Prerequisites:

Hadoop was installed to Ubuntu 20.04 virtual machine.

Guide used: https://phoenixnap.com/kb/install-hadoop-ubuntu

Hadoop version 3.2.4

New Hadoop user was created: hdoop

Steps 1-3:

Building a jar file.

```
hdoop@kyrbl-VirtualBox:~$ mkdir units
hdoop@kyrbl-VirtualBox:~$ javac -classpath hadoop-core-1.2.1.jar -d units ProcessUnits.java
hdoop@kyrbl-VirtualBox:~$ jar -cvf units.jar -C units/ .
added manifest
adding: hadoop/(in = 0) (out= 0)(stored 0%)
adding: hadoop/ProcessUnits.class(in = 1567) (out= 768)(deflated 50%)
adding: hadoop/ProcessUnits$E_EReduce.class(in = 1671) (out= 686)(deflated 58%)
adding: hadoop/ProcessUnits$E_EMapper.class(in = 1898) (out= 775)(deflated 59%)
```

```
hdoop@kyrbl-VirtualBox:~$ ls -l
total 484964
drwxrwxr-x  3 hdoop hdoop      4096 січ 17 13:09 dfsdata
drwxr-xr-x 10 hdoop hdoop      4096 січ 17 13:08 hadoop-3.2.4
-rw-rw-r--  1 hdoop hdoop 492368219 лип 22 05:06 hadoop-3.2.4.tar.gz
-rw-rw-r--  1 hdoop hdoop   4203713 лип 24  2013 hadoop-core-1.2.1.jar
-rw-rw-r--  1 hdoop hdoop      2975 січ 17 14:08 ProcessUnits.java
drwxrwxr-x  4 hdoop hdoop      4096 січ 17 13:10 tmpdata
drwxrwxr-x  3 hdoop hdoop      4096 січ 17 14:13 units
-rw-rw-r--  1 hdoop hdoop      3121 січ 17 14:13 units.jar
```

Steps 4-6:

Creating an input folder in Hadoop filesystem and putting out input there.

```
hdoop@kyrbl-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -mkdir -p input_dir
2023-01-17 14:28:44,855 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java clas
ses where applicable
```

```
hdoop@kyrbl-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -put sample.txt input_dir
2023-01-17 14:31:43,308 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java clas
ses where applicable
```

```
hdoop@kyrbl-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -ls input_dir/
2023-01-17 14:32:37,811 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java clas
ses where applicable
Found 1 items
-rw-r--r--   1 hdoop supergroup        219 2023-01-17 14:31 input_dir/sample.txt
```

Step 7

Running a given script.

hdoop@kyrbl-VirtualBox:~$ $HADOOP_HOME/bin/hadoop jar units.jar hadoop.ProcessUnits input_dir output_dir

2023-01-17 14:34:12,004 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

2023-01-17 14:34:13,036 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032

2023-01-17 14:34:13,424 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032

2023-01-17 14:34:13,731 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

2023-01-17 14:34:13,773 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_1673953831696_0001

2023-01-17 14:34:14,185 INFO mapred.FileInputFormat: Total input files to process : 1

2023-01-17 14:34:14,689 INFO mapreduce.JobSubmitter: number of splits:2

2023-01-17 14:34:14,962 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1673953831696_0001

2023-01-17 14:34:14,963 INFO mapreduce.JobSubmitter: Executing with tokens: []

2023-01-17 14:34:15,305 INFO conf.Configuration: resource-types.xml not found

2023-01-17 14:34:15,306 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

2023-01-17 14:34:15,635 INFO impl.YarnClientImpl: Submitted application application_1673953831696_0001

2023-01-17 14:34:15,775 INFO mapreduce.Job: The url to track the job: http://kyrbl-VirtualBox:8088/proxy/application_1673953831696_0001/

2023-01-17 14:34:15,797 INFO mapreduce.Job: Running job: job_1673953831696_0001

2023-01-17 14:34:29,291 INFO mapreduce.Job: Job job_1673953831696_0001 running in uber mode : false

2023-01-17 14:34:29,295 INFO mapreduce.Job: map 0% reduce 0%

2023-01-17 14:34:40,616 INFO mapreduce.Job: map 100% reduce 0%

2023-01-17 14:34:47,710 INFO mapreduce.Job: map 100% reduce 100%

2023-01-17 14:34:48,739 INFO mapreduce.Job: Job job_1673953831696_0001 completed successfully

2023-01-17 14:34:48,933 INFO mapreduce.Job: Counters: 54

File System Counters

FILE: Number of bytes read=61

FILE: Number of bytes written=713779

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=539

HDFS: Number of bytes written=75

HDFS: Number of read operations=11

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=2

Launched reduce tasks=1

Data-local map tasks=2

Total time spent by all maps in occupied slots (ms)=18242

Total time spent by all reduces in occupied slots (ms)=4184

Total time spent by all map tasks (ms)=18242

Total time spent by all reduce tasks (ms)=4184

Total vcore-milliseconds taken by all map tasks=18242

Total vcore-milliseconds taken by all reduce tasks=4184

Total megabyte-milliseconds taken by all map tasks=18679808

Total megabyte-milliseconds taken by all reduce tasks=4284416

Map-Reduce Framework

Map input records=5

Map output records=65

Map output bytes=585

Map output materialized bytes=67

Input split bytes=210

Combine input records=65

Combine output records=5

Reduce input groups=5

Reduce shuffle bytes=67

Reduce input records=5

Reduce output records=5

Spilled Records=10

Shuffled Maps =2

Failed Shuffles=0

Merged Map outputs=2

GC time elapsed (ms)=250

CPU time spent (ms)=1530

Physical memory (bytes) snapshot=606994432

Virtual memory (bytes) snapshot=7453040640

Total committed heap usage (bytes)=489889792

Peak Map Physical memory (bytes)=237015040

Peak Map Virtual memory (bytes)=2482147328

Peak Reduce Physical memory (bytes)=133443584

Peak Reduce Virtual memory (bytes)=2488745984

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=329

File Output Format Counters

Bytes Written=75

Steps 8-9:

Verifying results.

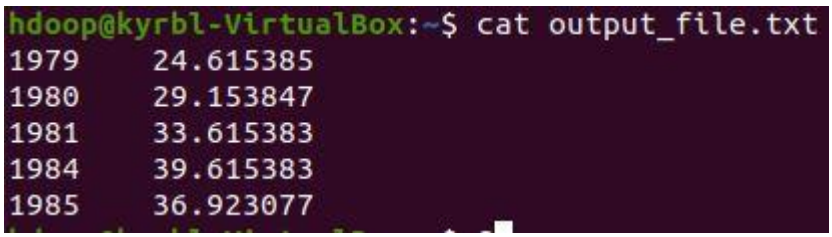
```
hdoop@kyrbl-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -ls output_dir/
2023-01-17 14:37:41,706 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 hdoop supergroup          0 2023-01-17 14:34 output_dir/_SUCCESS
-rw-r--r--   1 hdoop supergroup         75 2023-01-17 14:34 output_dir/part-00000
```

```
hdoop@kyrbl-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -cat output_dir/part-00000
2023-01-17 14:38:20,958 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1979    24.615385
1980    29.153847
1981    33.615383
1984    39.615383
1985    36.923077
```

Step 10:

Saving results to file system.

```
hdoop@kyrbl-VirtualBox:~$ $HADOOP_HOME/bin/hadoop fs -cat output_dir/part-00000 > output_file.txt
2023-01-17 14:41:29,263 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```


hdoop@kyrbl-VirtualBox:~$ cat output_file.txt
1979 24.615385
1980 29.153847
1981 33.615383
1984 39.615383
1985 36.923077