
BIG DATA ANALYTICS

22AIE312 – Labsheet 3

Mapreduce WordCount Java Program:

Execution Steps:

Step 1:hadoop@nisha-VirtualBox:~\$ hdfs dfs -mkdir /count

Step 2:hadoop@nisha-VirtualBox:~\$ hdfs dfs -touchz /count/outputcount

Step 3:hadoop@nisha-VirtualBox:~\$ hdfs dfs -appendToFile - /count/outputcount

one two three

three two one

four five six

ctrl+d Ctrl+d

Step 4:mkdir -p /home/hadoop/outputmap

```
hadoop@DESKTOP-07BKI3U:~$ hdfs dfs -mkdir /count
hadoop@DESKTOP-07BKI3U:~$ hdfs dfs -touchz /count/outputcount
hadoop@DESKTOP-07BKI3U:~$ hdfs dfs -appendToFile - /count/outputcount
two three
three two one
four five six
hadoop@DESKTOP-07BKI3U:~$ mkdir -p /home/hadoop/outputmap
```

Step 5:hadoop@nisha-VirtualBox:~\$ javac -classpath \$(hadoop classpath) -d /home/hadoop/outputmap /home/hadoop/WordCount.java

Step 6:hadoop@nisha-VirtualBox:~\$ jar -cvf wordcount.jar -C /home/hadoop/outputmap/ .

```

hadoop@DESKTOP-07BKI3U:~$ nano WordCount.java
hadoop@DESKTOP-07BKI3U:~$ javac -classpath $(hadoop classpath) -d ~/outputmap ~/WordCount.java
hadoop@DESKTOP-07BKI3U:~$ jar -cvf wordcount.jar -C /home/hadoop/outputmap/ .
added manifest
adding: WordCount$TokenizerMapper.class(in = 1862) (out= 777)(deflated 58%)
adding: WordCount$IntSumReducer.class(in = 1755) (out= 749)(deflated 57%)
adding: WordCount.class(in = 1676) (out= 933)(deflated 44%)
hadoop@DESKTOP-07BKI3U:~$ hadoop jar wordcount.jar WordCount /count /outputcount
2025-03-05 10:06:25,478 INFO client.DefaultNoHARMFaloverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2025-03-05 10:06:26,139 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement
the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-05 10:06:26,164 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/
hadoop/.staging/job_1741148130018_0001
2025-03-05 10:06:26,452 INFO input.FileInputFormat: Total input files to process : 1
2025-03-05 10:06:26,516 INFO mapreduce.JobSubmitter: number of splits:1
2025-03-05 10:06:26,741 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1741148130018_0001
2025-03-05 10:06:26,741 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-05 10:06:26,915 INFO conf.Configuration: resource-types.xml not found
2025-03-05 10:06:26,915 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-03-05 10:06:27,354 INFO impl.YarnClientImpl: Submitted application application_1741148130018_0001
2025-03-05 10:06:27,409 INFO mapreduce.Job: The url to track the job: http://DESKTOP-07BKI3U.:8088/proxy/application_174
1148130018_0001/
2025-03-05 10:06:27,410 INFO mapreduce.Job: Running job: job_1741148130018_0001
2025-03-05 10:06:34,512 INFO mapreduce.Job: Job job_1741148130018_0001 running in uber mode : false
2025-03-05 10:06:34,516 INFO mapreduce.Job: map 0% reduce 0%
2025-03-05 10:06:38,566 INFO mapreduce.Job: map 100% reduce 0%
2025-03-05 10:06:43,602 INFO mapreduce.Job: map 100% reduce 100%
2025-03-05 10:06:44,626 INFO mapreduce.Job: Job job_1741148130018_0001 completed successfully
2025-03-05 10:06:44,690 INFO mapreduce.Job: Counters: 54
File System Counters

```

Step 7:hadoop jar wordcount.jar WordCount /count /outputcount

Step 8:hadoop@nisha-VirtualBox:~\$ hadoop fs -ls /outputcount

Step 9:hadoop fs -cat /outputcount/part*

```

five      1
four      1
one       2
six       1
three     2
two       2

```

```

hadoop@DESKTOP-07BKI3U:~$ hadoop fs -ls /outputcount
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2025-03-05 10:06 /outputcount/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 40 2025-03-05 10:06 /outputcount/part-r-000000
hadoop@DESKTOP-07BKI3U:~$ hadoop fs -cat /outputcount/part*
five      1
four      1
one       2
six       1
three     2
two       2
hadoop@DESKTOP-07BKI3U:~$ |

```