

A1 Task 1

Start hadoop server

`./starthadoop.sh`

```
bigdata@bigdata-VirtualBox: ~  
bigdata@bigdata-VirtualBox:~$ ./starthadoop.sh  
./starthadoop.sh: line 1: fg: no job control  
Starting namenodes on [localhost]  
localhost: starting namenode, logging to /usr/share/hadoop-2.7.3/logs/hadoop-bigdata-namenode-bigdata-VirtualBox.out  
localhost: starting datanode, logging to /usr/share/hadoop-2.7.3/logs/hadoop-bigdata-datanode-bigdata-VirtualBox.out  
Starting secondary namenodes [0.0.0.0]  
0.0.0.0: starting secondarynamenode, logging to /usr/share/hadoop-2.7.3/logs/hadoop-bigdata-secondarynamenode-bigdata-VirtualBox.out  
starting yarn daemons  
starting resourcemanager, logging to /usr/share/hadoop-2.7.3/logs/yarn-bigdata-resourcemanager-bigdata-VirtualBox.out  
localhost: starting nodemanager, logging to /usr/share/hadoop-2.7.3/logs/yarn-bigdata-nodemanager-bigdata-VirtualBox.out  
starting historyserver, logging to /usr/share/hadoop-2.7.3/logs/mapred-bigdata-historyserver-bigdata-VirtualBox.out  
24148 ResourceManager  
3060 RemoteInterpreterServer  
24309 NodeManager  
23574 NameNode  
23942 SecondaryNameNode  
24554 JobHistoryServer  
24619 Jps  
23724 DataNode  
2957 RemoteInterpreterServer  
2734 ZeppelinServer  
bigdata@bigdata-VirtualBox:~$
```

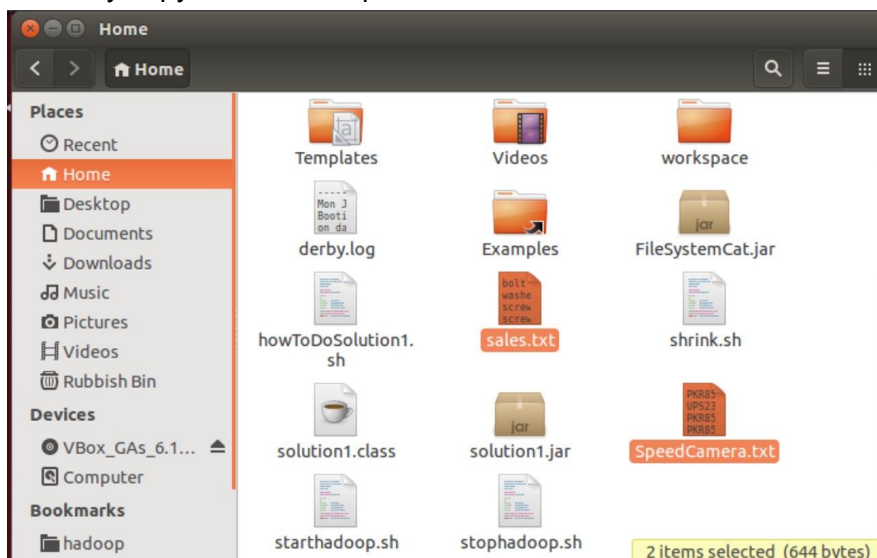
Create a folder called A1T1 and check if folder has been created correctly

`$HADOOP_HOME/bin/hadoop fs -mkdir A1T1`

`$HADOOP_HOME/bin/hadoop fs -ls`

```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -mkdir A1T1  
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -ls  
Found 4 items  
drwxr-xr-x - bigdata supergroup 0 2017-07-03 01:33 .sparkStaging  
drwxr-xr-x - bigdata supergroup 0 2022-10-11 15:54 A1T1  
drwxr-xr-x - bigdata supergroup 0 2022-10-07 17:39 myfolder  
drwxr-xr-x - bigdata supergroup 0 2022-10-11 15:13 test1  
bigdata@bigdata-VirtualBox:~$
```

Manually copy sales.txt & SpeedCamera.txt file to Home



Copy 2 .txt files from the local file system to HDFS

```
$HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/sales.txt A1T1
$HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/SpeedCamera.txt A1T1
```

```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/sales.txt A1T1
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/SpeedCamera.txt A1T1
bigdata@bigdata-VirtualBox:~$
```

Verify the results

```
ls $HADOOP_HOME/*.txt
```

```
bigdata@bigdata-VirtualBox:~$ ls $HADOOP_HOME/*.txt
/usr/share/hadoop/LICENSE.txt  /usr/share/hadoop/README.txt  /usr/share/hadoop/SpeedCamera.txt
/usr/share/hadoop/NOTICE.txt   /usr/share/hadoop/sales.txt
bigdata@bigdata-VirtualBox:~$
```

List the files in A1T1 folder in HDFS

```
$HADOOP_HOME/bin/hadoop fs -ls A1T1
```

```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -ls A1T1
Found 2 items
-rw-r--r--  1 bigdata supergroup      544 2022-10-11 15:59 A1T1/SpeedCamera.txt
-rw-r--r--  1 bigdata supergroup     100 2022-10-11 15:59 A1T1/sales.txt
bigdata@bigdata-VirtualBox:~$
```

Check data in both .txt files

```
$HADOOP_HOME/bin/hadoop fs -cat A1T1/sales.txt
```

```
$HADOOP_HOME/bin/hadoop fs -cat A1T1/SpeedCamera.txt
```

```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -cat A1T1/sales.txt
bolt 45
washer 3
screw 67
screw 23
nail 5
screw 78
bolt 5
bolt 1
drill 1
drill 1
washer 56
washer 7
bigdata@bigdata-VirtualBox:~$
```

```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -cat A1T1/SpeedCamera.txt
PKR856 14-NOV-2021 140
UPS234 20-FEB-2022 180
PKR856 20-MAR-2020 90
PKR856 17-JUN-2021 70
UPS234 22-SEP-2022 200
UPS234 03-AUG-2020 130
PKR856 14-NOV-2021 140
UPS234 20-FEB-2022 180
PKR856 20-MAR-2020 90
PKR856 17-JUN-2021 70
UPS234 22-SEP-2022 200
UPS234 03-AUG-2020 130
PKR856 14-NOV-2021 140
UPS234 20-FEB-2022 180
PKR856 20-MAR-2020 90
PKR856 17-JUN-2021 70
UPS234 22-SEP-2022 200
UPS234 03-AUG-2020 130
PKR856 14-NOV-2021 140
UPS234 20-FEB-2022 180
PKR856 20-MAR-2020 90
PKR856 17-JUN-2021 70
UPS234 22-SEP-2022 200
UPS234 03-AUG-2020 130
bigdata@bigdata-VirtualBox:~$
```

Use the following command to define the environment variable HADOOP_CLASSPATH

```
export HADOOP_CLASSPATH=$(HADOOP_HOME/bin/hadoop classpath)
javac -classpath ${HADOOP_CLASSPATH} solution1.java
jar cf solution2.jar solution1*.class
```

```
bigdata@bigdata-VirtualBox:~$ export HADOOP_CLASSPATH=$(HADOOP_HOME/bin/hadoop classpath)
bigdata@bigdata-VirtualBox:~$
```

```
bigdata@bigdata-VirtualBox:~$ javac -classpath $HADOOP_CLASSPATH solution1.java
bigdata@bigdata-VirtualBox:~$ jar cvf solution1.jar solution1.class
added manifest
adding: solution1.class(in = 1112) (out= 604)(deflated 45%)
bigdata@bigdata-VirtualBox:~$
```

Run the jar file by using the hadoop script and jar command.

A1T1/sales.txt is the 1st argument arg[0] from the java file

'A1T1/solution1.txt is the 2nd argument arg[1] from the java file.

solution1.txt is the 3rd argument arg[3] based on whatever name you want.

```
$HADOOP_HOME/bin/hadoop jar solution1.jar solution1 A1T1/sales.txt
```

```
A1T1/SpeedCamera.txt A1T1/solution.txt
```

```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop jar solution1.jar solution1 A1T1/s
ales.txt A1T1/SpeedCamera.txt A1T1/solution.txt
bigdata@bigdata-VirtualBox:~$
```

To see the results from processing of **solution3** applications, process the following command.

```
$HADOOP_HOME/bin/hadoop fs -cat /A1T1/output/*
```

```

bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -cat A1T1/solution.txt
bolt 45
washer 3
screw 67
screw 23
nail 5
screw 78
bolt 5
bolt 1
drill 1
drill 1
washer 56
washer 7
PKR856 14-NOV-2021 140
UPS234 20-FEB-2022 180
PKR856 20-MAR-2020 90
PKR856 17-JUN-2021 70
UPS234 22-SEP-2022 200
UPS234 03-AUG-2020 130
PKR856 14-NOV-2021 140
UPS234 20-FEB-2022 180
PKR856 20-MAR-2020 90
PKR856 17-JUN-2021 70
UPS234 22-SEP-2022 200
UPS234 03-AUG-2020 130
PKR856 14-NOV-2021 140
UPS234 20-FEB-2022 180
PKR856 20-MAR-2020 90
PKR856 17-JUN-2021 70
UPS234 22-SEP-2022 200
UPS234 03-AUG-2020 130
PKR856 14-NOV-2021 140
UPS234 20-FEB-2022 180
PKR856 20-MAR-2020 90
PKR856 17-JUN-2021 70
UPS234 22-SEP-2022 200
UPS234 03-AUG-2020 130
bigdata@bigdata-VirtualBox:~$ █

```

Shut down hadoop server

```

bigdata@bigdata-VirtualBox:~$ ./stophadoop.sh
./stophadoop.sh: line 1: fg: no job control
Stopping namenodes on [localhost]
localhost: stopping namenode
localhost: stopping datanode
Stopping secondary namenodes [0.0.0.0]
0.0.0.0: stopping secondarynamenode
stopping yarn daemons
stopping resourcemanager
localhost: stopping nodemanager
no proxyserver to stop
stopping historyserver
3060 RemoteInterpreterServer
2957 RemoteInterpreterServer
2734 ZeppelinServer
26143 Jps
bigdata@bigdata-VirtualBox:~$ █

```


A1 Task 2

Start hadoop

```
./starthadoop.sh
```

```
bigdata@bigdata-VirtualBox: ~  
bigdata@bigdata-VirtualBox:~$ ./starthadoop.sh  
./starthadoop.sh: line 1: fg: no job control  
Starting namenodes on [localhost]  
localhost: namenode running as process 26372. Stop it first.  
localhost: datanode running as process 26522. Stop it first.  
Starting secondary namenodes [0.0.0.0]  
0.0.0.0: secondarynamenode running as process 26740. Stop it first.  
starting yarn daemons  
resourcemanager running as process 26946. Stop it first.  
localhost: nodemanager running as process 27107. Stop it first.  
historyserver running as process 27350. Stop it first.  
30976 Jps  
26946 ResourceManager  
27107 NodeManager  
26372 NameNode  
26740 SecondaryNameNode  
3060 RemoteInterpreterServer  
27350 JobHistoryServer  
26522 DataNode  
2957 RemoteInterpreterServer  
2734 ZeppelinServer
```

Create a folder & check for the folder

```
$HADOOP_HOME/bin/hadoop fs -mkdir a1t2
```

```
$HADOOP_HOME/bin/hadoop fs -ls
```

```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -mkdir a1t2  
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -ls  
Found 8 items  
drwxr-xr-x - bigdata supergroup 0 2017-07-03 01:33 .sparkStaging  
drwxr-xr-x - bigdata supergroup 0 2022-10-11 15:59 A1T1  
drwxr-xr-x - bigdata supergroup 0 2022-10-14 16:27 a1t2  
drwxr-xr-x - bigdata supergroup 0 2022-10-14 14:11 a1t3  
drwxr-xr-x - bigdata supergroup 0 2022-10-12 02:10 input  
drwxr-xr-x - bigdata supergroup 0 2022-10-07 17:39 myfolder  
-rw-r--r-- 1 bigdata supergroup 100 2022-10-14 15:23 sales.txt  
drwxr-xr-x - bigdata supergroup 0 2022-10-11 16:16 test1  
bigdata@bigdata-VirtualBox:~$
```

Copy a file from the local file system to HDFS. change sales.txt to .txt name

```
$HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/speedCamera.txt a1t2
```

```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/SpeedCamera.txt a1t2  
bigdata@bigdata-VirtualBox:~$
```

Check the files inside a1t2 directory

```
$HADOOP_HOME/bin/hadoop fs -ls $HADOOP_HOME/ a1t3
```

```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -ls $HADOOP_HOME/ a1t2  
ls: '/usr/share/hadoop/': No such file or directory  
Found 1 items  
-rw-r--r-- 1 bigdata supergroup 544 2022-10-14 16:29 a1t2/SpeedCamera.txt  
bigdata@bigdata-VirtualBox:~$
```

Compile the java file solution3.java

```
export HADOOP_CLASSPATH=$(HADOOP_HOME/bin/hadoop classpath)
javac -classpath ${HADOOP_CLASSPATH} solution2.java
jar cf solution2.jar solution2*.class
```

```
bigdata@bigdata-VirtualBox:~$ export HADOOP_CLASSPATH=$(HADOOP_HOME/bin/hadoop classpath)
bigdata@bigdata-VirtualBox:~$ javac -classpath ${HADOOP_CLASSPATH} solution2.java
bigdata@bigdata-VirtualBox:~$ jar cf solution2.jar solution2*.class
bigdata@bigdata-VirtualBox:~$
```

Upload to HDFS home folder a file sales.txt and list its contents in HDFS.

```
$HADOOP_HOME/bin/hadoop fs -put SpeedCamera.txt
$HADOOP_HOME/bin/hadoop fs -cat SpeedCamera.txt
```

```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -put SpeedCamera.txt
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -cat SpeedCamera.txt
PKR856 14-NOV-2021 140
UPS234 20-FEB-2022 180
PKR856 20-MAR-2020 90
PKR856 17-JUN-2021 70
UPS234 22-SEP-2022 200
UPS234 03-AUG-2020 130
PKR856 14-NOV-2021 140
UPS234 20-FEB-2022 180
PKR856 20-MAR-2020 90
PKR856 17-JUN-2021 70
UPS234 22-SEP-2022 200
UPS234 03-AUG-2020 130
PKR856 14-NOV-2021 140
UPS234 20-FEB-2022 180
PKR856 20-MAR-2020 90
PKR856 17-JUN-2021 70
UPS234 22-SEP-2022 200
UPS234 03-AUG-2020 130
PKR856 14-NOV-2021 140
UPS234 20-FEB-2022 180
PKR856 20-MAR-2020 90
PKR856 17-JUN-2021 70
UPS234 22-SEP-2022 200
UPS234 03-AUG-2020 130
bigdata@bigdata-VirtualBox:~$
```

Finally, start the MinMax application in the following way.

```
$HADOOP_HOME/bin/hadoop jar solution2.jar solution2 SpeedCamera.txt
/user/output
```

```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop jar solution2.jar solution2 SpeedCamera.txt /user/output
22/10/14 17:03:06 INFO client.RMProxy: Connecting to ResourceManager at bigdata-VirtualBox/10.0.2.15:8032
22/10/14 17:03:06 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/10/14 17:03:06 INFO input.FileInputFormat: Total input paths to process : 1
22/10/14 17:03:07 INFO mapreduce.JobSubmitter: number of splits:1
22/10/14 17:03:07 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1665725230768_0004
```



```

22/10/14 17:22:51 INFO impl.YarnClientImpl: Submitted application application_1665725230768_0009
22/10/14 17:22:51 INFO mapreduce.Job: The url to track the job: http://bigdata-VirtualBox:8088/proxy/application_1665725230768_0009/
22/10/14 17:22:51 INFO mapreduce.Job: Running job: job_1665725230768_0009
22/10/14 17:22:57 INFO mapreduce.Job: Job job_1665725230768_0009 running in uber mode : false
22/10/14 17:22:57 INFO mapreduce.Job:  map 0% reduce 0%
22/10/14 17:23:01 INFO mapreduce.Job:  map 100% reduce 0%
22/10/14 17:23:06 INFO mapreduce.Job:  map 100% reduce 100%
22/10/14 17:23:07 INFO mapreduce.Job: Job job_1665725230768_0009 completed successfully
22/10/14 17:23:07 INFO mapreduce.Job: Counters: 49

```

File System Counters

```

FILE: Number of bytes read=32
FILE: Number of bytes written=237613
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=659
HDFS: Number of bytes written=22
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)=2001
Total time spent by all reduces in occupied slots (ms)=2302
Total time spent by all map tasks (ms)=2001
Total time spent by all reduce tasks (ms)=2302
Total vcore-milliseconds taken by all map tasks=2001
Total vcore-milliseconds taken by all reduce tasks=2302
Total megabyte-milliseconds taken by all map tasks=2049024
Total megabyte-milliseconds taken by all reduce tasks=2357248

```

Map-Reduce Framework

```

Map input records=24
Map output records=24
Map output bytes=264
Map output materialized bytes=32
Input split bytes=115
Combine input records=24
Combine output records=2
Reduce input groups=2
Reduce shuffle bytes=32
Reduce input records=2
Reduce output records=2
Spilled Records=4
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=99
CPU time spent (ms)=1430
Physical memory (bytes) snapshot=449232896
Virtual memory (bytes) snapshot=3843444736
Total committed heap usage (bytes)=348127232

```

```

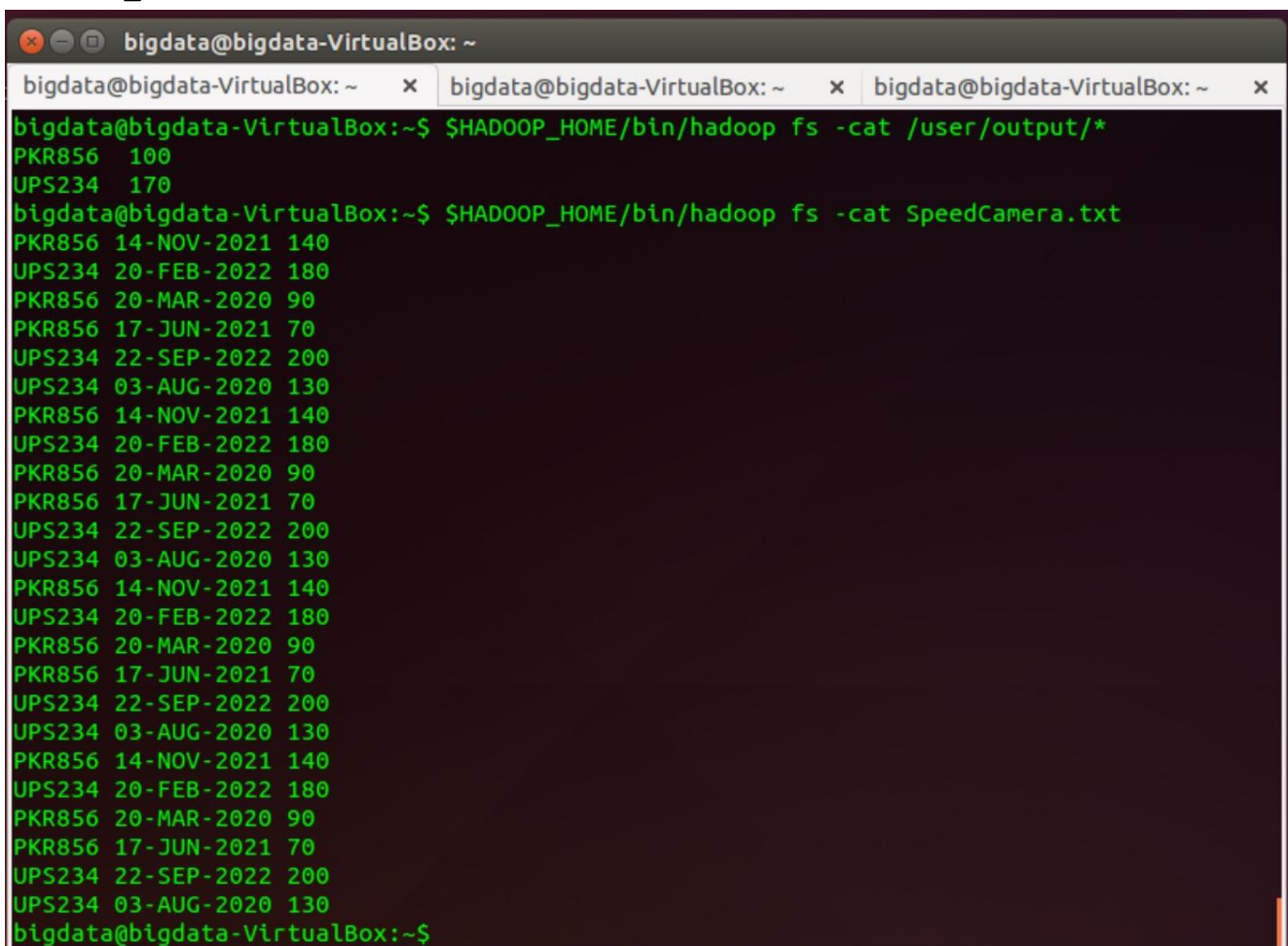
Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
File Input Format Counters
    Bytes Read=544
File Output Format Counters
    Bytes Written=22

```

To see the results from processing of **solution3** applications, process the following command.

```
$HADOOP_HOME/bin/hadoop fs -cat /user/output/*
```

```
$HADOOP_HOME/bin/hadoop fs -cat SpeedCamera.txt
```



```

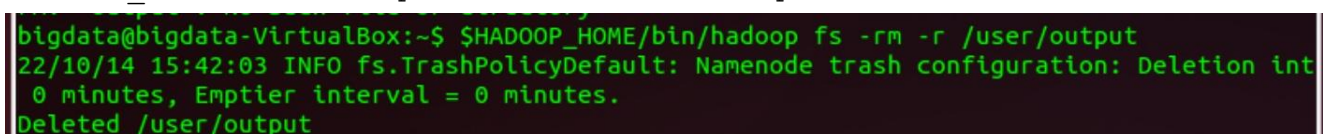
bigdata@bigdata-VirtualBox: ~
bigdata@bigdata-VirtualBox: ~ x bigdata@bigdata-VirtualBox: ~ x bigdata@bigdata-VirtualBox: ~ x
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -cat /user/output/*
PKR856 100
UPS234 170
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -cat SpeedCamera.txt
PKR856 14-NOV-2021 140
UPS234 20-FEB-2022 180
PKR856 20-MAR-2020 90
PKR856 17-JUN-2021 70
UPS234 22-SEP-2022 200
UPS234 03-AUG-2020 130
PKR856 14-NOV-2021 140
UPS234 20-FEB-2022 180
PKR856 20-MAR-2020 90
PKR856 17-JUN-2021 70
UPS234 22-SEP-2022 200
UPS234 03-AUG-2020 130
PKR856 14-NOV-2021 140
UPS234 20-FEB-2022 180
PKR856 20-MAR-2020 90
PKR856 17-JUN-2021 70
UPS234 22-SEP-2022 200
UPS234 03-AUG-2020 130
PKR856 14-NOV-2021 140
UPS234 20-FEB-2022 180
PKR856 20-MAR-2020 90
PKR856 17-JUN-2021 70
UPS234 22-SEP-2022 200
UPS234 03-AUG-2020 130
bigdata@bigdata-VirtualBox:~$

```

Remove the folders input and output from HDFS.

```
$HADOOP_HOME/bin/hadoop fs -rm -r /user/output/*
```

```
$HADOOP_HOME/bin/hadoop fs -rmdir /user/output
```



```

bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -rm -r /user/output
22/10/14 15:42:03 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion int
0 minutes, Emptier interval = 0 minutes.
Deleted /user/output

```


Shut down hadoop server

./stophadoop.sh

```
bigdata@bigdata-VirtualBox:~$ ./stophadoop.sh
./stophadoop.sh: line 1: fg: no job control
Stopping namenodes on [localhost]
localhost: stopping namenode
localhost: stopping datanode
Stopping secondary namenodes [0.0.0.0]
0.0.0.0: stopping secondarynamenode
stopping yarn daemons
stopping resourcemanager
localhost: stopping nodemanager
no proxyserver to stop
stopping historyserver
3060 RemoteInterpreterServer
2957 RemoteInterpreterServer
2734 ZeppelinServer
26143 Jps
bigdata@bigdata-VirtualBox:~$
```

A1 Task 3

Start hadoop

```
./starthadoop.sh
```

```
bigdata@bigdata-VirtualBox: ~  
bigdata@bigdata-VirtualBox:~$ ./starthadoop.sh  
./starthadoop.sh: line 1: fg: no job control  
Starting namenodes on [localhost]  
localhost: namenode running as process 26372. Stop it first.  
localhost: datanode running as process 26522. Stop it first.  
Starting secondary namenodes [0.0.0.0]  
0.0.0.0: secondarynamenode running as process 26740. Stop it first.  
starting yarn daemons  
resourcemanager running as process 26946. Stop it first.  
localhost: nodemanager running as process 27107. Stop it first.  
historyserver running as process 27350. Stop it first.  
30976 Jps  
26946 ResourceManager  
27107 NodeManager  
26372 NameNode  
26740 SecondaryNameNode  
3060 RemoteInterpreterServer  
27350 JobHistoryServer  
26522 DataNode  
2957 RemoteInterpreterServer  
2734 ZeppelinServer
```

Create a folder & check for the folder

```
$HADOOP_HOME/bin/hadoop fs -mkdir a1t3
```

```
$HADOOP_HOME/bin/hadoop fs -ls
```

```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -mkdir a1t3  
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -ls  
Found 6 items  
drwxr-xr-x - bigdata supergroup 0 2017-07-03 01:33 .sparkStaging  
drwxr-xr-x - bigdata supergroup 0 2022-10-11 15:59 A1T1  
drwxr-xr-x - bigdata supergroup 0 2022-10-14 14:09 a1t3  
drwxr-xr-x - bigdata supergroup 0 2022-10-12 02:10 input  
drwxr-xr-x - bigdata supergroup 0 2022-10-07 17:39 myfolder  
drwxr-xr-x - bigdata supergroup 0 2022-10-11 16:16 test1  
bigdata@bigdata-VirtualBox:~$
```

Copy a file from the local file system to HDFS. change sales.txt to .txt name

```
$HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/sales.txt a1t3
```

```
$HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/SpeedCamera.txt a1t3
```

```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/sales.txt a1t3  
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/SpeedCamera.txt a1t3  
bigdata@bigdata-VirtualBox:~$
```

Check the files inside a1t3 directory

```
$HADOOP_HOME/bin/hadoop fs -ls $HADOOP_HOME/ a1t3
```

```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -ls $HADOOP_HOME/ a1t3  
ls: '/usr/share/hadoop/': No such file or directory  
Found 2 items  
-rw-r--r-- 1 bigdata supergroup 544 2022-10-14 14:11 a1t3/SpeedCamera.txt  
-rw-r--r-- 1 bigdata supergroup 100 2022-10-14 14:11 a1t3/sales.txt  
bigdata@bigdata-VirtualBox:~$
```

Compile the java file solution3.java

```
export HADOOP_CLASSPATH=$(HADOOP_HOME/bin/hadoop classpath)
```

```
javac -classpath ${HADOOP_CLASSPATH} solution3.java
```

```
jar cf solution3.jar solution3*.class
```

```
bigdata@bigdata-VirtualBox:~$ export HADOOP_CLASSPATH=$(HADOOP_HOME/bin/hadoop classpath)
bigdata@bigdata-VirtualBox:~$ javac -classpath ${HADOOP_CLASSPATH} solution3.java
bigdata@bigdata-VirtualBox:~$ jar cf solution3.jar solution3*.class
bigdata@bigdata-VirtualBox:~$
```

Upload to HDFS home folder a file sales.txt and list its contents in HDFS.

```
$HADOOP_HOME/bin/hadoop fs -put sales.txt
```

```
$HADOOP_HOME/bin/hadoop fs -cat sales.txt
```

```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -put sales.txt
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -cat sales.txt
bolt 45
washer 3
screw 67
screw 23
nail 5
screw 78
bolt 5
bolt 1
drill 1
drill 1
washer 56
washer 7
bigdata@bigdata-VirtualBox:~$
```

Before processing a solution3 application, make sure that a folder /user/output does not exist in HDFS.

```
$HADOOP_HOME/bin/hadoop fs -ls /user
```

```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -ls /user
Found 2 items
drwxr-xr-x - bigdata supergroup 0 2022-10-14 15:23 /user/bigdata
drwxr-xr-x - bigdata supergroup 0 2017-07-03 01:29 /user/hive
bigdata@bigdata-VirtualBox:~$
```


Finally, start the MinMax application in the following way.

```
$HADOOP_HOME/bin/hadoop jar solution3.jar solution3 sales.txt /user/output
```

```
bigdata@bigdata-VirtualBox: ~
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop jar solution3.jar solution3 sales.txt /user/output
22/10/14 15:25:46 INFO client.RMPProxy: Connecting to ResourceManager at bigdata-VirtualBox/10.0.2.15:8032
22/10/14 15:25:46 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/10/14 15:25:46 INFO input.FileInputFormat: Total input paths to process : 1
22/10/14 15:25:47 INFO mapreduce.JobSubmitter: number of splits:1
22/10/14 15:25:47 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1665500816114_0002
22/10/14 15:25:47 INFO impl.YarnClientImpl: Submitted application application_1665500816114_0002
22/10/14 15:25:47 INFO mapreduce.Job: The url to track the job: http://bigdata-VirtualBox:8088/proxy/application_1665500816114_0002/
22/10/14 15:25:47 INFO mapreduce.Job: Running job: job_1665500816114_0002
22/10/14 15:25:53 INFO mapreduce.Job: Job job_1665500816114_0002 running in uber mode : false
22/10/14 15:25:53 INFO mapreduce.Job: map 0% reduce 0%
22/10/14 15:25:58 INFO mapreduce.Job: map 100% reduce 0%
22/10/14 15:26:03 INFO mapreduce.Job: map 100% reduce 100%
22/10/14 15:26:04 INFO mapreduce.Job: Job job_1665500816114_0002 completed successfully
22/10/14 15:26:04 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=242
    FILE: Number of bytes written=238027
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=209
    HDFS: Number of bytes written=171
    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)=2538
    Total time spent by all reduces in occupied slots (ms)=2333
    Total time spent by all map tasks (ms)=2538
    Total time spent by all reduce tasks (ms)=2333
    Total vcore-milliseconds taken by all map tasks=2538
    Total vcore-milliseconds taken by all reduce tasks=2333
    Total megabyte-milliseconds taken by all map tasks=2598912
    Total megabyte-milliseconds taken by all reduce tasks=2388992

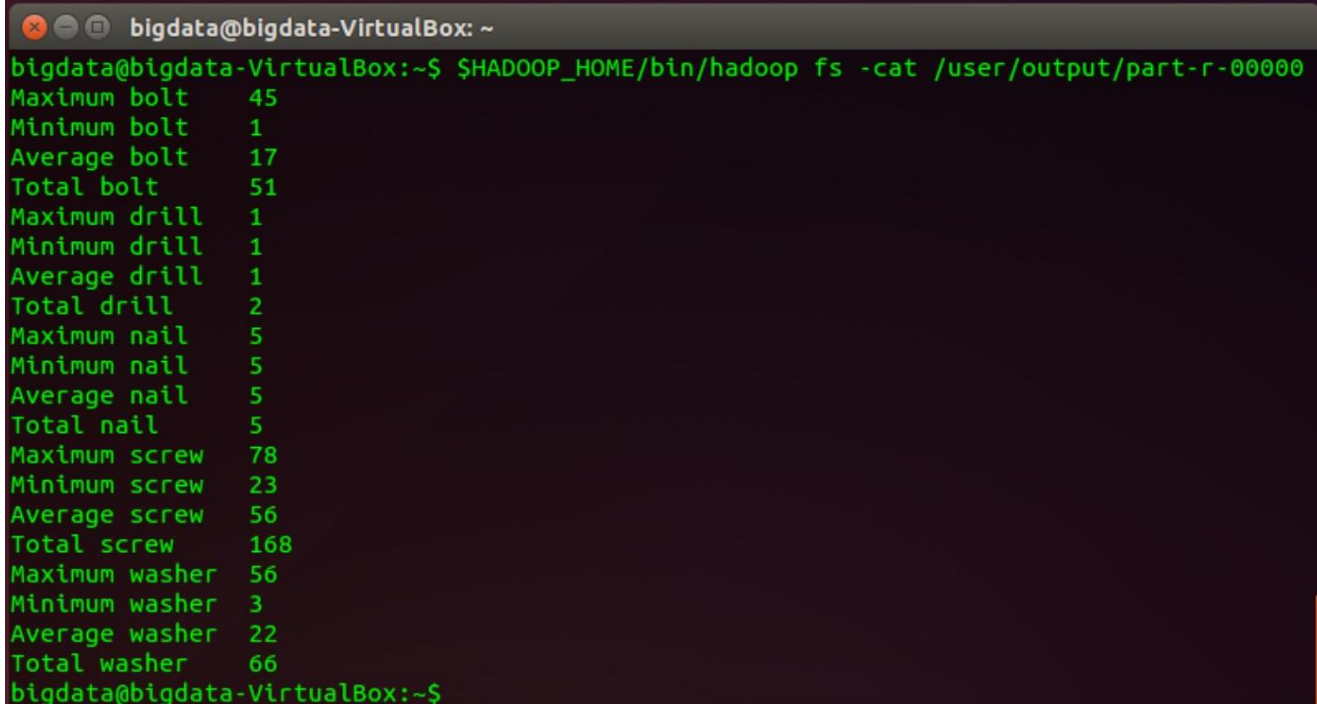
  Map-Reduce Framework
    Map input records=12
    Map output records=12
    Map output bytes=119
    Map output materialized bytes=242
    Input split bytes=109
    Combine input records=12
    Combine output records=20
    Reduce input groups=5
    Reduce shuffle bytes=242
    Reduce input records=20
    Reduce output records=20
    Spilled Records=40
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=123
    CPU time spent (ms)=1540
    Physical memory (bytes) snapshot=444555264
    Virtual memory (bytes) snapshot=3843600384
    Total committed heap usage (bytes)=347602944

  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=100
  File Output Format Counters
    Bytes Written=171
bigdata@bigdata-VirtualBox:~$
```

To see the results from processing of **solution3** applications, process the following command.

```
$HADOOP_HOME/bin/hadoop fs -cat /user/output/part-r-00000
```

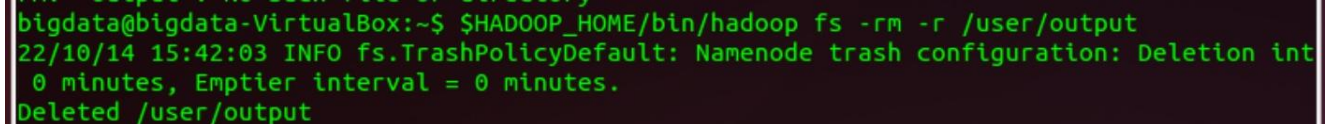


```
bigdata@bigdata-VirtualBox: ~  
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -cat /user/output/part-r-00000  
Maximum bolt      45  
Minimum bolt      1  
Average bolt      17  
Total bolt        51  
Maximum drill     1  
Minimum drill     1  
Average drill     1  
Total drill       2  
Maximum nail      5  
Minimum nail      5  
Average nail      5  
Total nail        5  
Maximum screw     78  
Minimum screw     23  
Average screw     56  
Total screw       168  
Maximum washer    56  
Minimum washer    3  
Average washer    22  
Total washer      66  
bigdata@bigdata-VirtualBox:~$
```

Remove the folders input and output from HDFS.

```
$HADOOP_HOME/bin/hadoop fs -rm -r /user/output/*
```

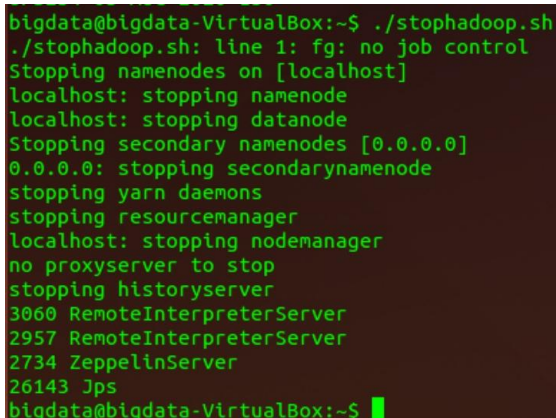
```
$HADOOP_HOME/bin/hadoop fs -rmdir /user/output
```



```
bigdata@bigdata-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -rm -r /user/output  
22/10/14 15:42:03 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion int  
0 minutes, Emptier interval = 0 minutes.  
Deleted /user/output
```

Shut down hadoop server

```
./stophadoop.sh
```



```
bigdata@bigdata-VirtualBox:~$ ./stophadoop.sh  
./stophadoop.sh: line 1: fg: no job control  
Stopping namenodes on [localhost]  
localhost: stopping namenode  
localhost: stopping datanode  
Stopping secondary namenodes [0.0.0.0]  
0.0.0.0: stopping secondarynamenode  
stopping yarn daemons  
stopping resourcemanager  
localhost: stopping nodemanager  
no proxyserver to stop  
stopping historyserver  
3060 RemoteInterpreterServer  
2957 RemoteInterpreterServer  
2734 ZeppelinServer  
26143 Jps  
bigdata@bigdata-VirtualBox:~$
```