**Hadoop MapReduce**

Summary of the program:

I implemented the MapReduce class using mrjob a library for easier running python scripts on Hadoop clusters. The class is divided into 3 parts :

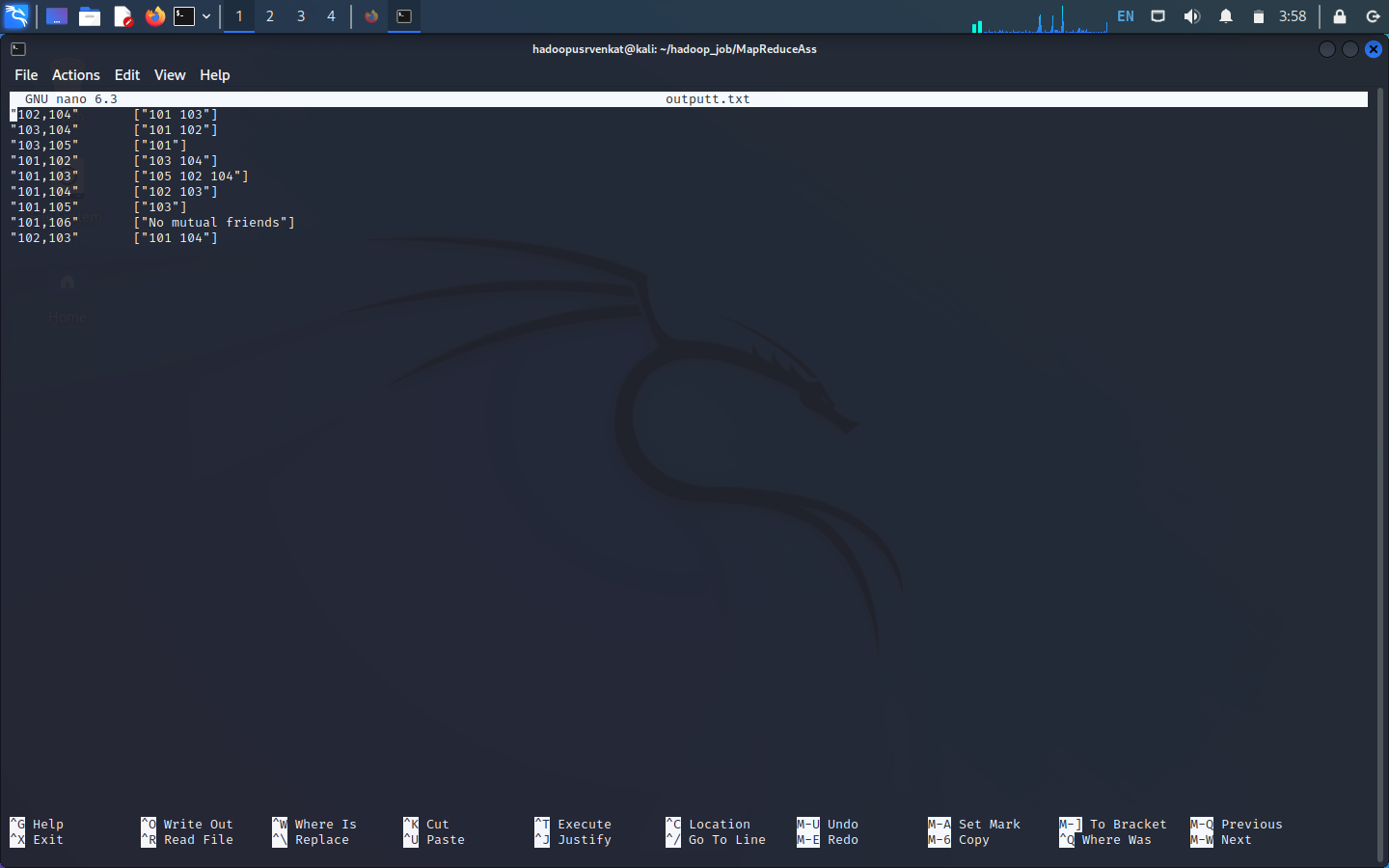
1. The steps to be taken ie the mapper function and the reducer function
2. The mapper function
   1. The mapper function takes in a line of input and splits it into user and friends.
   2. The friends are then split into a list.
   3. For each friend, a key-value pair is emitted.
   4. The key is the user and friend sorted in ascending order.
   5. The value is the line of input with the friend removed.
3. The reducer function
   1. The reducer function takes in a key and a list of values.
   2. The values are appended to a list.
   3. If there are two values, then there are two friends.
   4. The friends are split into lists.
   5. The intersection of the two lists is found.
   6. If there are mutual friends, the key value pair is emitted.
   7. If there are no mutual friends, the key value pair is emitted with a value of "No mutual friends".

Running the MapReduce mrjob on Hadoop clusters

python your\_mr\_job\_sub\_class.py -r hadoop < input > output

python com\_friends.py -r hadoop < sample.txt > output.txt

The output of the program

1. On the small dataset ie sample.txt
2. 

2.On the larger dataset ie sample2.txt

