

```
"Glamis 1
"God 2
"Good 1
"Havoc!" 1
"He 1
"Help 1
"Help, 2
"Here 1
"Hold, 2
"I 4
"Indeed!" 1
"King 1
"Liberty, 1
"Lo, 1
"Long 1
"Murder!" 2
"Neither 1
"Now 1
"O 2
"Peace, 1
"Pro- 1
"Project 1
"Right 1
"Shall 1
"Sing 2
"Sir, 1
"Sleep 2
"Small 2
"Speak, 1
"Sweet 1
"That 1
"The 1
"These 1
"They 2
"This 2
"Thus 2
"Tis 2
"Where 1
"Willow, 1
"You'll 1
"better"? 1
"hem," 1
"never." 1
"not" 1
"small 1
"then" 1
"thrusting" 1
"thy 1
"twas 1
```

```
Map output bytes=8546434
Map output materialized bytes=978916
Input split bytes=120
Combine input records=901325
Combine output records=67505
Reduce input groups=67505
Reduce shuffle bytes=978916
Reduce input records=67505
Reduce output records=67505
Spilled Records=135010
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=126
CPU time spent (ms)=3430
Physical memory (bytes) snapshot=359378944
Virtual memory (bytes) snapshot=3007582208
Total committed heap usage (bytes)=226365440

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=5458199
File Output Format Counters
  Bytes Written=717768
[cloudera@quickstart ~]$ hadoop fs -ls
Found 2 items
drwxr-xr-x  - cloudera cloudera      0 2020-09-10 02:23 out
-rw-r--r--  1 cloudera cloudera  5458199 2020-09-10 02:15 words.txt
[cloudera@quickstart ~]$ hadoop fs ls out
ls: Unknown command
Did you mean -ls? This command begins with a dash.
[cloudera@quickstart ~]$ clear

[cloudera@quickstart ~]$ hadoop fs -ls
Found 2 items
drwxr-xr-x  - cloudera cloudera      0 2020-09-10 02:23 out
-rw-r--r--  1 cloudera cloudera  5458199 2020-09-10 02:15 words.txt
[cloudera@quickstart ~]$ hadoop fs -ls out
Found 2 items
-rw-r--r--  1 cloudera cloudera      0 2020-09-10 02:23 out/_SUCCESS
-rw-r--r--  1 cloudera cloudera  717768 2020-09-10 02:23 out/part-r-00000
[cloudera@quickstart ~]$ hadoop fs -copyToLocal out/part-r-00000 local
[cloudera@quickstart ~]$ more local
"
  241
```

distbbp: A map/reduce program that uses a BBP-type formula to compute exact bits of Pi.
grep: A map/reduce program that counts the matches of a regex in the input.
join: A job that effects a join over sorted, equally partitioned datasets
multifilewc: A job that counts words from several files.
pentomino: A map/reduce tile laying program to find solutions to pentomino problems.
pi: A map/reduce program that estimates Pi using a quasi-Monte Carlo method.
randomtextwriter: A map/reduce program that writes 10GB of random textual data per node.
randomwriter: A map/reduce program that writes 10GB of random data per node.
secondarysort: An example defining a secondary sort to the reduce.
sort: A map/reduce program that sorts the data written by the random writer.
sudoku: A sudoku solver.
teragen: Generate data for the terasort
terasort: Run the terasort
teravalidate: Checking results of terasort
wordcount: A map/reduce program that counts the words in the input files.
wordmean: A map/reduce program that counts the average length of the words in the input files.
wordmedian: A map/reduce program that counts the median length of the words in the input files.
wordstandarddeviation: A map/reduce program that counts the standard deviation of the length of the words in the input files.

```
cloudera@quickstart ~]$ hadoop jar /usr/jars/hadoop-examples.jar wordcount words.txt out
20/09/10 02:22:51 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
20/09/10 02:22:52 INFO input.FileInputFormat: Total input paths to process : 1
20/09/10 02:22:52 INFO mapreduce.JobSubmitter: number of splits:1
20/09/10 02:22:53 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1599726937632_0002
20/09/10 02:22:53 INFO impl.YarnClientImpl: Submitted application application_1599726937632_0002
20/09/10 02:22:53 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1599726937632_0002/
20/09/10 02:22:53 INFO mapreduce.Job: Running job: job_1599726937632_0002
20/09/10 02:23:04 INFO mapreduce.Job: Job job_1599726937632_0002 running in uber mode : false
20/09/10 02:23:04 INFO mapreduce.Job: map 0% reduce 0%
20/09/10 02:23:12 INFO mapreduce.Job: map 100% reduce 0%
20/09/10 02:23:22 INFO mapreduce.Job: map 100% reduce 100%
20/09/10 02:23:22 INFO mapreduce.Job: Job job_1599726937632_0002 completed successfully
20/09/10 02:23:22 INFO mapreduce.Job: Counters: 49
```

File System Counters

FILE: Number of bytes read=978916
FILE: Number of bytes written=2178931
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=5458319
HDFS: Number of bytes written=717768
HDFS: Number of read operations=6
HDFS: Number of large read operations=0
HDFS: Number of write operations=2

Job Counters

Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1

cloudera@quickstart:~

File Edit View Search Terminal Help

[cloudera@quickstart ~]\$ hadoop fs -ls

Found 1 items

-rw-r--r-- 1 cloudera cloudera 5458199 2020-09-10 02:15 words.txt

[cloudera@quickstart ~]\$ hadoop jar /usr/ja

jars/ java/

[cloudera@quickstart ~]\$ hadoop jar /usr/ja

jars/ java/

[cloudera@quickstart ~]\$ hadoop jar /usr/ja

jars/ java/

[cloudera@quickstart ~]\$ hadoop jar /usr/jars/hadoop-examples.jar words.txt output_words

Unknown program 'words.txt' chosen.

Valid program names are:

aggregatewordcount: An Aggregate based map/reduce program that counts the words in the input files.

aggregatewordhist: An Aggregate based map/reduce program that computes the histogram of the words in the input files.

bbp: A map/reduce program that uses Bailey-Borwein-Plouffe to compute exact digits of Pi.

dbcount: An example job that count the pageview counts from a database.

distbbp: A map/reduce program that uses a BBP-type formula to compute exact bits of Pi.

grep: A map/reduce program that counts the matches of a regex in the input.

join: A job that effects a join over sorted, equally partitioned datasets

multifilewc: A job that counts words from several files.

pentomino: A map/reduce tile laying program to find solutions to pentomino problems.

pi: A map/reduce program that estimates Pi using a quasi-Monte Carlo method.

randomtextwriter: A map/reduce program that writes 10GB of random textual data per node.

randomwriter: A map/reduce program that writes 10GB of random data per node.

secondarysort: An example defining a secondary sort to the reduce.

sort: A map/reduce program that sorts the data written by the random writer.

sudoku: A sudoku solver.

teragen: Generate data for the terasort

terasort: Run the terasort

teravalidate: Checking results of terasort

wordcount: A map/reduce program that counts the words in the input files.

wordmean: A map/reduce program that counts the average length of the words in the input files.

wordmedian: A map/reduce program that counts the median length of the words in the input files.

wordstandarddeviation: A map/reduce program that counts the standard deviation of the length of the words in the input files.

[cloudera@quickstart ~]\$ clear

[cloudera@quickstart ~]\$ hadoop fs -ls

Found 1 items

-rw-r--r-- 1 cloudera cloudera 5458199 2020-09-10 02:15 words.txt

[cloudera@quickstart ~]\$ hadoop jar /usr/jars/hadoop-examples.jar words.txt out

Unknown program 'words.txt' chosen.

Valid program names are:

aggregatewordcount: An Aggregate based map/reduce program that counts the words in the input files.

aggregatewordhist: An Aggregate based map/reduce program that computes the histogram of the words in the input files.

les.

bbp: A map/reduce program that uses Bailey-Borwein-Plouffe to compute exact digits of Pi.

dbcount: An example job that count the pageview counts from a database.

cloudera@quickstart:~