

```

[root@sandbox ~]# hadoop jar /usr/hdp/2.4.0.0-169/hadoop-mapreduce/hadoop-streaming-2.7.1.2.4.0.0-169.jar -file /root/lab/wc_mapper-2.py -mapper wc_mapper-2.py -file /root/lab/wc_reducer-2.py -reducer wc_reducer-2.py -input /user/root/shakespeare_100.txt -output /user/root/word_streaming_out
WARNING: Use "yarn jar" to launch YARN applications.
16/10/04 16:36:07 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
packageJobJar: [/root/lab/wc_mapper-2.py, /root/lab/wc_reducer-2.py] [/usr/hdp/2.4.0.0-169/hadoop-mapreduce/hadoop-streaming-2.7.1.2.4.0.0-169.jar] /tmp/streamjob2661142778205118858.jar tmpDir=null
16/10/04 16:36:10 INFO impl.TimelineClientImpl: Timeline service address: http://sandbox.hortonworks.com:8188/ws/v1/timeline/
16/10/04 16:36:10 INFO client.RMPProxy: Connecting to ResourceManager at sandbox.hortonworks.com/10.0.2.15:8050
16/10/04 16:36:12 INFO impl.TimelineClientImpl: Timeline service address: http://sandbox.hortonworks.com:8188/ws/v1/timeline/
16/10/04 16:36:12 INFO client.RMPProxy: Connecting to ResourceManager at sandbox.hortonworks.com/10.0.2.15:8050
16/10/04 16:36:12 INFO mapred.FileInputFormat: Total input paths to process : 1
16/10/04 16:36:12 INFO mapreduce.JobSubmitter: number of splits:2
16/10/04 16:36:13 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1475591931336_0006
16/10/04 16:36:13 INFO impl.YarnClientImpl: Submitted application application_1475591931336_0006
16/10/04 16:36:14 INFO mapreduce.Job: The url to track the job: http://sandbox.hortonworks.com:8088/proxy/application_1475591931336_0006/
16/10/04 16:36:14 INFO mapreduce.Job: Running job: job_1475591931336_0006
16/10/04 16:36:24 INFO mapreduce.Job: Job job_1475591931336_0006 running in uber mode : false
16/10/04 16:36:24 INFO mapreduce.Job: map 0% reduce 0%
16/10/04 16:36:37 INFO mapreduce.Job: map 100% reduce 0%
16/10/04 16:36:53 INFO mapreduce.Job: map 100% reduce 100%
16/10/04 16:36:54 INFO mapreduce.Job: Job job_1475591931336_0006 completed successfully
16/10/04 16:36:54 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=8574785
    FILE: Number of bytes written=17556948
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=5721223
    HDFS: Number of bytes written=721004
    HDFS: Number of read operations=9
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=2
    Launched reduce tasks=1
    Data-local map tasks=2
    Total time spent by all maps in occupied slots (ms)=21940
    Total time spent by all reduces in occupied slots (ms)=13628
    Total time spent by all map tasks (ms)=21940
    Total time spent by all reduce tasks (ms)=13628
    Total vcore-seconds taken by all map tasks=21940
    Total vcore-seconds taken by all reduce tasks=13628
    Total megabyte-seconds taken by all map tasks=5485000
    Total megabyte-seconds taken by all reduce tasks=3407000
  Map-Reduce Framework
    Map input records=124788
    Map output records=904065
    Map output bytes=6766649
    Map output materialized bytes=8574791
    Input split bytes=234
    Combine input records=0
    Combine output records=0
    Reduce input groups=67782

```

```
[root@sandbox ~]# hadoop fs -cat /user/root/word_streaming_out/part* | sort -n -k2 -r | head -n20
```

the	23407
I	19540
and	18358
to	15682
of	15649
a	12586
my	10825
in	9633
you	9129
is	7874
that	7543
[And	7068
not	6967
[with	6771
his	6218
be	6017
your	6016
for	5629
have	5236
[it	4920

—