


```
Activities Terminal Apr 19 19:35 hadoop@saurabh-VirtualBox: ~/analogs

hadoop@saurabh-VirtualBox: ~/analogs
hadoop@saurabh-VirtualBox:~$ jps
2466 Jps
hadoop@saurabh-VirtualBox:~$ 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 [saurabh-VirtualBox]
Starting resource manager
Starting nodemanagers
hadoop@saurabh-VirtualBox:~$ cd analogs
hadoop@saurabh-VirtualBox:~/analogs$ SHADOOP_HOME/bin/hdfs dfs -put ~/ip/
hadoop@saurabh-VirtualBox:~/analogs$ SHADOOP_HOME/bin/hadoop jar analogs.jar /ip /op
2024-04-19 18:42:00,155 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-04-19 18:42:00,276 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-04-19 18:42:00,276 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-04-19 18:42:00,309 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2024-04-19 18:42:00,498 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and
execute your application with ToolRunner to remedy this.
2024-04-19 18:42:00,643 INFO mapreduce.JobSubmitter: Cleaning up the staging area file:/tmp/hadoop/mapred/staging/hadoop1451632081/.staging/jo
b_local1451632081_0001
org.apache.hadoop.mapred.InvalidInputException: Input path does not exist: hdfs://localhost:9000/ip
    at org.apache.hadoop.mapred.FileInputFormat.singleThreadedListStatus(FileInputFormat.java:304)
    at org.apache.hadoop.mapred.FileInputFormat.listStatus(FileInputFormat.java:244)
    at org.apache.hadoop.mapred.FileInputFormat.get_splits(FileInputFormat.java:332)
    at org.apache.hadoop.mapreduce.JobSubmitter.writeOld_splits(JobSubmitter.java:338)
    at org.apache.hadoop.mapreduce.JobSubmitter.write_splits(JobSubmitter.java:329)
    at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:200)
    at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1678)
```

```
Activities Terminal Apr 19 19:35 hadoop@saurabh-VirtualBox: ~/analogs

hadoop@saurabh-VirtualBox: ~/analogs
2024-04-19 18:52:40,912 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2024-04-19 18:52:40,912 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2024-04-19 18:52:40,922 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2024-04-19 18:52:41,076 INFO mapreduce.Job: Job job_local311841520_0001 running in uber mode : false
2024-04-19 18:52:41,079 INFO mapreduce.Job: map 0% reduce 0%
2024-04-19 18:52:41,265 INFO mapred.LocalJobRunner:
2024-04-19 18:52:41,265 INFO mapred.MapTask: Starting flush of map output
2024-04-19 18:52:41,265 INFO mapred.MapTask: Spilling map output
2024-04-19 18:52:41,265 INFO mapred.MapTask: bufstart = 0; bufend = 15743; bufvoid = 104857600
2024-04-19 18:52:41,265 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26210404(104841616); length = 3993/6553600
2024-04-19 18:52:41,324 INFO mapred.MapTask: Finished spill 0
2024-04-19 18:52:41,412 INFO mapred.Task: Task:attempt_local311841520_0001_m_000000_0 is done. And is in the process of committing
2024-04-19 18:52:41,420 INFO mapred.LocalJobRunner: hdfs://localhost:9000/ip/SalesJan2009.csv:0+123637
2024-04-19 18:52:41,421 INFO mapred.Task: Task 'attempt_local311841520_0001_m_000000_0' done.
2024-04-19 18:52:41,428 INFO mapred.Task: Final counters for attempt_local311841520_0001_m_000000_0: Counters: 23
```



```
Map input records=999
Map output records=999
Map output bytes=15743
Map output materialized bytes=17747
Input split bytes=93
Combine input records=0
Spilled Records=999
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=37
Total committed heap usage (bytes)=319815680
File Input Format Counters
  Bytes Read=123637
2024-04-19 18:52:41,431 INFO mapred.LocalJobRunner: Finishing task: attempt_local311841520_0001_m_000000_0
2024-04-19 18:52:41,431 INFO mapred.LocalJobRunner: map task executor complete.
2024-04-19 18:52:41,446 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2024-04-19 18:52:41,447 INFO mapred.LocalJobRunner: Starting task: attempt_local311841520_0001_r_000000_0
2024-04-19 18:52:41,459 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-04-19 18:52:41,459 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-04-19 18:52:41,459 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2024-04-19 18:52:41,463 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@73d5c0d2
2024-04-19 18:52:41,465 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2024-04-19 18:52:41,495 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=802632512, maxSingleShuffleLimit=200658128, mergeThreshold=529737472, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2024-04-19 18:52:41,500 INFO reduce.EventFetcher: attempt_local311841520_0001_r_000000_0 Thread: started EventFetcher for fetching Map output
```

```
hadoop@saurabh-VirtualBox:~$ jps
11808 NameNode
11937 DataNode
12150 SecondaryNameNode
13320 Jps
12472 NodeManager
12351 ResourceManager
hadoop@saurabh-VirtualBox:~$ cd analogs
hadoop@saurabh-VirtualBox:~/analogs$ ls
hadoop-mapreduce-example-file.txt  SalesCountryDriver.java  SalesMapper.java
Manifest.txt                        SalesCountryReducer.java
SalesCountry                        SalesJan2009.csv
hadoop@saurabh-VirtualBox:~/analogs$ jar -cfm analogs.jar Manifest.txt SalesCountry/*.txt
SalesCountry/*.txt : no such file or directory
hadoop@saurabh-VirtualBox:~/analogs$ jar -cfm analogs.jar Manifest.txt SalesCountry/*.java
SalesCountry/*.java : no such file or directory
hadoop@saurabh-VirtualBox:~/analogs$ jar -cfm analogs.jar Manifest.txt SalesCountry/*.class
hadoop@saurabh-VirtualBox:~/analogs$ ls
analogs.jar  SalesCountry  SalesJan2009.csv
hadoop-mapreduce-example-file.txt  SalesCountryDriver.java  SalesMapper.java
Manifest.txt                        SalesCountryReducer.java
hadoop@saurabh-VirtualBox:~/analogs$ sudo mkdir ~/ip
[sudo] password for hadoop:
hadoop@saurabh-VirtualBox:~/analogs$
```

```
hadoop@saurabh-VirtualBox: ~/
File Output Format Counters
  Bytes Written=661
2024-04-19 18:52:42,192 INFO mapred.LocalJobRunner: Finishing task: attempt_local311841520_0001_r_000000_0
2024-04-19 18:52:42,193 INFO mapred.LocalJobRunner: reduce task executor complete.
2024-04-19 18:52:43,086 INFO mapreduce.Job: map 100% reduce 100%
2024-04-19 18:52:43,087 INFO mapreduce.Job: Job job_local311841520_0001 completed successfully
2024-04-19 18:52:43,103 INFO mapreduce.Job: Counters: 36
File System Counters
  FILE: Number of bytes read=41770
  FILE: Number of bytes written=1336181
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=247274
  HDFS: Number of bytes written=661
  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=999
  Map output records=999
  Map output bytes=15743
```

```
Map input records=999
Map output records=999
Map output bytes=15743
Map output materialized bytes=17747
Input split bytes=93
Combine input records=0
Combine output records=0
Reduce input groups=58
Reduce shuffle bytes=17747
Reduce input records=999
Reduce output records=58
Spilled Records=1998
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=37
Total committed heap usage (bytes)=639631360
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=123637
File Output Format Counters
  Bytes Written=661
hadoop@saurabh-VirtualBox: ~/analog$ $HADOOP_HOME/bin/hdfs dfs -cat /op/part-00000
hadoop@saurabh-VirtualBox: ~/analog$ hadoop fs -ls /op/part-00000
```

Browse Directory

Show

25

▼

entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	0 B	Apr 19 18:52	1	128 MB	_SUCCESS	
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	661 B	Apr 19 18:52	1	128 MB	part-00000	

Showing 1 to 2 of 2 entries

Previous

1

Next

Browse Directory

Show

25

▼

entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	120.74 KB	Apr 19 18:50	1	128 MB	SalesJan2009.csv	

Showing 1 to 1 of 1 entries

Previous

1

Next