/\* Student Number: 10241485 \*/

/\* Name: Jeslyn Ho Ka Yan \*/  
/\* Date: 13 Oct 2024 \*/

/\* Assignment 1, Task 3 \*/

/\* \*/

/\* Setup and Configuration ----------------------------------------------------- \*/  
/\* Set Hadoop classpath for Java compilation \*/

bigdata@bigdata-VirtualBox:~$ export HADOOP\_CLASSPATH=$($HADOOP\_HOME/bin/hadoop classpath)

/\* Check if Hadoop classpath is set correctly \*/

bigdata@bigdata-VirtualBox:~$ 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/contrib/capacity-scheduler/\*.jar

/\* Check Hadoop home directory \*/

bigdata@bigdata-VirtualBox:~$ echo $HADOOP\_HOME

/usr/share/hadoop

/\* Project Directory Setup ----------------------------------------------------- \*/  
/\* Navigate to project directory \*/

bigdata@bigdata-VirtualBox:~$ cd ISIT312/A1/Task3

/\* List files in the project directory \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task3$ ls

sales.txt solution3.java solution3.java~

/\* HDFS Operations ------------------------------------------------------------- \*/

/\* Upload local files ‘sales.txt' to HDFS \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task3$ $HADOOP\_HOME/bin/hadoop fs -put sales.txt

/\* List HDFS contents to verify directory creation \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task3$ $HADOOP\_HOME/bin/hadoop fs -ls

Found 8 items

drwxr-xr-x - bigdata supergroup 0 2017-07-03 01:33 .sparkStaging

drwxr-xr-x - bigdata supergroup 0 2024-10-15 23:05 A1Task2OP.txt

-rw-r--r-- 1 bigdata supergroup 576 2024-10-15 22:52 SpeedCamera.txt

drwxr-xr-x - bigdata supergroup 0 2024-10-11 17:09 WordCountOutput

drwxr-xr-x - bigdata supergroup 0 2024-10-13 02:51 myfolderA

-rw-r--r-- 1 bigdata supergroup 230 2024-10-16 00:00 sales.txt

-rw-r--r-- 1 bigdata supergroup 1743 2024-10-13 02:45 solution1.java

-rw-r--r-- 1 bigdata supergroup 2811 2024-10-15 23:02 solution2.java

/\* View contents of sales.txt' from HDFS \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task3$ $HADOOP\_HOME/bin/hadoop fs -cat sales.txt

bolt 45

washer 7

bolt 51

washer 10

screw 48

screw 13

nail 50

washer 3

washer 56

screw 28

bolt 25

bolt 10

screw 67

screw 23

nail 5

screw 78

bolt 5

bolt 1

drill 1

drill 1

drill 14

drill 3

washer 36

washer 27

/\* Java Compilation and JAR Creation --------------------------------------- \*/  
/\* Compile Java program with Hadoop classpath \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task3$ javac -classpath ${HADOOP\_CLASSPATH} solution3.java

/\* Upload compiled Java file to HDFS \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task3$ $HADOOP\_HOME/bin/hadoop fs -put solution3.java

/\* Create a JAR file containing compiled Java class \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task3$ jar cvf solution3.jar solution3\*.class

added manifest

adding: solution3.class(in = 1454) (out= 795)(deflated 45%)

adding: solution3$IntSumReducer.class(in = 2116) (out= 974)(deflated 53%)

adding: solution3$TokenizerMapper.class(in = 1837) (out= 798)(deflated 56%)

/\* List HDFS files to verify Java file upload \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task3$ ls

sales.txt solution3$IntSumReducer.class solution3.java solution3$TokenizerMapper.class

solution3.class solution3.jar solution3.java~

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task3$ $HADOOP\_HOME/bin/hadoop fs -ls

Found 9 items

drwxr-xr-x - bigdata supergroup 0 2017-07-03 01:33 .sparkStaging

drwxr-xr-x - bigdata supergroup 0 2024-10-15 23:05 A1Task2OP.txt

-rw-r--r-- 1 bigdata supergroup 576 2024-10-15 22:52 SpeedCamera.txt

drwxr-xr-x - bigdata supergroup 0 2024-10-11 17:09 WordCountOutput

drwxr-xr-x - bigdata supergroup 0 2024-10-13 02:51 myfolderA

-rw-r--r-- 1 bigdata supergroup 230 2024-10-16 00:00 sales.txt

-rw-r--r-- 1 bigdata supergroup 1743 2024-10-13 02:45 solution1.java

-rw-r--r-- 1 bigdata supergroup 2811 2024-10-15 23:02 solution2.java

-rw-r--r-- 1 bigdata supergroup 3823 2024-10-16 00:03 solution3.java

/\* Running the Hadoop Job----------------------------------------------------- \*/  
/\* Run Hadoop job with the created JAR \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task3$ $HADOOP\_HOME/bin/hadoop jar solution3.jar solution3 sales.txt A1Task3OP.txt

24/10/16 00:06:14 INFO client.RMProxy: Connecting to ResourceManager at bigdata-VirtualBox/10.0.2.15:8032

24/10/16 00:06:15 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

24/10/16 00:06:16 INFO input.FileInputFormat: Total input paths to process : 1

24/10/16 00:06:16 INFO mapreduce.JobSubmitter: number of splits:1

24/10/16 00:06:17 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1728992865259\_0002

24/10/16 00:06:18 INFO impl.YarnClientImpl: Submitted application application\_1728992865259\_0002

24/10/16 00:06:18 INFO mapreduce.Job: The url to track the job: http://bigdata-VirtualBox:8088/proxy/application\_1728992865259\_0002/

24/10/16 00:06:18 INFO mapreduce.Job: Running job: job\_1728992865259\_0002

24/10/16 00:06:34 INFO mapreduce.Job: Job job\_1728992865259\_0002 running in uber mode : false

24/10/16 00:06:34 INFO mapreduce.Job: map 0% reduce 0%

24/10/16 00:06:42 INFO mapreduce.Job: map 100% reduce 0%

24/10/16 00:06:53 INFO mapreduce.Job: map 100% reduce 100%

24/10/16 00:06:55 INFO mapreduce.Job: Job job\_1728992865259\_0002 completed successfully

24/10/16 00:06:55 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=292

FILE: Number of bytes written=237829

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=339

HDFS: Number of bytes written=192

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)=6535

Total time spent by all reduces in occupied slots (ms)=7854

Total time spent by all map tasks (ms)=6535

Total time spent by all reduce tasks (ms)=7854

Total vcore-milliseconds taken by all map tasks=6535

Total vcore-milliseconds taken by all reduce tasks=7854

Total megabyte-milliseconds taken by all map tasks=6691840

Total megabyte-milliseconds taken by all reduce tasks=8042496

Map-Reduce Framework

Map input records=24

Map output records=24

Map output bytes=238

Map output materialized bytes=292

Input split bytes=109

Combine input records=0

Combine output records=0

Reduce input groups=5

Reduce shuffle bytes=292

Reduce input records=24

Reduce output records=5

Spilled Records=48

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=547

CPU time spent (ms)=2480

Physical memory (bytes) snapshot=442449920

Virtual memory (bytes) snapshot=3830812672

Total committed heap usage (bytes)=321388544

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=230

File Output Format Counters

Bytes Written=192

/\* List the contents of the output directory to check for results \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task3$ $HADOOP\_HOME/bin/hadoop fs -ls A1Task3OP.txtFound 2 items

-rw-r--r-- 1 bigdata supergroup 0 2024-10-16 00:06 A1Task3OP.txt/\_SUCCESS

-rw-r--r-- 1 bigdata supergroup 192 2024-10-16 00:06 A1Task3OP.txt/part-r-00000

/\* Check the output file \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task3$ $HADOOP\_HOME/bin/hadoop fs -cat A1Task3OP.txt/part-r-00000

bolt Max:51 Min:1 Avg:22.83 Total:137

drill Max:14 Min:1 Avg:4.75 Total:19

nail Max:50 Min:5 Avg:27.50 Total:55

screw Max:78 Min:13 Avg:42.83 Total:257

washer Max:56 Min:3 Avg:23.17 Total:139

/\* **Summary of Steps for Hadoop Job Execution**

**1. Setup and Configuration:** Make sure Hadoop is configured correctly by setting the HADOOP\_CLASSPATH and verifying its value. confirmed the path for HADOOP\_HOME.  
  
**2. Project Directory Setup:** Proceeded to the project directory and verified that your files were there.   
  
**3. HDFS Operations:** The SpeedCamera.txt file was successfully uploaded to HDFS.   
Listing and displaying the file's contents allowed me to confirm the upload.   
  
**4. Java Compilation and JAR Generation**: Made use of the Hadoop classpath to compile the Java code.   
Using your built classes, a JAR file was created, and its construction was verified.   
  
**5. Executing the Hadoop task:** Using the generated JAR file, I ran the Hadoop task, tracked its progress, and verified that it finished properly.   
The task output displayed the average speeds and speed infractions for particular cars.   
  
**6. Output Verification:** Examined the information in the output directory and showed the processing job's outcomes.

\*/