

CS 6847 : Cloud Computing
Assignment 7 : Trend Topic Using Storm

– Mahesh Gupta (CS12M018)

A. Code Structure

1. *Topic Spout*: TopicSpout class has been defined which has predefined set of 10 words out of which it emits a random word once in a 50 millisecond.

2. *Bolts*: Two bolts have been defined

2.1 *Counter Bolt* : This is attached to topic bolt and uses field grouping, thus each word is sent to a specific bolt. Each bolt keeps track of count of how many times it has seen the topic and emits the frequency of latest word.

2.2 *Global Counter Bolt*: This is a global bolt which receives data from all the counter bolt. As there is only one global counter bolt, it maintains most frequent word and its count. Every time it receives the tuple it compares it with current value and updates accordingly. It also maintains the last time when it emitted output and emits only if last output was emitted at least 10 seconds (a configurable value) before.

B. Execution

The topology has been defined as shown above, 5 instances of topic bolt and 10 instances of counter bolt and 1 instance of global counter bolt has been used.

Execution happens in a local cluster. Debugging has been turned off. The topology is allowed to execute for 100 seconds after which it is shut down.

C. Results

As the topics are generated randomly, there is no deterministic result. At every 10 seconds the output can be seen on the console. Storm processes the data as stream and results are generated every few time unit.