

```
[cloudera@quickstart FMA]$ chmod +x mapper.py reducer.py
[cloudera@quickstart FMA]$ cat Tweets.csv | ./mapper.py | sort | ./reducer.py
: No such file or directory
: No such file or directory
[cloudera@quickstart FMA]$ dos2unix mapper.py
dos2unix: converting file mapper.py to UNIX format ...
[cloudera@quickstart FMA]$ dos2unix reducer.py
dos2unix: converting file reducer.py to UNIX format ...
[cloudera@quickstart FMA]$ cat Tweets.csv | ./mapper.py | sort | ./reducer.py
Estimated distinct users: 8192
[cloudera@quickstart FMA]$ hdfs dfs -mkdir -p /user/cloudera/tweets_input
[cloudera@quickstart FMA]$ hdfs dfs -put Tweets.csv /user/cloudera/tweets_input/
[cloudera@quickstart FMA]$ hdfs dfs -rm -r /user/cloudera/tweets_output
rm: `/user/cloudera/tweets_output': No such file or directory
[cloudera@quickstart FMA]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming
.jar \
> -files mapper.py,reducer.py \
> -mapper ./mapper.py \
> -reducer ./reducer.py \
> -input /user/cloudera/tweets_input/Tweets.csv \
> -output /user/cloudera/tweets_output
```

```
virtual memory (bytes) snapshot=4700170000
Total committed heap usage (bytes)=874512384
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=3425527
File Output Format Counters
Bytes Written=32
25/09/01 04:50:23 INFO streaming.StreamJob: Output directory: /user/cloudera/twe
ets_output
[cloudera@quickstart FMA]$ hdfs dfs -cat /user/cloudera/tweets_output/part-00000
Estimated distinct users: 8192
```