**PRINCIPLES OF BIG DATA MANAGEMENT**

**PROJECT : PHASE 1**

**TOPIC : IPHONE7**

RAVILLA LAKSHMI SAI KRISHNA 16222098

SRAVAN VARMA 16221049

SAINATH REDDY SUDHIREDDY 16232714

SRI VIDYA JYOTHULA 16224836

**1) Program to collect tweets**

from tweepy import Stream

from tweepy import OAuthHandler

from tweepy.streaming import StreamListener

import time

import os

//Private

cKey="CN8nk2OCoUW8nUxTmh653z0kP"

cSecret="ChpPhWldC5MeWcfAkJ8bFzFbTkT4UEHsEA4T3WDhQkJD57zZ52"

aToken="4863039415-jIk8MOEtH5Z9m7avNPl9pTAfbPMxK9o7cSmeZJP"

aSecret="mqMV4orHuy4mq1nWA3z1CGcuRwxGublqzVfMmuXvbeyiv"

class Reciever(StreamListener):

def on\_data(Obj, data):

try:

writeFile= open('Data.json','a')

writeFile.write(data)

writeFile.write('\n')

writeFile.close()

return(True)

except Exception, e:

print 'failed on\_date,',str(e)

def on\_error(Obj, status):

pass

oAuth = OAuthHandler(cKey, cSecret)

oAuth.set\_access\_token(aToken, aSecret)

twitterData = Stream(oAuth, Reciever())

twitterData.filter(track=["iphone 7","iphone7","iphone-7", languages=['en'])

**2) Commands to map and Reduce**

val input=sc.textFile("iPhone7.json")

val words=input.flatMap(f=>f.split("\\W+"))

val kvpair=words.map(f=>(f,1))

val counts=kvpair.reduceByKey{case (x,y)=>x+y}

counts.foreach(f=>println(f))

counts.saveAsTextFile("output")

counts.saveAsTextFile("C:\Users\sravan\Documents\Results")

1. **Log File**

We uploaded Log file separately.

1. **Got word count File**

We uploaded Word count file separately.

1. **HDFS screenshots**





