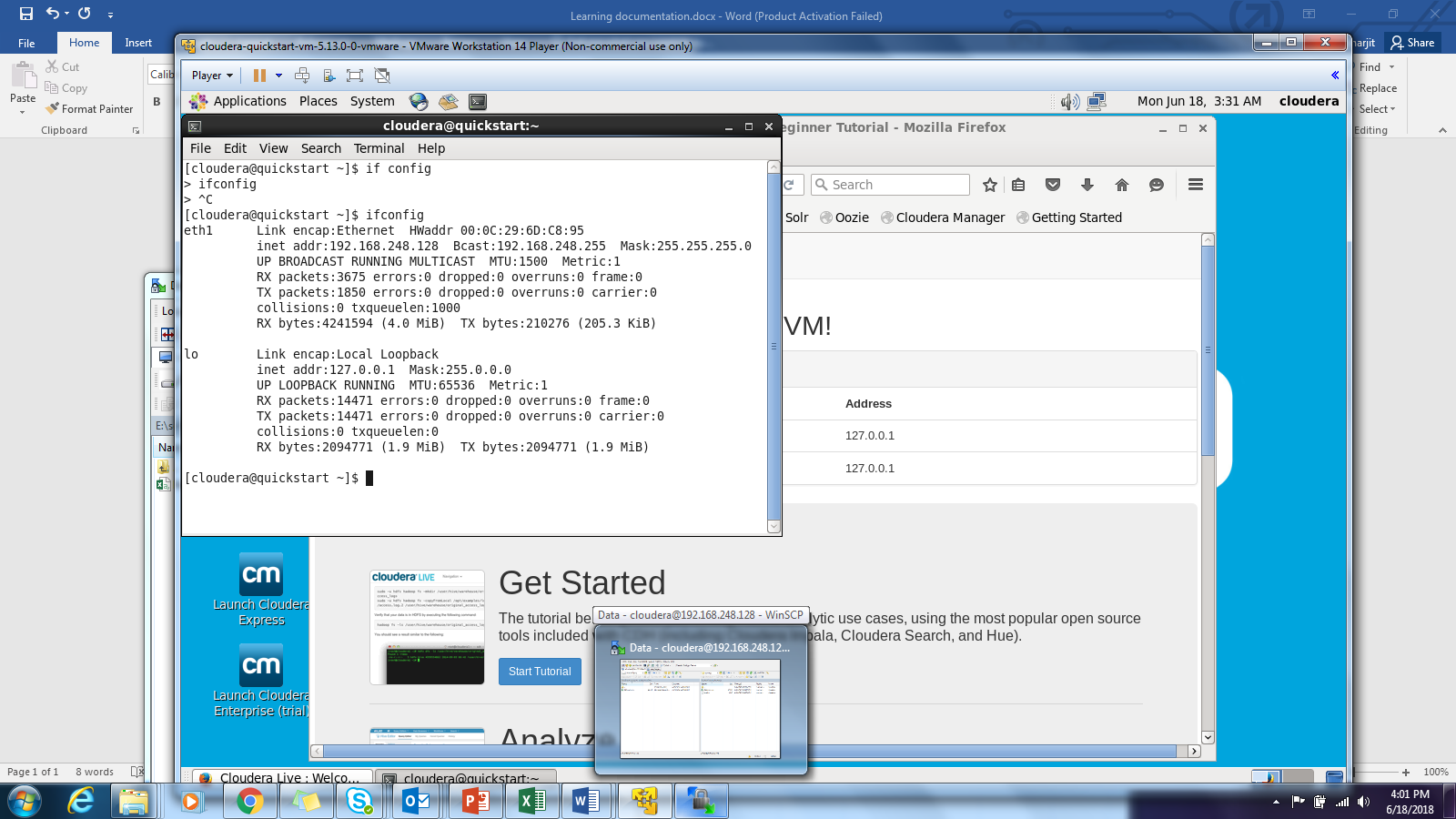
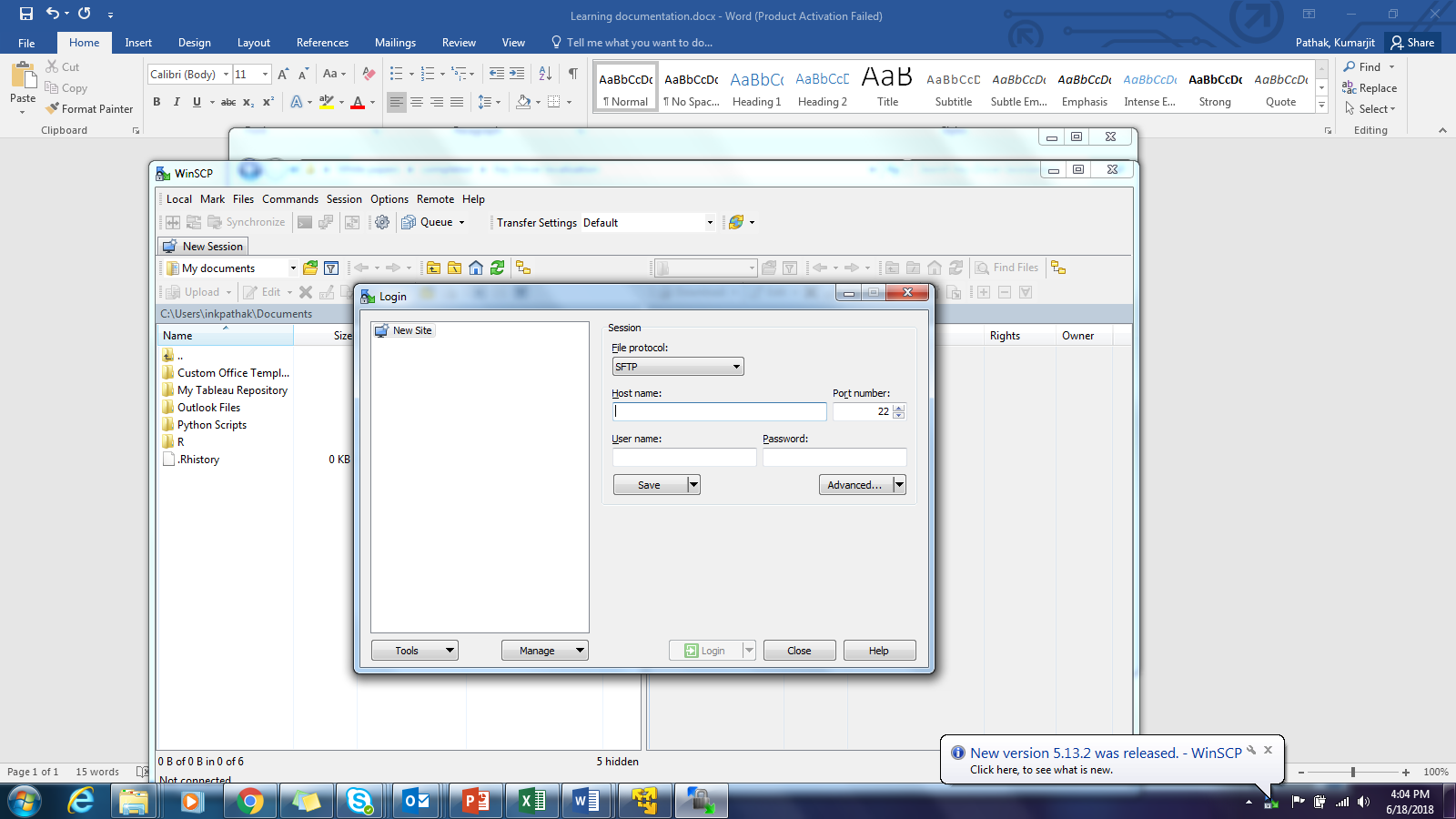
Comand prompt:

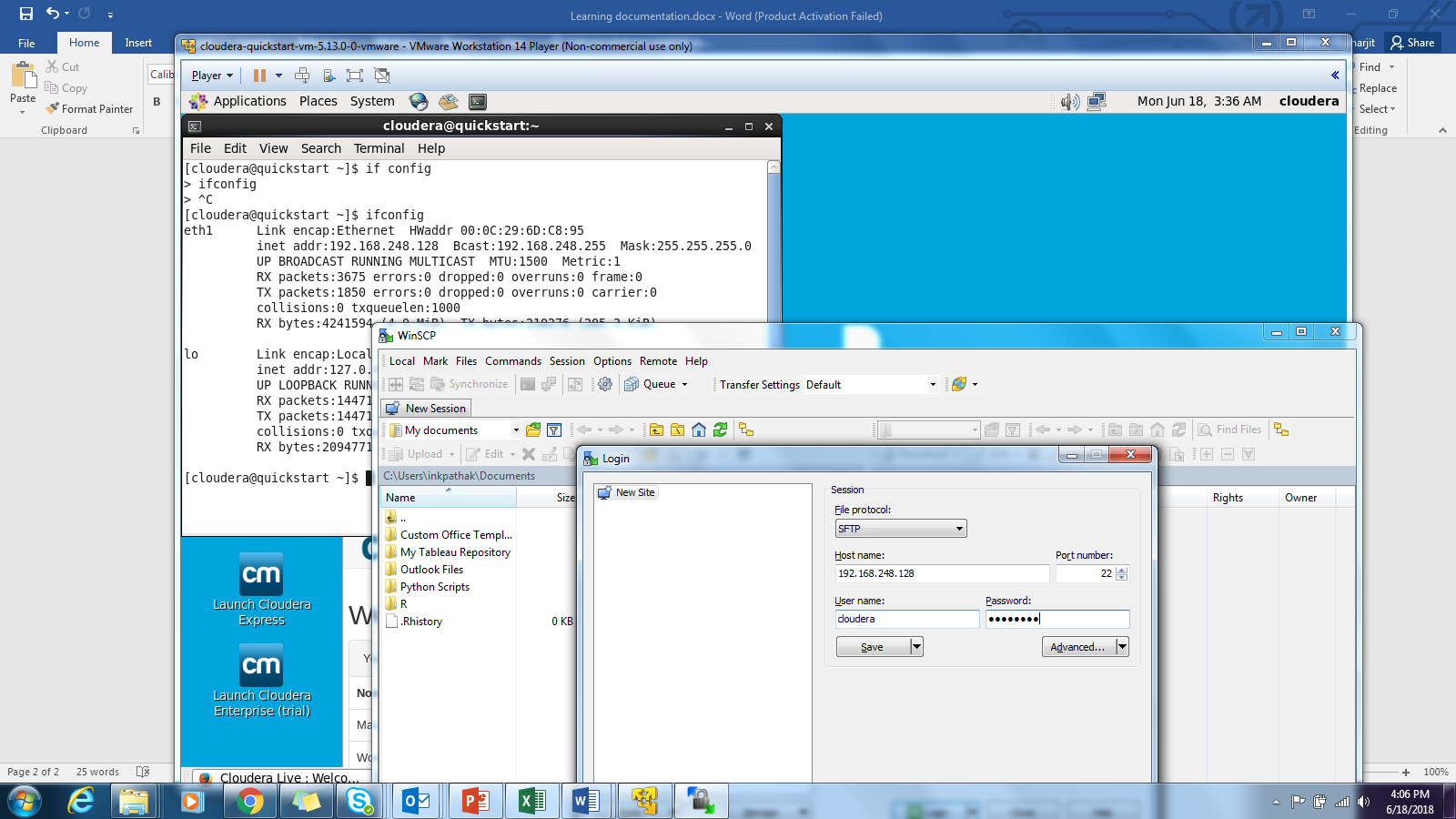
ifconfig : to understand IP address



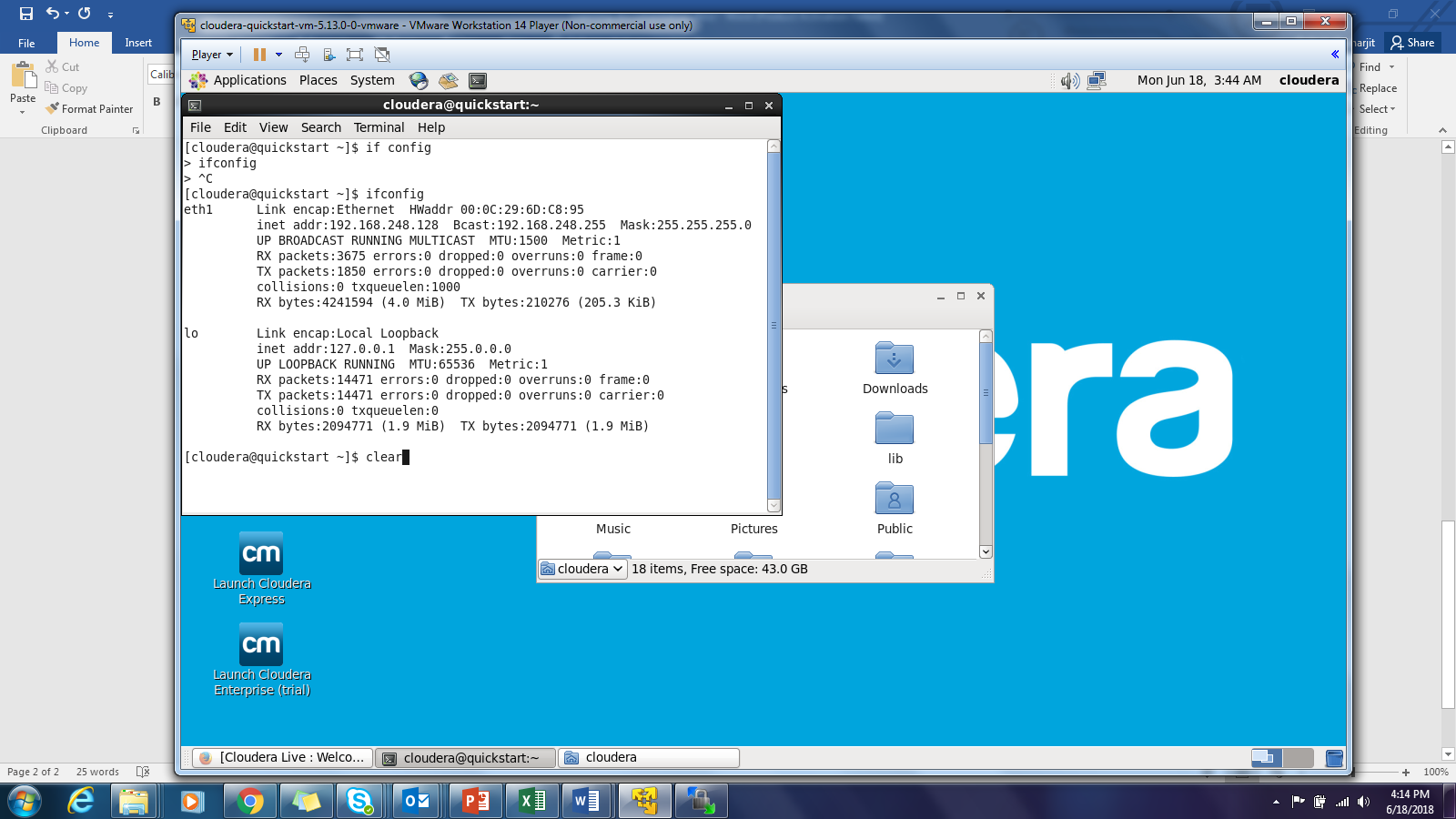
Copy the IP address from Inet Address

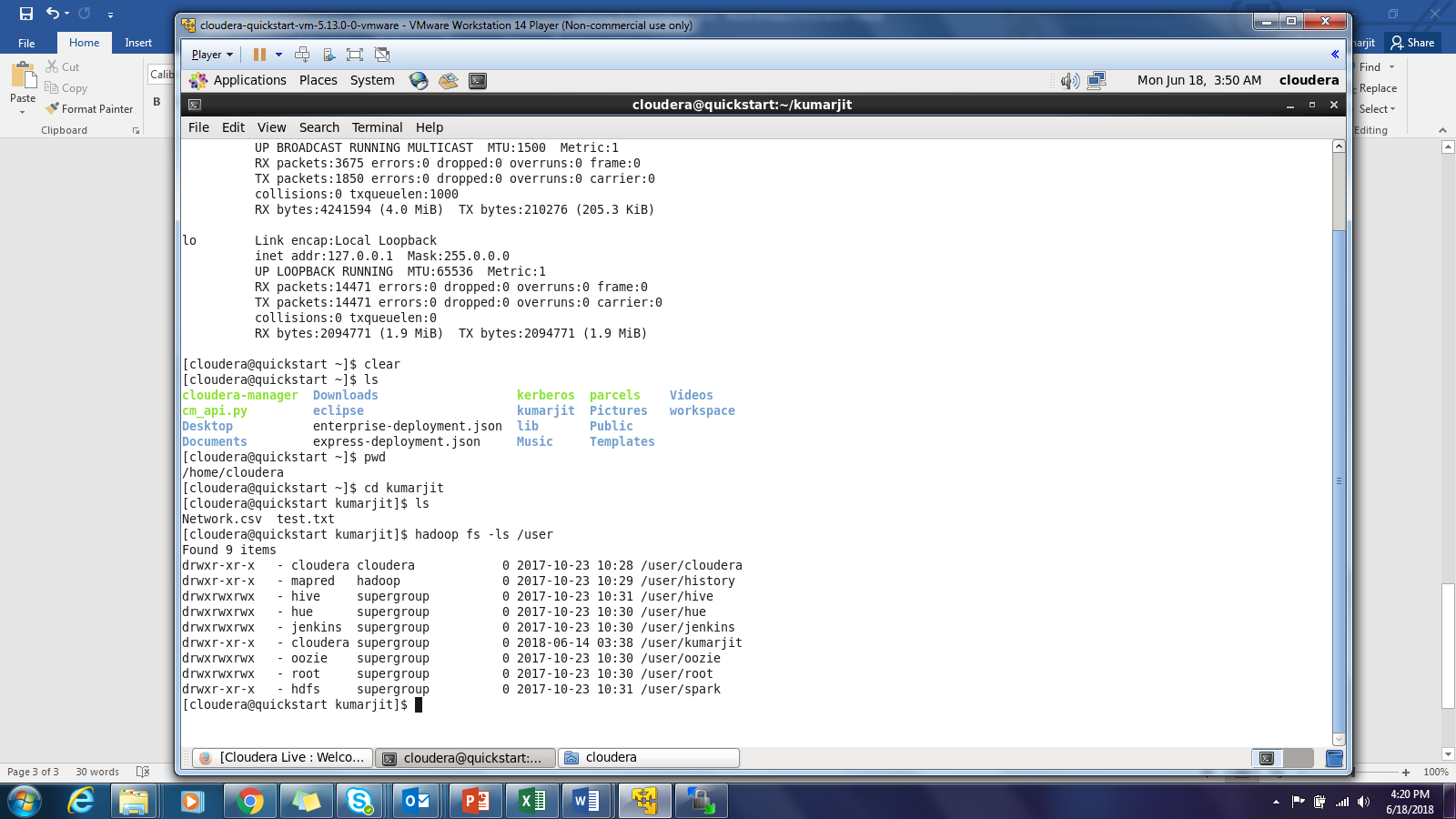


Put the address here and ID PASS both as **cloudera**

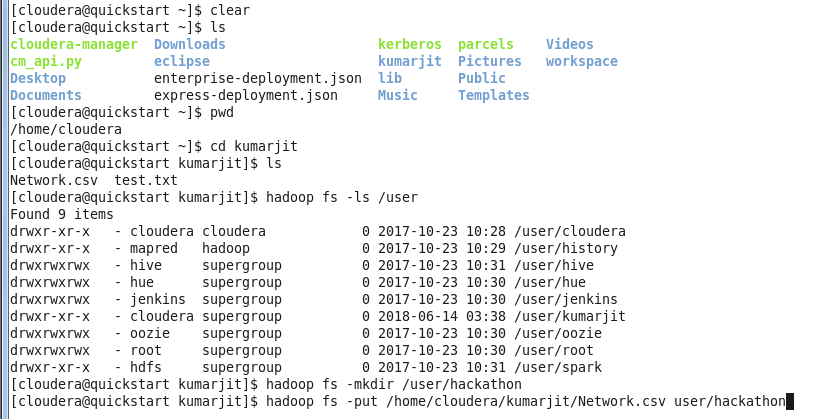


clear: to clear the screen

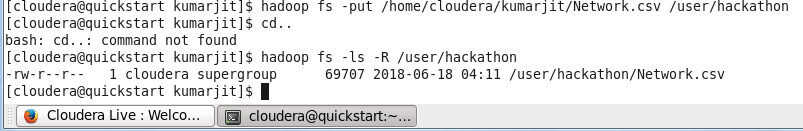


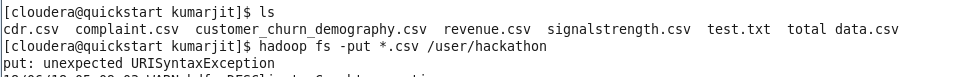


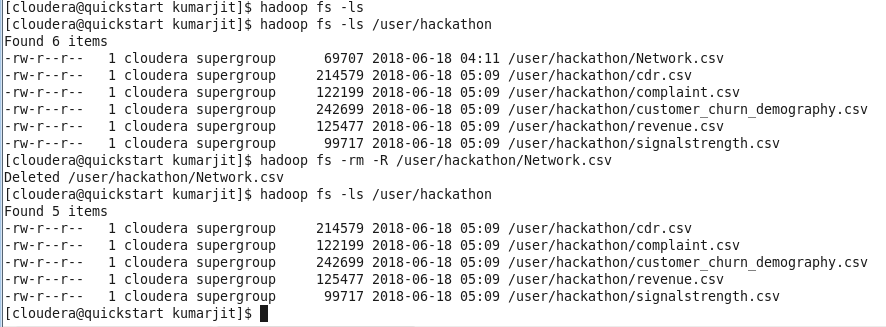
Hadoop fs – mkdir /user/hackathon

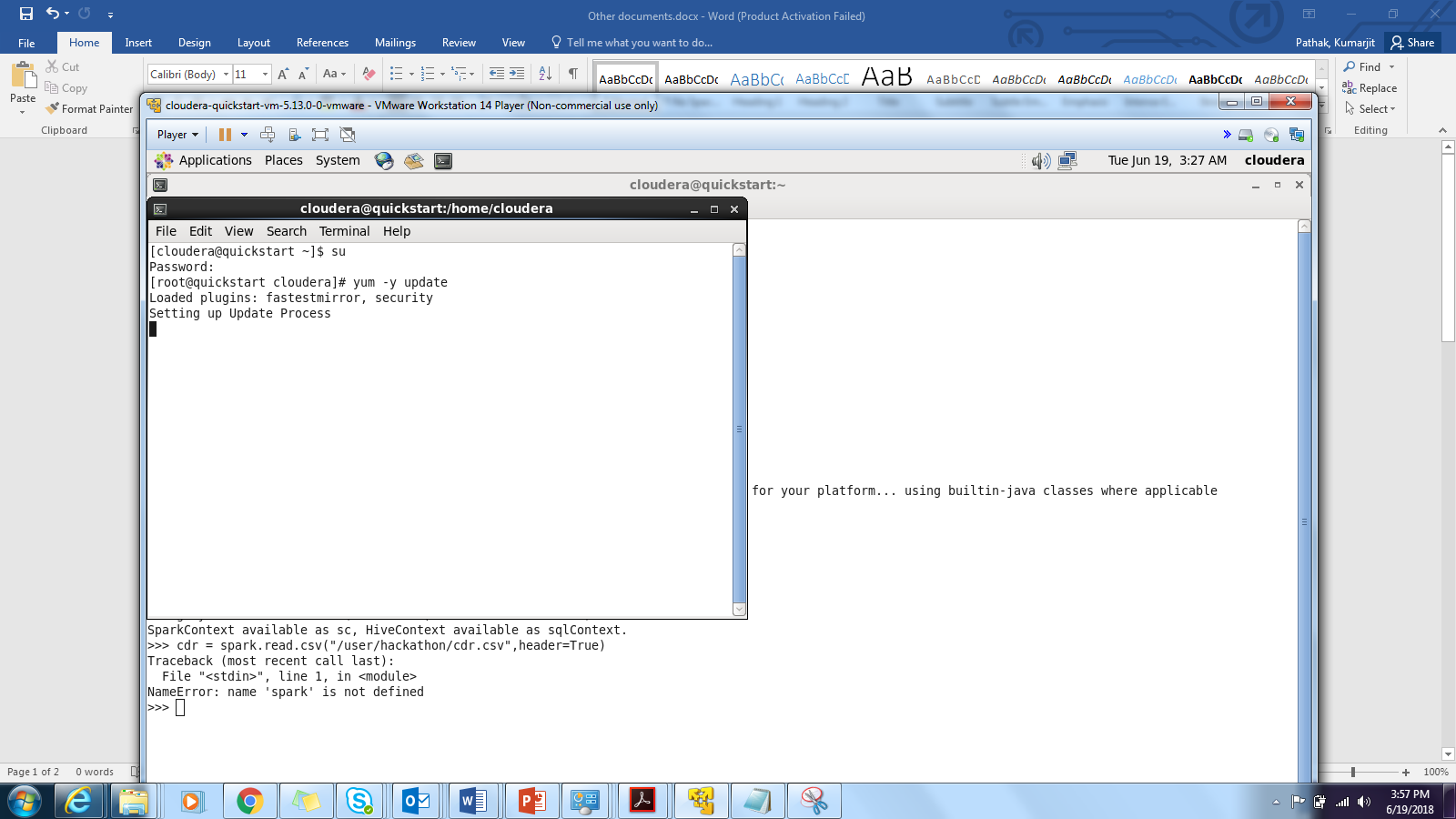


-R stands for recursive









[‎6/‎18/‎2018 3:48 PM] Agarwal, Rahul:

https://www.kaggle.com/c/telstra-recruiting-network

link for telstra problem on kaggle

We saved this conversation. You'll see it soon in the Conversations tab in Skype for Business and in the Conversation History folder in Outlook.

[‎6/‎19/‎2018 2:50 PM] Agarwal, Rahul:

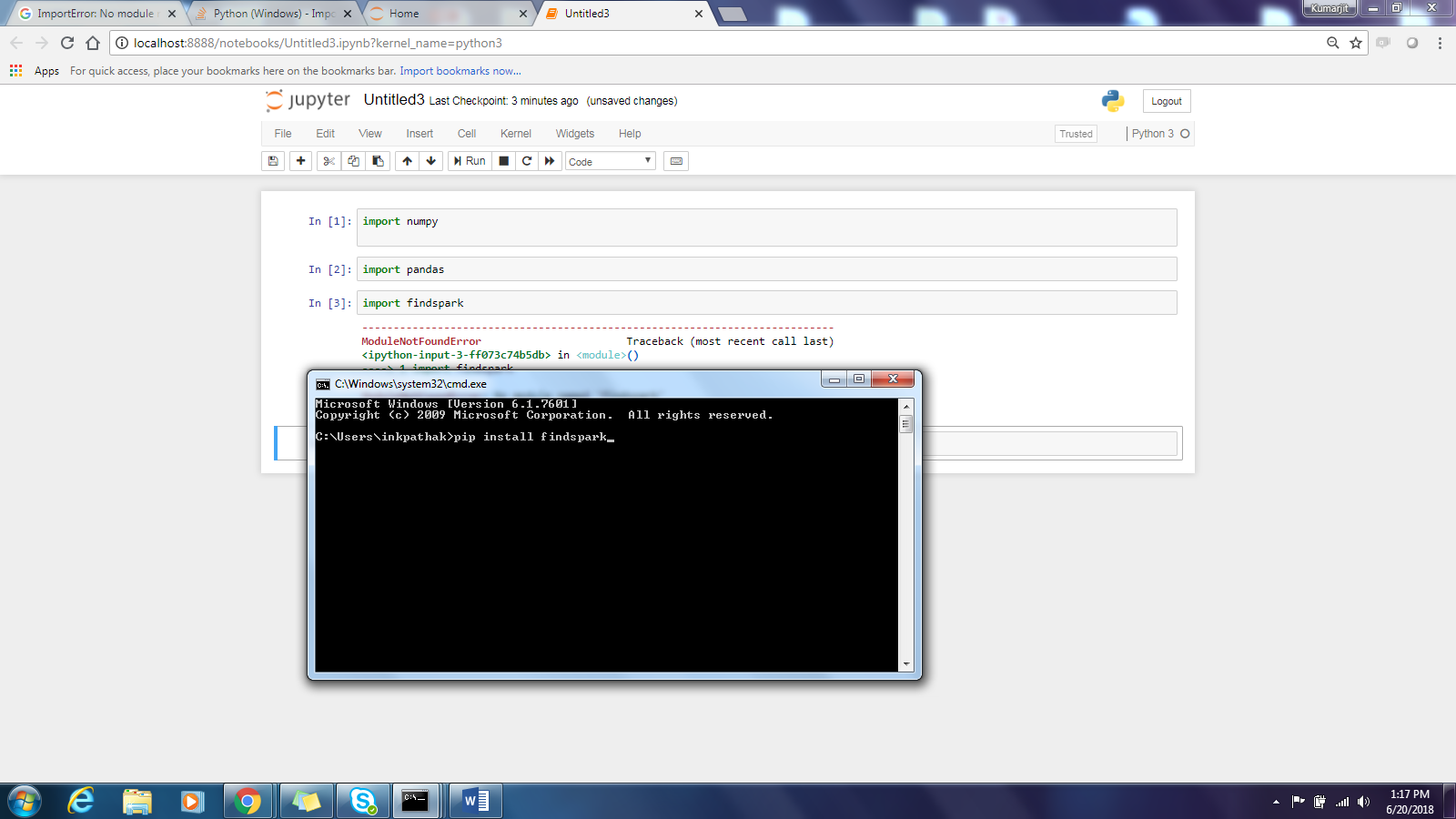
https://www.datacamp.com/community/tutorials/apache-spark-python

[‎6/‎19/‎2018 3:03 PM] Agarwal, Rahul:

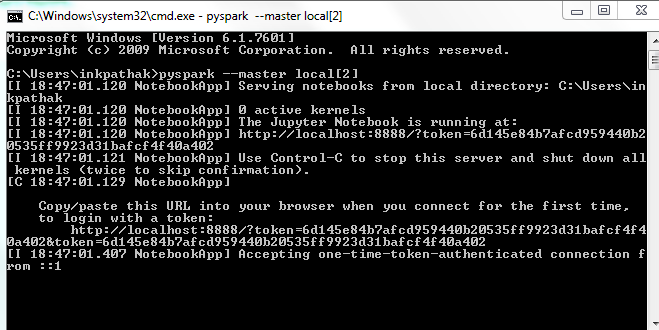
D:\Users\inragarwal20\AppData\Local\Continuum\anaconda3 (file:///D:/Users/inragarwal20/AppData/Local/Continuum/anaconda3)

import pyspark   
from pyspark import SQLContext   
sqlcon = SQLContext(sc)   
df = sqlcon.read.csv("D:\\Telco\\CustomerChurn.csv")

To connect SPARK



While running pyspark please give atleast 2 cores to pyspark



Goto Anaconda prompt and install a package:”spark\_df\_profiling”



Then you will see a warning msg:Need to install “msgpack”

