* Create an instance in amazon aws
* Connect

cd /Users/apple/Documents/assignments/bigdata ln

ssh -i "bigdata01.pem" [ec2-user@ec2-52-36-18-172.us-west-2.compute.amazonaws.com](mailto:ec2-user@ec2-52-36-18-172.us-west-2.compute.amazonaws.com)

Note: To copy the data from system to instance use below command

bigdata tejasvi$ scp -i "bigdata01.pem" bigdata01.pem [ec2-user@ec2-52-36-18-172.us-west-2.compute.amazonaws.com:~/](mailto:ec2-user@ec2-52-36-18-172.us-west-2.compute.amazonaws.com:~/)

$scp -i ~/Desktop/amazon.pem [ubuntu@ec2-54-166-128-20.compute-1.amazonaws.com:/data/ecoli\_ref-5m-trim.fastq.gz ~/Download/](mailto:ubuntu@ec2-54-166-128-20.compute-1.amazonaws.com:/data/ecoli_ref-5m-trim.fastq.gz%20~/Download/)

scp –i ~/newone.pem  [**ec2-user@ec2-52-11-232-63.us-west-2.compute.amazonaws.com**:/home/ec2-user/hadoop-3.0.0-alpha2/bin/Driver.java](mailto:%20ec2-user@ec2-52-11-232-63.us-west-2.compute.amazonaws.com:/home/ec2-user/hadoop-3.0.0-alpha2/bin/Driver.java) /Users/apple/Documents/assignments/bigdata

* To download Hadoop package :

Wget <http://mirrors.ibiblio.org/apache/hadoop/common/hadoop-3.0.0-alpha2/hadoop-3.0.0-alpha2.tar.gzto>

wget http://apache.mirrors.lucidnetworks.net/hadoop/common/hadoop-3.0.0-alpha2/hadoop-3.0.0-alpha2.tar.gz

* To exact it :

tar –xvf hadoop-3.0.0-alpha2.tar.gz

* To check the present java version:

java –version

* Download latest version:

sudo yum install java-1.8.0-openjdk ---- this gives just java run time only

sudo yum install java-1.8.0-openjdk-devel ---- this gives jdk also. Run this one also

* To make default version

sudo /usr/sbin/alternatives --config java

* Need to set the paths to run java for Hadoop :

Go to home directory by giving “CD”

Open .bashrc in nano mode –

nano .bashrc

* Copy the below two statements into that bash file

export JAVA\_HOME="/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.121-0.b13.29.amzn1.x86\_64"

export HADOOP\_CLASSPATH=${JAVA\_HOME}/lib/tools.jar

* After saving it to reset it by –------

source .bashrc

* Also java\_home has to be set in the Hadoop.env.sh which is present in the below path :

/home/ec2-user/hadoop-3.0.0-alpha2/etc/hadoop/hadoop.env.sh

* Now move to bin folder of hadoop :

cd /home/ec2-user/hadoop-3.0.0-alpha2/bin

* Once you are into this folder do the below command

nano WordMapper.java

* Editor will open
* Copy the programme into it and save the file.
* DO the same for the rest two.
* Write programmes here for example WordMapper.java, SumReducer.java, WordCount.java
* Compile the above java file with the below command from bin path only to create class files

./hadoop com.sun.tools.javac.Main \*.java

* Create the jar file with the below command from the same path

jar cf wc.jar \*.class

* Create an input directory to store the input file from bin directory

mkdir input

cd input

nano teja

* write some data and save it
* Come out of input directory now with the below command

Cd ..

* Now run the final command to run the mapreduce job

if

./hadoop jar wc.jar WordCount input output

* Once the job is done check the output file in output folder
* 2 files will bepresent
* if success is present then it is successful
* and second file contains output to view give below command

cat part-r-00000

* To delete the outputfile

rm -rf output

* to delete the folder

rm input