

STEP 3

```
PS D:\desktop\BPOC\lab4> javac -classpath hadoop-core-1.2.1.jar -d units ProcessUnits.java
```

```
PS D:\desktop\BPOC\lab4> jar -cvf units.jar -C units/
```

Error parsing file arguments

Try `jar --help' for more information.

```
PS D:\desktop\BPOC\lab4> jar -cvf units.jar -C units/ .
```

added manifest

adding: hadoop/(in = 0) (out= 0)(stored 0%)

adding: hadoop/ProcessUnits\$E_EMapper.class(in = 1914) (out= 785)(deflated 58%)

adding: hadoop/ProcessUnits\$E_EReduce.class(in = 1690) (out= 697)(deflated 58%)

adding: hadoop/ProcessUnits.class(in = 1593) (out= 783)(deflated 50%)

Step 4

```
PS D:\desktop\BPOC\lab4> docker exec -it namenode /bin/bash
```

```
root@9564840172e7:/# hdfs dfs -mkdir input_dir
```

mkdir: `input_dir': File exists

Step 5

```
root@9564840172e7:/# hdfs dfs -put /tmp/sample.txt input_dir
```

Step 6

```
root@9564840172e7:/# hdfs dfs -ls input_dir/
```

Found 1 items

```
-rw-r--r--  3 root supergroup    222 2022-12-18 21:31 input_dir/sample.txt
```

Step 7

```
root@9564840172e7:/# hadoop jar units1.jar hadoop.ProcessUnits input_dir output_dir
```

```
22/12/18 21:54:02 INFO Configuration.deprecation: session.id is deprecated. Instead, use
dfs.metrics.session-id
```

```
22/12/18 21:54:02 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
```

```
22/12/18 21:54:02 INFO jvm.JvmMetrics: Cannot initialize JVM Metrics with processName=JobTracker,
sessionId= - already initialized
```

22/12/18 21:54:02 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

22/12/18 21:54:02 INFO mapred.FileInputFormat: Total input paths to process : 1

22/12/18 21:54:02 INFO mapreduce.JobSubmitter: number of splits:1

22/12/18 21:54:02 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local692046146_0001

22/12/18 21:54:02 INFO mapreduce.Job: The url to track the job: http://localhost:8080/

22/12/18 21:54:02 INFO mapred.LocalJobRunner: OutputCommitter set in config null

22/12/18 21:54:02 INFO mapreduce.Job: Running job: job_local692046146_0001

22/12/18 21:54:02 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter

22/12/18 21:54:02 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

22/12/18 21:54:02 INFO mapred.LocalJobRunner: Waiting for map tasks

22/12/18 21:54:02 INFO mapred.LocalJobRunner: Starting task:
attempt_local692046146_0001_m_000000_0

22/12/18 21:54:02 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

22/12/18 21:54:02 INFO mapred.Task: Using ResourceCalculatorProcessTree : []

22/12/18 21:54:02 INFO mapred.MapTask: Processing split:
hdfs://namenode:8020/user/root/input_dir/sample.txt:0+222

22/12/18 21:54:02 INFO mapred.MapTask: numReduceTasks: 1

22/12/18 21:54:02 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)

22/12/18 21:54:02 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100

22/12/18 21:54:02 INFO mapred.MapTask: soft limit at 83886080

22/12/18 21:54:02 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600

22/12/18 21:54:02 INFO mapred.MapTask: kvstart = 26214396; length = 6553600

22/12/18 21:54:02 INFO mapred.MapTask: Map output collector class =
org.apache.hadoop.mapred.MapTask\$MapOutputBuffer

22/12/18 21:54:02 INFO mapred.LocalJobRunner:

22/12/18 21:54:02 INFO mapred.MapTask: Starting flush of map output

22/12/18 21:54:02 INFO mapred.MapTask: Spilling map output

22/12/18 21:54:02 INFO mapred.MapTask: bufstart = 0; bufend = 585; bufvoid = 104857600

22/12/18 21:54:02 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214140(104856560);
length = 257/6553600

22/12/18 21:54:02 INFO mapred.MapTask: Finished spill 0

22/12/18 21:54:02 INFO mapred.Task: Task:attempt_local692046146_0001_m_000000_0 is done. And is in the process of committing

22/12/18 21:54:02 INFO mapred.LocalJobRunner:
hdfs://namenode:8020/user/root/input_dir/sample.txt:0+222

22/12/18 21:54:02 INFO mapred.Task: Task 'attempt_local692046146_0001_m_000000_0' done.

22/12/18 21:54:02 INFO mapred.LocalJobRunner: Finishing task:
attempt_local692046146_0001_m_000000_0

22/12/18 21:54:02 INFO mapred.LocalJobRunner: map task executor complete.

22/12/18 21:54:02 INFO mapred.LocalJobRunner: Waiting for reduce tasks

22/12/18 21:54:02 INFO mapred.LocalJobRunner: Starting task:
attempt_local692046146_0001_r_000000_0

22/12/18 21:54:03 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

22/12/18 21:54:03 INFO mapred.Task: Using ResourceCalculatorProcessTree : []

22/12/18 21:54:03 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin:
org.apache.hadoop.mapreduce.task.reduce.Shuffle@7b76943a

22/12/18 21:54:03 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=334338464,
maxSingleShuffleLimit=83584616, mergeThreshold=220663392, ioSortFactor=10,
memToMemMergeOutputsThreshold=10

22/12/18 21:54:03 INFO reduce.EventFetcher: attempt_local692046146_0001_r_000000_0 Thread started:
EventFetcher for fetching Map Completion Events

22/12/18 21:54:03 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map
attempt_local692046146_0001_m_000000_0 decomp: 57 len: 61 to MEMORY

22/12/18 21:54:03 INFO reduce.InMemoryMapOutput: Read 57 bytes from map-output for
attempt_local692046146_0001_m_000000_0

22/12/18 21:54:03 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 57,
inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->57

22/12/18 21:54:03 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning

22/12/18 21:54:03 INFO mapred.LocalJobRunner: 1 / 1 copied.

22/12/18 21:54:03 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and
0 on-disk map-outputs

22/12/18 21:54:03 INFO mapred.Merger: Merging 1 sorted segments

22/12/18 21:54:03 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 50
bytes

22/12/18 21:54:03 INFO reduce.MergeManagerImpl: Merged 1 segments, 57 bytes to disk to satisfy reduce memory limit

22/12/18 21:54:03 INFO reduce.MergeManagerImpl: Merging 1 files, 61 bytes from disk

22/12/18 21:54:03 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce

22/12/18 21:54:03 INFO mapred.Merger: Merging 1 sorted segments

22/12/18 21:54:03 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 50 bytes

22/12/18 21:54:03 INFO mapred.LocalJobRunner: 1 / 1 copied.

22/12/18 21:54:03 INFO mapred.Task: Task:attempt_local692046146_0001_r_000000_0 is done. And is in the process of committing

22/12/18 21:54:03 INFO mapred.LocalJobRunner: 1 / 1 copied.

22/12/18 21:54:03 INFO mapred.Task: Task attempt_local692046146_0001_r_000000_0 is allowed to commit now

22/12/18 21:54:03 INFO output.FileOutputCommitter: Saved output of task 'attempt_local692046146_0001_r_000000_0' to
hdfs://namenode:8020/user/root/output_dir/_temporary/0/task_local692046146_0001_r_000000

22/12/18 21:54:03 INFO mapred.LocalJobRunner: reduce > reduce

22/12/18 21:54:03 INFO mapred.Task: Task 'attempt_local692046146_0001_r_000000_0' done.

22/12/18 21:54:03 INFO mapred.LocalJobRunner: Finishing task:
attempt_local692046146_0001_r_000000_0

22/12/18 21:54:03 INFO mapred.LocalJobRunner: reduce task executor complete.

22/12/18 21:54:03 INFO mapreduce.Job: Job job_local692046146_0001 running in uber mode : false

22/12/18 21:54:03 INFO mapreduce.Job: map 100% reduce 100%

22/12/18 21:54:03 INFO mapreduce.Job: Job job_local692046146_0001 completed successfully

22/12/18 21:54:03 INFO mapreduce.Job: Counters: 35

File System Counters

FILE: Number of bytes read=6706

FILE: Number of bytes written=586079

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=444

HDFS: Number of bytes written=75

HDFS: Number of read operations=13

HDFS: Number of large read operations=0

HDFS: Number of write operations=4

Map-Reduce Framework

Map input records=5

Map output records=65

Map output bytes=585

Map output materialized bytes=61

Input split bytes=103

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)=7

Total committed heap usage (bytes)=652738560

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=222

File Output Format Counters

Bytes Written=75

Step 8:

```
root@9564840172e7:/# hdfs dfs -ls output_dir/
```

Found 2 items

```
-rw-r--r--  3 root supergroup      0 2022-12-18 21:54 output_dir/_SUCCESS
-rw-r--r--  3 root supergroup    75 2022-12-18 21:54 output_dir/part-00000
```

Step 9:

```
root@9564840172e7:/# hdfs dfs -cat output_dir/part-00000
```

1979 24.615385

1980 29.153847

1981 33.615383

1984 39.615383

1985 36.923077

Step 10:

```
root@9564840172e7:/# hdfs dfs -get output_dir /home/hadoop
```

```
root@9564840172e7:/# cd home/hadoop
```

```
root@9564840172e7:/home/hadoop# ls
```

_SUCCESS part-00000

PS D:\desktop\BPOC\lab4> docker cp namenode:/home/hadoop results