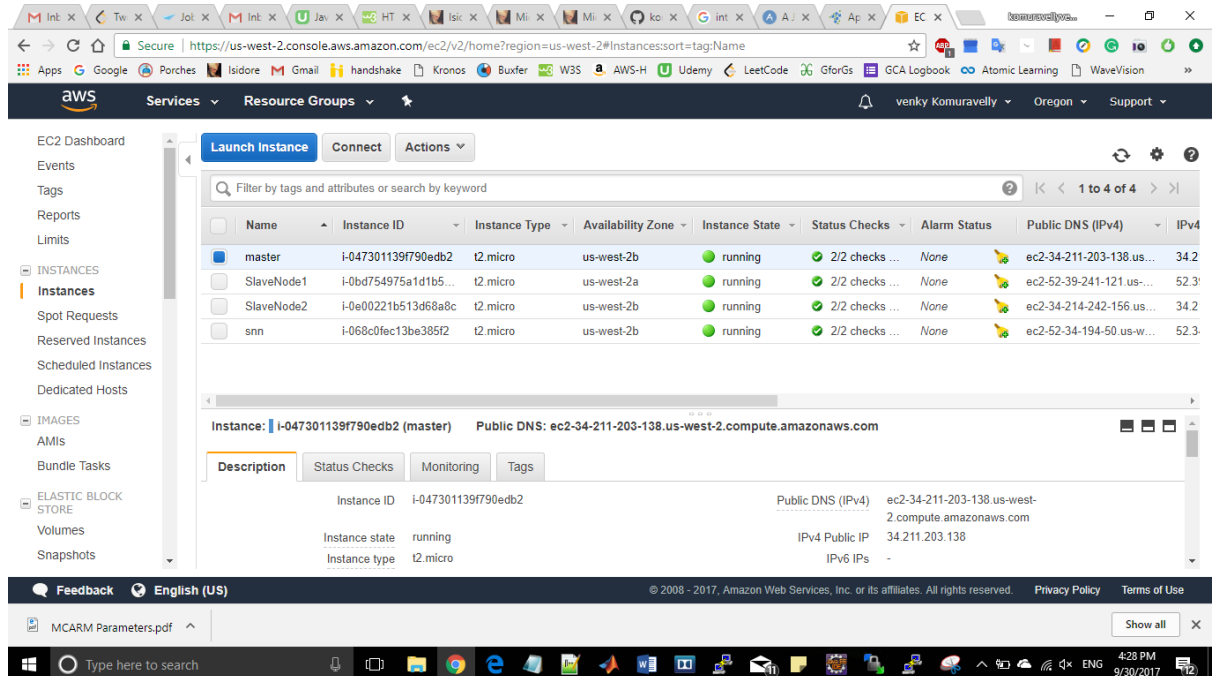


## Assignment-2 In-Map Combining Methods

For the following assignment I used Stock Overflow and Google as resources.

Started the AWS-Instances: -



Formatted the Name Node space:-

```
ubuntu@ip-172-31-38-5: ~/hadoop/conf
17/09/30 20:27:06 INFO namenode.FSNamesystem: fsOwner=ubuntu
17/09/30 20:27:06 INFO namenode.FSNamesystem: supergroup=supergroup
17/09/30 20:27:06 INFO namenode.FSNamesystem: isPermissionEnabled=false
17/09/30 20:27:06 INFO namenode.FSNamesystem: dfs.block.invalidate.limit=100
17/09/30 20:27:06 INFO namenode.FSNamesystem: isAccessTokenEnabled=false accessK
eyUpdateInterval=0 min(s), accessTokenLifetime=0 min(s)
17/09/30 20:27:06 INFO namenode.FSEditLog: dfs.namenode.edits.toleration.length
= 0
17/09/30 20:27:06 INFO namenode.NameNode: Caching file names occurring more than
10 times
17/09/30 20:27:06 INFO common.Storage: Image file /home/ubuntu/hdfstmp/dfs/name/
current/fsimage of size 112 bytes saved in 0 seconds.
17/09/30 20:27:06 INFO namenode.FSEditLog: closing edit log: position=4, editlog
=/home/ubuntu/hdfstmp/dfs/name/current/edits
17/09/30 20:27:07 INFO namenode.FSEditLog: close success: truncate to 4, editlog
=/home/ubuntu/hdfstmp/dfs/name/current/edits
17/09/30 20:27:07 INFO common.Storage: Storage directory /home/ubuntu/hdfstmp/df
s/name has been successfully formatted.
17/09/30 20:27:07 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at ip-172-31-38-5/172.31.38.5
*****/
```

Now, Started the Hadoop by using start-all.sh command as shown below. By giving the JPS command we can confirm that Name node, Secondary Name node ,Job tracker has started on Master Instance.

```
ubuntu@ip-172-31-38-5: ~/hadoop/conf
ubuntu@ip-172-31-38-5:~/hadoop$ clear
ubuntu@ip-172-31-38-5:~/hadoop$ jps
1755 Jps
ubuntu@ip-172-31-38-5:~/hadoop$ cd conf
ubuntu@ip-172-31-38-5:~/hadoop/conf$ start-all.sh
starting namenode, logging to /home/ubuntu/hadoop/libexec/../../logs/hadoop-ubuntu-
namenode-ip-172-31-38-5.out
172.31.33.156: starting datanode, logging to /home/ubuntu/hadoop/libexec/../../logs/
hadoop-ubuntu-datanode-ip-172-31-33-156.out
172.31.29.24: starting datanode, logging to /home/ubuntu/hadoop/libexec/../../logs/
hadoop-ubuntu-datanode-ip-172-31-29-24.out
172.31.36.158: starting secondarynamenode, logging to /home/ubuntu/hadoop/libexe
c/../../logs/hadoop-ubuntu-secondarynamenode-ip-172-31-36-158.out
172.31.38.5: starting secondarynamenode, logging to /home/ubuntu/hadoop/libexec/
../../logs/hadoop-ubuntu-secondarynamenode-ip-172-31-38-5.out
localhost: secondarynamenode running as process 2067. Stop it first.
starting jobtracker, logging to /home/ubuntu/hadoop/libexec/../../logs/hadoop-ubunt
u-jobtracker-ip-172-31-38-5.out
172.31.29.24: starting tasktracker, logging to /home/ubuntu/hadoop/libexec/../../lo
gs/hadoop-ubuntu-tasktracker-ip-172-31-29-24.out
172.31.33.156: starting tasktracker, logging to /home/ubuntu/hadoop/libexec/../../l
ogs/hadoop-ubuntu-tasktracker-ip-172-31-33-156.out
ubuntu@ip-172-31-38-5:~/hadoop/conf$ jps
2162 JobTracker
2067 SecondaryNameNode
2278 Jps
1823 NameNode
ubuntu@ip-172-31-38-5:~/hadoop/conf$
```

Created wordcountdir in HDFS system using *hadoop fs -mkdir wordcountdir* Then copied Input Folder from Ubuntu File System to HDFS created folder. This, we can see by *hadoop fs -lsr* . Jar files we can see by ls command in hadoop folder.

```
ubuntu@ip-172-31-38-5: ~/hadoop
ubuntu@ip-172-31-38-5:~/hadoop/conf$ cd ..
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop fs -lsr
lsr: Cannot access .: No such file or directory.
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop fs -mkdir wordcountdir
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop fs -lsr
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:32 /user/ubuntu/wordcountdir
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop fs -put /home/ubuntu/Inputs /user/ubuntu/wordcountdir
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop fs -lsr
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:34 /user/ubuntu/wordcountdir
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs
-rw-r--r-- 2 ubuntu supergroup 95 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs/wordinput1.txt
-rw-r--r-- 2 ubuntu supergroup 1513 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs/wordinput2.txt
-rw-r--r-- 2 ubuntu supergroup 4873 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs/wordinput3.txt
ubuntu@ip-172-31-38-5:~/hadoop$ ls
bin docs hadoop-test-1.2.1.jar LICENSE.txt README.txt votecountMR.jar
build.xml hadoop-ant-1.2.1.jar hadoop-tools-1.2.1.jar logs sbin webapps
c++ hadoop-client-1.2.1.jar ivy MaxTemperature.jar share WordCountProgram.jar
CHANGES.txt hadoop-core-1.2.1.jar ivy.xml NOTICE.txt tempapplication.jar WordCountProgramsw.jar
conf hadoop-examples-1.2.1.jar lib MyProgram.jar tempinput1.txt WordCountProgramv2.jar
contrib hadoop-minicluster-1.2.1.jar libexec NOTICE.txt tempinput1.txt WordCountProgramv3.jar
```

Here we ran version1 of the system by giving the Folder containing the input files as input and storing the output in outputv1.

Note:: Comparison between Version1,2,3 is shown in the last slide.

Komuravelly Venkateshwarlu  
Komuravellyv1@udayton.edu  
1015237060

```
ubuntu@ip-172-31-38-5: ~/hadoop
17/10/01 01:55:14 INFO mapred.JobClient: Bytes Written=5195
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop jar WordCountProgramv2.jar WordCountProgramv2.WordCountJob /user/ubuntu/wordcountdir/Inputs /user/ubuntu/outputmv2
Total time ran=685
17/10/01 02:03:10 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
17/10/01 02:03:10 INFO input.FileInputFormat: Total input paths to process : 3
17/10/01 02:03:10 INFO util.NativeCodeLoader: Loaded the native-hadoop library
17/10/01 02:03:10 WARN snappy.LoadSnappy: Snappy native library not loaded
17/10/01 02:03:11 INFO mapred.JobClient: Running job: job_201710010152_0002
17/10/01 02:03:12 INFO mapred.JobClient: map 0% reduce 0%
17/10/01 02:03:18 INFO mapred.JobClient: map 33% reduce 0%
17/10/01 02:03:19 INFO mapred.JobClient: map 100% reduce 0%
17/10/01 02:03:26 INFO mapred.JobClient: map 100% reduce 33%
17/10/01 02:03:27 INFO mapred.JobClient: map 100% reduce 100%
17/10/01 02:03:28 INFO mapred.JobClient: Job complete: job_201710010152_0002
17/10/01 02:03:28 INFO mapred.JobClient: Counters: 29
17/10/01 02:03:28 INFO mapred.JobClient: Map-Reduce Framework
17/10/01 02:03:28 INFO mapred.JobClient: Spilled Records=1576
17/10/01 02:03:28 INFO mapred.JobClient: Map output materialized bytes=13238
17/10/01 02:03:28 INFO mapred.JobClient: Reduce input records=788
17/10/01 02:03:28 INFO mapred.JobClient: Virtual memory (bytes) snapshot=7612669952
17/10/01 02:03:28 INFO mapred.JobClient: Map input records=37
17/10/01 02:03:28 INFO mapred.JobClient: SPLIT_RAW_BYTES=405
17/10/01 02:03:28 INFO mapred.JobClient: Map output bytes=11644
17/10/01 02:03:28 INFO mapred.JobClient: Reduce shuffle bytes=13238
17/10/01 02:03:28 INFO mapred.JobClient: Physical memory (bytes) snapshot=597327872
17/10/01 02:03:28 INFO mapred.JobClient: Reduce input groups=528
17/10/01 02:03:28 INFO mapred.JobClient: Combine output records=0
17/10/01 02:03:28 INFO mapred.JobClient: Reduce output records=528
17/10/01 02:03:28 INFO mapred.JobClient: Map output records=788
17/10/01 02:03:28 INFO mapred.JobClient: Combine input records=0
17/10/01 02:03:28 INFO mapred.JobClient: CPU time spent (ms)=1110
17/10/01 02:03:28 INFO mapred.JobClient: Total committed heap usage (bytes)=498544640
17/10/01 02:03:28 INFO mapred.JobClient: File Input Format Counters
17/10/01 02:03:28 INFO mapred.JobClient: Bytes Read=6481
17/10/01 02:03:28 INFO mapred.JobClient: FileSystemCounters
17/10/01 02:03:28 INFO mapred.JobClient: HDFS_BYTES_READ=6886
17/10/01 02:03:28 INFO mapred.JobClient: FILE_BYTES_WRITTEN=53921
17/10/01 02:03:28 INFO mapred.JobClient: FILE_BYTES_READ=13238
17/10/01 02:03:28 INFO mapred.JobClient: HDFS_BYTES_WRITTEN=5195
17/10/01 02:03:28 INFO mapred.JobClient: Job Counters
17/10/01 02:03:28 INFO mapred.JobClient: Launched map tasks=3
17/10/01 02:03:28 INFO mapred.JobClient: Launched reduce tasks=1
17/10/01 02:03:28 INFO mapred.JobClient: SLOTS_MILLIS_REDUCES=9240
17/10/01 02:03:28 INFO mapred.JobClient: Total time spent by all reduces waiting after reserving slots (ms)=0
```

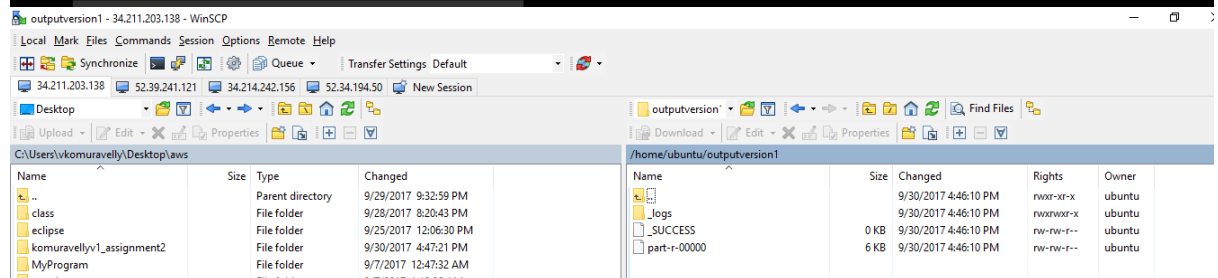
```
17/09/30 20:37:15 INFO mapred.JobClient: Job Counters
17/09/30 20:37:15 INFO mapred.JobClient: Launched map tasks=3
17/09/30 20:37:15 INFO mapred.JobClient: Launched reduce tasks=1
17/09/30 20:37:15 INFO mapred.JobClient: SLOTS_MILLIS_REDUCES=8993
17/09/30 20:37:15 INFO mapred.JobClient: Total time spent by all reduces waiting after reserving slots (ms)=0
17/09/30 20:37:15 INFO mapred.JobClient: SLOTS_MILLIS_MAPS=13416
17/09/30 20:37:15 INFO mapred.JobClient: Total time spent by all maps waiting after reserving slots (ms)=0
17/09/30 20:37:15 INFO mapred.JobClient: Data-local map tasks=3
17/09/30 20:37:15 INFO mapred.JobClient: File Output Format Counters
17/09/30 20:37:15 INFO mapred.JobClient: Bytes Written=5195
ubuntu@ip-172-31-38-5:~/hadoop$
```

```
ubuntu@ip-172-31-38-5: ~/hadoop
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop fs -lsr
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:37 /user/ubuntu/outputv1
-rw-r--r-- 2 ubuntu supergroup 0 2017-09-30 20:37 /user/ubuntu/outputv1/_SUCCESS
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:36 /user/ubuntu/outputv1/_logs
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:36 /user/ubuntu/outputv1/_logs/history
-rw-r--r-- 2 ubuntu supergroup 19989 2017-09-30 20:36 /user/ubuntu/outputv1/_logs/history/job_201709302027_0001_1506803818322_ubuntu_WordCountProblem-v1
-rw-r--r-- 2 ubuntu supergroup 49134 2017-09-30 20:36 /user/ubuntu/outputv1/_logs/history/job_201709302027_0001_conf.xml
-rw-r--r-- 2 ubuntu supergroup 5195 2017-09-30 20:37 /user/ubuntu/outputv1/part-r-00000
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:34 /user/ubuntu/wordcountdir
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs
-rw-r--r-- 2 ubuntu supergroup 96 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs/wordinput1.txt
-rw-r--r-- 2 ubuntu supergroup 1513 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs/wordinput2.txt
-rw-r--r-- 2 ubuntu supergroup 4873 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs/wordinput3.txt
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop fs -cat /user/ubuntu/outputv1/part-r-00000
18
(but 1
, 2
3rd-party 1
A 4
API 2
API, 1
An 1
Append. 1
B 2
Because 1
Big 1
But 1
C#, 1
C++, 1
Clients 1
Cloudera, 1
Cocoa, 1
College 1
Data 2
Datadog, 1
Dayton 1
Due 1
Each 1
Erlang, 1
Failure 1
Federation, 1
File 1
Filesystem 1
```

```
ubuntu@ip-172-31-38-5: ~/hadoop
trade-off      1
traffic 1
transactional  1
transfer.      1
tweets 2
two 1
types 1
typically      1
underlying     2
unnecessary    1
up 1
up-to-date     1
use 2
used 2
useful. 1
uses 1
using 2
utilize 1
value 1
variety 1
various 2
venkat 1
version 1
via 1
vibration,     1
virtual 1
voltage 1
was 1
web 1
when 1
whether 1
which 2
whole 1
widespread     1
with 7
within 1
without 1
work 1
would 1
write-operations. 1
written 1
you 1
ubuntu@ip-172-31-38-5:~/hadoop$
```

Copied the output file created to Ubuntu system by using get command.

```
write-operations.      1
written 1
you 1
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop fs -get /user/ubuntu/outputv1 /home/ubuntu/outputversion1
ubuntu@ip-172-31-38-5:~/hadoop$
```



The screenshot shows the WinSCP interface with two panes. The left pane displays the local file system at 'C:\Users\vkumaravelly\Desktop\aws', and the right pane displays the remote file system at '/home/ubuntu/outputversion1'. Both panes show a list of files and folders with columns for Name, Size, Type, Changed, Rights, and Owner.

Name	Size	Type	Changed	Rights	Owner
..		Parent directory	9/29/2017 9:32:59 PM		
class		File folder	9/28/2017 8:20:43 PM		ubuntu
eclipse		File folder	9/25/2017 12:06:30 PM		ubuntu
komuravellyv1_assignment2		File folder	9/30/2017 4:47:21 PM		ubuntu
MyProgram		File folder	9/7/2017 12:47:32 AM		ubuntu
nmn1		File folder	0/7/2017 1:13:35 AM		

Name	Size	Changed	Rights	Owner
..		9/30/2017 4:46:10 PM	rw-r-xr-x	ubuntu
_logs		9/30/2017 4:46:10 PM	rw-rw-r-x	ubuntu
_SUCCESS	0 KB	9/30/2017 4:46:10 PM	rw-rw-r--	ubuntu
part-r-00000	6 KB	9/30/2017 4:46:10 PM	rw-rw-r--	ubuntu

## Version-2

Like same as version 1, Version 2 is ran. But in version 2 I used HashMap to store the map output instead of sending the map output directly to reducer. By using HashMap method I made the Key as unique so that key-value pairs are reduced. Adding the value for the same key makes few number of key-value pairs.

Komuravelly Venkateshwarlu  
Komuravellyv1@udayton.edu  
1015237060

```
ubuntu@ip-172-31-38-5: ~/hadoop
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop jar WordCountProgram.jar WordCountProgram.WordCountJob /user/ubuntu/wordcountdir/Inputs /user/ubuntu/outputmv1
Total time ran671
17/10/01 01:54:56 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
17/10/01 01:54:56 INFO input.FileInputFormat: Total input paths to process : 3
17/10/01 01:54:56 INFO util.NativeCodeLoader: Loaded the native-hadoop library
17/10/01 01:54:56 WARN snappy.LoadSnappy: Snappy native library not loaded
17/10/01 01:54:57 INFO mapred.JobClient: Running job: job_201710010152_0001
17/10/01 01:54:58 INFO mapred.JobClient: map 0% reduce 0%
17/10/01 01:55:04 INFO mapred.JobClient: map 33% reduce 0%
17/10/01 01:55:06 INFO mapred.JobClient: map 100% reduce 0%
17/10/01 01:55:12 INFO mapred.JobClient: map 100% reduce 33%
17/10/01 01:55:13 INFO mapred.JobClient: map 100% reduce 100%
17/10/01 01:55:14 INFO mapred.JobClient: Job completer: job_201710010152_0001
17/10/01 01:55:14 INFO mapred.JobClient: Counters: 29
17/10/01 01:55:14 INFO mapred.JobClient: Map-Reduce Framework
17/10/01 01:55:14 INFO mapred.JobClient: Spilled Records=2084
17/10/01 01:55:14 INFO mapred.JobClient: Map output materialized bytes=16883
17/10/01 01:55:14 INFO mapred.JobClient: Reduce input records=1042
17/10/01 01:55:14 INFO mapred.JobClient: Virtual memory (bytes) snapshot=7612661760
17/10/01 01:55:14 INFO mapred.JobClient: Map input records=37
17/10/01 01:55:14 INFO mapred.JobClient: SPLIT_RAW_BYTES=405
17/10/01 01:55:14 INFO mapred.JobClient: Map output bytes=14781
17/10/01 01:55:14 INFO mapred.JobClient: Reduce shuffle bytes=16883
17/10/01 01:55:14 INFO mapred.JobClient: Physical memory (bytes) snapshot=597430272
17/10/01 01:55:14 INFO mapred.JobClient: Reduce input groups=529
17/10/01 01:55:14 INFO mapred.JobClient: Combine output records=0
17/10/01 01:55:14 INFO mapred.JobClient: Reduce output records=529
17/10/01 01:55:14 INFO mapred.JobClient: Map output records=1042
17/10/01 01:55:14 INFO mapred.JobClient: Combine input records=0
17/10/01 01:55:14 INFO mapred.JobClient: CPU time spent (ms)=21050
17/10/01 01:55:14 INFO mapred.JobClient: Total Committed heap usage (bytes)=498544640
17/10/01 01:55:14 INFO mapred.JobClient: File Input Format Counters
17/10/01 01:55:14 INFO mapred.JobClient: Bytes Read=6481
17/10/01 01:55:14 INFO mapred.JobClient: FileSystemCounters
17/10/01 01:55:14 INFO mapred.JobClient: HDFS_BYTES_READ=6886
17/10/01 01:55:14 INFO mapred.JobClient: FILE_BYTES_WRITTEN=261175
17/10/01 01:55:14 INFO mapred.JobClient: FILE_BYTES_READ=16871
17/10/01 01:55:14 INFO mapred.JobClient: HDFS_BYTES_WRITTEN=5195
17/10/01 01:55:14 INFO mapred.JobClient: Job Counters
17/10/01 01:55:14 INFO mapred.JobClient: Launched map tasks=3
17/10/01 01:55:14 INFO mapred.JobClient: Launched reduce tasks=1
17/10/01 01:55:14 INFO mapred.JobClient: SLOTS_MILLIS_REDUCES=9122
17/10/01 01:55:14 INFO mapred.JobClient: Total time spent by all reduces waiting after reserving slots (ms)=0
17/10/01 01:55:14 INFO mapred.JobClient: SLOTS_MILLIS_MAPS=14827

17/09/30 20:50:53 INFO mapred.JobClient: Job Counters
17/09/30 20:50:53 INFO mapred.JobClient: Launched map tasks=3
17/09/30 20:50:53 INFO mapred.JobClient: Launched reduce tasks=1
17/09/30 20:50:53 INFO mapred.JobClient: SLOTS_MILLIS_REDUCES=8963
17/09/30 20:50:53 INFO mapred.JobClient: Total time spent by all reduces waiting after reserving slots (ms)=0
17/09/30 20:50:53 INFO mapred.JobClient: SLOTS_MILLIS_MAPS=12331
17/09/30 20:50:53 INFO mapred.JobClient: Total time spent by all maps waiting after reserving slots (ms)=0
17/09/30 20:50:53 INFO mapred.JobClient: Data-local map tasks=3
17/09/30 20:50:53 INFO mapred.JobClient: File Output Format Counters
17/09/30 20:50:53 INFO mapred.JobClient: Bytes Written=5191
ubuntu@ip-172-31-38-5:~/hadoop$
```

```
ubuntu@ip-172-31-38-5: ~/hadoop
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop fs -lsr
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:37 /user/ubuntu/outputv1
-rw-r--r-- 2 ubuntu supergroup 0 2017-09-30 20:37 /user/ubuntu/outputv1/_SUCCESS
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:36 /user/ubuntu/outputv1/_logs
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:36 /user/ubuntu/outputv1/_logs/history
-rw-r--r-- 2 ubuntu supergroup 19989 2017-09-30 20:36 /user/ubuntu/outputv1/_logs/history/job_201709302027_0001_1506803818322_ubuntu_WordCountProblem-v1
-rw-r--r-- 2 ubuntu supergroup 49134 2017-09-30 20:36 /user/ubuntu/outputv1/_logs/history/job_201709302027_0001_conf.xml
-rw-r--r-- 2 ubuntu supergroup 5195 2017-09-30 20:37 /user/ubuntu/outputv1/part-r-00000
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:50 /user/ubuntu/outputv2
-rw-r--r-- 2 ubuntu supergroup 0 2017-09-30 20:50 /user/ubuntu/outputv2/_SUCCESS
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:50 /user/ubuntu/outputv2/_logs
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:50 /user/ubuntu/outputv2/_logs/history
-rw-r--r-- 2 ubuntu supergroup 19975 2017-09-30 20:50 /user/ubuntu/outputv2/_logs/history/job_201709302027_0002_1506804637367_ubuntu_WordCountProblem-v2
-rw-r--r-- 2 ubuntu supergroup 49142 2017-09-30 20:50 /user/ubuntu/outputv2/_logs/history/job_201709302027_0002_conf.xml
-rw-r--r-- 2 ubuntu supergroup 5191 2017-09-30 20:50 /user/ubuntu/outputv2/part-r-00000
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:34 /user/ubuntu/wordcountdir
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs
-rw-r--r-- 2 ubuntu supergroup 95 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs/wordinput1.txt
-rw-r--r-- 2 ubuntu supergroup 1513 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs/wordinput2.txt
-rw-r--r-- 2 ubuntu supergroup 4873 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs/wordinput3.txt
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop fs -cat /user/ubuntu/outputv2/part-r-00000
(But
1
,
2
3rd-party
A
4
API
2
API,
1
An
1
Append.
1
B
2
Because
1
Big
1
But
1
C#,
1
C++,
1
Clients
1
Cloudera,
1
Cocoa,
1
College
1
Data
2
Datadog,
1
Dayton
1
Due
1
```

Komuravelly Venkateshwarlu  
Komuravellyv1@udayton.edu  
1015237060

```
ubuntu@ip-172-31-38-5: ~/hadoop
trade-off 1
traffic 1
transactional 1
transfer 1
tweets 2
two 1
types 1
typically 1
underlying 2
unnecessary 1
up 1
up-to-date 1
use 2
used 2
useful 1
uses 1
using 2
utilize 1
value 1
variety 1
various 2
venkat 1
version 1
via 1
vibration, 1
virtual 1
voltage 1
was 1
web 1
when 1
whether 1
which 2
whole 1
widespread 1
with 7
within 1
without 1
work 1
would 1
write-operations. 1
written 1
you 1
ubuntu@ip-172-31-38-5:~/hadoop$
```

```
work 1
would 1
write-operations. 1
written 1
you 1
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop fs -get /user/ubuntu/outputv2 /home/ubuntu/outputversion2
ubuntu@ip-172-31-38-5:~/hadoop$
```

outputversion2 - 34.211.203.138 - WinSCP

C:\Users\vkumaravelly\Desktop\aws					/home/ubuntu/outputversion2				
Name	Size	Type	Changed		Name	Size	Changed	Rights	Owner
.		Parent directory	9/29/2017 9:32:59 PM		._logs		9/30/2017 4:57:15 PM	rwxr-xr-x	ubuntu
class		File folder	9/28/2017 8:20:43 PM		._SUCCESS	0 KB	9/30/2017 4:57:15 PM	rw-rw-r--	ubuntu
eclipse		File folder	9/30/2017 12:06:30 PM		part-r-00000	6 KB	9/30/2017 4:57:15 PM	rw-rw-r--	ubuntu
komuravellyv1_assignment2		File folder	9/30/2017 4:47:21 PM						
MyProgram		File folder	9/7/2017 12:47:32 AM						
pgm1		File folder	9/7/2017 1:13:25 AM						
.									

## Version-3

In Version 3 , Using CleanUp Method, result of reduced key-value pairs from map (used HashMap method) are stored in MemoryBuffer. Same mapper is used for the next input chunk. In this way In total there will be few key-value pairs for the reducer to process.



Komuravelly Venkateshwarlu  
Komuravellyv1@udayton.edu  
1015237060

```
ubuntu@ip-172-31-38-5: ~/hadoop
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop jar WordCountProgram3.jar WordCountProgram3.WordCountJob /user/ubuntu/wordcountdir/Inputs /user/ubuntu/outputmv3
Total time ran620
17/10/01 02:07:43 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
17/10/01 02:07:43 INFO input.FileInputFormat: Total input paths to process : 3
17/10/01 02:07:43 INFO util.NativeCodeLoader: Loaded the native-hadoop library
17/10/01 02:07:43 WARN snappy.LoadSnappy: Snappy native library not loaded
17/10/01 02:07:43 INFO mapred.JobClient: Running job: job_201710010152_0003
17/10/01 02:07:44 INFO mapred.JobClient: map 0% reduce 0%
17/10/01 02:07:49 INFO mapred.JobClient: map 33% reduce 0%
17/10/01 02:07:51 INFO mapred.JobClient: map 100% reduce 0%
17/10/01 02:07:57 INFO mapred.JobClient: map 100% reduce 33%
17/10/01 02:07:59 INFO mapred.JobClient: map 100% reduce 100%
17/10/01 02:08:00 INFO mapred.JobClient: Job completer: job_201710010152_0003
17/10/01 02:08:00 INFO mapred.JobClient: Counters: 29
17/10/01 02:08:00 INFO mapred.JobClient: Map-Reduce Framework
17/10/01 02:08:00 INFO mapred.JobClient: Spilled Records=1166
17/10/01 02:08:00 INFO mapred.JobClient: Map output materialized bytes=10239
17/10/01 02:08:00 INFO mapred.JobClient: Reduce input records=583
17/10/01 02:08:00 INFO mapred.JobClient: Virtual memory (bytes) snapshot=7612932096
17/10/01 02:08:00 INFO mapred.JobClient: Map input records=37
17/10/01 02:08:00 INFO mapred.JobClient: SPLIT_RAW_BYTES=405
17/10/01 02:08:00 INFO mapred.JobClient: Map output bytes=9055
17/10/01 02:08:00 INFO mapred.JobClient: Reduce shuffle bytes=10239
17/10/01 02:08:00 INFO mapred.JobClient: Physical memory (bytes) snapshot=598740992
17/10/01 02:08:00 INFO mapred.JobClient: Reduce input groups=528
17/10/01 02:08:00 INFO mapred.JobClient: Combine output records=0
17/10/01 02:08:00 INFO mapred.JobClient: Reduce output records=528
17/10/01 02:08:00 INFO mapred.JobClient: Map output records=583
17/10/01 02:08:00 INFO mapred.JobClient: Combine input records=0
17/10/01 02:08:00 INFO mapred.JobClient: CPU time spent (ms)=21160
17/10/01 02:08:00 INFO mapred.JobClient: Total Committed heap usage (bytes)=49854640
17/10/01 02:08:00 INFO mapred.JobClient: File Input Format Counters
17/10/01 02:08:00 INFO mapred.JobClient: Bytes Read=6481
17/10/01 02:08:00 INFO mapred.JobClient: FileSystemCounters
17/10/01 02:08:00 INFO mapred.JobClient: HDFS_BYTES_READ=6886
17/10/01 02:08:00 INFO mapred.JobClient: FILE_BYTES_WRITTEN=247923
17/10/01 02:08:00 INFO mapred.JobClient: FILE_BYTES_READ=10237
17/10/01 02:08:00 INFO mapred.JobClient: HDFS_BYTES_WRITTEN=5191
17/10/01 02:08:00 INFO mapred.JobClient: Job Counters
17/10/01 02:08:00 INFO mapred.JobClient: Launched map tasks=3
17/10/01 02:08:00 INFO mapred.JobClient: Launched reduce tasks=1
17/10/01 02:08:00 INFO mapred.JobClient: SLOTS_MILLIS_REDUCES=9177
17/10/01 02:08:00 INFO mapred.JobClient: Total time spent by all reduces waiting after reserving slots (ms)=0
17/10/01 02:08:00 INFO mapred.JobClient: SLOTS_MILLIS_MAPS=12824
17/09/30 21:32:15 INFO mapred.JobClient: Job Counters
17/09/30 21:32:15 INFO mapred.JobClient: Launched map tasks=3
17/09/30 21:32:15 INFO mapred.JobClient: Launched reduce tasks=1
17/09/30 21:32:15 INFO mapred.JobClient: SLOTS_MILLIS_REDUCES=8796
17/09/30 21:32:15 INFO mapred.JobClient: Total time spent by all reduces waiting after reserving slots (ms)=0
17/09/30 21:32:15 INFO mapred.JobClient: SLOTS_MILLIS_MAPS=12695
17/09/30 21:32:15 INFO mapred.JobClient: Total time spent by all maps waiting after reserving slots (ms)=0
17/09/30 21:32:15 INFO mapred.JobClient: Data-local map tasks=3
17/09/30 21:32:15 INFO mapred.JobClient: File Output Format Counters
17/09/30 21:32:15 INFO mapred.JobClient: Bytes Written=5191
ubuntu@ip-172-31-38-5:~/hadoop$
```

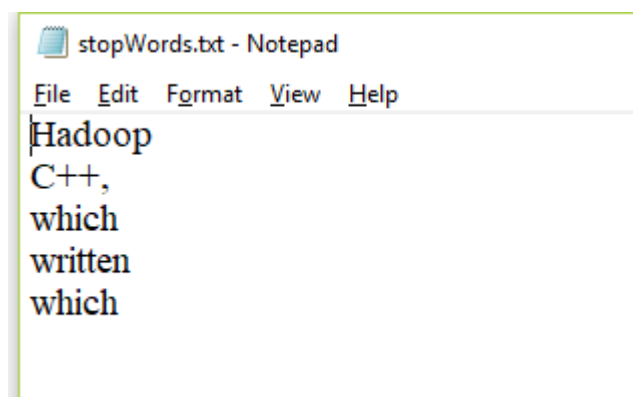
```
ubuntu@ip-172-31-38-5: ~
ubuntu@ip-172-31-38-5:~$ jps
2162 JobTracker
2067 SecondaryNameNode
3530 Jps
1823 NameNode
ubuntu@ip-172-31-38-5:~$ hadoop fs -lsr
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:37 /user/ubuntu/outputv1
-rw-r--r-- 2 ubuntu supergroup 0 2017-09-30 20:37 /user/ubuntu/outputv1/_SUCCESS
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:36 /user/ubuntu/outputv1/_logs
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:36 /user/ubuntu/outputv1/_logs/history
19989 2017-09-30 20:36 /user/ubuntu/outputv1/_logs/history/job_201709302027_0001_1506803818322_ubuntu_WordCountProblem-v1
49134 2017-09-30 20:36 /user/ubuntu/outputv1/_logs/history/job_201709302027_0001_conf.xml
5195 2017-09-30 20:37 /user/ubuntu/outputv1/part-r-00000
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:50 /user/ubuntu/outputv2
-rw-r--r-- 2 ubuntu supergroup 0 2017-09-30 20:50 /user/ubuntu/outputv2/_SUCCESS
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:50 /user/ubuntu/outputv2/_logs
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:50 /user/ubuntu/outputv2/_logs/history
19975 2017-09-30 20:50 /user/ubuntu/outputv2/_logs/history/job_201709302027_0002_1506804637367_ubuntu_WordCountProblem-v2
49142 2017-09-30 20:50 /user/ubuntu/outputv2/_logs/history/job_201709302027_0002_conf.xml
5191 2017-09-30 20:50 /user/ubuntu/outputv2/part-r-00000
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:59 /user/ubuntu/outputv3
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:59 /user/ubuntu/outputv3/_logs
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:59 /user/ubuntu/outputv3/_logs/history
20622 2017-09-30 20:59 /user/ubuntu/outputv3/_logs/history/job_201709302027_0003_1506805165327_ubuntu_WordCountProblem-v3
49142 2017-09-30 20:59 /user/ubuntu/outputv3/_logs/history/job_201709302027_0003_conf.xml
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 21:32 /user/ubuntu/outputv3/_SUCCESS
-rw-r--r-- 2 ubuntu supergroup 0 2017-09-30 21:32 /user/ubuntu/outputv3/_logs
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 21:31 /user/ubuntu/outputv3/_logs/history
19973 2017-09-30 21:31 /user/ubuntu/outputv3/_logs/history/job_201709302027_0005_1506807118905_ubuntu_WordCountProblem-v3
49144 2017-09-30 21:31 /user/ubuntu/outputv3/_logs/history/job_201709302027_0005_conf.xml
5191 2017-09-30 21:32 /user/ubuntu/outputv3/part-r-00000
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs
-rw-r--r-- 2 ubuntu supergroup 95 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs/wordinput1.txt
-rw-r--r-- 2 ubuntu supergroup 1513 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs/wordinput2.txt
-rw-r--r-- 2 ubuntu supergroup 4873 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs/wordinput3.txt
ubuntu@ip-172-31-38-5:~$ hadoop fs -cat /user/ubuntu/outputv3/part-r-00000
```

```
ubuntu@ip-172-31-38-5: ~  
tracking 1  
trade-off 1  
traffic 1  
transactional 1  
transfer 1  
tweets 2  
two 1  
types 1  
typically 1  
underlying 2  
unnecessary 1  
up 1  
up-to-date 1  
use 2  
used 2  
useful 1  
uses 1  
using 2  
utilize 1  
value 1  
variety 1  
various 2  
venkat 1  
version 1  
via 1  
vibration, 1  
virtual 1  
voltage 1  
was 1  
web 1  
when 1  
whether 1  
which 2  
whole 1  
widespread 1  
with 7  
within 1  
without 1  
work 1  
would 1  
write-operations. 1  
written 1  
you 1  
ubuntu@ip-172-31-38-5:~$
```

### StopWords:-

From the output of version 2 , I used below keywords as stopwords and processed using distributed file cache API to remove these words from the output.

Reference: <https://examples.javacodegeeks.com/enterprise-java/apache-hadoop/apache-hadoop-distributed-cache-example/>



stopWords.txt file is stored in HDFS folder.

```
drwxr-xr-x - ubuntu supergroup 0 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs  
-rw-r--r-- 2 ubuntu supergroup 95 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs/wordinput1.txt  
-rw-r--r-- 2 ubuntu supergroup 1513 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs/wordinput2.txt  
-rw-r--r-- 2 ubuntu supergroup 4873 2017-09-30 20:34 /user/ubuntu/wordcountdir/Inputs/wordinput3.txt  
-rw-r--r-- 2 ubuntu supergroup 35 2017-09-30 22:00 /user/ubuntu/wordcountdir/stopWords.txt
```

Ran the MapReduce program for stopWords below.



Komuravelly Venkateshwarlu  
Komuravellyv1@udayton.edu  
1015237060

```
ubuntu@ip-172-31-38-5: ~/hadoop
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop jar WordCountProgramaw.jar WordCountProgramaw.WordCountJob /user/ubuntu/wordcountdir/Inputs /user/ubuntu/outputsw3 /user/ubuntu/wordcountdir/stopWords.txt
17/09/30 22:18:27 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
17/09/30 22:18:27 INFO input.FileInputFormat: Total input paths to process : 3
17/09/30 22:18:27 INFO util.NativeCodeLoader: Loaded the native-hadoop library
17/09/30 22:18:27 WARN snappy.LoadSnappy: Snappy native library not loaded
17/09/30 22:18:28 INFO mapred.JobClient: Running job: job_201709302027_0012
17/09/30 22:18:29 INFO mapred.JobClient: map 0% reduce 0%
17/09/30 22:18:34 INFO mapred.JobClient: map 33% reduce 0%
17/09/30 22:18:36 INFO mapred.JobClient: map 100% reduce 0%
17/09/30 22:18:42 INFO mapred.JobClient: map 100% reduce 33%
17/09/30 22:18:43 INFO mapred.JobClient: map 100% reduce 100%
17/09/30 22:18:44 INFO mapred.JobClient: Job complete: job_201709302027_0012

ubuntu@ip-172-31-38-5: ~/hadoop
17/09/30 22:18:27 WARN snappy.LoadSnappy: Snappy native library not loaded
17/09/30 22:18:28 INFO mapred.JobClient: Running job: job_201709302027_0012
17/09/30 22:18:29 INFO mapred.JobClient: map 0% reduce 0%
17/09/30 22:18:34 INFO mapred.JobClient: map 33% reduce 0%
17/09/30 22:18:36 INFO mapred.JobClient: map 100% reduce 0%
17/09/30 22:18:42 INFO mapred.JobClient: map 100% reduce 33%
17/09/30 22:18:43 INFO mapred.JobClient: map 100% reduce 100%
17/09/30 22:18:44 INFO mapred.JobClient: Job complete: job_201709302027_0012
17/09/30 22:18:44 INFO mapred.JobClient: Counters: 29
17/09/30 22:18:44 INFO mapred.JobClient: Map-Reduce Framework
17/09/30 22:18:44 INFO mapred.JobClient: Spilled Records=2032
17/09/30 22:18:44 INFO mapred.JobClient: Map output materialized bytes=12488
17/09/30 22:18:44 INFO mapred.JobClient: Reduce input records=1016
17/09/30 22:18:44 INFO mapred.JobClient: Virtual memory (bytes) snapshot=7612641280
17/09/30 22:18:44 INFO mapred.JobClient: Map input records=37
17/09/30 22:18:44 INFO mapred.JobClient: SPLIT_RAW_BYTES=405
17/09/30 22:18:44 INFO mapred.JobClient: Map output bytes=10438
17/09/30 22:18:44 INFO mapred.JobClient: Reduce shuffle bytes=12488
17/09/30 22:18:44 INFO mapred.JobClient: Physical memory (bytes) snapshot=599322624
17/09/30 22:18:44 INFO mapred.JobClient: Reduce input groups=524
17/09/30 22:18:44 INFO mapred.JobClient: Combine output records=0
17/09/30 22:18:44 INFO mapred.JobClient: Reduce output records=524
17/09/30 22:18:44 INFO mapred.JobClient: Map output records=1016
17/09/30 22:18:44 INFO mapred.JobClient: Combine input records=0
17/09/30 22:18:44 INFO mapred.JobClient: CPU time spent (ms)=1050
17/09/30 22:18:44 INFO mapred.JobClient: Total committed heap usage (bytes)=498544640
17/09/30 22:18:44 INFO mapred.JobClient: File Input Format Counters
17/09/30 22:18:44 INFO mapred.JobClient: Bytes Read=6481
17/09/30 22:18:44 INFO mapred.JobClient: FileSystemCounters
17/09/30 22:18:44 INFO mapred.JobClient: HDFS_BYTES_READ=6886
17/09/30 22:18:44 INFO mapred.JobClient: FILE_BYTES_WRITTEN=254379
17/09/30 22:18:44 INFO mapred.JobClient: FILE_BYTES_READ=12476
17/09/30 22:18:44 INFO mapred.JobClient: HDFS_BYTES_WRITTEN=5157
17/09/30 22:18:44 INFO mapred.JobClient: Job Counters
17/09/30 22:18:44 INFO mapred.JobClient: Launched map tasks=3
17/09/30 22:18:44 INFO mapred.JobClient: Launched reduce tasks=1
17/09/30 22:18:44 INFO mapred.JobClient: SLOTS_MILLIS_REDUCES=8948
17/09/30 22:18:44 INFO mapred.JobClient: Total time spent by all reduces waiting after reserving slots (ms)=0
17/09/30 22:18:44 INFO mapred.JobClient: SLOTS_MILLIS_MAPS=12486
17/09/30 22:18:44 INFO mapred.JobClient: Total time spent by all maps waiting after reserving slots (ms)=0
17/09/30 22:18:44 INFO mapred.JobClient: Data-local map tasks=3
17/09/30 22:18:44 INFO mapred.JobClient: File Output Format Counters
17/09/30 22:18:44 INFO mapred.JobClient: Bytes Written=5157
Job was successful

drwxr-xr-x 2 ubuntu supergroup 0 2017-09-30 22:18 /user/ubuntu/outputsw3
-rw-r--r-- 2 ubuntu supergroup 0 2017-09-30 22:18 /user/ubuntu/outputsw3/_SUCCESS
drwxr-xr-x 2 ubuntu supergroup 0 2017-09-30 22:18 /user/ubuntu/outputsw3/_logs
drwxr-xr-x 2 ubuntu supergroup 0 2017-09-30 22:18 /user/ubuntu/outputsw3/_logs/history
-rw-r--r-- 20002 2017-09-30 22:18 /user/ubuntu/outputsw3/_logs/history/job_201709302027_0012_1506809907989_ubuntu_Word+Counter+With+Stop+WordsRemoval
-rw-r--r-- 2 ubuntu supergroup 49277 2017-09-30 22:18 /user/ubuntu/outputsw3/_logs/history/job_201709302027_0012_conf.xml
-rw-r--r-- 2 ubuntu supergroup 5157 2017-09-30 22:18 /user/ubuntu/outputsw3/part-r-00000

ubuntu@ip-172-31-38-5:~/hadoop$ hadoop fs -get /user/ubuntu/outputsw3 /home/ubuntu/outputstopwords
ubuntu@ip-172-31-38-5:~/hadoop$
```

Here, We can see the words removed from output.

The image shows two Notepad++ windows side-by-side, both titled 'C:\Users\vkomuravelly\Desktop\av'. The left window has tabs for 'hadoop\_basic\_examples' and 'java\_c', and the right window has tabs for 'hadoop\_basic\_examples' and 'java\_c'. Both windows display a list of words and their counts, with some lines highlighted in blue.

Left Window (hadoop\_basic\_examples):

```

1 (but 1
2 , 2
3 3rd-party 1
4 A 4
5 API 2
6 API, 1
7 An 1
8 Append. 1
9 B 2
10 Because 1
11 Big 1
12 But 1
13 C#, 1
14 C++, 1
15 Clients 1
16 Cloudera, 1
17 Cocoa, 1
18 College 1
19 Data 2
20 Datadog. 1
21 Dayton 1
22 Due 1
23 Each 1
24 Failure 1
25 via 1
510 vibration, 1
511 virtual 1
512 voltage 1
513 was 1
514 web 1
515 when 1
516 whether 1
517 which 2
518 whole 1
519 widespread 1
520 with 7
521 within 1
522 without 1
523 work 1
524 would 1
525 write-operations. 1
526 written 1
527 you 1
528
529
  
```

Right Window (hadoop\_basic\_examples):

```

1 (but 1
2 , 2
3 3rd-party 1
4 A 4
5 API 2
6 API, 1
7 An 1
8 Append. 1
9 B 2
10 Because 1
11 Big 1
12 But 1
13 C#, 1
14 Clients 1
15 Cloudera, 1
16 Cocoa, 1
17 College 1
18 Data 2
19 Datadog. 1
20 Dayton 1
21 Due 1
22 Each 1
23 Erlang, 1
24 Failure 1
  
```

Below the right window, there is a partial view of another window showing lines 510-525:

```

510 virtual 1
511 voltage 1
512 was 1
513 web 1
514 when 1
515 whether 1
516 whole 1
517 widespread 1
518 with 7
519 within 1
520 without 1
521 work 1
522 would 1
523 write-operations. 1
524 you 1
525
  
```

Comparison between v1,v2,v3.  
 Execution times=  
 V1→685msec V2→671msec V3→620msec

Komuravelly Venkateshwarlu  
Komuravellyv1@udayton.edu  
1015237060

```
ubuntu@ip-172-31-38-5: ~/hadoop
17/10/01 01:55:14 INFO mapred.JobClient: Bytes Written=5195
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop jar WordCountProgramv2.jar WordCountProgramv2.WordC
Total time ran 685
17/10/01 02:03:10 WARN mapred.JobClient: Use GenericOptionsParser for parsing the argument
17/10/01 02:03:10 INFO input.FileInputFormat: Total input paths to process : 3
17/10/01 02:03:10 INFO util.NativeCodeLoader: Loaded the native-hadoop library
17/10/01 02:03:10 WARN snappy.LoadSnappy: Snappy native library not loaded
17/10/01 02:03:11 INFO mapred.JobClient: Running job: job_201710010152_0002
17/10/01 02:03:12 INFO mapred.JobClient: map 0% reduce 0%
17/10/01 02:03:18 INFO mapred.JobClient: map 33% reduce 0%
17/10/01 02:03:19 INFO mapred.JobClient: map 100% reduce 0%
17/10/01 02:03:26 INFO mapred.JobClient: map 100% reduce 33%
17/10/01 02:03:27 INFO mapred.JobClient: map 100% reduce 100%
17/10/01 02:03:28 INFO mapred.JobClient: Job complete: job_201710010152_0002
17/10/01 02:03:28 INFO mapred.JobClient: Counters: 29
17/10/01 02:03:28 INFO mapred.JobClient: Map-Reduce Framework
17/10/01 02:03:28 INFO mapred.JobClient: Spilled Records=1576
17/10/01 02:03:28 INFO mapred.JobClient: Map output materialized bytes=13238
17/10/01 02:03:28 INFO mapred.JobClient: Reduce input records=788
17/10/01 02:03:28 INFO mapred.JobClient: Virtual memory (bytes) snapshot=7612669952
17/10/01 02:03:28 INFO mapred.JobClient: Map input records=37
17/10/01 02:03:28 INFO mapred.JobClient: SPLIT_RAW_BYTES=405
17/10/01 02:03:28 INFO mapred.JobClient: Map output bytes=11644
17/10/01 02:03:28 INFO mapred.JobClient: Reduce shuffle bytes=13238
17/10/01 02:03:28 INFO mapred.JobClient: Physical memory (bytes) snapshot=597327872
17/10/01 02:03:28 INFO mapred.JobClient: Reduce input groups=528
17/10/01 02:03:28 INFO mapred.JobClient: Combine output records=0
17/10/01 02:03:28 INFO mapred.JobClient: Reduce output records=528
17/10/01 02:03:28 INFO mapred.JobClient: Map output records=788
17/10/01 02:03:28 INFO mapred.JobClient: Combine input records=0
17/10/01 02:03:28 INFO mapred.JobClient: CPU time spent (ms)=1110
17/10/01 02:03:28 INFO mapred.JobClient: Total committed heap usage (bytes)=498544640
17/10/01 02:03:28 INFO mapred.JobClient: File Input Format Counters
17/10/01 02:03:28 INFO mapred.JobClient: Bytes Read=6481
17/10/01 02:03:28 INFO mapred.JobClient: FileSystemCounters
17/10/01 02:03:28 INFO mapred.JobClient: HDFS_BYTES_READ=6886
17/10/01 02:03:28 INFO mapred.JobClient: FILE_BYTES_WRITTEN=253921
17/10/01 02:03:28 INFO mapred.JobClient: FILE_BYTES_READ=13226
17/10/01 02:03:28 INFO mapred.JobClient: HDFS_BYTES_WRITTEN=5191
17/10/01 02:03:28 INFO mapred.JobClient: Job Counters
17/10/01 02:03:28 INFO mapred.JobClient: Launched map tasks=3
```

Komuravelly Venkateshwarlu  
Komuravellyv1@udayton.edu  
1015237060

```
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop jar WordCountProgram.jar WordCountProgram.WordC
Total time ran 671
17/10/01 01:54:56 WARN mapred.JobClient: Use GenericOptionsParser for parsing the argum
17/10/01 01:54:56 INFO input.FileInputFormat: Total input paths to process : 3
17/10/01 01:54:56 INFO util.NativeCodeLoader: Loaded the native-hadoop library
17/10/01 01:54:56 WARN snappy.LoadSnappy: Snappy native library not loaded
17/10/01 01:54:57 INFO mapred.JobClient: Running job: job_201710010152_0001
17/10/01 01:54:58 INFO mapred.JobClient: map 0% reduce 0%
17/10/01 01:55:04 INFO mapred.JobClient: map 33% reduce 0%
17/10/01 01:55:06 INFO mapred.JobClient: map 100% reduce 0%
17/10/01 01:55:12 INFO mapred.JobClient: map 100% reduce 33%
17/10/01 01:55:13 INFO mapred.JobClient: map 100% reduce 100%
17/10/01 01:55:14 INFO mapred.JobClient: Job complete: job_201710010152_0001
17/10/01 01:55:14 INFO mapred.JobClient: Counters: 29
17/10/01 01:55:14 INFO mapred.JobClient:   Map-Reduce Framework
17/10/01 01:55:14 INFO mapred.JobClient:     Spilled Records=2084
17/10/01 01:55:14 INFO mapred.JobClient:     Map output materialized bytes=16883
17/10/01 01:55:14 INFO mapred.JobClient:     Reduce input records=1042
17/10/01 01:55:14 INFO mapred.JobClient:     Virtual memory (bytes) snapshot=761266176
17/10/01 01:55:14 INFO mapred.JobClient:     Map input records=37
17/10/01 01:55:14 INFO mapred.JobClient:     SPLIT_RAW_BYTES=405
17/10/01 01:55:14 INFO mapred.JobClient:     Map output bytes=14781
17/10/01 01:55:14 INFO mapred.JobClient:     Reduce shuffle bytes=16883
17/10/01 01:55:14 INFO mapred.JobClient:     Physical memory (bytes) snapshot=59743027
17/10/01 01:55:14 INFO mapred.JobClient:     Reduce input groups=529
17/10/01 01:55:14 INFO mapred.JobClient:     Combine output records=0
17/10/01 01:55:14 INFO mapred.JobClient:     Reduce output records=529
17/10/01 01:55:14 INFO mapred.JobClient:     Map output records=1042
17/10/01 01:55:14 INFO mapred.JobClient:     Combine input records=0
17/10/01 01:55:14 INFO mapred.JobClient:     CPU time spent (ms)=1050
17/10/01 01:55:14 INFO mapred.JobClient:     Total committed heap usage (bytes)=498544
17/10/01 01:55:14 INFO mapred.JobClient: File Input Format Counters
17/10/01 01:55:14 INFO mapred.JobClient:   Bytes Read=6481
17/10/01 01:55:14 INFO mapred.JobClient: FileSystemCounters
17/10/01 01:55:14 INFO mapred.JobClient:   HDFS_BYTES_READ=6886
17/10/01 01:55:14 INFO mapred.JobClient:   FILE_BYTES_WRITTEN=261175
17/10/01 01:55:14 INFO mapred.JobClient:   FILE_BYTES_READ=16871
17/10/01 01:55:14 INFO mapred.JobClient:   HDFS_BYTES_WRITTEN=5195
17/10/01 01:55:14 INFO mapred.JobClient: Job Counters
17/10/01 01:55:14 INFO mapred.JobClient:   Launched map tasks=3
17/10/01 01:55:14 INFO mapred.JobClient:   Launched reduce tasks=1
17/10/01 01:55:14 INFO mapred.JobClient:   SLOTS_MILLIS_REDUCES=9123
17/10/01 01:55:14 INFO mapred.JobClient:   Total time spent by all reduces waiting a
17/10/01 01:55:14 INFO mapred.JobClient:   SLOTS_MILLIS_MAPS=14827
```



Komuravelly Venkateshwarlu  
Komuravellyv1@udayton.edu  
1015237060

```
ubuntu@ip-172-31-38-5:~/hadoop$ hadoop jar WordCountProgramv3.jar WordCountProgramv3.WordCountJob
Total time ran 620
17/10/01 02:07:43 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Appli
17/10/01 02:07:43 INFO input.FileInputFormat: Total input paths to process : 3
17/10/01 02:07:43 INFO util.NativeCodeLoader: Loaded the native-hadoop library
17/10/01 02:07:43 WARN snappy.LoadSnappy: Snappy native library not loaded
17/10/01 02:07:43 INFO mapred.JobClient: Running job: job_201710010152_0003
17/10/01 02:07:44 INFO mapred.JobClient: map 0% reduce 0%
17/10/01 02:07:49 INFO mapred.JobClient: map 33% reduce 0%
17/10/01 02:07:51 INFO mapred.JobClient: map 100% reduce 0%
17/10/01 02:07:57 INFO mapred.JobClient: map 100% reduce 33%
17/10/01 02:07:59 INFO mapred.JobClient: map 100% reduce 100%
17/10/01 02:08:00 INFO mapred.JobClient: Job complete: job_201710010152_0003
17/10/01 02:08:00 INFO mapred.JobClient: Counters: 29
17/10/01 02:08:00 INFO mapred.JobClient: Map-Reduce Framework
17/10/01 02:08:00 INFO mapred.JobClient: Spilled Records=1166
17/10/01 02:08:00 INFO mapred.JobClient: Map output materialized bytes=10239
17/10/01 02:08:00 INFO mapred.JobClient: Reduce input records=583
17/10/01 02:08:00 INFO mapred.JobClient: Virtual memory (bytes) snapshot=7612932096
17/10/01 02:08:00 INFO mapred.JobClient: Map input records=37
17/10/01 02:08:00 INFO mapred.JobClient: SPLIT_RAW_BYTES=405
17/10/01 02:08:00 INFO mapred.JobClient: Map output bytes=9055
17/10/01 02:08:00 INFO mapred.JobClient: Reduce shuffle bytes=10239
17/10/01 02:08:00 INFO mapred.JobClient: Physical memory (bytes) snapshot=598740992
17/10/01 02:08:00 INFO mapred.JobClient: Reduce input groups=528
17/10/01 02:08:00 INFO mapred.JobClient: Combine output records=0
17/10/01 02:08:00 INFO mapred.JobClient: Reduce output records=528
17/10/01 02:08:00 INFO mapred.JobClient: Map output records=583
17/10/01 02:08:00 INFO mapred.JobClient: Combine input records=0
17/10/01 02:08:00 INFO mapred.JobClient: CPU time spent (ms)=1160
17/10/01 02:08:00 INFO mapred.JobClient: Total committed heap usage (bytes)=498544640
17/10/01 02:08:00 INFO mapred.JobClient: File Input Format Counters
17/10/01 02:08:00 INFO mapred.JobClient: Bytes Read=6481
17/10/01 02:08:00 INFO mapred.JobClient: FileSystemCounters
17/10/01 02:08:00 INFO mapred.JobClient: HDFS_BYTES_READ=6886
17/10/01 02:08:00 INFO mapred.JobClient: FILE_BYTES_WRITTEN=247923
17/10/01 02:08:00 INFO mapred.JobClient: FILE_BYTES_READ=10227
17/10/01 02:08:00 INFO mapred.JobClient: HDFS_BYTES_WRITTEN=5191
17/10/01 02:08:00 INFO mapred.JobClient: Job Counters
17/10/01 02:08:00 INFO mapred.JobClient: Launched map tasks=3
17/10/01 02:08:00 INFO mapred.JobClient: Launched reduce tasks=1
17/10/01 02:08:00 INFO mapred.JobClient: SLOTS_MILLIS_REDUCES=9177
17/10/01 02:08:00 INFO mapred.JobClient: Total time spent by all reduces waiting after reserv
17/10/01 02:08:00 INFO mapred.JobClient: SLOTS_MILLIS_MAPS=12824
```