**Hadoop\_Assignment\_20.2**

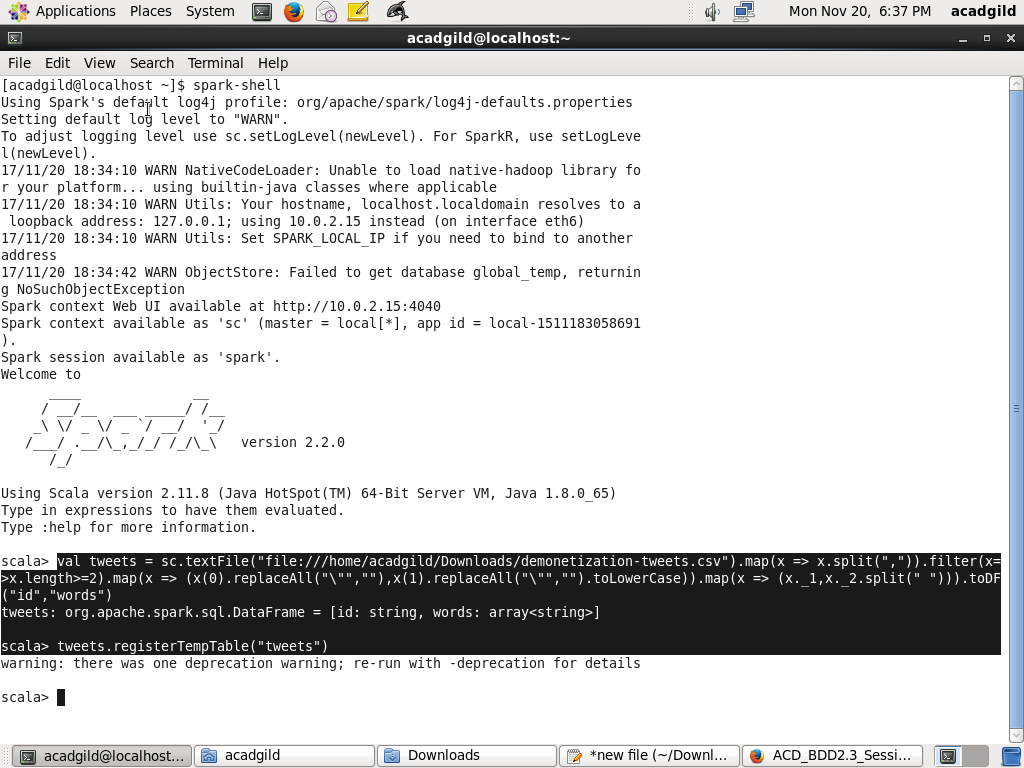
To find out the views of different people on the demonetization by analysing the tweets from twitter.

Dataset(Demonetization-tweets.csv) Contains twitter tweets gathered in CSV format.

1. we will load the data using **sc.textfile** as the Dataset is in CSV format split words by using ‘,’ Delimeter and convert it to DF by ID and Words and Register the Temptable using registerTempTable.

Sol:

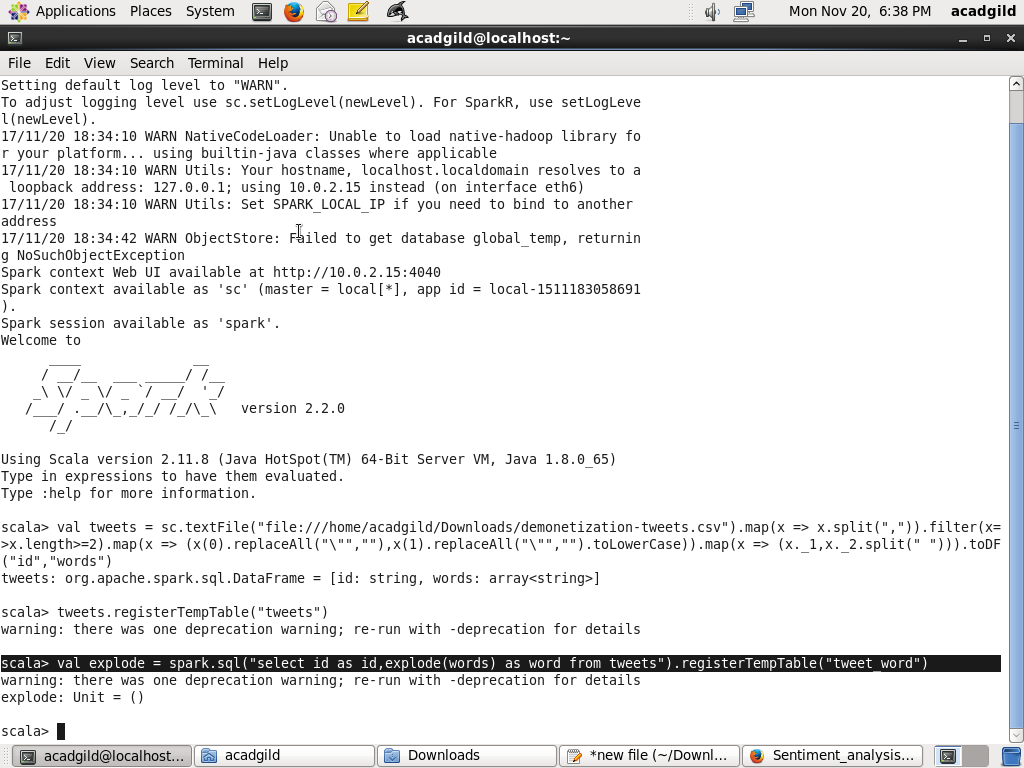
val tweets = sc.textFile("file:///home/acadgild/Downloads/demonetization-tweets.csv").map(x => x.split(",")).filter(x=>x.length>=2).map(x => (x(0).replaceAll("\"",""),x(1).replaceAll("\"","").toLowerCase)).map(x => (x.\_1,x.\_2.split(" "))).toDF("id","words")tweets.registerTempTable("tweets")



1. Connect to SparkSql and Perform Query Operation.

Sol:

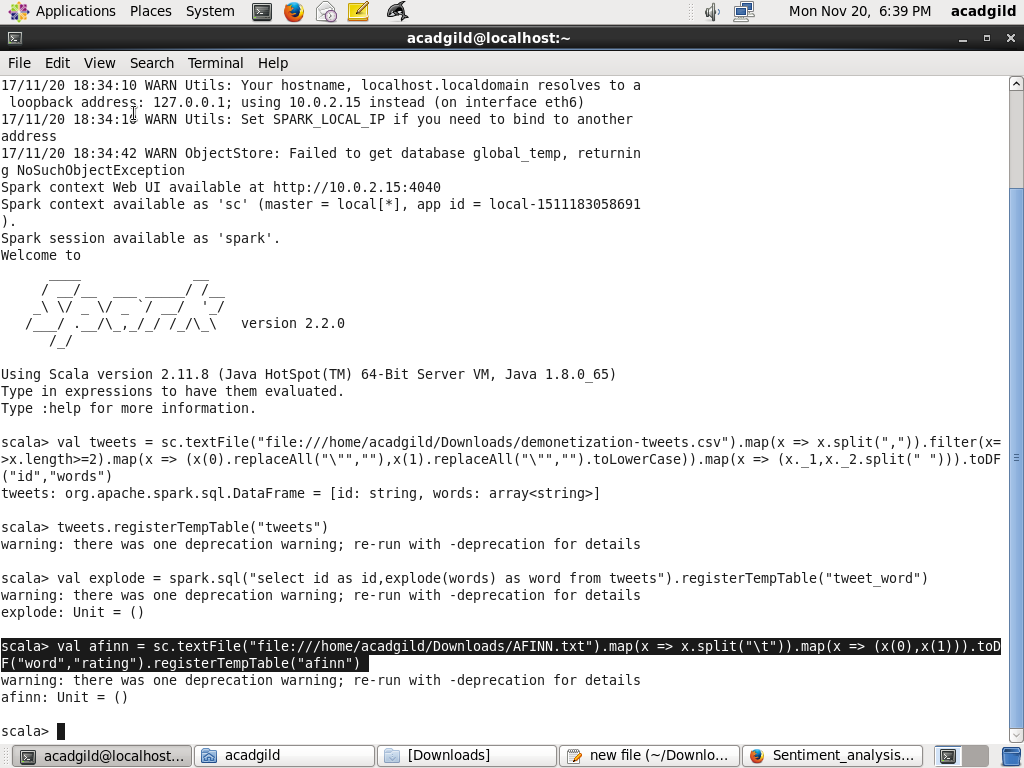
val explode = spark.sql("select id as id,explode(words) as word from tweets").registerTempTable("tweet\_word")



1. Now, we have to analyse the Sentiment for the tweet by using the words in the text. We will rate the word as per its meaning from +5 to -5 using the dictionary **AFINN**. The AFINN is a dictionary which consists of 2500 words which are rated from +5 to -5 depending on their meaning.

Sol:

val afinn = sc.textFile("file:///home/acadgild/Downloads/AFINN.txt").map(x => x.split("\t")).map(x => (x(0),x(1))).toDF("word","rating").registerTempTable("afinn")



1. Find out the Avg Rating and perform Join operation with results obtained from above query .

Sol:

val join = spark.sql("select t.id,AVG(a.rating) as rating from tweet\_word t join afinn a on t.word=a.word group by t.id order by rating desc").show

Output:

