**IoT Lab Report 1**

**03.10.2017**

**Mert Ozer**

**Prerequisites to run the code**

1. Hadoop 2.7.4
2. Java >7
3. WordCount.java (provided in the folder)
4. Items file (provided in the folder)

**How to run step by step**

1. Open the terminal.
2. **mert@mert-Q550LF**:~/hadoop-2.7.4$ bin/hadoop com.sun.tools.javac.Main WordCount.java
3. **mert@mert-Q550LF:**~/hadoop-2.7.4$ jar cf wc.jar WordCount\*.class
4. **mert@mert-Q550LF:**~/hadoop-2.7.4$ sudo bin/hadoop jar wc.jar WordCount /user/root/input/items /user/root/output/hw
5. **mert@mert-Q550LF**:~/hadoop-2.7.4$ bin/hadoop fs -cat /user/root/output/hw/p\*

**Idea behind the code**

(Reminder) Our input looked like this.

book12, book34, cd12, cd42, dvd32

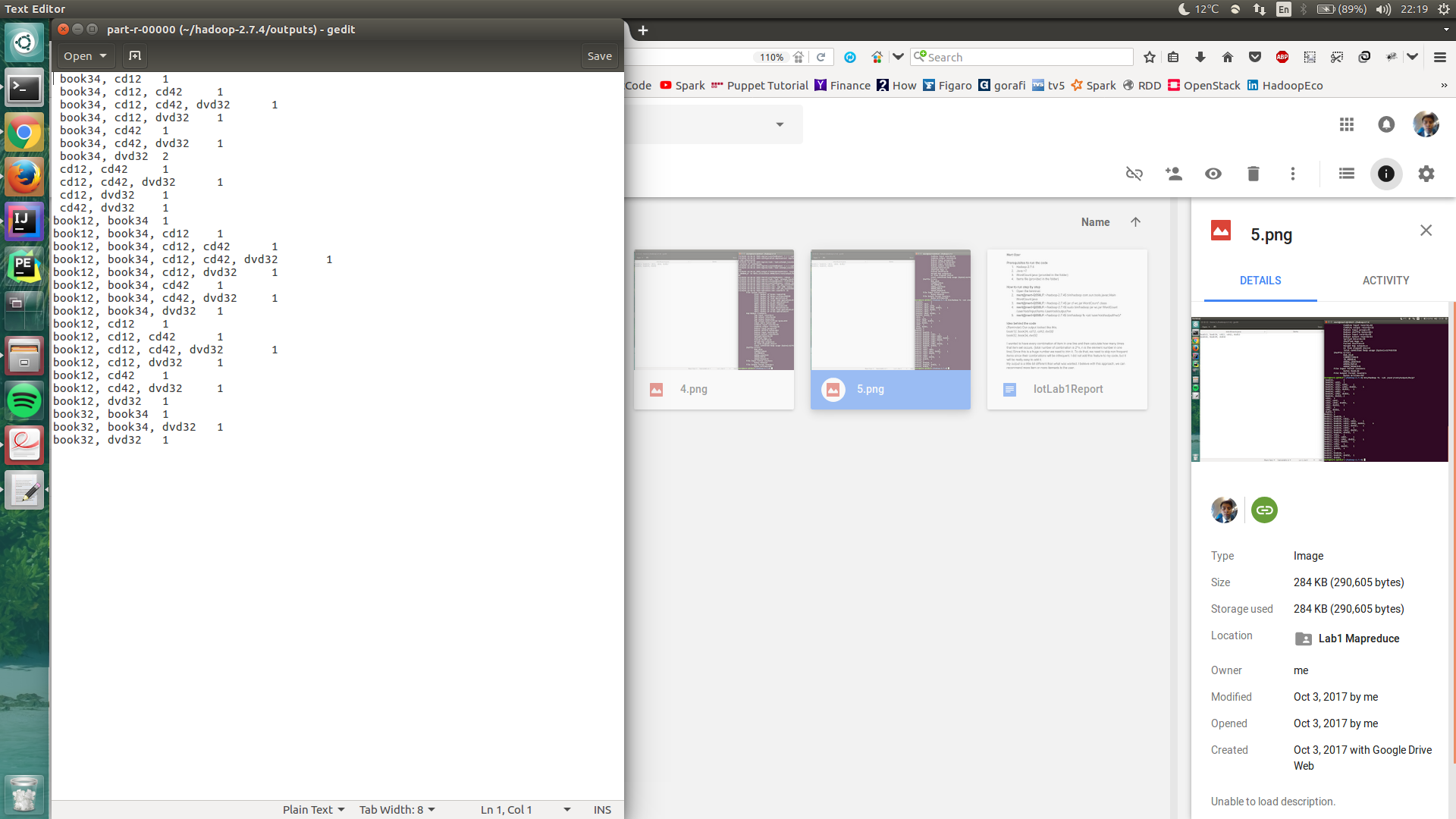
book32, book34, dvd32

I wanted to have every combination of item in one line and then calculate how many times that item set occurs. (total number of combination is 2^n, n is the element number in one line) Since this is a huge number we need to trim it. To do that, we need to skip non frequent items since their combinations will be infrequent.

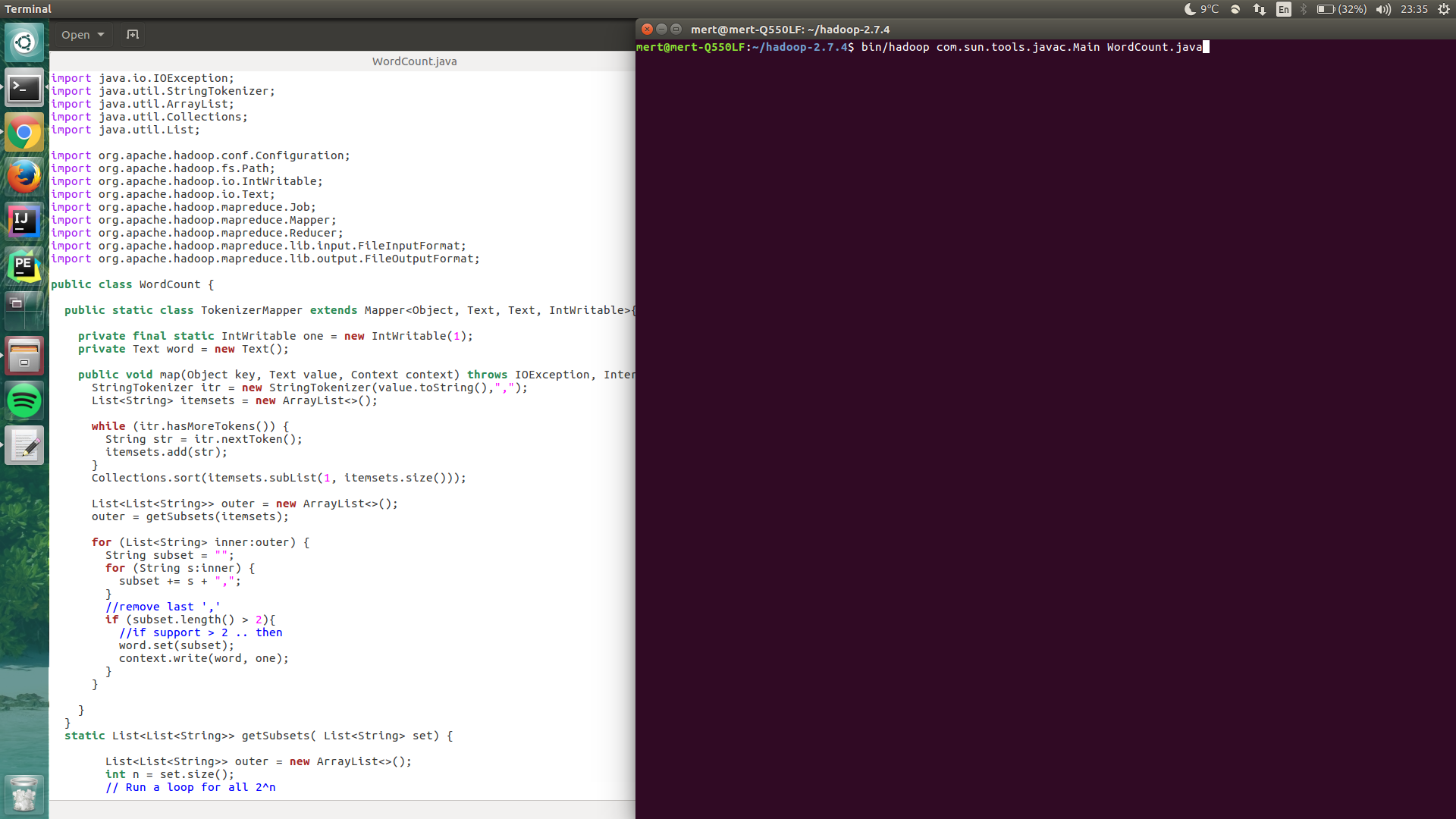
My output is a little bit different than what was wanted. Tasks output was “key-value pairs in which the key is the item and the value is the list of the items most commonly bought by customers who also bought this item”.

I believe with my approach, we can recommend more items or more itemsets to the user. Also, knowing how many times that items occurred together could be used later on.

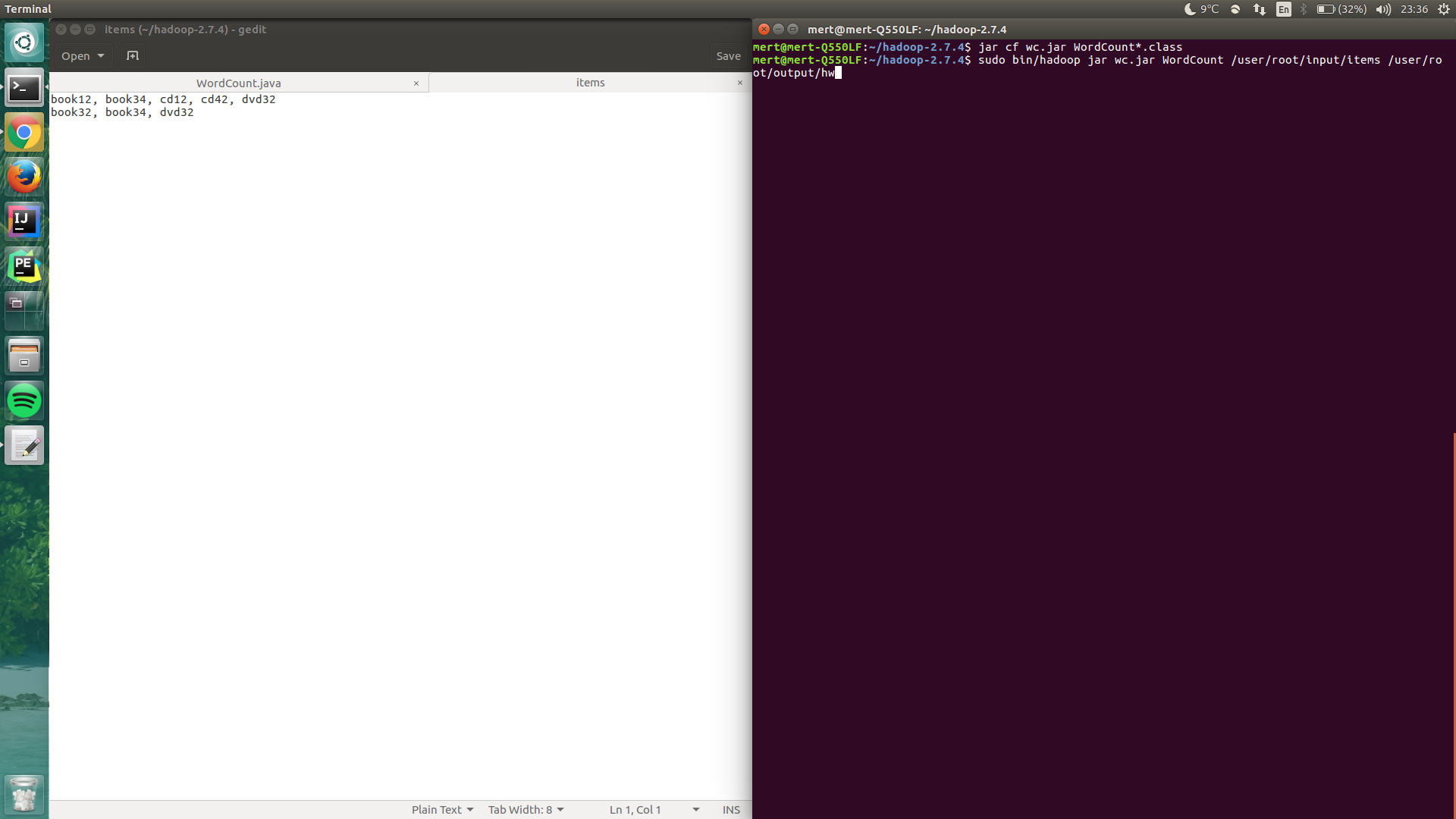
**Outputs (You can find them in the folder with the screenshots)**

****

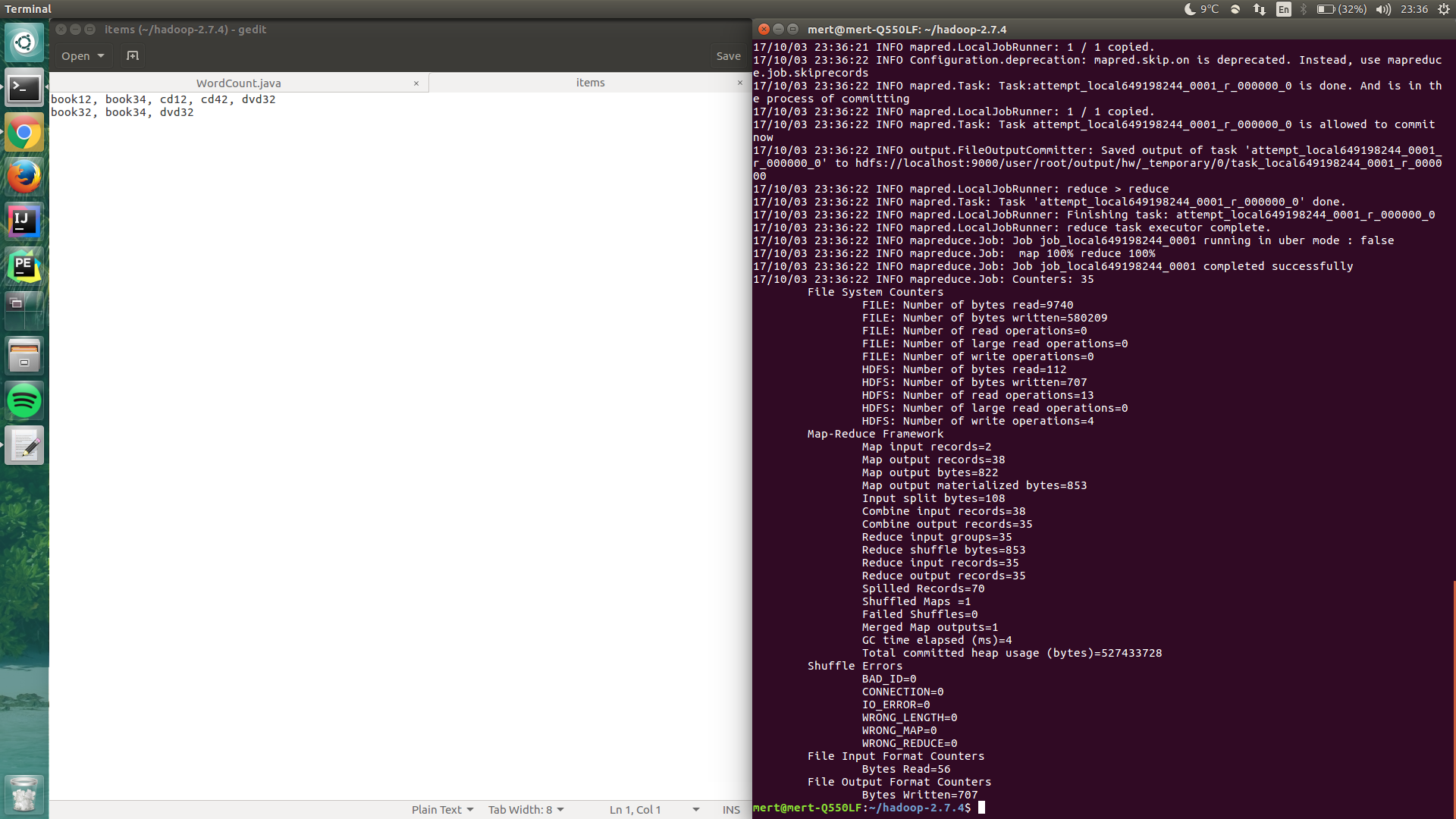
**1.png**

****

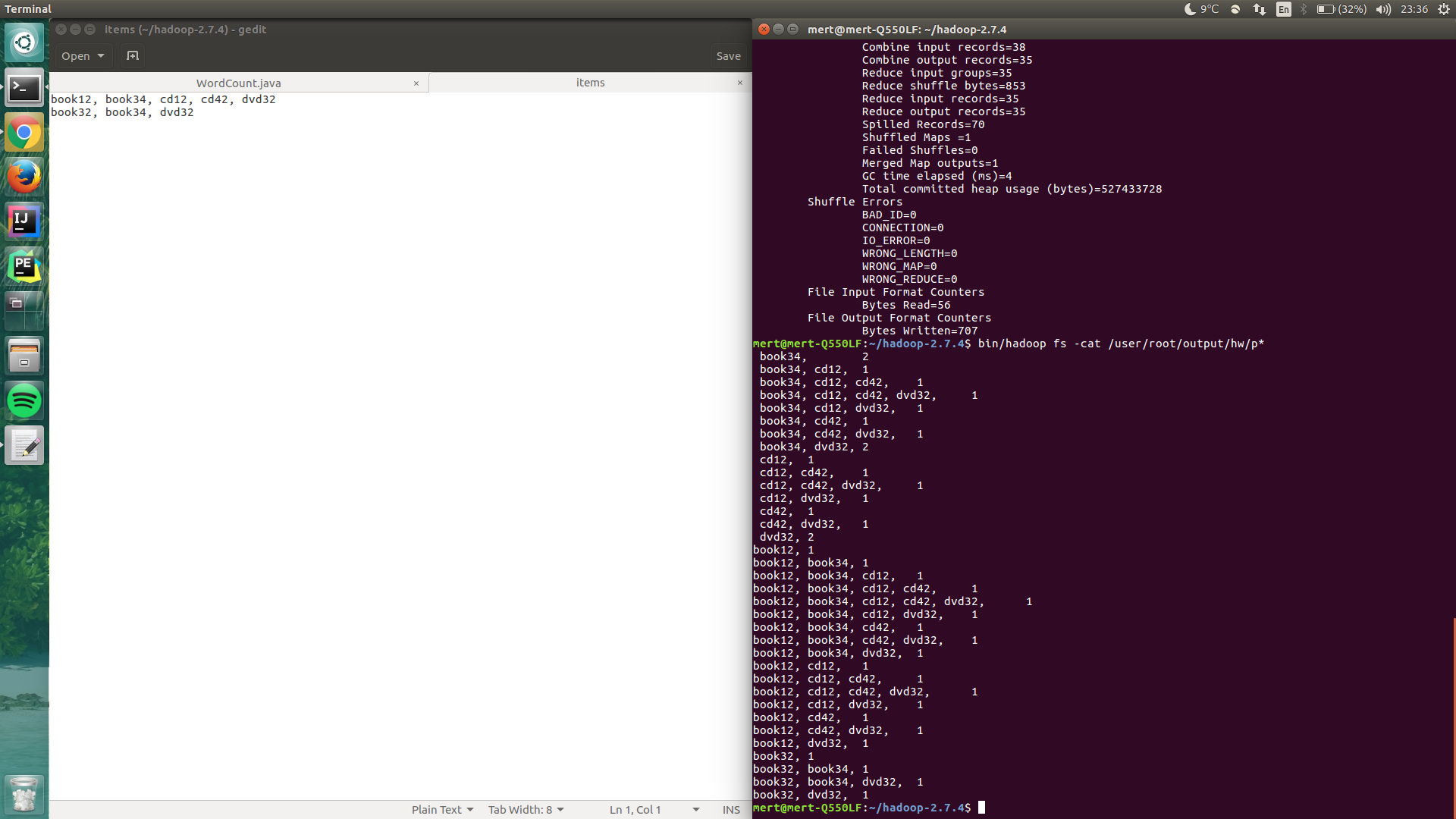
**2.png**

****

**3.png**

****

**4.png**

****

**5.png**