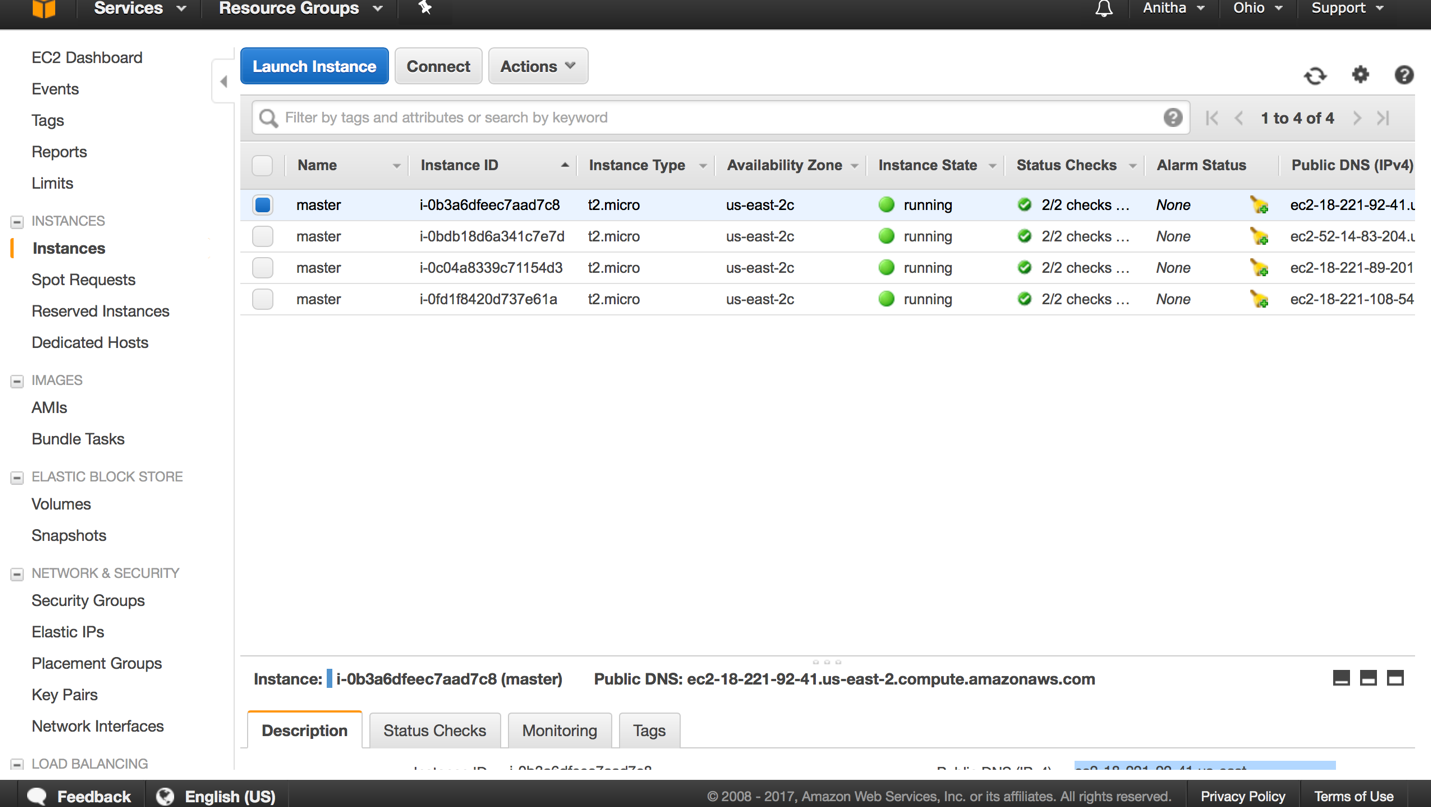
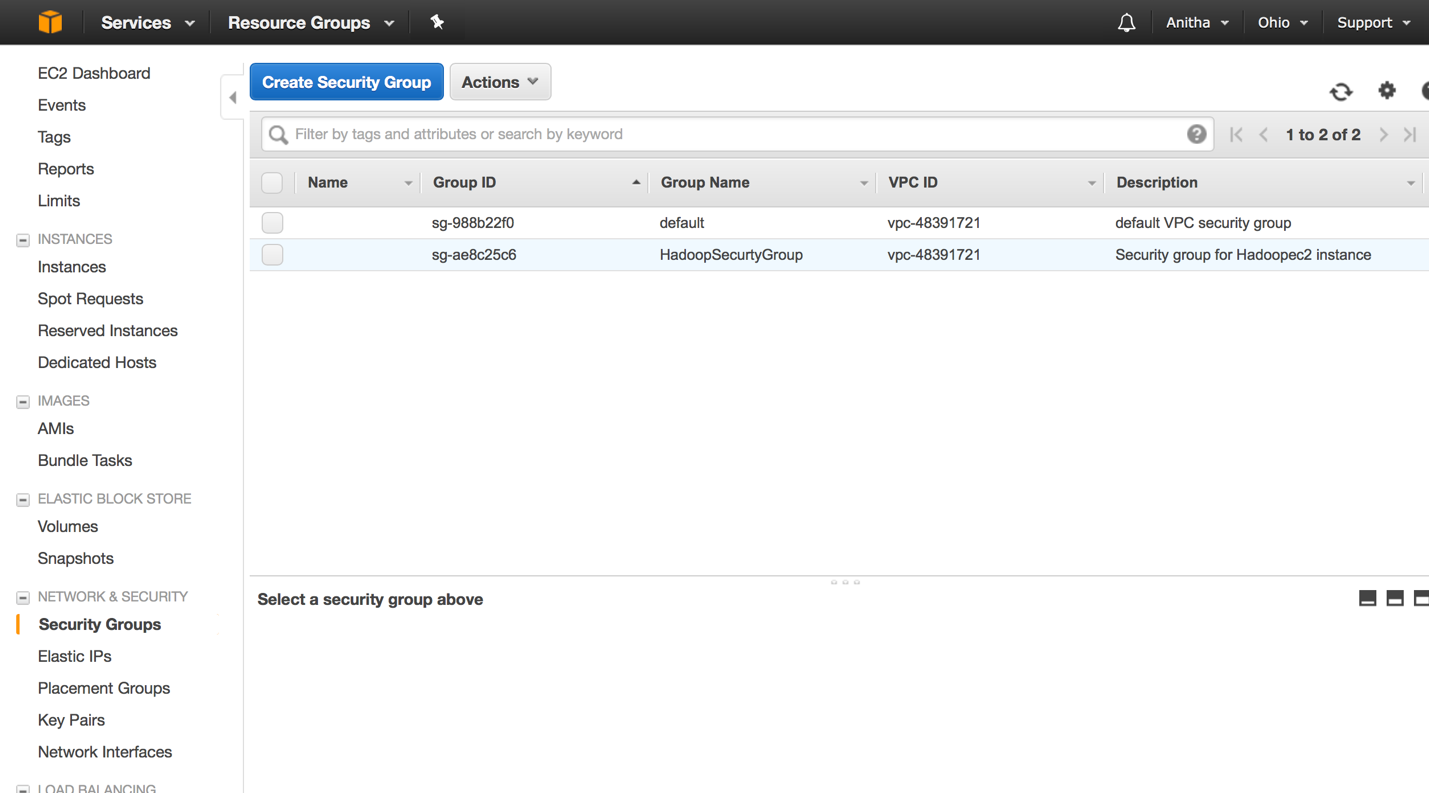
**README**

I have included three projects for task1 , task2 and task3 respectively and run them on Amazon AWS using EC2. The project ‘WordCountMapReduce’ is for task one where we count one-word frequency as in the word count example. The project ‘DoubleWordCountMapReduce’ counts double-word frequency and the project ‘WordCountMapReducePatterSearch1’ uses Distributed Cache (URI cache) to find frequency of one-words in another given list.

Each program has a mapper which maps every occurrence of the word as key to a count 1. There is a reducer which acts on the sorted and grouped <K,V> pair i.e. <word,1(count)> to reduce the mapping to its actual count. However for the third task, I have used a Hash Map to keep a track of words in the pattern.txt file so that, it just has to traverse through the ‘bible’ file which is big and check if any of the words in ‘bible’ are present in the pattern file.

**4 instances created: (1 Namenode 3 data node)**

****

****

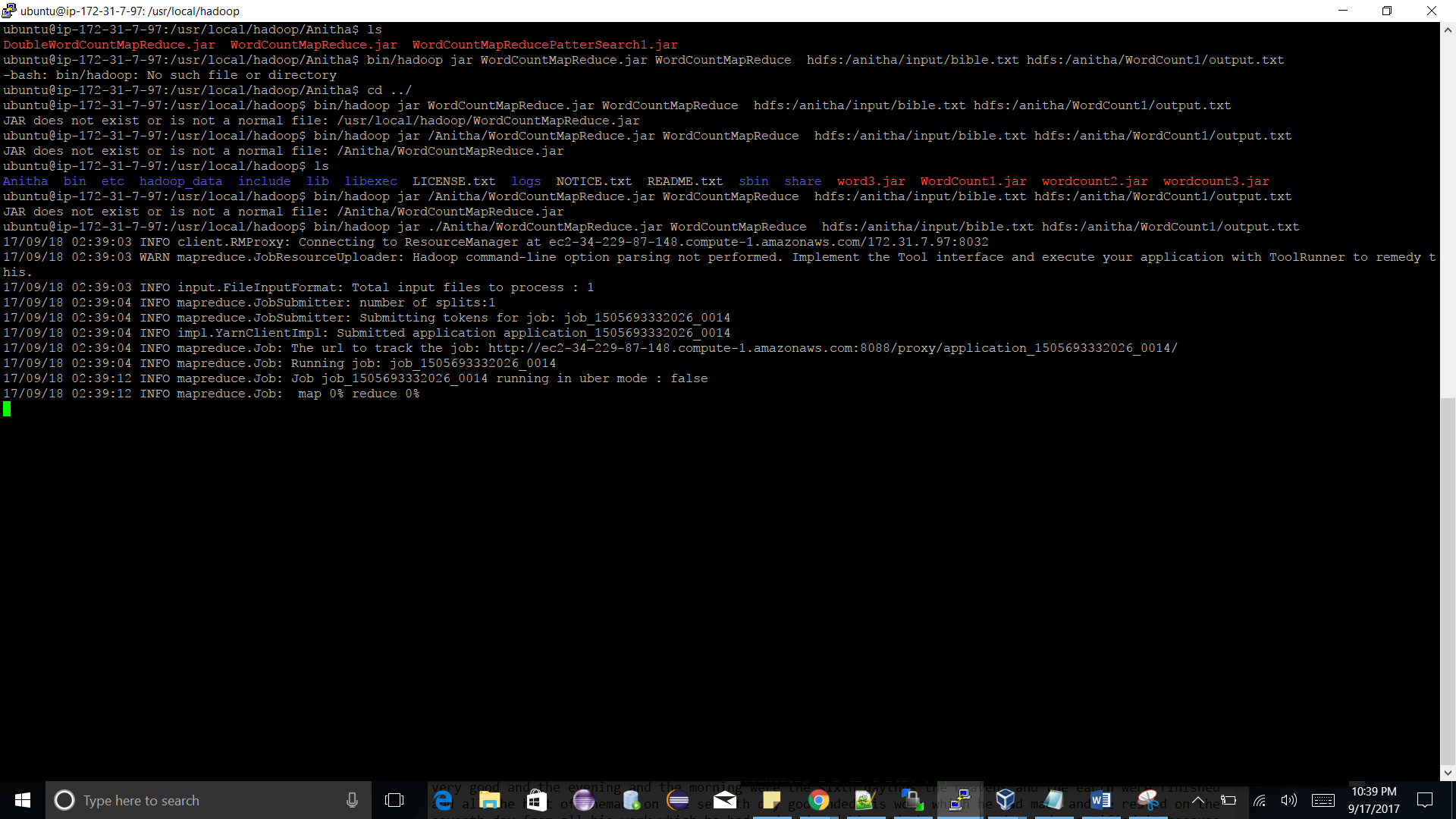
**Running the tasks:**

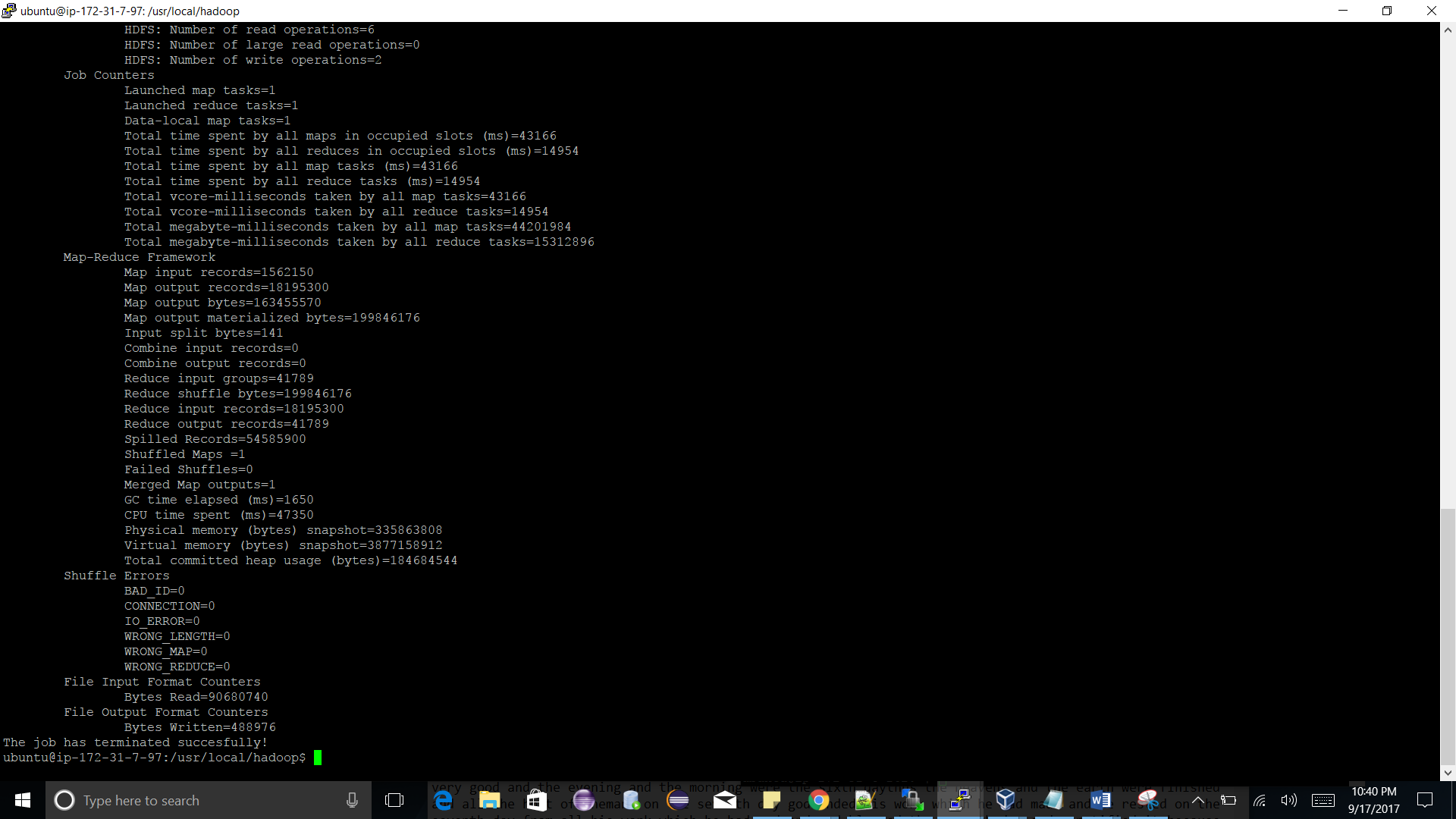
Task 1:

$ bin/hadoop jar ./Anitha/WordCountMapReduce.jar WordCountMapReduce hdfs:/anitha/input/bible.txt hdfs:/anitha/WordCount1/output.txt

Here we run the jar file from its stipulated path and the output is sent to the Anitha/WordCount1/output.txt

Please Note: The ouput.txt might show as though its 0 bytes but you need to open it and check the ‘part-r-0000’ file to check the actual output.

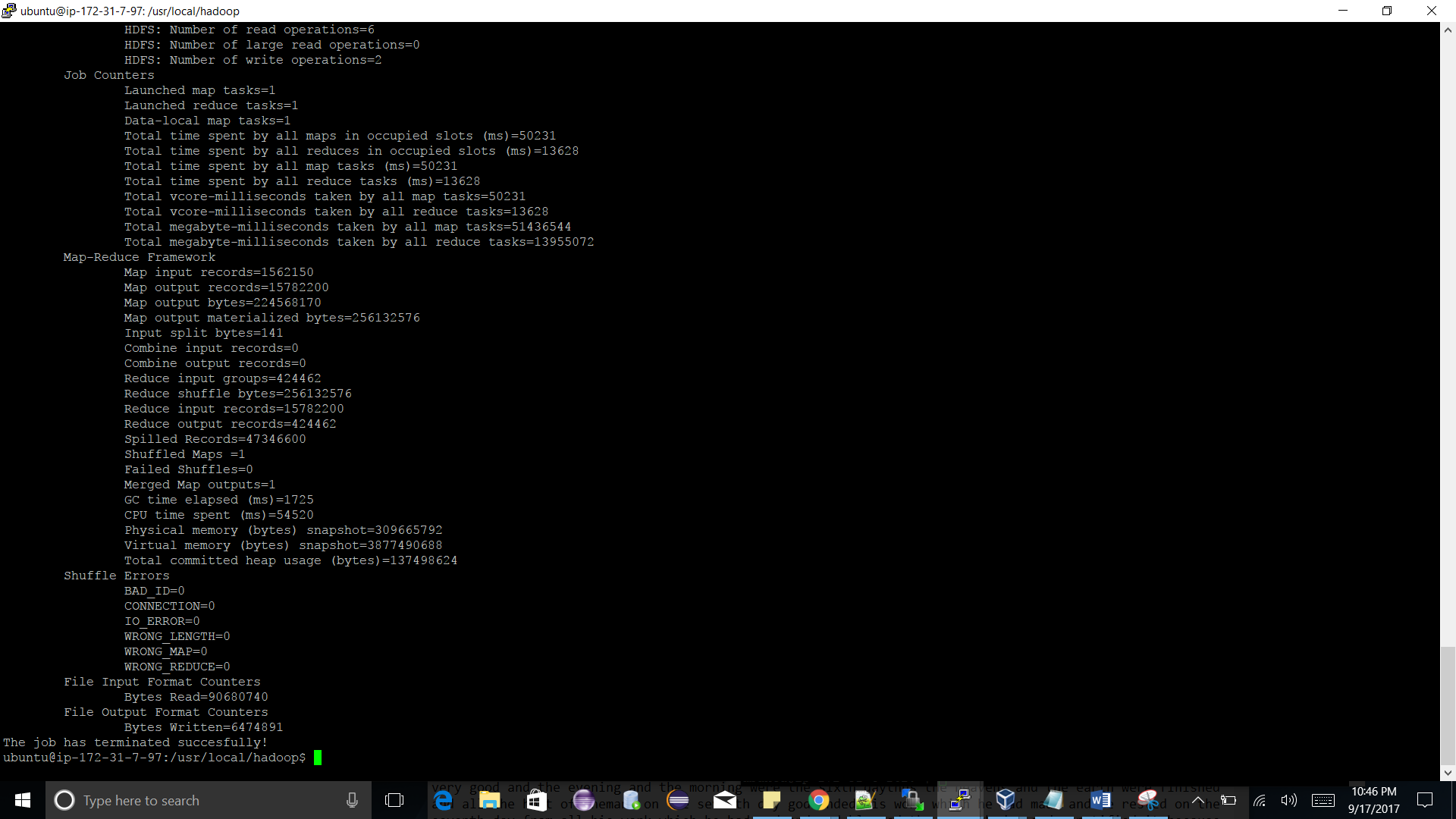


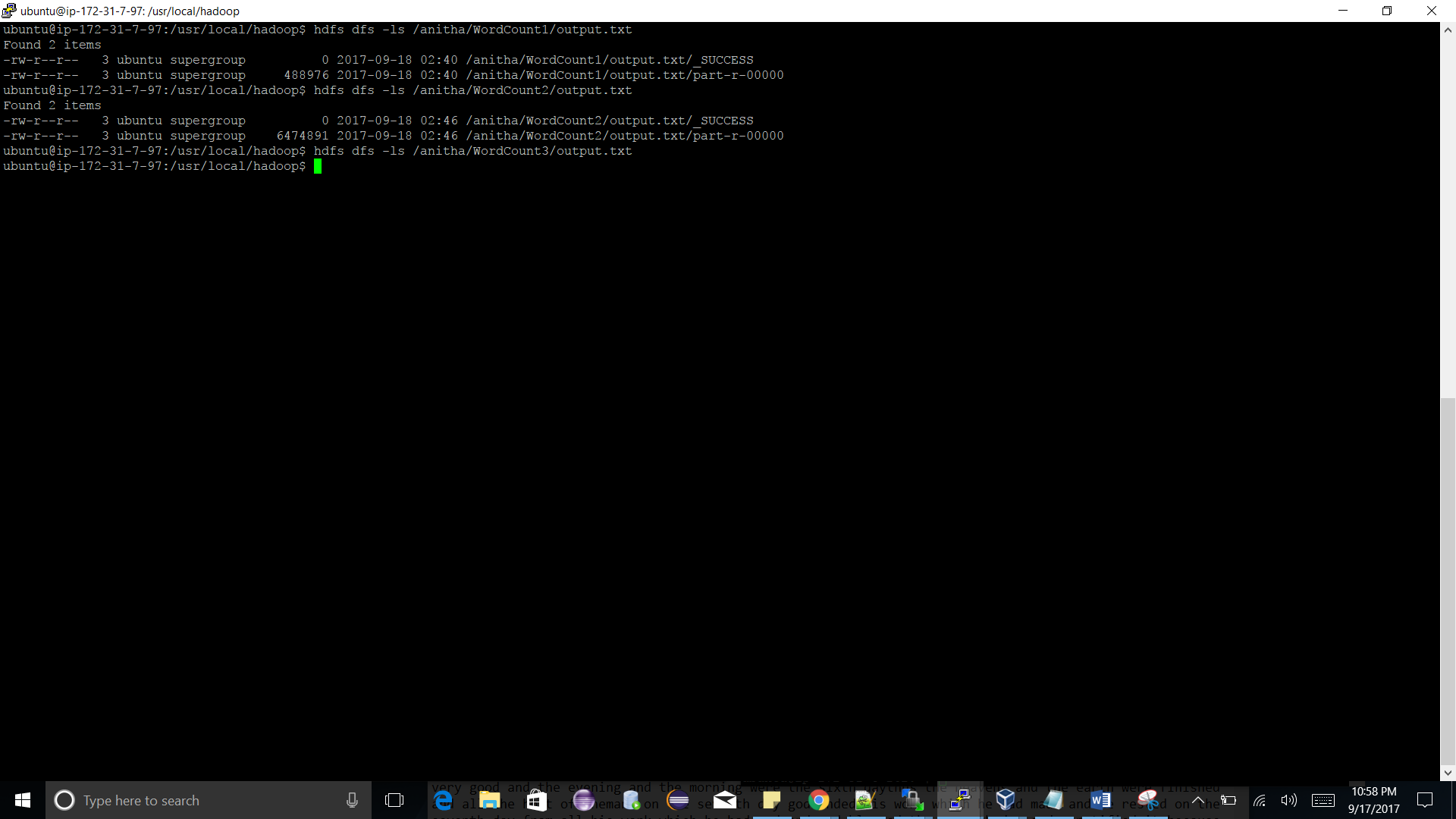


Task 2:

bin/hadoop jar ./Anitha/ DoubleWordCountMapReduce.jar DoubleWordCountMapReduce hdfs:/anitha/input/bible.txt hdfs:/anitha/WordCount2/output.txt

Here we run the jar file from its stipulated path and the output is sent to the Anitha/WordCount2/output.txt



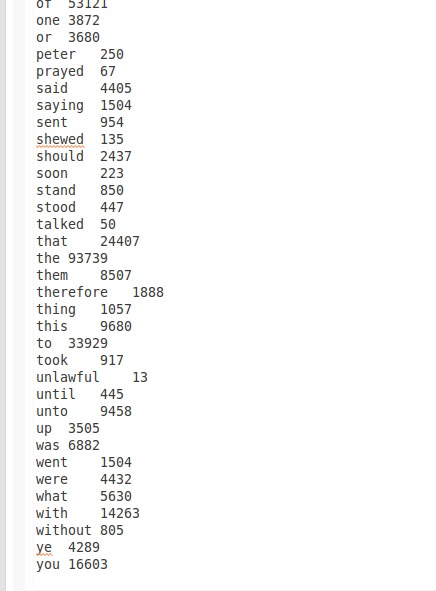


Task 3:

bin/hadoop jar ./Anitha/ DoubleWordCountMapReduce.jar DoubleWordCountMapReduce hdfs:/anitha/input/bible.txt hdfs:/anitha/WordCount2/output.txt

Here we run the jar file from its stipulated path and the output is sent to the Anitha/WordCount2/output.txt

**Please Note:** Each output is in internal files. Please refer to output of each task according to readme.



**Link to the code:** <https://github.com/Anitha19r/HadoopMapReduceAssignment1>

**Link to the output:** <http://ec2-34-229-87-148.compute-1.amazonaws.com:50070/explorer.html#/anitha>

Task1: <http://ec2-34-229-87-148.compute-1.amazonaws.com:50070/explorer.html#/anitha/WordCount1>

Task2: <http://ec2-34-229-87-148.compute-1.amazonaws.com:50070/explorer.html#/anitha/WordCount2>

Task2: <http://ec2-34-229-87-148.compute-1.amazonaws.com:50070/explorer.html#/anitha/WordCount3>