

```
duser@student-ThinkCentre-M700:~$ hdfs dfs -mkdir /ksm
2024-04-23 14:30:31,610 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes
where applicable
hduser@student-ThinkCentre-M700:~$ hdfs dfs -copyFromLocal trees.txt
/ksm/
2024-04-23 14:31:48,860 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes
where applicable
hduser@student-ThinkCentre-M700:~$ hadoop jar w_c.jar driver
/k_data/trees.txt /k_data/output02/
hduser@student-ThinkCentre-M700:~$ hadoop jar w_c.jar driver
/ksm/trees.txt /k_data/output05/
hduser@student-ThinkCentre-M700:~$ hadoop jar wordcount.jar WDriver
/ksm/trees.txt /ksm/outputr/
2024-04-23 14:50:56,740 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes
where applicable
2024-04-23 14:50:57,327 INFO impl.MetricsConfig: Loaded properties from
hadoop-metrics2.properties
2024-04-23 14:50:57,392 INFO impl.MetricsSystemImpl: Scheduled Metric
snapshot period at 10 second(s).
2024-04-23 14:50:57,392 INFO impl.MetricsSystemImpl: JobTracker metrics
system started
2024-04-23 14:50:57,524 WARN mapreduce.JobResourceUploader: Hadoop
command-line option parsing not performed. Implement the Tool interface
and execute your application with ToolRunner to remedy this.
2024-04-23 14:50:57,609 INFO input.FileInputFormat: Total input files to
process : 1
2024-04-23 14:50:57,682 INFO mapreduce.JobSubmitter: number of splits:1
2024-04-23 14:50:57,784 INFO mapreduce.JobSubmitter: Submitting tokens
for job: job_local598663902_0001
2024-04-23 14:50:57,784 INFO mapreduce.JobSubmitter: Executing with
tokens: []
2024-04-23 14:50:57,918 INFO mapreduce.Job: The url to track the job:
http://localhost:8080/
2024-04-23 14:50:57,918 INFO mapreduce.Job: Running job:
job_local598663902_0001
2024-04-23 14:50:57,919 INFO mapred.LocalJobRunner: OutputCommitter set
in config null
2024-04-23 14:50:57,930 INFO output.FileOutputCommitter: File Output
Committer Algorithm version is 2
2024-04-23 14:50:57,930 INFO output.FileOutputCommitter:
FileOutputCommitter skip cleanup _temporary folders under output
directory:false, ignore cleanup failures: false
2024-04-23 14:50:57,931 INFO mapred.LocalJobRunner: OutputCommitter is
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-04-23 14:50:58,031 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-04-23 14:50:58,032 INFO mapred.LocalJobRunner: Starting task:
attempt_local598663902_0001_m_000000_0
2024-04-23 14:50:58,052 INFO output.FileOutputCommitter: File Output
Committer Algorithm version is 2
2024-04-23 14:50:58,052 INFO output.FileOutputCommitter:
FileOutputCommitter skip cleanup _temporary folders under output
directory:false, ignore cleanup failures: false
2024-04-23 14:50:58,067 INFO mapred.Task: Using
ResourceCalculatorProcessTree : [ ]
```

```

2024-04-23 14:50:58,071 INFO mapred.MapTask: Processing split:
hdfs://localhost:9000/ksm/trees.txt:0+1256
2024-04-23 14:50:58,157 INFO mapred.MapTask: (EQUATOR) 0 kvi
26214396(104857584)
2024-04-23 14:50:58,157 INFO mapred.MapTask: mapreduce.task.io.sort.mb:
100
2024-04-23 14:50:58,157 INFO mapred.MapTask: soft limit at 83886080
2024-04-23 14:50:58,157 INFO mapred.MapTask: bufstart = 0; bufvoid =
104857600
2024-04-23 14:50:58,157 INFO mapred.MapTask: kvstart = 26214396; length =
6553600
2024-04-23 14:50:58,160 INFO mapred.MapTask: Map output collector class =
org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2024-04-23 14:50:58,354 INFO mapred.LocalJobRunner:
2024-04-23 14:50:58,355 INFO mapred.MapTask: Starting flush of map output
2024-04-23 14:50:58,355 INFO mapred.MapTask: Spilling map output
2024-04-23 14:50:58,356 INFO mapred.MapTask: bufstart = 0; bufend = 2104;
bufvoid = 104857600
2024-04-23 14:50:58,356 INFO mapred.MapTask: kvstart =
26214396(104857584); kvend = 26213552(104854208); length = 845/6553600
2024-04-23 14:50:58,520 INFO mapred.MapTask: Finished spill 0
2024-04-23 14:50:58,537 INFO mapred.Task:
Task:attempt_local598663902_0001_m_000000_0 is done. And is in the
process of committing
2024-04-23 14:50:58,541 INFO mapred.LocalJobRunner: map
2024-04-23 14:50:58,541 INFO mapred.Task: Task
'attempt_local598663902_0001_m_000000_0' done.
2024-04-23 14:50:58,551 INFO mapred.Task: Final Counters for
attempt_local598663902_0001_m_000000_0: Counters: 23
  File System Counters
    FILE: Number of bytes read=4071
    FILE: Number of bytes written=642411
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=1256
    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=4
    Map output records=212
    Map output bytes=2104
    Map output materialized bytes=2534
    Input split bytes=100
    Combine input records=0
    Spilled Records=212
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=8
    Total committed heap usage (bytes)=293601280
  File Input Format Counters
    Bytes Read=1256
2024-04-23 14:50:58,551 INFO mapred.LocalJobRunner: Finishing task:
attempt_local598663902_0001_m_000000_0

```

2024-04-23 14:50:58,553 INFO mapred.LocalJobRunner: map task executor complete.

2024-04-23 14:50:58,556 INFO mapred.LocalJobRunner: Waiting for reduce tasks

2024-04-23 14:50:58,557 INFO mapred.LocalJobRunner: Starting task: attempt\_local598663902\_0001\_r\_000000\_0

2024-04-23 14:50:58,563 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2

2024-04-23 14:50:58,563 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup \_temporary folders under output directory:false, ignore cleanup failures: false

2024-04-23 14:50:58,563 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]

2024-04-23 14:50:58,566 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@4b369a58

2024-04-23 14:50:58,567 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!

2024-04-23 14:50:58,595 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=1280101504, maxSingleShuffleLimit=320025376, mergeThreshold=844867008, ioSortFactor=10, memToMemMergeOutputsThreshold=10

2024-04-23 14:50:58,598 INFO reduce.EventFetcher: attempt\_local598663902\_0001\_r\_000000\_0 Thread started: EventFetcher for fetching Map Completion Events

2024-04-23 14:50:58,628 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt\_local598663902\_0001\_m\_000000\_0 decomp: 2530 len: 2534 to MEMORY

2024-04-23 14:50:58,640 INFO reduce.InMemoryMapOutput: Read 2530 bytes from map-output for attempt\_local598663902\_0001\_m\_000000\_0

2024-04-23 14:50:58,641 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 2530, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->2530

2024-04-23 14:50:58,642 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning

2024-04-23 14:50:58,643 INFO mapred.LocalJobRunner: 1 / 1 copied.

2024-04-23 14:50:58,643 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs

2024-04-23 14:50:58,653 INFO mapred.Merger: Merging 1 sorted segments

2024-04-23 14:50:58,653 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 2524 bytes

2024-04-23 14:50:58,657 INFO reduce.MergeManagerImpl: Merged 1 segments, 2530 bytes to disk to satisfy reduce memory limit

2024-04-23 14:50:58,657 INFO reduce.MergeManagerImpl: Merging 1 files, 2534 bytes from disk

2024-04-23 14:50:58,658 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce

2024-04-23 14:50:58,658 INFO mapred.Merger: Merging 1 sorted segments

2024-04-23 14:50:58,658 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 2524 bytes

2024-04-23 14:50:58,659 INFO mapred.LocalJobRunner: 1 / 1 copied.

2024-04-23 14:50:58,706 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords

2024-04-23 14:50:58,780 INFO mapred.Task: Task:attempt\_local598663902\_0001\_r\_000000\_0 is done. And is in the process of committing

2024-04-23 14:50:58,782 INFO mapred.LocalJobRunner: 1 / 1 copied.

2024-04-23 14:50:58,783 INFO mapred.Task: Task  
attempt\_local598663902\_0001\_r\_000000\_0 is allowed to commit now  
2024-04-23 14:50:58,831 INFO output.FileOutputCommitter: Saved output of  
task 'attempt\_local598663902\_0001\_r\_000000\_0' to  
hdfs://localhost:9000/ksm/output  
2024-04-23 14:50:58,832 INFO mapred.LocalJobRunner: reduce > reduce  
2024-04-23 14:50:58,832 INFO mapred.Task: Task  
'attempt\_local598663902\_0001\_r\_000000\_0' done.  
2024-04-23 14:50:58,833 INFO mapred.Task: Final Counters for  
attempt\_local598663902\_0001\_r\_000000\_0: Counters: 30  
File System Counters  
FILE: Number of bytes read=9171  
FILE: Number of bytes written=644945  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=1256  
HDFS: Number of bytes written=361  
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=42  
Reduce shuffle bytes=2534  
Reduce input records=212  
Reduce output records=42  
Spilled Records=212  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=0  
Total committed heap usage (bytes)=293601280  
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=361  
2024-04-23 14:50:58,833 INFO mapred.LocalJobRunner: Finishing task:  
attempt\_local598663902\_0001\_r\_000000\_0  
2024-04-23 14:50:58,833 INFO mapred.LocalJobRunner: reduce task executor  
complete.  
2024-04-23 14:50:58,923 INFO mapreduce.Job: Job job\_local598663902\_0001  
running in uber mode : false  
2024-04-23 14:50:58,924 INFO mapreduce.Job: map 100% reduce 100%  
2024-04-23 14:50:58,925 INFO mapreduce.Job: Job job\_local598663902\_0001  
completed successfully  
2024-04-23 14:50:58,934 INFO mapreduce.Job: Counters: 36  
File System Counters  
FILE: Number of bytes read=13242  
FILE: Number of bytes written=1287356  
FILE: Number of read operations=0

```
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=2512
HDFS: Number of bytes written=361
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=4
  Map output records=212
  Map output bytes=2104
  Map output materialized bytes=2534
  Input split bytes=100
  Combine input records=0
  Combine output records=0
  Reduce input groups=42
  Reduce shuffle bytes=2534
  Reduce input records=212
  Reduce output records=42
  Spilled Records=424
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=8
  Total committed heap usage (bytes)=587202560
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=1256
File Output Format Counters
  Bytes Written=361
```

```
hduser@student-ThinkCentre-M700:~$ hdfs dfs -cat /ksm/output/*
2024-04-23 15:11:35,904 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes
where applicable
The      1
Trees    1
a        1
above    1
and      1
angiosperms 1
are      1
as       1
branches 1
but      1
compete  1
consist  1
evolved  1
for      1
group    1
gymnosperms 1
hardwoods; 1
```

```

have 1
independently 1
majority 1
many 1
monophyletic 1
not 1
of 1
or 1
other 1
plant 1
plants 1
rest, 1
softwoods. 1
species 1
sunlight. 1
taxonomic 1
that 1
the 1
to 1
tower 1
tree 1
trunk 1
variety 1
way 1
wide 1

```

