

CS5229 - Big Data Analytics Technologies
AWS Mahout Assignment
 K.A.T.S. Jayathilaka - 209338R

GitHub: [tsj1992/CS5229-BDA-AWS-Mahout-Project](https://github.com/tsj1992/CS5229-BDA-AWS-Mahout-Project): Author: 209338R KATS Jayathilaka - This repository is to store and update development of an AWS Mahout Project under the module of Big Data Analytics in MSc in Computer Science, University of Moratuwa (github.com)

Summary of the work done.

1. According to the tutorial, I created a Hadoop cluster and then connected SSH into it using putty with my keypair.
2. There were some changes that need to be adopted to the commands when doing the assignment which is given below. All the commands that I have entered are in file **commands.txt** and **commands.sh** in the repo.
 - a. I used **hdfs dfs** instead of **hadoop fs**
3. Since the cluster's python version is python3, **easy_install** was not successful, so I used **pip3 install** command for installing **twisted**, **klein** and **redis** libraries.
4. The following changes were also done to the web service written in python2 version due to the version conflict. File is in repo as **hello8081.py**.
 - a. Line 9: Changed due to python version conflict
 - i. From: `p = os.popen("hdfs dfs -cat recommendations/part*")`
 - ii. To: `p = os.popen('hdfs dfs -cat recommendations/part*', 'r')`
 - b. Line 18: Changed due to incorrect splitting of key value pair
 - i. From: `k,v = i.split('t')`
 - ii. To: `k,v = i.split('\t')`
 - c. Line 28-29: Changed due to python version conflict
 - i. From:
 1. `v = r.get(id)`
 2. `return 'The recommendations for user '+id+' are '+v`
 - ii. To:
 1. `v = r.get(str(id))`
 2. `return 'The recommendations for user '+str(id)+' are '+str(v)`
 - d. Line 40: Changed due to 8080 port already in use
 - i. From: `run("localhost", 8081)`
 - ii. To: `run("localhost", 8081)`
5. The demo video of the implementation is also uploaded to the GitHub repo as **209338R_Mahout_demo.gif**.