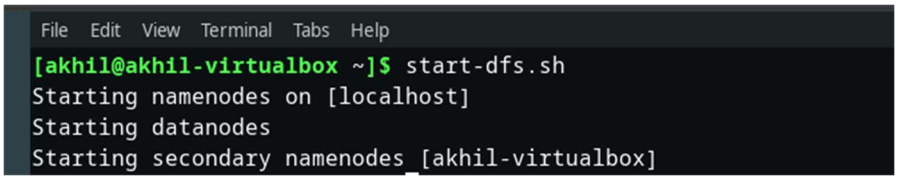
To run my program Please follow the below instructions.

1. Unzip the provided file.
2. Move the farmersMarket directory to the location you want on your Linux machine.
3. Now cd to the farmersMarket directory and open a terminal window.
4. Use **start-dfs.sh** to start Hadoop cluster. (If using Ubuntu this command might be different). Once your Hadoop cluster is ready to use.



5. Please run below commands:
   (Assuming data preprocessing is done).
   hdfs dfs -mkdir -p /user/akhil
   hdfs dfs -put fm_data.csv

   This will help us to place our input files to the Hadoop cluster.
   If your terminal current working directory is not in the homeWork3 then this might not work as expected.
6. If you have run the hdfs dfs -mkdir -p /user/akhil command, then no need to make any changes in **script.sh**. If not done, then you must update the input and output paths accordingly.
   - Please cross check Hadoop jar path provided in the script with yours and change accordingly.
7. Please check whether **script.sh** has execute rights or not if execute rights are not present please provide using chmod +x mutual_friends_script.sh .

8. Now run the script file.
   ./script.sh

   After successfully running the script you will see some thing like this

```
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=3053007
        File Output Format Counters
                Bytes Written=180728
2022-12-03 10:18:25,325 INFO streaming.StreamJob: Output directory: /user/akhil/fm
```

If not seen output like this, then something might went bad.

9. Now our output is ready we need to copy it from Hadoop cluster to our local machine and It is ready to use. To do this use below command
   hdfs dfs -get fm output

```
[akhil@akhil-virtualbox farmersMarket]$ hdfs dfs -get fm output
```

- If already have the output file in your directory, please remove it and create a new one.

Use cat output/* to see the data.

```
port townsend,Washington     1 , 1
punta gorda,Florida     1 , 33
rainer,Oregon     1 , 2
rockville centre,New York     1 , 17
roselle,Illinois     1 , 13
sauk city,Wisconsin     1 , 0
sorrento,Florida     1 , 25
spartanburg,South Carolina     2,31
stevenson,Washington     1 , 19
sun city west,Arizona     2,2
sunland park,New Mexico     1 , 29
toano,Virginia     1 , 19
tularosa,New Mexico     1 , 9
volcano,Hawaii     1 , 21
wheatland,Wyoming     1 , 12
woodstock,Illinois     2,26
```

Snap of output

That's all we are pretty much done Thanks for Reading this.

- Uploaded my output file in output.zip