

```
[cloudera@quickstart bloom]$ chmod +x mapper.py reducer.py
[cloudera@quickstart bloom]$ dos2unix mapper.py reducer.py
dos2unix: converting file mapper.py to UNIX format ...
dos2unix: converting file reducer.py to UNIX format ...
[cloudera@quickstart bloom]$ cat weblog.csv | head -50 | python mapper.py | pyth
on reducer.py
Invalid IP format: IP
Duplicate detected: 10.128.2.1
Duplicate detected: 10.128.2.1
Duplicate detected: 10.131.2.1
Duplicate detected: 10.131.2.1
Duplicate detected: 10.131.2.1
Duplicate detected: 10.130.2.1
Duplicate detected: 10.130.2.1
Duplicate detected: 10.130.2.1
Duplicate detected: 10.130.2.1
Duplicate detected: 10.130.2.1
Duplicate detected: 10.131.2.1
Duplicate detected: 10.131.2.1
Duplicate detected: 10.128.2.1
Duplicate detected: 10.131.0.1
Duplicate detected: 10.130.2.1
```

```
25/09/01 21:29:27 INFO mapreduce.Job: Total unique IP records: 10
[cloudera@quickstart bloom]$ hadoop fs -mkdir -p /user/cloudera/input
[cloudera@quickstart bloom]$ hadoop fs -mkdir -p /user/cloudera/output
[cloudera@quickstart bloom]$ hadoop fs -put weblog.csv /user/cloudera/input/
[cloudera@quickstart bloom]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streami
ng.jar \
> -files mapper.py,reducer.py \
> -mapper ./mapper.py \
> -reducer ./reducer.py \
> -input /user/cloudera/input/weblog.csv \
> -output /user/cloudera/output/weblog_unique
packageJobJar: [] [/usr/jars/hadoop-streaming-2.6.0-cdh5.4.2.jar] /tmp/streamjob
5937629864985890142.jar tmpDir=null
```

```
25/09/01 21:29:27 INFO mapreduce.Job: map 0% reduce 0%
25/09/01 21:29:33 INFO mapreduce.Job: map 50% reduce 0%
25/09/01 21:29:35 INFO mapreduce.Job: map 100% reduce 0%
25/09/01 21:29:38 INFO mapreduce.Job: map 100% reduce 100%
25/09/01 21:29:39 INFO mapreduce.Job: Job job_1756783182874_0004 completed succe
```

```
25/09/01 21:29:39 INFO streaming.StreamJob: Output directory: /user/cloudera/output/weblog_unique
[cloudera@quickstart bloom]$ hadoop fs -cat /user/cloudera/output/weblog_unique/part-00000 | head -10
10.128.2.1,[25/Nov/2017:15:58:03,GET /home.php HTTP/1.1,200
10.128.2.1,[25/Nov/2017:15:57:16,POST /compile.php HTTP/1.1,200
10.128.2.1,[29/Nov/2017:06:59:02,POST /process.php HTTP/1.1,302
10.128.2.1,[29/Nov/2017:06:58:55,GET /login.php HTTP/1.1,200
10.129.2.1,[29/Nov/2017:13:38:20,GET /css/main.css HTTP/1.1,200
10.129.2.1,[29/Nov/2017:13:38:03,POST /process.php HTTP/1.1,302
10.129.2.1,[25/Nov/2017:16:47:37,POST /compile.php HTTP/1.1,200
10.129.2.1,[25/Nov/2017:16:34:18,GET / HTTP/1.1,200
10.130.2.1,[25/Nov/2017:17:23:37,GET /home.php HTTP/1.1,200
10.130.2.1,[25/Nov/2017:16:03:11,POST /compile.php HTTP/1.1,200
[cloudera@quickstart bloom]$ █
```