```
/* Name: Elroy Chua
Tutorial Grp: T04F
UOW ID:7431673
Task 1--------------------------------------------------------
Comments: export and echo before running the ls command */
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ $HADOOP_HOME/bin/
hadoop fs -ls
Found 2 items
drwxr-xr-x   - bigdata supergroup          0 2017-07-03
01:33 .sparkStaging
-rw-r--r--   1 bigdata supergroup         91 2023-04-29 17:55
grep.txt
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ $HADOOP_HOME/bin/
hadoop fs -mkdir myfolderA
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ $HADOOP_HOME/bin/
hadoop fs -mkdir myfolderB
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ $HADOOP_HOME/bin/
hadoop fs -ls
Found 4 items
drwxr-xr-x   - bigdata supergroup          0 2017-07-03
01:33 .sparkStaging
-rw-r--r--   1 bigdata supergroup         91 2023-04-29 17:55
grep.txt
drwxr-xr-x   - bigdata supergroup          0 2023-04-29 23:21
myfolderA
drwxr-xr-x   - bigdata supergroup          0 2023-04-29 23:21
myfolderB
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ $HADOOP_HOME/bin/
hadoop fs -put /home/bigdata/test.txt myfolderA
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ $HADOOP_HOME/bin/
hadoop fs -cat myfolderA/test.txt
Test txt for Assignment 1 Solution 1a.
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ javac -cp
$HADOOP_CLASSPATH solution1.java
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ jar cvf
solution1.jar solution1*.class
added manifest
adding: solution1.class(in = 1578) (out= 853)(deflated 45%)
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ $HADOOP_HOME/bin/
hadoop jar /home/bigdata/Desktop/A1_BigData/FileSystemPut.jar
FileSystemPut myfolderA/test.txt myfolderB/testCopy.txt
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ $HADOOP_HOME/bin/
hadoop fs -ls myfolderA
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ $HADOOP_HOME/bin/
hadoop fs -ls myfolderB
Found 1 items
-rw-r--r--   1 bigdata supergroup         40 2023-04-29 23:32
myfolderB/testCopy.txt
```

```
/* Name: Elroy Chua
Tutorial Grp: T04F
UOW ID:7431673
Task 2——————————————————————————————————————————————————*/
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ export
HADOOP_CLASSPATH=$($HADOOP_HOME/bin/hadoop classpath)
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ echo
$HADOOP_CLASSPATH
/usr/share/hadoop/etc/hadoop:/usr/share/hadoop-2.7.3/share/hadoop/
common/lib/*:/usr/share/hadoop-2.7.3/share/hadoop/common/*:/usr/
share/hadoop-2.7.3/share/hadoop/hdfs:/usr/share/hadoop-2.7.3/share/
hadoop/hdfs/lib/*:/usr/share/hadoop-2.7.3/share/hadoop/hdfs/*:/usr/
share/hadoop-2.7.3/share/hadoop/yarn/lib/*:/usr/share/hadoop-2.7.3/
share/hadoop/yarn/*:/usr/share/hadoop-2.7.3/share/hadoop/mapreduce/
lib/*:/usr/share/hadoop-2.7.3/share/hadoop/mapreduce/*:/usr/share/
hadoop/etc/hadoop:/usr/share/hadoop-2.7.3/share/hadoop/common/lib/
*:/usr/share/hadoop-2.7.3/share/hadoop/common/*:/usr/share/
hadoop-2.7.3/share/hadoop/hdfs:/usr/share/hadoop-2.7.3/share/hadoop/
hdfs/lib/*:/usr/share/hadoop-2.7.3/share/hadoop/hdfs/*:/usr/share/
hadoop-2.7.3/share/hadoop/yarn/lib/*:/usr/share/hadoop-2.7.3/share/
hadoop/yarn/*:/usr/share/hadoop-2.7.3/share/hadoop/mapreduce/lib/*:/
usr/share/hadoop-2.7.3/share/hadoop/mapreduce/*:/usr/share/hadoop/
etc/hadoop:/usr/share/hadoop-2.7.3/share/hadoop/common/lib/*:/usr/
share/hadoop-2.7.3/share/hadoop/common/*:/usr/share/hadoop-2.7.3/
share/hadoop/hdfs:/usr/share/hadoop-2.7.3/share/hadoop/hdfs/lib/*:/
usr/share/hadoop-2.7.3/share/hadoop/hdfs/*:/usr/share/hadoop-2.7.3/
share/hadoop/yarn/lib/*:/usr/share/hadoop-2.7.3/share/hadoop/yarn/
*:/usr/share/hadoop-2.7.3/share/hadoop/mapreduce/lib/*:/usr/share/
hadoop-2.7.3/share/hadoop/mapreduce/*:/usr/share/hadoop/etc/hadoop:/
usr/share/hadoop-2.7.3/share/hadoop/common/lib/*:/usr/share/
hadoop-2.7.3/share/hadoop/common/*:/usr/share/hadoop-2.7.3/share/
hadoop/hdfs:/usr/share/hadoop-2.7.3/share/hadoop/hdfs/lib/*:/usr/
share/hadoop-2.7.3/share/hadoop/hdfs/*:/usr/share/hadoop-2.7.3/
share/hadoop/yarn/lib/*:/usr/share/hadoop-2.7.3/share/hadoop/yarn/
*:/usr/share/hadoop-2.7.3/share/hadoop/mapreduce/lib/*:/usr/share/
hadoop-2.7.3/share/hadoop/mapreduce/*:/usr/share/hadoop/etc/hadoop:/
usr/share/hadoop-2.7.3/share/hadoop/common/lib/*:/usr/share/
hadoop-2.7.3/share/hadoop/common/*:/usr/share/hadoop-2.7.3/share/
hadoop/hdfs:/usr/share/hadoop-2.7.3/share/hadoop/hdfs/lib/*:/usr/
share/hadoop-2.7.3/share/hadoop/hdfs/*:/usr/share/hadoop-2.7.3/
share/hadoop/yarn/lib/*:/usr/share/hadoop-2.7.3/share/hadoop/yarn/
*:/usr/share/hadoop-2.7.3/share/hadoop/mapreduce/lib/*:/usr/share/
hadoop-2.7.3/share/hadoop/mapreduce/*:/usr/share/hadoop/contrib/
capacity-scheduler/*.jar:/usr/share/hadoop/contrib/capacity-
scheduler/*.jar:/usr/share/hadoop/contrib/capacity-scheduler/*.jar:/
usr/share/hadoop/contrib/capacity-scheduler/*.jar:/usr/share/hadoop/
contrib/capacity-scheduler/*.jar:/usr/share/hadoop/contrib/capacity-
```

```
scheduler/*.jar
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ $HADOOP_HOME/bin/
hadoop fs -put aus_records.txt
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ $HADOOP_HOME/bin/
hadoop fs -ls
Found 5 items
drwxr-xr-x   - bigdata supergroup          0 2017-07-03
01:33 .sparkStaging
-rw-r--r--   1 bigdata supergroup        242 2023-04-30 01:43
aus_records.txt
-rw-r--r--   1 bigdata supergroup         91 2023-04-29 17:55
grep.txt
drwxr-xr-x   - bigdata supergroup          0 2023-04-29 23:32
myfolderA
drwxr-xr-x   - bigdata supergroup          0 2023-04-29 23:32
myfolderB
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ javac -cp
$HADOOP_CLASSPATH solution2.java
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ jar cvf
solution2.jar solution2*.class
added manifest
adding: solution2.class(in = 1468) (out= 796)(deflated 45%)
adding: solution2$StateRainfallReducer.class(in = 2572) (out= 1165)
(deflated 54%)
adding: solution2$TokenizeMapper.class(in = 2087) (out= 930)
(deflated 55%)
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ $HADOOP_HOME/bin/
hadoop jar solution2.jar solution2 aus_records.txt /user/output/
23/04/30 01:44:22 INFO client.RMProxy: Connecting to ResourceManager
at bigdata-VirtualBox/10.0.2.15:8032
23/04/30 01:44:23 WARN mapreduce.JobResourceUploader: Hadoop
command-line option parsing not performed. Implement the Tool
interface and execute your application with ToolRunner to remedy
this.
23/04/30 01:44:23 INFO input.FileInputFormat: Total input paths to
process : 1
23/04/30 01:44:23 INFO mapreduce.JobSubmitter: number of splits:1
23/04/30 01:44:23 INFO mapreduce.JobSubmitter: Submitting tokens for
job: job_1682739043984_0014
23/04/30 01:44:24 INFO impl.YarnClientImpl: Submitted application
application_1682739043984_0014
23/04/30 01:44:24 INFO mapreduce.Job: The url to track the job:
http://bigdata-VirtualBox:8088/proxy/application_1682739043984_0014/
23/04/30 01:44:24 INFO mapreduce.Job: Running job:
job_1682739043984_0014
23/04/30 01:44:31 INFO mapreduce.Job: Job job_1682739043984_0014
running in uber mode : false
23/04/30 01:44:31 INFO mapreduce.Job:  map 0% reduce 0%
23/04/30 01:44:35 INFO mapreduce.Job:  map 100% reduce 0%
23/04/30 01:44:41 INFO mapreduce.Job:  map 100% reduce 100%
23/04/30 01:44:42 INFO mapreduce.Job: Job job_1682739043984_0014
completed successfully
23/04/30 01:44:42 INFO mapreduce.Job: Counters: 49
        File System Counters
```

```
                FILE: Number of bytes read=189
                FILE: Number of bytes written=237621
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=357
                HDFS: Number of bytes written=165
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots
(ms)=2443
                Total time spent by all reduces in occupied slots
(ms)=4048
                Total time spent by all map tasks (ms)=2443
                Total time spent by all reduce tasks (ms)=4048
                Total vcore-milliseconds taken by all map tasks=2443
                Total vcore-milliseconds taken by all reduce
tasks=4048
                Total megabyte-milliseconds taken by all map
tasks=2501632
                Total megabyte-milliseconds taken by all reduce
tasks=4145152
        Map-Reduce Framework
                Map input records=10
                Map output records=10
                Map output bytes=163
                Map output materialized bytes=189
                Input split bytes=115
                Combine input records=0
                Combine output records=0
                Reduce input groups=5
                Reduce shuffle bytes=189
                Reduce input records=10
                Reduce output records=5
                Spilled Records=20
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=172
                CPU time spent (ms)=700
                Physical memory (bytes) snapshot=438366208
                Virtual memory (bytes) snapshot=3824607232
                Total committed heap usage (bytes)=311951360
        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=242
        File Output Format Counters
                Bytes Written=165
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ $HADOOP_HOME/bin/
hadoop fs -ls /user/output
Found 2 items
-rw-r--r--   1 bigdata supergroup          0 2023-04-30 01:44 /user/
output/_SUCCESS
-rw-r--r--   1 bigdata supergroup        165 2023-04-30 01:44 /user/
output/part-r-00000
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ $HADOOP_HOME/bin/
hadoop fs -cat /user/output/part-r-00000
NSW                     900 900 900
Queensland               75  50  25
South_Australia         300 300 300
Victoria                225 125  10
Western_Australia       410 200  10
```

```
/* Name: Elroy Chua
Tutorial Grp: T04F
UOW ID:7431673
Task 3---------------------------------------------------------*/
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ javac -cp
$HADOOP_CLASSPATH solution3.java
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ jar cf
solution3.jar solution3*.class
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ $HADOOP_HOME/bin/
hadoop jar solution3.jar solution3 grep.txt /user/output
23/04/29 18:52:40 INFO client.RMProxy: Connecting to ResourceManager
at bigdata-VirtualBox/10.0.2.15:8032
23/04/29 18:52:40 WARN mapreduce.JobResourceUploader: Hadoop
command-line option parsing not performed. Implement the Tool
interface and execute your application with ToolRunner to remedy
this.
23/04/29 18:52:41 INFO input.FileInputFormat: Total input paths to
process : 1
23/04/29 18:52:41 INFO mapreduce.JobSubmitter: number of splits:1
23/04/29 18:52:42 INFO mapreduce.JobSubmitter: Submitting tokens for
job: job_1682739043984_0009
23/04/29 18:52:42 INFO impl.YarnClientImpl: Submitted application
application_1682739043984_0009
23/04/29 18:52:42 INFO mapreduce.Job: The url to track the job:
http://bigdata-VirtualBox:8088/proxy/application_1682739043984_0009/
23/04/29 18:52:42 INFO mapreduce.Job: Running job:
job_1682739043984_0009
23/04/29 18:52:48 INFO mapreduce.Job: Job job_1682739043984_0009
running in uber mode : false
23/04/29 18:52:48 INFO mapreduce.Job:  map 0% reduce 0%
23/04/29 18:52:53 INFO mapreduce.Job:  map 100% reduce 0%
23/04/29 18:52:58 INFO mapreduce.Job:  map 100% reduce 100%
23/04/29 18:52:58 INFO mapreduce.Job: Job job_1682739043984_0009
completed successfully
23/04/29 18:52:58 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=1163
                FILE: Number of bytes written=239535
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=199
                HDFS: Number of bytes written=114
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots
(ms)=2246
                Total time spent by all reduces in occupied slots
(ms)=2801
```

```
                        Total time spent by all map tasks (ms)=2246
                        Total time spent by all reduce tasks (ms)=2801
                        Total vcore-milliseconds taken by all map tasks=2246
                        Total vcore-milliseconds taken by all reduce
tasks=2801
                        Total megabyte-milliseconds taken by all map
tasks=2299904
                        Total megabyte-milliseconds taken by all reduce
tasks=2868224
        Map-Reduce Framework
                        Map input records=12
                        Map output records=78
                        Map output bytes=1001
                        Map output materialized bytes=1163
                        Input split bytes=108
                        Combine input records=0
                        Combine output records=0
                        Reduce input groups=6
                        Reduce shuffle bytes=1163
                        Reduce input records=78
                        Reduce output records=6
                        Spilled Records=156
                        Shuffled Maps =1
                        Failed Shuffles=0
                        Merged Map outputs=1
                        GC time elapsed (ms)=98
                        CPU time spent (ms)=730
                        Physical memory (bytes) snapshot=434446336
                        Virtual memory (bytes) snapshot=3821424640
                        Total committed heap usage (bytes)=309329920
        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=91
        File Output Format Counters
                        Bytes Written=114
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$ $HADOOP_HOME/bin/
hadoop fs -cat /user/output/part-r-00000
X short:        4 words
short:          3 words
medium:         3 words
long:           2 words
X long:         0 words
XX long:        1 words
bigdata@bigdata-VirtualBox:~/Desktop/A1_BigData$
```