

From: Madhushree S Shetty madhushree.cs22@bmsce.ac.in
Subject: Lab 7 - hadoop wordcount
Date: 6 May 2025 at 2:50 PM
To: Madhushree S Shetty madhushreesetty2003@gmail.com

MS

```
hadoop@bmscece-HP-Elite-Tower-800-G9-Desktop-PC: ~/De
hadoop@bmscece-HP-Elite-Tower-800-G9-Desktop-PC:~$ cd ./Desktop/
hadoop@bmscece-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [bmscece-HP-Elite-Tower-800-G9-Desktop-PC]
Starting resourcemanager
Starting nodemanagers
hadoop@bmscece-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ jps
12080 NodeManager
11906 ResourceManager
5142 org.eclipse.equinox.launcher_1.6.1000.v20250227-1734.jar
11304 DataNode
11611 SecondaryNameNode
12492 Jps
11134 NameNode
hadoop@bmscece-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop fs -mkdir /rgs
mkdir: '/rgs': File exists
hadoop@bmscece-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop fs -ls /
Found 8 items
drwxr-xr-x - hadoop supergroup 0 2025-04-15 14:31 /Lab6
drwxr-xr-x - hadoop supergroup 0 2024-05-14 14:59 /hadibrahim
drwxr-xr-x - hadoop supergroup 0 2024-05-14 15:15 /newdirectory
drwxr-xr-x - hadoop supergroup 0 2024-05-21 15:30 /output
drwxr-xr-x - hadoop supergroup 0 2024-05-21 15:22 /rgs
drwxr-xr-x - hadoop supergroup 0 2025-04-21 14:48 /sample
-rw-r--r-- 1 hadoop supergroup 35 2025-04-21 14:58 /sample1
drwxr-xr-x - hadoop supergroup 0 2025-04-15 14:26 /test_Lab6
hadoop@bmscece-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop fs -copyFromLocal D:/sample.txt /rgs/test.txt
copyFromLocal: '/rgs/test.txt': File exists
```

```
hadoop@bmscece-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop fs -mkdir -p /user/hadoop/input
hadoop@bmscece-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop fs -put ~/Desktop/sample.txt /user/hadoop/input/
hadoop@bmscece-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop fs -ls /user/hadoop/input
Found 1 items
-rw-r--r-- 1 hadoop supergroup 90 2025-05-06 14:45 /user/hadoop/input/sample.txt
```

```
hadoop@bmscece-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop jar /home/hadoop/eclipse-workspace/WordCount.jar WCDriver input output
2025-05-06 14:45:31,601 INFO Impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-05-06 14:45:31,636 INFO Impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-05-06 14:45:31,636 INFO Impl.MetricsSystemImpl: JobTracker metrics system started
2025-05-06 14:45:31,695 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2025-05-06 14:45:31,738 INFO Input.FileInputFormat: Total input files to process : 1
2025-05-06 14:45:31,765 INFO mapreduce.JobSubmitter: number of splits:1
2025-05-06 14:45:31,825 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1385456850_0001
2025-05-06 14:45:31,825 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-05-06 14:45:31,887 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-05-06 14:45:31,887 INFO mapreduce.Job: Running job: job_local1385456850_0001
2025-05-06 14:45:31,888 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-05-06 14:45:31,892 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-05-06 14:45:31,893 INFO output.FileOutputCommitter: File OutputCommitter Algorithm version is 2
2025-05-06 14:45:31,893 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup_temporary folders under output directory:false, ignore cleanup failures:
2025-05-06 14:45:31,893 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2025-05-06 14:45:31,925 INFO mapred.LocalJobRunner: Waiting for map tasks
2025-05-06 14:45:31,925 INFO mapred.LocalJobRunner: Starting task: attempt_local1385456850_0001_m_000000_0
2025-05-06 14:45:31,935 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-05-06 14:45:31,935 INFO output.FileOutputCommitter: File OutputCommitter Algorithm version is 2
2025-05-06 14:45:31,935 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup_temporary folders under output directory:false, ignore cleanup failures:
2025-05-06 14:45:31,943 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2025-05-06 14:45:31,945 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/user/hadoop/input/sample.txt:0+90
2025-05-06 14:45:31,978 INFO mapred.MapTask: (EQUATOR) 0 kvt 26214396(104857584)
2025-05-06 14:45:31,978 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2025-05-06 14:45:31,978 INFO mapred.MapTask: soft limit at 83886080
2025-05-06 14:45:31,979 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2025-05-06 14:45:31,979 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2025-05-06 14:45:31,980 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2025-05-06 14:45:32,049 INFO mapred.LocalJobRunner:
2025-05-06 14:45:32,050 INFO mapred.MapTask: Starting flush of map output
2025-05-06 14:45:32,050 INFO mapred.MapTask: Spilling map output
2025-05-06 14:45:32,050 INFO mapred.MapTask: bufstart = 0; bufend = 174; bufvoid = 104857600
2025-05-06 14:45:32,050 INFO mapred.MapTask: kvstart = 26214396(104857584); kwend = 26214316(104857264); length = 81/6553600
2025-05-06 14:45:32,052 INFO mapred.MapTask: Finished spill 0
2025-05-06 14:45:32,057 INFO mapred.Task: Task:attempt_local1385456850_0001_m_000000_0 is done. And is in the process of committing
2025-05-06 14:45:32,059 INFO mapred.LocalJobRunner: map
2025-05-06 14:45:32,059 INFO mapred.Task: Task 'attempt_local1385456850_0001_m_000000_0' done.
2025-05-06 14:45:32,061 INFO mapred.Task: Final Counters for attempt_local1385456850_0001_m_000000_0: Counters: 23
File System Counters
FILE: Number of bytes read=201
FILE: Number of bytes written=641533
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
```

```
hadoop@bmscece-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop fs -ls /output/
Found 2 items
```

```
-rw-r--r-- 1 hadoop supergroup 0 2024-05-21 15:30 /output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 69 2024-05-21 15:30 /output/part-00000
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop fs -cat /output/part-00000
are 1
brother 1
family 1
hi 1
how 5
is 4
job 1
sister 1
you 1
your 4
```