

Workshop 6 – Detail Hands on Hadoop

1. Hadoop Directory Structure [hdfs](#)

[dfs -mkdir -p /user/hadoop](#)

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~  
dfsdata Downloads hadoop-3.2.2.tar.gz Pictures snap tmpdata  
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ cd hadoop-3.2.2/sbin  
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/sbin$ ./start-d  
fs.sh  
Starting namenodes on [localhost]  
Starting datanodes  
Starting secondary namenodes [rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx]  
2024-04-10 07:20:10,669 WARN util.NativeCodeLoader: Unable to load native-hadoop  
library for your platform... using builtin-java classes where applicable  
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/sbin$ ./start-y  
arn.sh  
Starting resourcemanager  
Starting nodemanagers  
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/sbin$ jps  
5776 NodeManager  
5974 Jps  
5239 SecondaryNameNode  
5450 ResourceManager  
5036 DataNode  
4780 NameNode  
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/sbin$ hdfs dfs  
-mkdir -p /user/hadoop  
2024-04-10 07:21:52,929 WARN util.NativeCodeLoader: Unable to load native-hadoop  
library for your platform... using builtin-java classes where applicable
```

[hdfs dfs -ls /user/hadoop](#)

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~  
Generic options supported are:  
-conf <configuration file> specify an application configuration file  
-D <property=value> define a value for a given property  
-fs <file:///hdfs://namenode:port> specify default filesystem URL to use, overr  
ides 'fs.defaultFS' property from configurations.  
-jt <local|resourcemanager:port> specify a ResourceManager  
-files <file1,...> specify a comma-separated list of files to be  
copied to the map reduce cluster  
-libjars <jar1,...> specify a comma-separated list of jar files to  
be included in the classpath  
-archives <archive1,...> specify a comma-separated list of archives to  
be unarchived on the compute machines  
  
The general command line syntax is:  
command [genericOptions] [commandOptions]  
  
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ hdfs dfs -ls  
2024-04-10 07:26:26,227 WARN util.NativeCodeLoader: Unable to load native-hadoop  
library for your platform... using builtin-java classes where applicable  
Found 2 items  
drwxr-xr-x - hadoop supergroup 0 2024-03-29 08:53 input  
drwxr-xr-x - hadoop supergroup 0 2024-03-29 09:12 output  
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$
```

2. Example Programs

yarn jar \$YARN_EXAMPLES/hadoop-mapreduce-examples.jar

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/ha...
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/share/hadoop/m...
preduce$ yarn jar hadoop-mapreduce-examples-3.2.2.jar
An example program must be given as the first argument.
Valid program names are:
  aggregatwordcount: An Aggregate based map/reduce program that counts the word
s in the input files.
  aggregatwordhist: An Aggregate based map/reduce program that computes the his
togram of the words in the input files.
  bbp: A map/reduce program that uses Bailey-Borwein-Plouffe to compute exact di
gits 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 bi
ts 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 pro
blems.
  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.
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/share/hadoop/m...
preduce$
```

2.1 Pi program

yarn jar \$YARN_EXAMPLES/hadoop-mapreduce-examples.jar pi 16 1000

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/share/hadoop/m...
preduce$ yarn jar hadoop-mapreduce-examples-3.2.2.jar pi 16 1000
Number of Maps = 16
Samples per Map = 1000
2024-04-10 07:53:45,644 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Wrote input for Map #0
Wrote input for Map #1
Wrote input for Map #2
Wrote input for Map #3
Wrote input for Map #4
File Input Format Counters
  Bytes Read=1888
File Output Format Counters
  Bytes Written=97
Job Finished in 3.493 seconds
Estimated value of Pi is 3.142500000000000000000000
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/share/hadoop/m...
preduce$
```

3. Word Count Version 1 3.1 Word

Count.java `mkdir wordcount-v1 cd`

`wordcount-v1`

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/share/hadoop/ma
preduce$ cd
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ mkdir wordcount-v1
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ ls
Desktop    Downloads  Music      snap       Videos
dfsdata    hadoop-3.2.2  Pictures  Templates  wordcount-v1
Documents  hadoop-3.2.2.tar.gz  Public    tmpdata
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ cd wordcount-v1
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$
```

3.2 Running the Word Count Program

1. `javac -classpath $(hadoop classpath) WordCount.java`

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ mkdir wordcount-v1
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ ls
Desktop    Downloads  Music      snap       Videos
dfsdata    hadoop-3.2.2  Pictures  Templates  wordcount-v1
Documents  hadoop-3.2.2.tar.gz  Public    tmpdata
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ cd wordcount-v1
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ javac -classpa
th $(hadoop classpath) WordCount.java
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$
```

2. `jar cf wordcount.jar Word*.class`

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ mkdir wordcount-v1
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ ls
Desktop    Downloads  Music      snap       Videos
dfsdata    hadoop-3.2.2  Pictures  Templates  wordcount-v1
Documents  hadoop-3.2.2.tar.gz  Public    tmpdata
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ cd wordcount-v1
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ javac -classpath $(hadoop classpat
h) WordCount.java
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ jar cf wordcount.jar Word*.class
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$
```

3. `hdfs dfs -mkdir input_word`

```
dfsdata    hadoop-3.2.2  Pictures  Templates  wordcount-v1
Documents  hadoop-3.2.2.tar.gz  Public    tmpdata
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ cd wordcount-v1
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ javac -classpath $(hadoop classpat
h) WordCount.java
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ jar cf wordcount.jar Word*.class
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ hdfs dfs -mkdir input_word
2024-04-10 08:10:50,559 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your pl
atform... using builtin-java classes where applicable
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$
```

4. `echo A long time ago in a galaxy far far away > testfile1`

`echo Another episode of Star Wars > testfile2`


```

2024-04-10 08:10:50,559 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ echo A long time ago in a galaxy far far away > testfile1
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ echo Another episode of Star Wars > testfile2
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ echo A long time ago in a galaxy far far away > testfile1
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ echo Another episode of Star Wars > testfile2
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$

```

5. `hdfs dfs -put testfile? input_word`

```

hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ echo A long time ago in a galaxy far far away > testfile1
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ echo Another episode of Star Wars > testfile2
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ hdfs dfs -put testfile? input_word
2024-04-10 08:14:53,416 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$

```

6. `hdfs dfs -rm -R output_word`

```

hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ echo Another episode of Star Wars > testfile2
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ hdfs dfs -put testfile? input_word
2024-04-10 08:14:53,416 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ hdfs dfs -rm -R output_word
2024-04-10 08:16:27,007 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
rm: 'output_word': No such file or directory
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$

```

7. `hadoop jar wordcount.jar WordCount input_word output_word`

```

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=70
File Output Format Counters
    Bytes Written=94
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$

```

```

hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ hadoop jar wordcount.jar WordCount
input_word output_word
2024-04-10 08:21:48,865 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-04-10 08:21:49,268 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-04-10 08:21:49,329 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-04-10 08:21:49,329 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-04-10 08:21:49,432 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2024-04-10 08:21:49,504 INFO input.FileInputFormat: Total input files to process : 2
2024-04-10 08:21:49,538 INFO mapreduce.JobSubmitter: number of splits:2
2024-04-10 08:21:49,628 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1421412149_0001
2024-04-10 08:21:49,628 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-04-10 08:21:49,716 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-04-10 08:21:49,716 INFO mapreduce.Job: Running job: job_local1421412149_0001
2024-04-10 08:21:49,717 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-04-10 08:21:49,722 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-04-10 08:21:49,722 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-04-10 08:21:49,722 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-04-10 08:21:49,747 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-04-10 08:21:49,748 INFO mapred.LocalJobRunner: Starting task: attempt_local1421412149_0001_m_000000_0
2024-04-10 08:21:49,763 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-04-10 08:21:49,763 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false

```

8. hdfs dfs -ls output_word

```

Bytes Written=94
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ hdfs dfs -ls output_word
2024-04-10 08:24:24,942 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 hadoop supergroup          0 2024-04-10 08:21 output_word/_SUCCESS
-rw-r--r-- 1 hadoop supergroup        94 2024-04-10 08:21 output_word/part-r-000000
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$

```

9. hdfs dfs -cat output_word/part-r-000000

```

hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ hdfs dfs -cat output_word/part-r-000000
2024-04-10 08:26:30,633 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
A 1
Another 1
Star 1
Wars 1
a 1
ago 1
away 1
episode 1
far 2
galaxy 1
in 1
long 1
of 1
time 1
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$

```

10. hdfs dfs -get output_word/part-r-000000 word-results.txt

```

time 1
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ hdfs dfs -get output_word/part-r-000000 word-results.txt
2024-04-10 08:31:09,315 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$

```

3.3 Rerunning the Word Count

`hdfs dfs -rm -R output_word hadoop jar wordcount.jar`

`WordCount input_word output_word`

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ hdfs dfs -rm -R output_word
2024-04-10 08:54:46,441 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Deleted output_word
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ hadoop jar wordcount.jar WordCount
input_word output_word
2024-04-10 10:41:39,325 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-04-10 10:41:39,688 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-04-10 10:41:39,733 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-04-10 10:41:39,733 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-04-10 10:41:39,827 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not p
```

`hdfs dfs -cat output_word/part-r-00000`

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ hdfs dfs -cat output_word/part-r-00000
2024-04-10 10:46:09,985 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
cat: 'output_word/part-r-00000': No such file or directory
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$
```

3.4 Using a larger dataset

1. Hadoop will work on any file in the input directory, so remove the previous testfiles:

`hdfs dfs -rm input_word/testfile?`

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ hdfs dfs -rm input_word/testfile?
2024-04-10 10:55:33,548 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Deleted input_word/testfile1
Deleted input_word/testfile2
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$
```

2. Save the new file to the input directory:

`hdfs dfs -put shakespeare.txt input_word`

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ hdfs dfs -put t8.shakespeare.txt input_word
2024-04-10 11:01:58,776 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$
```

3. By now you will have created an output directory in the previous run of the program, so you need to delete this:

`hdfs dfs -rm -R output_word`

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ hdfs dfs -rm -R output_word
2024-04-10 11:04:05,702 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Deleted output_word
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$
```

4. Run the Map Reduce program:

`hadoop jar wordcount.jar WordCount input_word output_word`

```

hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ hadoop jar wordcount.jar WordCount
input_word output_word
2024-04-10 11:20:50,309 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-04-10 11:20:50,731 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-04-10 11:20:50,777 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-04-10 11:20:50,777 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-04-10 11:20:50,879 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
    BAD_IO=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
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$

```

5. Check what files are in the output directory:

`hdfs dfs -ls output_word`

```

hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ hdfs dfs -ls output_word
2024-04-10 11:23:45,778 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2024-04-10 11:20 output_word/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 717768 2024-04-10 11:20 output_word/part-r-00000
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$

```

6. See what is in the output file:

`hdfs dfs -cat output_word/part-r-00000`

```

zip 1
zir, 1
zir. 1
zo 1
zodiac 1
zodiacs 1
zone, 1
zounds! 1
zounds, 1
zwagger'd 1
} 2
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$

```

7. This time there will be a bigger results set, so to view it properly, copy the file locally:

`hdfs dfs -get output_word/part-r-00000 shakespeare-results.txt`

```

hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ hdfs dfs -get output_word/part-r-00000 shakespeare-results.txt
2024-04-10 11:27:11,954 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$

```

8. Use the Linux command `more` to view the results file in the Operating System:

`more shakespeare-results.txt`

```

zounds, 1
zwagger'd 1
} 2
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$

```

4. Word Count Version 2

The results file is quite large, so you can use the Linux command `grep` to search for a particular word:

`grep anon shakespeare-results.txt`

```
hdoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ grep anon shakespeare-results.txt
Canonized, 1
abroad-anon 1
anon 30
anon! 2
anon, 12
anon- 2
anon. 41
anon.- 1
anon; 5
anon? 1
anonymous 1
Canon 2
Canon, 1
Canon. 2
Canon; 1
Canoniz'd 1
Canoniz'd, 1
Canonize 1
Canonized 1
Canons, 1
hdoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$
```

`grep Anon shakespeare-results.txt`

```
hdoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ grep Anon shakespeare-results.txt
'Anon 1
'Anon! 1
'Anon, 1
Anon 9
Anon! 1
Anon, 18
Anon. 2
hdoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$
```

4.1 Making Word Count Case Insensitive

Some extra code is needed to make the word count not case sensitive.

Return to your home directory and create a new directory to keep the versions separate, for example:

`cd`

`mkdir wordcount-v2`

Change to this new directory:

`cd wordcount-v2`

Then copy the new version here.

Compile the file and create a Jar file as seen in steps 1 and 2 of Section 4.2


```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v1$ cd
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ mkdir wordcount-v2
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ cd wordcount-v2
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v2$ javac -classpath $(hadoop classpath) WordCount.java
javac: file not found: WordCount.java
Usage: javac <options> <source files>
use -help for a list of possible options
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v2$ javac -classpath $(hadoop classpath) WordCount.java
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v2$
```

Remove the output directory: `hdfs`

`dfs -rm -R output_word`

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v2$ hdfs dfs -rm -R output_word
2024-04-10 11:52:47,540 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Deleted output_word
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v2$
```

Assuming that the `shakespeare.txt` file is still stored in the hdfs run the program:

`hadoop jar wordcount.jar WordCount input_word output_word`

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v2$ hadoop jar wordcount.jar WordCount input_word output_word
2024-04-10 11:56:19,317 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-04-10 11:56:19,697 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-04-10 11:56:19,742 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 seconds
```

If everything has worked correctly, retrieve the results file from the hdfs:

`hdfs dfs -get output_word/part-r-00000 shakespeare-results.txt`

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v2$ hdfs dfs -get output_word/part-r-00000 shakespeare-results.txt
2024-04-10 12:01:19,981 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v2$
```

Try using `grep` to search for `anon` again:

`grep anon shakespeare-results.txt`

```

hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v2$ grep anon shakespeare-results.txt
Canonized,      1
abroad-anon     1
anon           30
anon!           2
anon,          12
anon-           2
anon.          41
anon.-         1
anon;           5
anon?           1
anonymous      1
anon           2
anon,          1
anon.          2
anon;          1
anoniz'd       1
anoniz'd,      1
anonize        1
anonized       1
anons,         1
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v2$

```

grep Anon shakespeare-results.txt

```

hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v2$ grep Anon shakespeare-results.txt
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v2$

```

5. Word Count Version 3

5.1 Removing Punctuation from Words

This final version of Word Count will use the Pattern class from java.util.regex

Return to your home directory and create another new directory to keep this version separate:

```

cd mkdir wordcount-
v3

```

```

hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v2$ grep Anon shakespeare-results.txt
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v2$ cd
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ mkdir wordcount-v3
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$

```

Change to this new directory: `cd`

`wordcount-v3`

```

hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~$ cd wordcount-v3
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$

```

Then copy the new version here.

Compile the file and create a Jar file as seen in steps 1 and 2 earlier.

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ cd wordcount-v3
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ javac -classpath $(hadoop classpath) WordCount.java
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ jar cf wordcount.jar Word*.class
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$
```

Don't forget to remove the output directory:

```
hdfs dfs -rm -R output_word
```

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ hdfs dfs -rm -R output_word
2024-04-10 12:38:00,440 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Deleted output_word
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$
```

Assuming that the shakespeare.txt file is still stored in the hdfs run the program:

```
hadoop jar wordcount.jar WordCount input_word output_word
```

```
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=245910
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$
```

If everything has worked correctly, retrieve the results file from the hdfs:

```
hdfs dfs -get output_word/part-r-00000 shakespeare-results.csv
```

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ hdfs dfs -get output_word/part-r-00000 shakespeare-results.csv
2024-04-10 13:06:00,324 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$
```

Try using grep to search for anon again: `grep`

```
anon shakespeare-results.csv
```

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ grep anon shakespeare-results.csv
anon,128
anonymous,1
anon,6
anon,2
anon,1
anon,2
anon,1
anon,1
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$
```

This time anon will only appear once:

```
anon,128
```

```
anonymous,1 anon,6
```


canoniz,2 canonize,1

canonized,2 canons,1

The results are still case insensitive too, this should produce no results: [grep](#)

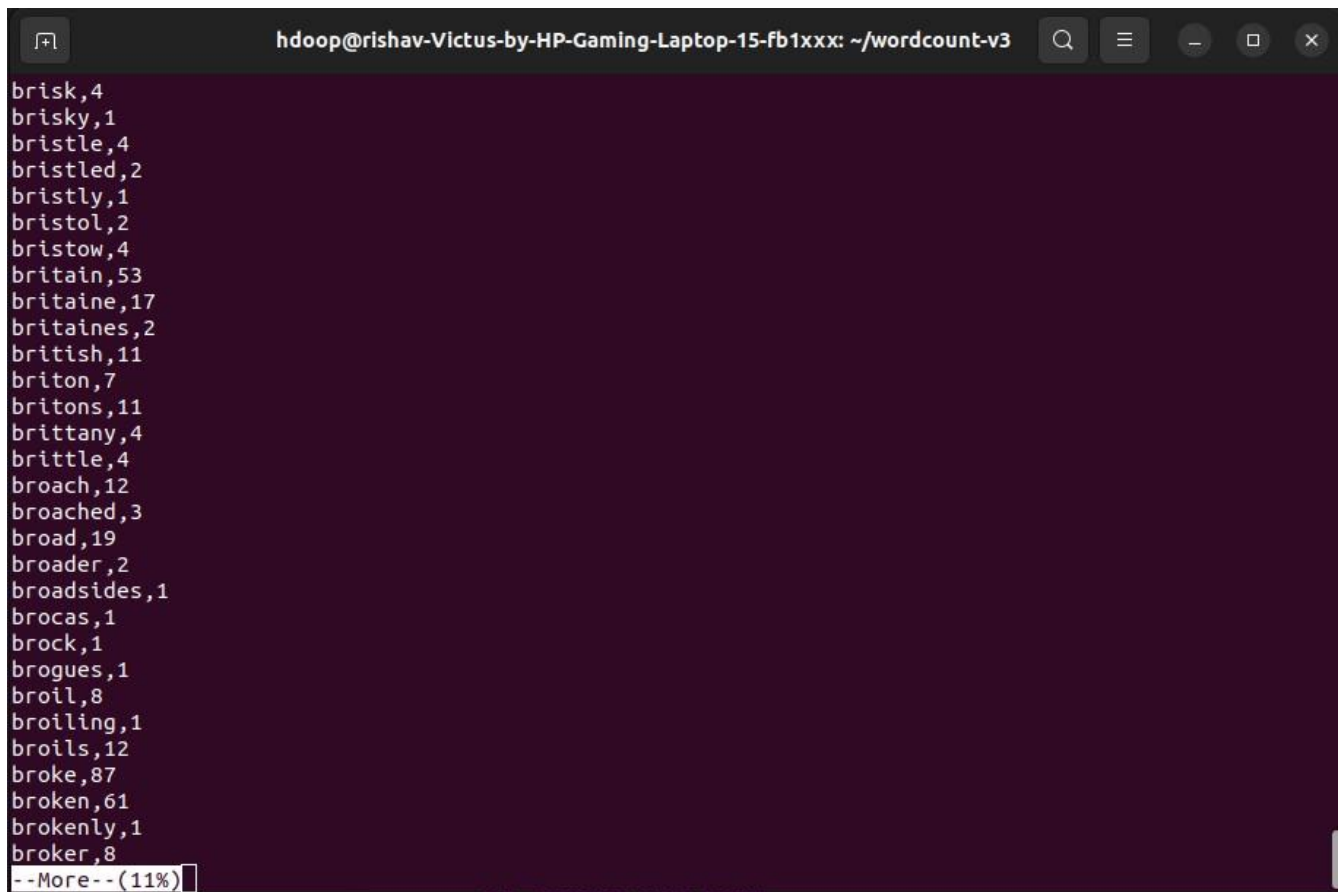
[Anon shakespeare-results.csv](#)

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ grep Anon shakespeare-results.csv
grep: shakespeare-results.csv : No such file or directory
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$
```

6. Exercises to do

The results are still not perfect. If you use more to list the whole file: [more](#)

[shakespeare-results.csv](#)



```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/wordcount-v3
brisk,4
brisky,1
bristle,4
bristled,2
bristly,1
bristol,2
bristow,4
britain,53
britannie,17
britannes,2
british,11
briton,7
britons,11
brittany,4
brittle,4
broach,12
broached,3
broad,19
broader,2
broadsides,1
brocas,1
brock,1
brogues,1
broil,8
broiling,1
broils,12
broke,87
broken,61
brokenly,1
broker,8
--More-- (11%)
```

6.2 Sample Programs

Section 3 Introduced the sample programs that come with Hadoop. As a reminder to see the full list type:

[yarn jar \\$YARN_EXAMPLES/hadoop-mapreduce-examples.jar](#)

```

hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/hadoop-3.2.2/share/hadoop/...
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/share/hadoop$ cd mapreduce/
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/share/hadoop/mapreduce$ yarn jar ha
doop-mapreduce-examples-3.2.2.jar
An example program must be given as the first argument.
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 th
e words in the input files.
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/share/hadoop/mapreduce$ 

```

To find out what options are required, add one of the above to the command, for example, if you are a Sudoku fan:

`yarn jar $YARN_EXAMPLES/hadoop-mapreduce-examples.jar sudoku` Which tells

you to: Include a puzzle on the command line.

```

e words in the input files.
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/share/hadoop/mapreduce$ yarn jar ha
doop-mapreduce-examples-3.2.2.jar sudoku
Include a puzzle on the command line.
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/share/hadoop/mapreduce$ 

```

To run Sudoku with this puzzle:

`yarn jar $YARN_EXAMPLES/hadoop-mapreduce-examples.jar sudoku puzzle.txt`


```

hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/share/hadoop/mapreduce$ yarn jar ha
doop-mapreduce-examples-3.2.2.jar sudoku
Include a puzzle on the command line.
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/share/hadoop/mapreduce$ yarn jar ha
doop-mapreduce-examples-3.2.2.jar sudoku puzzle.txt
Solving puzzle.txt
8 5 1 3 9 2 6 4 7
4 3 2 6 7 8 1 9 5
7 9 6 5 1 4 3 8 2
6 1 4 8 2 3 7 5 9
5 7 8 9 6 1 4 2 3
3 2 9 4 5 7 8 1 6
9 4 7 2 8 6 5 3 1
1 8 5 7 3 9 2 6 4
2 6 3 1 4 5 9 7 8

Found 1 solutions
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/hadoop-3.2.2/share/hadoop/mapreduce$

```

6.3 Shell Script

```

hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/wordcount-v3
#!/bin/bash
# Word Count Script
# Usage: ./wordcount-v3.sh <input_dir> <output_dir>

# Check if input and output directories are provided
if [ $# -ne 2 ]; then
    echo "Usage: ./wordcount-v3.sh <input_dir> <output_dir>"
    exit 1
fi

input_dir=$1
output_dir=$2

# Create output directory if it doesn't exist
mkdir -p $output_dir

# Run the wordcount program
hadoop jar $HADOOP_HOME/share/hadoop-mapreduce/hadoop-mapreduce-examples-3.2.2.jar wordcount $input_dir $output_dir

# Check if the job was successful
if [ $? -eq 0 ]; then
    echo "Word count completed successfully."
else
    echo "Word count failed."
fi

```

```

hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ if [ -f word-results.txt ];
then rm word-results.txt;
fi
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ hdfs dfs -get output_word/part-r-00000 word-results.txt
2024-04-10 19:06:59,217 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ chmod u+x runProg
chmod: cannot access 'runProg': No such file or directory
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ chmod u+x runProg
chmod: changing permissions of 'runProg': Operation not permitted
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ ^C
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ ls -l runProg
-rw-r--r-- 1 root root 521 मॉर्च  8 2019 runProg
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ sudo chmod u+x runProg
[sudo] password for hadoop:
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ ./runProg
bash: ./runProg: Permission denied
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ ^C
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ chmod +x runProg
chmod: changing permissions of 'runProg': Operation not permitted
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ sudo chmod +x runProg
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/wordcount-v3$ ./runProg
2024-04-10 19:13:32,261 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-04-10 19:13:33,048 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Deleted output_word
2024-04-10 19:13:33,905 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-04-10 19:13:34,259 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-04-10 19:13:34,311 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-04-10 19:13:34,311 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-04-10 19:13:34,405 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2024-04-10 19:13:34,475 INFO input.FileInputFormat: Total input files to process : 1
2024-04-10 19:13:34,506 INFO mapreduce.JobSubmitter: number of splits:1
2024-04-10 19:13:34,592 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local2060250944_0001
2024-04-10 19:13:34,592 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-04-10 19:13:34,660 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-04-10 19:13:34,660 INFO mapreduce.Job: Running job: job_local2060250944_0001
2024-04-10 19:13:34,661 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-04-10 19:13:34,666 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-04-10 19:13:34,666 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup_temporary folders under output directory:false, ignore cleanup failures: false
2024-04-10 19:13:34,666 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-04-10 19:13:34,688 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-04-10 19:13:34,689 INFO mapred.LocalJobRunner: Starting task: attempt_local2060250944_0001_m_000000_0
2024-04-10 19:13:34,703 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-04-10 19:13:34,703 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup_temporary folders under output directory:false, ignore cleanup failures: false
2024-04-10 19:13:34,716 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2024-04-10 19:13:34,718 INFO mapred.MapTask: Processing split: hdfs://127.0.0.1:9000/user/hadoop/input_word/t8.shakespeare.txt:0+5458199
2024-04-10 19:13:34,765 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2024-04-10 19:13:34,765 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2024-04-10 19:13:34,765 INFO mapred.MapTask: soft limit at 83886080

```



```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/wordcount-v3
WRONG_REDUCE=0
File Output Format Counters
  Bytes Written=245910
2024-04-10 19:13:36,635 INFO mapred.LocalJobRunner: Finishing task: attempt_local2060250944_0001_r_000000_0
2024-04-10 19:13:36,635 INFO mapred.LocalJobRunner: reduce task executor complete.
2024-04-10 19:13:36,677 INFO mapreduce.Job: map 100% reduce 100%
2024-04-10 19:13:36,678 INFO mapreduce.Job: Job job_local2060250944_0001 completed successfully
2024-04-10 19:13:36,683 INFO mapreduce.Job: Counters: 36
File System Counters
  FILE: Number of bytes read=676208
  FILE: Number of bytes written=2098668
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=10916398
  HDFS: Number of bytes written=245910
  HDFS: Number of read operations=15
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=4
  HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Map input records=124456
  Map output records=984497
  Map output bytes=8770768
  Map output materialized bytes=334066
  Input split bytes=127
  Combine input records=984497
  Combine output records=23722
  Reduce input groups=23722
  Reduce shuffle bytes=334066
  Reduce input records=23722
  Reduce output records=23722
  Spilled Records=47444
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=14
  Total committed heap usage (bytes)=1064384640
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=245910
2024-04-10 19:13:37,158 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2024-04-10 19:13 output_word/ SUCCESS
-rw-r--r-- 1 hadoop supergroup 245910 2024-04-10 19:13 output_word/part-r-000000
2024-04-10 19:13:37,956 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/wordcount-v3
```

7. Python Code

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/python
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: $ hdfs dfs -ls
2024-04-11 08:52:04,185 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 4 items
drwxr-xr-x 1 hadoop supergroup 0 2024-03-29 08:53 input
drwxr-xr-x 1 hadoop supergroup 0 2024-04-10 11:01 input_word
drwxr-xr-x 1 hadoop supergroup 0 2024-03-29 09:12 output
drwxr-xr-x 1 hadoop supergroup 0 2024-04-10 19:13 output_word
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/python$ ^C
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/python$ hdfs dfs -rm -R output
2024-04-11 08:53:16,231 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
deleted output
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/python$ ls
hadoop-streaming-3.2.2.jar runpy wordcount_mapper.py wordcount_reducer.py
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/python$ ^C
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/python$ hadoop jar /admin:///home/hadoop/hadoop-3.2.2/share/hadoop/tools/lib/hadoop-streaming-3.2.2.jar \
-input myHadoop_input \
-output myHadoop_output \
-mapper $PWD/wordcount_mapper.py \
-reducer $PWD/wordcount_reducer.py
JAR does not exist or is not a normal file: /admin:/home/hadoop/hadoop-3.2.2/share/hadoop/tools/lib/hadoop-streaming-3.2.2.jar
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/python$ hadoop jar /home/hadoop/hadoop-3.2.2/share/hadoop/tools/lib/hadoop-streaming-3.2.2.jar \
-input myHadoop_input \
-output myHadoop_output \
-mapper $PWD/wordcount_mapper.py \
-reducer $PWD/wordcount_reducer.py
2024-04-11 08:58:41,553 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-04-11 08:58:42,002 INFO Impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-04-11 08:58:42,058 INFO Impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-04-11 08:58:42,058 INFO Impl.MetricsSystemImpl: JobTracker metrics system started
2024-04-11 08:58:42,068 WARN Impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2024-04-11 08:58:42,208 INFO mapreduce.JobSubmitter: Cleaning up the staging area file:/tmp/hadoop/mapred/staging/hadoop1881238094/.staging/job_local1881238094_0001
2024-04-11 08:58:42,209 ERROR streaming.StreamJob: Error Launching job : Input path does not exist: hdfs://127.0.0.1:9000/user/hadoop/myHadoop_input
Streaming Command Failed!
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/python$ hdfs dfs -ls myHadoop_input
ls: 'myHadoop_input': No such file or directory
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/python$ hdfs dfs -ls
2024-04-11 08:59:44,116 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 3 items
drwxr-xr-x 1 hadoop supergroup 0 2024-03-29 08:53 input
drwxr-xr-x 1 hadoop supergroup 0 2024-04-10 11:01 input_word
drwxr-xr-x 1 hadoop supergroup 0 2024-04-10 19:13 output_word
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/python$ hdfs dfs -mkdir myHadoop_input
2024-04-11 09:00:10,090 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/python$ hdfs dfs -put /home/hadoop/python/wordcount_mapper.py myHadoop_input
2024-04-11 09:01:39,804 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
put: '/home/python/wordcount_mapper.py': No such file or directory
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/python$ hdfs dfs -put /home/hadoop/python/wordcount_mapper.py myHadoop_input
2024-04-11 09:02:07,213 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/python$ hdfs dfs -put /home/hadoop/python/wordcount_reducer.py myHadoop_input
2024-04-11 09:02:14,370 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/python$ hadoop jar /home/hadoop/hadoop-3.2.2/share/hadoop/tools/lib/hadoop-streaming-3.2.2.jar -input myHadoop_input -output myHadoop_output -mapper $PWD
/wordcount_mapper.py -reducer $PWD/wordcount_reducer.py
2024-04-11 09:02:31,457 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-04-11 09:02:31,920 INFO Impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-04-11 09:02:31,958 INFO Impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
```

```
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx: ~/python
Bytes Written=1902
2024-04-11 09:02:32,730 INFO mapred.LocalJobRunner: Finishing task: attempt_local338679690_0001_r_000000_0
2024-04-11 09:02:32,730 INFO mapred.LocalJobRunner: reduce task executor complete.
2024-04-11 09:02:33,297 INFO mapreduce.Job: Job job_local338679690_0001 running in uber mode : false
2024-04-11 09:02:33,309 INFO mapreduce.Job: map 100% reduce 100%
2024-04-11 09:02:33,310 INFO mapreduce.Job: Job job_local338679690_0001 completed successfully
2024-04-11 09:02:33,316 INFO mapreduce.Job: Counters: 36
File System Counters
  FILE: Number of bytes read=539833
  FILE: Number of bytes written=2189679
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=8494
  HDFS: Number of bytes written=1902
  HDFS: Number of read operations=24
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=5
  HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Map input records=77
  Map output records=378
  Map output bytes=3477
  Map output materialized bytes=4245
  Input split bytes=239
  Combine input records=0
  Combine output records=0
  Reduce input groups=189
  Reduce shuffle bytes=4245
  Reduce input records=378
  Reduce output records=189
  Spilled Records=756
  Shuffled Maps =2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=5
  Total committed heap usage (bytes)=1036517376
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=3317
File Output Format Counters
  Bytes Written=1902
2024-04-11 09:02:33,316 INFO streaming.StreamJob: Output directory: myHadoop_output
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/python$ hdfs dfs -ls myHadoop_output
2024-04-11 09:03:03,943 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 hadoop supergroup          0 2024-04-11 09:02 myHadoop_output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup      1902 2024-04-11 09:02 myHadoop_output/part-00000
hadoop@rishav-Victus-by-HP-Gaming-Laptop-15-fb1xxx:~/python$
```