**AWS Hadoop Assignment Guide**

1. Connect to your aws account on url:<https://www.awseducate.com/student/s/classrooms>

2. Choose “**Big Data Platform**” class

3. Choose “**AWS Console**”

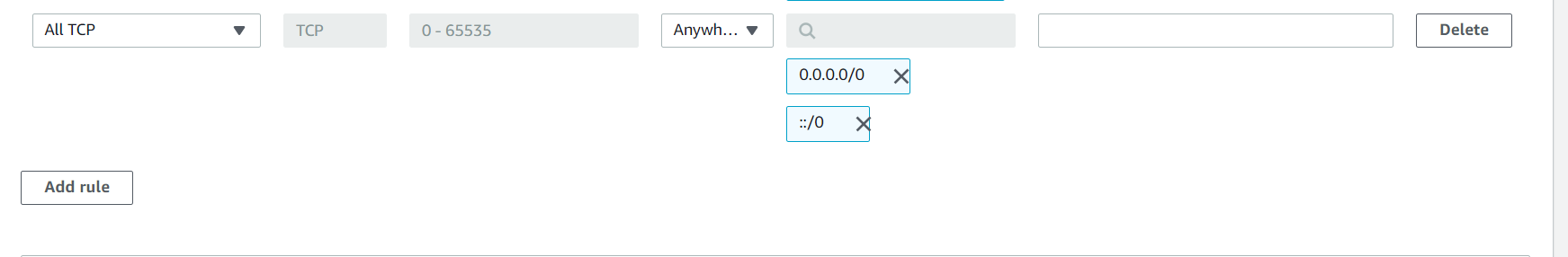
4. Under “**Services**”, go to “**EC2**”

a. Choose **Key Pairs** resource  
b. Choose pem **file format**c. Click on “**Create key pair**”  
d. Select the pem “**File format**” according to your system

**IMPORTANT:** Save the .pem file and don’t delete it! (We’ll use it soon)

5. Under “**Services**”, go to “**EMR**”

a. Choose **Create Cluster**

i. Go to Advanced options if you want use spot instances  
 ii. Wait until your cluster is in “Waiting” state (it takes about 10 minutes)  
 iii. Scroll down the page and choose **“Security groups for Master”** iv. Choose “ElasticMapReduce-master” and click “edit Inbound Rules”  
 v. Create new rule like that for allow jupyter-lab on your browser:

6. connect to your master-node instance:

1. Mac: ssh -i *my-key.pem* hadoop@*master-node-dns*
2. Windows: connect with [mobaXterm](https://mobaxterm.mobatek.net/download-home-edition.html) by create new session:
   1. host: *master-node-dns*
   2. user: hadoop
   3. choose advanced options and choose your key pair, which you create in the section above

**IMPORTANT:** replace my-key.pem with your key, and master-node-dns to yours

7. After connect the master node run the following commands:

*a.* *sudo python3 -m pip install ipykernel jupyterlab*

b. *jupyter lab --no-browser --port=40000 --ip=$HOSTNAME*

i. You should get output like this:

*http://127.0.0.1:8888 ?token=ec0ad5afba127fadbdb2ed10ed763945d10283b0f4a68db5)*

ii. open your computer browser and go to that address **by replace 127.0.0.1 with your *master-node ip***

8. Pay attention that when you terminate your EMR cluster, all the data on the cluster is lost.