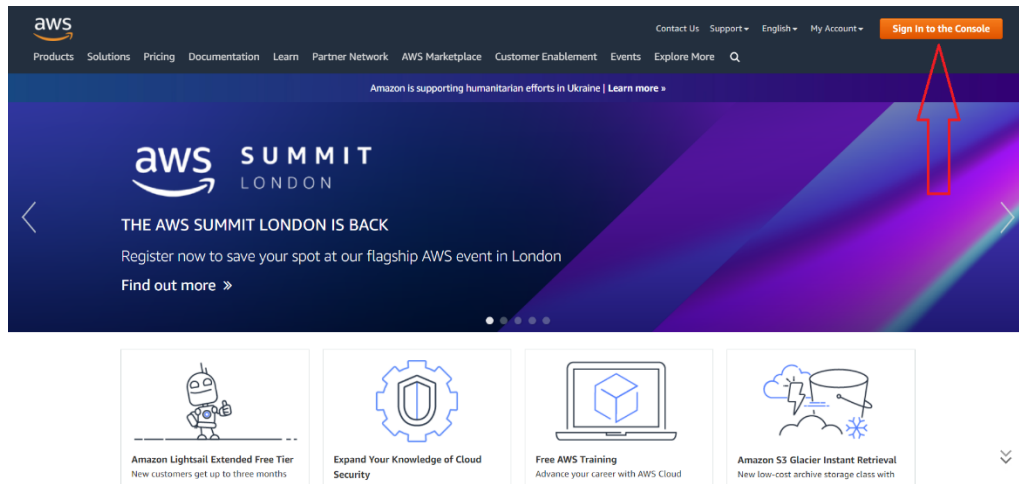


Figure 1

## 2. Sign in:

Use your personal information which are account id, username, and password to sign in to your account. (see Figure 2)

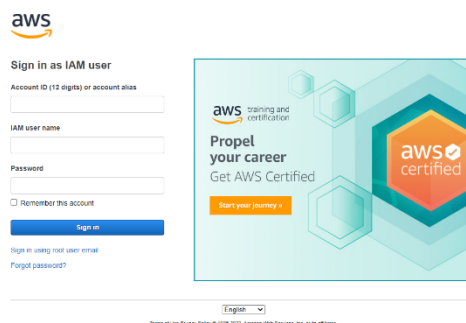


Figure 2

### 3. Create EC2 keypair:

EC2 keypair is used while creating EMR, thus it must be initially created. In order to create EC2 follow the steps from A to I.

A. First, write ec2 inside the search bar. Then, click on EC2 icon as it is shown in figure 3.

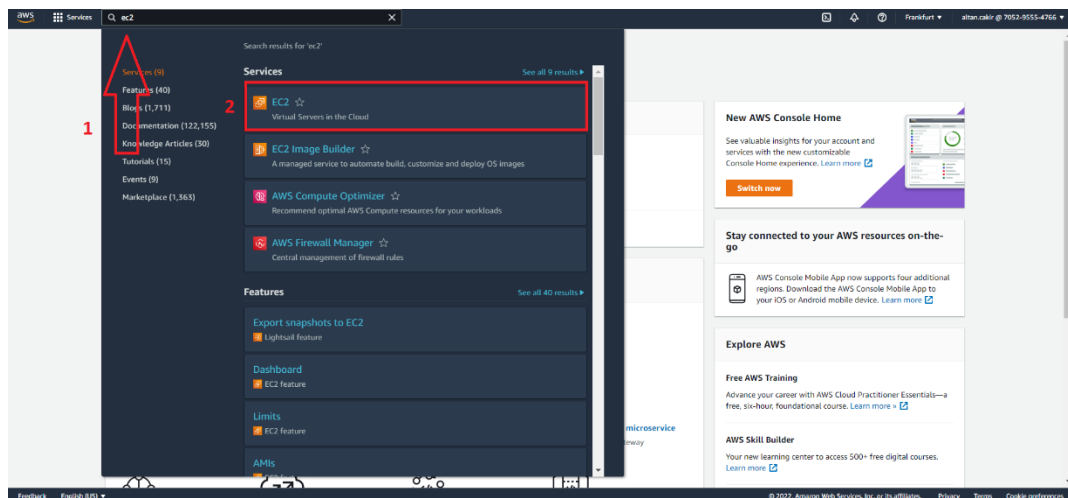


Figure 3

B. Go to EC2 dashboard, and click on launch instance button as it is clearly seen in figure 4

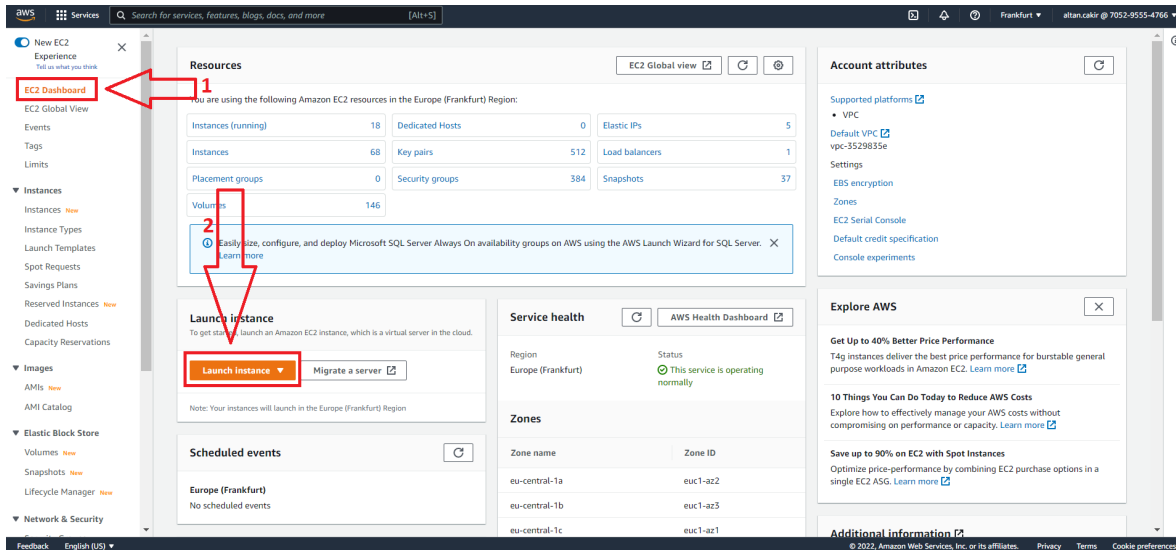


Figure 4

C. Select your machine operating system by clicking on select button that can be found next to the machines. (see Figure5)

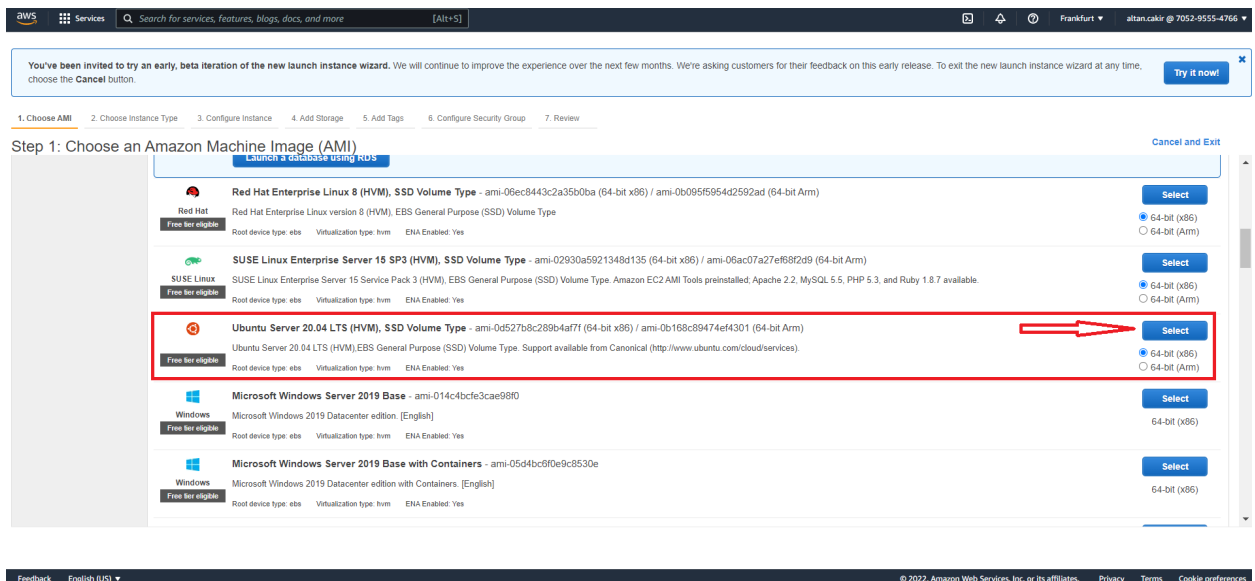


Figure5

D. Select the amount of resources of that machine by clicking on the square at the left side of each type. Then, click on Review and Lunch button below as it is explained in figure 6. Some other configurations such as security and storage can be modified by clicking on Next. Configure Instance Details.

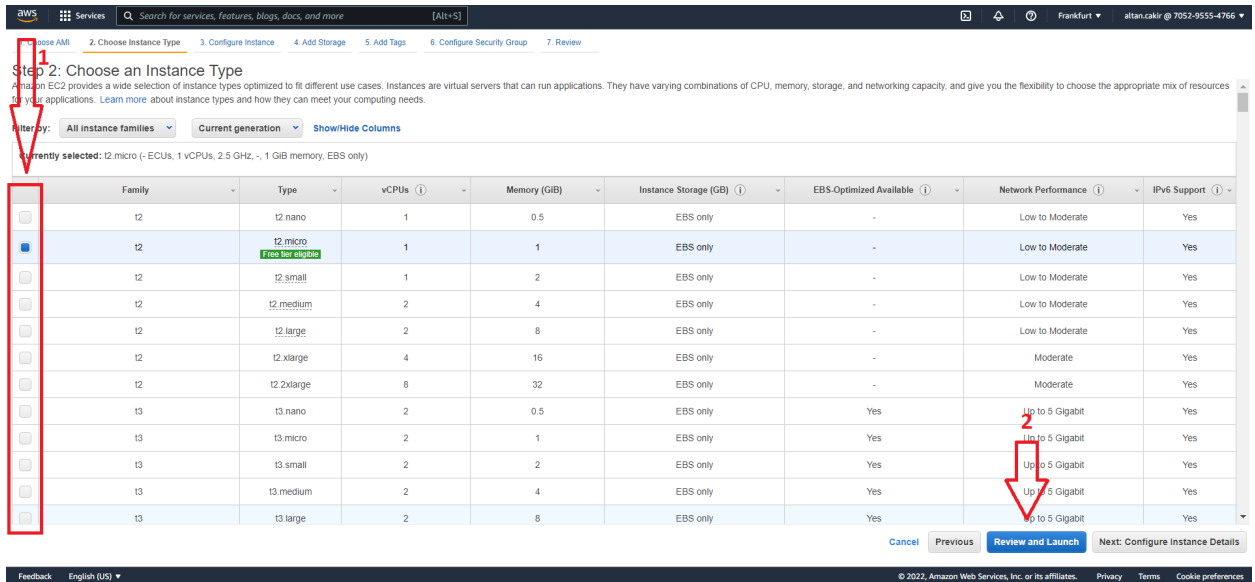


Figure 6

E. Click on lunch to go to the final step as it is clear in Figure 7.

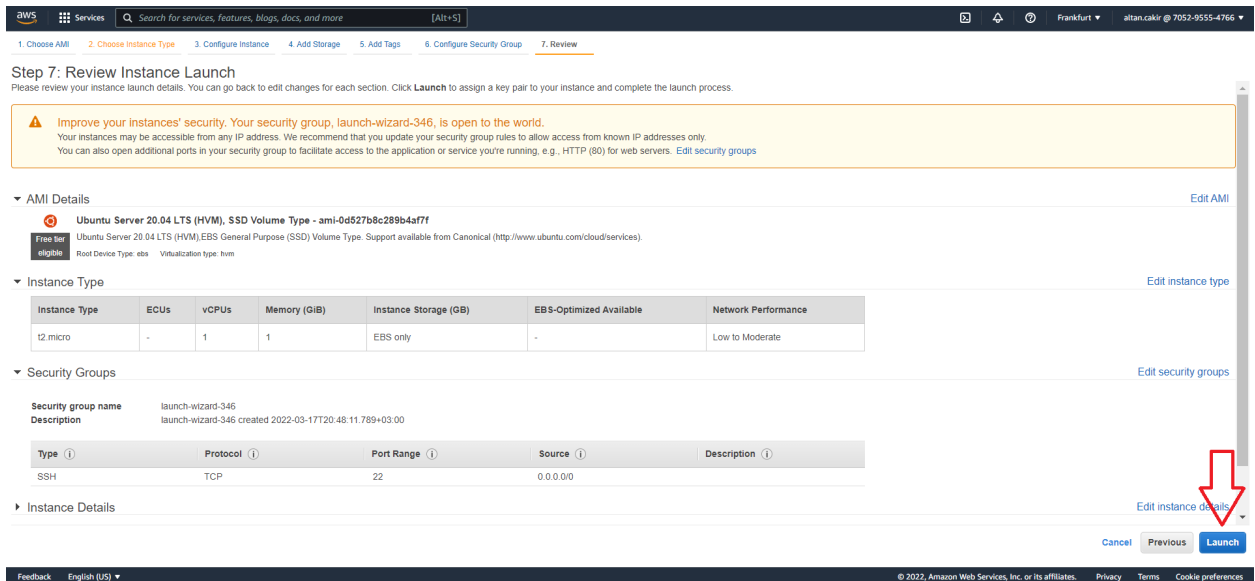


Figure 7

F. From the first bar select create a new key pair. After that, give a name to your key pair. Then, click on Download key pair button. Eventually, lunch the instance by clicking on Lunch instance button. (see Figure 8)

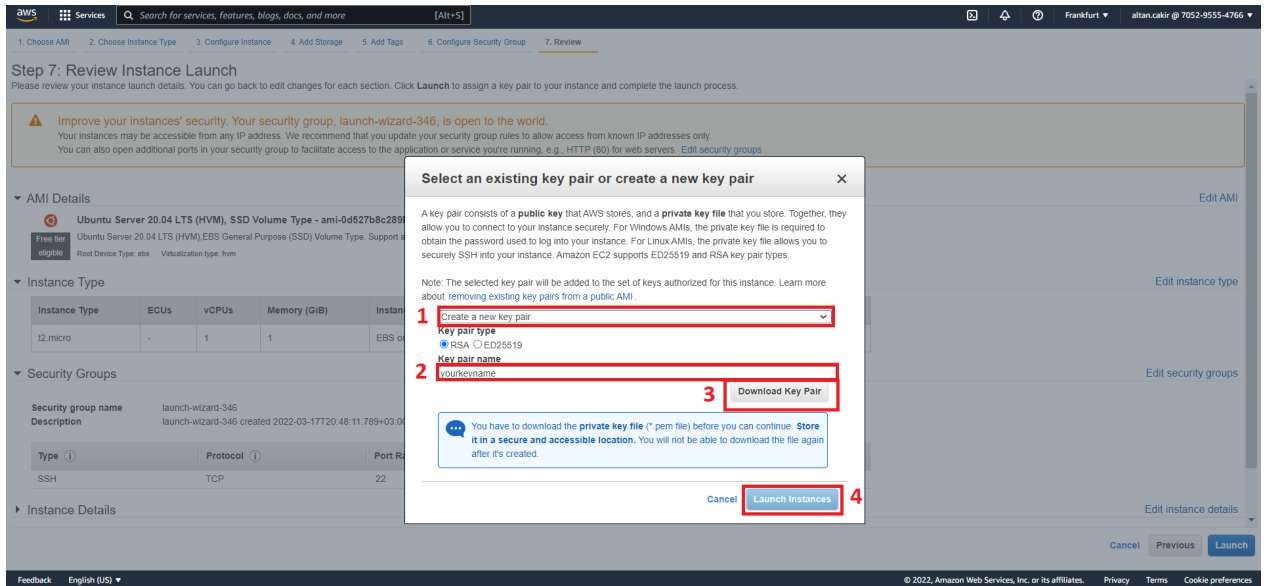
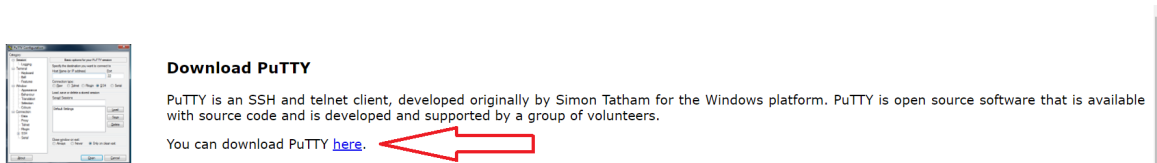
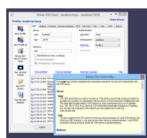


Figure 8

G. Your key pair must be in ppk format. One way to convert file from pem to ppk format is using Putty program. Download Putty from the following link <https://www.putty.org/> (see Figure 9).



Below suggestions are independent of the authors of PuTTY. They are *not* to be seen as endorsements by the PuTTY project.

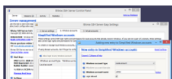


#### Bitvise SSH Client

Bitvise SSH Client is an SSH and SFTP client for Windows. It is developed and supported professionally by Bitvise. The SSH Client is robust, easy to install, easy to use, and supports all features supported by PuTTY, as well as the following:

- graphical SFTP file transfer;
- single-click Remote Desktop tunneling;
- auto-reconnecting capability;
- dynamic port forwarding through an integrated proxy;
- an FTP-to-SFTP protocol bridge.

Bitvise SSH Client is **free to use**. You can [download it here](#).



#### Bitvise SSH Server

Bitvise SSH Server is an SSH, SFTP and SCP server for Windows. It is robust, easy to install, easy to use, and works well with a variety of SSH clients, including Bitvise SSH Client, OpenSSH, and PuTTY. The SSH Server is developed and supported professionally by Bitvise.

Figure 9

H. Lucnh Puttygen, and upload your pem file that is generated in step F. (see Figure 10)

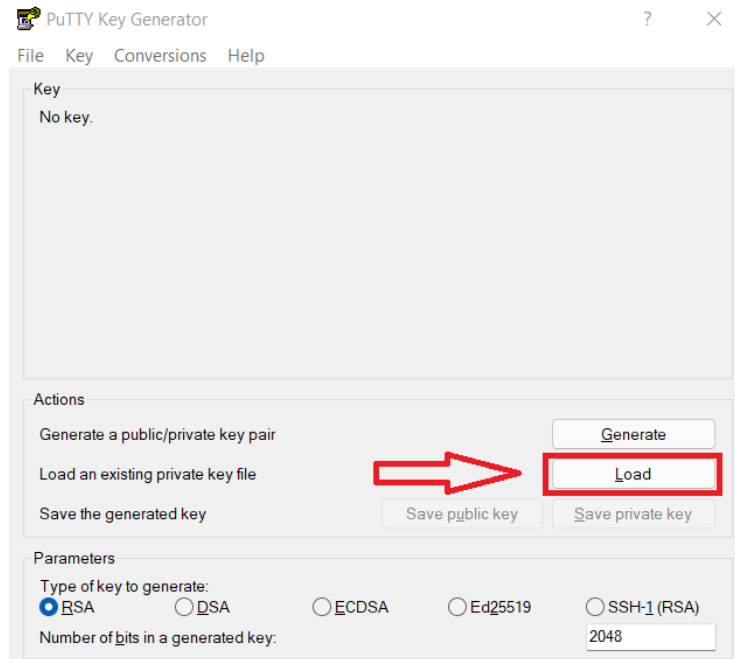


Figure10

- I. Save it as private key which is in ppk format by clicking on Save private key button as it is shown in figure 11.

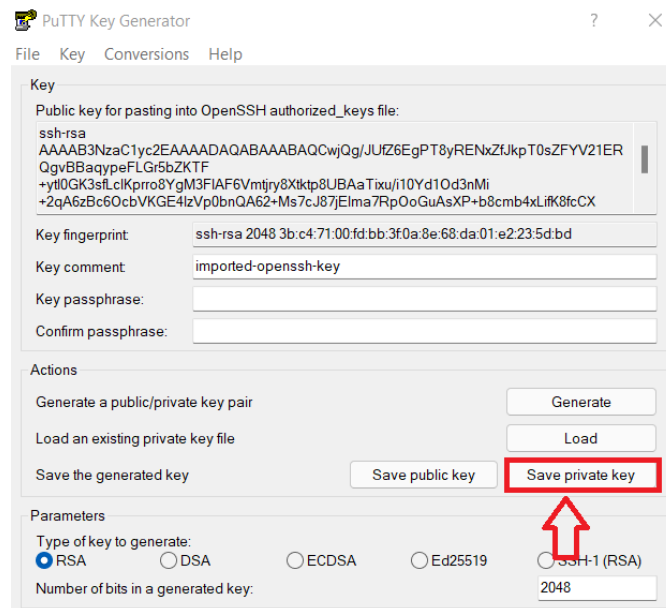


Figure 11

#### 4. AWS EMR

Write EMR inside the search bar and then click on its icon as it can be seen in figure 12.

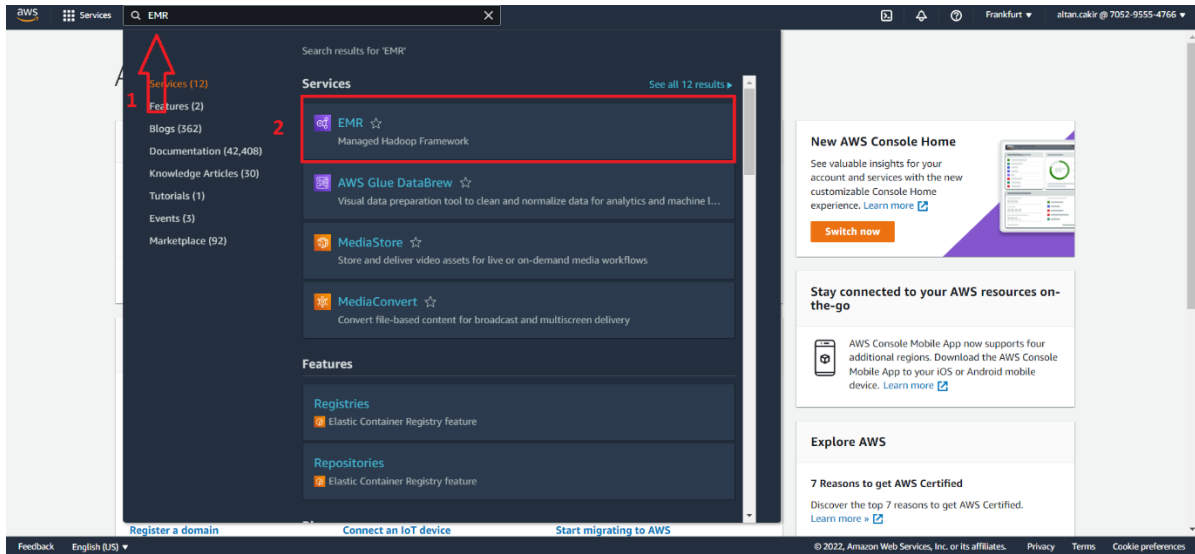


Figure 12

4. A. Start creating your cluster by clicking on Create cluster button (see Figure 13)

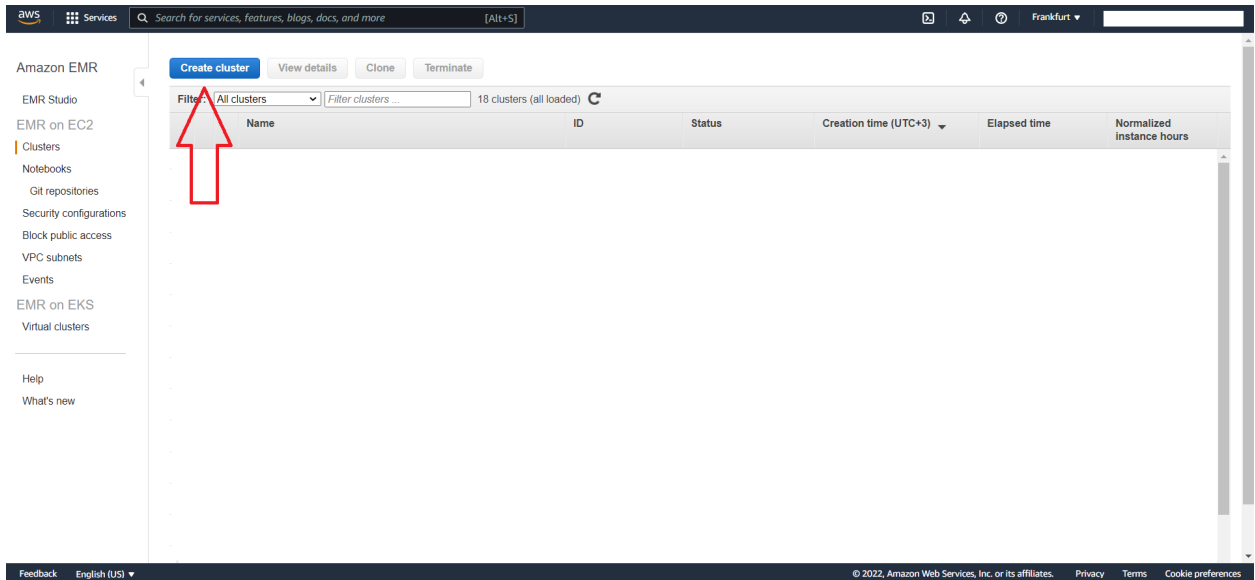


Figure 13

4. B. Now, click on Go to advanced option. (see Figure 14)

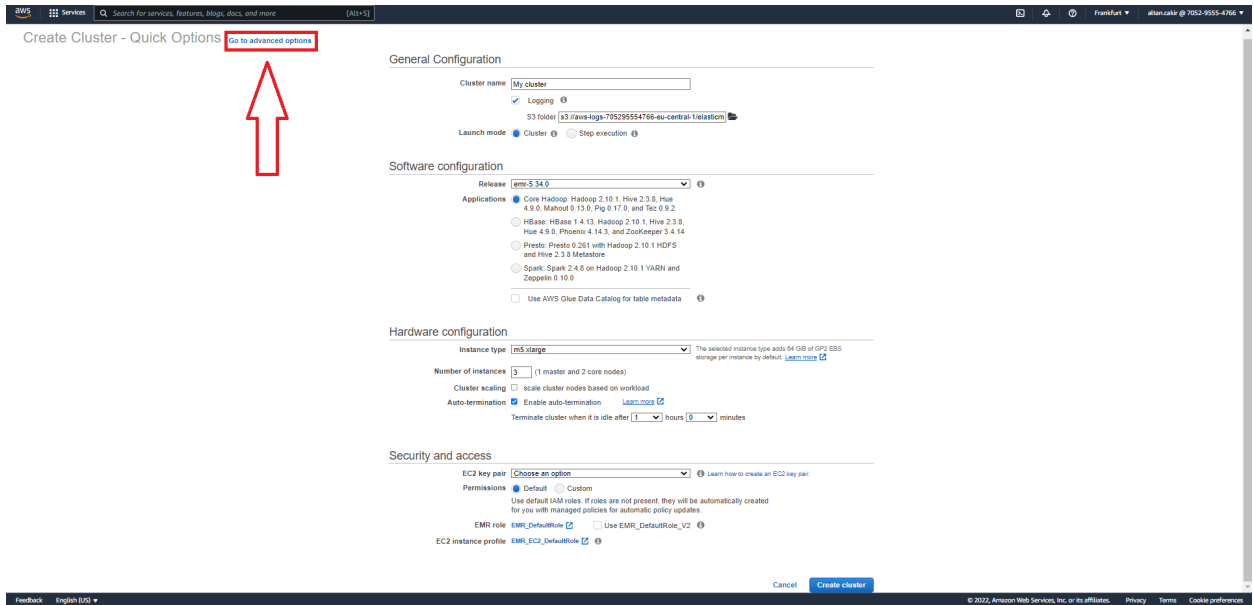


Figure 14

4. C. First of all, select the services that will be established directly by AWS. After that, load your spark configuration file by clicking on load Json from s3 and select your configuration file. Configuration files can be found in S3 at the following path ***s3://turknet-dns/spark\_configurations/***. Then, click on next as it is obvious in Figure 15.

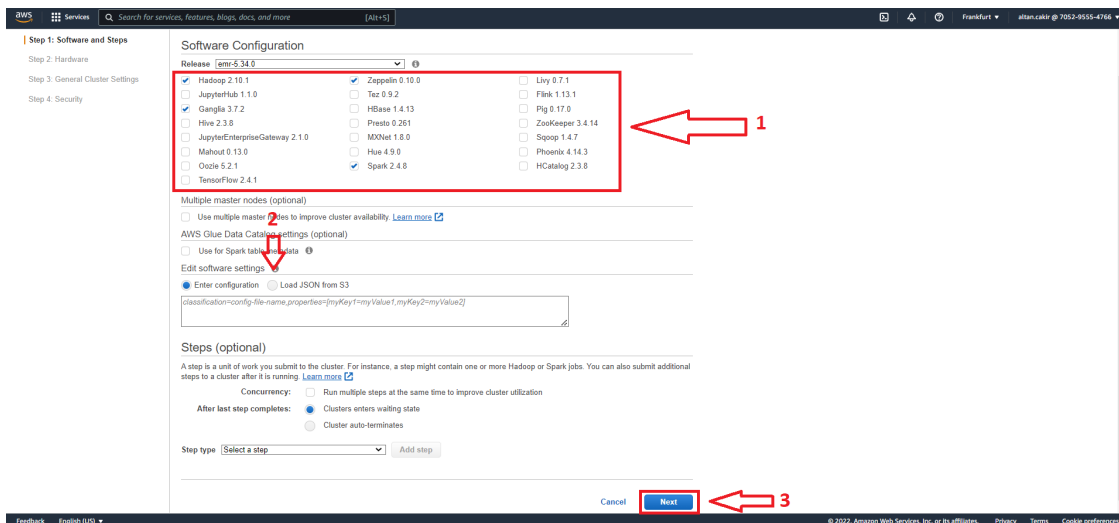


Figure 15

4. D. Then, Change the default VPC to vpc-01056db5d0f29c379 (10.0.0.0/16) | yousefVPC. This is required due to security issues. After that, select the desired machine type and number of master(s) and number of node(s) (see Figure 16)



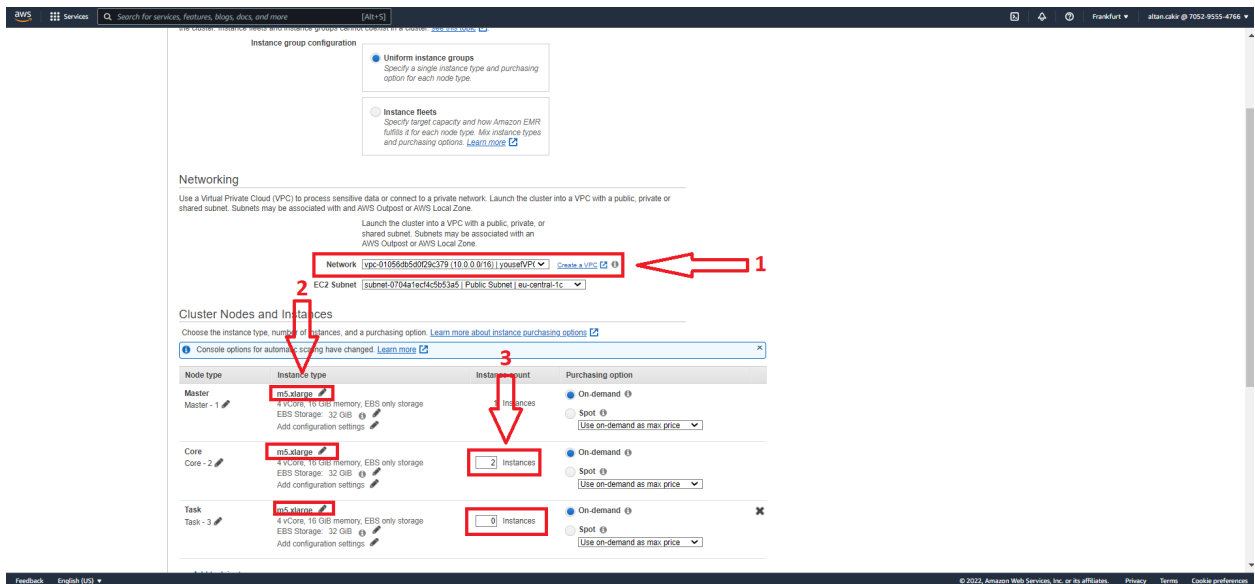


Figure 16

4. E. Give a name for your cluster, and click on Next button (see Figure17)

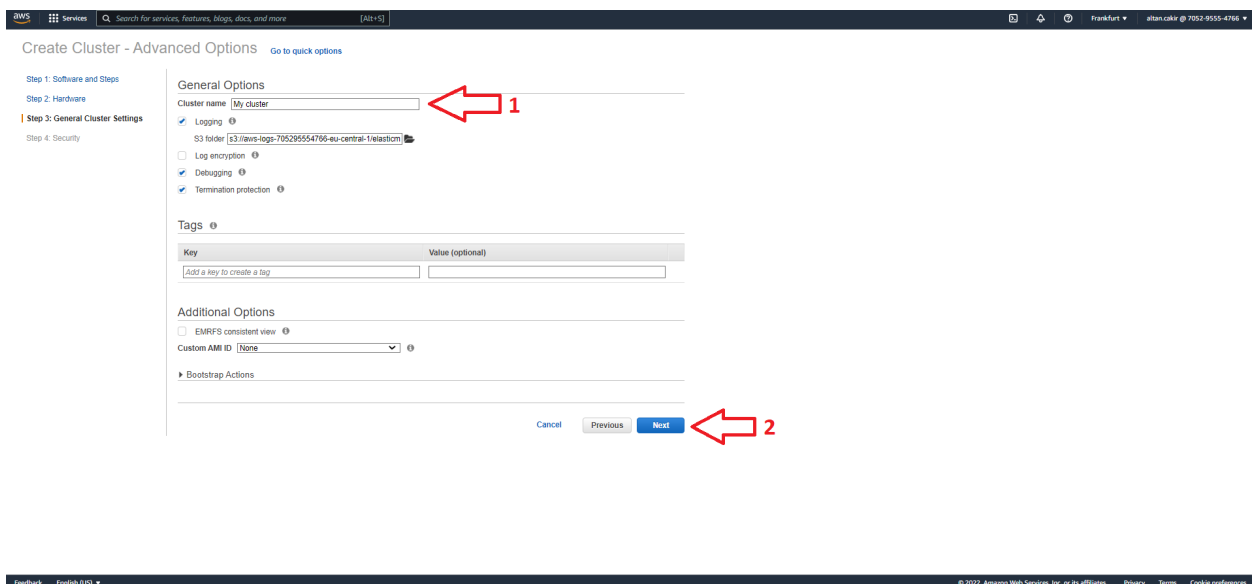


Figure 17

4. F. Select the keypair that you have created before in step 2.F and click on Create cluster button (see Figure 18)

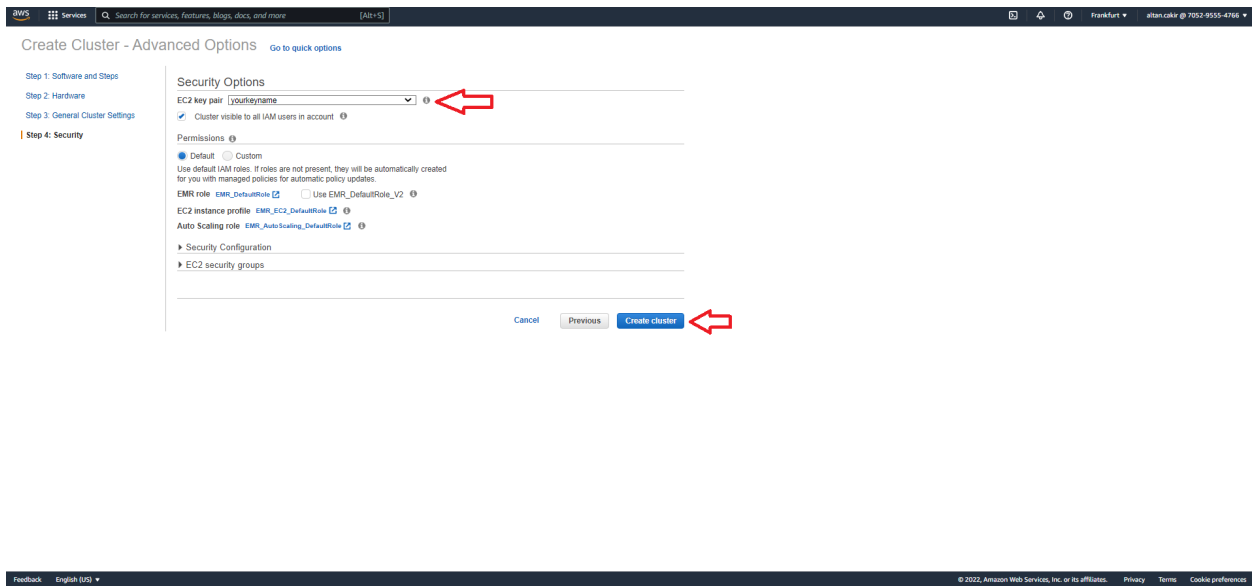


Figure 18

4. G. Click on Enable an SSH connection as it is seen figure 19.

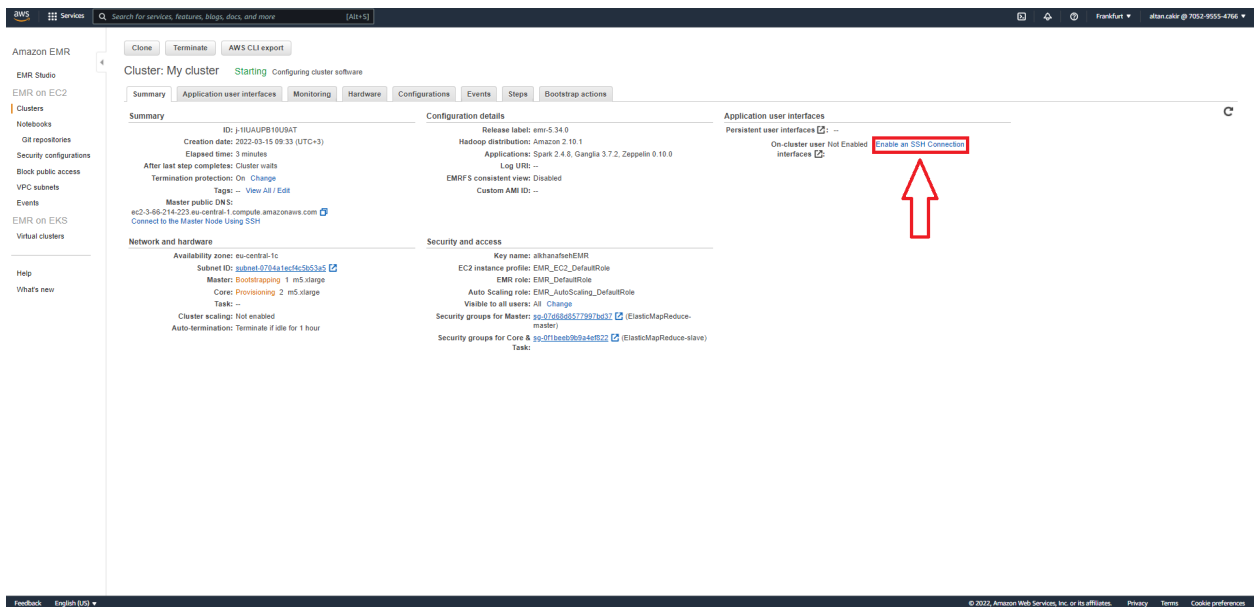


Figure 19

4. H. Follow the given informations in order to start ssh with the created cluster. Moreover, these informations must be applied in order to prepare your browser to open the interfaces of the cluster services. (see Figure 20)

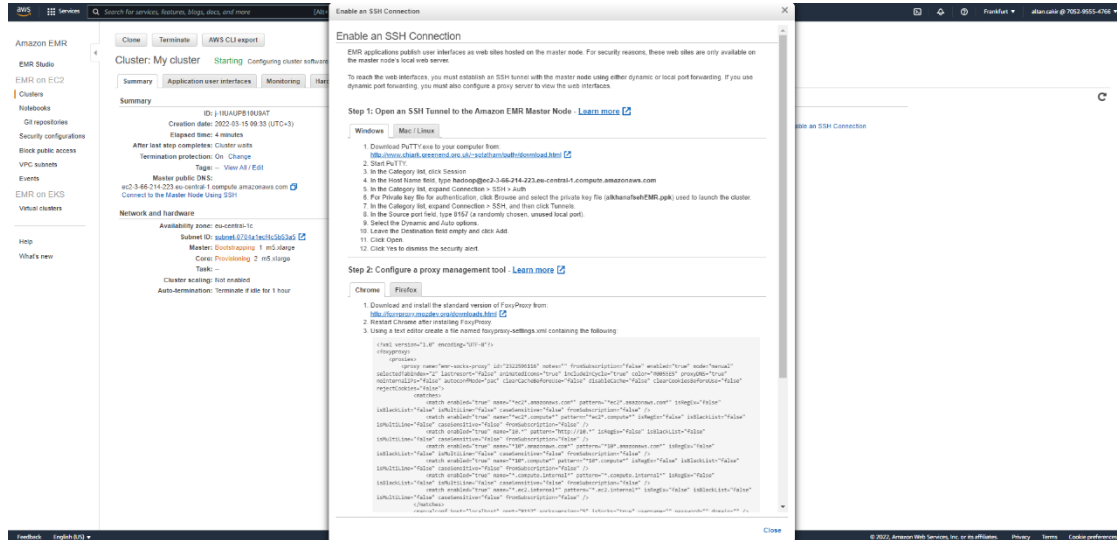


Figure 20

4. I. After enabling SSH connection with EMR master, the following screen must be appeared (see Figure 21).

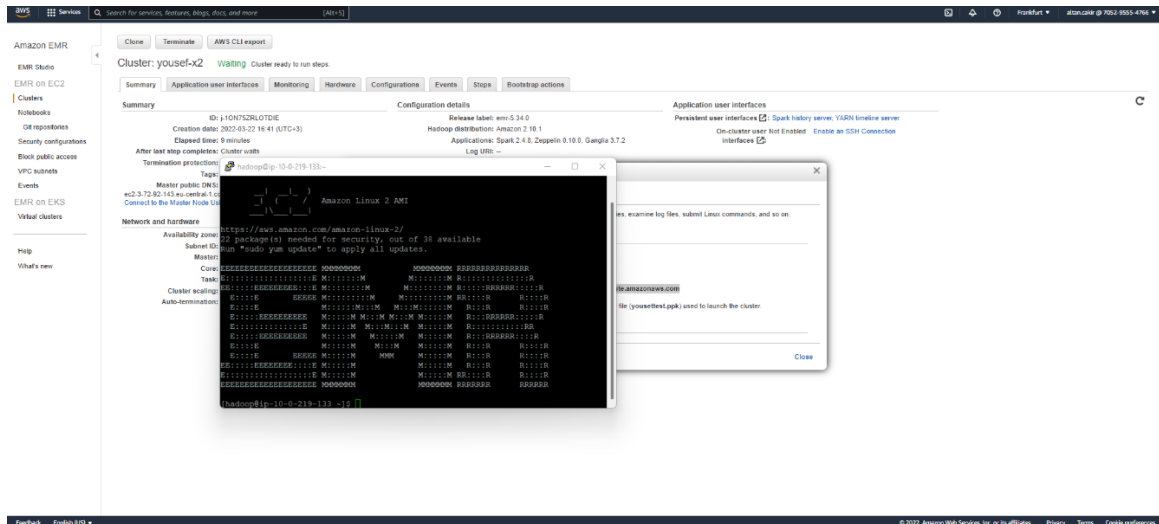


Figure 21

4. J. Click on Application user interfaces in order to see your established services (see figure 22).

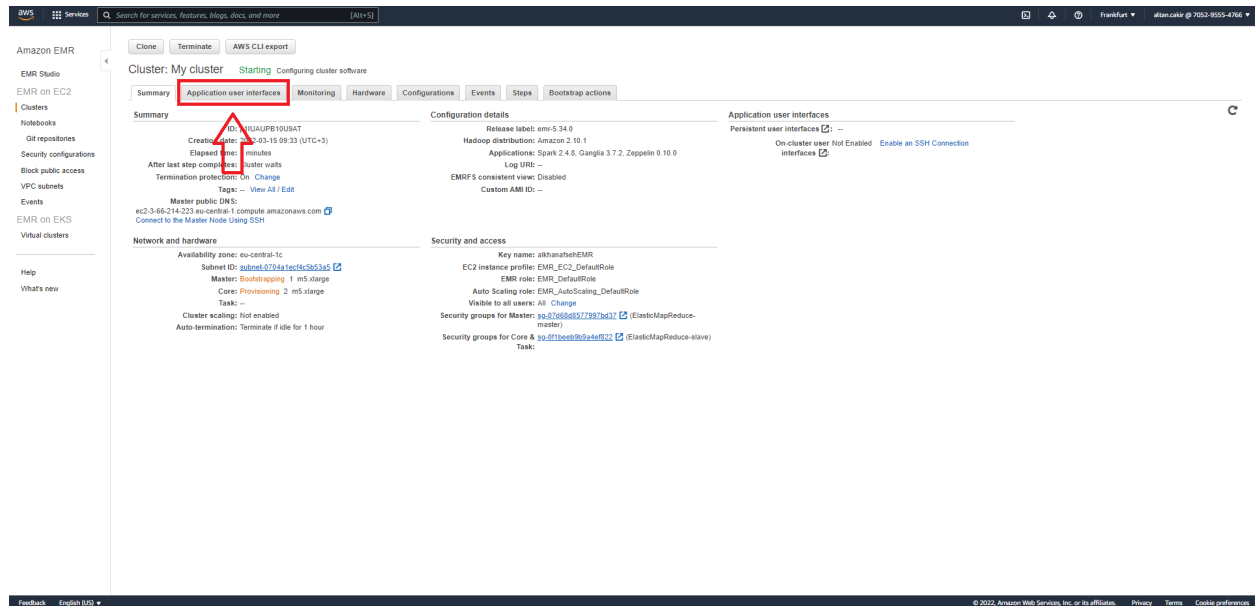


Figure 22

4. K. Go to the given links in order to open their interfaces as it is shown in Figure 23.

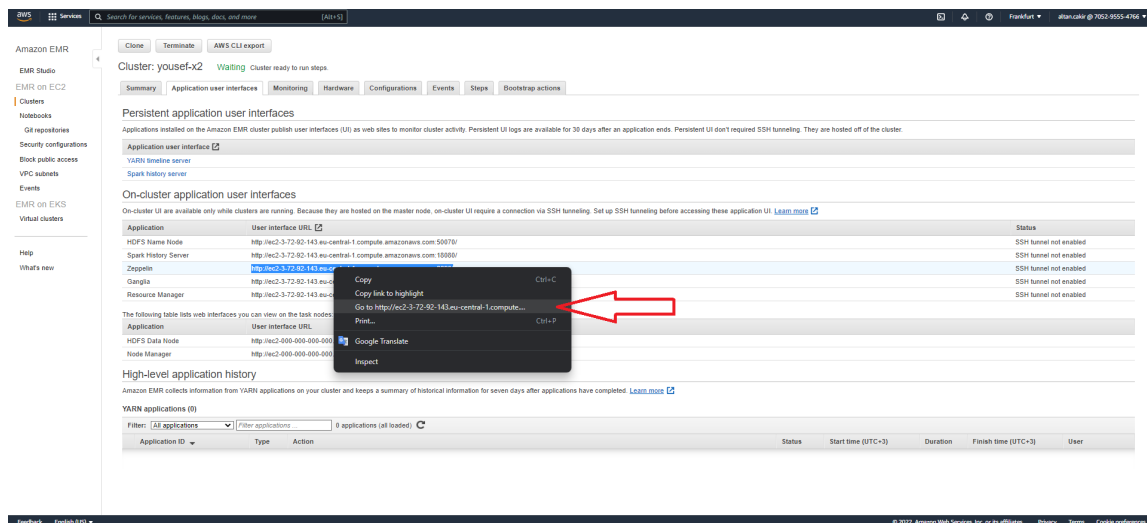


Figure 23

4. L. After opening Zeppelin Notebook, you can either import note or create new note. In the example in Figure 24, a ready note was uploaded.

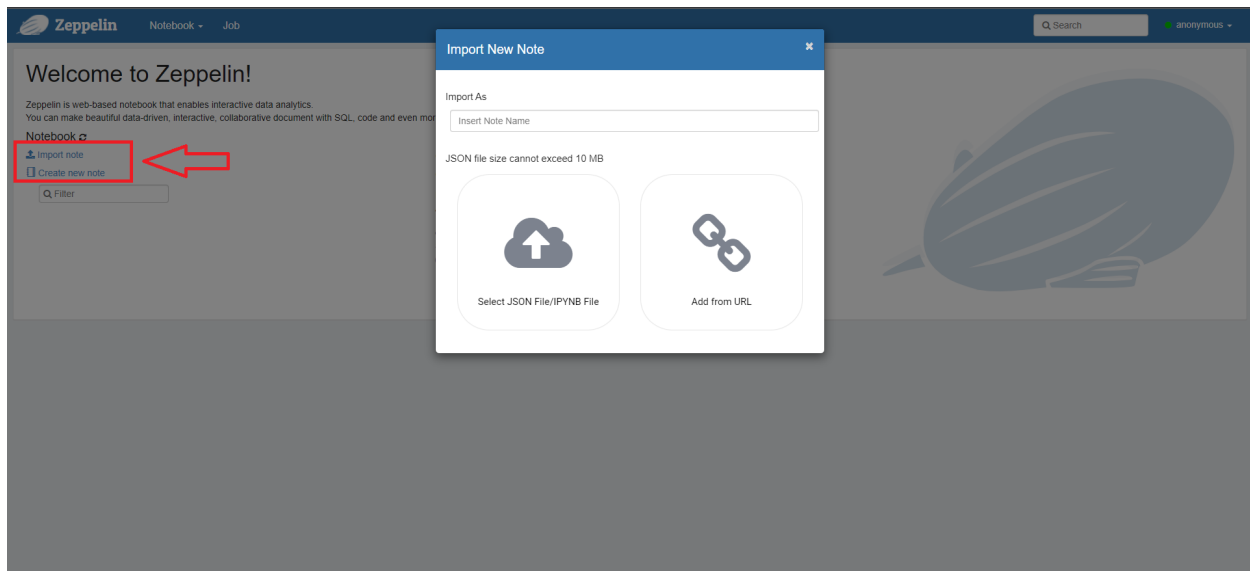


Figure 24

4. M. Start your coding on Zeppelin Notebook. An example can be seen in Figure 25

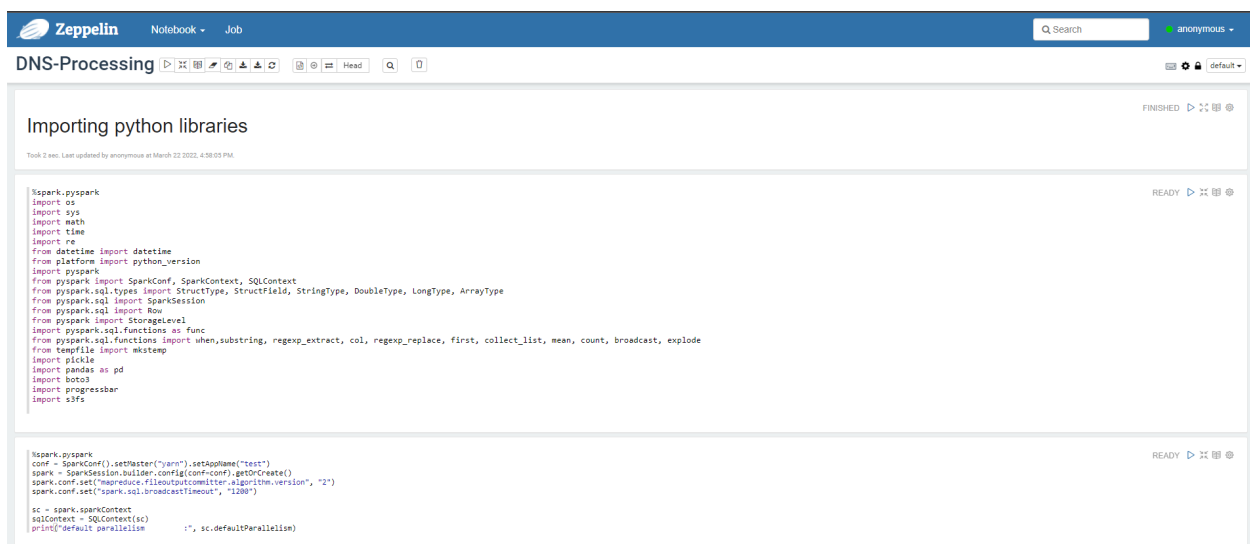


Figure 25