

Description

Xiaoyu Zheng

1. Directory structure

In Solution directories, there are two folders, which contain the python one and the scala one separately. In the python folder, there are two python files, while in the scala folder, there are two scala files and an executable jar. There is no OutputFiles folder, as no output files will be created.

2. Versions

Spark: 2.2.1 Python: 2.7.15 Hadoop: 2.7 Scala: 2.11.8 SBT: 0.13.8

3. Execution rules

- 1) Put my **executable files** (including all python files and the scala jar file) directly **in the root directory of Spark**.
- 2) To utilize my work, please **get into the root directory of spark** with terminal or in the command line.
- 3) Then to execute my work, you need the following commands. Note: These commands will overwrite the files which have the same name as my output files in the root directory of Spark.

Task1:

Python:

bin/spark-submit Xiaoyu_Zheng_TwitterStreaming.py

Scala:

bin/spark-submit --class TwitterStreaming Xiaoyu_Zheng_hw5.jar

Task2:

Python:

bin/spark-submit Xiaoyu_Zheng_DGIMAlgorithm.py

Scala:

bin/spark-submit --class DGIMAlgorithm Xiaoyu_Zheng_hw5.jar

- 4) The output txt files will also be in the root directory of Spark.