**Assignment 20.2**

Hello team,

Apologies, I could not submit the assignment 20 since I have been hospitalized yesterday for my health issue.

As of now please accept my blank document. I will make sure to complete the assignment as soon as I am cured.

Regards,

1. Real-time set of offensive words

Find the word count of the offensive words inputted by the user as per the

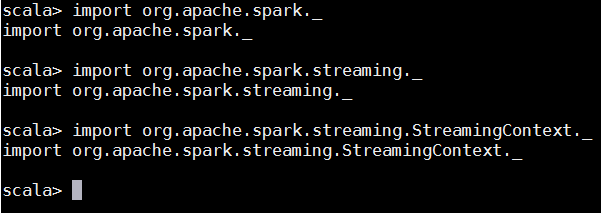
real-time set of offensive words.

We will first import the streaming packages

import org.apache.spark.\_

import org.apache.spark.streaming.\_

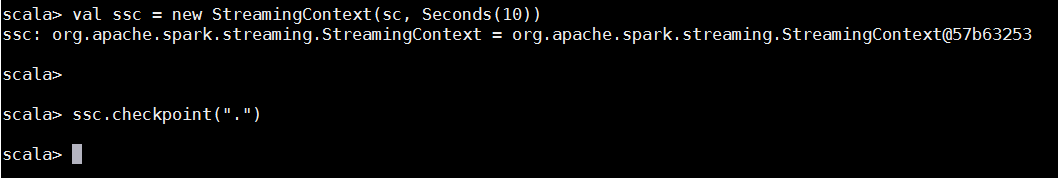
import org.apache.spark.streaming.StreamingContext.\_



Now, we will create a real-time streaming context with a window of 10 seconds

**val ssc = new StreamingContext(sc, Seconds(10))**

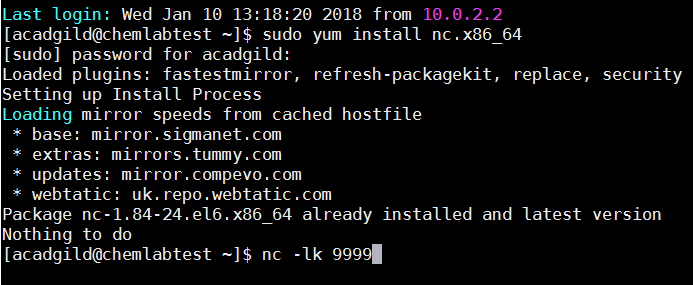
**ssc.checkpoint(".")**

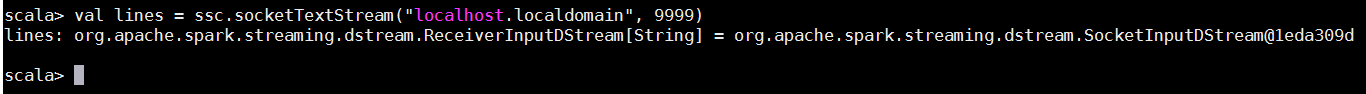


Now, we will insert lines of offensive word in **nc –lk 9999** after installing **sudo yum install nc.x86\_64**

The following command will be used to read those inserted lines

**val lines = ssc.socketTextStream("localhost.localdomain", 9999)**





Then, we will split the words with a space and count the number of times the word has been inserted

**val words = lines.flatMap(\_.split(" "))**

**val wordDstream = words.map(word => (word, 1))**

The following commands will create an RDD for the words and no. of times the word has been inserted and give us an output.

**val initialRDD = ssc.sparkContext.parallelize(List[(String, Int)]())**

**val mappingFunc = (word: String, one: Option[Int], state: State[Int]) => {**

**val sum = one.getOrElse(0) + state.getOption.getOrElse(0)**

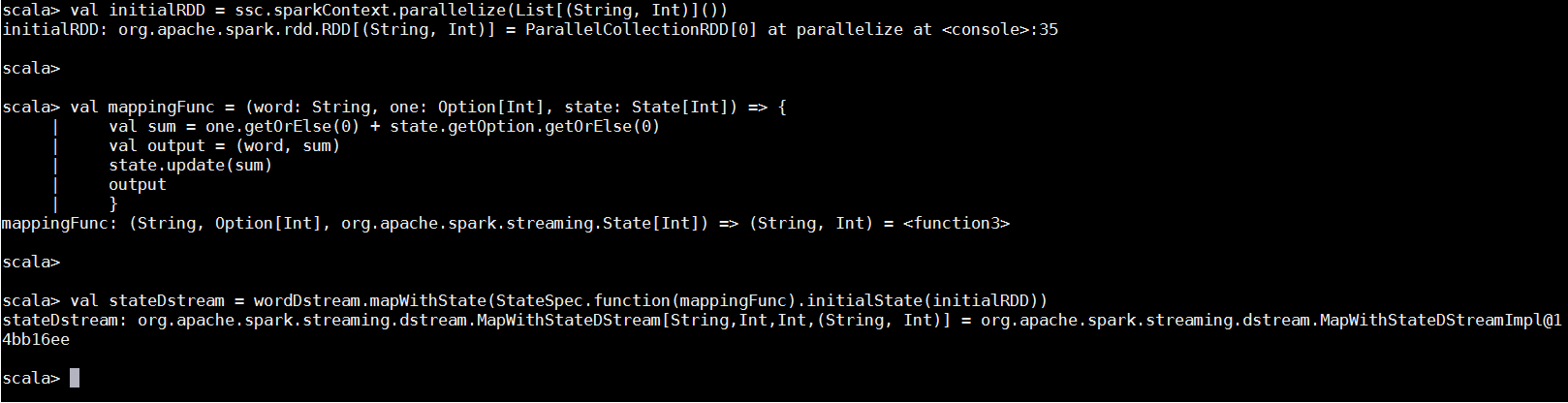
**val output = (word, sum)**

**state.update(sum)**

**output**

**}**

**val stateDstream = wordDstream.mapWithState(StateSpec.function(mappingFunc).initialState(initialRDD))**

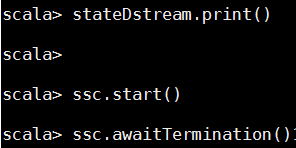


The following commands will start the real-time streaming

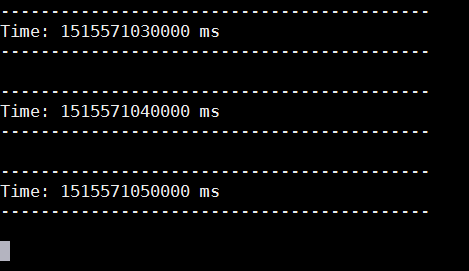
**stateDstream.print()**

**ssc.start()**

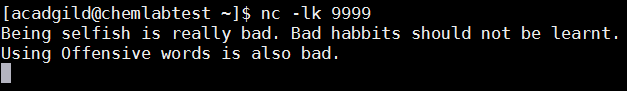
**ssc.awaitTermination()**



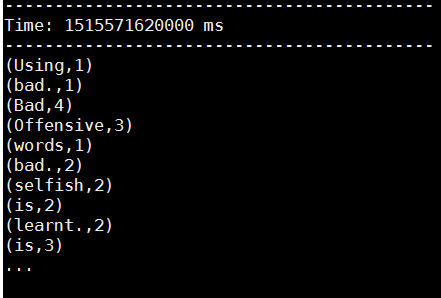
The streaming will look as following:



Following is the sentence inserted taking bad as an offensive word.



**Output:**



As you can see in the output, each word has a count of no. of times the word has been inserted. However the word “Bad” has occurred 4 times.