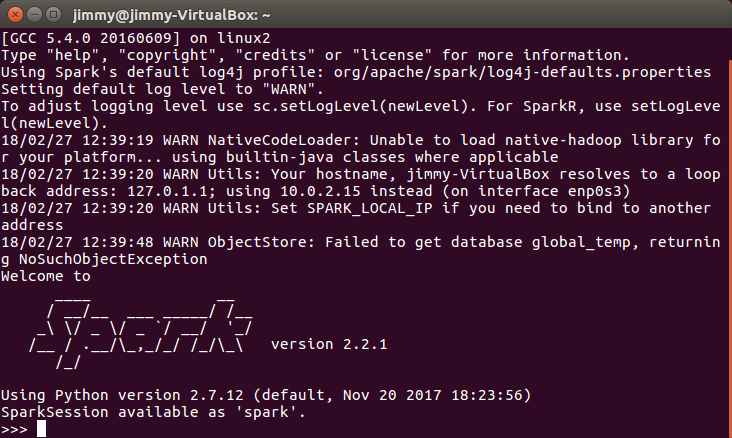
Jimmy Tran

CSC 4760

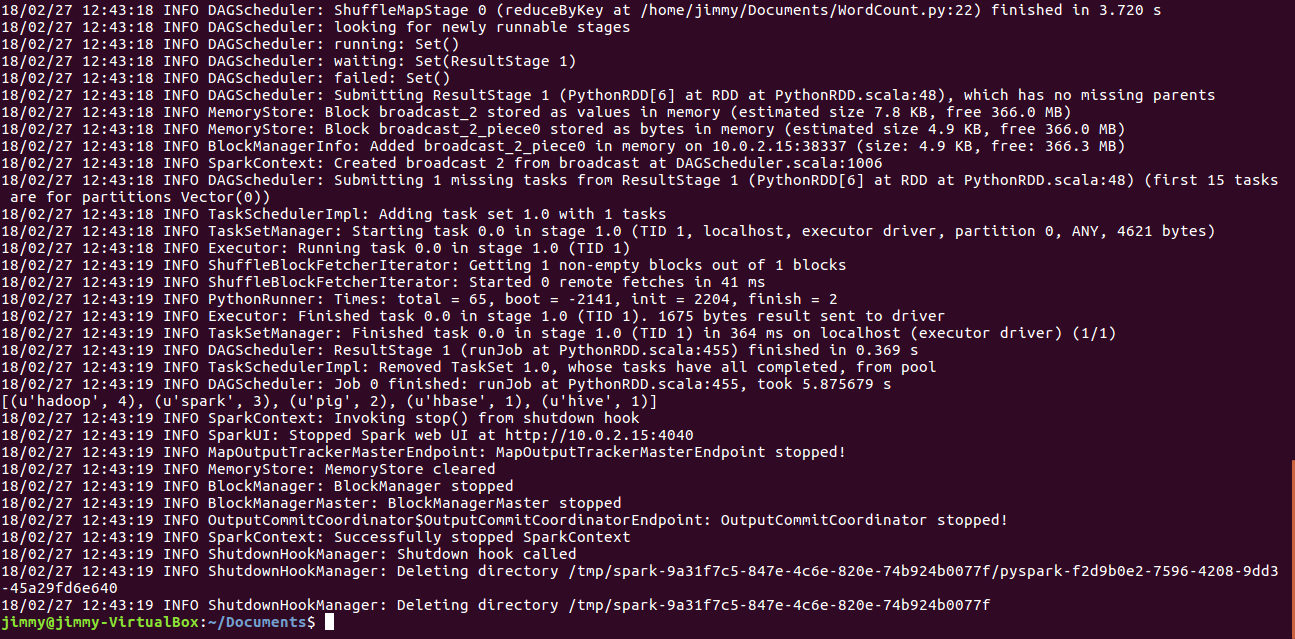
28 February 2018

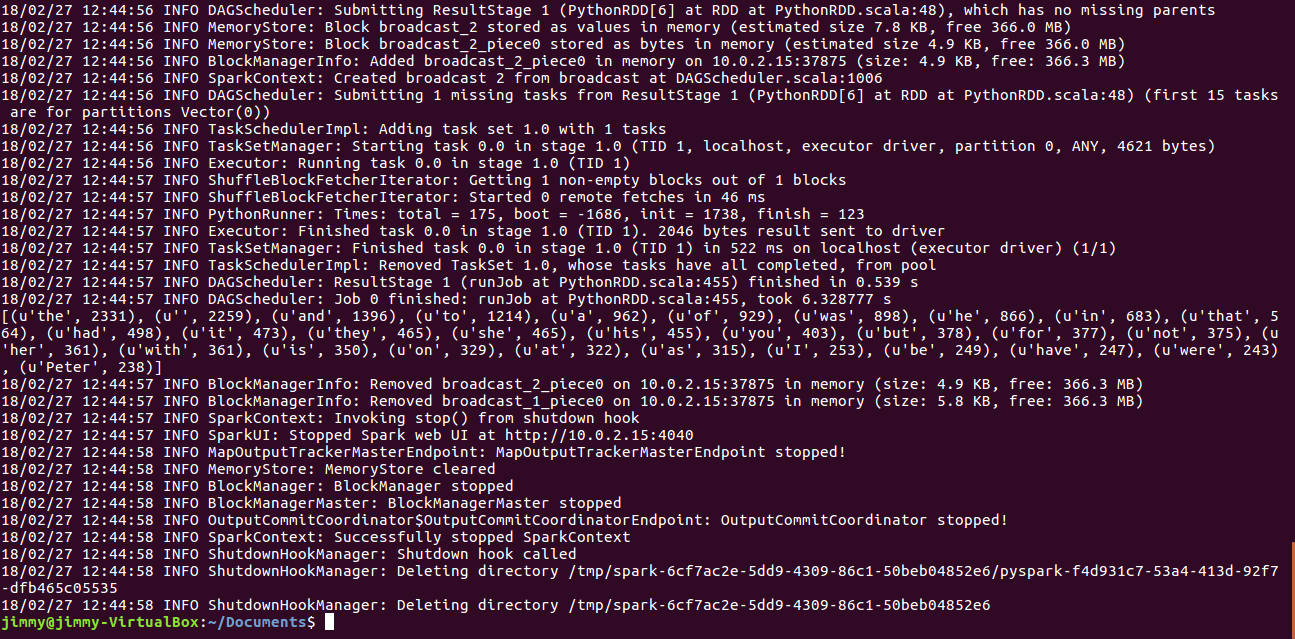
Assignment 3 Report

1) Spark installed successfully



2) test.txt output



peterpan.txt output

3) Report

The commands passed in the terminal were in the format “pyspark (.py file) (.txt file location) (top # of words, where # is the number the user inputs).

pyspark is the process/program called to run.

The .py file is the file being run by pyspark.

The .txt file is the location of the .txt file that the .py file is run on.

The last parameter is the number of top words the user wants returned or output after the program completes. For example, entering 15 would mean the user wants the 15 most common/repeated words found by the program in the given .txt file. In the case of the two files above, the top 5 and top 30 were output by the program in the middle of each screenshot in the format (word, count).