EXPERIMENT-6

Implement Wordcount program on Hadoop framework

Code with Screenshots:

```
resourcemanager is running as process 8687.  Stop it first and ensure /tmp/hadoop-hadoop-resourcemanager.pid file is empty before retry.
Starting nodemanagers
localhost: nodemanager is running as process 8852.  Stop it first and ensure /tmp/hadoop-hadoop-nodemanager.pid file is empty before retry.
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ jps
8400 SecondaryNameNode
15538 Jps
8852 NodeManager
7941 NameNode
8118 DataNode
10892 org.eclipse.equinox.launcher_1.6.1000.v20250227-1734.jar
8687 ResourceManager
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ cd ~/Desktop  # or navigate to any other directory where you want to create the file
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ nano sample.txt
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ cd ~
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ jps
8400 SecondaryNameNode
8852 NodeManager
7941 NameNode
8118 DataNode
15593 Jps
10892 org.eclipse.equinox.launcher_1.6.1000.v20250227-1734.jar
8687 ResourceManager
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -mkdir /rgs
mkdir: `/rgs': File exists
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -copyFromLocal ~/Desktop/sample.txt /rgs/test.txt
copyFromLocal: `/rgs/test.txt': File exists
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -ls /rgs
Found 1 items
-rw-r--r--   1 hadoop supergroup         90 2025-04-15 14:19 /rgs/test.txt
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -cat /rgs/test.txt
hi how are you
how is your job
how is your family
how is your brother
how is your sister

hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop jar /home/hadoop/Desktop/WC.jar MyDriver /rgs /output
2025-04-29 15:29:14,897 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-04-29 15:29:14,935 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-04-29 15:29:14,935 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-04-29 15:29:14,941 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2025-04-29 15:29:14,996 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-04-29 15:29:15,059 INFO mapred.FileInputFormat: Total input files to process : 1
2025-04-29 15:29:15,064 INFO mapreduce.JobSubmitter: number of splits:1
2025-04-29 15:29:15,126 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local537170479_0001
2025-04-29 15:29:15,126 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-04-29 15:29:15,182 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-04-29 15:29:15,183 INFO mapreduce.Job: Running job: job_local537170479_0001
2025-04-29 15:29:15,183 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-04-29 15:29:15,184 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter
2025-04-29 15:29:15,186 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-04-29 15:29:15,187 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-04-29 15:29:15,221 INFO mapred.LocalJobRunner: Waiting for map tasks
2025-04-29 15:29:15,222 INFO mapred.LocalJobRunner: Starting task: attempt_local537170479_0001_m_000000_0
2025-04-29 15:29:15,233 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-04-29 15:29:15,233 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
```

```
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop jar /home/hadoop/Desktop/WC.jar MyDriver /rgs /output
2025-04-29 15:29:14,897 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-04-29 15:29:14,935 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-04-29 15:29:14,935 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-04-29 15:29:14,941 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2025-04-29 15:29:14,996 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-04-29 15:29:15,059 INFO mapred.FileInputFormat: Total input files to process : 1
2025-04-29 15:29:15,064 INFO mapreduce.JobSubmitter: number of splits:1
2025-04-29 15:29:15,126 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local537170479_0001
2025-04-29 15:29:15,126 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-04-29 15:29:15,182 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-04-29 15:29:15,183 INFO mapreduce.Job: Running job: job_local537170479_0001
2025-04-29 15:29:15,183 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-04-29 15:29:15,184 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter
2025-04-29 15:29:15,186 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-04-29 15:29:15,187 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-04-29 15:29:15,221 INFO mapred.LocalJobRunner: Waiting for map tasks
2025-04-29 15:29:15,222 INFO mapred.LocalJobRunner: Starting task: attempt_local537170479_0001_m_000000_0
2025-04-29 15:29:15,233 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-04-29 15:29:15,233 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-04-29 15:29:15,239 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
2025-04-29 15:29:15,242 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/rgs/test.txt:0+90
2025-04-29 15:29:15,252 INFO mapred.MapTask: numReduceTasks: 1
2025-04-29 15:29:15,281 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2025-04-29 15:29:15,281 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2025-04-29 15:29:15,281 INFO mapred.MapTask: soft limit at 83886080
2025-04-29 15:29:15,281 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2025-04-29 15:29:15,281 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2025-04-29 15:29:15,283 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2025-04-29 15:29:15,316 INFO mapred.LocalJobRunner:
2025-04-29 15:29:15,316 INFO mapred.MapTask: Starting flush of map output
2025-04-29 15:29:15,316 INFO mapred.MapTask: Spilling map output
2025-04-29 15:29:15,316 INFO mapred.MapTask: bufstart = 0; bufend = 169; bufvoid = 104857600
2025-04-29 15:29:15,316 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214320(104857280); length = 77/6553600
2025-04-29 15:29:15,319 INFO mapred.MapTask: Finished spill 0
2025-04-29 15:29:15,325 INFO mapred.Task: Task:attempt_local537170479_0001_m_000000_0 is done. And is in the process of committing
2025-04-29 15:29:15,327 INFO mapred.LocalJobRunner: hdfs://localhost:9000/rgs/test.txt:0+90
2025-04-29 15:29:15,327 INFO mapred.Task: Task 'attempt_local537170479_0001_m_000000_0' done.
2025-04-29 15:29:15,329 INFO mapred.Task: Final Counters for attempt_local537170479_0001_m_000000_0: Counters: 23
        File System Counters
                FILE: Number of bytes read=4156
                FILE: Number of bytes written=642471
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=90
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=5
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=1
                HDFS: Number of bytes read erasure-coded=0
        Map-Reduce Framework
                Map input records=6
                Map output records=20
                Map output bytes=169
                Map output materialized bytes=215
```

hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~

```
2025-04-29 15:29:15,316 INFO mapred.MapTask: bufstart = 0; bufend = 169; bufvoid = 104857600
2025-04-29 15:29:15,316 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214320(104857280); length = 77/6553600
2025-04-29 15:29:15,319 INFO mapred.MapTask: Finished spill 0
2025-04-29 15:29:15,325 INFO mapred.Task: Task:attempt_local537170479_0001_m_000000_0 is done. And is in the process of committing
2025-04-29 15:29:15,327 INFO mapred.LocalJobRunner: hdfs://localhost:9000/rgs/test.txt:0+90
2025-04-29 15:29:15,327 INFO mapred.Task: Task 'attempt_local537170479_0001_m_000000_0' done.
2025-04-29 15:29:15,327 INFO mapred.Task: Final Counters for attempt_local537170479_0001_m_000000_0: Counters: 23
        File System Counters
                FILE: Number of bytes read=4156
                FILE: Number of bytes written=642471
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=90
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=5
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=1
                HDFS: Number of bytes read erasure-coded=0
        Map-Reduce Framework
                Map input records=6
                Map output records=20
                Map output bytes=169
                Map output materialized bytes=215
                Input split bytes=86
                Combine input records=0
                Spilled Records=20
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=0
                Total committed heap usage (bytes)=633339904
        File Input Format Counters
                Bytes Read=90
2025-04-29 15:29:15,329 INFO mapred.LocalJobRunner: Finishing task: attempt_local537170479_0001_m_000000_0
2025-04-29 15:29:15,330 INFO mapred.LocalJobRunner: map task executor complete.
2025-04-29 15:29:15,331 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2025-04-29 15:29:15,331 INFO mapred.LocalJobRunner: Starting task: attempt_local537170479_0001_r_000000_0
2025-04-29 15:29:15,335 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-04-29 15:29:15,335 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-04-29 15:29:15,335 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
2025-04-29 15:29:15,336 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@27a9203f
2025-04-29 15:29:15,337 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2025-04-29 15:29:15,344 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=5827985408, maxSingleShuffleLimit=1456996352, mergeThreshold=3846470400, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2025-04-29 15:29:15,345 INFO reduce.EventFetcher: attempt_local537170479_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2025-04-29 15:29:15,356 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local537170479_0001_m_000000_0 decomp: 211 len: 215 to MEMORY
2025-04-29 15:29:15,357 INFO reduce.InMemoryMapOutput: Read 211 bytes from map-output for attempt_local537170479_0001_m_000000_0
2025-04-29 15:29:15,358 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 211, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->211
2025-04-29 15:29:15,359 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
2025-04-29 15:29:15,359 INFO mapred.LocalJobRunner: 1 / 1 copied.
2025-04-29 15:29:15,359 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
2025-04-29 15:29:15,362 INFO mapred.Merger: Merging 1 sorted segments
2025-04-29 15:29:15,362 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 205 bytes
2025-04-29 15:29:15,362 INFO reduce.MergeManagerImpl: Merged 1 segments, 211 bytes to disk to satisfy reduce memory limit
2025-04-29 15:29:15,362 INFO reduce.MergeManagerImpl: Merging 1 files, 215 bytes from disk
2025-04-29 15:29:15,363 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2025-04-29 15:29:15,363 INFO mapred.Merger: Merging 1 sorted segments
```

hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~

```
                Bytes Read=90
2025-04-29 15:29:15,329 INFO mapred.LocalJobRunner: Finishing task: attempt_local537170479_0001_m_000000_0
2025-04-29 15:29:15,330 INFO mapred.LocalJobRunner: map task executor complete.
2025-04-29 15:29:15,331 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2025-04-29 15:29:15,331 INFO mapred.LocalJobRunner: Starting task: attempt_local537170479_0001_r_000000_0
2025-04-29 15:29:15,335 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-04-29 15:29:15,335 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-04-29 15:29:15,335 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
2025-04-29 15:29:15,336 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@27a9203f
2025-04-29 15:29:15,337 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2025-04-29 15:29:15,344 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=5827985408, maxSingleShuffleLimit=1456996352, mergeThreshold=3846470400, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2025-04-29 15:29:15,345 INFO reduce.EventFetcher: attempt_local537170479_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2025-04-29 15:29:15,356 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local537170479_0001_m_000000_0 decomp: 211 len: 215 to MEMORY
2025-04-29 15:29:15,357 INFO reduce.InMemoryMapOutput: Read 211 bytes from map-output for attempt_local537170479_0001_m_000000_0
2025-04-29 15:29:15,358 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 211, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->211
2025-04-29 15:29:15,359 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
2025-04-29 15:29:15,359 INFO mapred.LocalJobRunner: 1 / 1 copied.
2025-04-29 15:29:15,359 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
2025-04-29 15:29:15,362 INFO mapred.Merger: Merging 1 sorted segments
2025-04-29 15:29:15,362 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 205 bytes
2025-04-29 15:29:15,362 INFO reduce.MergeManagerImpl: Merged 1 segments, 211 bytes to disk to satisfy reduce memory limit
2025-04-29 15:29:15,362 INFO reduce.MergeManagerImpl: Merging 1 files, 215 bytes from disk
2025-04-29 15:29:15,363 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2025-04-29 15:29:15,363 INFO mapred.Merger: Merging 1 sorted segments
2025-04-29 15:29:15,363 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 205 bytes
2025-04-29 15:29:15,363 INFO mapred.LocalJobRunner: 1 / 1 copied.
2025-04-29 15:29:15,415 INFO mapred.Task: Task:attempt_local537170479_0001_r_000000_0 is done. And is in the process of committing
2025-04-29 15:29:15,416 INFO mapred.LocalJobRunner: 1 / 1 copied.
2025-04-29 15:29:15,416 INFO mapred.Task: Task attempt_local537170479_0001_r_000000_0 is allowed to commit now
2025-04-29 15:29:15,423 INFO output.FileOutputCommitter: Saved output of task 'attempt_local537170479_0001_r_000000_0' to hdfs://localhost:9000/output
2025-04-29 15:29:15,424 INFO mapred.LocalJobRunner: reduce > reduce
2025-04-29 15:29:15,424 INFO mapred.Task: Task 'attempt_local537170479_0001_r_000000_0' done.
2025-04-29 15:29:15,424 INFO mapred.Task: Final Counters for attempt_local537170479_0001_r_000000_0: Counters: 30
        File System Counters
                FILE: Number of bytes read=4618
                FILE: Number of bytes written=642686
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=90
                HDFS: Number of bytes written=69
                HDFS: Number of read operations=10
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=3
                HDFS: Number of bytes read erasure-coded=0
        Map-Reduce Framework
                Combine input records=0
                Combine output records=0
                Reduce input groups=10
                Reduce shuffle bytes=215
                Reduce input records=20
                Reduce output records=10
                Spilled Records=20
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
```

hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~

```
        File System Counters
                FILE: Number of bytes read=4618
                FILE: Number of bytes written=642686
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=90
                HDFS: Number of bytes written=69
                HDFS: Number of read operations=10
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=3
                HDFS: Number of bytes read erasure-coded=0
        Map-Reduce Framework
                Combine input records=0
                Combine output records=0
                Reduce input groups=10
                Reduce shuffle bytes=215
                Reduce input records=20
                Reduce output records=10
                Spilled Records=20
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=0
                Total committed heap usage (bytes)=633339904
        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=69
2025-04-29 15:29:15,424 INFO mapred.LocalJobRunner: Finishing task: attempt_local537170479_0001_r_000000_0
2025-04-29 15:29:15,424 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-04-29 15:29:16,185 INFO mapreduce.Job: Job job_local537170479_0001 running in uber mode : false
2025-04-29 15:29:16,186 INFO mapreduce.Job:  map 100% reduce 100%
2025-04-29 15:29:16,186 INFO mapreduce.Job: Job job_local537170479_0001 completed successfully
2025-04-29 15:29:16,190 INFO mapreduce.Job: Counters: 36
        File System Counters
                FILE: Number of bytes read=8774
                FILE: Number of bytes written=1285157
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=180
                HDFS: Number of bytes written=69
                HDFS: Number of read operations=15
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=4
                HDFS: Number of bytes read erasure-coded=0
        Map-Reduce Framework
                Map input records=6
                Map output records=20
                Map output bytes=169
```

```
                Map output materialized bytes=215
                Input split bytes=86
                Combine input records=0
                Combine output records=0
                Reduce input groups=10
                Reduce shuffle bytes=215
                Reduce input records=20
                Reduce output records=10
                Spilled Records=40
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=0
                Total committed heap usage (bytes)=1266679808
        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=90
        File Output Format Counters
                Bytes Written=69
0
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -ls /output
Found 2 items
-rw-r--r--   1 hadoop supergroup          0 2025-04-29 15:29 /output/_SUCCESS
-rw-r--r--   1 hadoop supergroup         69 2025-04-29 15:29 /output/part-00000
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -cat /output/part-00000
are     1
brother 1
family  1
hi      1
```

hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~

```
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=180
                HDFS: Number of bytes written=69
                HDFS: Number of read operations=15
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=4
                HDFS: Number of bytes read erasure-coded=0
        Map-Reduce Framework
                Map input records=6
                Map output records=20
                Map output bytes=169
                Map output materialized bytes=215
                Input split bytes=86
                Combine input records=0
                Combine output records=0
                Reduce input groups=10
                Reduce shuffle bytes=215
                Reduce input records=20
                Reduce output records=10
                Spilled Records=40
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=0
                Total committed heap usage (bytes)=1266679808
        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=90
        File Output Format Counters
                Bytes Written=69
0
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -ls /output
Found 2 items
-rw-r--r--   1 hadoop supergroup          0 2025-04-29 15:29 /output/_SUCCESS
-rw-r--r--   1 hadoop supergroup         69 2025-04-29 15:29 /output/part-00000
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -cat /output/part-00000
are     1
brother 1
family  1
hi      1
how     5
is      4
job     1
sister  1
you     4
your    4
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$
```

eclipse-workspace - WordCount/src/MyDriver.java - Eclipse IDE

File Edit Source Refactor Navigate Search Project Run Window Help

Package Explorer
- WordCount
  - JRE System Library [JavaSE-1.8]
  - src
    - (default package)
      - MyDriver.java
      - MyMapper.java
      - MyReducer.java
  - Referenced Libraries
  - .settings
  - .classpath
  - .project

MyMapper.java   MyReducer.java   MyDriver.java

```java
1  import java.io.IOException;
2  import org.apache.hadoop.conf.Configured;
3  import org.apache.hadoop.fs.Path;
4  import org.apache.hadoop.io.IntWritable;
5  import org.apache.hadoop.io.Text;
6  import org.apache.hadoop.mapred.FileInputFormat;
7  import org.apache.hadoop.mapred.FileOutputFormat;
8  import org.apache.hadoop.mapred.JobClient;
9  import org.apache.hadoop.mapred.JobConf;
10 import org.apache.hadoop.util.Tool;
11 import org.apache.hadoop.util.ToolRunner;
12
13 public class MyDriver extends Configured implements Tool {
14     public int run(String[] args) throws IOException {
15         if (args.length < 2) {
16             System.out.println("Please provide input and output paths");
17             return -1;
18         }
19         JobConf conf = new JobConf(MyDriver.class);
20         FileInputFormat.setInputPaths(conf, new Path(args[0]));
21         FileOutputFormat.setOutputPath(conf, new Path(args[1]));
22         conf.setMapperClass(MyMapper.class);
23         conf.setReducerClass(MyReducer.class);
24         conf.setMapOutputKeyClass(Text.class);
25         conf.setMapOutputValueClass(IntWritable.class);
26         conf.setOutputKeyClass(Text.class);
27         conf.setOutputValueClass(IntWritable.class);
28         JobClient.runJob(conf);
29         return 0;
30     }
31
32     public static void main(String[] args) throws Exception {
33         int exitCode = ToolRunner.run(new MyDriver(), args);
34         System.out.println(exitCode);
35     }
36 }
37
```

Task List
Find   All Activate...

Outline
- MyDriver

Problems   Javadoc   Declaration   Install Java 24 Support

Java 24 has been released.

Would you like to install Eclipse Java Development Tools support for Java 24 into your IDE?

[Install Java 24 Support] [Remind Me Later] [Don't Ask Again]

Would you like to download a Java 24 JDK?

[Download Temurin JDK 24] [Download OpenJDK 24]

WordCount

eclipse-workspace - WordCount/src/MyMapper.java - Eclipse IDE

File Edit Source Refactor Navigate Search Project Run Window Help

Package Explorer
- WordCount
  - JRE System Library [JavaSE-1.8]
  - src
    - (default package)
      - MyDriver.java
      - MyMapper.java
      - MyReducer.java
  - Referenced Libraries
  - .settings
  - .classpath
  - .project

MyMapper.java MyReducer.java MyDriver.java

```java
1 import java.io.IOException;
2 import org.apache.hadoop.io.IntWritable;
3 import org.apache.hadoop.io.LongWritable;
4 import org.apache.hadoop.io.Text;
5 import org.apache.hadoop.mapred.MapReduceBase;
6 import org.apache.hadoop.mapred.Mapper;
7 import org.apache.hadoop.mapred.OutputCollector;
8 import org.apache.hadoop.mapred.Reporter;
9
10 public class MyMapper extends MapReduceBase implements Mapper<LongWritable, Text, Text, IntWritable> {
11     public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output, Reporter rep) throws IOExcepti
12         String line = value.toString();
13         for (String word : line.split(" ")) {
14             if (word.length() > 0) {
15                 output.collect(new Text(word), new IntWritable(1));
16             }
17         }
18     }
19 }
20
21
```

Problems Javadoc Declaration Install Java 24 Support

Java 24 has been released.

Would you like to install Eclipse Java Development Tools support for Java 24 into your IDE?

Install Java 24 Support   Remind Me Later   Don't Ask Again

Would you like to download a Java 24 JDK?

Download Temurin JDK 24   Download OpenJDK 24

Outline
- MyMapper

Writable   Smart Insert   21 : 1 : 779

---

eclipse-workspace - WordCount/src/MyReducer.java - Eclipse IDE

File Edit Source Refactor Navigate Search Project Run Window Help

Package Explorer
- WordCount
  - JRE System Library [JavaSE-1.8]
  - src
    - (default package)
      - MyDriver.java
      - MyMapper.java
      - MyReducer.java
  - Referenced Libraries
  - .settings
  - .classpath
  - .project

MyMapper.java MyReducer.java MyDriver.java

```java
1 import java.io.IOException;
2 import java.util.Iterator;
3 import org.apache.hadoop.io.IntWritable;
4 import org.apache.hadoop.io.Text;
5 import org.apache.hadoop.mapred.MapReduceBase;
6 import org.apache.hadoop.mapred.OutputCollector;
7 import org.apache.hadoop.mapred.Reducer;
8 import org.apache.hadoop.mapred.Reporter;
9
10 public class MyReducer extends MapReduceBase implements Reducer<Text, IntWritable, Text, IntWritable> {
11     public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text, IntWritable> output, Reporter rep) th
12         int count = 0;
13         while (values.hasNext()) {
14             count += values.next().get();
15         }
16         output.collect(key, new IntWritable(count));
17     }
18 }
19
```

Problems Javadoc Declaration Install Java 24 Support

Java 24 has been released.

Would you like to install Eclipse Java Development Tools support for Java 24 into your IDE?

Install Java 24 Support   Remind Me Later   Don't Ask Again

Would you like to download a Java 24 JDK?

Download Temurin JDK 24   Download OpenJDK 24

Outline
- MyReducer

Writable   Smart Insert   19 : 1 : 726

eclipse-workspace - WordCount/src/MyDriver.java - Eclipse IDE

File Edit Source Refactor Navigate Search Project Run Window Help

Package Explorer ×   MyMapper.java   MyReducer.java   MyDriver.java ×

```
 1  import java.io.IOException;
 2  import org.apache.hadoop.conf.Configured;
 3  import org.apache.hadoop.fs.Path;
 4  import org.apache.hadoop.io.IntWritable;
 5  import org.apache.hadoop.io.Text;
 6  import org.apache.hadoop.mapred.FileInputFormat;
 7  import org.apache.hadoop.mapred.FileOutputFormat;
 8  import org.apache.hadoop.mapred.JobClient;
 9  import org.apache.hadoop.mapred.JobConf;
10  import org.apache.hadoop.util.Tool;
11  import org.apache.hadoop.util.ToolRunner;
12
13  public class MyDriver extends Configured implements Tool {
14      public int run(String[] args) throws IOException {
15          if (args.length < 2) {
16              System.out.println("Please provide input and output paths");
17              return -1;
18          }
19          JobConf conf = new JobConf(MyDriver.class);
20          FileInputFormat.setInputPaths(conf, new Path(args[0]));
21          FileOutputFormat.setOutputPath(conf, new Path(args[1]));
22          conf.setMapperClass(MyMapper.class);
23          conf.setReducerClass(MyReducer.class);
24          conf.setMapOutputKeyClass(Text.class);
25          conf.setMapOutputValueClass(IntWritable.class);
26          conf.setOutputKeyClass(Text.class);
27          conf.setOutputValueClass(IntWritable.class);
28          JobClient.runJob(conf);
29          return 0;
30      }
31
32      public static void main(String[] args) throws Exception {
33          int exitCode = ToolRunner.run(new MyDriver(), args);
34          System.out.println(exitCode);
35      }
36  }
37
```

Problems ● Javadoc ■ Declaration ① Install Java 24 Support ×

Java 24 has been released.

Would you like to install Eclipse Java Development Tools support for Java 24 into your IDE?

Install Java 24 Support   Remind Me Later   Don't Ask Again

Would you like to download a Java 24 JDK?

Download Temurin JDK 24   Download OpenJDK 24

Writable   Smart Insert   37 : 1 : 1391

---

eclipse-workspace - WordCount/src/MyDriver.java - Eclipse IDE

File Edit Source Refactor Navigate Search Project Run Window Help

Package Explorer ×   MyMapper.java   MyReducer.java   MyDriver.java ×

JAR Export   ✕

JAR File Specification

Define which resources should be exported into the JAR.

Select the resources to export:

- ☑ WordCount   ☑ .classpath
                       ☑ .project

☑ Export generated class files and resources
☐ Export all output folders for checked projects
☐ Export Java source files and resources
☐ Export refactorings for checked projects. Select refactorings...

Select the export destination:

JAR file: /home/hadoop/Desktop/WC.jar   ▾   Browse...

Options:
☑ Compress the contents of the JAR file
☐ Add directory entries
☐ Overwrite existing files without warning

⑦   < Back   Next >   Cancel   Finish

Java 24 has been released.

Would you like to install

Install Java 24 Support   Remind Me Later   Don't Ask Again

Would you like to download a Java 24 JDK?

Download Temurin JDK 24   Download OpenJDK 24

WordCount