

Solution For Saavn Song Recommendation Clustering Problem

1. Spark Command To Execute The Program On Spark 2.3.0 Cluster (On EC2):

```
spark2-submit -class
com.upgrad.bigdataanalytics.clustering.SaavnSongRecommendation --master yarn --
deploy-mode client --name "SaavnSongRecommendation" --num-executors 6 --driver-
memory 5g --executor-memory 10g /home/ec2-
user/Song_Recommendation_Using_Clustering-0.0.1-SNAPSHOT-jar-with-
dependencies.jar s3a://<Access key ID>:<Secret access
key>@bigdataanalyticsupgrad/activity/sample100mb.csv
s3a://<Access key ID>:<Secret access key>@bigdataanalyticsupgrad/newmetadata/*
s3a://<Access key ID>:<Secret access
key>@bigdataanalyticsupgrad/notification_actor/notification.csv s3a://<Access key
ID>:<Secret access key>@bigdataanalyticsupgrad/notification_clicks/*
hdfs:///user/ec2-user/Song_Recommendation_Using_Clustering/output/user_clusters/
hdfs:///user/ec2-user/Song_Recommendation_Using_Clustering/output/ctr_data/
```

2. Note:

- a) In the above command, please replace the following text:
 <Access key ID> with the AWS access key ID
 <Secret access key> with AWS Secret access key
- b) This program is tested on an Eclipse installed on Ubuntu linux. This program has Not Been Tested On A Windows OS.
- c) The program is tested and executed on the Spark 2.3.0 Cluster setup on an AWS EC2 instance.
- d) The program is tested, executed and the output is generated using the 100MB dataset.
- e) The program is developed using the ALS model approach.
- f) Instead of hardcoding the access key id and the secret access key in the program, I have chosen to pass both of them as an argument while specifying the path to data file(s) present on an S3 bucket. Please make sure to pass the access key id and secret access key as argument to the file location on S3 for the program to run. The command used for execution can be considered.
- g) Below are the HDFS commands used for creating output folders on HDFS:
 sudo su hdfs

 hadoop fs -mkdir -p /user/ec2-
 user/Song_Recommendation_Using_Clustering/output/user_clusters/

```
hadoop fs -mkdir -p /user/ec2-  
user/Song_Recommendation_Using_Clustering/output/ctr_data/
```

```
hadoop fs -chown -R ec2-user:ec2-user /user/ec2-  
user/Song_Recommendation_Using_Clustering/output/user_clusters/
```

```
hadoop fs -chown -R ec2-user:ec2-user /user/ec2-  
user/Song_Recommendation_Using_Clustering/output/ctr_data/
```

```
hadoop fs -chmod -R 777 /user/ec2-  
user/Song_Recommendation_Using_Clustering/output/user_clusters/
```

```
hadoop fs -chmod -R 777 /user/ec2-  
user/Song_Recommendation_Using_Clustering/output/ctr_data/
```