MAP REDUCE WORD COUNT DOCS

TERMINAL:

hadoop@ubuntu:~$ hadoop version
Hadoop 3.3.6
Source code repository https://github.com/apache/hadoop.git -r
1be78238728da9266a4f88195058f08fd012bf9c
Compiled by ubuntu on 2023-06-18T08:22Z
Compiled on platform linux-x86_64
Compiled with protoc 3.7.1
From source with checksum 5652179ad55f76cb287d9c633bb53bbd
This command was run using
/home/hadoop/hadoop/share/hadoop/common/hadoop-common-3.3.6.jar
hadoop@ubuntu:~$ ls
Desktop   Downloads hadoop-3.3.6.tar.gz Music   Public Templates
Documents hadoop   hadoopdata   Pictures snap   Videos
hadoop@ubuntu:~$ cd Desktop/
hadoop@ubuntu:~/Desktop$ touch WordCount.java
hadoop@ubuntu:~/Desktop$ cd input_data/
hadoop@ubuntu:~/Desktop/input_data$ touch input.txt
**hadoop@ubuntu:~/Desktop/input_data$ export
HADOOP_CLASSPATH=$(hadoop classpath)**
hadoop@ubuntu:~/Desktop/input_data$ echo $HADOOP_CLASSPATH
/home/hadoop/hadoop/etc/hadoop:/home/hadoop/hadoop/share/hadoop/common/li
b/*:/home/hadoop/hadoop/share/hadoop/common/*:/home/hadoop/hadoop/share/h
adoop/hdfs:/home/hadoop/hadoop/share/hadoop/hdfs/lib/*:/home/hadoop/hadoop/s
hare/hadoop/hdfs/*:/home/hadoop/hadoop/share/hadoop/mapreduce/*:/home/hado
op/hadoop/share/hadoop/yarn:/home/hadoop/hadoop/share/hadoop/yarn/lib/*:/hom
e/hadoop/hadoop/share/hadoop/yarn/*
**hadoop@ubuntu:~/Desktop/input_data$ hadoop fs -mkdir
/WordCountTutorial**
mkdir: `/WordCountTutorial': File exists
**hadoop@ubuntu:~/Desktop/input_data$ hadoop fs -mkdir
/WordCountTutorial/Input**
mkdir: `/WordCountTutorial/Input': File exists
**hadoop@ubuntu:~/Desktop/input_data$ hadoop fs -put
'/home/hadoop/Desktop/input_data/input.txt' /WordCountTutorial/Input**
**hadoop@ubuntu:~/Desktop/input_data$ cd
/home/hadoop/Desktop/WordCountTutorial**
bash: cd: /home/hadoop/Desktop/WordCountTutorial: No such file or directory
hadoop@ubuntu:~/Desktop/input_data$ cd /home/hadoop/Desktop/

hadoop@ubuntu:~/Desktop$ javac -classpath
$(HADOOP_CLASSPATH) -d '/home/hadoop/Desktop/tutorial_classes'
'/home/hadoop/Desktop/WordCount.java'
HADOOP_CLASSPATH: command not found
javac: invalid flag: /home/hadoop/Desktop/tutorial_classes
Usage: javac <options> <source files>
use -help for a list of possible options
hadoop@ubuntu:~/Desktop$ javac -classpath
$(HADOOP_CLASSPATH)-d '/home/hadoop/Desktop/tutorial_classes'
'/home/hadoop/Desktop/WordCount.java'
HADOOP_CLASSPATH: command not found
javac: invalid flag: /home/hadoop/Desktop/tutorial_classes
Usage: javac <options> <source files>
use -help for a list of possible options
hadoop@ubuntu:~/Desktop$ ^C
hadoop@ubuntu:~/Desktop$ ^C
hadoop@ubuntu:~/Desktop$ ^C
hadoop@ubuntu:~/Desktop$ echo $HADOOP_CLASSPATH
/home/hadoop/hadoop/etc/hadoop:/home/hadoop/hadoop/share/hadoop/common/li
b/*:/home/hadoop/hadoop/share/hadoop/common/*:/home/hadoop/hadoop/share/h
adoop/hdfs:/home/hadoop/hadoop/share/hadoop/hdfs/lib/*:/home/hadoop/hadoop/s
hare/hadoop/hdfs/*:/home/hadoop/hadoop/share/hadoop/mapreduce/*:/home/hado
op/hadoop/share/hadoop/yarn:/home/hadoop/hadoop/share/hadoop/yarn/lib/*:/hom
e/hadoop/hadoop/share/hadoop/yarn/*
hadoop@ubuntu:~/Desktop$ javac -classpath
($HADOOP_CLASSPATH)-d '/home/hadoop/Desktop/tutorial_classes'
'/home/hadoop/Desktop/WordCount.java'
bash: syntax error near unexpected token `('
hadoop@ubuntu:~/Desktop$ javac -classpath $(HADOOP_CLASSPATH)
-d '/home/hadoop/Desktop/tutorial_classes'
'/home/hadoop/Desktop/WordCount.java'
HADOOP_CLASSPATH: command not found
javac: invalid flag: /home/hadoop/Desktop/tutorial_classes
Usage: javac <options> <source files>
use -help for a list of possible options
hadoop@ubuntu:~/Desktop$ javac -classpath
$HADOOP_CLASSPATH -d '/home/hadoop/Desktop/tutorial_classes'
'/home/hadoop/Desktop/WordCount.java'
hadoop@ubuntu:~/Desktop$ jar -cvf firstTutorial.jar -C
tutorial_classes/ .
added manifest
adding: WordCount$IntSumReducer.class(in = 1739) (out= 740)(deflated 57%)
adding: WordCount.class(in = 1491) (out= 816)(deflated 45%)
adding: WordCount$TokenizerMapper.class(in = 1736) (out= 755)(deflated 56%)

hadoop@ubuntu:~/Desktop$ hadoop jar '/home/hadoop/Desktop/firstTutorial.jar' WordCount /WordCountTutorial/Input /WordCountTutorial/Output

2024-09-12 19:42:20,894 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties

2024-09-12 19:42:21,142 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).

2024-09-12 19:42:21,142 INFO impl.MetricsSystemImpl: JobTracker metrics system started

2024-09-12 19:42:21,346 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

2024-09-12 19:42:21,516 INFO input.FileInputFormat: Total input files to process : 1

2024-09-12 19:42:21,551 INFO mapreduce.JobSubmitter: number of splits:1

2024-09-12 19:42:21,680 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local144329354_0001

2024-09-12 19:42:21,680 INFO mapreduce.JobSubmitter: Executing with tokens: []

2024-09-12 19:42:21,929 INFO mapreduce.Job: The url to track the job: http://localhost:8080/

2024-09-12 19:42:21,931 INFO mapreduce.Job: Running job: job_local144329354_0001

2024-09-12 19:42:21,933 INFO mapred.LocalJobRunner: OutputCommitter set in config null

2024-09-12 19:42:21,978 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory

2024-09-12 19:42:21,980 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2

2024-09-12 19:42:21,980 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false

2024-09-12 19:42:21,981 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter

2024-09-12 19:42:22,110 INFO mapred.LocalJobRunner: Waiting for map tasks

2024-09-12 19:42:22,111 INFO mapred.LocalJobRunner: Starting task: attempt_local144329354_0001_m_000000_0

2024-09-12 19:42:22,196 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory

2024-09-12 19:42:22,196 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2

2024-09-12 19:42:22,196 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false

2024-09-12 19:42:22,304 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]

2024-09-12 19:42:22,320 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/WordCountTutorial/Input/input.txt:0+121

2024-09-12 19:42:22,402 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)

2024-09-12 19:42:22,402 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100

2024-09-12 19:42:22,402 INFO mapred.MapTask: soft limit at 83886080

2024-09-12 19:42:22,402 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600

2024-09-12 19:42:22,402 INFO mapred.MapTask: kvstart = 26214396; length = 6553600

2024-09-12 19:42:22,415 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer

2024-09-12 19:42:22,592 INFO mapred.LocalJobRunner:

2024-09-12 19:42:22,597 INFO mapred.MapTask: Starting flush of map output

2024-09-12 19:42:22,598 INFO mapred.MapTask: Spilling map output

2024-09-12 19:42:22,598 INFO mapred.MapTask: bufstart = 0; bufend = 196; bufvoid = 104857600

2024-09-12 19:42:22,598 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214324(104857296); length = 73/6553600

2024-09-12 19:42:22,643 INFO mapred.MapTask: Finished spill 0

2024-09-12 19:42:22,664 INFO mapred.Task: Task:attempt_local144329354_0001_m_000000_0 is done. And is in the process of committing

2024-09-12 19:42:22,670 INFO mapred.LocalJobRunner: map

2024-09-12 19:42:22,671 INFO mapred.Task: Task 'attempt_local144329354_0001_m_000000_0' done.

2024-09-12 19:42:22,688 INFO mapred.Task: Final Counters for attempt_local144329354_0001_m_000000_0: Counters: 24
      File System Counters
      FILE: Number of bytes read=3251
      FILE: Number of bytes written=641373
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=121
      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=20

        Map output records=19
        Map output bytes=196
        Map output materialized bytes=189
        Input split bytes=120
        Combine input records=19
        Combine output records=15
        Spilled Records=15
        Failed Shuffles=0
        Merged Map outputs=0
        GC time elapsed (ms)=64
        Total committed heap usage (bytes)=254803968
        File Input Format Counters
        Bytes Read=121
2024-09-12 19:42:22,688 INFO mapred.LocalJobRunner: Finishing task: attempt_local144329354_0001_m_000000_0
2024-09-12 19:42:22,689 INFO mapred.LocalJobRunner: map task executor complete.
2024-09-12 19:42:22,715 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2024-09-12 19:42:22,720 INFO mapred.LocalJobRunner: Starting task: attempt_local144329354_0001_r_000000_0
2024-09-12 19:42:22,785 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2024-09-12 19:42:22,786 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-09-12 19:42:22,786 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-09-12 19:42:22,787 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2024-09-12 19:42:22,811 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@2ee7ccdf
2024-09-12 19:42:22,816 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2024-09-12 19:42:22,864 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=1265788544, maxSingleShuffleLimit=316447136, mergeThreshold=835420480, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2024-09-12 19:42:22,869 INFO reduce.EventFetcher: attempt_local144329354_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2024-09-12 19:42:22,930 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local144329354_0001_m_000000_0 decomp: 185 len: 189 to MEMORY

2024-09-12 19:42:22,934 INFO reduce.InMemoryMapOutput: Read 185 bytes from map-output for attempt_local144329354_0001_m_000000_0
2024-09-12 19:42:22,948 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 185, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->185
2024-09-12 19:42:22,957 INFO mapreduce.Job: Job job_local144329354_0001 running in uber mode : false
2024-09-12 19:42:22,958 INFO mapreduce.Job: map 100% reduce 0%
2024-09-12 19:42:22,978 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
2024-09-12 19:42:23,009 INFO mapred.LocalJobRunner: 1 / 1 copied.
2024-09-12 19:42:23,012 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
2024-09-12 19:42:23,027 INFO mapred.Merger: Merging 1 sorted segments
2024-09-12 19:42:23,033 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 173 bytes
2024-09-12 19:42:23,037 INFO reduce.MergeManagerImpl: Merged 1 segments, 185 bytes to disk to satisfy reduce memory limit
2024-09-12 19:42:23,038 INFO reduce.MergeManagerImpl: Merging 1 files, 189 bytes from disk
2024-09-12 19:42:23,039 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2024-09-12 19:42:23,040 INFO mapred.Merger: Merging 1 sorted segments
2024-09-12 19:42:23,042 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 173 bytes
2024-09-12 19:42:23,043 INFO mapred.LocalJobRunner: 1 / 1 copied.
2024-09-12 19:42:23,119 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords
2024-09-12 19:42:23,206 INFO mapred.Task: Task:attempt_local144329354_0001_r_000000_0 is done. And is in the process of committing
2024-09-12 19:42:23,210 INFO mapred.LocalJobRunner: 1 / 1 copied.
2024-09-12 19:42:23,212 INFO mapred.Task: Task attempt_local144329354_0001_r_000000_0 is allowed to commit now
2024-09-12 19:42:23,297 INFO output.FileOutputCommitter: Saved output of task 'attempt_local144329354_0001_r_000000_0' to hdfs://localhost:9000/WordCountTutorial/Output
2024-09-12 19:42:23,298 INFO mapred.LocalJobRunner: reduce > reduce
2024-09-12 19:42:23,298 INFO mapred.Task: Task 'attempt_local144329354_0001_r_000000_0' done.
2024-09-12 19:42:23,299 INFO mapred.Task: Final Counters for attempt_local144329354_0001_r_000000_0: Counters: 30
        File System Counters
        FILE: Number of bytes read=3661

FILE: Number of bytes written=641562
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=121
HDFS: Number of bytes written=123
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=15
Reduce shuffle bytes=189
Reduce input records=15
Reduce output records=15
Spilled Records=15
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=0
Total committed heap usage (bytes)=254803968
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=123
2024-09-12 19:42:23,299 INFO mapred.LocalJobRunner: Finishing task:
attempt_local144329354_0001_r_000000_0
2024-09-12 19:42:23,300 INFO mapred.LocalJobRunner: reduce task executor
complete.
2024-09-12 19:42:23,972 INFO mapreduce.Job: map 100% reduce 100%
2024-09-12 19:42:23,974 INFO mapreduce.Job: Job job_local144329354_0001
completed successfully
2024-09-12 19:42:23,990 INFO mapreduce.Job: Counters: 36
File System Counters
FILE: Number of bytes read=6912
FILE: Number of bytes written=1282935
FILE: Number of read operations=0

FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=242
HDFS: Number of bytes written=123
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=20
Map output records=19
Map output bytes=196
Map output materialized bytes=189
Input split bytes=120
Combine input records=19
Combine output records=15
Reduce input groups=15
Reduce shuffle bytes=189
Reduce input records=15
Reduce output records=15
Spilled Records=30
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=64
Total committed heap usage (bytes)=509607936
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=121
File Output Format Counters
Bytes Written=123

OUTPUT:

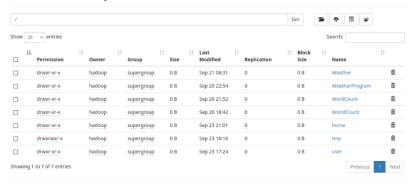hadoop@ubuntu:~/Desktop$ hadoop dfs -cat /WordCountTutorial/Output/*
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

```
accenture       1
badam           1
cse     1
cts     1
hello   1
infosys         1
kesari 1
pistah 1
raj     1
tcs     1
vimal 2
vinoth3
vishal 2
vishva          1
world 1
hadoop@ubuntu:~/Desktop$ ^C
hadoop@ubuntu:~/Desktop$
```

Browse Directory

/	Go!

Show 25 ⌄ entries   Search: [   ]

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
☐	drwxr-xr-x	hadoop	supergroup	0 B	Sep 21 08:31	0	0 B	Weather	🗑
☐	drwxr-xr-x	hadoop	supergroup	0 B	Sep 20 22:54	0	0 B	WeatherProgram	🗑
☐	drwxr-xr-x	hadoop	supergroup	0 B	Sep 20 21:52	0	0 B	WordCount	🗑
☐	drwxr-xr-x	hadoop	supergroup	0 B	Sep 20 18:42	0	0 B	WorldCount	🗑
☐	drwxr-xr-x	hadoop	supergroup	0 B	Sep 23 21:01	0	0 B	home	🗑
☐	drwxrwxr-x	hadoop	supergroup	0 B	Sep 23 18:16	0	0 B	tmp	🗑
☐	drwxr-xr-x	hadoop	supergroup	0 B	Sep 23 17:24	0	0 B	user	🗑

Showing 1 to 7 of 7 entries

Previous 1 Next

Hadoop, 2023.