Step3

```
[root@23052bbaa7cd:/home/university/data/my4lab# jar -cvf units.jar -C units/ .
added manifest
adding: hadoop/(in = 0) (out= 0)(stored 0%)
adding: hadoop/ProcessUnits$E_EMapper.class(in = 1898) (out= 775)(deflated 59%)
adding: hadoop/ProcessUnits$E_EReduce.class(in = 1671) (out= 686)(deflated 58%)
adding: hadoop/ProcessUnits.class(in = 1567) (out= 768)(deflated 50%)
```

Install Hadoop

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

```
[root@23052bbaa7cd:/home/university/data# wget https://downloads.apache.org/hadoo
p/common/hadoop-3.2.1/hadoop-3.2.1.tar.gz
--2020-12-10 19:20:19--  https://downloads.apache.org/hadoop/common/hadoop-3.2.1
/hadoop-3.2.1.tar.gz
Resolving downloads.apache.org (downloads.apache.org)... 88.99.95.219, 2a01:4f8:
10a:201a::2
Connecting to downloads.apache.org (downloads.apache.org)|88.99.95.219|:443... c
onnected.
HTTP request sent, awaiting response... 200 OK
Length: 359196911 (343M) [application/x-gzip]
Saving to: 'hadoop-3.2.1.tar.gz'

hadoop-3.2.1.tar.gz  22%[===>                 ]  77.77M   745KB/s    eta 8m 9s
```

```
#Hadoop Related Options

export HADOOP_HOME=/home/hdoop/hadoop-3.2.1
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```

```
sudo nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

Step 6

```
hdoop@23052bbaa7cd:~$ $HADOOP_HOME/bin/hadoop fs -ls input_dir/
Found 1 items
-rw-r--r--   1 hdoop hdoop        218 2020-12-10 23:42 input_dir/sample.txt
```

Step 7

```
hdoop@23052bbaa7cd:~$ $HADOOP_HOME/bin/hadoop jar units.jar hadoop.ProcessUnits input_dir output_dir
2020-12-10 23:43:55,581 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2020-12-10 23:43:55,627 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2020-12-10 23:43:55,627 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2020-12-10 23:43:55,634 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2020-12-10 23:43:55,666 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRu
nner to remedy this.
2020-12-10 23:43:55,697 INFO mapred.FileInputFormat: Total input files to process : 1
2020-12-10 23:43:55,705 INFO mapreduce.JobSubmitter: number of splits:1
2020-12-10 23:43:55,766 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local2127575678_0001
2020-12-10 23:43:55,766 INFO mapreduce.JobSubmitter: Executing with tokens: []
2020-12-10 23:43:55,822 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2020-12-10 23:43:55,822 INFO mapreduce.Job: Running job: job_local2127575678_0001
2020-12-10 23:43:55,823 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2020-12-10 23:43:55,824 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter
2020-12-10 23:43:55,828 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2020-12-10 23:43:55,828 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2020-12-10 23:43:55,844 INFO mapred.LocalJobRunner: Waiting for map tasks
2020-12-10 23:43:55,847 INFO mapred.LocalJobRunner: Starting task: attempt_local2127575678_0001_m_000000_0
2020-12-10 23:43:55,864 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2020-12-10 23:43:55,864 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2020-12-10 23:43:55,871 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
2020-12-10 23:43:55,879 INFO mapred.MapTask: Processing split: file:/home/hdoop/input_dir/sample.txt:0+218
2020-12-10 23:43:55,885 INFO mapred.MapTask: numReduceTasks: 1
2020-12-10 23:43:55,904 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2020-12-10 23:43:55,904 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2020-12-10 23:43:55,904 INFO mapred.MapTask: soft limit at 83886080
2020-12-10 23:43:55,904 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2020-12-10 23:43:55,904 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2020-12-10 23:43:55,906 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2020-12-10 23:43:55,910 INFO mapred.LocalJobRunner:
2020-12-10 23:43:55,910 INFO mapred.MapTask: Starting flush of map output
2020-12-10 23:43:55,910 INFO mapred.MapTask: Spilling map output
2020-12-10 23:43:55,910 INFO mapred.MapTask: bufstart = 0; bufend = 585; bufvoid = 104857600
2020-12-10 23:43:55,910 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214140(104856560); length = 257/6553600
2020-12-10 23:43:55,914 INFO mapred.MapTask: Finished spill 0
2020-12-10 23:43:55,918 INFO mapred.Task: Task:attempt_local2127575678_0001_m_000000_0 is done. And is in the process of committing
2020-12-10 23:43:55,919 INFO mapred.LocalJobRunner: file:/home/hdoop/input_dir/sample.txt:0+218
2020-12-10 23:43:55,919 INFO mapred.Task: Task 'attempt_local2127575678_0001_m_000000_0' done.
2020-12-10 23:43:55,922 INFO mapred.Task: Final Counters for attempt_local2127575678_0001_m_000000_0: Counters: 18
```

```
2020-12-10 23:43:55,919 INFO mapred.Task: Task 'attempt_local2127575678_0001_m_000000_0' done.
2020-12-10 23:43:55,922 INFO mapred.Task: Final Counters for attempt_local2127575678_0001_m_000000_0: Counters: 18
        File System Counters
                FILE: Number of bytes read=3503
                FILE: Number of bytes written=524562
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
        Map-Reduce Framework
                Map input records=5
                Map output records=65
                Map output bytes=585
                Map output materialized bytes=61
                Input split bytes=89
                Combine input records=65
                Combine output records=5
                Spilled Records=5
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=5
                Total committed heap usage (bytes)=1012924416
        File Input Format Counters
                Bytes Read=238
2020-12-10 23:43:55,922 INFO mapred.LocalJobRunner: Finishing task: attempt_local2127575678_0001_m_000000_0
2020-12-10 23:43:55,922 INFO mapred.LocalJobRunner: map task executor complete.
2020-12-10 23:43:55,924 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2020-12-10 23:43:55,924 INFO mapred.LocalJobRunner: Starting task: attempt_local2127575678_0001_r_000000_0
2020-12-10 23:43:55,928 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2020-12-10 23:43:55,928 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2020-12-10 23:43:55,929 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
2020-12-10 23:43:55,931 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@6a97b575
2020-12-10 23:43:55,932 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2020-12-10 23:43:55,940 INFO reduce.MergeManagerImpl: The max number of bytes for a single in-memory shuffle cannot be larger than Integer.MAX_VALUE. Setting it to Integer.MAX_VAL
UE
2020-12-10 23:43:55,940 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=10505787392, maxSingleShuffleLimit=2147483647, mergeThreshold=6933819904, ioSortFactor=10, memToMe
mMergeOutputsThreshold=10
2020-12-10 23:43:55,941 INFO reduce.EventFetcher: attempt_local2127575678_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2020-12-10 23:43:55,952 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local2127575678_0001_m_000000_0 decomp: 57 len: 61 to MEMORY
2020-12-10 23:43:55,953 INFO reduce.InMemoryMapOutput: Read 57 bytes from map-output for attempt_local2127575678_0001_m_000000_0
2020-12-10 23:43:55,954 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 57, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->57
2020-12-10 23:43:55,954 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
2020-12-10 23:43:55,954 INFO mapred.LocalJobRunner: 1 / 1 copied.
2020-12-10 23:43:55,955 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
2020-12-10 23:43:55,958 INFO mapred.Merger: Merging 1 sorted segments
2020-12-10 23:43:55,958 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 50 bytes
2020-12-10 23:43:55,959 INFO reduce.MergeManagerImpl: Merged 1 segments, 57 bytes to disk to satisfy reduce memory limit
2020-12-10 23:43:55,959 INFO reduce.MergeManagerImpl: Merging 1 files, 61 bytes from disk
2020-12-10 23:43:55,959 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2020-12-10 23:43:55,959 INFO mapred.Merger: Merging 1 sorted segments
2020-12-10 23:43:55,960 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 50 bytes
```

```
2020-12-10 23:43:55,959 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2020-12-10 23:43:55,959 INFO mapred.Merger: Merging 1 sorted segments
2020-12-10 23:43:55,960 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 50 bytes
2020-12-10 23:43:55,960 INFO mapred.LocalJobRunner: 1 / 1 copied.
2020-12-10 23:43:55,965 INFO mapred.Task: Task:attempt_local2127575678_0001_r_000000_0 is done. And is in the process of committing
2020-12-10 23:43:55,965 INFO mapred.LocalJobRunner: 1 / 1 copied.
2020-12-10 23:43:55,965 INFO mapred.Task: Task attempt_local2127575678_0001_r_000000_0 is allowed to commit now
2020-12-10 23:43:55,966 INFO output.FileOutputCommitter: Saved output of task 'attempt_local2127575678_0001_r_000000_0' to file:/home/hdoop/output_dir
2020-12-10 23:43:55,966 INFO mapred.LocalJobRunner: reduce > reduce
2020-12-10 23:43:55,966 INFO mapred.Task: Task 'attempt_local2127575678_0001_r_000000_0' done.
2020-12-10 23:43:55,967 INFO mapred.Task: Final Counters for attempt_local2127575678_0001_r_000000_0: Counters: 24
        File System Counters
                FILE: Number of bytes read=3657
                FILE: Number of bytes written=524710
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
        Map-Reduce Framework
                Combine input records=0
                Combine output records=0
                Reduce input groups=5
                Reduce shuffle bytes=61
                Reduce input records=5
                Reduce output records=5
                Spilled Records=5
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=0
                Total committed heap usage (bytes)=1012924416
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Output Format Counters
                Bytes Written=87
2020-12-10 23:43:55,967 INFO mapred.LocalJobRunner: Finishing task: attempt_local2127575678_0001_r_000000_0
2020-12-10 23:43:55,967 INFO mapred.LocalJobRunner: reduce task executor complete.
2020-12-10 23:43:56,825 INFO mapreduce.Job: Job job_local2127575678_0001 running in uber mode : false
2020-12-10 23:43:56,826 INFO mapreduce.Job:  map 100% reduce 100%
2020-12-10 23:43:56,827 INFO mapreduce.Job: Job job_local2127575678_0001 completed successfully
2020-12-10 23:43:56,833 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=7160
                FILE: Number of bytes written=1049272
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
```

```
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Output Format Counters
                Bytes Written=87
2020-12-10 23:43:55,967 INFO mapred.LocalJobRunner: Finishing task: attempt_local2127575678_0001_r_000000_0
2020-12-10 23:43:55,967 INFO mapred.LocalJobRunner: reduce task executor complete.
2020-12-10 23:43:56,825 INFO mapreduce.Job: Job job_local2127575678_0001 running in uber mode : false
2020-12-10 23:43:56,826 INFO mapreduce.Job:  map 100% reduce 100%
2020-12-10 23:43:56,827 INFO mapreduce.Job: Job job_local2127575678_0001 completed successfully
2020-12-10 23:43:56,833 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=7160
                FILE: Number of bytes written=1049272
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
        Map-Reduce Framework
                Map input records=5
                Map output records=65
                Map output bytes=585
                Map output materialized bytes=61
                Input split bytes=89
                Combine input records=65
                Combine output records=5
                Reduce input groups=5
                Reduce shuffle bytes=61
                Reduce input records=5
                Reduce output records=5
                Spilled Records=10
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=5
                Total committed heap usage (bytes)=2025848832
        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=238
        File Output Format Counters
                Bytes Written=87
hdoop@23052bbaa7cd:~$
```

STep 8

```
                Bytes Written=87
hdoop@23052bbaa7cd:~$ $HADOOP_HOME/bin/hadoop fs -ls output_dir/
Found 2 items
-rw-r--r--   1 hdoop hdoop          0 2020-12-10 23:43 output_dir/_SUCCESS
-rw-r--r--   1 hdoop hdoop         75 2020-12-10 23:43 output_dir/part-00000
hdoop@23052bbaa7cd:~$
```

Step 9

```
hdoop@23052bbaa7cd:~$ $HADOOP_HOME/bin/hadoop fs -cat output_dir/part-00000
1979    24.615385
1980    29.153847
1981    33.615383
1984    39.615383
1985    36.923077
hdoop@23052bbaa7cd:~$
```

Step 10

```
hdoop@23052bbaa7cd:~$ $HADOOP_HOME/bin/hadoop fs -cat output_dir/part-00000 > result.txt
hdoop@23052bbaa7cd:~$ ls
ProcessUnits.java  hadoop-3.2.1  hadoop-core-1.2.1.jar  hadoop-core-3.2.1.tar.gz  input_dir  output_dir  result.txt  sample.txt  units  units.jar
hdoop@23052bbaa7cd:~$ cat result.txt
1979    24.615385
1980    29.153847
1981    33.615383
1984    39.615383
1985    36.923077
hdoop@23052bbaa7cd:~$
```