/\* Student Number: 10241485 \*/

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

/\* Assignment 1, Task 2 \*/

/\* \*/

/\* 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 /home/bigdata/ISIT312/A1/Task2  
  
/\* List files in the project directory \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task2$ ls

solution2.java solution2.java~ SpeedCamera.txt

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

/\* Upload local files 'SpeedCamera.txt' to HDFS \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task2$ $HADOOP\_HOME/bin/hadoop fs -put SpeedCamera.txt  
  
/\* List HDFS contents to verify directory creation \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task2$ $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 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 1743 2024-10-13 02:45 solution1.java

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

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task2$ $HADOOP\_HOME/bin/hadoop fs -cat SpeedCamera.txt

PKR856 AYE 14-NOV-2021 80

UPS234 CTE 20-FEB-2022 110

PKR856 PIE 20-MAR-2020 90

PKR856 PIE 17-JUN-2021 100

UPS234 CTE 22-SEP-2022 100

UPS234 CTE 03-AUG-2020 90

XYZ123 CTE 10-OCT-2021 95

XYZ123 AYE 11-OCT-2021 85

ABC789 PIE 05-JAN-2022 120

LMN456 CTE 15-FEB-2023 100

LMN456 AYE 20-FEB-2022 70

ABC789 PIE 22-FEB-2023 150

XYZ123 CTE 28-FEB-2021 105

DEF234 AYE 01-MAR-2022 80

GHI345 CTE 10-MAR-2022 110

LMN456 PIE 12-MAR-2022 95

JKL678 AYE 14-MAR-2022 85

MNO789 CTE 15-MAR-2022 115

JKL678 PIE 20-MAR-2022 120

PQR123 AYE 25-MAR-2022 75

STU456 CTE 30-MAR-2022 130

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

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task2$ javac -classpath ${HADOOP\_CLASSPATH} solution2.java  
  
/\* Upload compiled Java file to HDFS \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task2$ $HADOOP\_HOME/bin/hadoop fs -put solution2.java

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

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task2$ jar cvf solution2.jar solution2\*.class

added manifest

adding: solution2.class(in = 1454) (out= 795)(deflated 45%)

adding: solution2$SpeedMapper.class(in = 1963) (out= 866)(deflated 55%)

adding: solution2$SpeedReducer.class(in = 2137) (out= 975)(deflated 54%)  
  
/\* List files to confirm the JAR was created \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task2$ ls

solution2.class solution2.java solution2$SpeedMapper.class SpeedCamera.txt

solution2.jar solution2.java~ solution2$SpeedReducer.class  
  
/\* List HDFS files to verify Java file upload \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task2$ $HADOOP\_HOME/bin/hadoop fs -ls

Found 6 items

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

-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 1743 2024-10-13 02:45 solution1.java

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

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

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task2$ $HADOOP\_HOME/bin/hadoop jar solution2.jar solution2 SpeedCamera.txt A1Task2OP.txt

24/10/15 23:05:07 INFO client.RMProxy: Connecting to ResourceManager at bigdata-VirtualBox/10.0.2.15:8032

24/10/15 23:05:09 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

24/10/15 23:05:11 INFO input.FileInputFormat: Total input paths to process : 1

24/10/15 23:05:11 INFO mapreduce.JobSubmitter: number of splits:1

24/10/15 23:05:12 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1728992865259\_0001

24/10/15 23:05:13 INFO impl.YarnClientImpl: Submitted application application\_1728992865259\_0001

24/10/15 23:05:14 INFO mapreduce.Job: The url to track the job: http://bigdata-VirtualBox:8088/proxy/application\_1728992865259\_0001/

24/10/15 23:05:14 INFO mapreduce.Job: Running job: job\_1728992865259\_0001

24/10/15 23:05:35 INFO mapreduce.Job: Job job\_1728992865259\_0001 running in uber mode : false

24/10/15 23:05:35 INFO mapreduce.Job: map 0% reduce 0%

24/10/15 23:05:46 INFO mapreduce.Job: map 100% reduce 0%

24/10/15 23:05:59 INFO mapreduce.Job: map 100% reduce 100%

24/10/15 23:06:00 INFO mapreduce.Job: Job job\_1728992865259\_0001 completed successfully

24/10/15 23:06:00 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=447

FILE: Number of bytes written=238161

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=691

HDFS: Number of bytes written=558

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)=9605

Total time spent by all reduces in occupied slots (ms)=9602

Total time spent by all map tasks (ms)=9605

Total time spent by all reduce tasks (ms)=9602

Total vcore-milliseconds taken by all map tasks=9605

Total vcore-milliseconds taken by all reduce tasks=9602

Total megabyte-milliseconds taken by all map tasks=9835520

Total megabyte-milliseconds taken by all reduce tasks=9832448

Map-Reduce Framework

Map input records=21

Map output records=21

Map output bytes=399

Map output materialized bytes=447

Input split bytes=115

Combine input records=0

Combine output records=0

Reduce input groups=16

Reduce shuffle bytes=447

Reduce input records=21

Reduce output records=10

Spilled Records=42

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=761

CPU time spent (ms)=2490

Physical memory (bytes) snapshot=438763520

Virtual memory (bytes) snapshot=3829075968

Total committed heap usage (bytes)=324534272

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=576

File Output Format Counters

Bytes Written=558

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

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task2$ $HADOOP\_HOME/bin/hadoop fs -ls A1Task2OP.txt

Found 2 items

-rw-r--r-- 1 bigdata supergroup 0 2024-10-15 23:05 A1Task2OP.txt/\_SUCCESS

-rw-r--r-- 1 bigdata supergroup 558 2024-10-15 23:05 A1Task2OP.txt/part-r-00000

/\* Check the output file \*/

bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task2$ $HADOOP\_HOME/bin/hadoop fs -cat A1Task3OP.txt/part-r-00000

ABC789 PIE Speed limit exceeded!! Average Speed: 135.00

GHI345 CTE Speed limit exceeded!! Average Speed: 110.00

JKL678 PIE Speed limit exceeded!! Average Speed: 120.00

LMN456 CTE Speed limit exceeded!! Average Speed: 100.00

LMN456 PIE Speed limit exceeded!! Average Speed: 95.00

MNO789 CTE Speed limit exceeded!! Average Speed: 115.00

PKR856 PIE Speed limit exceeded!! Average Speed: 95.00

STU456 CTE Speed limit exceeded!! Average Speed: 130.00

UPS234 CTE Speed limit exceeded!! Average Speed: 100.00

XYZ123 CTE Speed limit exceeded!! Average Speed: 100.00

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

1. **Setup and Configuration**:

Set the HADOOP\_CLASSPATH and checked its value, confirming that Hadoop is configured properly. verified the HADOOP\_HOME path.

**2. Project Directory Setup**:

Navigated to the project directory and confirmed the presence of your files.

**3. HDFS Operations**:

Successfully uploaded SpeedCamera.txt to HDFS.

Verified the upload by listing and displaying the contents of the file.

4. **Java Compilation and JAR Creation**:

Compiled the Java code using the Hadoop classpath.

Created a JAR file from your compiled classes and confirmed the JAR file creation.

5. **Running the Hadoop Job**:

Executed the Hadoop job using the created JAR file, monitored the job's progress, and confirmed that it completed successfully.

The job output indicated the speed violations and average speeds for specific vehicles.

6. **Output Verification**:

Checked the contents of the output directory and displayed the results of the processing job.

\*/