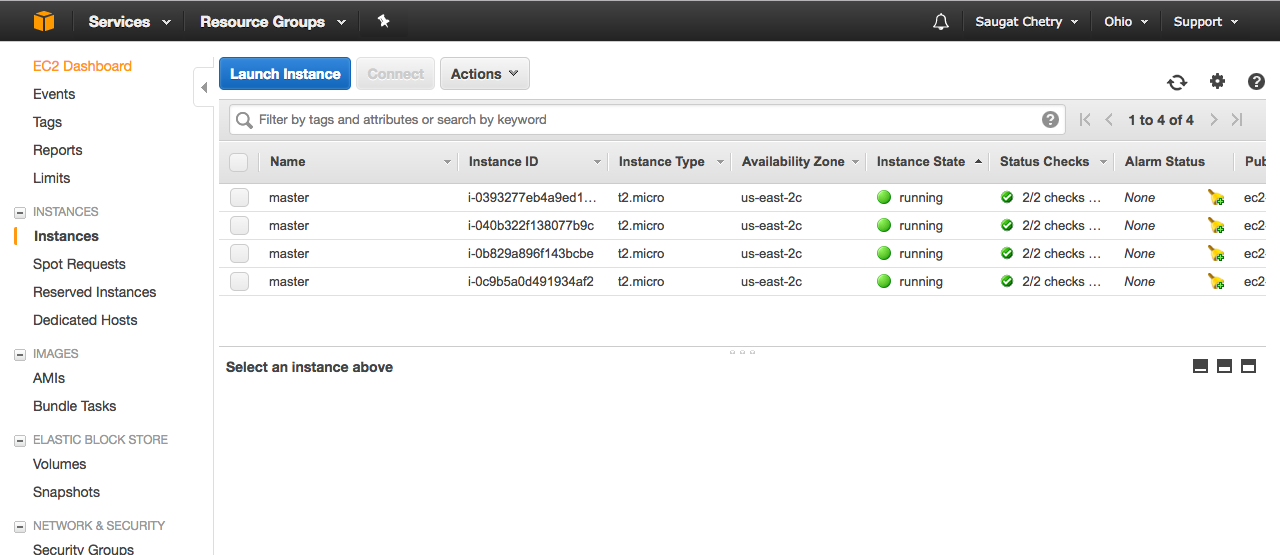
Read Me :-

Below are the three task which were part of the assignment. All these tasks are coded in JAVA and are executed/run on AWS with 3 datanodes and 1 master node

Below is the screen shot of the 4 nodes in AWS –



1. SimpleWordCount-

This program calculates the frequencies of the word present in the input file. There are 3 classes in it – SimpleWordCount.java, WordMapper.java, WordReducer.java. Below is the screen shot -

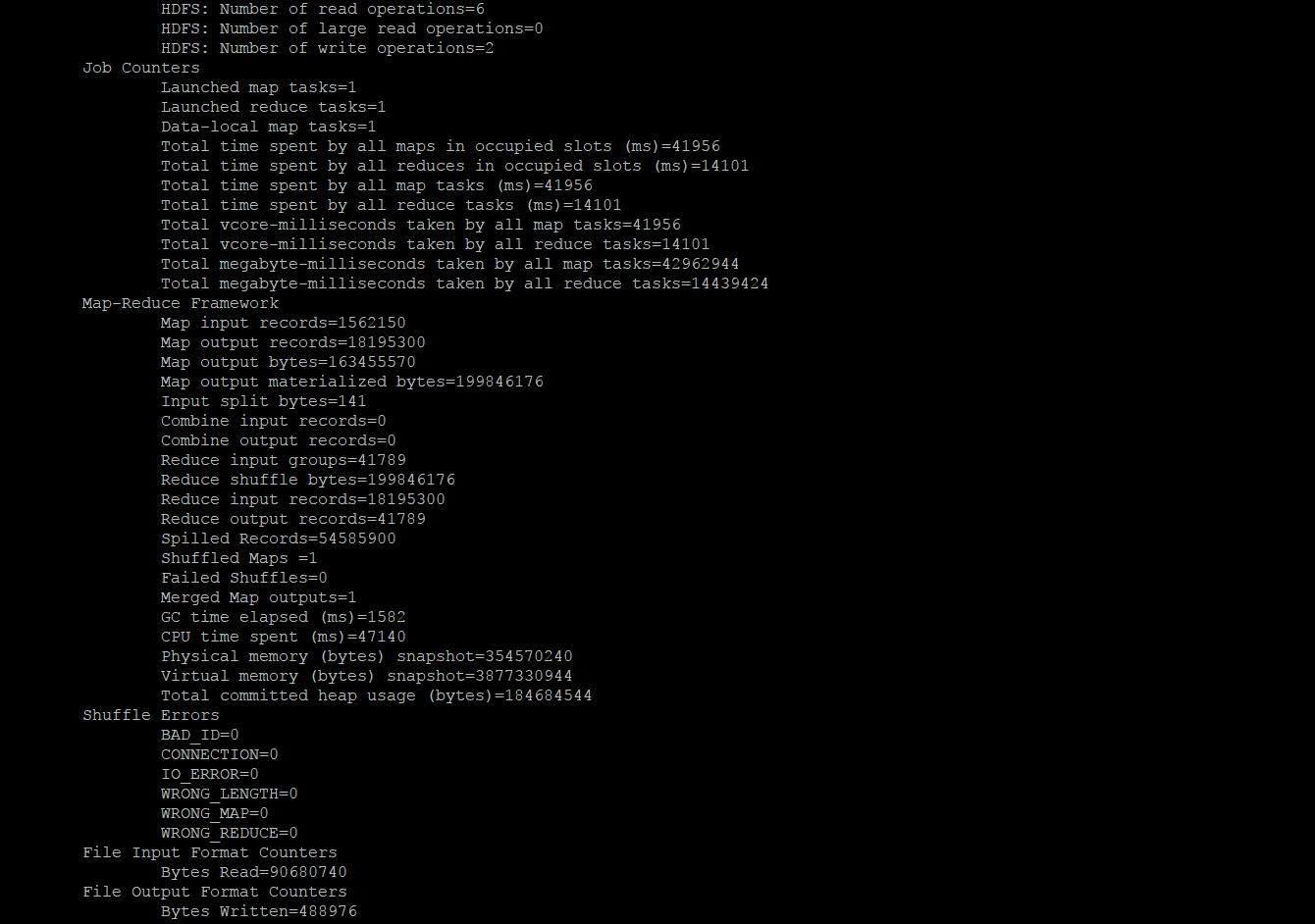


Fig. Successful Execution of the SimpleWordCount.java program

Outfile Link - <https://tinyurl.com/y92nvumq>

1. PairWordCount –

The program calculates the frequencies of all the words in pairs. The words are paired if they are separated by a blank space. The programs first maps all the pairs that are formed in from the given input text. Then the reducer iterates through each of the pairs and calculates the frequency of each pair of the words.

The screen shot of the execution of the program is given below –

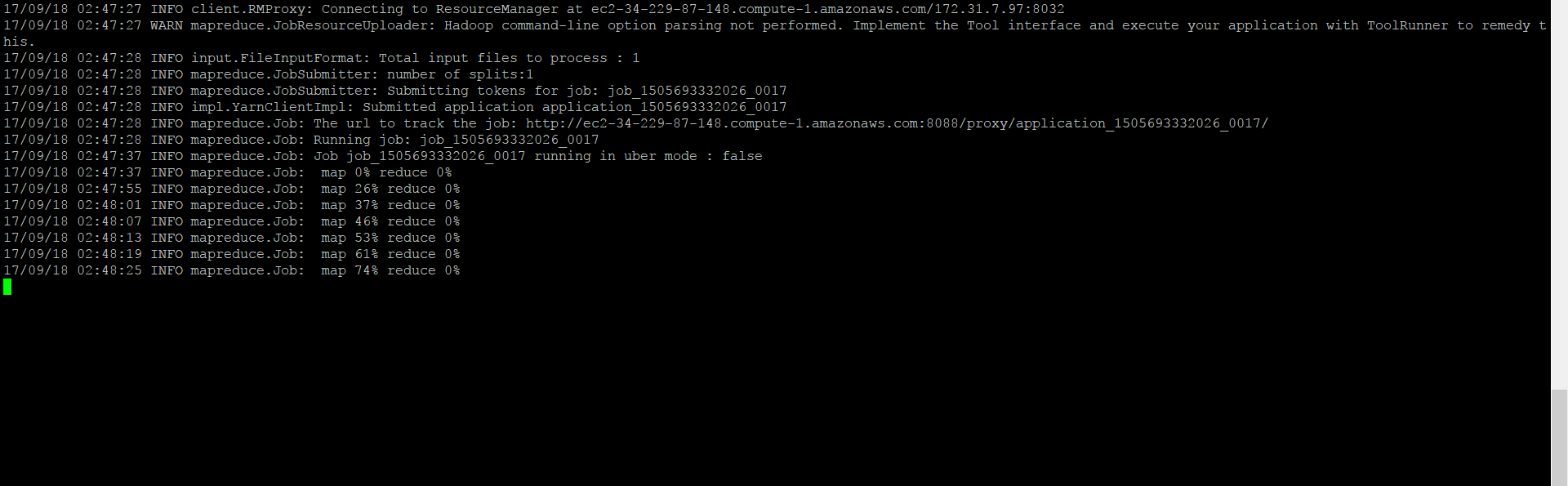
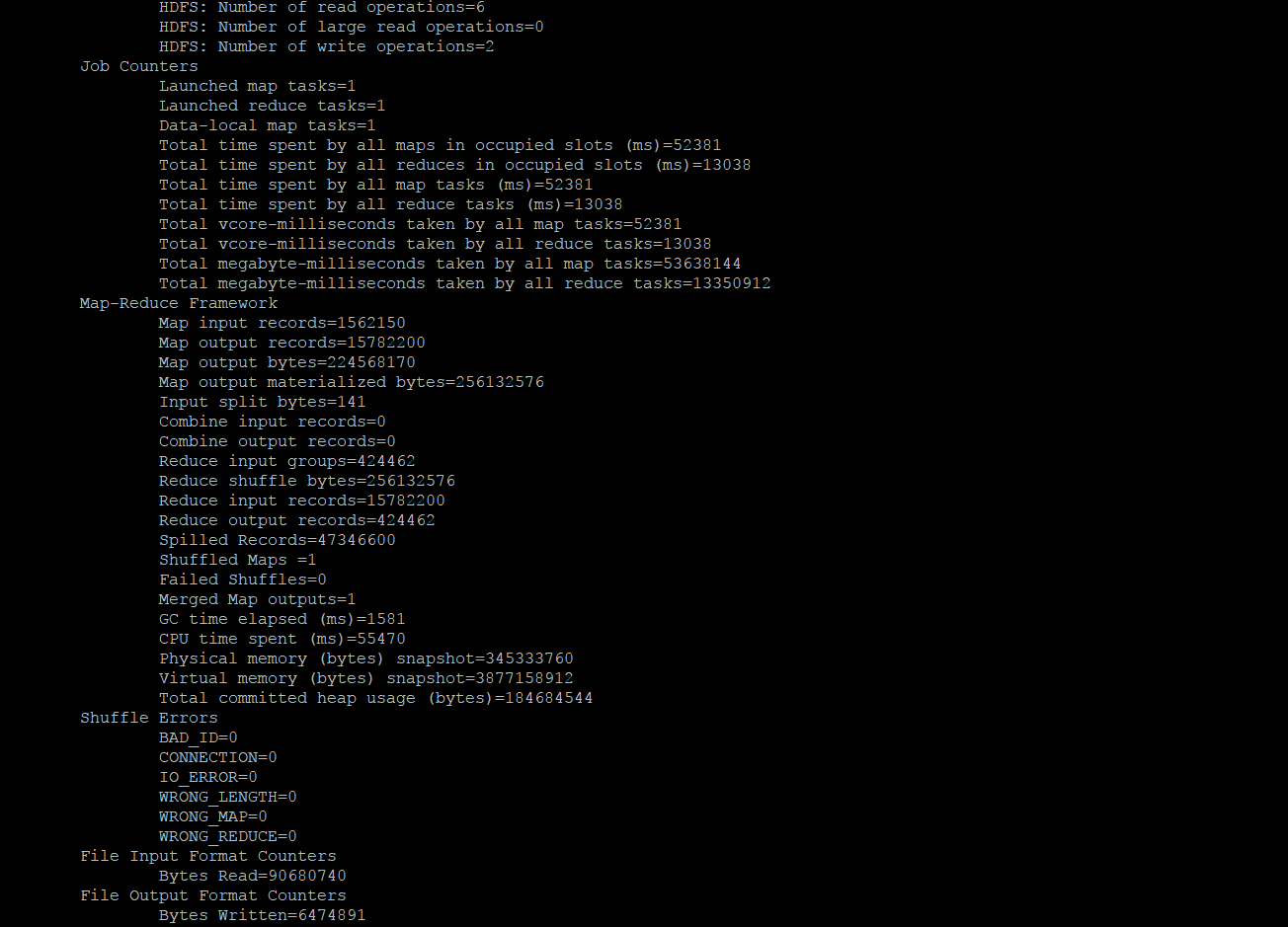


fig. Running of the second task.

The PairWordCount has completed successfully in the below screenshot –

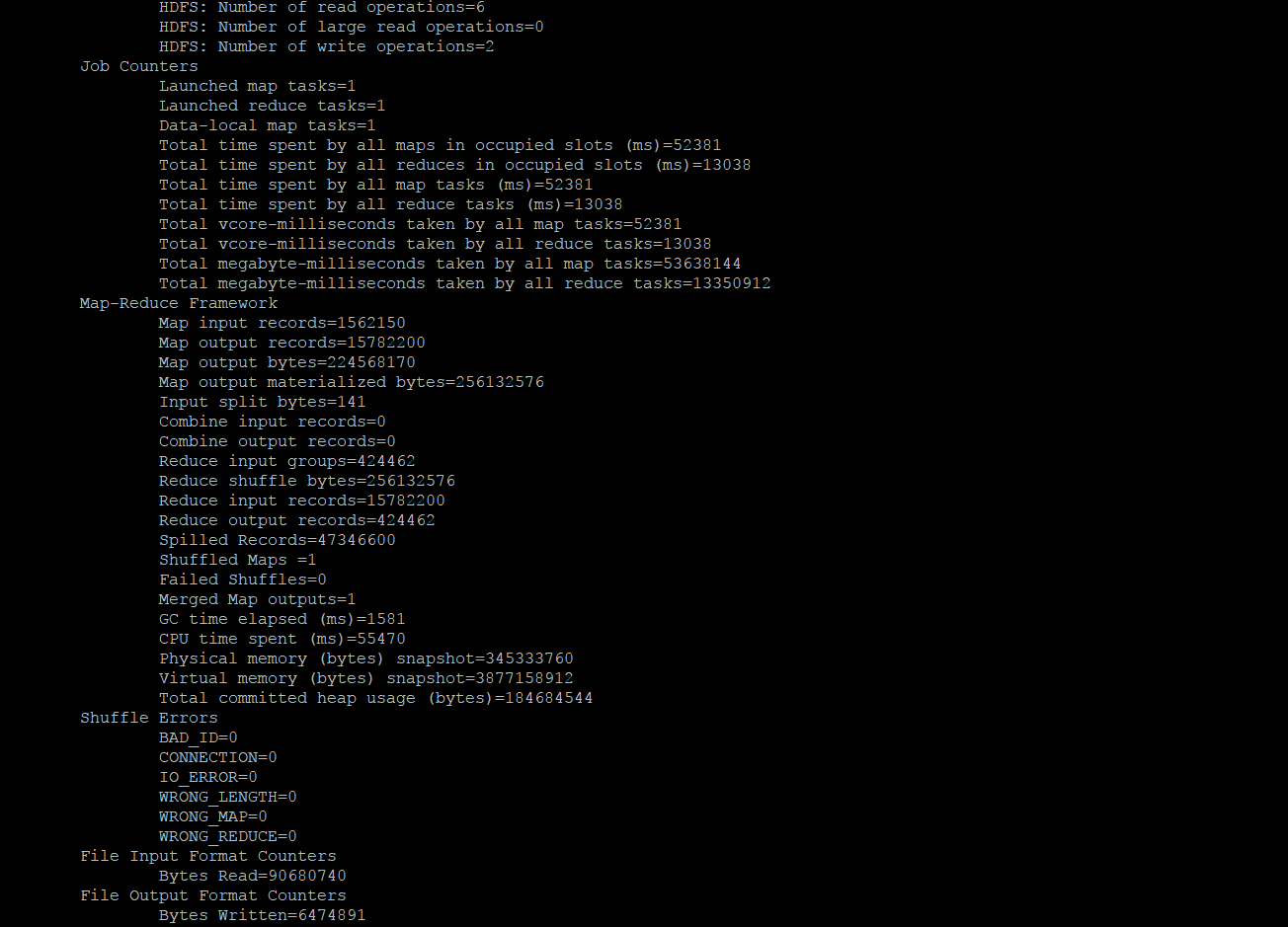


Output- <https://tinyurl.com/y8fb3b7o>

1. ListWordCount-

This programs takes 2 arguments/inputs – a text file (bible+shakes.nopunc) and a word-pattern file (word-pattern.txt). The program then iterates thorough the entire file and counts all the words that appear in the text file with all the words that are present in the pattern file. The mapper maps all the words that are present in both the input files and the reducer will count the frequency of all the words found.

Below is the screen shot of the execution of the this program –



Links – Below are the links to the code base in github and also the output in the aws.

1.<https://github.com/saugatchetry/WordCount_Assignment.git>

2.[http://ec2-34-229-87-148.compute-1.amazonaws.com:50070/explorer.html - /saugat](http://ec2-34-229-87-148.compute-1.amazonaws.com:50070/explorer.html#/saugat)