**CLOUD COMPUTING (P3) – *Hadoop for Real Problems***

***Twitter data question 3:***

How does @PrezOno’s tweet length compare to the average of all others? What is his average length? All others?

***Required data to answer the questions:***

1. User data in twitter data set
   1. Only screen name is required to answer the question.
2. Twitter text – Only length of the text.

According to the question, we get to know that we have to find tweet length of each person as well as President Ono’s tweet length. At the end we have to calculate the average tweet length of President Ono and a cumulative tweet length of all others. This can be done using map reduce by following the below approach.

***Approach:***

Using Hadoop streaming method we can solve this question. Find below the steps followed in solving the question:

* Initially, we create a mapper which parses each input line using json library and extracts the Screen Name from user data and Twitter text.
* In the mapper we iterate through each user along with his tweet from the beginning and when the mapper encounters the screen name of President Ono, it calculates the text length as well as it increases the counter of each tweet.
* When the mapper encounters individuals other that President Ono, it calculates the text length as well as it increases the counter of each tweet.
* Mapper sends the output to the reducer and reducer considers them as a key value pair for President Ono’s tweets and other’s tweets.
* Reducer sums up the all the values corresponding to the specific key and performs the union operation on them. At end we get total tweet length text of president Ono along with the number of tweets and also we get the total tweet length of all others (cumulative data) along with the number of tweets.
* Used only one reducer and a combiner to complete the job faster.
* We can calculate the average twitter text length of president Ono when compared with all others (cumulative data of others)

***Results:***

ocount 6079961

osum 496716166

pcount 341

psum 35578

***Final result:***

Average twitter text length of President Ono – psum/pcount = 35578/341 = **104.33**

Average twitter text length of all others (cumulative data) – osum/ocount = 496716166/6079961 = **81.697**

*When compared to other’s president Ono’s tweet text length is* ***more than*** *all other’s tweet length*