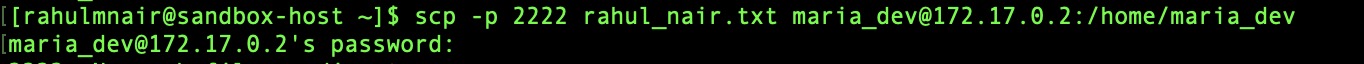
**Big Data Homework-2**

Task:-

1. Copy the file rahul\_nair.txt to the maria\_dev account.

First transfer the file to azure account. 

Then , transfer it to maria\_dev account.



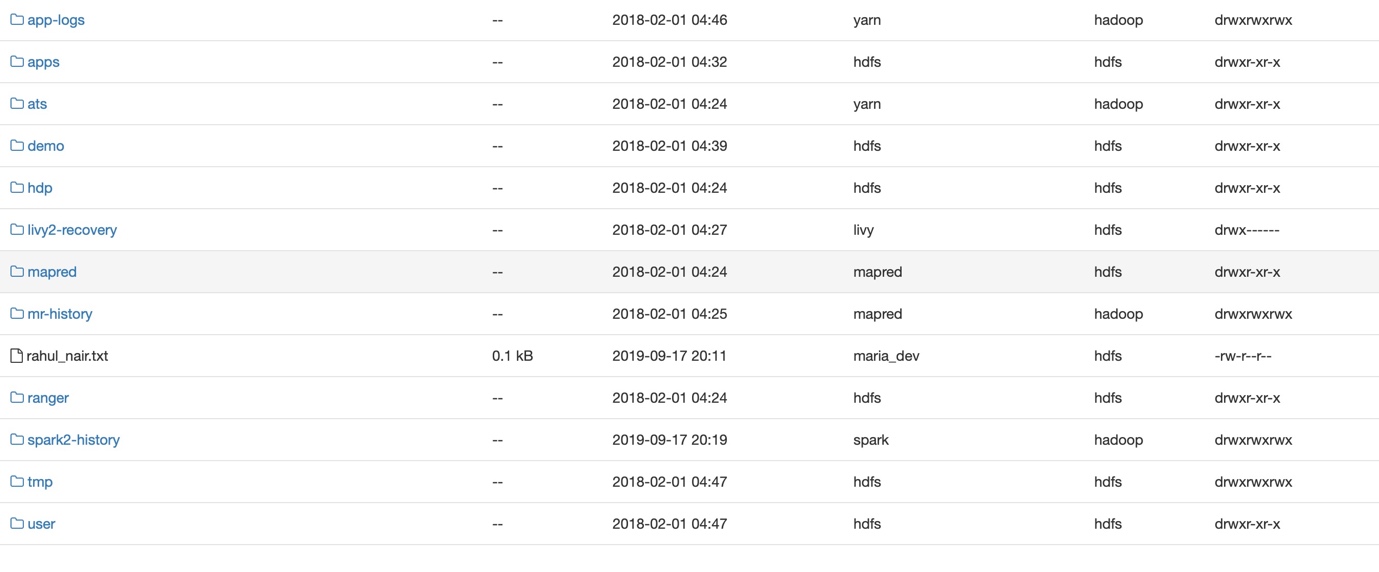
1. Execute appropriate command to view the file on azure computer.



1. Execute appropriate command to move the file to move the file on HDFS.



1. Open Ambari account and locate the file.



1. Research the appropriate command to view the file contants.





1. Make a copy of the file and rename it to rahul\_nair\_2.txt.



1. View both the files on Hadoop.







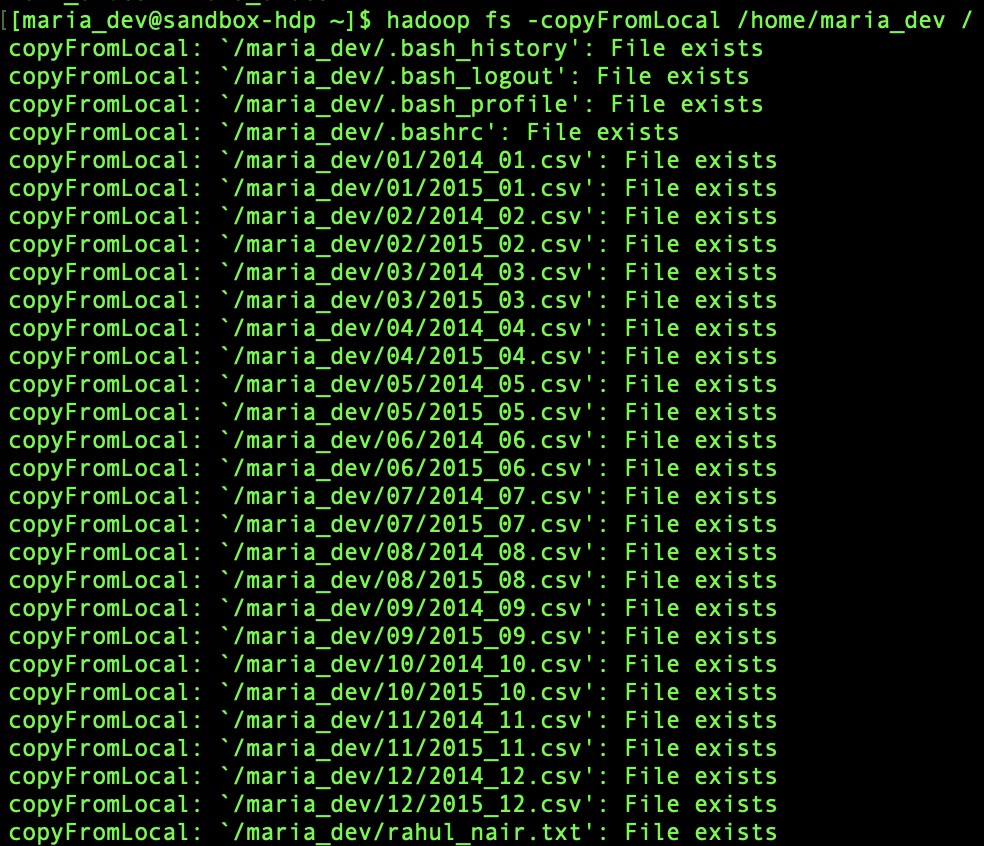
1. Using your favorite programming language, take the data set [climbing\_statistics.csv](https://drive.google.com/file/d/1xI08NL11uq9xLvpwPkkVAkJuiLubhocV/view?usp=sharing) and for each month in the data set (under the date column), create a directory (labeled by month number).

Using the same data set, split the data into one file per month year combo. Save the file labeled as “yyyy\_mm”.csv.

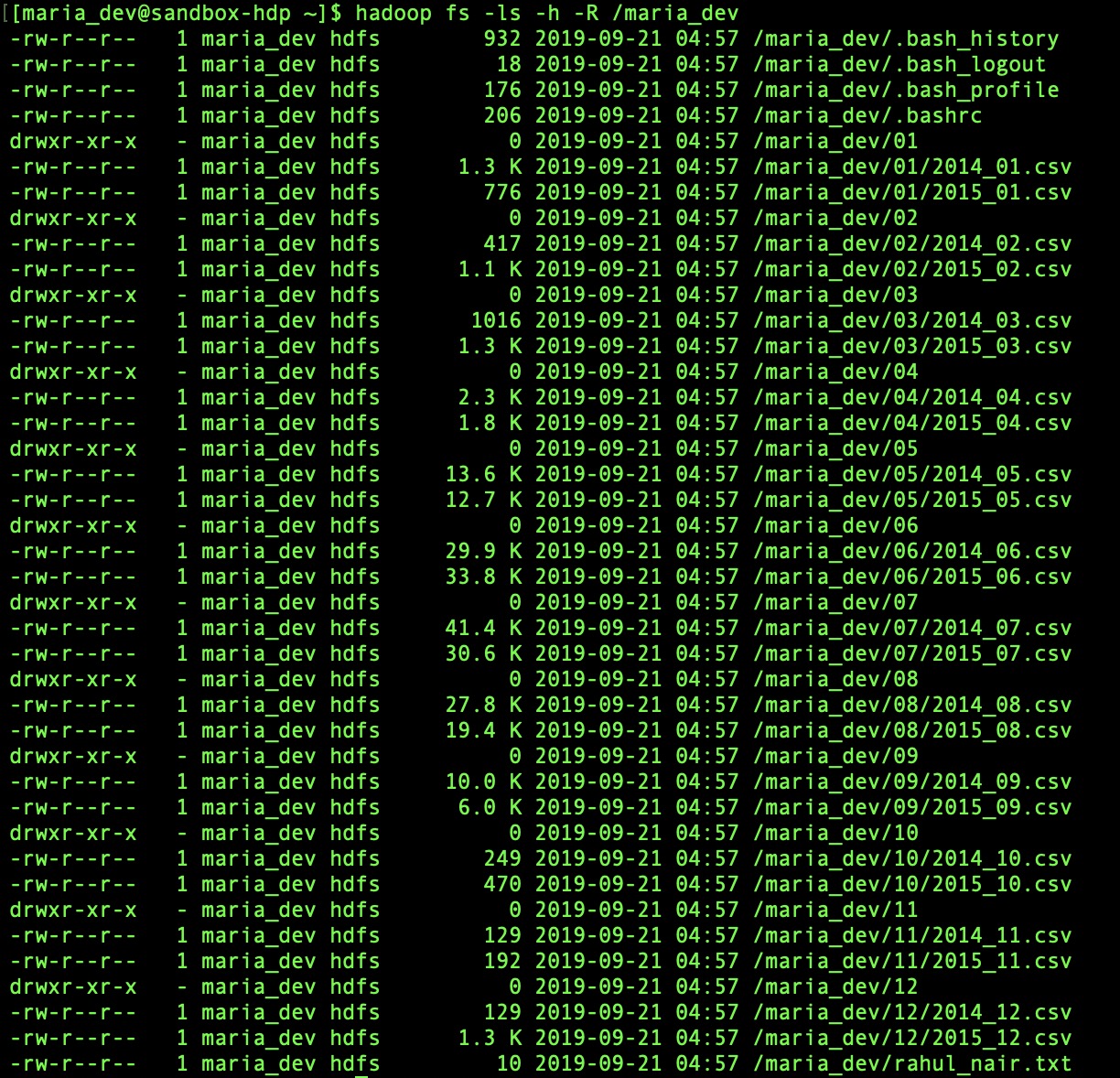
Using the files created in the previous step, place them into the appropriate directories for that month.

import pandas as pd  
import os  
  
  
df = pd.read\_csv('/Users/rahulnair/desktop/climbing\_statistics.csv')  
  
  
def getYearMonth(s):  
 return s.split("/")[0]+"/"+s.split("/")[2]  
  
  
os.mkdir('/Users/rahulnair/desktop/01')  
os.mkdir('/Users/rahulnair/desktop/02')  
os.mkdir('/Users/rahulnair/desktop/03')  
os.mkdir('/Users/rahulnair/desktop/04')  
os.mkdir('/Users/rahulnair/desktop/05')  
os.mkdir('/Users/rahulnair/desktop/06')  
os.mkdir('/Users/rahulnair/desktop/07')  
os.mkdir('/Users/rahulnair/desktop/08')  
os.mkdir('/Users/rahulnair/desktop/09')  
os.mkdir('/Users/rahulnair/desktop/10')  
os.mkdir('/Users/rahulnair/desktop/11')  
os.mkdir('/Users/rahulnair/desktop/12')  
  
df['YearMonth']= df['Date'].apply(lambda x: getYearMonth(x))  
data = pd.DataFrame(df)  
data\_114 = data[data.YearMonth=='1/2014']  
data\_214 = data[data.YearMonth=='2/2014']  
data\_314 = data[data.YearMonth=='3/2014']  
data\_414 = data[data.YearMonth=='4/2014']  
data\_514 = data[data.YearMonth=='5/2014']  
data\_614 = data[data.YearMonth=='6/2014']  
data\_714 = data[data.YearMonth=='7/2014']  
data\_814 = data[data.YearMonth=='8/2014']  
data\_914 = data[data.YearMonth=='9/2014']  
data\_1014 = data[data.YearMonth=='10/2014']  
data\_115 = data[data.YearMonth=='1/2015']  
data\_215 = data[data.YearMonth=='2/2015']  
data\_315 = data[data.YearMonth=='3/2015']  
data\_415 = data[data.YearMonth=='4/2015']  
data\_515 = data[data.YearMonth=='5/2015']  
data\_615 = data[data.YearMonth=='6/2015']  
data\_715 = data[data.YearMonth=='7/2015']  
data\_815 = data[data.YearMonth=='8/2015']  
data\_915 = data[data.YearMonth=='9/2015']  
data\_1015 = data[data.YearMonth=='10/2015']  
data\_1115 = data[data.YearMonth=='11/2015']  
data\_1215 = data[data.YearMonth=='12/2015']  
data\_1114 = data[data.YearMonth=='11/2014']  
data\_1214 = data[data.YearMonth=='12/2014']

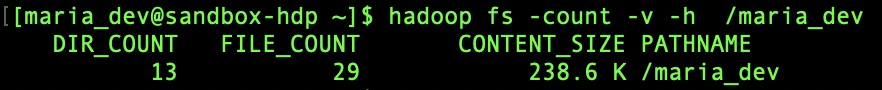
1. Research and execute the appropriate commands to migrate the files created in the previous question to HDFS in your Azure machine.



1. Research and execute the appropriate commands to print the full directory structure including files contents.



1. Research and execute the appropriate commands to count all of the files and their size in maria\_dev’s usr folder. Ensure file sizes are formatted in a human readable fashion.



1. Research and execute the appropriate commands to delete all of the files used in this assignment from your Azure machine

