**Question**: What day of the week does PrezOno tweet the most on average? (Sneha Jinturkar)

**Solution and method:**

I used Java to work on this problem. I only had to install eclipse in addition to the existing features in the OS.

1. I set up the environment and the VPN to start with, following the steps provided.
2. Next, in the root folder, I created a new folder called files, in which I created the folders ‘classes’ to later store the class files.
3. From the twitter dataset, I got a part on my local machine to analyze the structure of the data, using the following command

Hadoop fs -get /data/twitter/part-00192

1. The data is in JSON format and so, I used the **org.json library** in my code to parse through the data.
2. I created a .java file tweetTwo, in which I coded the logic to extract the necessary details. Below is the logic for the mapper class of the same file -

public static class TweetClass extends Mapper<Object, Text, Text, LongWritable>

{

public void map(Object key, Text value, Context context)

throws IOException, InterruptedException

{

String[] tweets = value.toString().split("\\n");

for( String tweet : tweets){

JSONObject obj=new JSONObject(tweet);

String scrname = obj.getJSONObject("user").getString("screen\_name");

String name = obj.getJSONObject("user").getString("name");

if((scrname.equalsIgnoreCase("PrezOno")))

{

String date =obj.getString("created\_at");

String subs = date.substring(0, 3);

context.write(new Text(subs), new LongWritable(1));

}

}

}

6. To execute this, I had to change the Hadoop\_classpath to add the .jar file of json.

export HADOOP\_CLASSPATH=$JAVA\_HOME/lib/tools.jar:/root/Downloads/json-20140107.jar

7. I also had to combine the jar files (json and tweetTwo) to run the program. Below are the commands I used to do it:

hadoop com.sun.tools.javac.Main TweetTwo.java -d classes

jar -cvf TweetTwo.jar -C classes .

mkdir tmp //To combine the two jar files

(cd tmp; jar -xf ../TweetTwo.jar)

(cd tmp; jar -xf ../json-20140107.jar)

jar -cvf combined.jar -C tmp .

And to run the program and store the output in a folder,

hadoop jar combined.jar TweetTwo /data/twitter OutputExpected

hadoop fs -get OutputExpected

**Output:**

Fri 39

Mon 48

Sat 53

Sun 59

Thu 54

Tue 33

Wed 55

**Conclusion:** From my analysis, I conclude that PrezOno tweeted the most on Sundays, even though the result was pretty close to the other days.

Find on Github,

1. Output files
2. .java file
3. .class files related to the above solution.