Project 1 - Apache Spark—Real Time Project—Marketing Analysis

# Pre-requisites:

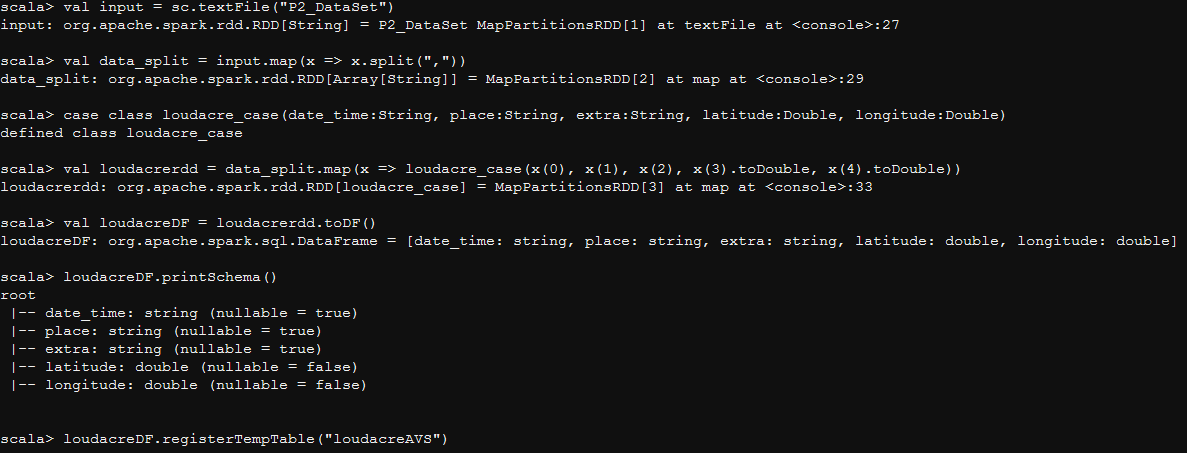
The data set file was uploaded to cloudlab using FTP service.

Then they were uploaded to Hadoop FS using the command:

***hadoop fs –put P2\_DataSet .***

The spark shell is then launched and the data processing starts.

# Load data and create Spark data frame



# K-Means Working

# 

**Solution:**

**This is the required information of latitude and longitude and the three clusters of users found with k-means algorithm is as below:**

**[34.52886579878435,-116.34531611913462]**

**[0.0,0.0]**

**[39.57392651941122,-121.24864484001667]**