

Directory:

Surbhi_Batra_hw1

- OutputFiles
 - Surbhi_Batra_task1_big.csv
 - Surbhi_Batra_task1_small.csv
 - Surbhi_Batra_task2_big.csv
 - Surbhi_Batra_task2_small.csv
 - Jar_output
 - Surbhi_Batra_task1_big.csv
 - Surbhi_Batra_task1_small.csv
 - Surbhi_Batra_task2_big.csv
 - Surbhi_Batra_task2_small.csv
- Solution
 - surbhi_batra_hw1.jar
 - Surbhi_Batra_task1.py
 - Surbhi_Batra_task2.py
 - Surbhi_Batra_task1.scala
 - Surbhi_Batra_task2.scala
- Surbhi_Batra
 - Surbhi_Batra_Description

=====

Command to execute: cd/to/Surbhi_Batra_hw1

=====

Python:

Task1:

ml-latest-small :

\$SPARK_HOME/bin/spark-submit Solution/Surbhi_Batra_task1.py path/to/data/ratings.csv
path/to/outputfile

ml-20m :

\$SPARK_HOME/bin/spark-submit Solution/Surbhi_Batra_task1.py path/to/data/ratings.csv
path/to/outputfile

Task2:

ml-latest-small : \$SPARK_HOME/bin/spark-submit Solution/Surbhi_Batra_task2.py
path/to/ml-latest-small/ratings.csv path/to/ml-latest-small/tags.csv path/to/outputfile.csv

ml-20m : \$SPARK_HOME/bin/spark-submit Solution/Surbhi_Batra_task2.py
path/to/ml-20m/ratings.csv path/to/ml-20m/tags.csv path/to/Outputfile.csv

=====

Scala:

Rename the file Surbhi_Batra_task1.scala as Task1.scala and Surbhi_Batra_task2.scala as Task2.scala before running.

Task1:

ml-latest-small : spark-submit --class Task1 Solution/surbhi_batra_hw1.jar
path/to/ml-latest-small/ratings.csv path/to/outputfile.csv

ml-20m : spark-submit --class Task1 Solution/surbhi_batra_hw1.jar path/to/ml-20m/ratings.csv
path/to/outputfile

Task2:

ml-latest-small : spark-submit --class Task2 Solution/surbhi_batra_hw1.jar
path/to/ml-latest-small/ratings.csv path/to/ml-latest-small/tags.csv path/to/outputfile

ml-20m : spark-submit --class Task2 Solution/surbhi_batra_hw1.jar path/to/ml-20m/ratings.csv
path/to/ml-20m/tags.csv path/to/Output