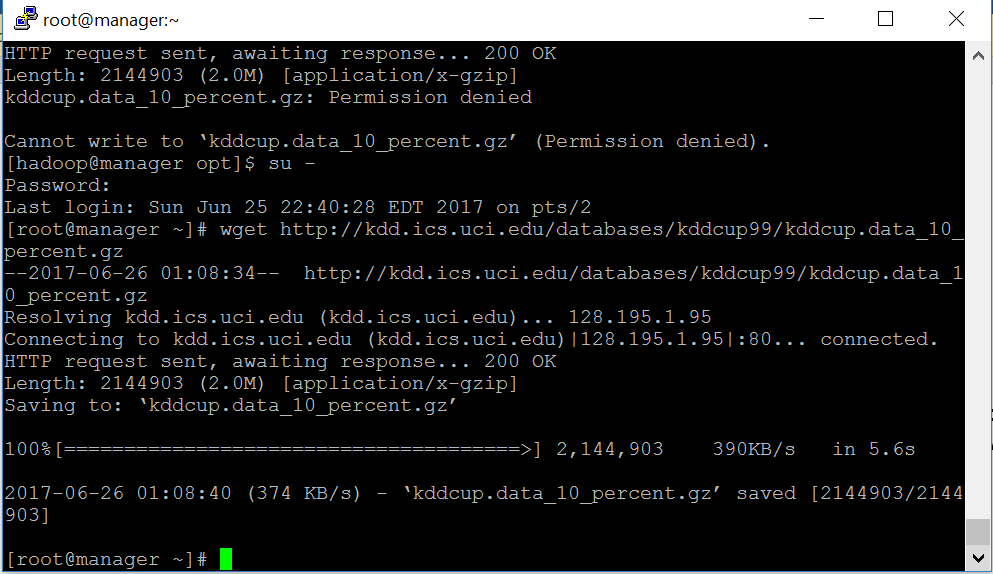
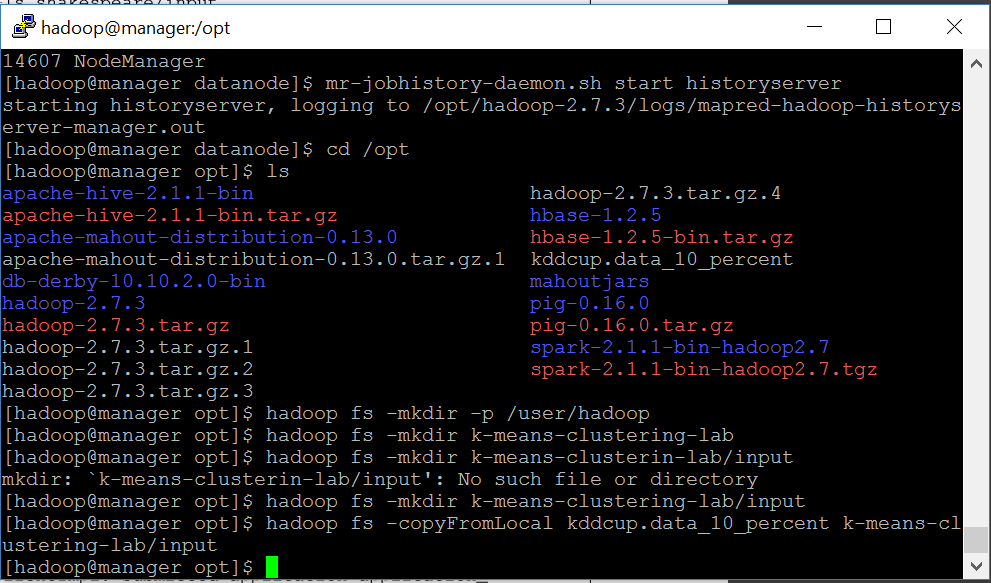
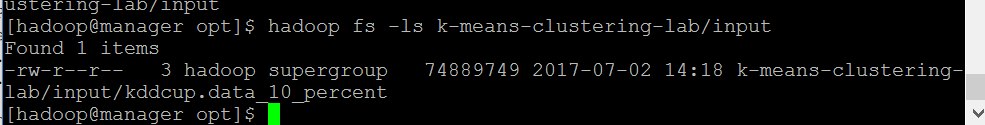
**Apache Spark Lab: K-means clustering**

Succesfully downloaded 10 percent data

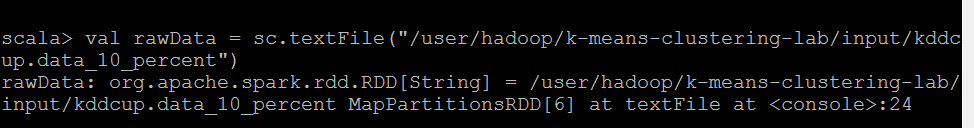


Copied 10 percent file into HDFS

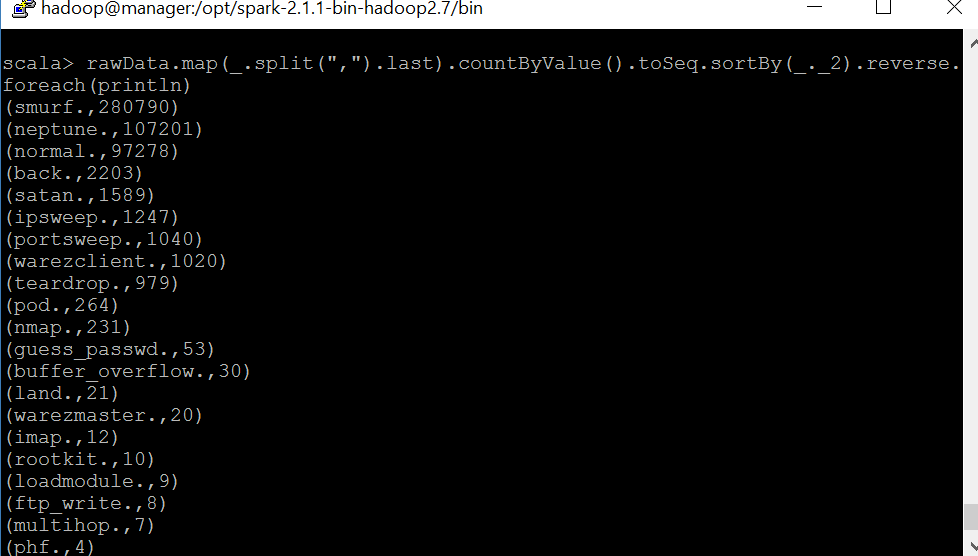




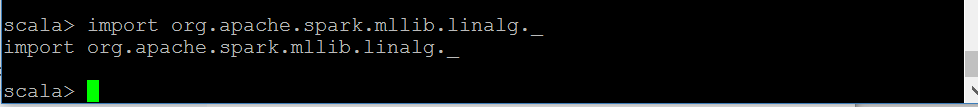
Spark commands loading textfile as source



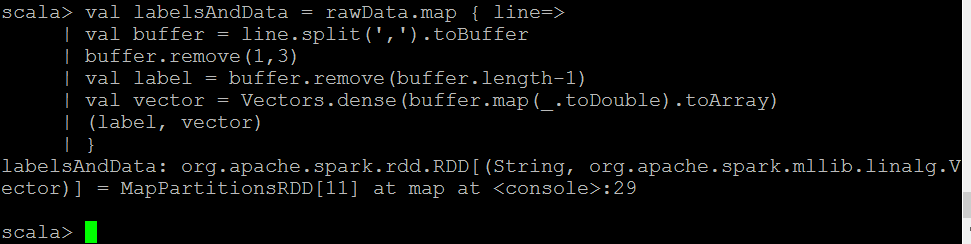
Data mapping manipulation on text file

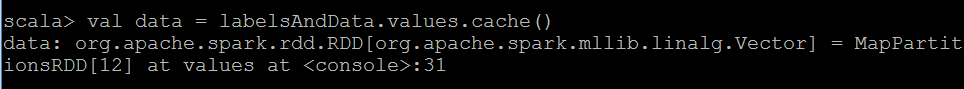


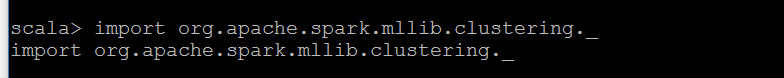
Imported library of functions

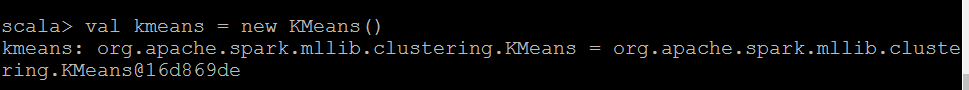


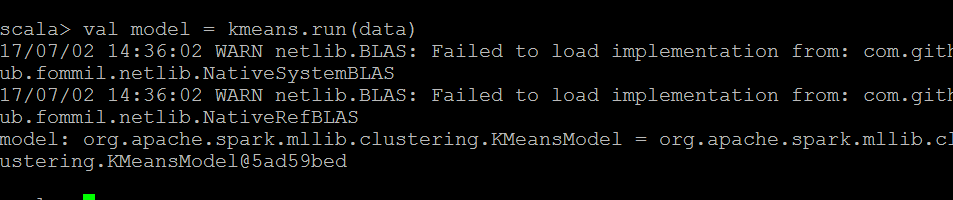
Scala functions for finding centroids and cluster anomalies

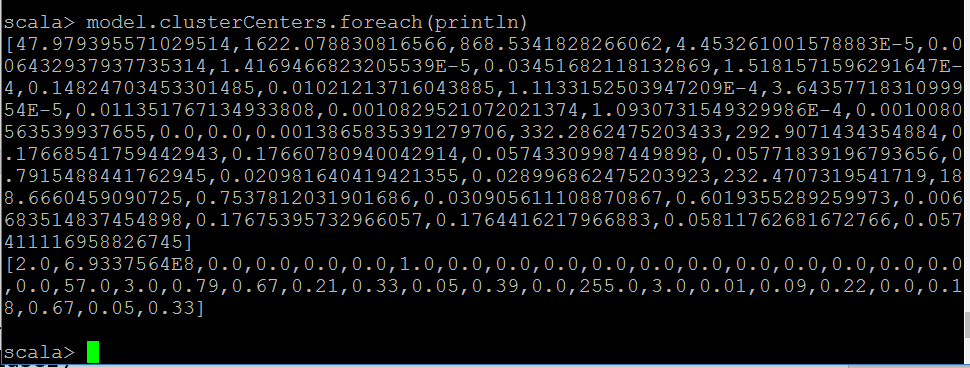


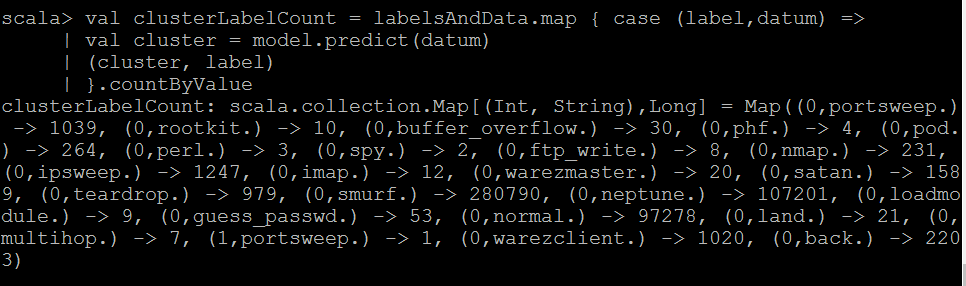


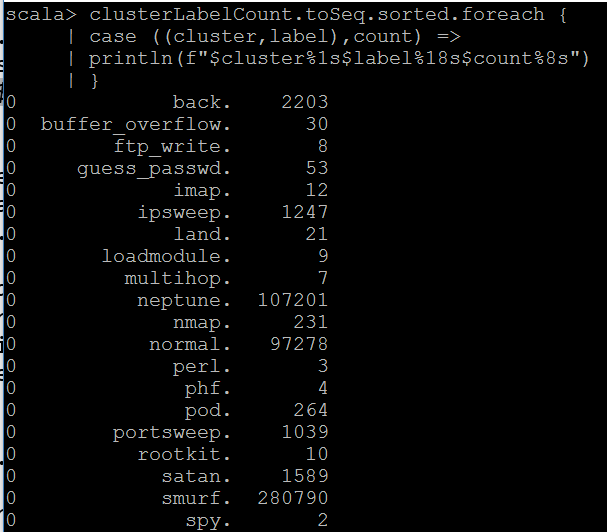


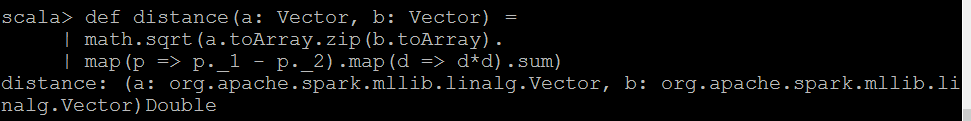


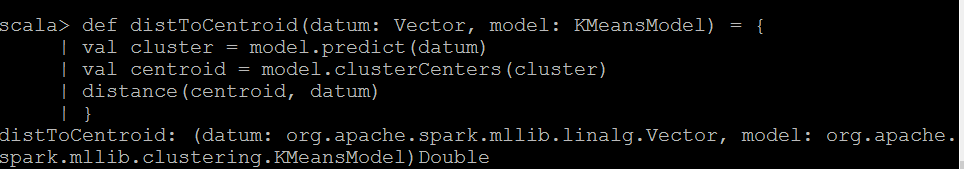


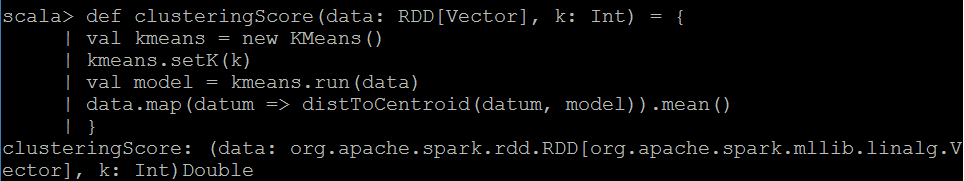


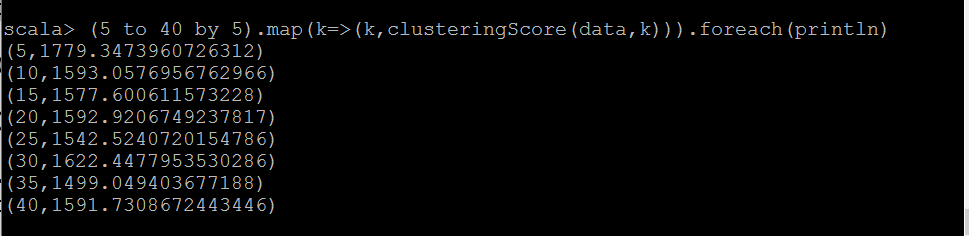


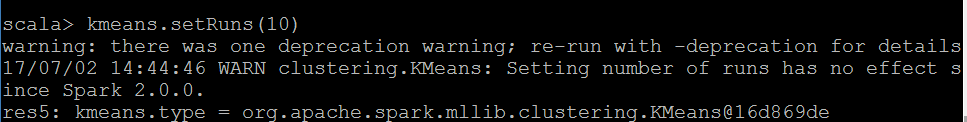


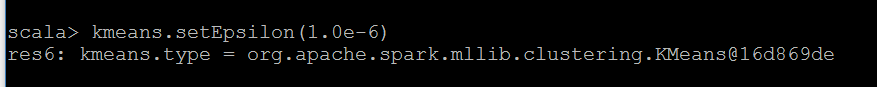


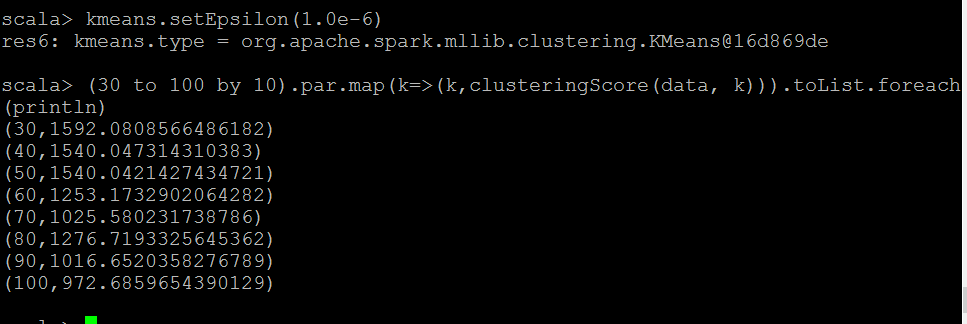












Thanks Professor Busch for all the hands on help this semester!

I learned a lot