Screenshots

Execution steps:

```
● ● ●                    🏠 mangalakhandekar — ssh localhost — 129×58
Mangalas-MBP:WordCountV3 mangalakhandekar$ cd ..
Mangalas-MBP:HadoopPrograms mangalakhandekar$ cd WordCountV4
Mangalas-MBP:WordCountV4 mangalakhandekar$ javac WordCountV4.java -cp $(hadoop classpath)
Mangalas-MBP:WordCountV4 mangalakhandekar$ jar cf WordCountV4.jar WordCount*.class
Mangalas-MBP:WordCountV4 mangalakhandekar$ time hadoop jar WordCountV4.jar WordCountV4 /user/mangalakhandekar/inputLarge /user/ma
ngalakhandekar/outputLarge/wcV4Output -skip /user/mangalakhandekar/stop_words.txt
2018-09-18 19:44:54,907 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
2018-09-18 19:44:55,745 INFO impl.MetricsConfig: loaded properties from hadoop-metrics2.properties
2018-09-18 19:44:55,875 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2018-09-18 19:44:55,875 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2018-09-18 19:44:56,030 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool
interface and execute your application with ToolRunner to remedy this.
2018-09-18 19:44:56,157 INFO input.FileInputFormat: Total input files to process : 1
2018-09-18 19:44:56,222 INFO mapreduce.JobSubmitter: number of splits:1
2018-09-18 19:44:56,422 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1037996150_0001
2018-09-18 19:44:56,424 INFO mapreduce.JobSubmitter: Executing with tokens: []
2018-09-18 19:44:56,713 INFO mapred.LocalDistributedCacheManager: Creating symlink: /usr/local/Cellar/hadoop/hdfs/tmp/mapred/loca
l/1533314296504/stop_words.txt <- /Users/mangalakhandekar/HadoopPrograms/WordCountV4/stop_words.txt
2018-09-18 19:44:56,720 INFO mapred.LocalDistributedCacheManager: Localized hdfs://localhost:9000/user/mangalakhandekar/stop_word
s.txt as file:/usr/local/Cellar/hadoop/hdfs/tmp/mapred/local/1533314296504/stop_words.txt
2018-09-18 19:44:56,796 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2018-09-18 19:44:56,797 INFO mapreduce.Job: Running job: job_local1037996150_0001
2018-09-18 19:44:56,798 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2018-09-18 19:44:56,804 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2018-09-18 19:44:56,804 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directo
ry:false, ignore cleanup failures: false
2018-09-18 19:44:56,805 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2018-09-18 19:44:56,856 INFO mapred.LocalJobRunner: Waiting for map tasks
2018-09-18 19:44:56,857 INFO mapred.LocalJobRunner: Starting task: attempt_local1037996150_0001_m_000000_0
2018-09-18 19:44:56,879 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2018-09-18 19:44:56,879 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directo
ry:false, ignore cleanup failures: false
2018-09-18 19:44:56,886 INFO util.ProcfsBasedProcessTree: ProcfsBasedProcessTree currently is supported only on Linux.
2018-09-18 19:44:56,887 INFO mapred.Task:  Using ResourceCalculatorProcessTree : null
2018-09-18 19:44:56,891 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/user/mangalakhandekar/inputLarge/largeFile1.
txt:0+408778
2018-09-18 19:44:56,962 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2018-09-18 19:44:56,962 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2018-09-18 19:44:56,962 INFO mapred.MapTask: soft limit at 83886080
2018-09-18 19:44:56,962 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2018-09-18 19:44:56,962 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2018-09-18 19:44:56,966 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2018-09-18 19:44:57,129 INFO mapred.LocalJobRunner:
2018-09-18 19:44:57,129 INFO mapred.MapTask: Starting flush of map output
2018-09-18 19:44:57,129 INFO mapred.MapTask: Spilling map output
2018-09-18 19:44:57,129 INFO mapred.MapTask: bufstart = 0; bufend = 581669; bufvoid = 104857600
2018-09-18 19:44:57,129 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 25984740(103938960); length = 229657/6553600
2018-09-18 19:44:57,274 INFO mapred.MapTask: Finished spill 0
2018-09-18 19:44:57,292 INFO mapred.Task: Task:attempt_local1037996150_0001_m_000000_0 is done. And is in the process of committi
ng
2018-09-18 19:44:57,295 INFO mapred.LocalJobRunner: map
2018-09-18 19:44:57,295 INFO mapred.Task: Task 'attempt_local1037996150_0001_m_000000_0' done.
2018-09-18 19:44:57,302 INFO mapred.Task: Final Counters for attempt_local1037996150_0001_m_000000_0: Counters: 23
        File System Counters
                FILE: Number of bytes read=4522
                FILE: Number of bytes written=636758
                FILE: Number of read operations=0
```

Input file creation on local and copying it to HDFS:

```
Mangalas-MBP:WordCountV1 mangalakhandekar$ hadoop fs -mkdir -p /user/mangalakhandekar/inputSimple
2018-09-18 18:50:36,744 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
Mangalas-MBP:WordCountV1 mangalakhandekar$ cd
Mangalas-MBP:~ mangalakhandekar$ hadoop fs -copyFromLocal Desktop/simpleFile1.txt /user/mangalakhandekar/inputSimple
2018-09-18 18:51:14,102 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
Mangalas-MBP:~ mangalakhandekar$ hadoop fs -copyFromLocal Desktop/simpleFile2.txt /user/mangalakhandekar/inputSimple
2018-09-18 18:51:22,630 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
Mangalas-MBP:~ mangalakhandekar$ hadoop fs -copyFromLocal Desktop/simpleFile3.txt /user/mangalakhandekar/inputSimple
2018-09-18 18:51:32,010 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
Mangalas-MBP:~ mangalakhandekar$ ▊
```

Output Screenshots:
Below screenshots are samples of output:

Dataset1:
Word Count Program V1:

```
        Bytes Written=100

real 0m3.847s
user 0m9.809s
sys 0m0.608s
Mangalas-MBP:WordCountV1 mangalakhandekar$ hadoop fs -cat /user/mangalakhandekar/
outputSimple/wcV1Output/part-r-00000
2018-09-18 19:08:38,294 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
!	1
Bye	1
Goodbye	1
Hadoop	2
Hello	2
World	2
and	1
bye	1
goodbye	1
hadoop	2
hello	2
world	2
.	1
Mangalas-MBP:WordCountV1 mangalakhandekar$
```

Word Count Program V2:

```
        Bytes Written=100

real 0m3.853s
user 0m9.778s
sys 0m0.615s
Mangalas-MBP:WordCountV2 mangalakhandekar$ hadoop fs -cat /user/mangalakhandekar/
outputSimple/wcV2Output/part-r-00000
2018-09-18 19:10:20,514 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
!	1
Bye	1
Goodbye	1
Hadoop	2
Hello	2
World	2
and	1
bye	1
goodbye	1
hadoop	2
hello	2
world	2
.	1
Mangalas-MBP:WordCountV2 mangalakhandekar$
```

Word Count Program V3:

```
        Bytes Written=59

real0m4.092s
user0m10.139s
sys 0m0.663s
Mangalas-MBP:WordCountV3 mangalakhandekar$ hadoop fs -cat /user/mangalakhandekar/outputSimple/
wcV3Output/part-r-00000
2018-09-18 19:12:17,859 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
!    1
and 1
bye 2
goodbye 2
hadoop   4
hello    4
world    4
.    1
```

Word Count Program V4:

```
        Bytes Written=55

real0m4.019s
user0m9.783s
sys 0m0.659s
Mangalas-MBP:WordCountV4 mangalakhandekar$ hadoop fs -cat /user/mangalakhandekar/
outputSimple/wcV4Output/part-r-00000
2018-09-18 19:15:47,698 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
bye 2
goodbye 2
hadoop   4
hello    4
world    4
Mangalas-MBP:WordCountV4 mangalakhandekar$
```

DataSet2:

Word Count Program V1:

```
        Bytes Written=476

real0m3.926s
user0m9.720s
sys 0m0.653s
Mangalas-MBP:WordCountV1 mangalakhandekar$ hadoop fs -cat /user/mangalakhandekar/
wcV1Output/part-r-00000
2018-09-18 15:57:37,685 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
,      1
.      1
A      2
An     2
And    1
Are    1
Because,1
But    1
Elephant 1
HDFS   1
Hadoop  2
Hadoop. 1
He     2
Hive,   1
Impala, 1
King!   1
Or     1
Sqoop.  1
The    1
Useful  1
a      1
an     2
and    4
anything 1
at     1
bad,   1
but    1
cling!  1
core    1
```

Word Count Program V2:

```
        bytes Read=425
    File Output Format Counters
        Bytes Written=446

real0m3.839s
user0m8.994s
sys 0m0.623s
Mangalas-MBP:WordCountV2 mangalakhandekar$ hadoop fs -cat /user/mangalakhandekar/wcV2Output/part-r-00000
2018-09-18 16:00:29,940 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
!      2
,      8
.      8
A      2
An     2
And    1
Are    1
Because 1
But    1
Elephant 1
HDFS   1
Hadoop   3
He     2
Hive   1
Impala   1
King   1
Or     1
Sqoop    1
The    1
Useful   1
a      1
an     2
and    4
anything 1
at     1
bad    1
but    1
cling    1
core   1
data   1
does   1
elegant  2
element  1
elephant 2
extraneous   1
fellow   1
forgets  1
gentle   1
gets   1
```

Word Count Program V3:

```
File Input Format Counters
        Bytes Read=425
File Output Format Counters
        Bytes Written=391
```

```
Mangalas-MBP:WordCountV3 mangalakhandekar$ hadoop fs -cat /user/mangalakhandekar/wcV3Output/part-r-00000
2018-09-18 16:03:13,870 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
!       2
,       8
.       8
a       3
an      4
and     5
anything 1
are  1
at      1
bad  1
because  1
but  2
cling      1
core 1
data 1
does 1
elegant  2
element  1
elephant 3
extraneous   1
fellow   1
forgets  1
gentle   1
gets 1
group      1
hadoop   3
hdfs 1
he      3
helps    1
him  1
his  1
hive 1
impala   1
in      1
is      5
king 2
lets 1
mad  1
mellow   1
never    2
or      3
plays    1
sqoop      1
the  4
thing    1
thrive   1
```

Word Count Program V4:

```
File Output Format Counters
        Bytes Written=341
```

```
Mangalas-MBP:WordCountV4 mangalakhandekar$ hadoop fs -cat /user/mangalakhandekar/wcV10Output/part-r-00000
2018-09-18 18:29:36,556 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
!       2
anything 1
are  1
at      1
bad  1
because  1
cling      1
core 1
data 1
does 1
elegant  2
element  1
elephant 3
extraneous   1
fellow   1
forgets  1
gentle   1
gets 1
group      1
hadoop   3
hdfs 1
he      3
helps    1
him  1
his  1
hive 1
impala   1
in      1
king 2
lets 1
mad  1
mellow   1
never    2
plays    1
sqoop      1
thing    1
thrive   1
useful   1
well 1
what 1
with 1
wonderful    1
yellow   2
```