**Report**

**Configuration Experience**

In this project I made use of one Mapper to load the Input table and one MapReduce to process the KMeans algorithm. At first it was difficult to do the setup as it took me around 2-3 days to complete the whole setup. I had to download the hadoop and Hbase 0.98 again and do the whole setup again from the start. I found the tutorials from tutorialpoints the most useful for the installation of Hbase and hadoop and followed each and every steps carefully. I was struggling with errors in configuration in the first few days and also few days after the configuration was setup.

**Implementation**

I first analyzed the dataset and just created the “Input” table in the main function according to the Schema required. Then I passed my dataset as input to the first mapper. I imported the dataset from the text file into the “Input” table by defining a particular row sequence and adding the data. In this each mapper inserts the data in each row of “Input” table while there is no output for this Mapper, its job is just to insert the data into the “Input” table. Then I created the “Center” table and Imported the First k(number of clusters) values into the Center table. My schema for the “Center” Table has only one column family named centroids and rows from X1-X10. But it does not effect the code as I access both the Input table and Center table as per the requirement. I gave input as the “Input” table to the 2nd Mapper and accessed the “Center” table from the HBase in the setup of the second Mapper. Then in the Map function I calculated the distance between the Center Table values and the Input table values and gave the Center Table as the key and calculated values as the value by converting them into array of Strings at the output of Mapper. The Reducer takes this as the input Calculates the average and updates it into the Center Table at the output. I kept a counter which keeps on converging after every iteration. The iteration is terminated depending on this counter as well when the minimum difference in distance is obtained and updated in the Center Table.

**Running the program**

In my eclipse also after the configuration it was showing the errors related to zookeeper but it was due to some start up issues. So I had to restart the hadoop and hbase again and the errors were solved after that. The same problem caused me after making the jar file and running it but it was solved after several attempts. Now my program runs on both Eclipse as well as Terminal.