```
[cloudera@quickstart Bloom_Filter]$ hdfs dfs -mkdir -p /user/cloudera/flights_in
put
[cloudera@quickstart Bloom_Filter]$ hdfs dfs -put flights.csv /user/cloudera/fli
ghts_input/
put: `/user/cloudera/flights_input/flights.csv': File exists
[cloudera@quickstart Bloom_Filter]$ hdfs dfs -ls /user/cloudera/flights_input
Found 1 items
-rw-r--r--   1 cloudera cloudera       1990 2025-09-01 05:02 /user/cloudera/flig
hts_input/flights.csv
[cloudera@quickstart Bloom_Filter]$ chmod +x mapper.py
[cloudera@quickstart Bloom_Filter]$ chmod +x reducer.py
[cloudera@quickstart Bloom_Filter]$ dos2unix mapper.py
dos2unix: converting file mapper.py to UNIX format ...
[cloudera@quickstart Bloom_Filter]$ dos2unix reducer.py
dos2unix: converting file reducer.py to UNIX format ...
[cloudera@quickstart Bloom_Filter]$ cat flights.csv | ./mapper.py | sort | ./red
ucer.py > bloom_output.txt
[cloudera@quickstart Bloom_Filter]$ head bloom_output.txt
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000001000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
01000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000001000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000001000
00000000010000000000000010000000000000000000000000000000000000000000000000000000
00000000000000000100010000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000
```

```
[cloudera@quickstart Bloom_Filter]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-
streaming.jar \
>   -files mapper.py,reducer.py \
>   -mapper ./mapper.py \
>   -reducer ./reducer.py \
>   -input /user/cloudera/flights_input/flights.csv \
>   -output /user/cloudera/flights_output
packageJobJar: [] [/usr/lib/hadoop-mapreduce/hadoop-streaming-2.6.0-cdh5.13.0.ja
r] /tmp/streamjob1554896106606651814.jar tmpDir=null
25/09/01 05:06:59 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0
:8032
25/09/01 05:07:00 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0
:8032
```

```
                Bytes Written=1002
25/09/01 05:10:32 INFO streaming.StreamJob: Output directory: /user/cloudera/flights_output
[cloudera@quickstart Bloom_Filter]$ hdfs dfs -cat /user/cloudera/flights_output/part-00000 | head -20
00000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000100
00000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000010000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000
```