Part B

In this part, we developed a Storm application to process streaming data from Twitter.

1 Final submitted code:

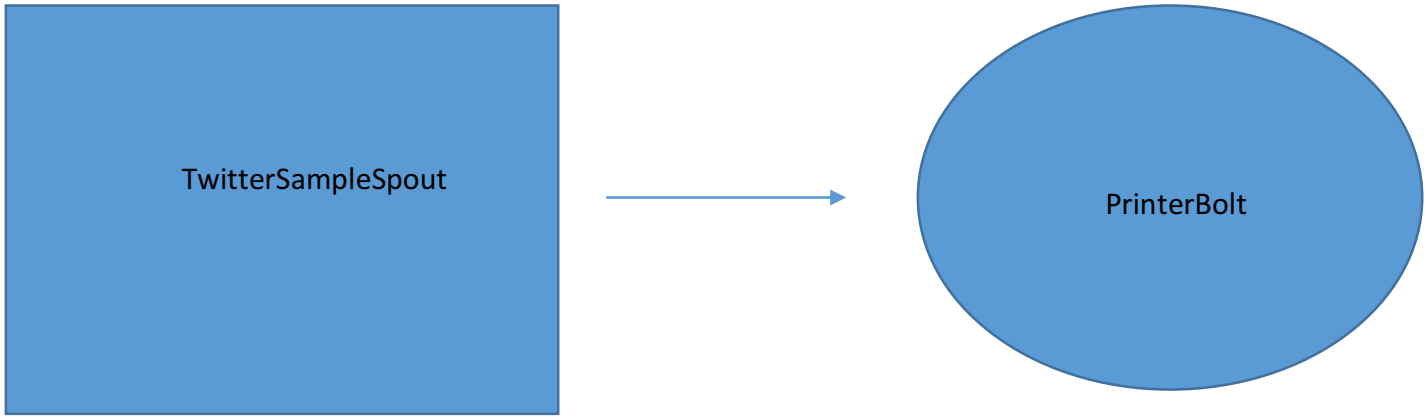
a. Question 1: run PartB/PartBQuestion1.sh to see the result. It will run locally by default. Uncomment the corresponding line then you can run it remotely, however you need to start nimbus and supervisor before remotely running it. The results are stored in local disk /home/ubuntu/q1_tweets.txt in the machine runs it.
Related files: Question1.java, TwitterSampleSpout.java, PrinterBolt.java.

b. Question 2: run PartB/PartBQuestion2.sh to see result. It will run locally by default. Uncomment the corresponding line then you can run it remotely, however you need to start nimbus and supervisor before remotely running it. The results are stored in local disk /home/ubuntu/q2_tweets.txt and /home/ubuntu/q2_ranking.txt in the machine runs it.
Related files: Question2.java, TwitterSampleSpout.java, HashtagGenerator,java, ThresholdGenerator.java, TweetsPrinterBolt.java, RollingCountBolt.java, IntermediateRankingBolt.java, TotalRankingsBolt.java, PrintRankBolt.java.

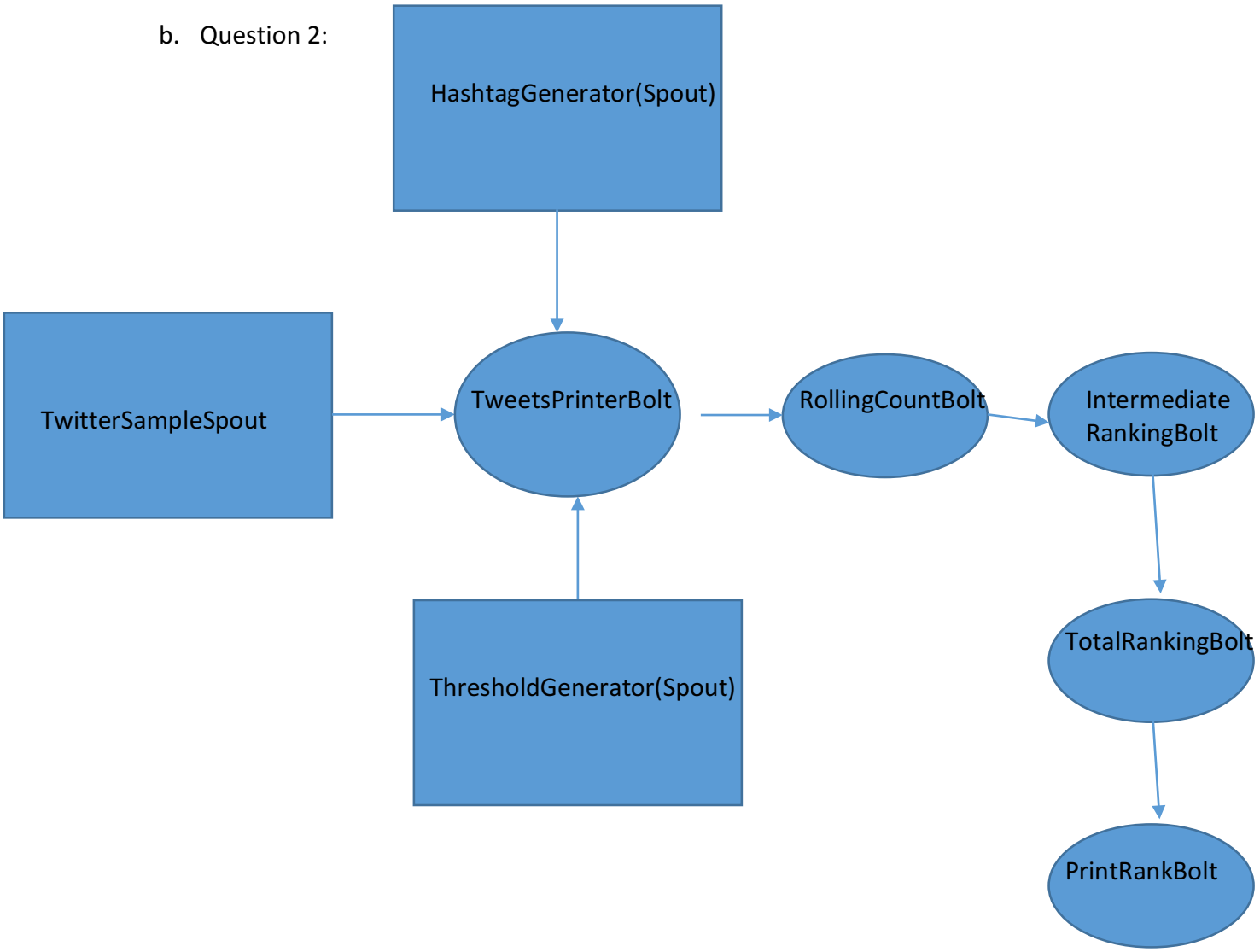
Noted that the popular hashtags now may be not that popular when you grade it.

2 Topology graphs:

a. Question 1:



b. Question 2:


HashtagGenerator(Spout)
TwitterSampleSpout
TweetsPrinterBolt
RollingCountBolt
Intermediate RankingBolt
ThresholdGenerator(Spout)
TotalRankingBolt
PrintRankBolt