

```

/* Student Number: 10241485 */
/* Name: Jeslyn Ho Ka Yan */
/* Date: 13 Oct 2024 */
/* Assignment 1, Task 1 */

/* */
/* 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

/* List files in the project directory */
bigdata@bigdata-VirtualBox:~/ISIT312/A1$ ls
file1 file2 solution1.java

/* HDFS Operations ----- */
/* Create a folder 'myfolderA' in HDFS */
bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP_HOME/bin/hadoop fs -mkdir myfolderA

/* List HDFS contents to verify directory creation */
bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP_HOME/bin/hadoop fs -ls
Found 3 items
drwxr-xr-x - bigdata supergroup          0 2017-07-03 01:33 .sparkStaging
drwxr-xr-x - bigdata supergroup          0 2024-10-11 17:09 WordCountOutput
drwxr-xr-x - bigdata supergroup          0 2024-10-13 02:36 myfolderA

/* Upload local files 'file1' and 'file2' to 'myfolderA' in HDFS */
bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP_HOME/bin/hadoop fs -put
/home/bigdata/ISIT312/A1/file1 myfolderA/
bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP_HOME/bin/hadoop fs -put
/home/bigdata/ISIT312/A1/file2 myfolderA/

/* List contents of 'myfolderA' to confirm files are uploaded */
bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP_HOME/bin/hadoop fs -ls myfolderA
Found 2 items
-rw-r--r-- 1 bigdata supergroup          27 2024-10-13 02:40 myfolderA/file1
-rw-r--r-- 1 bigdata supergroup          26 2024-10-13 02:40 myfolderA/file2

```

```

/* View contents of 'file1' from HDFS */
bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP_HOME/bin/hadoop fs -cat
myfolderA/file1
This is file1.
Halloween

/* View contents of 'file2' from HDFS */
bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP_HOME/bin/hadoop fs -cat
myfolderA/file2
This is file2.
Is coming!

/* Java Compilation and JAR Creation ----- */
/* Compile Java program with Hadoop classpath */
bigdata@bigdata-VirtualBox:~/ISIT312/A1$ javac -classpath ${HADOOP_CLASSPATH}
solution1.java

/* Upload compiled Java file to HDFS */
bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP_HOME/bin/hadoop fs -put
solution1.java

/* Create a JAR file containing compiled Java class */
bigdata@bigdata-VirtualBox:~/ISIT312/A1$ jar cvf solution1.jar solution1*.class
added manifest
adding: solution1.class(in = 1526) (out= 806)(deflated 47%)

/* List files to confirm the JAR was created */
bigdata@bigdata-VirtualBox:~/ISIT312/A1$ ls
file1 file2 solution1.class solution1.jar solution1.java

/* List HDFS files to verify Java file upload */
bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP_HOME/bin/hadoop fs -ls
Found 4 items
drwxr-xr-x - bigdata supergroup 0 2017-07-03 01:33 .sparkStaging
drwxr-xr-x - bigdata supergroup 0 2024-10-11 17:09 WordCountOutput
drwxr-xr-x - bigdata supergroup 0 2024-10-13 02:40 myfolderA
-rw-r--r-- 1 bigdata supergroup 1743 2024-10-13 02:45 solution1.java

/* Running the Hadoop Job----- */
/* Run Hadoop job with the created JAR */
bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP_HOME/bin/hadoop jar solution1.jar
solution1 myfolderA/file1 myfolderA/file2 myfolderA/mergedOutput.txt

/* Check the output of the merged file */
bigdata@bigdata-VirtualBox:~/ISIT312/A1$ $HADOOP_HOME/bin/hadoop fs -cat
myfolderA/mergedOutput.txt
This is file1.
Halloween

This is file2.
Is coming!

```

`/* Summary of Steps for Hadoop Job Execution`

- `1. Set Up Hadoop Classpath: Configure the classpath using `export HADOOP_CLASSPATH=${HADOOP_HOME}/bin/hadoop classpath` and verify with `echo $HADOOP_CLASSPATH`.`
- `2. Navigate to Project Directory: Change to the project directory with `cd ISIT312/A1`.`
- `3. Create HDFS Directory: Create a new folder in HDFS using `$HADOOP_HOME/bin/hadoop fs -mkdir myfolderA` and confirm its creation by listing HDFS contents.`
- `4. Upload Files*: Transfer local files 'file1' and 'file2' to HDFS with `$HADOOP_HOME/bin/hadoop fs -put`.`
- `5. View Uploaded Files: Check the contents of the uploaded files using `$HADOOP_HOME/bin/hadoop fs -cat`.`
- `6. Compile Java Program: Compile `solution1.java` with the Hadoop classpath using `javac -classpath ${HADOOP_CLASSPATH} solution1.java`.`
- `7. Create and Upload JAR File*: Package the compiled classes into a JAR file with `jar cvf solution1.jar solution1*.class` and upload it to HDFS.`
- `8. Run Hadoop Job: Execute the job using the command:
$HADOOP_HOME/bin/hadoop jar solution1.jar solution1 myfolderA/file1 myfolderA/file2
myfolderA/mergedOutput.txt`
- `9. Check Output: Verify the merged output file's contents with
$HADOOP_HOME/bin/hadoop fs
-cat myfolderA/mergedOutput.txt`.`

`*/`

```

/* 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 A1Task20P.txt
24/10/15 23:05:07 INFO client.RMPProxy: 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
A1Task20P.txt
Found 2 items
-rw-r--r--    1 bigdata supergroup          0 2024-10-15 23:05 A1Task20P.txt/_SUCCESS
-rw-r--r--    1 bigdata supergroup      558 2024-10-15 23:05 A1Task20P.txt/part-r-
000000

/* Check the output file */
bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task2$ $HADOOP_HOME/bin/hadoop fs -cat
A1Task30P.txt/part-r-000000
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.

*/

```

```

/* 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 A1Task20P.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 A1Task20P.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 A1Task30P.txt
24/10/16 00:06:14 INFO client.RMPProxy: 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
```

```
A1Task30P.txtFound 2 items
```

```
-rw-r--r--  1 bigdata supergroup          0 2024-10-16 00:06 A1Task30P.txt/_SUCCESS
-rw-r--r--  1 bigdata supergroup       192 2024-10-16 00:06 A1Task30P.txt/part-r-
000000
```

```
/* Check the output file */
```

```
bigdata@bigdata-VirtualBox:~/ISIT312/A1/Task3$ $HADOOP_HOME/bin/hadoop fs -cat
```

```
A1Task30P.txt/part-r-000000
```

```
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.

```
*/
```