

Jiayue Shi's Description

1. Version

Spark version: 2.2.1

Scala version: 2.11

Hadoop version: 2.7

2. How to run

First of all, the input files should under the data folder.

Task1

```
./bin/spark-submit --class TrackHashTags Jiayue_Shi_hw5.jar  
SxbNloG7GiilAwbylHEtYJgNt  
HvdhruYesRozcQbAy7oJYokObwtdvbAfQrglbP4ykKJearN0X6 986059857311641601-  
xy14aoe6esJVtCePE2jVvZjfFLHCagd  
DaTpBqFXHISgfELAsUHyUs1t5iOA5fGblp24D6Drqcoez
```

```
./bin/spark-submit --class AveTweetLength Jiayue_Shi_hw5.jar  
SxbNloG7GiilAwbylHEtYJgNt  
HvdhruYesRozcQbAy7oJYokObwtdvbAfQrglbP4ykKJearN0X6 986059857311641601-  
xy14aoe6esJVtCePE2jVvZjfFLHCagd  
DaTpBqFXHISgfELAsUHyUs1t5iOA5fGblp24D6Drqcoez
```

Task2

```
java -cp StreamingSimulation.jar StreamingSimulation data/userId.txt 9999 100
```

```
./bin/spark-submit --class UniqueUserCount Jiayue_Shi_hw5.jar localhost 9999
```