

1. Download [hadoop-3.2.3.tar.gz](#)

```
theia@theiadocker-rajendraabro:/home/project$ curl https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz --output hadoop-3.3.6.tar.gz
% Total    % Received % Xferd  Average Speed   Time
Time      Time Current
                                         Dload Upload Total Spent   Left
Speed
0     0     0     0     0     0     0     0 0--:- 0     0     0     0     0     0     0
0     0 0--:- 0     0     0     0     0     0     0 0--:- 0     0     0     0     0     0
0     0     0     0     0 0--:- 0     0     0     0     0     0     0     0     0 0--:-
0     0     0     0     0     0     0     0 0--:- 0     0     0     0     0     0     0
0     0 0--:- 0     0     0     0     0     0     0     0 0--:- 0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     0     0 0--:- 0     0     0
0     0     0     0     0 0--:- 0     0     0     0     0     0     0     0     0 0--:-
0     0     0     0     0     0     0     0     0     0     0 0--:- 0     0     0     0
0     0 0--:- 0     0     0     0     0     0     0     0     0 0--:- 0     0     0     0
0     0     0     0     0     0     0     0     0     0     0     0     0 0--:- 0     0
0     0 0--:- 0     0     0     0     0     0     0     0     0 0--:- 0     0     0     0
0     0     0     0     0 0--:- 0     0     0     0     0     0     0     0     0 0--:-
0     0     0     0     0     0     0     0     0     0     0     0     0 0--:- 0     0
0     0 0--:- 0     0     0     0     0     0     0     0     0 0--:- 0     0     0     0
0     0     0     0     0 0--:- 0     0     0     0     0     0     0     0     0 0--:-
0     0     0     0     0     0     0     0     0     0     0     0     0 0--:- 0     0
0     0 0--:- 0     0     0     0     0     0     0     0     0 0--:- 0 696M 0
19274 0     0 1060 0 7d 32 696M 32 223M 0     0
11.9M 0 0:074 696M 74 520M 0     0 26.3M 0
0:096 696M 96 670M 0     0 32.2M 0 0:0100 696M
100 696M 0     0 32.8M
```

2.Extract the tar file in the currently directory.

```
theia@theiadocker-rajendraabro:/home/project$ tar -xvf  
hadoop-3.3.6.tar.gz
```

```
theia@theiadocker-rajendraabro:/home/project$ cd  
hadoop-3.3.6
```

Check the hadoop command to see if it is setup

```
theia@theiadocker-rajendraabro:/home/project/hadoop  
p-3.3.6$ bin/hadoop
```

Usage: hadoop [OPTIONS] SUBCOMMAND
[SUBCOMMAND OPTIONS]
or hadoop [OPTIONS] CLASSNAME [CLASSNAME
OPTIONS]
where CLASSNAME is a user-provided Java class

OPTIONS is none or any of:

| | |
|--------------------------------|--|
| --config dir | Hadoop config directory |
| --debug | turn on shell script debug mode |
| --help | usage information |
| buildpaths | attempt to add class files from build tree |
| hostnames list[,of,host,names] | hosts to use in worker mode |
| hosts filename | list of hosts |

to use in
worker mode

loglevel level set the log4j
 level for this
 command

workers turn on worker
 mode

SUBCOMMAND is one of:

Admin Commands:

daemonlog get/set the log level for each
 daemon

Client Commands:

archive create a Hadoop archive

checknative check native Hadoop and
 compression libraries availability

classpath prints the class path needed to
 get the Hadoop jar and the
 required libraries

conftest validate configuration XML files

credential interact with credential providers

distch distributed metadata changer
distcp copy file or directories
recursively
dtutil operations related to delegation
tokens
envvars display computed Hadoop
environment variables
fs run a generic filesystem user
client
gridmix submit a mix of synthetic job,
modeling a profiled from
production load
jar <jar> run a jar file. NOTE: please use
"yarn jar" to launch YARN
applications, not this command.
jnopath prints the java.library.path
kdiag Diagnose Kerberos Problems
kerbname show auth_to_local principal
conversion
key manage keys via the KeyProvider
rumenfolder scale a rumen input trace
rumentrace convert logs into a rumen trace
s3guard S3 Commands
trace view and modify Hadoop tracing
settings
version print the version

Daemon Commands:

kms run KMS, the Key Management Server

registrydns run the registry DNS server

SUBCOMMAND may print help when invoked w/o parameters or with

6. Run the following command to download data.txt to your current directory

p-3.3.6\$ curl

```
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-BD0225EN-SkillsNetwork/labs/data/data.txt  
--output data.txtta.txt --output data.txt
```

SkillsN

| Skills Network Transfer Monitor | | | | | | | |
|---------------------------------|------|----------|-------|-----------|--------|-----------------------|--------|
| File | | Transfer | | Bandwidth | | System | |
| File | Size | Time | Speed | Dload | Upload | Total | Spent |
| | | | | 0 | 0 | 0 --:--:--:--:100 | 47 100 |
| | | | | 47 | 0 | 0 364 --:--:--:--:00m | |

7. Run the Map reduce application for wordcount on data.txt and store the output in /user/root/output

```
$ bin/hadoop jar  
share/hadoop/mapreduce/hadoop-mapreduce-examples-3.  
3.6.jar wordcount data.txt output
```

2025-12-23 11:35:39,598 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties

2025-12-23 11:35:39,796 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).

2025-12-23 11:35:39,796 INFO impl.MetricsSystemImpl: JobTracker metrics system started

2025-12-23 11:35:40,008 INFO input.FileInputFormat: Total input files to process : 1

2025-12-23 11:35:40,048 INFO mapreduce.JobSubmitter: number of splits:1

2025-12-23 11:35:40,319 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local211696964_0001

2025-12-23 11:35:40,320 INFO mapreduce.JobSubmitter: Executing with tokens: []

2025-12-23 11:35:40,503 INFO mapreduce.Job: The url to track the job: http://localhost:8080/

2025-12-23 11:35:40,504 INFO mapred.LocalJobRunner: OutputCommitter set in config null

2025-12-23 11:35:40,504 INFO mapreduce.Job: Running job: job_local211696964_0001

2025-12-23 11:35:40,509 INFO
output.PathOutputCommitterFactory: No output committer
factory defined, defaulting to FileOutputStreamCommitterFactory
2025-12-23 11:35:40,510 INFO output.FileOutputCommitter:
File Output Committer Algorithm version is 2
2025-12-23 11:35:40,510 INFO output.FileOutputCommitter:
FileOutputCommitter skip cleanup _temporary folders under
output directory:false, ignore cleanup failures: false
2025-12-23 11:35:40,511 INFO mapred.LocalJobRunner:
OutputCommitter is
org.apache.hadoop.mapreduce.lib.output.FileOutputCommit
ter
2025-12-23 11:35:40,543 INFO mapred.LocalJobRunner:
Waiting for map tasks
2025-12-23 11:35:40,544 INFO mapred.LocalJobRunner:
Starting task: attempt_local211696964_0001_m_000000_0
2025-12-23 11:35:40,576 INFO
output.PathOutputCommitterFactory: No output committer
factory defined, defaulting to FileOutputStreamCommitterFactory
2025-12-23 11:35:40,576 INFO output.FileOutputCommitter:
File Output Committer Algorithm version is 2
2025-12-23 11:35:40,576 INFO output.FileOutputCommitter:
FileOutputCommitter skip cleanup _temporary folders under
output directory:false, ignore cleanup failures: false
2025-12-23 11:35:40,601 INFO mapred.Task: Using
ResourceCalculatorProcessTree : []

2025-12-23 11:35:40,607 INFO mapred.MapTask:
Processing split:
file:/home/project/hadoop-3.3.6/data.txt:0+47
2025-12-23 11:35:40,695 INFO mapred.MapTask:
(EQUATOR) 0 kvi 26214396(104857584)
2025-12-23 11:35:40,695 INFO mapred.MapTask:
mapreduce.task.io.sort.mb: 100
2025-12-23 11:35:40,695 INFO mapred.MapTask: soft limit
at 83886080
2025-12-23 11:35:40,695 INFO mapred.MapTask: bufstart =
0; bufvoid = 104857600
2025-12-23 11:35:40,695 INFO mapred.MapTask: kvstart =
26214396; length = 6553600
2025-12-23 11:35:40,700 INFO mapred.MapTask: Map
output collector class =
org.apache.hadoop.mapred.MapTask\$MapOutputBuffer
2025-12-23 11:35:40,710 INFO mapred.LocalJobRunner:
2025-12-23 11:35:40,711 INFO mapred.MapTask: Starting
flush of map output
2025-12-23 11:35:40,711 INFO mapred.MapTask: Spilling
map output
2025-12-23 11:35:40,711 INFO mapred.MapTask: bufstart =
0; bufend = 71; bufvoid = 104857600
2025-12-23 11:35:40,711 INFO mapred.MapTask: kvstart =
26214396(104857584); kvend = 26214376(104857504);
length = 21/6553600
2025-12-23 11:35:40,725 INFO mapred.MapTask: Finished
spill 0

2025-12-23 11:35:40,737 INFO mapred.Task:
Task:attempt_local211696964_0001_m_000000_0 is done.

And is in the process of committing

2025-12-23 11:35:40,739 INFO mapred.LocalJobRunner:
map

2025-12-23 11:35:40,739 INFO mapred.Task: Task
'attempt_local211696964_0001_m_000000_0' done.

2025-12-23 11:35:40,744 INFO mapred.Task: Final
Counters for attempt_local211696964_0001_m_000000_0:
Counters: 18

File System Counters

FILE: Number of bytes read=281554
FILE: Number of bytes written=918295
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0

Map-Reduce Framework

Map input records=3
Map output records=6
Map output bytes=71
Map output materialized bytes=59
Input split bytes=105
Combine input records=6
Combine output records=4
Spilled Records=4
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=9

Total committed heap usage (bytes)=188743680
File Input Format Counters
Bytes Read=47

2025-12-23 11:35:40,744 INFO mapred.LocalJobRunner:
Finishing task: attempt_local211696964_0001_m_000000_0
2025-12-23 11:35:40,745 INFO mapred.LocalJobRunner:
map task executor complete.
2025-12-23 11:35:40,748 INFO mapred.LocalJobRunner:
Waiting for reduce tasks
2025-12-23 11:35:40,748 INFO mapred.LocalJobRunner:
Starting task: attempt_local211696964_0001_r_000000_0
2025-12-23 11:35:40,756 INFO
output.PathOutputCommitterFactory: No output committer
factory defined, defaulting to FileOutputStreamCommitterFactory
2025-12-23 11:35:40,756 INFO output.FileOutputCommitter:
File Output Committer Algorithm version is 2
2025-12-23 11:35:40,756 INFO output.FileOutputCommitter:
FileOutputCommitter skip cleanup _temporary folders under
output directory:false, ignore cleanup failures: false
2025-12-23 11:35:40,757 INFO mapred.Task: Using
ResourceCalculatorProcessTree : []
2025-12-23 11:35:40,759 INFO mapred.ReduceTask: Using
ShuffleConsumerPlugin:
org.apache.hadoop.mapreduce.task.reduce.Shuffle@3fd9d
a61
2025-12-23 11:35:40,761 WARN impl.MetricsSystemImpl:
JobTracker metrics system already initialized!

2025-12-23 11:35:40,785 INFO reduce.MergeManagerImpl:
MergerManager: memoryLimit=700239040,
maxSingleShuffleLimit=175059760,
mergeThreshold=462157792, ioSortFactor=10,
memToMemMergeOutputsThreshold=10

2025-12-23 11:35:40,787 INFO reduce.EventFetcher:
attempt_local211696964_0001_r_000000_0 Thread started:
EventFetcher for fetching Map Completion Events

2025-12-23 11:35:41,357 INFO reduce.LocalFetcher:
localfetcher#1 about to shuffle output of map
attempt_local211696964_0001_m_000000_0 decomp: 55
len: 59 to MEMORY

2025-12-23 11:35:41,361 INFO
reduce.InMemoryMapOutput: Read 55 bytes from
map-output for

attempt_local211696964_0001_m_000000_0

2025-12-23 11:35:41,364 INFO reduce.MergeManagerImpl:
closeInMemoryFile -> map-output of size: 55,
inMemoryMapOutputs.size() -> 1, commitMemory -> 0,
usedMemory ->55

2025-12-23 11:35:41,365 INFO reduce.EventFetcher:
EventFetcher is interrupted.. Returning

2025-12-23 11:35:41,367 INFO mapred.LocalJobRunner: 1 /
1 copied.

2025-12-23 11:35:41,367 INFO reduce.MergeManagerImpl:
finalMerge called with 1 in-memory map-outputs and 0
on-disk map-outputs

2025-12-23 11:35:41,375 INFO mapred.Merger: Merging 1 sorted segments

2025-12-23 11:35:41,376 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 45 bytes

2025-12-23 11:35:41,377 INFO reduce.MergeManagerImpl: Merged 1 segments, 55 bytes to disk to satisfy reduce memory limit

2025-12-23 11:35:41,378 INFO reduce.MergeManagerImpl: Merging 1 files, 59 bytes from disk

2025-12-23 11:35:41,378 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce

2025-12-23 11:35:41,379 INFO mapred.Merger: Merging 1 sorted segments

2025-12-23 11:35:41,380 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 45 bytes

2025-12-23 11:35:41,380 INFO mapred.LocalJobRunner: 1 / 1 copied.

2025-12-23 11:35:41,383 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords

2025-12-23 11:35:41,385 INFO mapred.Task: Task:attempt_local211696964_0001_r_000000_0 is done. And is in the process of committing

2025-12-23 11:35:41,386 INFO mapred.LocalJobRunner: 1 / 1 copied.

2025-12-23 11:35:41,386 INFO mapred.Task: Task attempt_local211696964_0001_r_000000_0 is allowed to commit now

2025-12-23 11:35:41,509 INFO mapreduce.Job: Job job_local211696964_0001 running in uber mode : false

2025-12-23 11:35:41,511 INFO mapreduce.Job: map 100% reduce 0%

2025-12-23 11:35:42,031 INFO output.FileOutputCommitter: Saved output of task

'attempt_local211696964_0001_r_000000_0' to file:/home/project/hadoop-3.3.6/output

2025-12-23 11:35:42,032 INFO mapred.LocalJobRunner: reduce > reduce

2025-12-23 11:35:42,032 INFO mapred.Task: Task 'attempt_local211696964_0001_r_000000_0' done.

2025-12-23 11:35:42,033 INFO mapred.Task: Final Counters for attempt_local211696964_0001_r_000000_0: Counters: 24

File System Counters

FILE: Number of bytes read=281704

FILE: Number of bytes written=918403

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

Map-Reduce Framework

Combine input records=0

Combine output records=0

Reduce input groups=4

Reduce shuffle bytes=59
Reduce input records=4
Reduce output records=4
Spilled Records=4
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=0
Total committed heap usage (bytes)=188743680

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=49

2025-12-23 11:35:42,033 INFO mapred.LocalJobRunner:
Finishing task: attempt_local211696964_0001_r_000000_0

2025-12-23 11:35:42,033 INFO mapred.LocalJobRunner:
reduce task executor complete.

2025-12-23 11:35:42,512 INFO mapreduce.Job: map 100%
reduce 100%

2025-12-23 11:35:42,513 INFO mapreduce.Job: Job
job_local211696964_0001 completed successfully

2025-12-23 11:35:42,520 INFO mapreduce.Job: Counters:

File System Counters

FILE: Number of bytes read=563258
FILE: Number of bytes written=1836698
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0

Map-Reduce Framework

Map input records=3
Map output records=6
Map output bytes=71
Map output materialized bytes=59
Input split bytes=105
Combine input records=6
Combine output records=4
Reduce input groups=4
Reduce shuffle bytes=59
Reduce input records=4
Reduce output records=4
Spilled Records=8
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=9
Total committed heap usage (bytes)=377487360

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=47
File Output Format Counters
[00m\$ bin/hadoop jar
share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.
.6.jar wordcount data.txt output
2025-12-23 11:36:00,948 INFO impl.MetricsConfig: Loaded
properties from hadoop-metrics2.properties
2025-12-23 11:36:01,116 INFO impl.MetricsSystemImpl:
Scheduled Metric snapshot period at 10 second(s).
2025-12-23 11:36:01,116 INFO impl.MetricsSystemImpl:
JobTracker metrics system started
org.apache.hadoop.mapred.FileAlreadyExistsException:
Output directory file:/home/project/hadoop-3.3.6/output
already exists
at
org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.
checkOutputSpecs(FileOutputFormat.java:164)
at
org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(J
obSubmitter.java:277)
at
org.apache.hadoop.mapreduce.JobSubmitter.submitJobInte
rnal(JobSubmitter.java:143)

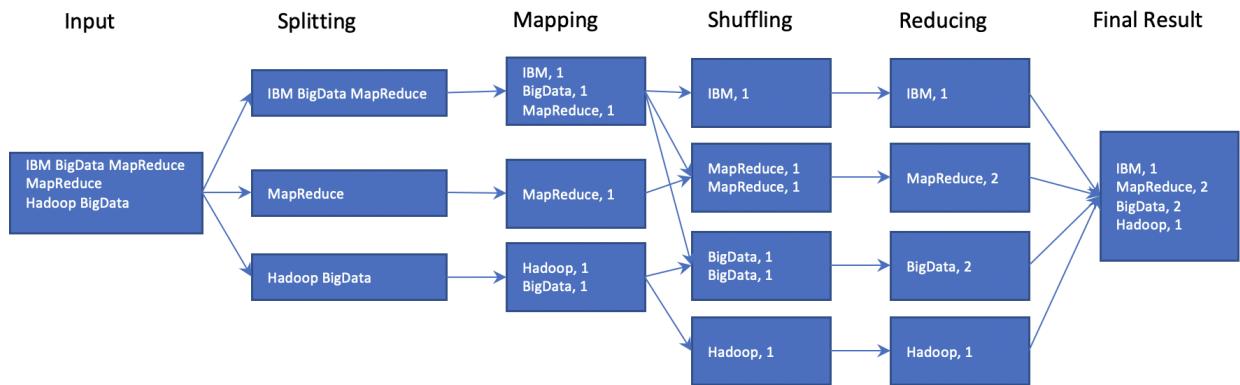
at
org.apache.hadoop.mapreduce.Job\$11.run(Job.java:1678)
at
org.apache.hadoop.mapreduce.Job\$11.run(Job.java:1675)
at
java.base/java.security.AccessController.doPrivileged(Acces
sController.java:714)
at
java.base/javax.security.auth.Subject.doAs(Subject.java:525
)
at
org.apache.hadoop.security.UserGroupInformation.doAs(Us
erGroupInformation.java:1899)
at
org.apache.hadoop.mapreduce.Job.submit(Job.java:1675)
at
org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.
java:1696)
at
org.apache.hadoop.examples.WordCount.main(WordCount.
java:87)
at
java.base/jdk.internal.reflect.DirectMethodHandleAccessor.i
nvoke(DirectMethodHandleAccessor.java:103)
at
java.base/java.lang.reflect.Method.invoke(Method.java:580)

```
at  
org.apache.hadoop.util.ProgramDriver$ProgramDescription.  
invoke(ProgramDriver.java:71)  
at  
org.apache.hadoop.util.ProgramDriver.run(ProgramDriver.ja  
va:144)  
at  
org.apache.hadoop.examples.ExampleDriver.main(Example  
Driver.java:74)  
at  
java.base/jdk.internal.reflect.DirectMethodHandleAccessor.i  
nvoke(DirectMethodHandleAccessor.java:103)  
at  
java.base/java.lang.reflect.Method.invoke(Method.java:580)  
at org.apache.hadoop.util.RunJar.run(RunJar.java:328)  
at  
org.apache.hadoop.util.RunJar.main(RunJar.java:241)
```

You should see part-r-00000 with _SUCCESS
indicating that the wordcount has been done.

theia@theiadocker-rajendraabro:/home/project/hadoo
p-3.3.6\$ ls output

_SUCCESS part-r-00000



```

theia@theiadocker-rajendraabro:/home/project/hadoop
p-3.3.6$ cat output/part-r-00000
BigData 2
Hadoop 1
IBM 1
MapReduce 2
  
```