

ASSIGNMENT ON DATAFRAMES DATASET

Consider the (complete) LastFM dataset discussed in the RDD and DF labs (see RDDLab and DFLab on Aulaweb).

ASSIGNMENT

Write RDD and Dataframe pyspark scripts for the following queries:

Q1: Select the list of users who listened to more than 100 songs in 2016

Q2: Select the top 5 songs among the most listened songs of friends of user

Q3: Select all songs that have been listened by user and one of his/her friends.