**Link to AWS output data**

**Task 1**

<http://ec2-13-59-120-93.us-east-2.compute.amazonaws.com:50070/explorer.html#/output/WordCountBible>

**Task 2**

<http://ec2-13-59-120-93.us-east-2.compute.amazonaws.com:50070/explorer.html#/output/TwoWordCountBible>

**Task 3**

Result of word-patterns.txt

<http://ec2-13-59-120-93.us-east-2.compute.amazonaws.com:50070/explorer.html#/output/WordPatterns>

Final Result

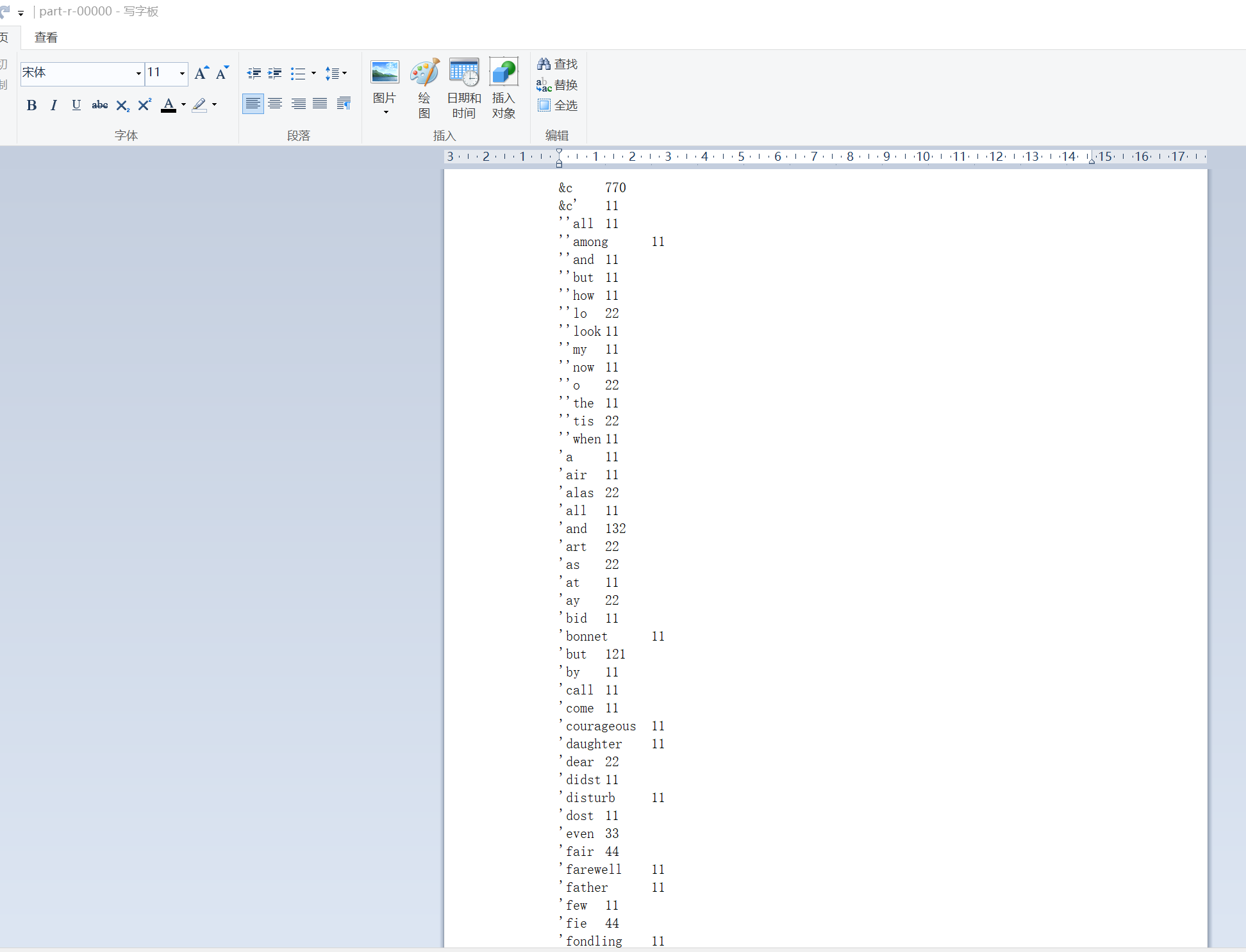
<http://ec2-13-59-120-93.us-east-2.compute.amazonaws.com:50070/explorer.html#/output/DCWordCount1>

**GitHub Link**

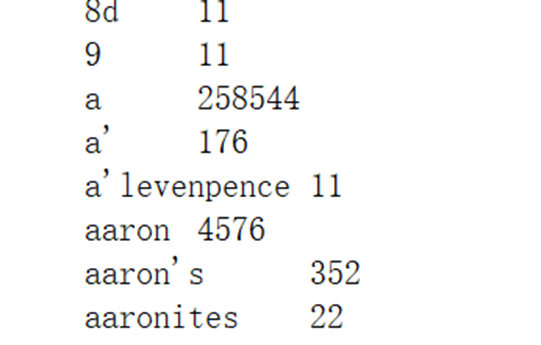
https://github.com/boubouplanet/CloudComputingAssignment1

**Results and Explanation**

**Task 1**

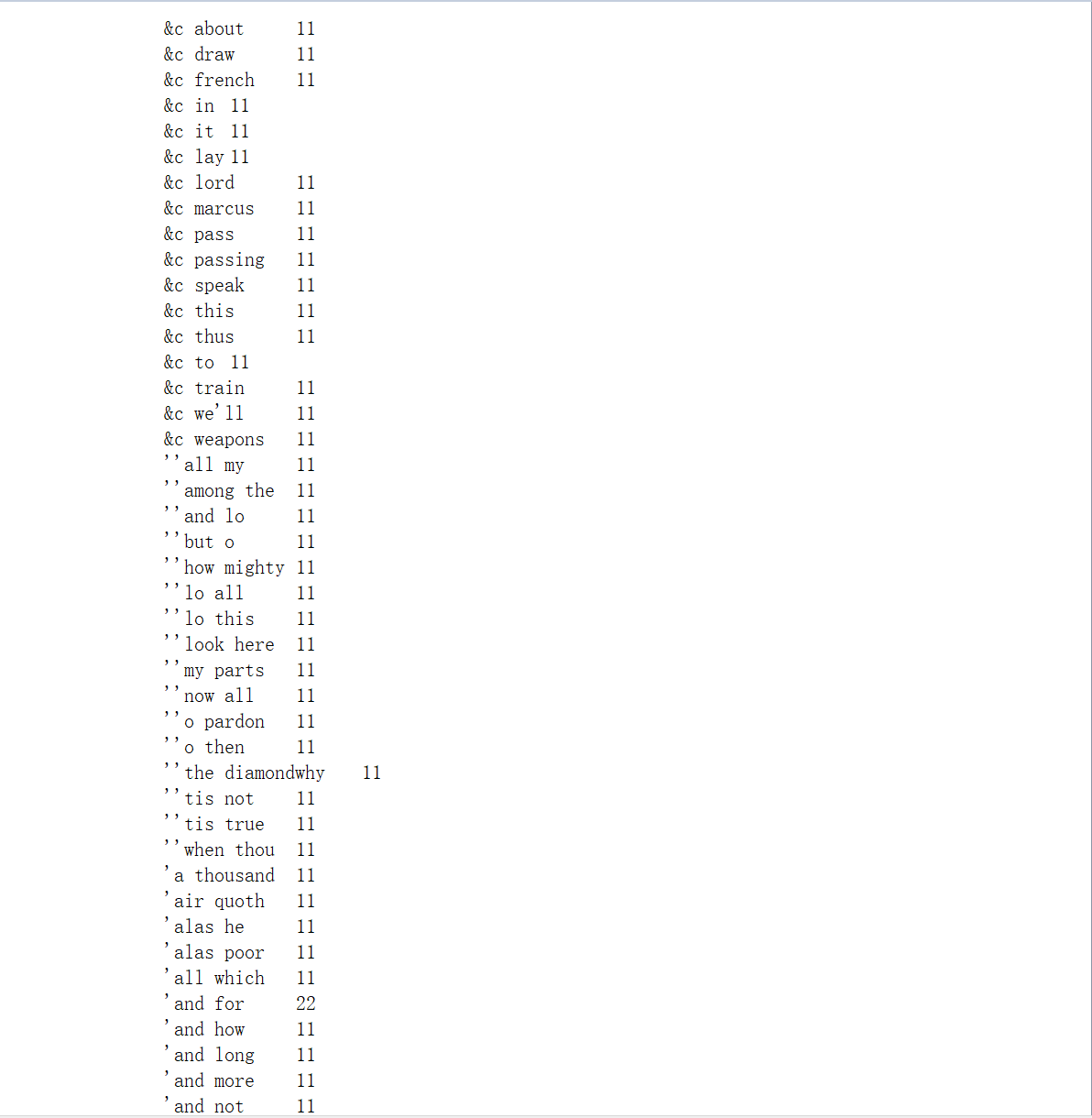


I copied the input file 11 times. So, most of the strange word have the count number 11 (one time per bible). Some of them have the number of 22. That means each word shows 2 times in one bible.



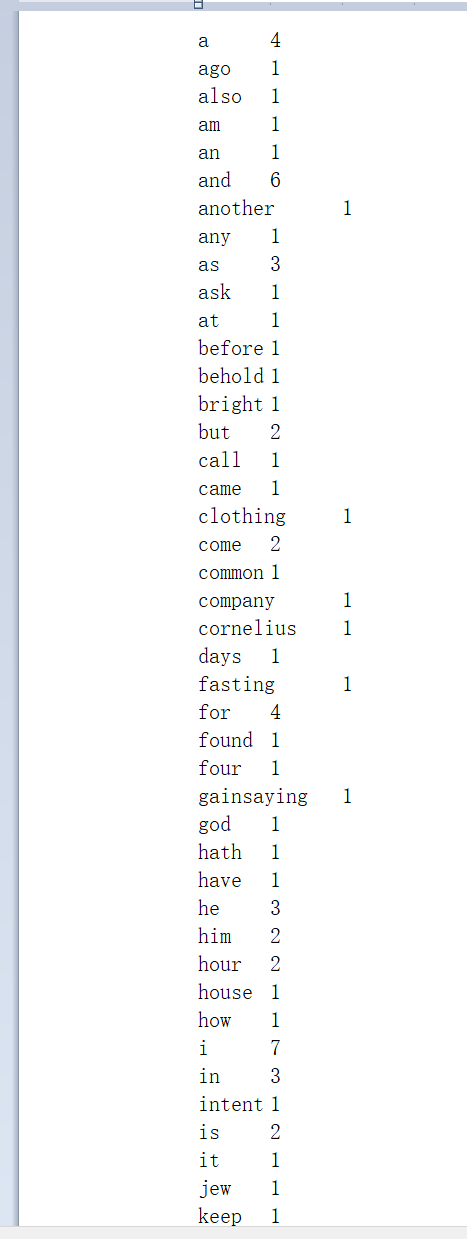
Then, let’s look at another screenshot example. In task1 we got 258544 times “a”. We will need this number in later.

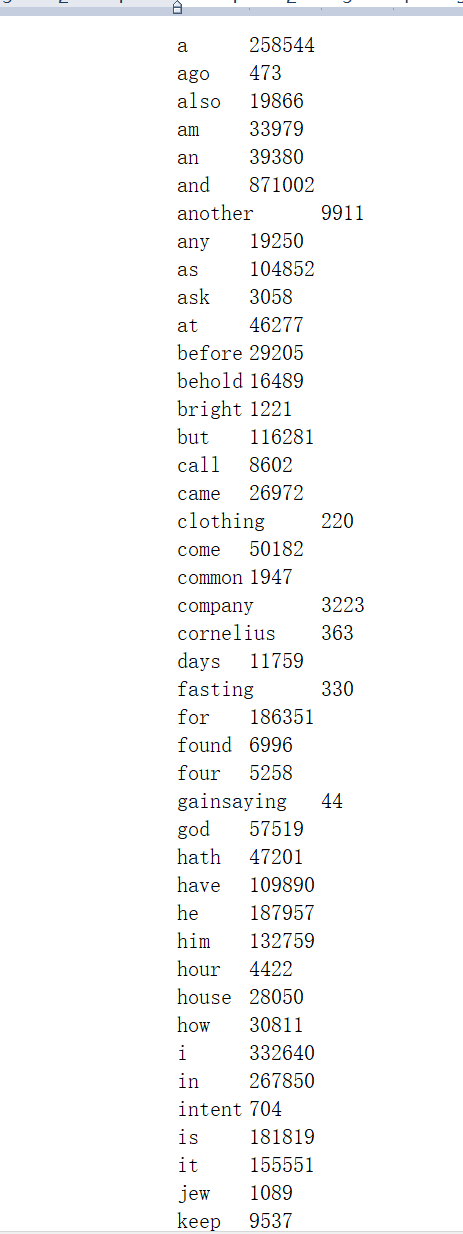
**Task 2**



We could see the number of most strange words are still 11. So, the task has been done successfully.

**Task 3**



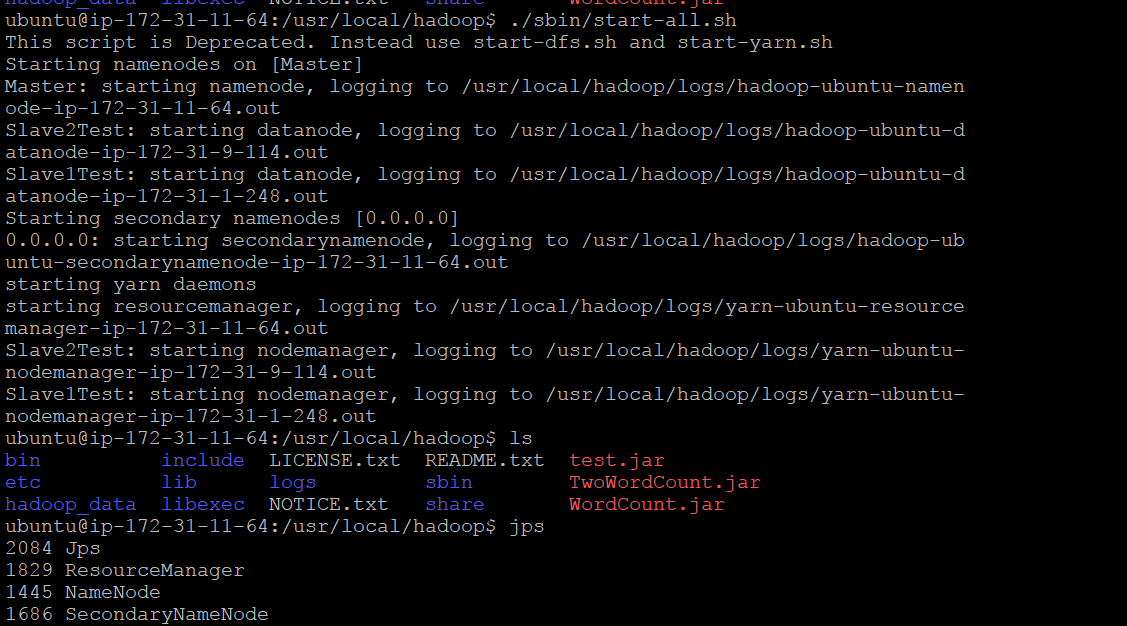


From the two screenshots of task3. We could easily find that all the words are the same. Also, we could notice that the number of “a” is 258544. And this number equals to the number in the task1. So, the results are correct.

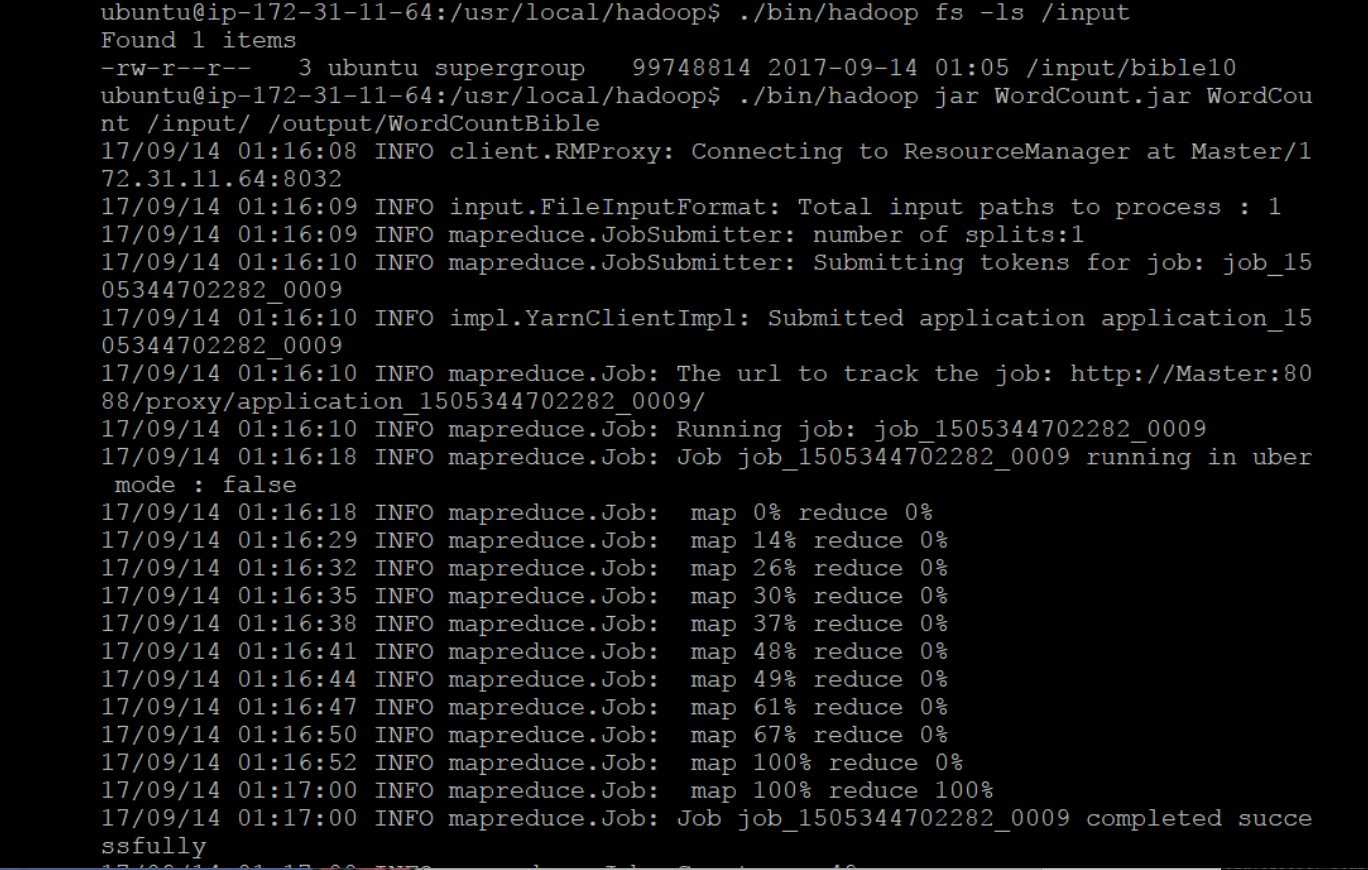
**Screenshots of the key steps on AWS**



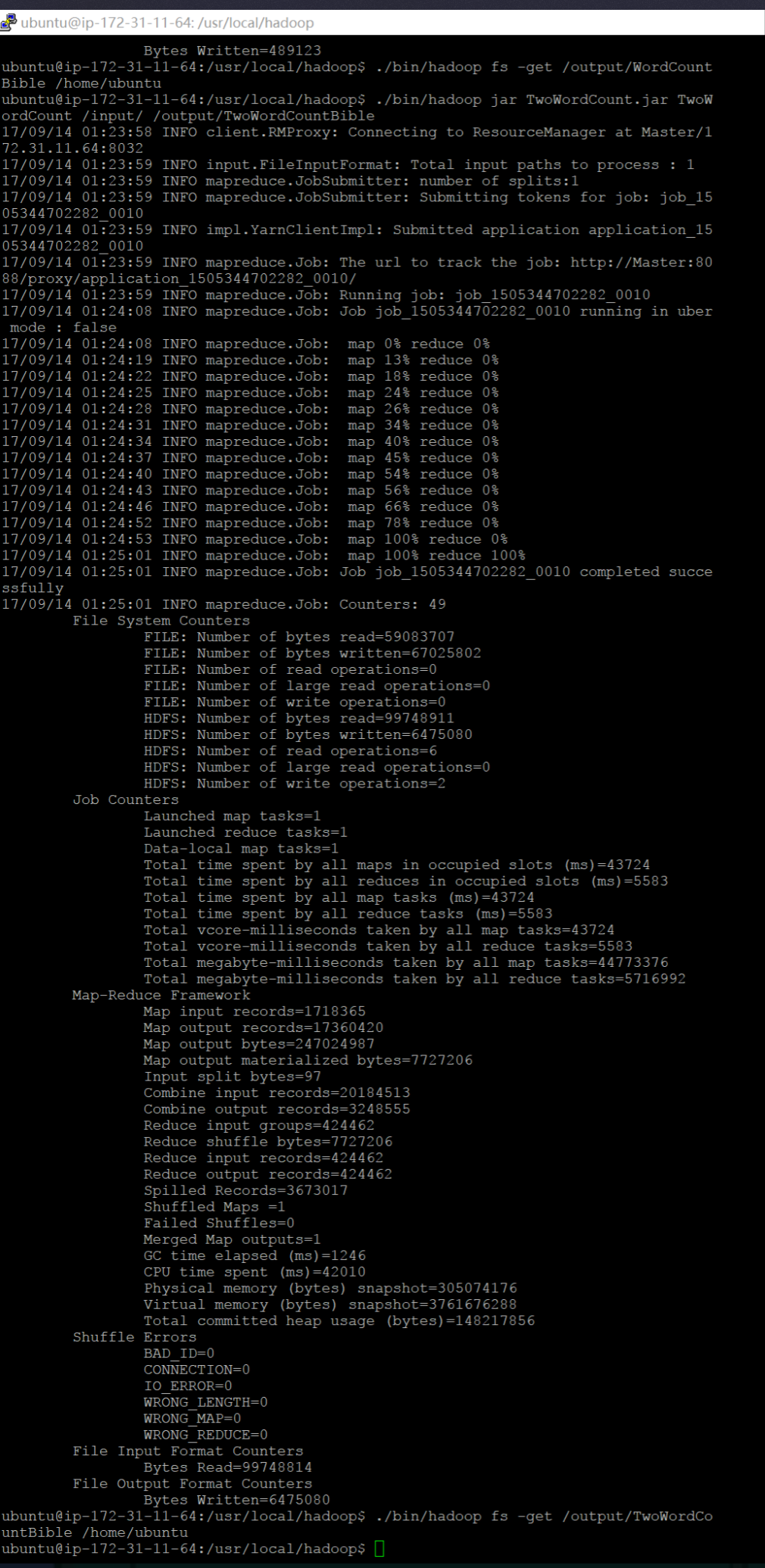
The three VMs on AWS EC2.



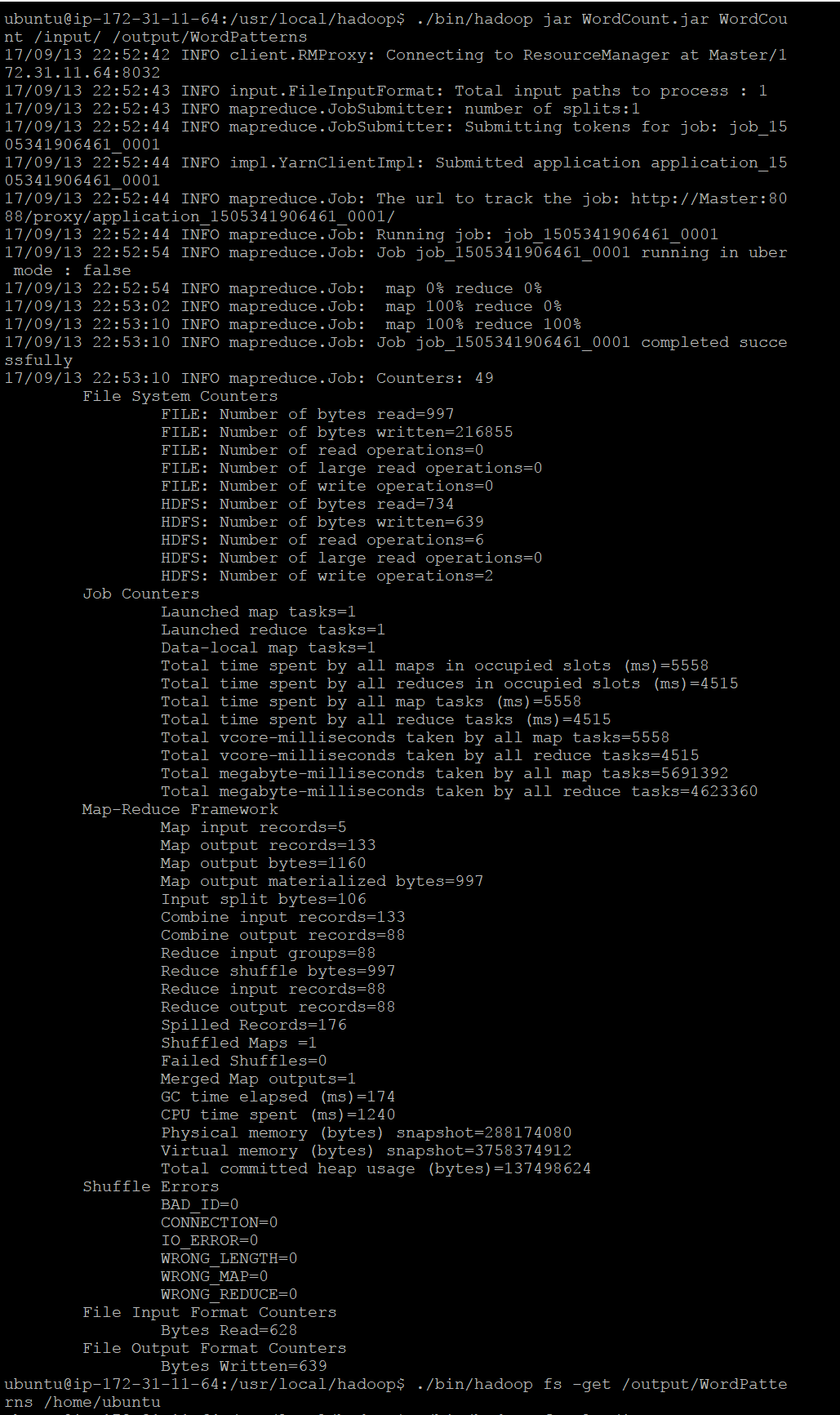
Start the Hadoop services



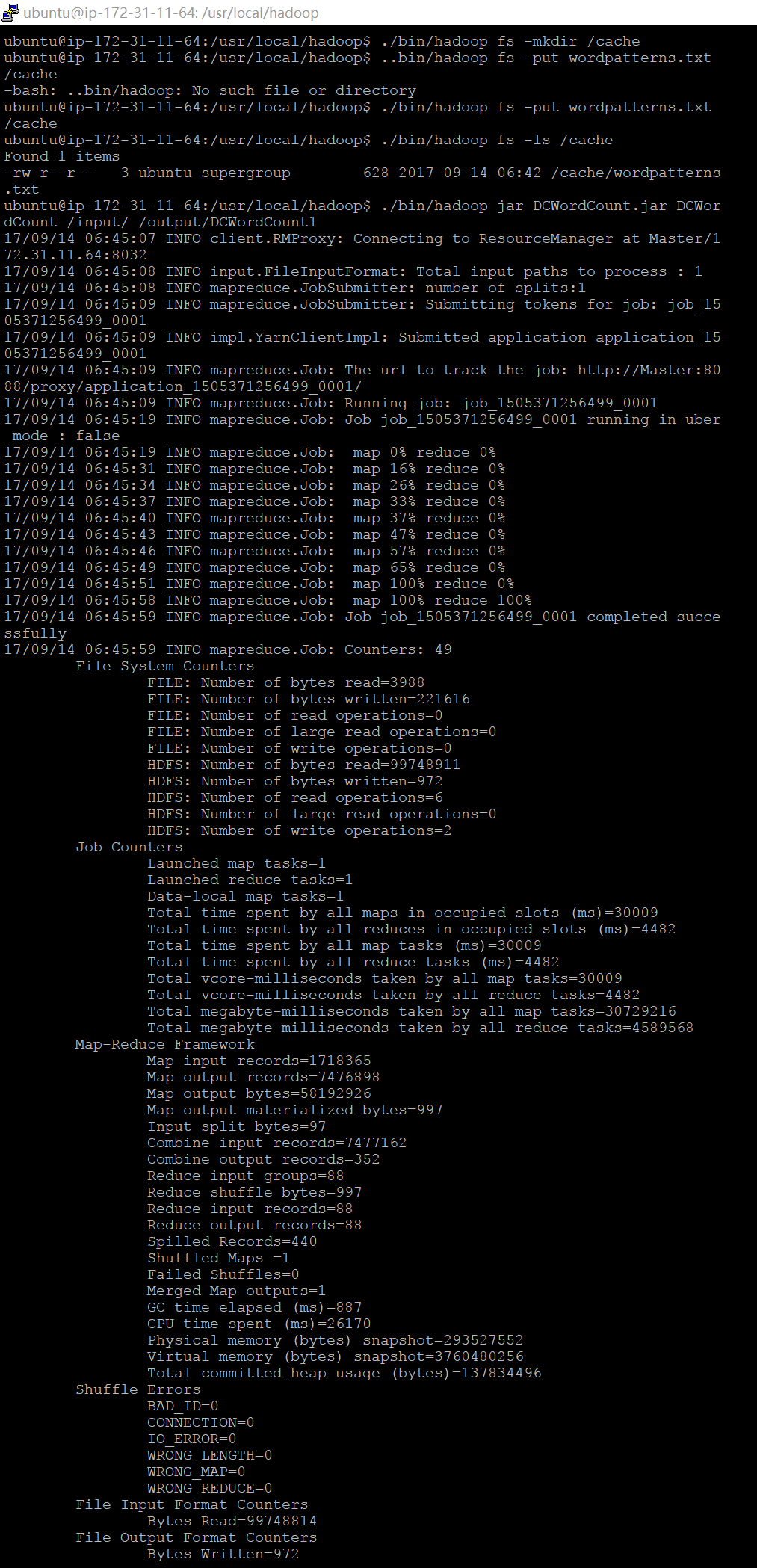
Run task1 and success



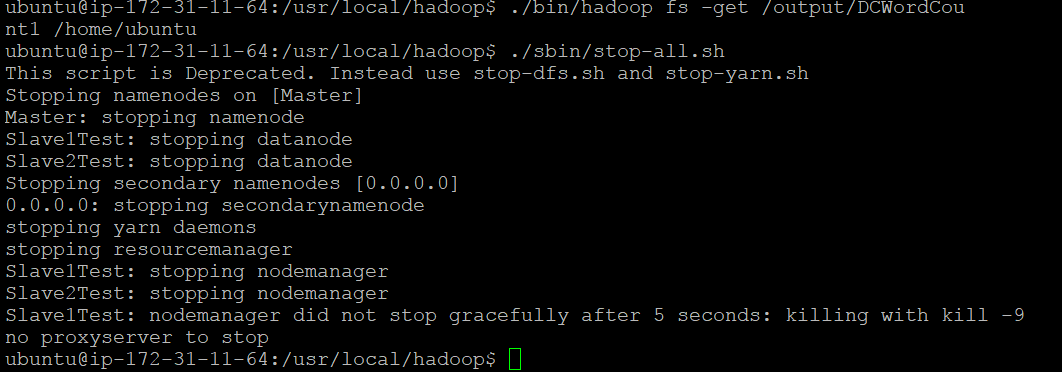
Run task2 and success



Prepare for task3, get the frequency of the small list



Send the file to the cache, run task3 and success



Last thing, get the file and shut down the Hadoop services