Steps to compile and run the project

The project is basically of two phases.

Phase 1- general recommendations using scala

Phase 2-specific recommendations using pig

Initial steps:

* Login to your remote server and upload the following data files

Users.csv, songs.csv, my.csv

* Also upload the two shell scripts, scala script and the pig script.
* Finally use the following command to see if all the 7 files are present on

Your remote server.

Hdfs dfs -ls

Phase 1-execution:

* Run the following shell script command to execute the phase 1 of the project and

Towards the end, you will a list recommended songs for the user on the spark shell

sh shell\_scala.sh

* Once you see that, exit from the spark-shell. I did not add a command to exit from spark

Shell to make sure that you can see the output on the screen.

Phase 2-execution:

* Execute the following command to see specific recommendations

sh shell\_pig.sh

* The specific recommendations will be displayed on the screen and also stored in the folders. If you want to check, u can check it using the cat command.

Contents of the submission:

* Data files - users.csv, songs.csv, my.csv
* Shell scripts – shell\_scala.sh, shell\_pig.sh
* Scala script file- recom2.scala
* Pig script file- recommendations.pig
* Output screenshots folder
* Analysis plots folder

Note:

* If you want to view the source code in the scala script or the pig script, use an editor to open the file and then view it
* Please be patient while running the pig part of the project as it takes around 10 mins.