1.
```
17122 NodeManager
16514 DataNode
16371 NameNode
17866 Jps
```

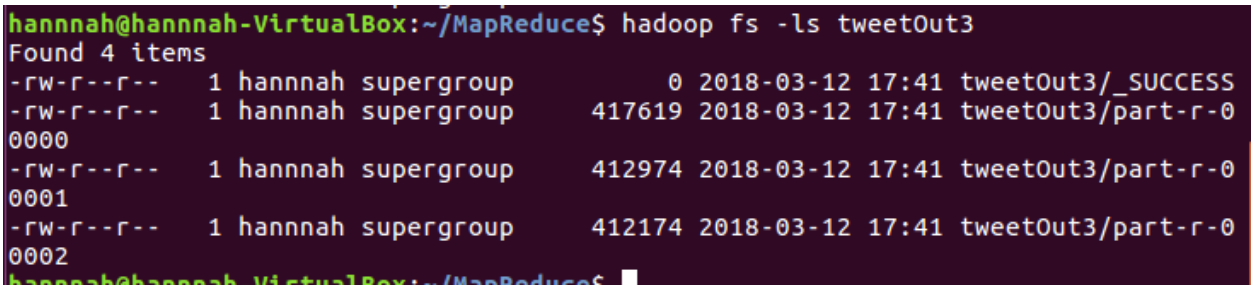
2.
```
hannnah@hannnah-VirtualBox:~$ hdfs fsck tweets.txt
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value
Connecting to namenode via http://localhost:9870/fsck?ugi=hannnah&pa
FSCK started by hannnah (auth:SIMPLE) from /127.0.0.1 for path /user
 PDT 2018

Status: HEALTHY
 Number of data-nodes:  1
 Number of racks:               1
 Total dirs:                    0
 Total symlinks:                0

Replicated Blocks:
 Total size:    482508953 B
 Total files:   1
 Total blocks (validated):      4 (avg. block size 120627238 B)
 Minimally replicated blocks:   4 (100.0 %)
 Over-replicated blocks:        0 (0.0 %)
 Under-replicated blocks:       0 (0.0 %)
 Mis-replicated blocks:         0 (0.0 %)
 Default replication factor:    1
 Average block replication:     1.0
 Missing blocks:                0
 Corrupt blocks:                0
 Missing replicas:              0 (0.0 %)
```

3. 4 blocks

```
        HDFS: Number of write oper
Job Counters
        Killed map tasks=1
        Launched map tasks=4
        Launched reduce tasks=1
        Data-local map tasks=4
```

4 block means 4 map task

4.
```
Job Counters
        Killed reduce tasks=1
        Launched map tasks=4
        Launched reduce tasks=3
        Data-local map tasks=4
```



Run using
Hadoop 3.0.0
Java 1.8.4

TO COMPILE
cd into any Question * file and execute the following commands (WHERE HADOOP_HOME is the path to your Hadoop location)

javac -classpath `$HADOOP_HOME/bin/hadoop classpath` *.java
jar cvf [CLASS NAME].jar *.class

Default .jar include with each file is

mr.jar

QUESTIONS 5 AND 6
Creates intermediate file "outputfile" which will need to be deleted between runs.

$HADOOP_HOME/bin/hadoop jar [CLASSNAME].jar Driver tweets.txt [OUTPUTFILE 1]

QUESTION 7

Creates intermediate files "tweetCountFile", "joinFile", and "countFile" which will need to be deleted between runs

$HADOOP_HOME/bin/hadoop jar [CLASSNAME].jar Driver tweets.txt users.txt [OUTPUTFILE 1]

Retrieve results with

hadoop fs -get [OUTPUTFILE 1]/part-r-00000