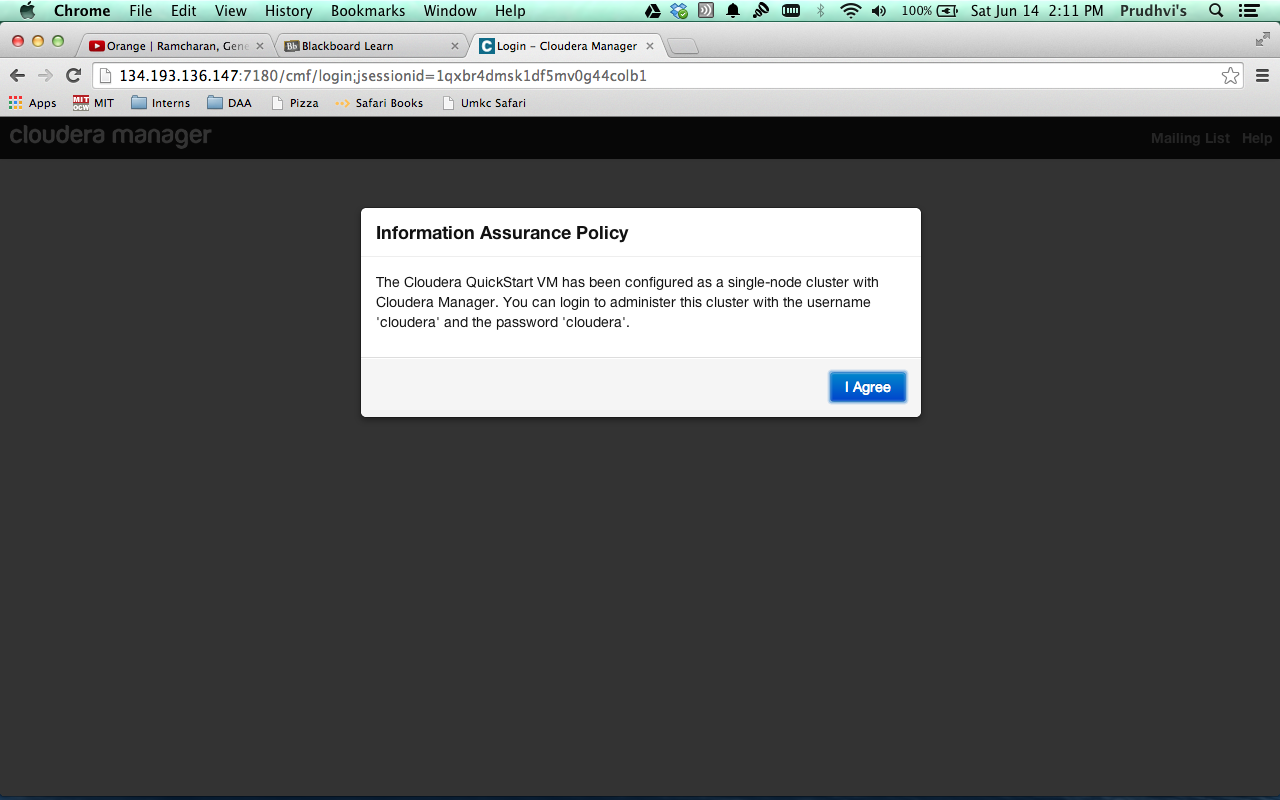
Lab 1 Assignment

Prudhvi Nalluri (16163411)

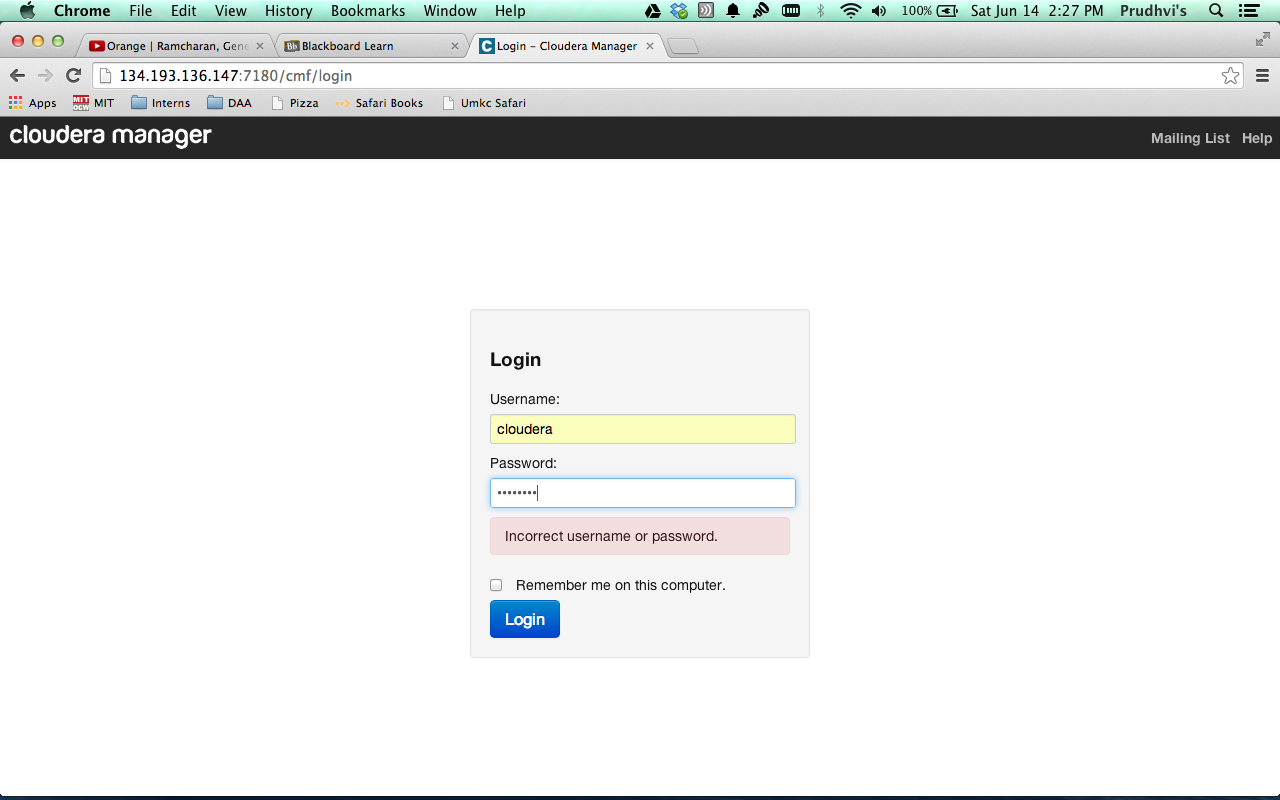
Task 2:

SubTask 1: To access UMKC Cloudera Servers

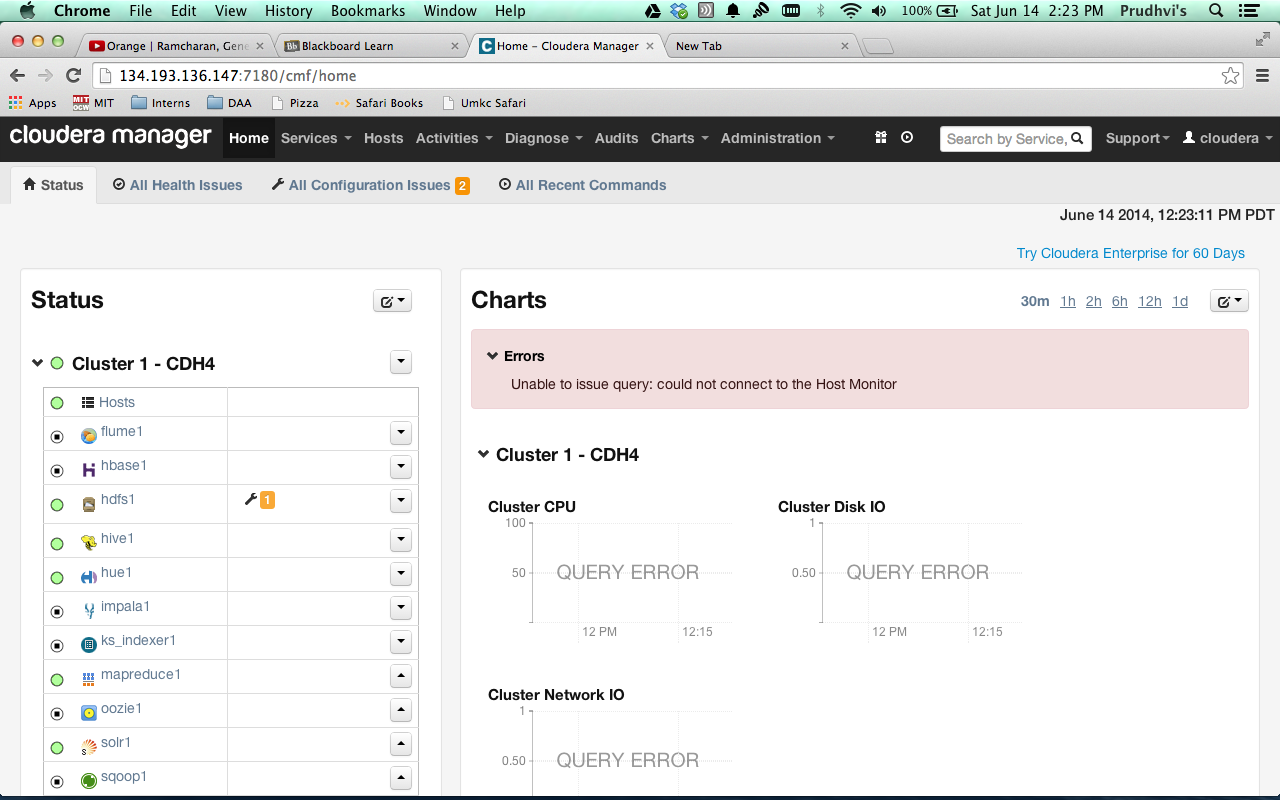
To Login into the Cloudera Manager, first connect to UMKC VPN(if you are outside UMKC) and go this address <http://134.193.136.147:7180> where 7180 is the port number to reach Cloudera manager page which looks like below image



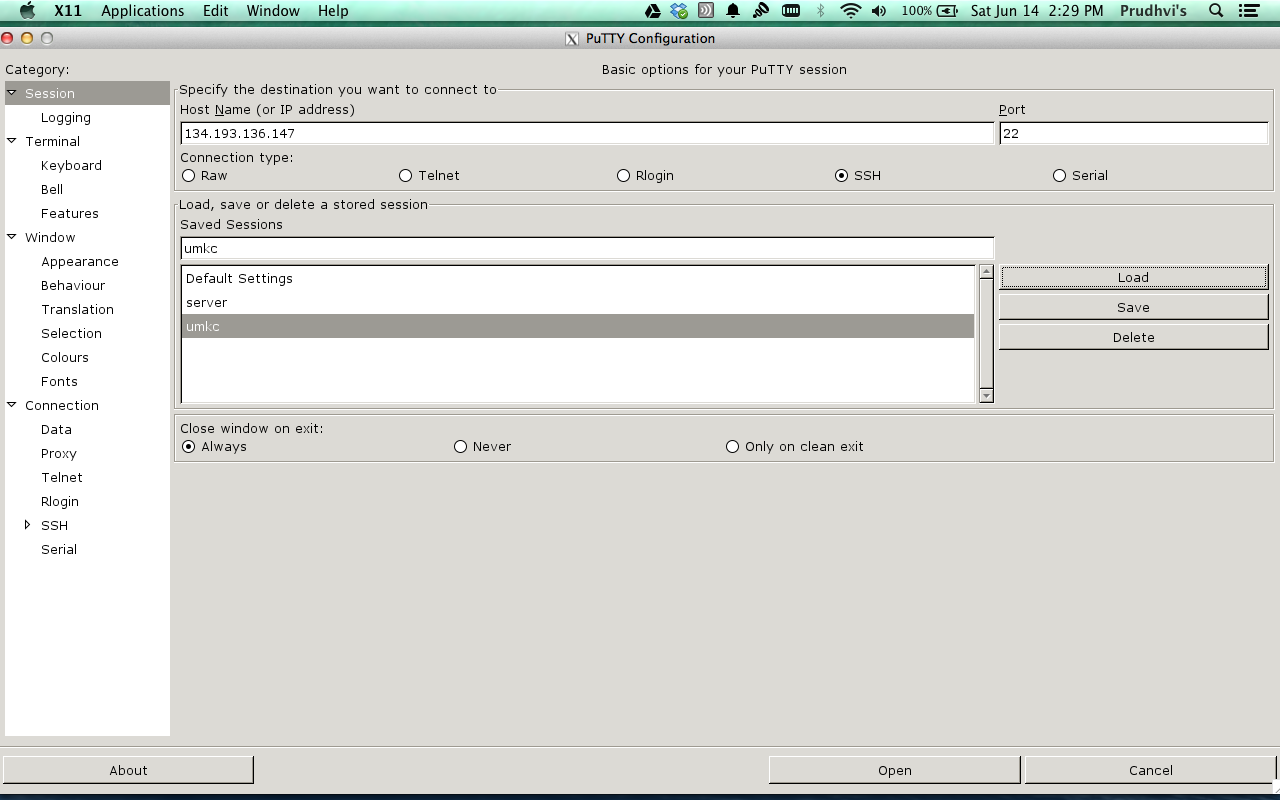
After clicking “I Agree” it will redirect to login page , login with given credentials



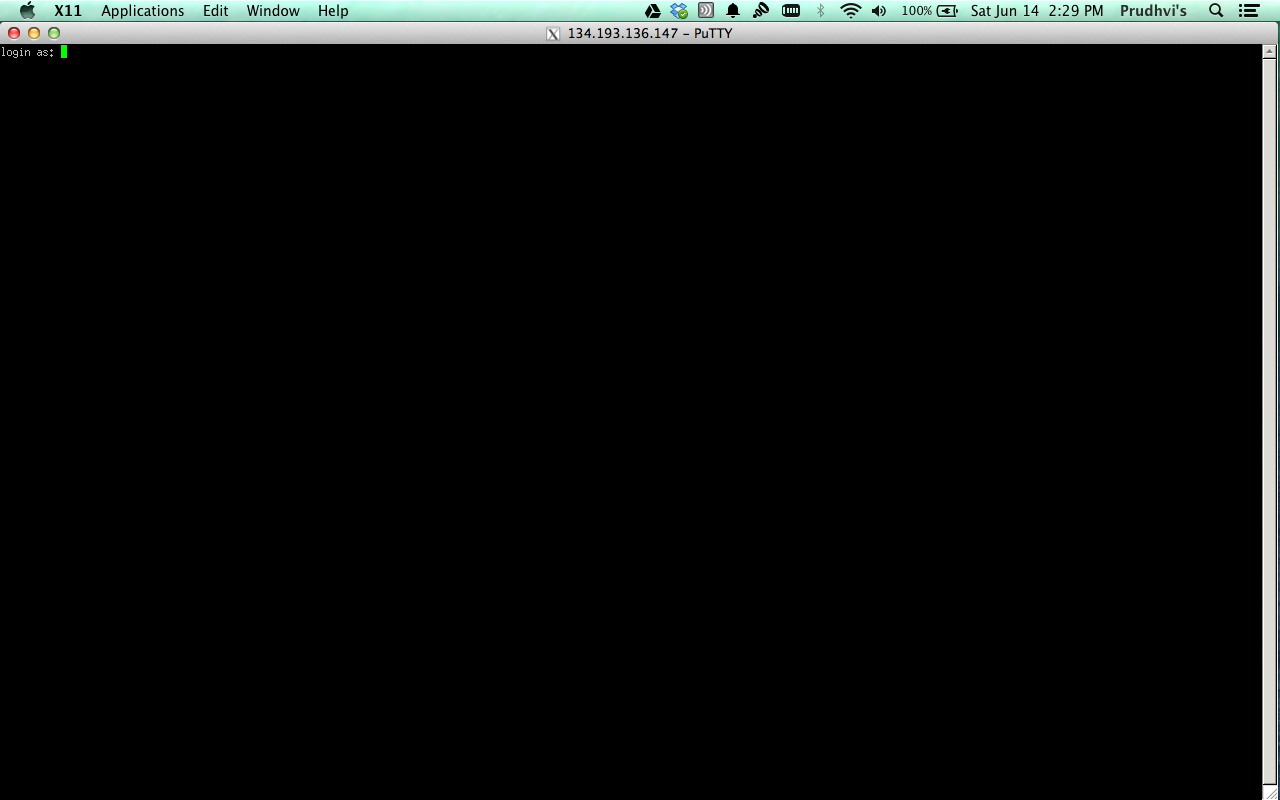
after logging into cloud manager, it will show home page of your cloudera servers



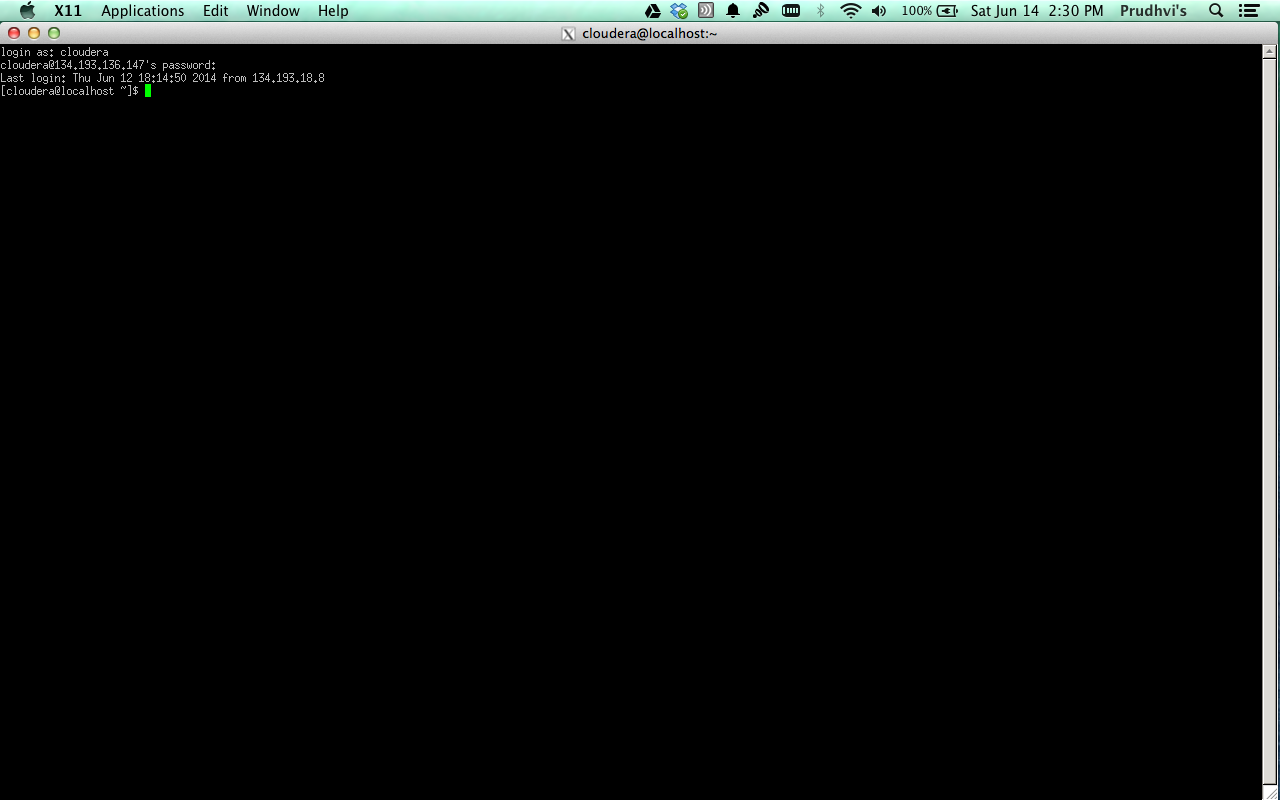
To access to the remote machine terminal, download the putty and open and go the session tab on the left side and enter the Host address, port number and save the session for the future use and click on open.



This will open a remote machine terminal where we have to our credentials as shown below

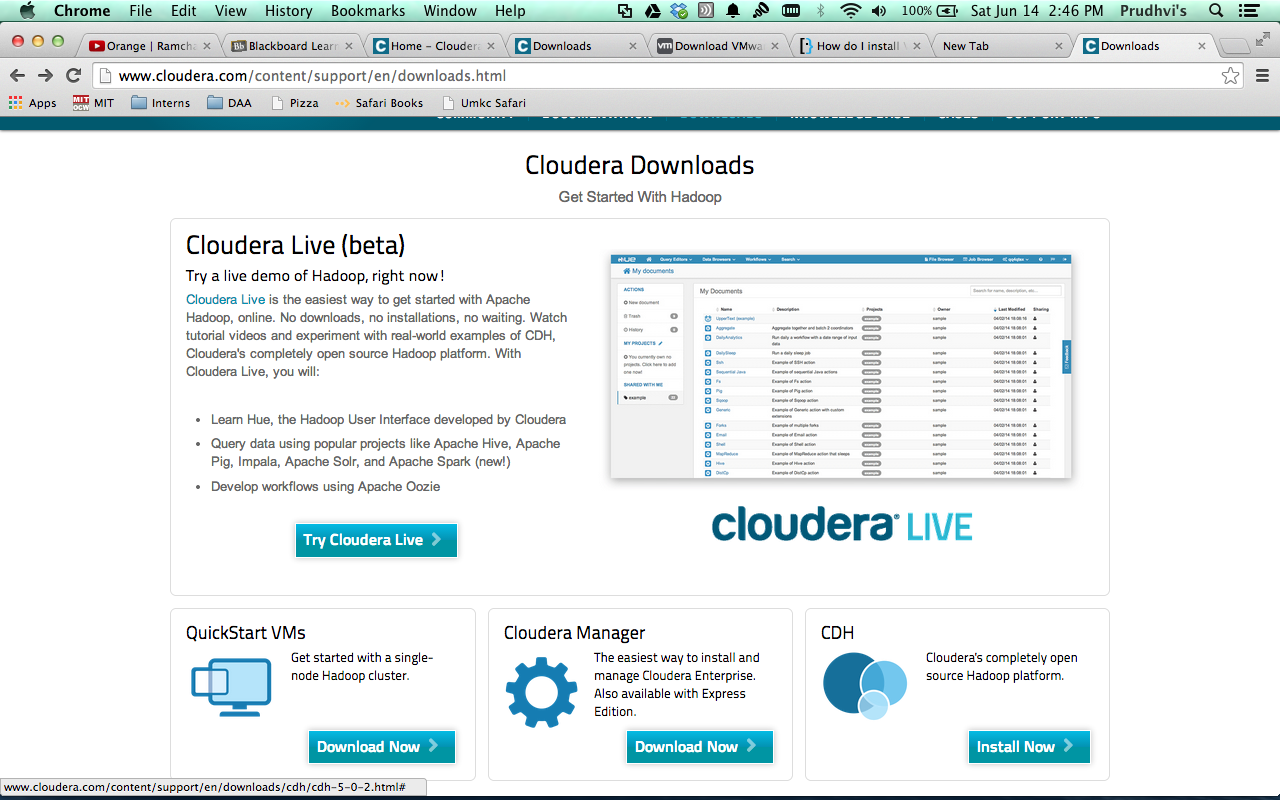


After successfully logging in we can access to our respective remote machine

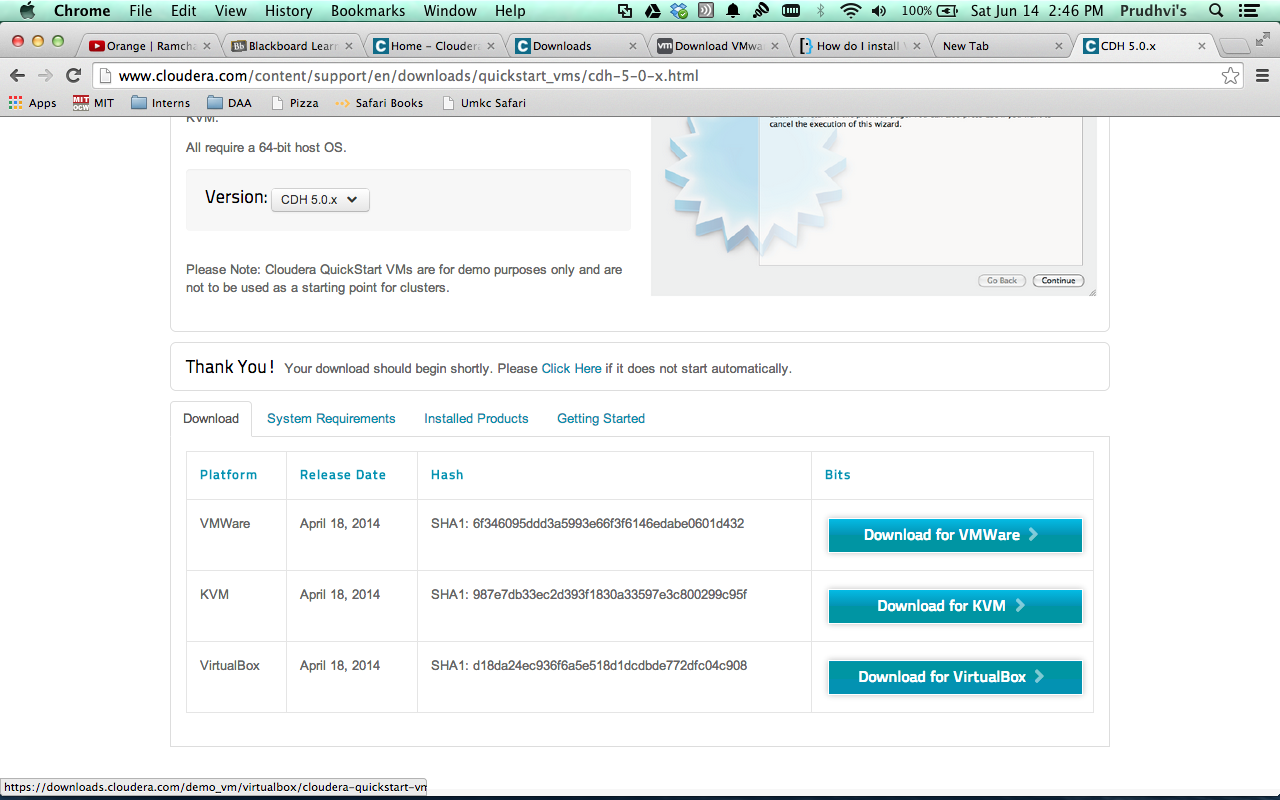


Subtask 2: installing own Cloudera Server

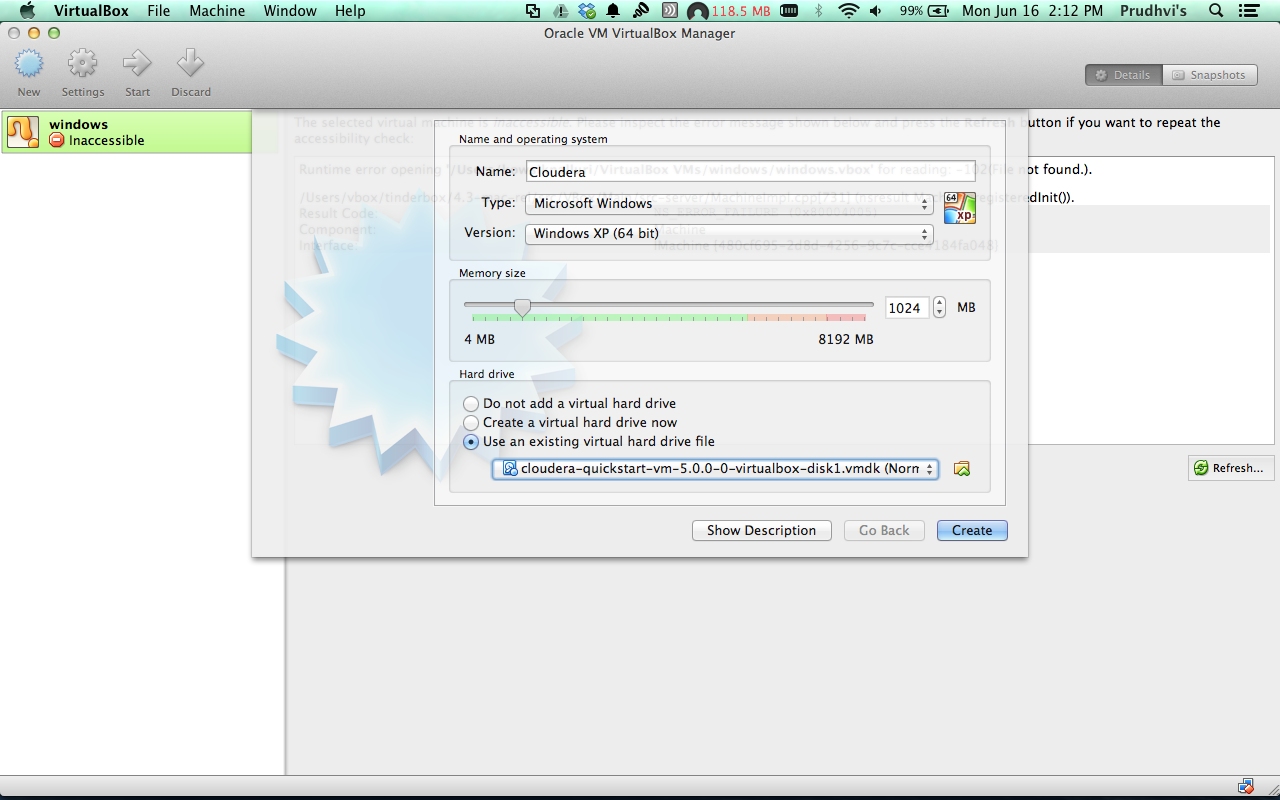
Download the QuickStart VMS based upon your platform(virtual box or vmware)



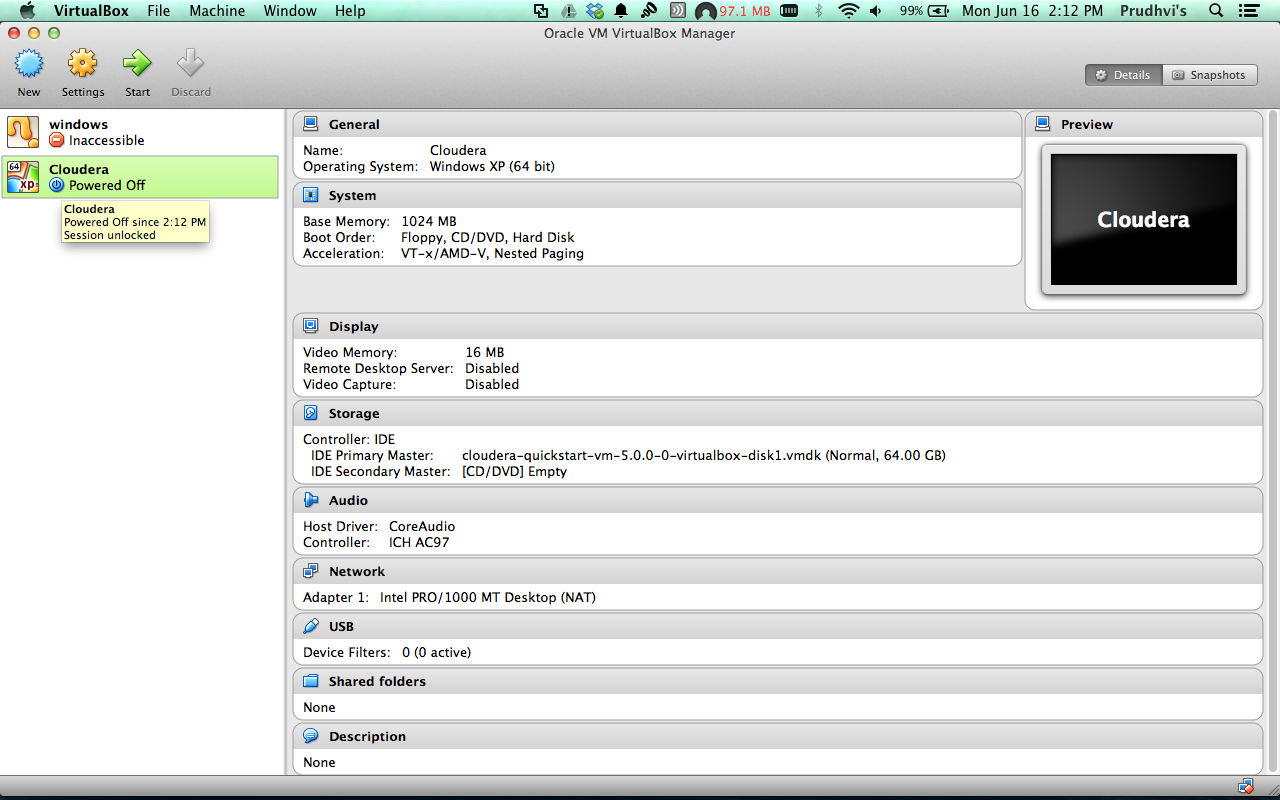
select the version and click download based upon your Platform



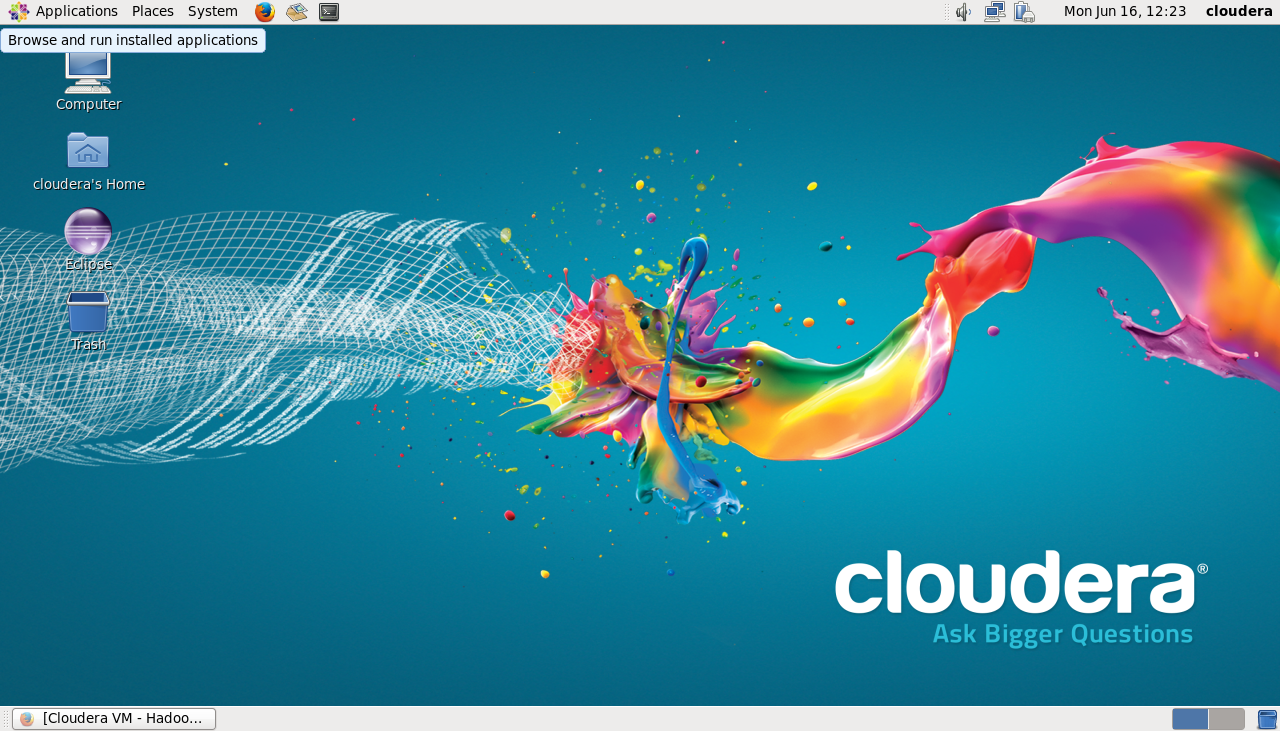
Open the virtual box, click new and assign name and memory size and browse cloudera image which we downloaded earlier and click create.



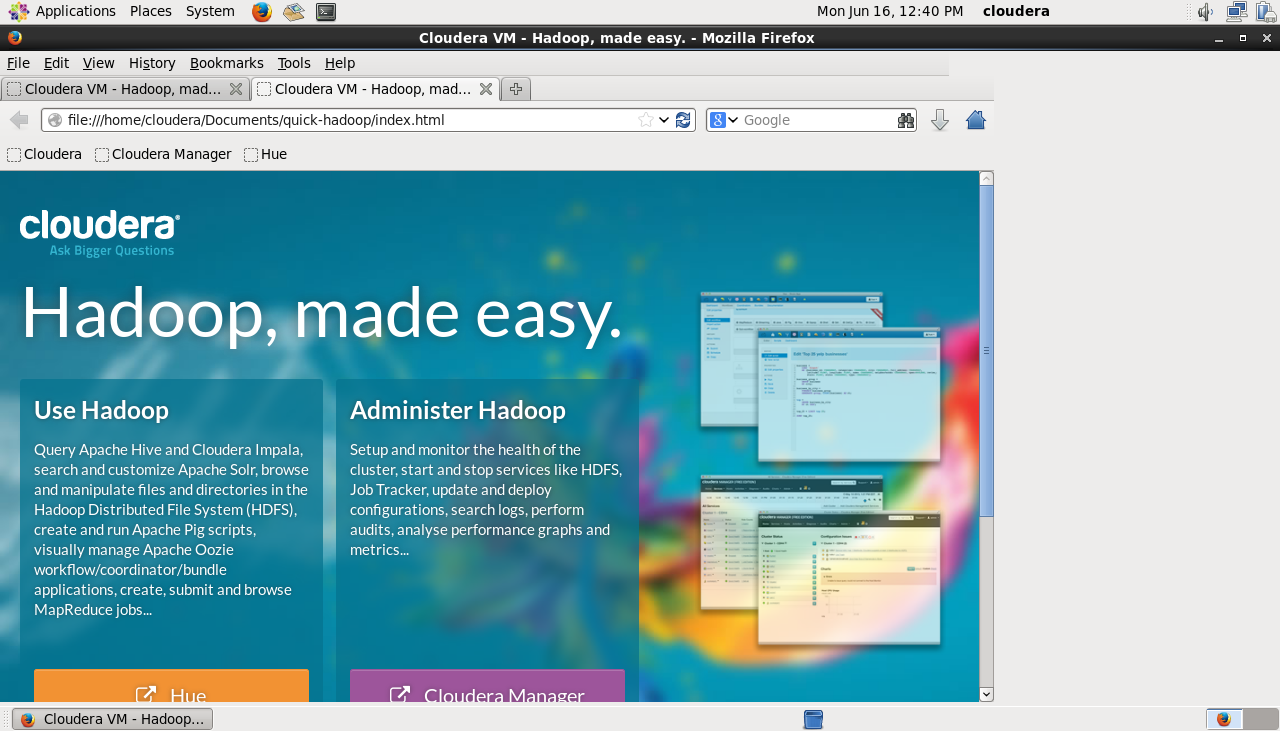
This will create a virtual machine which will in powered off mode and click on start to start the virtual machine.



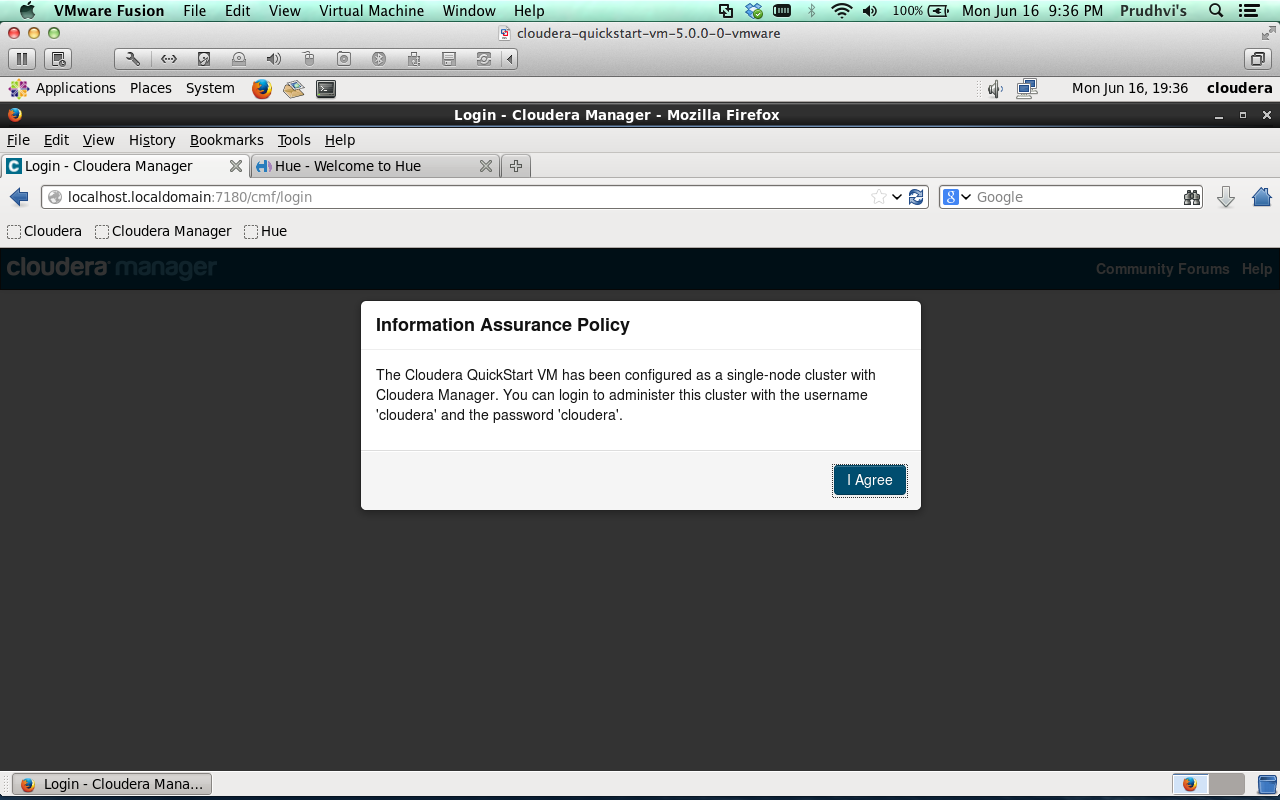
after the virtual machine started this will show the desktop as shown below



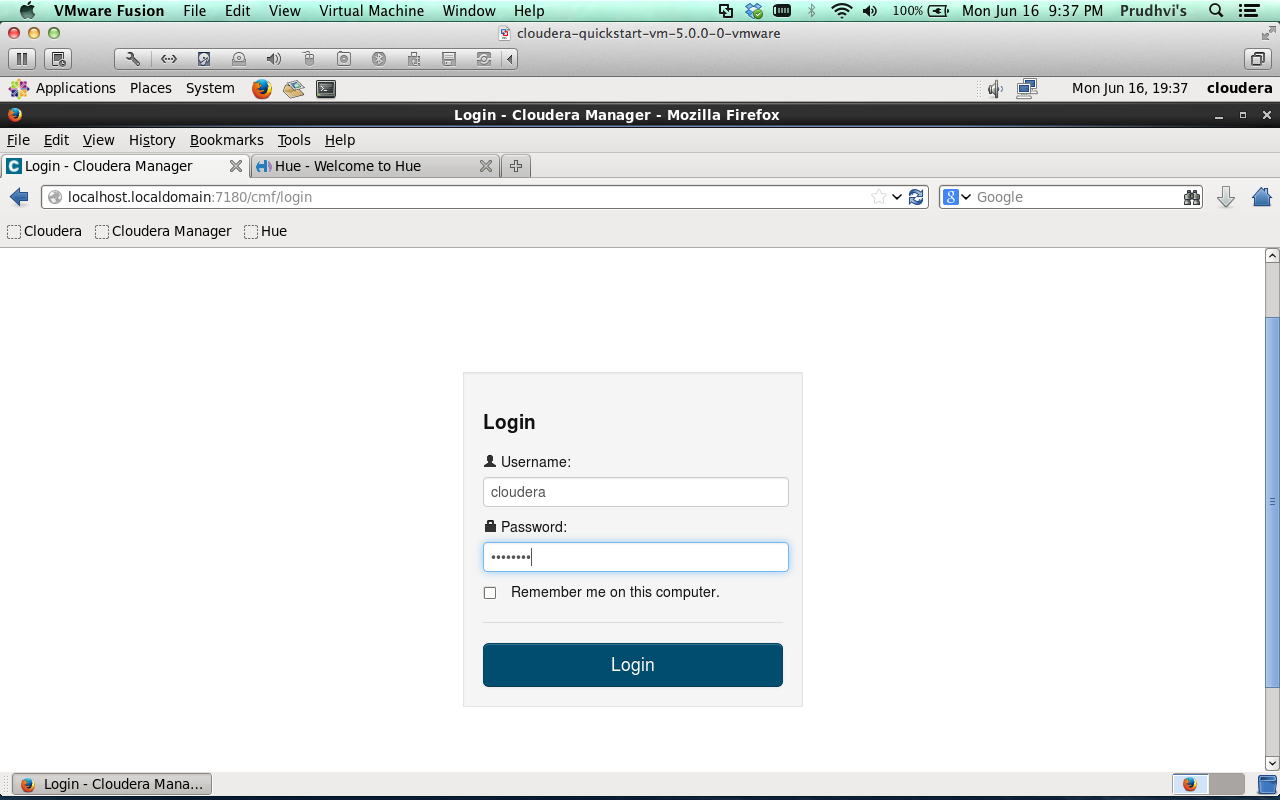
and by default it will firefox which will contain cloudera manager and hue tabs



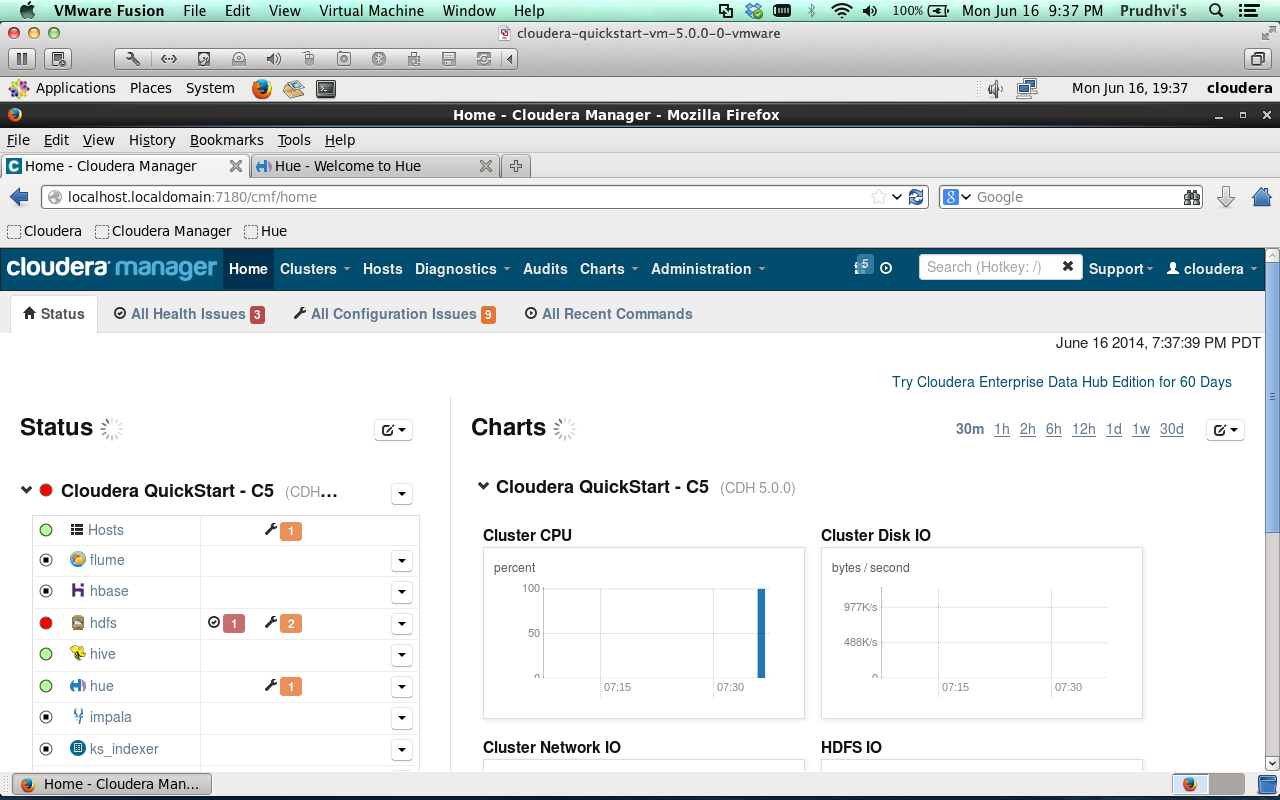
Click on the cloudera manager, which will pop up a alert and click on “I Agree”



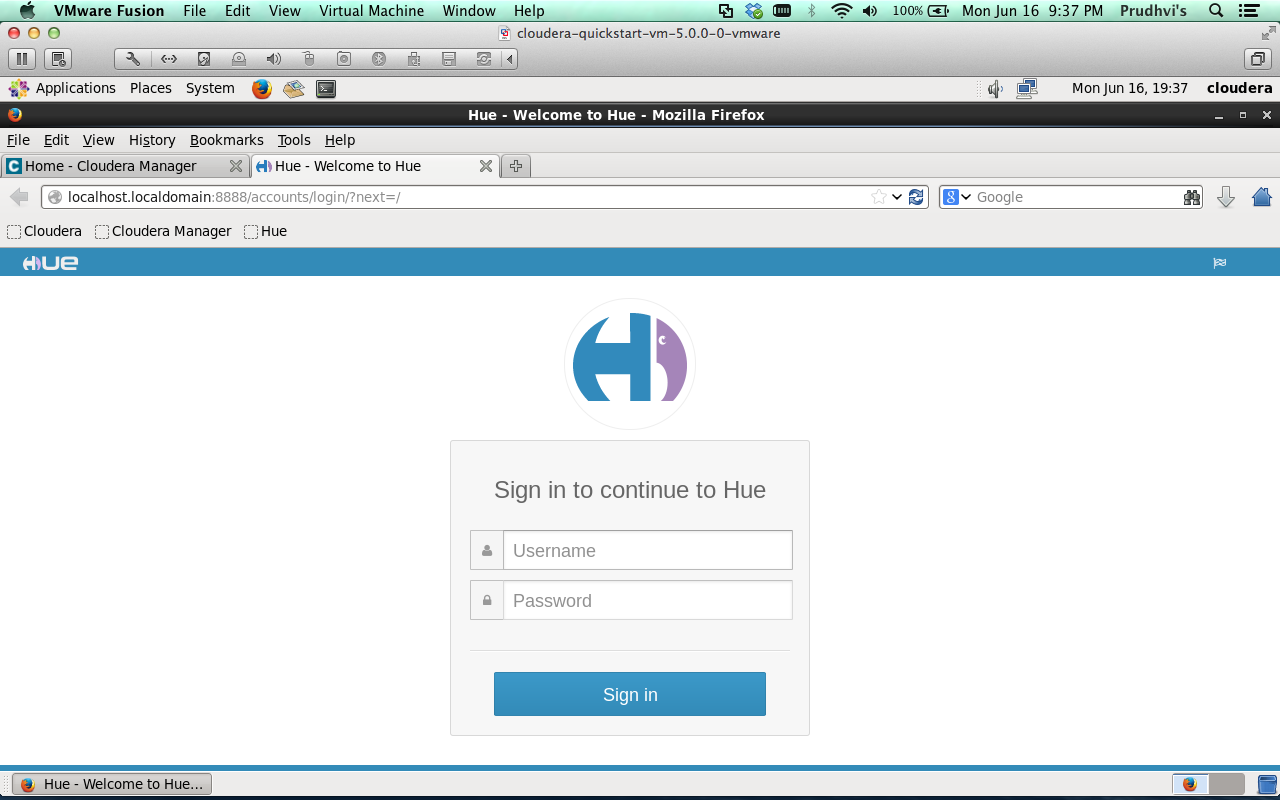
and login with username and password “cloudera”



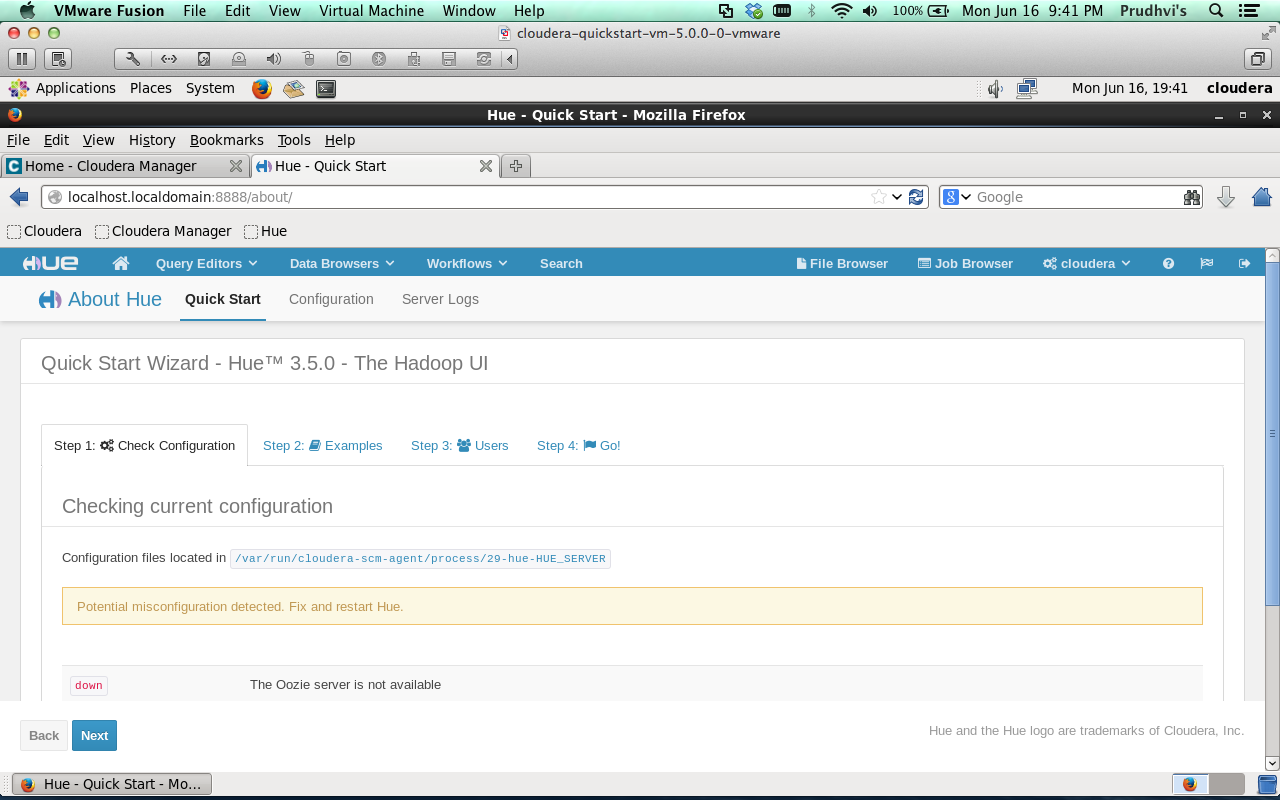
and it will redirect to your cloudera manager home.



and in the same way we can login to the hue with same username and password

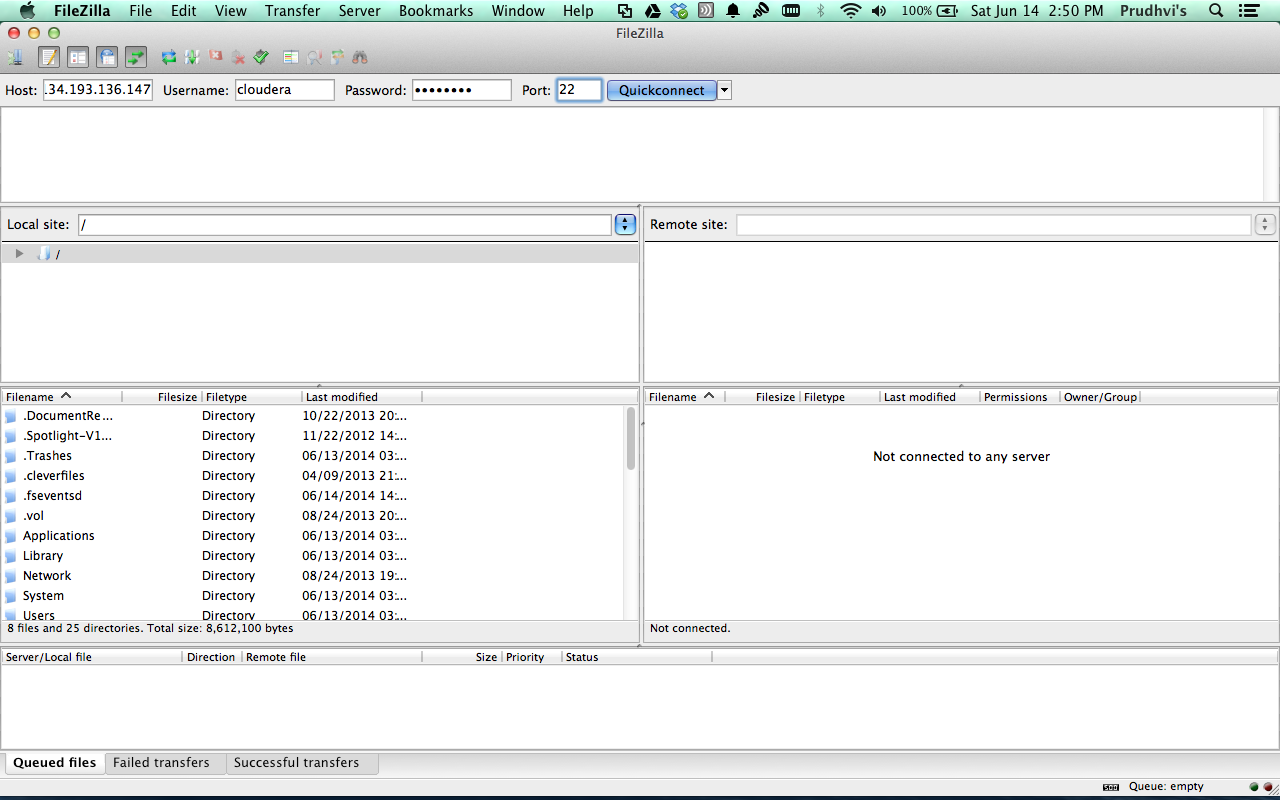


It will redirect to hue home which is a Hadoop UI (we don’t worry about using the command line).

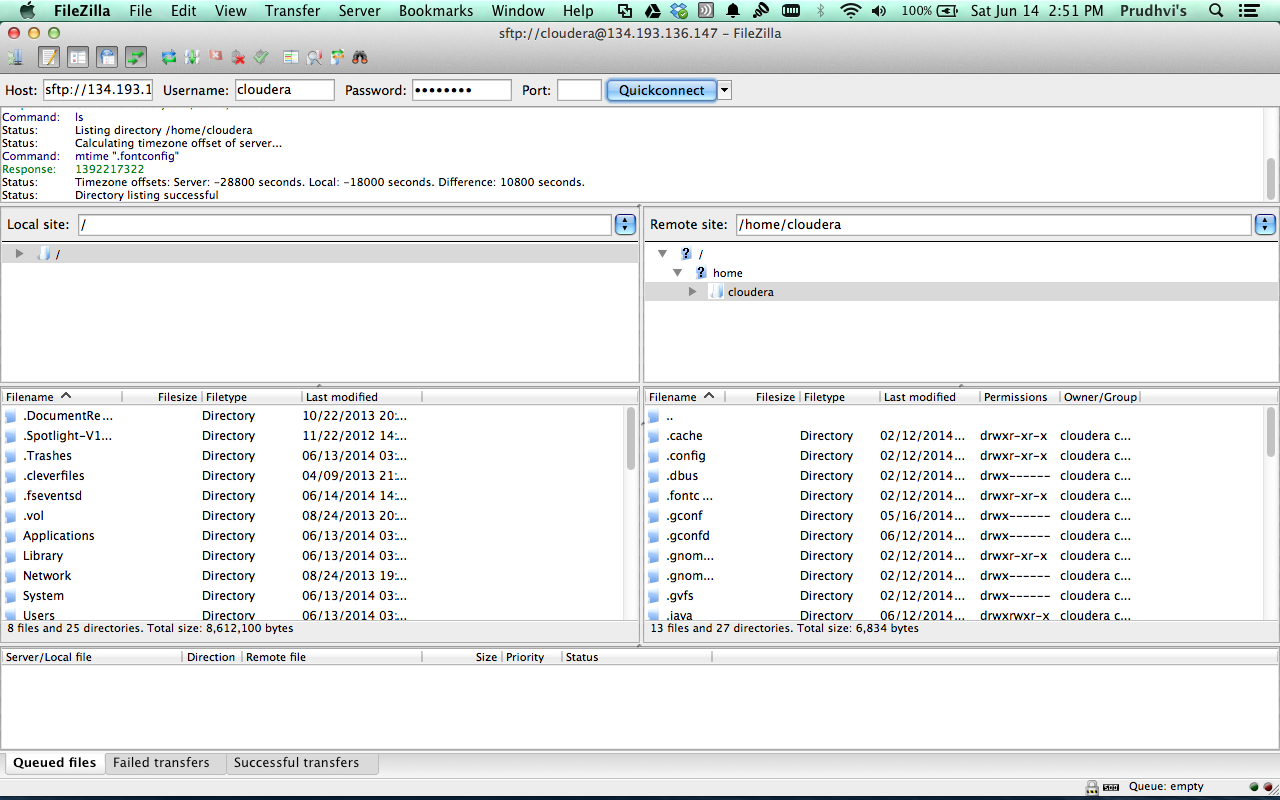


SubTask 3: To Transfer the files to cloudera using FileZilla

Download the FileZilla, install it and to transfer files through FileZilla the Host address, Username, Password and port are entered and click Connect and the FileZilla is connected to the server.

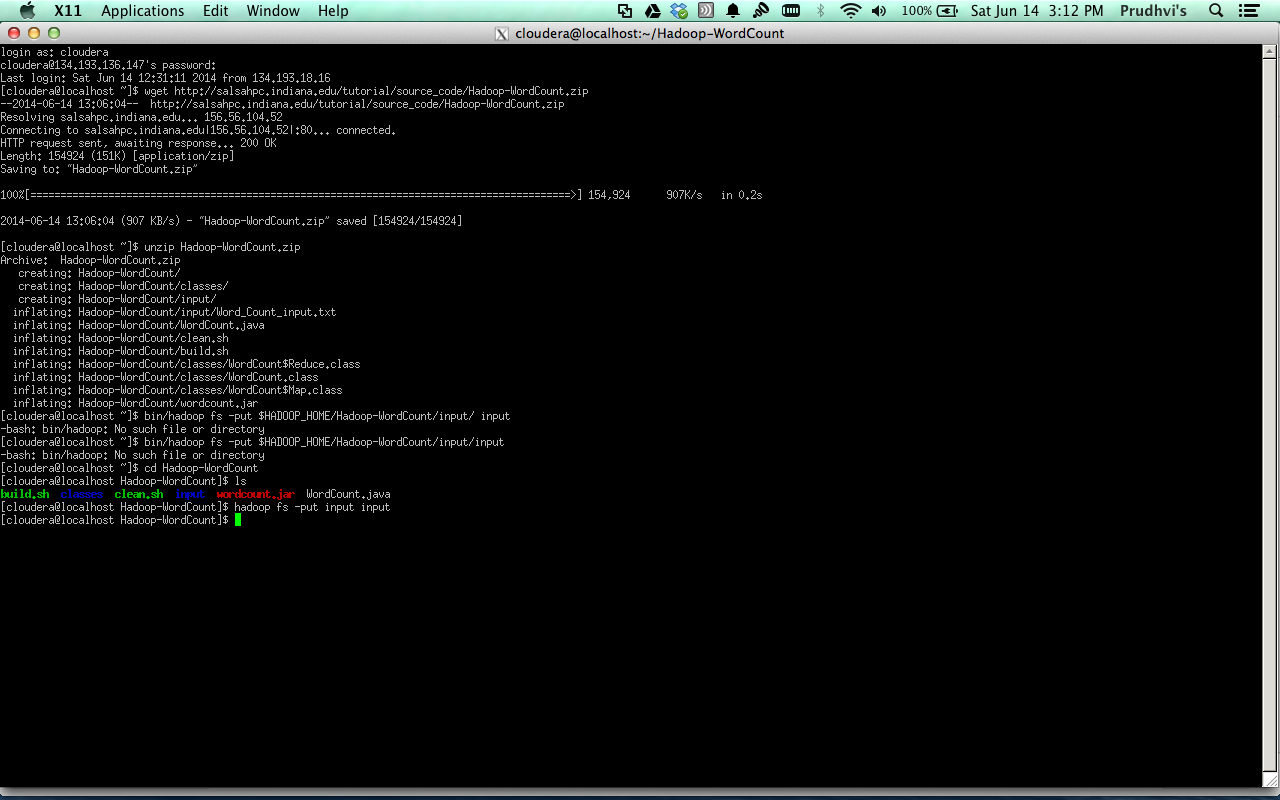


Here the FileZila is connected to Remote site, on the left side is your local site and right side is your remote site and we can easily transfer files by using drag and drop.



SubTask 4: Word Count

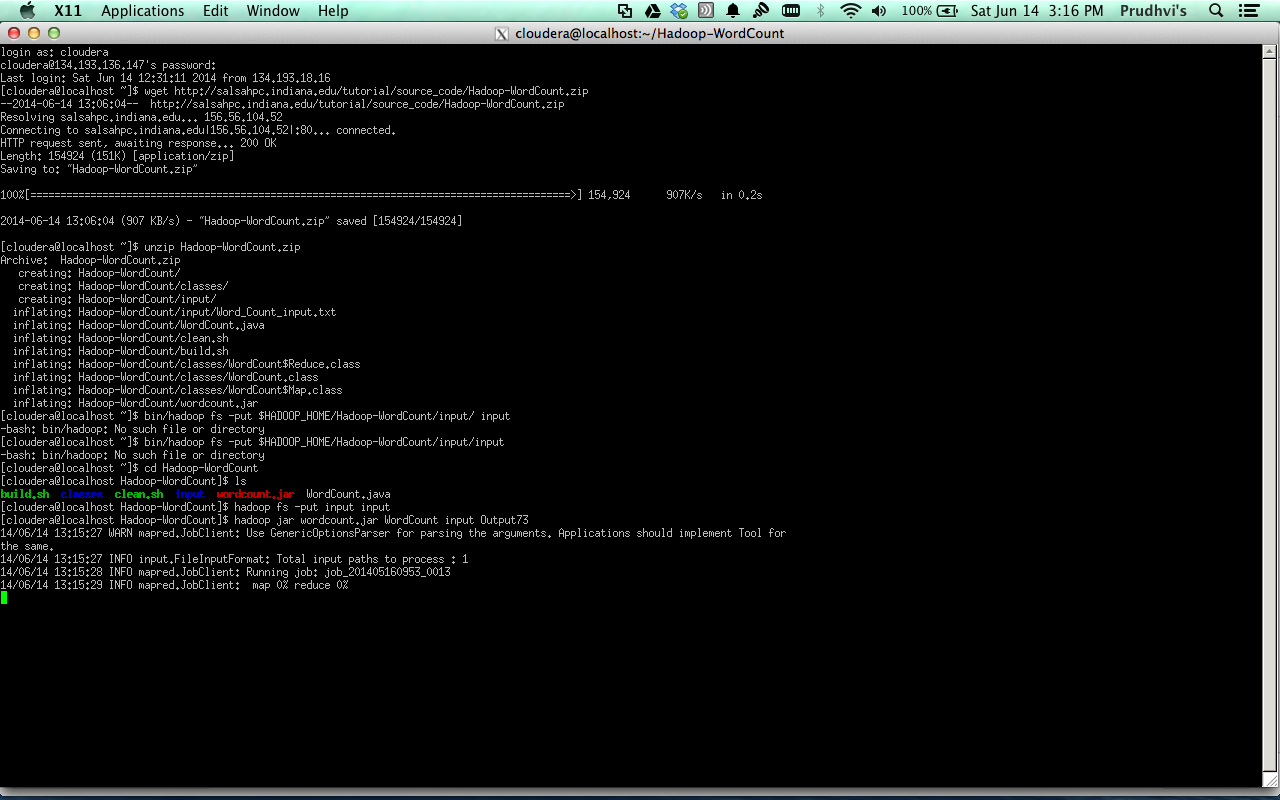
1. Download the Wordcount example from <https://portal.futuregrid.org/manual/hadoop-wordcount> and unzip it. It will contain the Wordcount java program which will use map –reduce to count and jar file and some other classes.
2. And open the putty connect to umkc session which we saved earlier and login with your credential’s and go to the Hadoop-WordCount folder which we downloaded in step 1
3. And run “hadoop fs –put input input” here we are uploading our input to the hadoop system (Hdfs) as input.



1. We uploaded the input into Hdfs and now run the WordCount program by using the command

“hadoop jar wordcount.jar input output73”

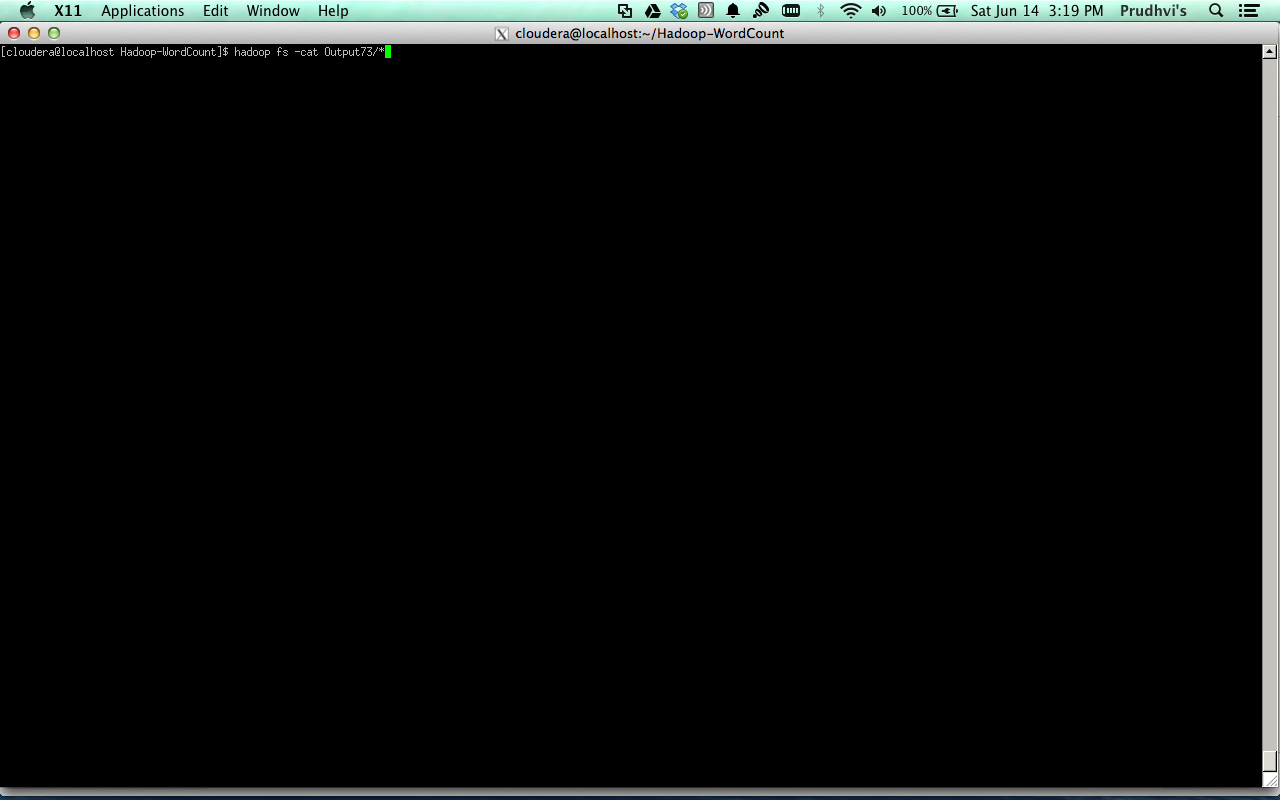
where input is the input to our program and the output73 is stored in output folder.



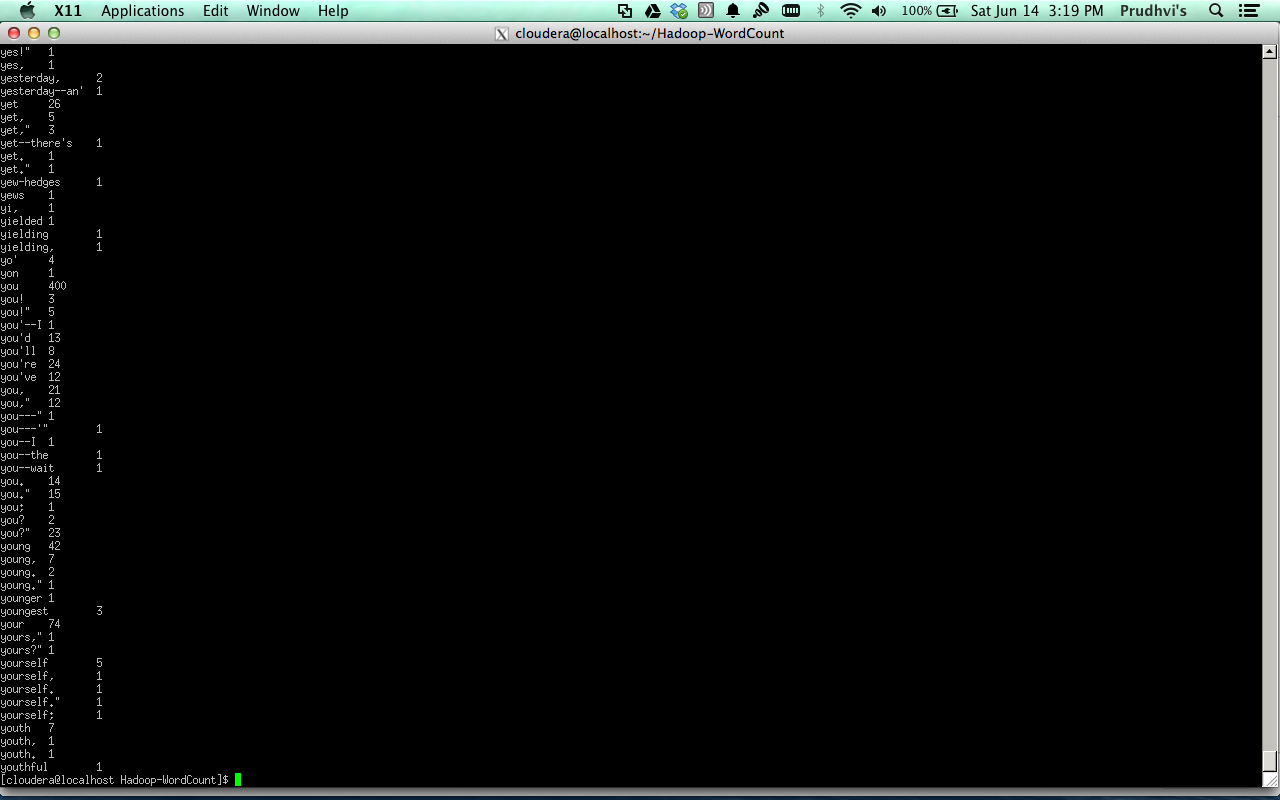
5) After running the wordcount program, now we have the run the command to check the output

“hadoop fs -cat output73/\*”

This command will show the output in the output73 folder (where we saved the output).

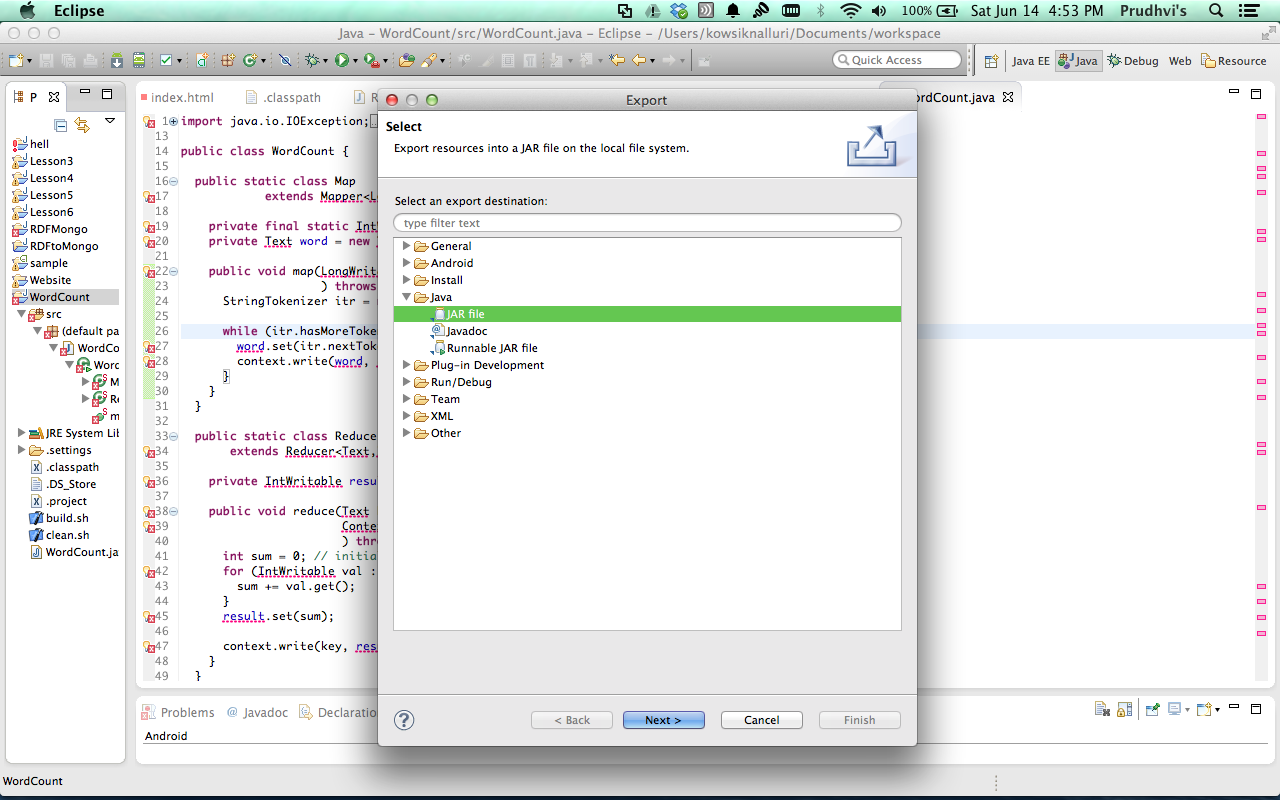


This is our output of the wordcount program.

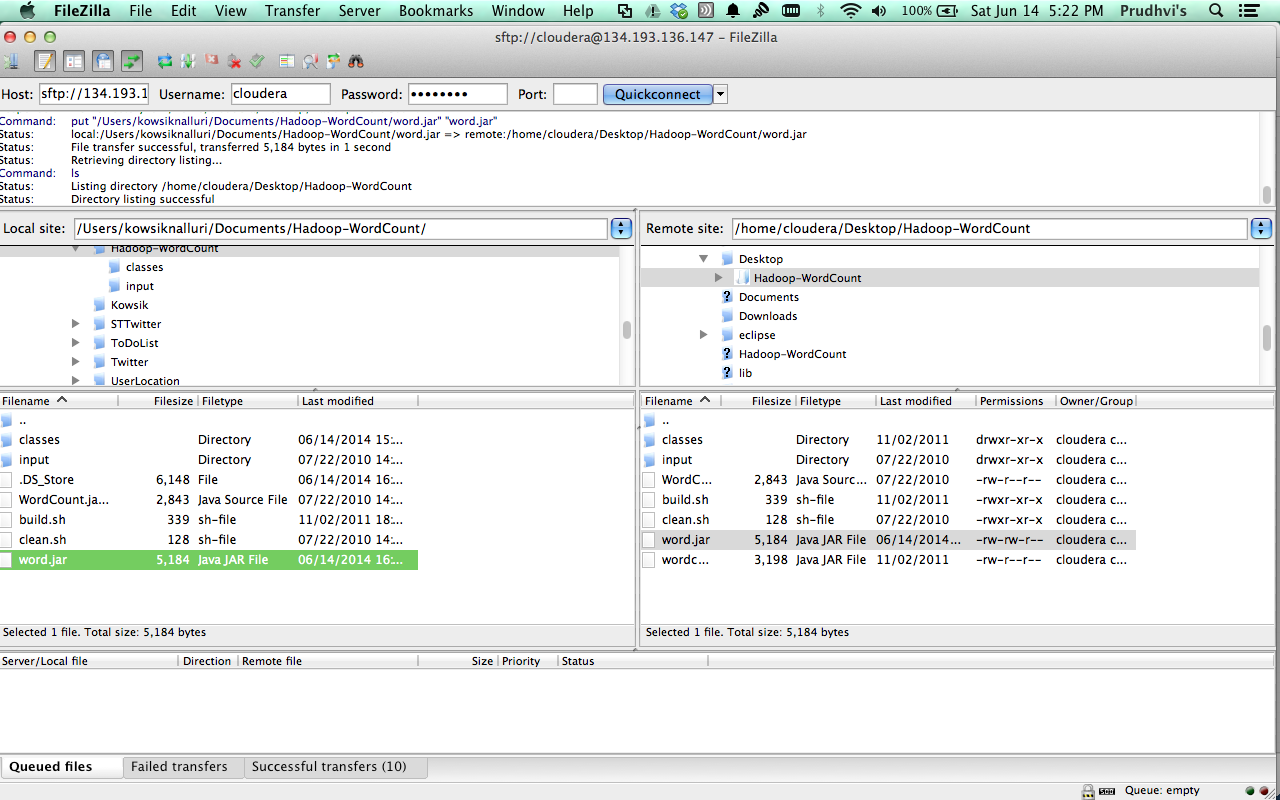


SubTask 5: How to make a hadoop jar and run it in cloudera

1. Open Eclipse and import the WordCount project which we downloaded before and open file and click on the export , it will create a jar file and save the jar file in local.



1. Now open the FileZilla to transfer the jar file into the remote site. Drag the jar file in the local site to the remote and drop it.



Now open putty and connect to remote machine terminal and go to the folder where we saved the jar file and run the command

“hadoop jar word.jar input output13”

this will create the output.

