Project for submission –

k-means-Real Time Project-LoudAcre Mobile

**Write a k-means algorithm –**

LoudAcre Mobile is a mobile phone service provider which has introduced a new Open Network campaign. As part of this campaign, the company has invited users to raise a request to initiate a complaint about the towers in their locality, if they face issues with their mobile network. LoudAcre has collected the dataset of users who had raised the complaint. The fourth and the fifth field of dataset has latitude and longitude of users which is an important information for the company. You have to find this information of latitude and longitude on the basis of available dataset and create three clusters of users with k-means algorithm. This will help Loudacre maximize the coverage for its users.

# Pre-requisites:

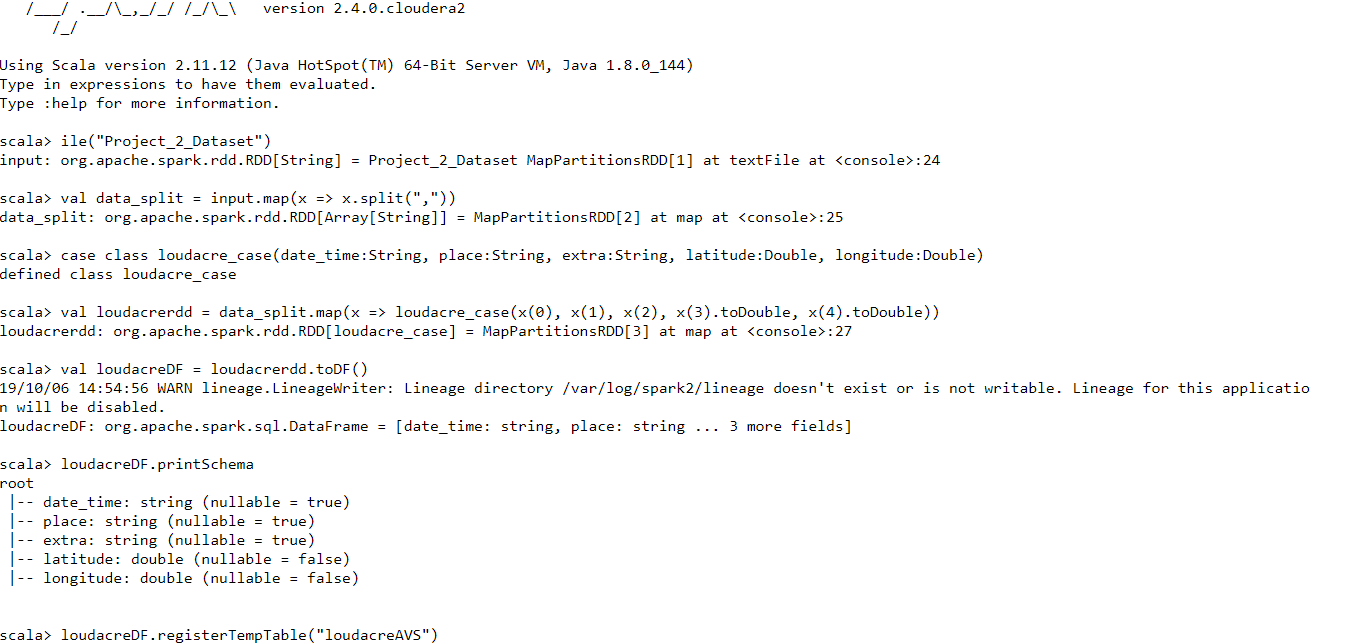
The data set file was uploaded to cloudlab using FTP service.

Then they were uploaded to Hadoop FS using the command:

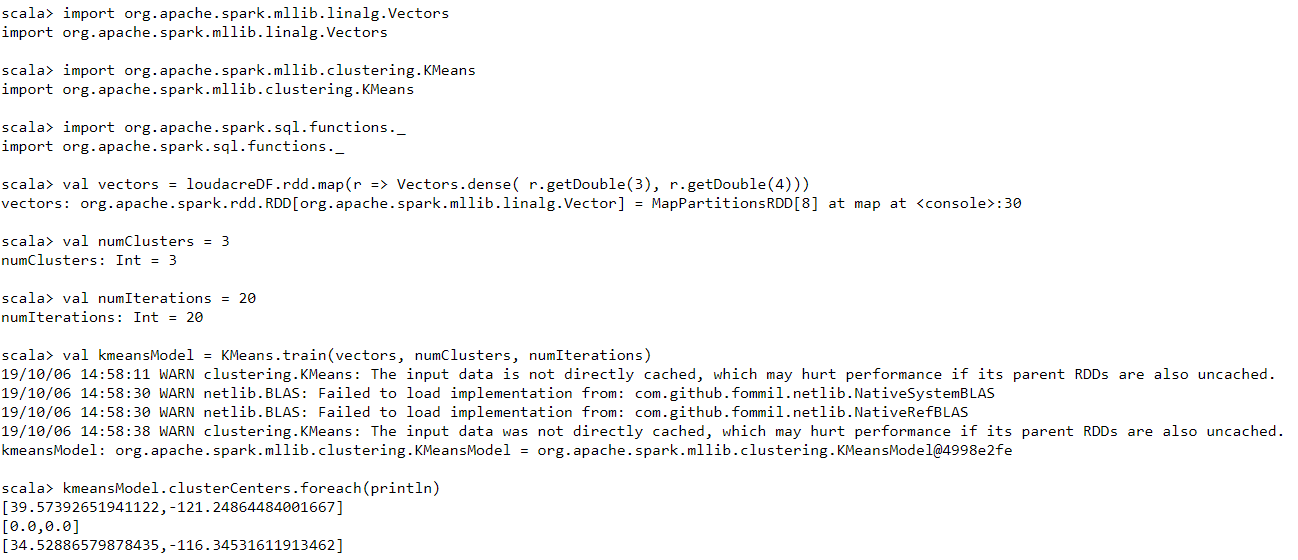
***hadoop fs –put Project\_2\_Dataset***

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

# Load data and create Spark data frame -



# K-Means -



**Solution:**

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

**[39.57392651941122,-121.24864484001667]**

**[0.0,0.0]**

**[34.52886579878435,-116.34531611913462]**