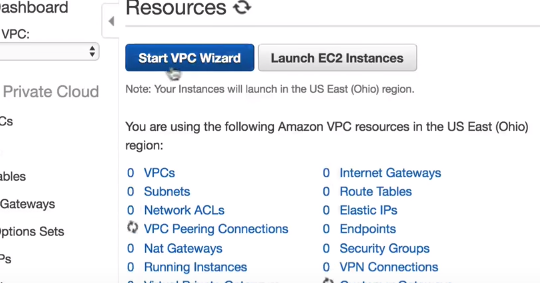
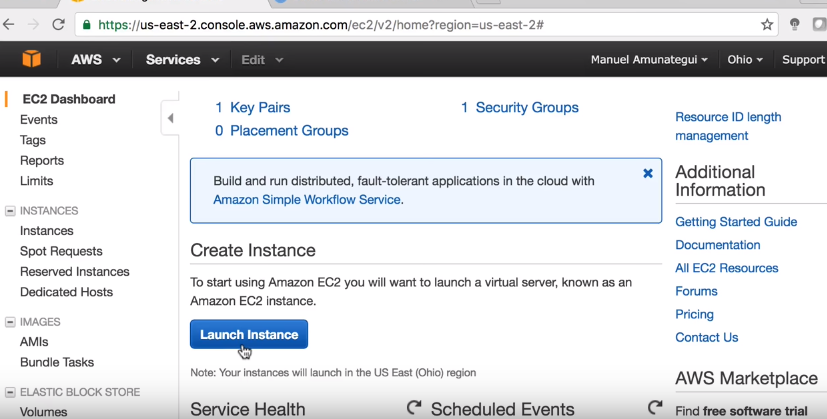
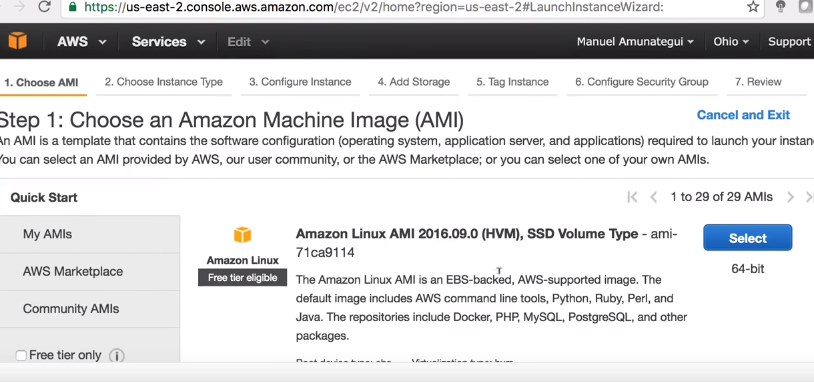
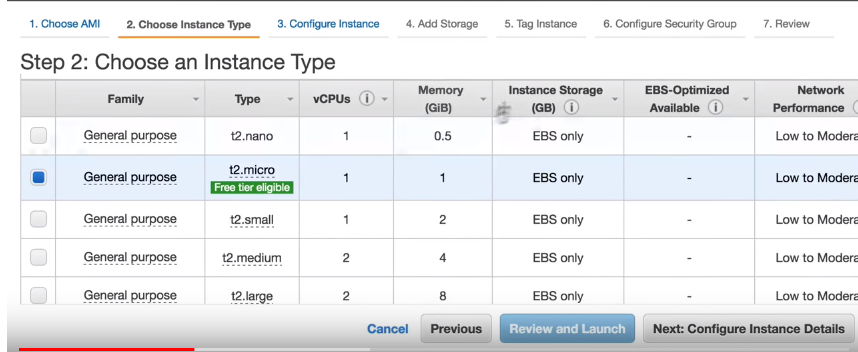
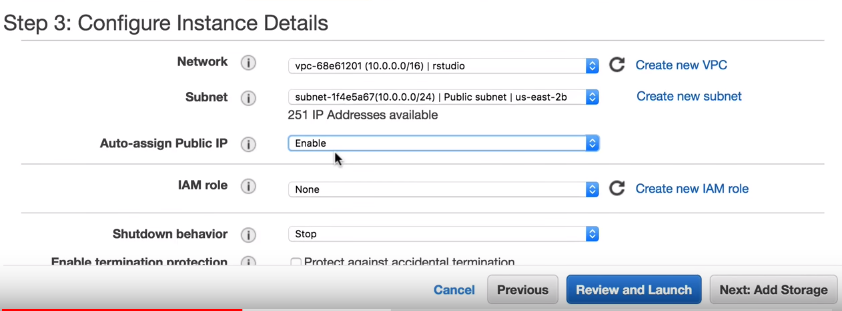
Running h2o on aws

Based on tutorial: <https://www.youtube.com/watch?v=zJuFpqB01u4>

1. Go to aws→VPC→Start VPC wizard  
   
2. Click select
3. Give a name, then create VPC
4. Go to EC2→Launch instances  
   
5. Choose Amazon linux machine  
   
6. Choose machine type then click next  
   
7. Choose Network to be the VPC just created, and enable public IP  
   
8. Scroll down to advanced details and input the following commands.   
   Get the red commands from here: <https://www.rstudio.com/products/rstudio/download-server/> Go to the RedHat CentOS tab and get the commands. Note the username and password below  
     
     
   #!/bin/bash

yum install -y R   
wget https://download2.rstudio.org/rstudio-server-rhel-1.0.153-x86\_64.rpm

yes | sudo yum install --nogpgcheck rstudio-server-rhel-1.0.153-x86\_64.rpm  
yum install -y curl-devel   
useradd mohamed  
echo mohamed:hamada | chpasswd

cd /home/ec2-user

sudo AWS\_ACCESS\_KEY\_ID=AKIAJHHVRLCYUDSB2EQQ AWS\_SECRET\_ACCESS\_KEY=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 aws s3 cp s3://abolfadl/h2o.jar h2o.jar

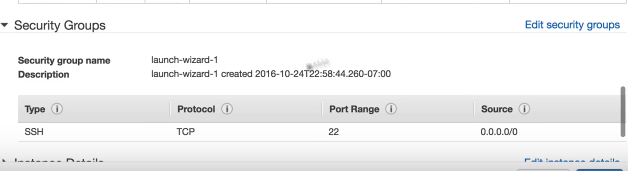
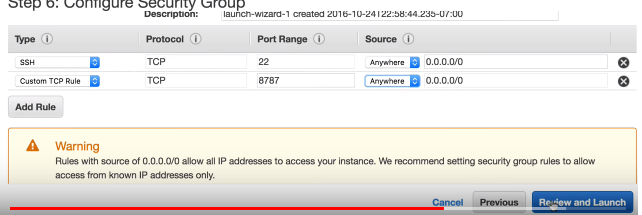
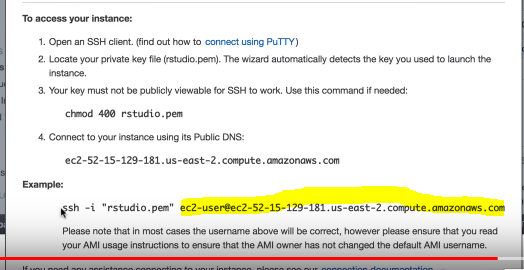
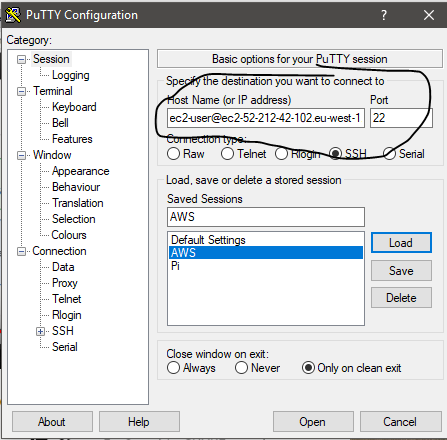
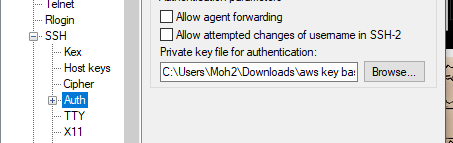
java –jar h2o.jar

java –Xmx1g -jar h2o.jar -flatfile flatfile.txt

java –Xmx1g -jar h2o.jar -network [10.0.0.0/24](http://192.168.1.0/24) -hdfs\_config core-site.xml

AWS\_ACCESS\_KEY\_ID=AKIAJHHVRLCYUDSB2EQQ AWS\_SECRET\_ACCESS\_KEY=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 aws s3 cp s3://abolfadl/flatfile.txt flatfile.txt

AWS\_ACCESS\_KEY\_ID=AKIAJHHVRLCYUDSB2EQQ AWS\_SECRET\_ACCESS\_KEY=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 aws s3 cp s3://abolfadl/data.csv data.csv

1. Click next and see if you need storage for large data
2. Click next and edit security groups  
   
3. Add TCP custom rule with port 8787 and make it anywhere as well as SSH anywhere. ALSO ADD ANOTHER TCP CUSTOM WITH PORT 54321 FOR FLOW!!!  
   
4. Click Launch
5. Create new pair→Download .pem file→open PuTTyGen →load .pem (make all files visible .\*)→Check RSA→save
6. Open PuTTy in host name put the username from the connect guide  
     
   
7. Go to SSH→Auth→Browse and select the ppk file created earlier  
   
8. Get public IP of the machine…put it in browser with port of rstudio and FLOW  
   xx.xx.xx.xx:8787
9. If rstudio doesn’t open rerun the following command while giving y as an answer  
   yes | sudo yum install --nogpgcheck rstudio-server-rhel-1.0.153-x86\_64.rpm
10. To get data from s3 go to the SSH console and type  
    AWS\_ACCESS\_KEY\_ID=AKIAJHHVRLCYUDSB2EQQ AWS\_SECRET\_ACCESS\_KEY=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 aws s3 cp s3://abolfadl/data.csv data.csv
11. In rstudio console install h2o, use it and initialize an h2o instance  
    install.packages(“h2o”)  
    library(h2o)  
    h2o.init()
12. Open the browser and launch h2o FLOW  
    xx.xx.xx.xx:54321