

Running MR2WordCount Job on Command Line (Linux Terminal Window)

```
[cloudera@quickstart target]$ hadoop jar practice_2-1.0.0-SNAPSHOT.jar \  
> wordcount.MR2WordCount \  
> /user/cloudera/shakespeare/histories?* /user/cloudera/practice2_mr2out  
  
16/11/01 16:32:01 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032  
  
16/11/01 16:32:03 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not  
performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.  
  
16/11/01 16:32:03 INFO mapreduce.JobSubmitter: Cleaning up the staging area /tmp/hadoop-  
yarn/staging/cloudera/.staging/job_1478022784726_0001  
  
16/11/01 16:32:03 WARN security.UserGroupInformation: PrivilegedActionException as:cloudera  
(auth:SIMPLE) cause:org.apache.hadoop.mapreduce.lib.input.InvalidInputException: Input Pattern  
hdfs://quickstart.cloudera:8020/user/cloudera/shakespeare/histories?* matches 0 files  
  
Exception in thread "main" org.apache.hadoop.mapreduce.lib.input.InvalidInputException: Input  
Pattern hdfs://quickstart.cloudera:8020/user/cloudera/shakespeare/histories?* matches 0 files  
  
    at  
org.apache.hadoop.mapreduce.lib.input.FileInputFormat.singleThreadedListStatus(FileInputFormat.java  
:323)  
  
    at org.apache.hadoop.mapreduce.lib.input.FileInputFormat.listStatus(FileInputFormat.java:265)  
    at org.apache.hadoop.mapreduce.lib.input.FileInputFormat.getSplits(FileInputFormat.java:387)  
    at org.apache.hadoop.mapreduce.JobSubmitter.writeNewSplits(JobSubmitter.java:305)  
    at org.apache.hadoop.mapreduce.JobSubmitter.writeSplits(JobSubmitter.java:322)  
    at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:200)  
    at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1307)  
    at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1304)  
    at java.security.AccessController.doPrivileged(Native Method)  
    at javax.security.auth.Subject.doAs(Subject.java:422)  
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1693)  
    at org.apache.hadoop.mapreduce.Job.submit(Job.java:1304)  
    at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1325)  
    at wordcount.MR2WordCount.main(MR2WordCount.java:128)  
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
```

```
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
[cloudera@quickstart target]$ pwd
/home/cloudera/workspace/practice_2/target
[cloudera@quickstart target]$ hadoop jar practice_2-1.0.0-SNAPSHOT.jar \
> wordcount.MR2WordCount \
> /user/cloudera/shakespeare/histories/* /user/cloudera/practice2_mr2out
16/11/01 17:38:38 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
16/11/01 17:38:39 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not
performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
16/11/01 17:38:40 INFO input.FileInputFormat: Total input paths to process : 10
16/11/01 17:38:40 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1249)
    at java.lang.Thread.join(Thread.java:1323)
    at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:862)
    at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:600)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:789)
16/11/01 17:38:40 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1249)
    at java.lang.Thread.join(Thread.java:1323)
```

at
org.apache.hadoop.hdfs.DFSOutputStream\$DataStreamer.closeResponder(DFSOutputStream.java:862)

at
org.apache.hadoop.hdfs.DFSOutputStream\$DataStreamer.endBlock(DFSOutputStream.java:600)

at org.apache.hadoop.hdfs.DFSOutputStream\$DataStreamer.run(DFSOutputStream.java:789)

16/11/01 17:38:40 INFO mapreduce.JobSubmitter: number of splits:10

16/11/01 17:38:40 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1478022784726_0002

16/11/01 17:38:41 INFO impl.YarnClientImpl: Submitted application application_1478022784726_0002

16/11/01 17:38:41 INFO mapreduce.Job: The url to track the job:
http://quickstart.cloudera:8088/proxy/application_1478022784726_0002/

16/11/01 17:38:41 INFO mapreduce.Job: Running job: job_1478022784726_0002

16/11/01 17:38:54 INFO mapreduce.Job: Job job_1478022784726_0002 running in uber mode : false

16/11/01 17:38:54 INFO mapreduce.Job: map 0% reduce 0%

16/11/01 17:39:26 INFO mapreduce.Job: map 7% reduce 0%

16/11/01 17:39:28 INFO mapreduce.Job: map 10% reduce 0%

16/11/01 17:39:30 INFO mapreduce.Job: map 20% reduce 0%

16/11/01 17:39:45 INFO mapreduce.Job: map 60% reduce 0%

16/11/01 17:40:03 INFO mapreduce.Job: map 80% reduce 0%

16/11/01 17:40:15 INFO mapreduce.Job: map 100% reduce 3%

16/11/01 17:40:19 INFO mapreduce.Job: map 100% reduce 8%

16/11/01 17:40:21 INFO mapreduce.Job: map 100% reduce 12%

16/11/01 17:40:29 INFO mapreduce.Job: map 100% reduce 18%

16/11/01 17:40:30 INFO mapreduce.Job: map 100% reduce 24%

16/11/01 17:40:46 INFO mapreduce.Job: map 100% reduce 35%

16/11/01 17:40:52 INFO mapreduce.Job: map 100% reduce 47%

16/11/01 17:41:00 INFO mapreduce.Job: map 100% reduce 53%

16/11/01 17:41:02 INFO mapreduce.Job: map 100% reduce 59%

16/11/01 17:41:16 INFO mapreduce.Job: map 100% reduce 65%

16/11/01 17:41:18 INFO mapreduce.Job: map 100% reduce 71%

16/11/01 17:41:23 INFO mapreduce.Job: map 100% reduce 76%
16/11/01 17:41:24 INFO mapreduce.Job: map 100% reduce 82%
16/11/01 17:41:28 INFO mapreduce.Job: map 100% reduce 88%
16/11/01 17:41:29 INFO mapreduce.Job: map 100% reduce 94%
16/11/01 17:41:31 INFO mapreduce.Job: map 100% reduce 100%
16/11/01 17:41:32 INFO mapreduce.Job: Job job_1478022784726_0002 completed successfully
16/11/01 17:41:32 INFO mapreduce.Job: Counters: 51

File System Counters

FILE: Number of bytes read=2995429
FILE: Number of bytes written=9136192
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=1480383
HDFS: Number of bytes written=151717
HDFS: Number of read operations=81
HDFS: Number of large read operations=0
HDFS: Number of write operations=34

Job Counters

Killed map tasks=2
Killed reduce tasks=1
Launched map tasks=10
Launched reduce tasks=18
Data-local map tasks=10
Total time spent by all maps in occupied slots (ms)=364752
Total time spent by all reduces in occupied slots (ms)=492506
Total time spent by all map tasks (ms)=364752
Total time spent by all reduce tasks (ms)=492506
Total vcore-seconds taken by all map tasks=364752

Total vcore-seconds taken by all reduce tasks=492506

Total megabyte-seconds taken by all map tasks=373506048

Total megabyte-seconds taken by all reduce tasks=504326144

Map-Reduce Framework

Map input records=45989

Map output records=268574

Map output bytes=2458179

Map output materialized bytes=2996347

Input split bytes=1448

Combine input records=0

Combine output records=0

Reduce input groups=15307

Reduce shuffle bytes=2996347

Reduce input records=268574

Reduce output records=15307

Spilled Records=537148

Shuffled Maps =170

Failed Shuffles=0

Merged Map outputs=170

GC time elapsed (ms)=5481

CPU time spent (ms)=31380

Physical memory (bytes) snapshot=4141338624

Virtual memory (bytes) snapshot=40621019136

Total committed heap usage (bytes)=2688917504

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=1478935

File Output Format Counters

Bytes Written=151717

[cloudera@quickstart target]\$