

Anmol Chawla Description File

Spark Version: 2.2.1

Python Version: 2.7

Task1

1. Create a new folder and copy the file Anmol_Chawla_TwitterStreaming.py
2. Open the command terminal in the same folder and write the following line ("spark-submit Anmol_Chawla_TwitterStreaming.py")
3. After you run the python file you will see the following printed on the terminal: Number of twitter from the beginning, Top five hashtags with their counts and the average length of a twitter till that point.

Task2

1. Create a new folder and copy the file Anmol_Chawla_DGIMAlgorithm.py and streaming4inf553assignment5
2. Open two command terminals in the same folder and write the following line in the first terminal ("java -cp streaming4inf553assignment5.jar StreamingSimulation 9999 100") and in the second write the following ("spark-submit Anmol_Chawla_TwitterStreaming.py")
3. After you run the python file you will see the following printed on the terminal: Estimated number of ones in the last 1000 bits , then followed by Actual number of ones in the last 1000 bits.