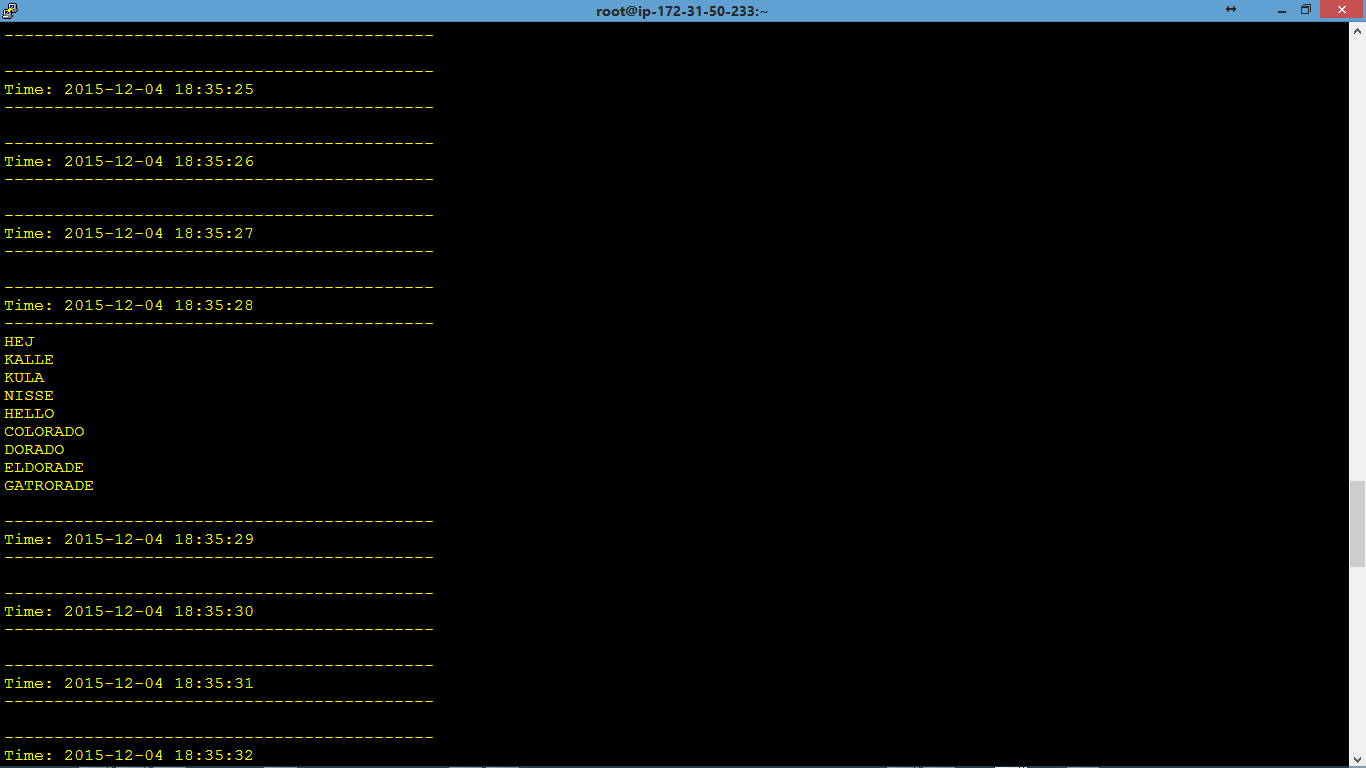
**MIDS W205 Lab # 12 Lab Spark Streaming Introduction**

Alan Wang

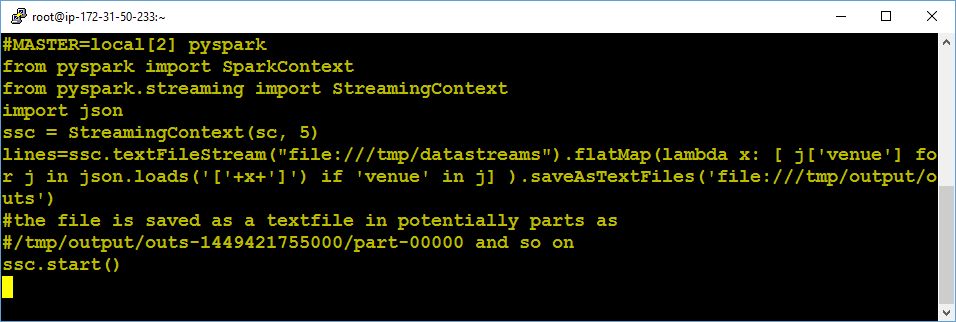
**SUBMISSION 1: Provide a screenshot of the output from the Spark Streaming process**



**SUBMISSION 2: Change the code so that you save the venue components to a text file. Submit you code.**

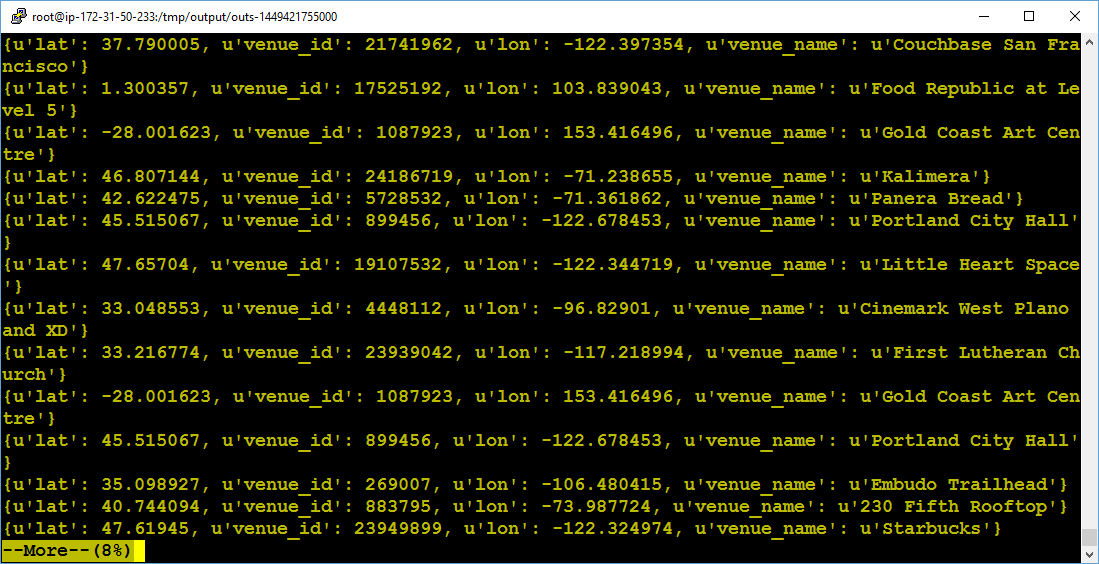
$vi lab12.s2.py

Also I attached the py file in the same ‘lab12’ folder

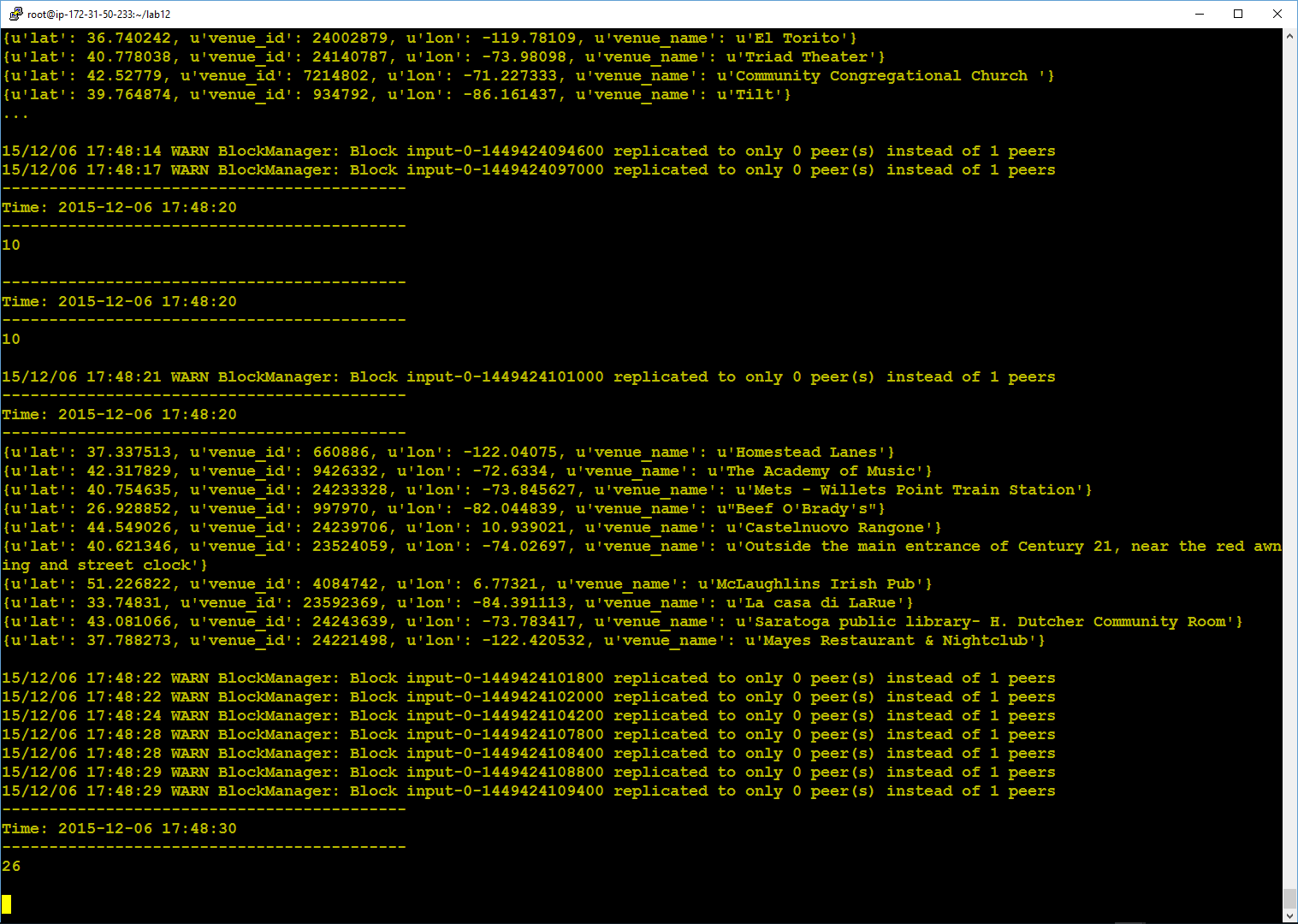


Saved Text File:

$more /tmp/output/outs-1449421755000/part-00000

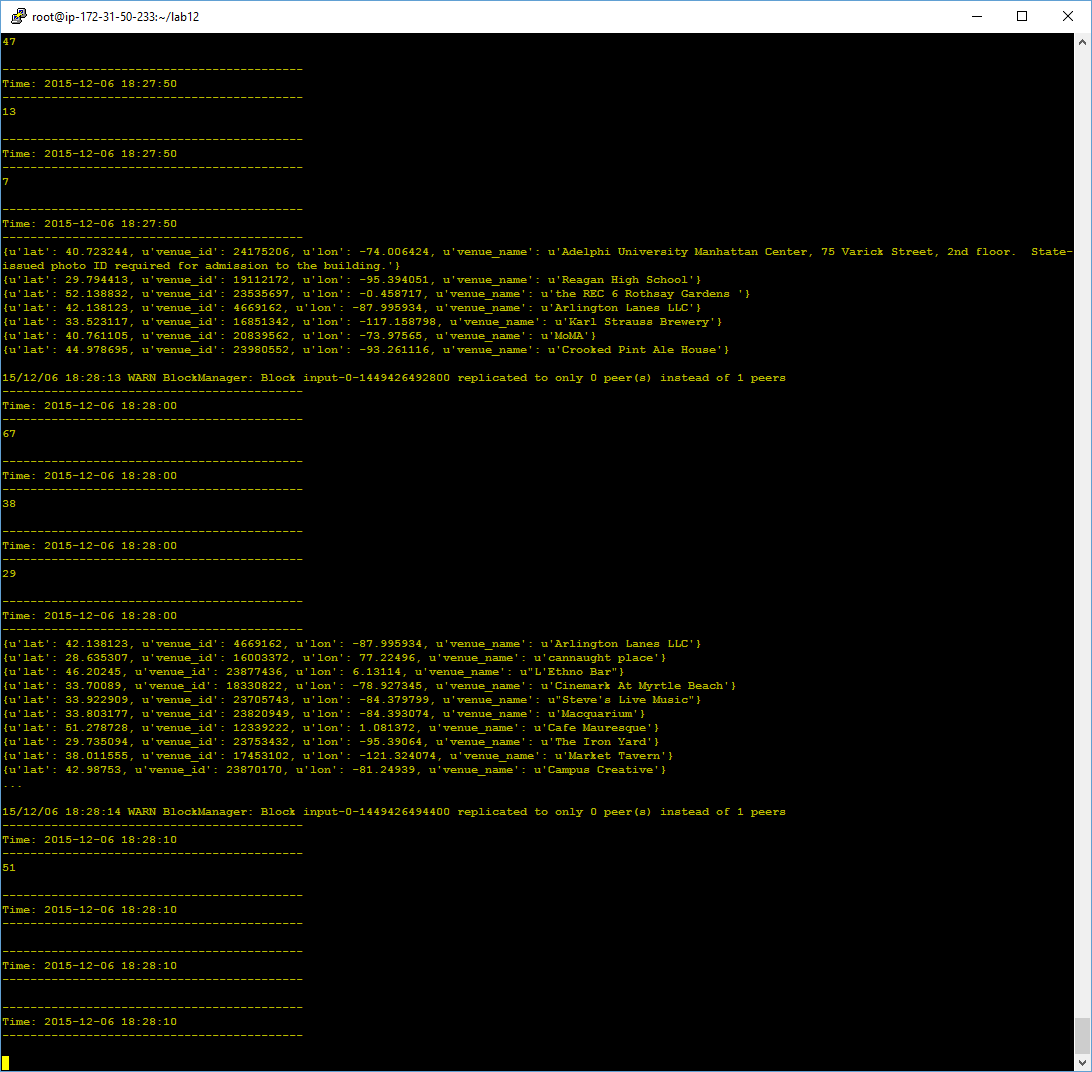


**SUBMISSION 3: Provide a screenshot showing the running Spark Streaming application.**

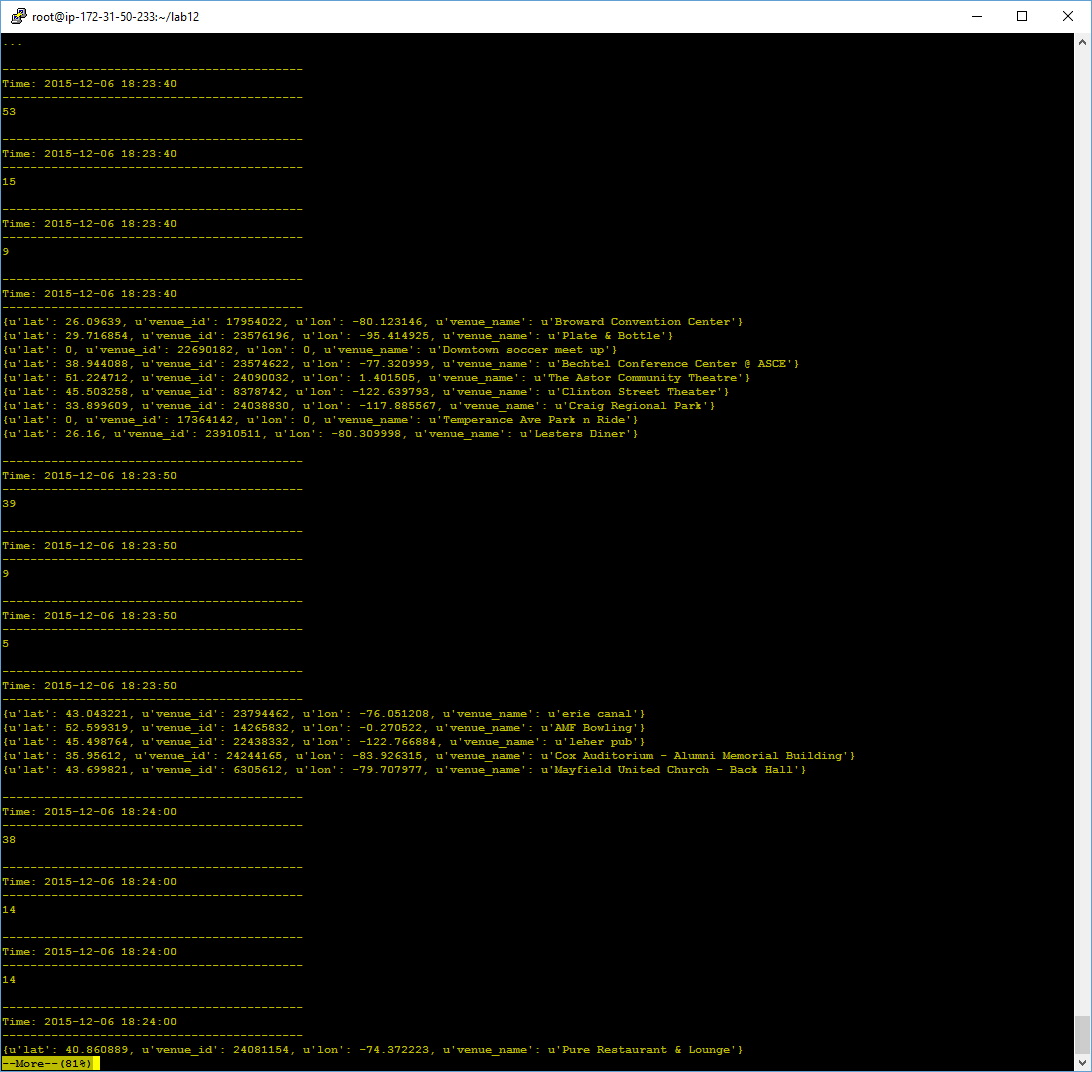


**SUBMISSION 4a: Provide a screenshot of the running Spark Streaming application that shows that the CountByWindow indeed provides an sum of the counts from the 3 latest batches. See example screenshot below.**

Screen shot with the running Spark streaming, but can’t show a window of 3 because of the warning messages:



Now, I redirect the output to a file and virtually weeded out the warning messages, and was able to catch a 3-batch window as below: **15 + 9 + 14 = 38 from 18:23:40, 18:23:50, and 18:24:00 resepectively**



**SUBMISSION 4b: Also explain what the difference is between having 10 sec batches with a 30 sec sliding window and a 30 second batch length.**

For 10-sec batch + 30 sliding window, each window covers 3 batches, and for 30-sec batch + 30 sliding window each window covers only 1 batch.