* Advanced option
* SSH
  + Specify pem file in security options while setting up aws
* ssh -i C:/Users/arali/Desktop/BrainStation/Lecture/Cloud/bstn-aq2.pem hadoop@ec2-18-221-181-171.us-east-2.compute.amazonaws.com
* hadoop fs -mkdir /user/hadoop/eng\_1M\_1gram
* hadoop distcp s3://brainstation-dsft/eng\_1M\_1gram.csv /user/hadoop/eng\_1M\_1gram
* ssh -i C:/Users/arali/Desktop/BrainStation/Lecture/Cloud/bstn-aq2.pem -ND 8157 hadoop@ec2-18-221-181-171.us-east-2.compute.amazonaws.com
* df = spark.read.csv('/user/hadoop/eng\_1M\_1gram', header=True)
* df.printSchema()
* df.count()
* df.createOrReplaceTempView('google\_ngrams')
* df1 = spark.sql("SELECT \* FROM google\_ngrams WHERE Token='data'")
* df1.show(10)
* df1.count()
* df1.write.csv('/user/hadoop/eng\_data\_1gram /filter\_data\_1gram')
* hadoop fs -ls
* hadoop fs -getmerge /user/hadoop/eng\_data\_1gram collected\_data\_1gram.csv
* aws s3 cp collected\_data\_1gram.csv s3://lastassignmentbucket/aws\_data\_1gram.csv