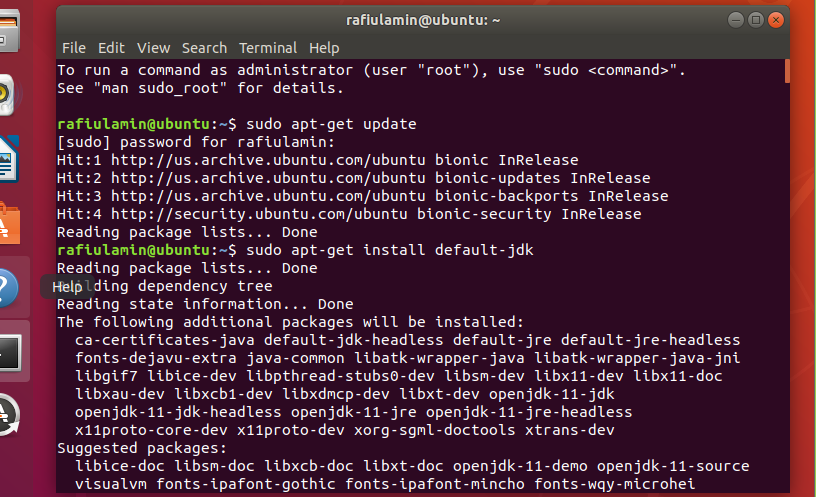
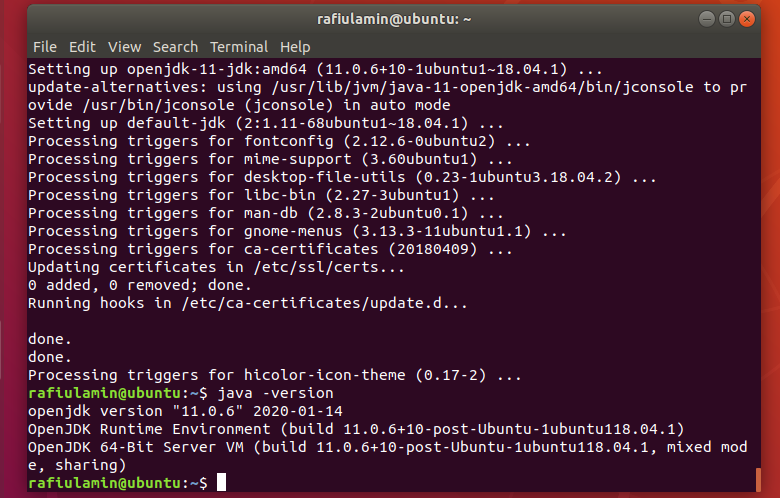
# Project 1

Name: Md. Rafiul Amin Student ID: 1637374

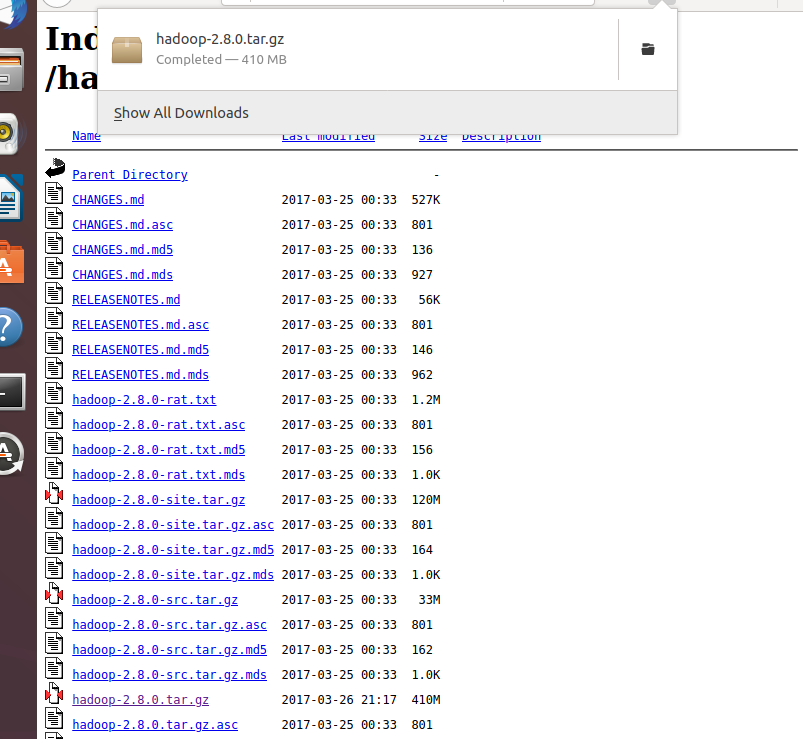
## Installing Java:



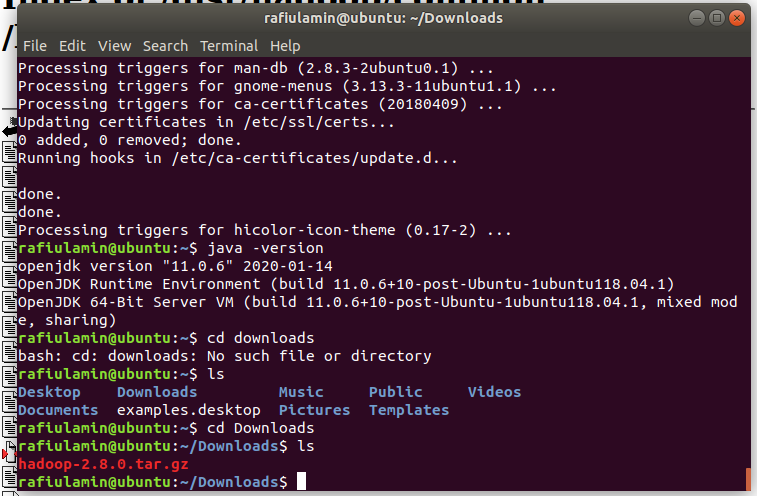
## Checking Java version:

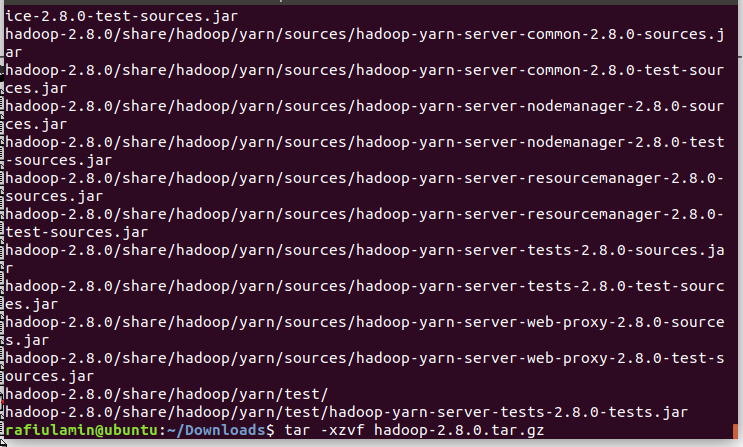


## Downloading Hadoop:

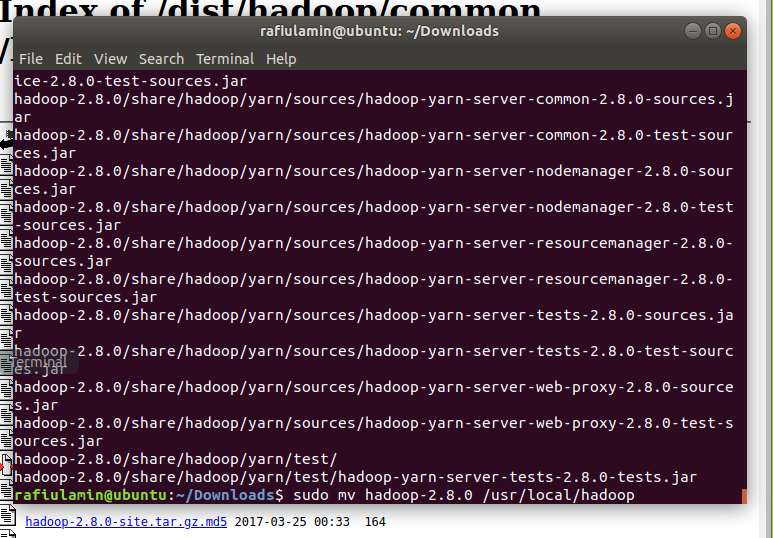
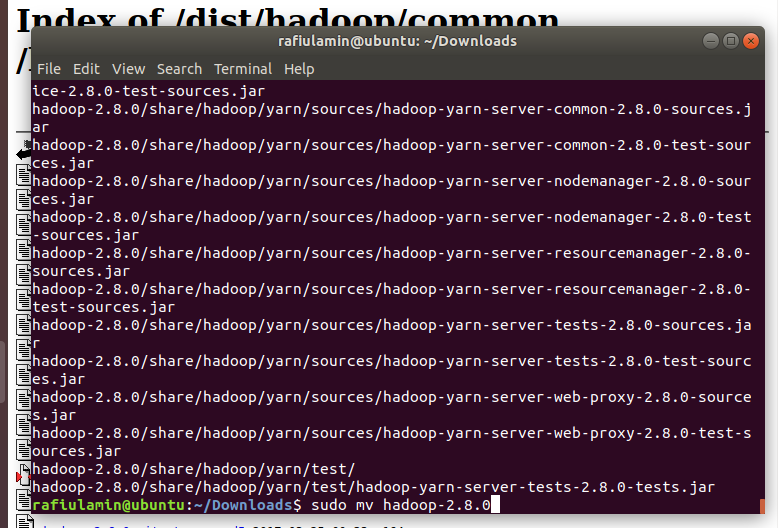


## Extract Hadoop files:

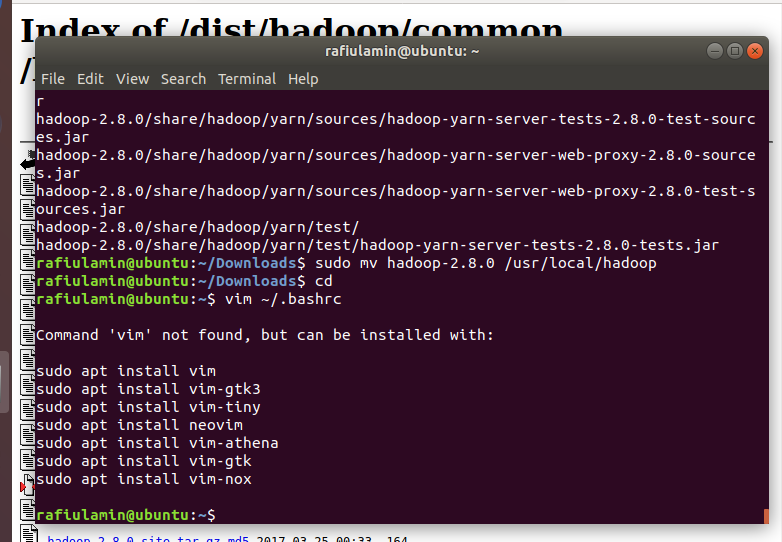




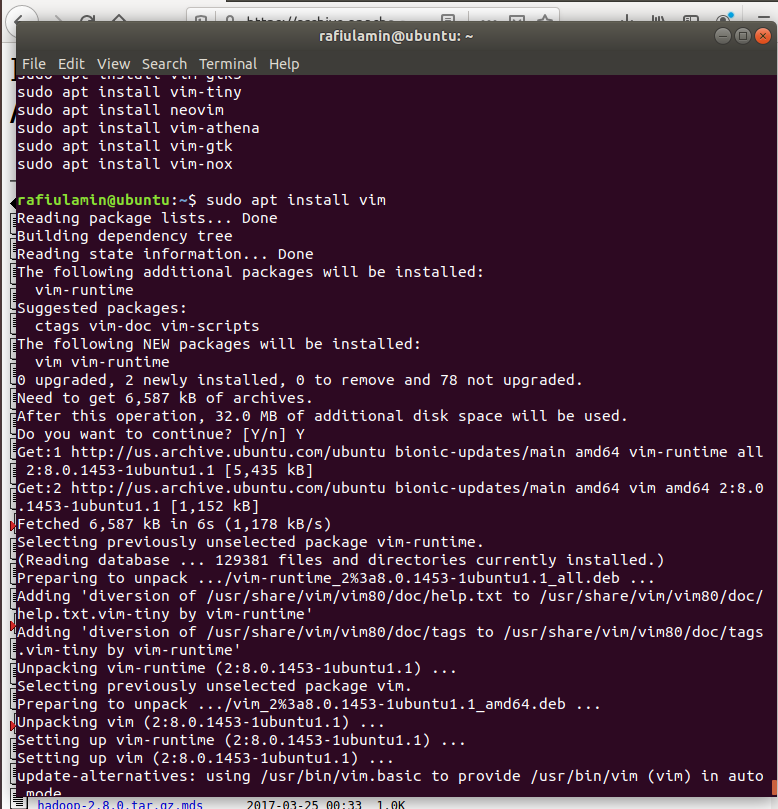
## Move Hadoop files to the new directory (/usr/local/hadoop):

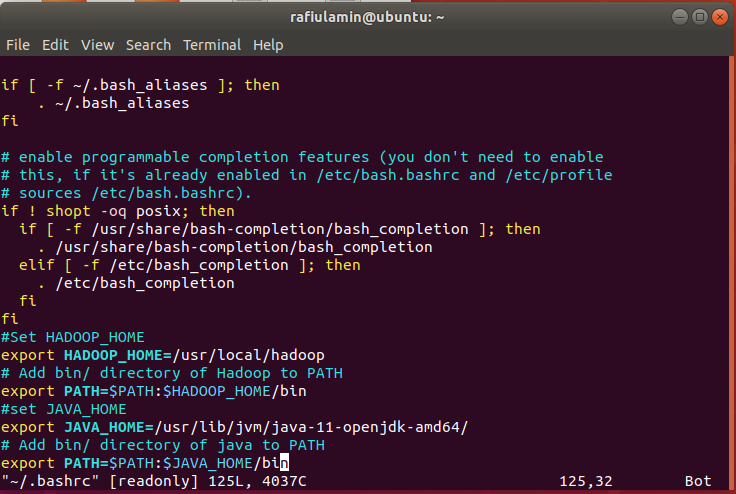


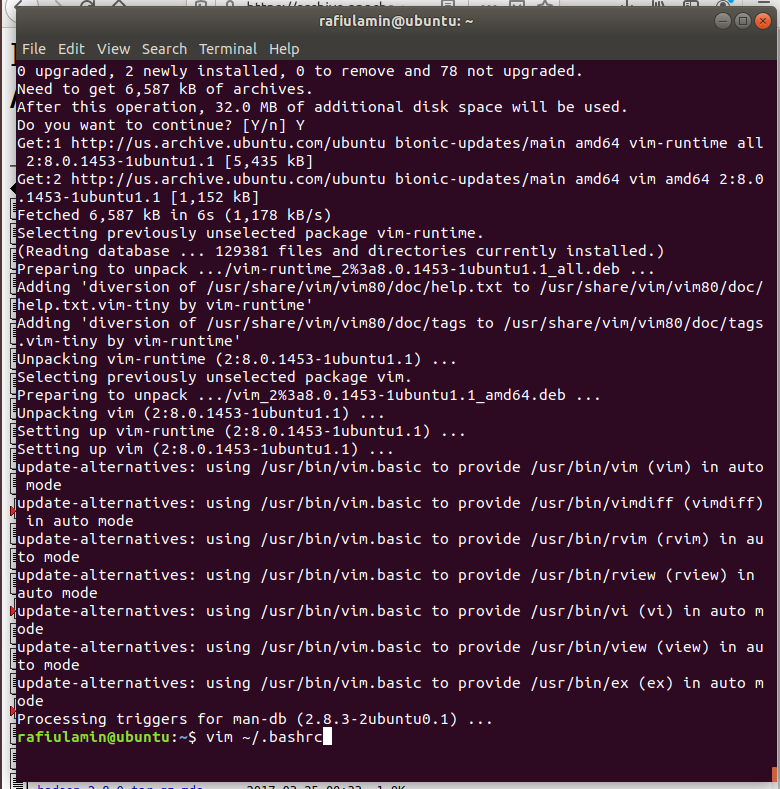
## Edit File “bashrc.sh” to set paths:



Install vim in order to open and edit the files.

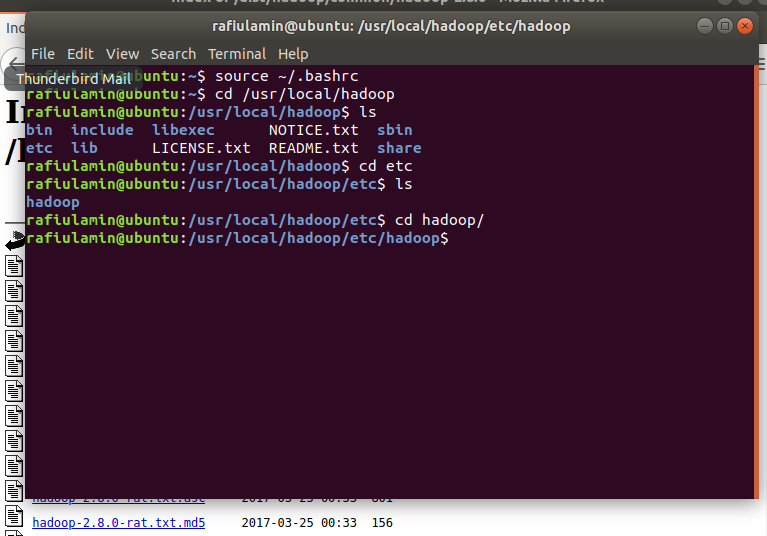


Now we can edit “bashrc.sh” file to set paths. 

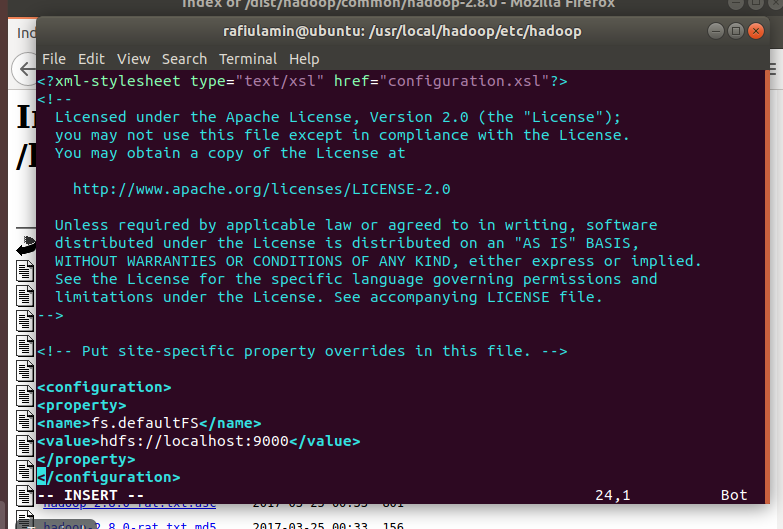
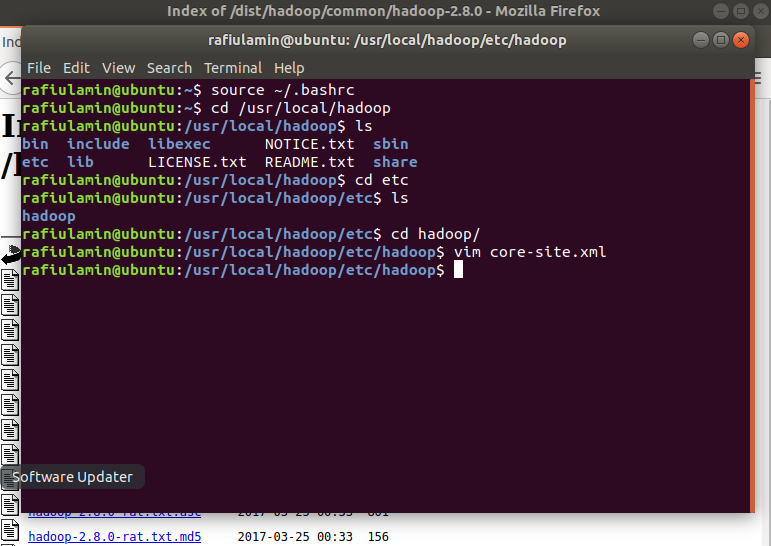


## Edit files for configuration:

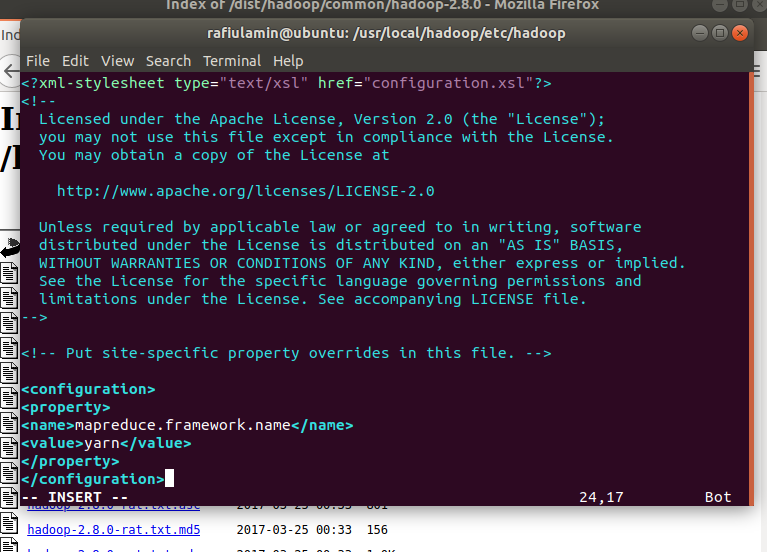
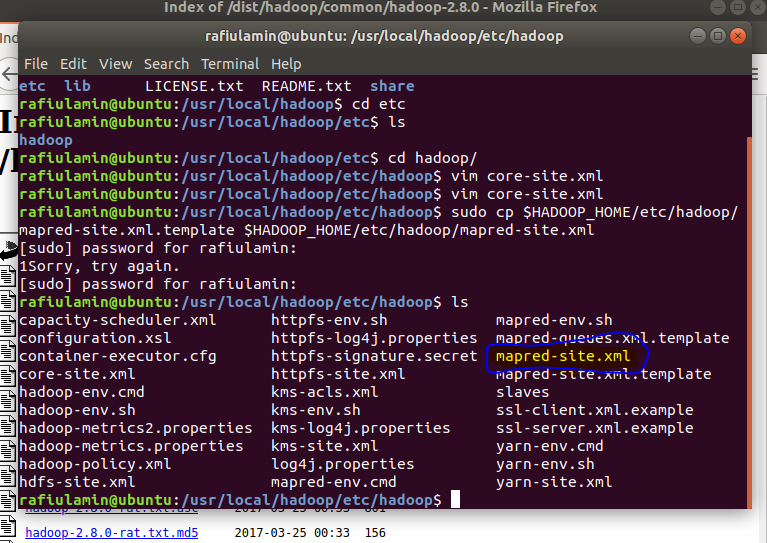
Change directory to “/usr/local/Hadoop/etc/hadoop”.



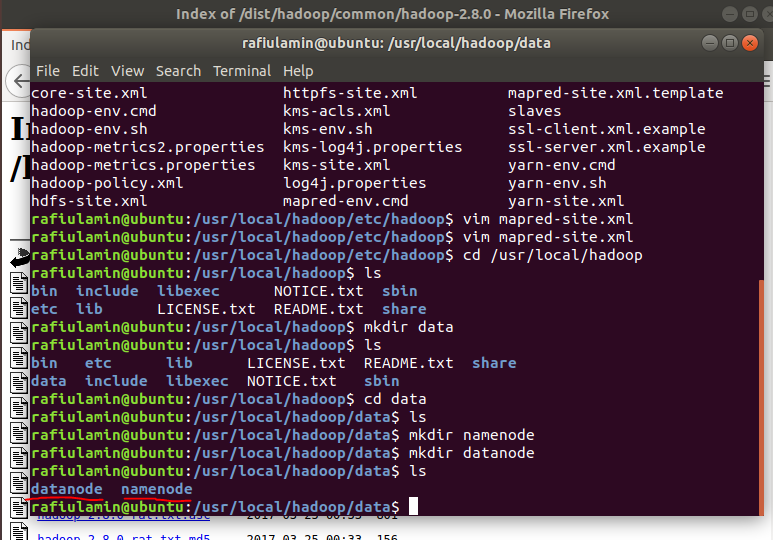
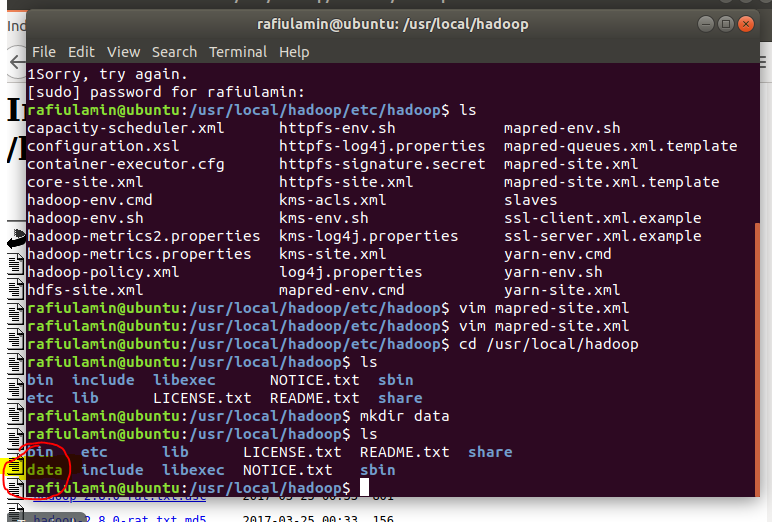
Configuring “core-site.xml” file.



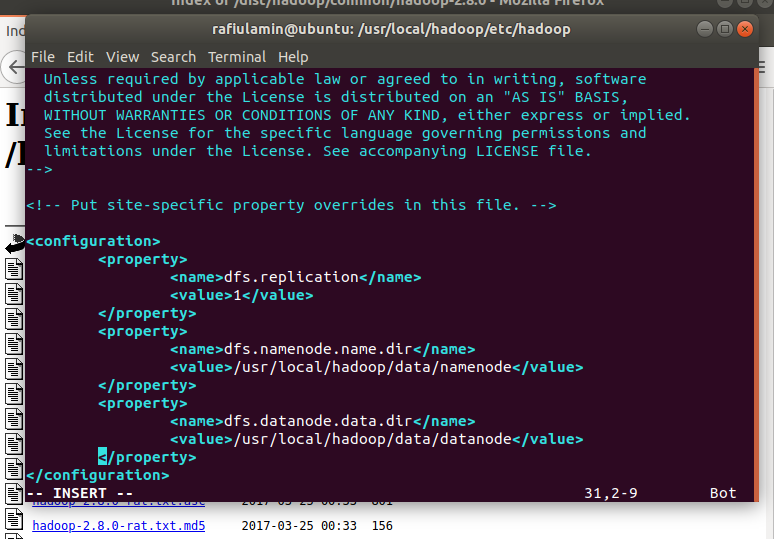
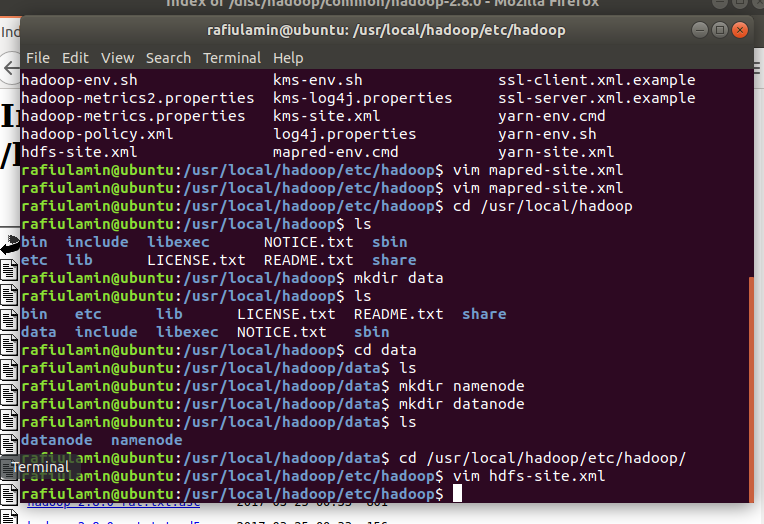
Make “mapred-site.xml” file. We copy the template and then edit it.



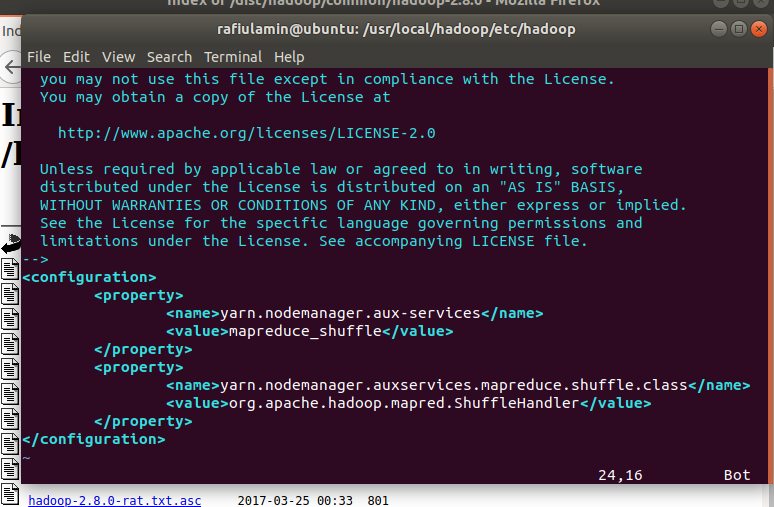
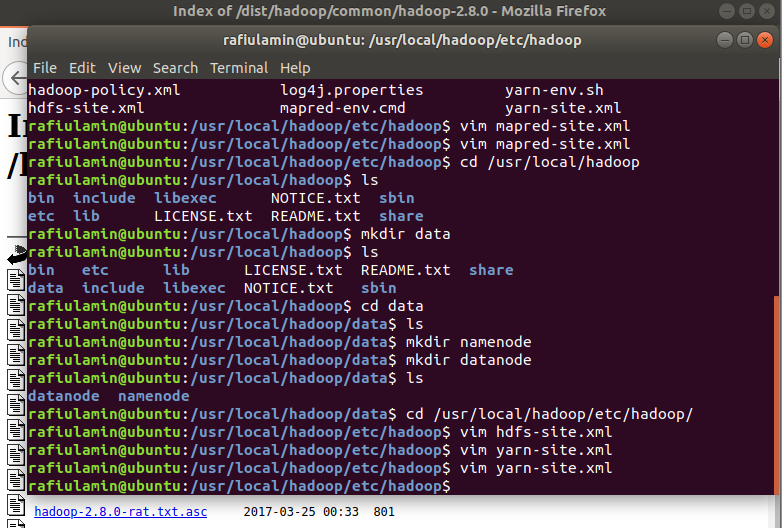
## Create “data” Directory and Corresponding Sub-directory:



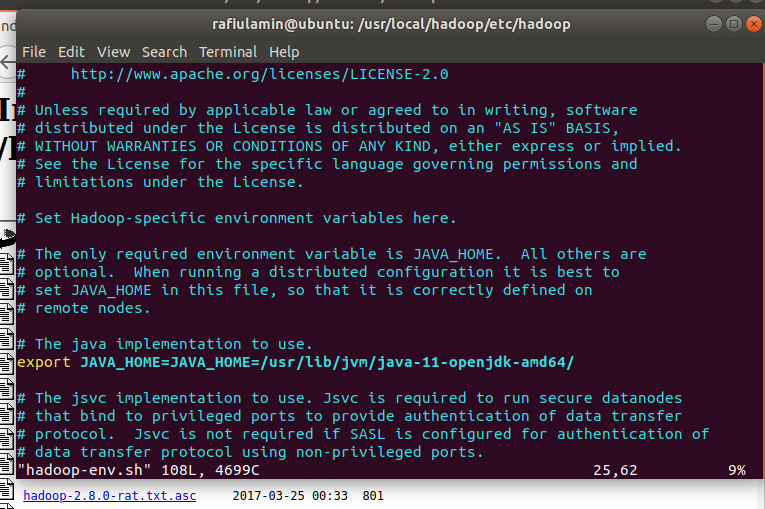
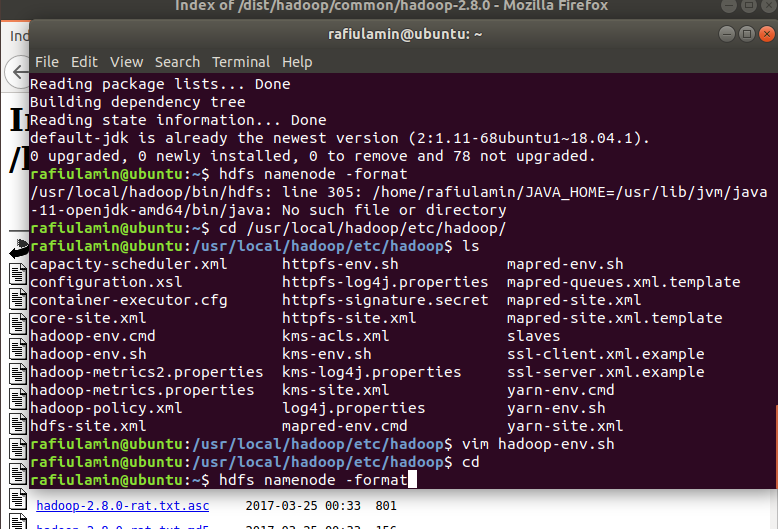
## We configure “hdfs-site.xml” to set the path for datanode and namenode:



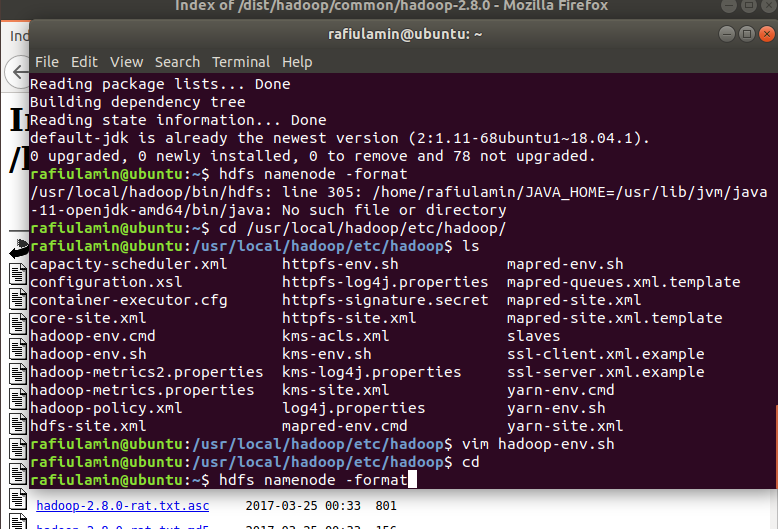
## Configure “yarn-site.xml”:



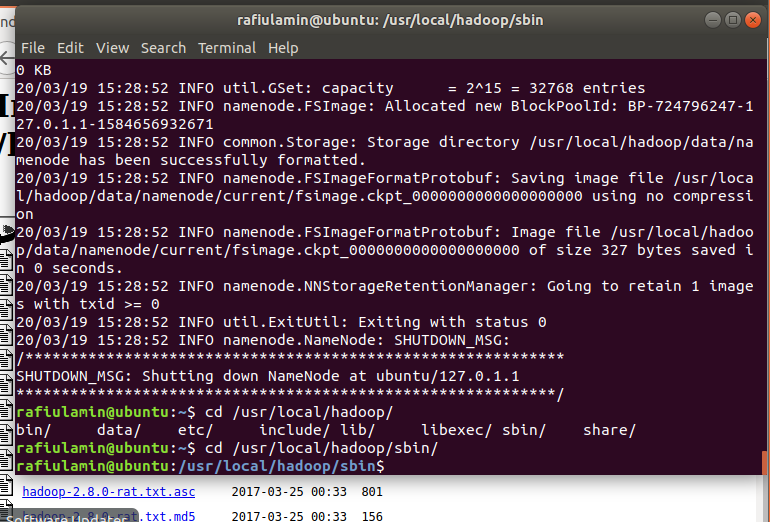
## Then we configure “hadoop-env.sh” and cha.



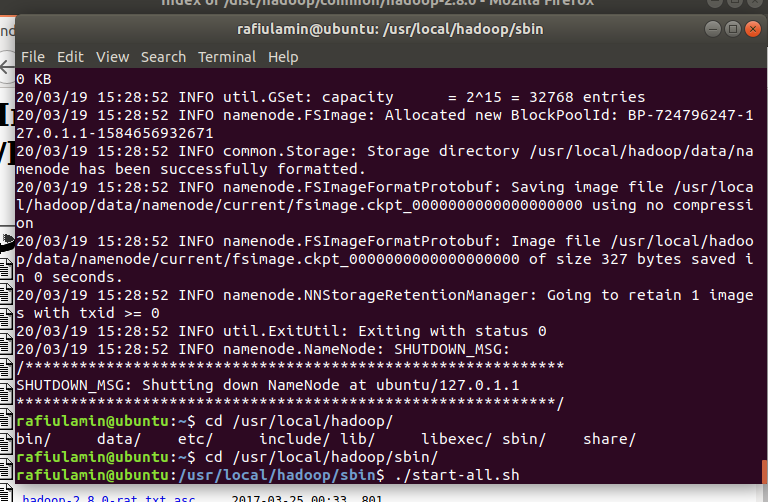
## Then we change directory and format namenode:



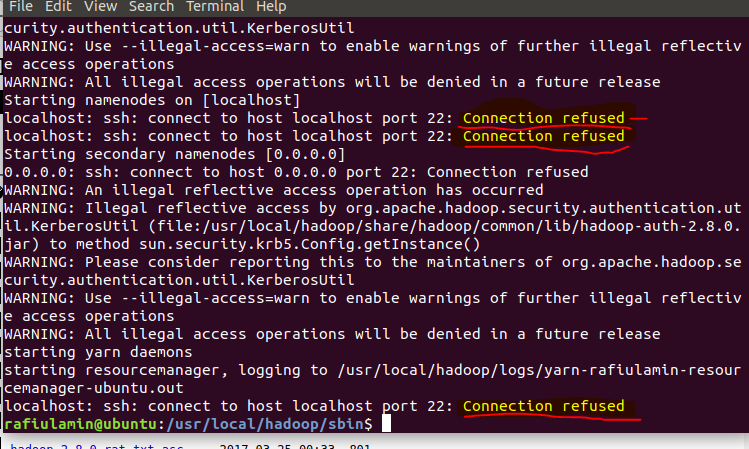
## We finally change directory to Hadoop “sbin” in order to start all the clusters:



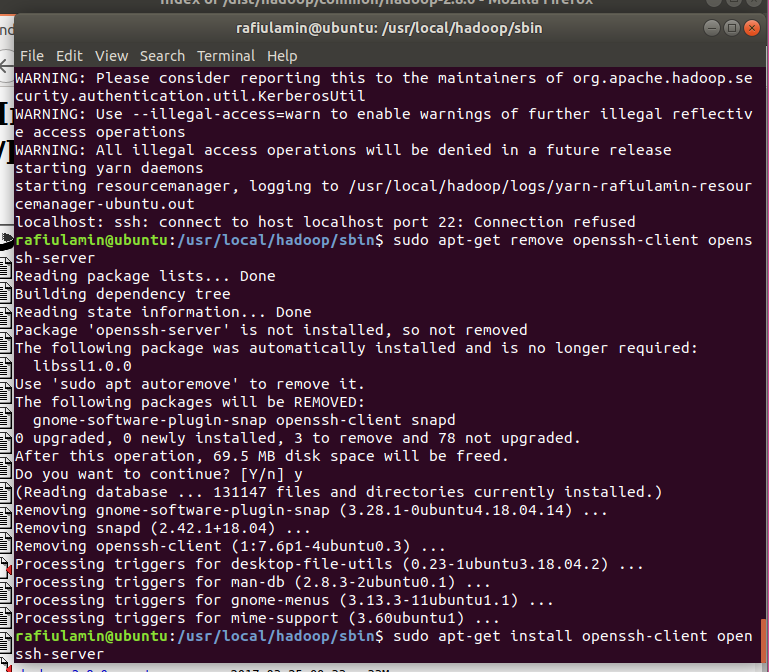
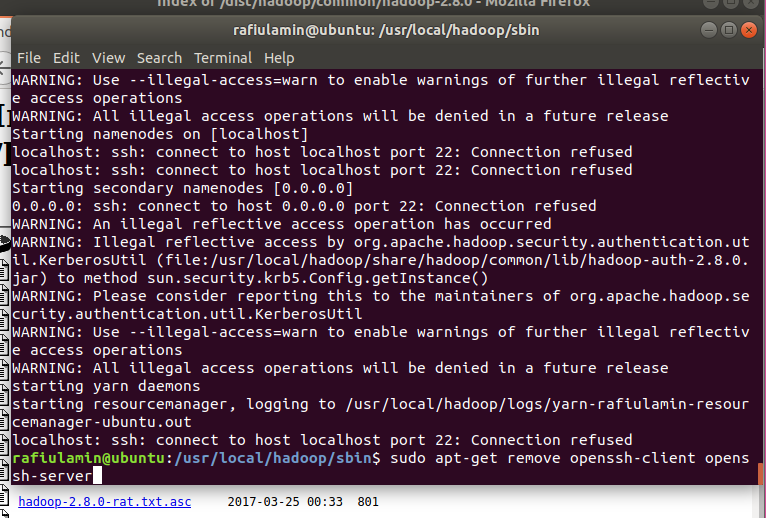
## We run “./start.sh”:



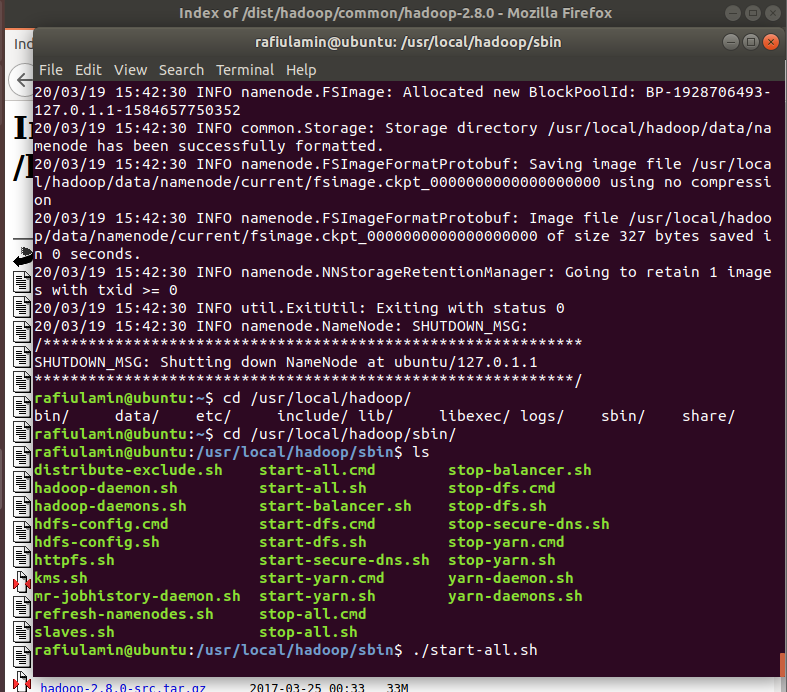
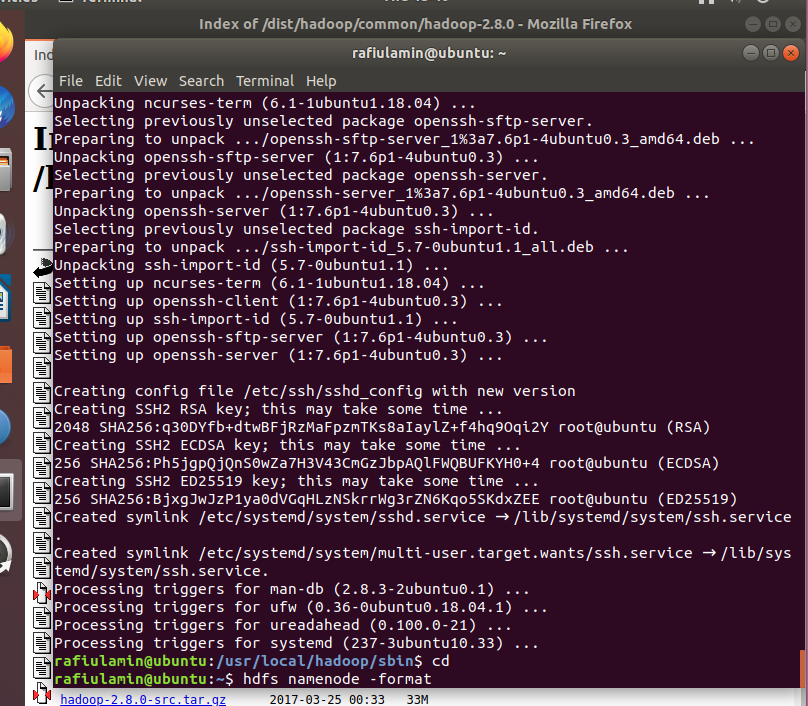
## Facing problem with connection refused:



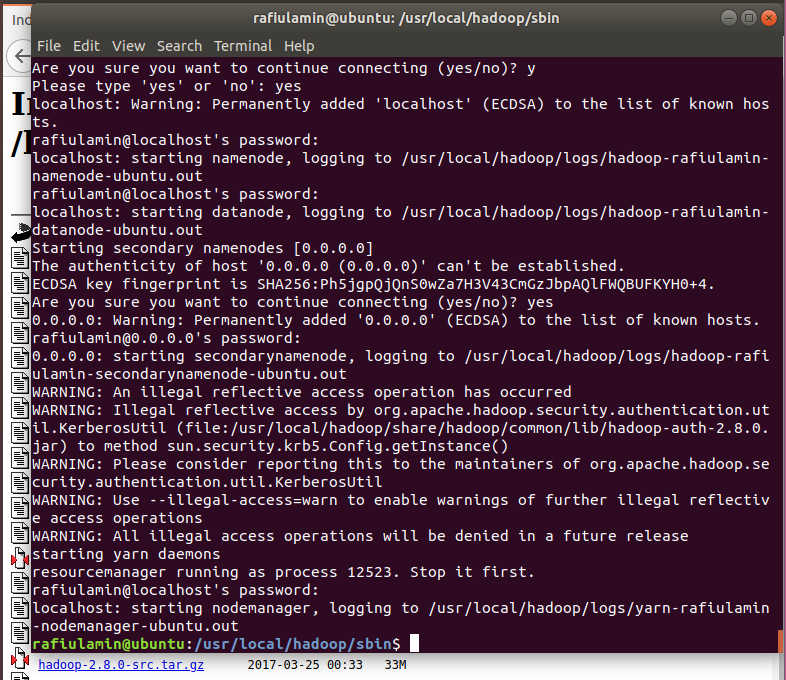
## Therefore we re-install open-ssh:



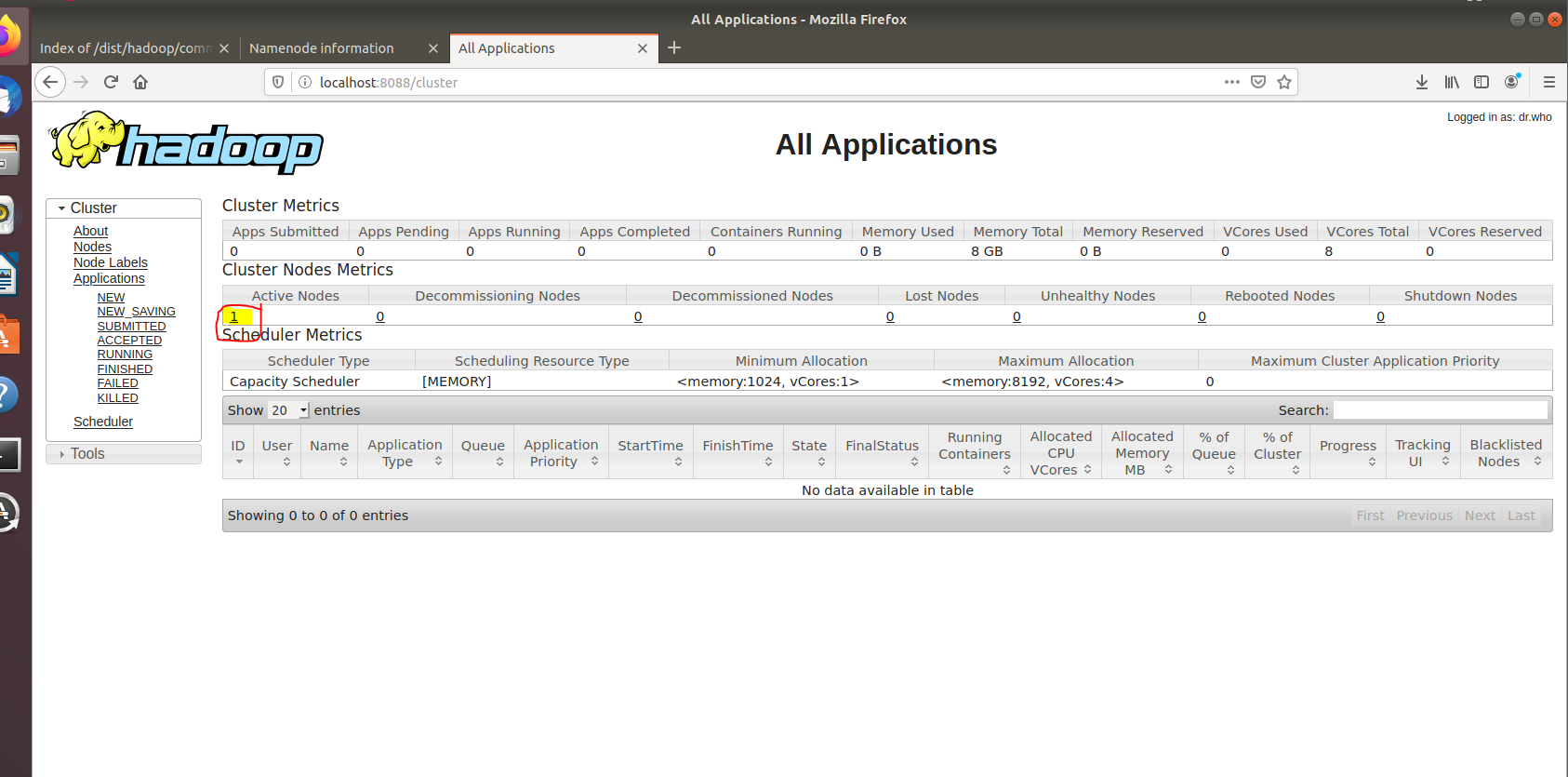
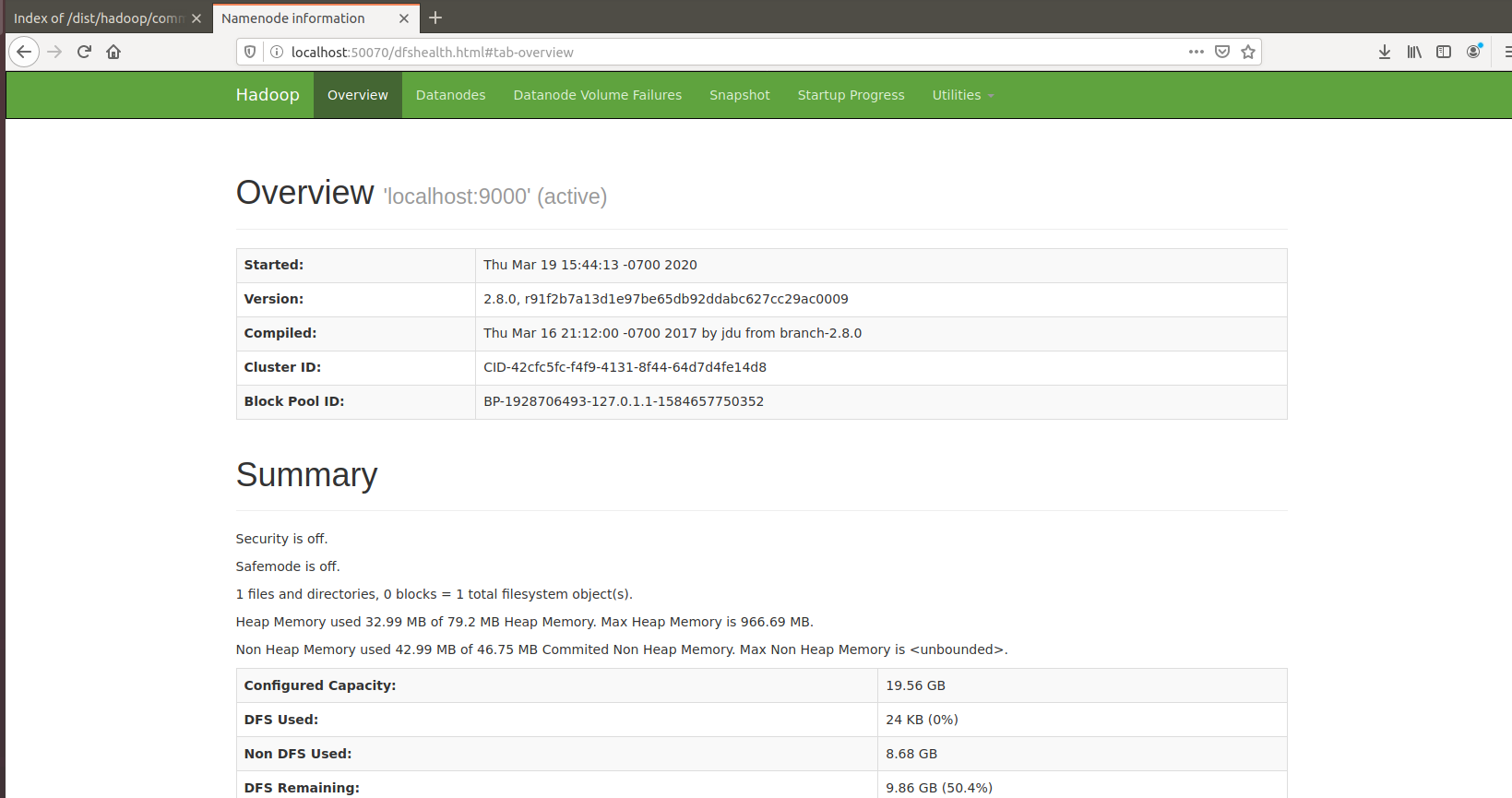
## We do the “namenode” formatting again and the run “start-all.sh” from Hadoop “sbin”:



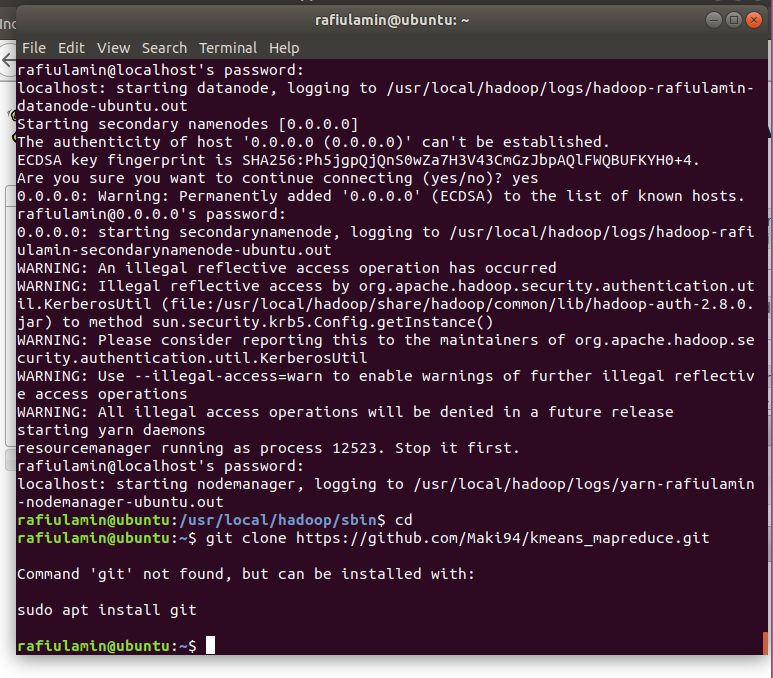
## It asks for passwords to log into different nodes. Finally all connections are established.



## We check the sites as well.



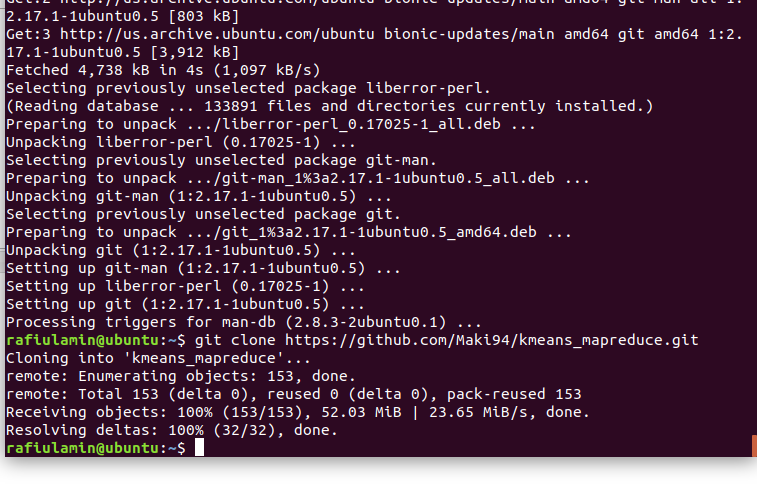
## Now we install git in order to clone kmean-mapreduce code:



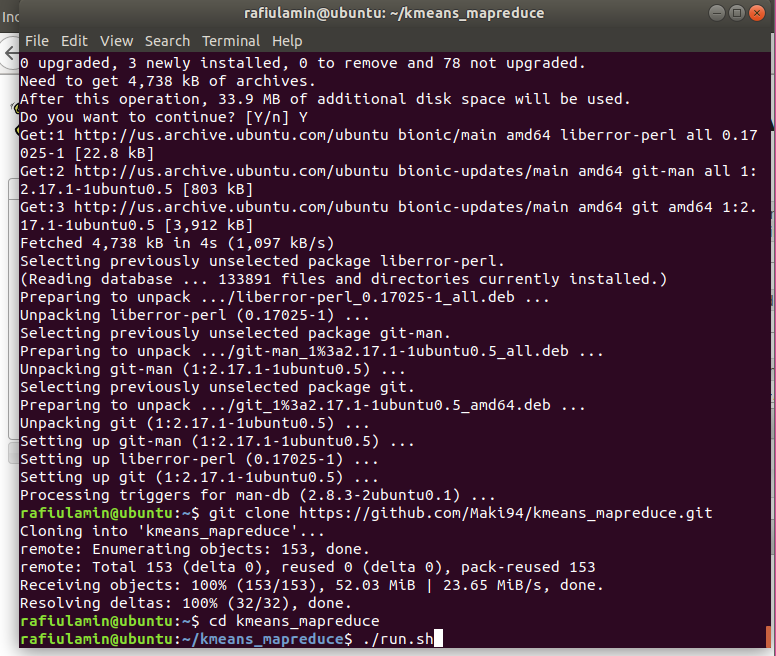
We use the following command to install git:

$sudo apt install git

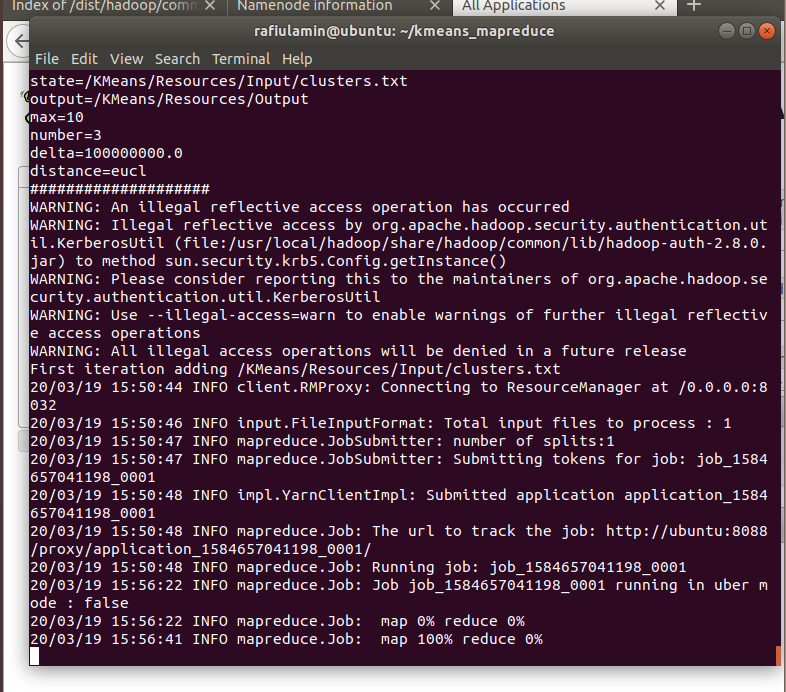
## Then we clone it from the specified git-hub link:



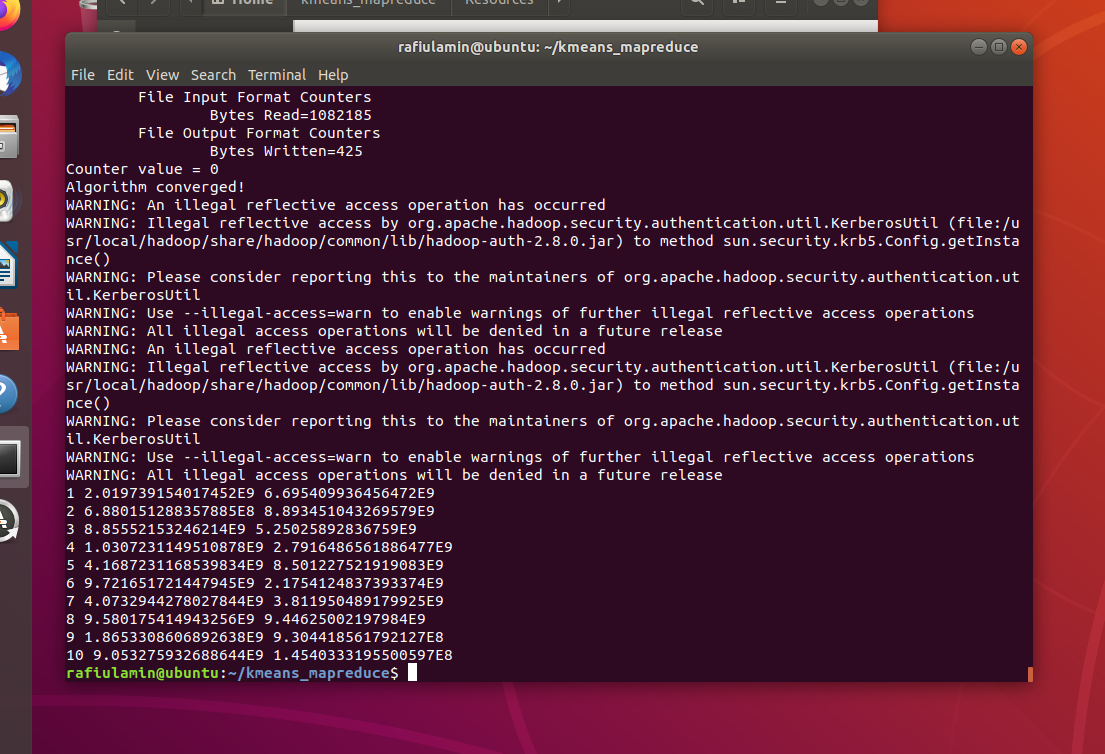
## Now is the time to run it!



## Kmeans is running in Hadoop:



## Finally it finishes running:



## We also check the site:

