**Twitter Question 7:**

**Required data to answer the questions:**

1. Average of Left Co-ordinate of geo component
2. Average of Right Co-ordinate of geo component
3. Number of tweets having geo data
4. Number of tweets having no geo data

**Approach:**

We used Hadoop streaming method to solve these questions. The steps followed in mapper are as below:

1. Mapper parses each input line using json library and extracts the geo data if present.
2. If geo data is present, then ‘yes’ value is incremented and left and right coordinates are added to the present coordinates.
3. If geo data is not present then ‘no’ value is incremented.
4. After all the lines in input to a mapper are exhausted we print all the variables as key and value pair.
5. Reducer add all the values corresponding to a key.

**Results:**

left 72845727.274578

no 4212000.000000

right -154046999.285180

yes 1868302.000000

**Centroid = ((left/yes),(right/yes))**

****

**Proportion of Tweets:**

Proportion = yes/no = 1868302.000000/4212000.000000 = 0.4435