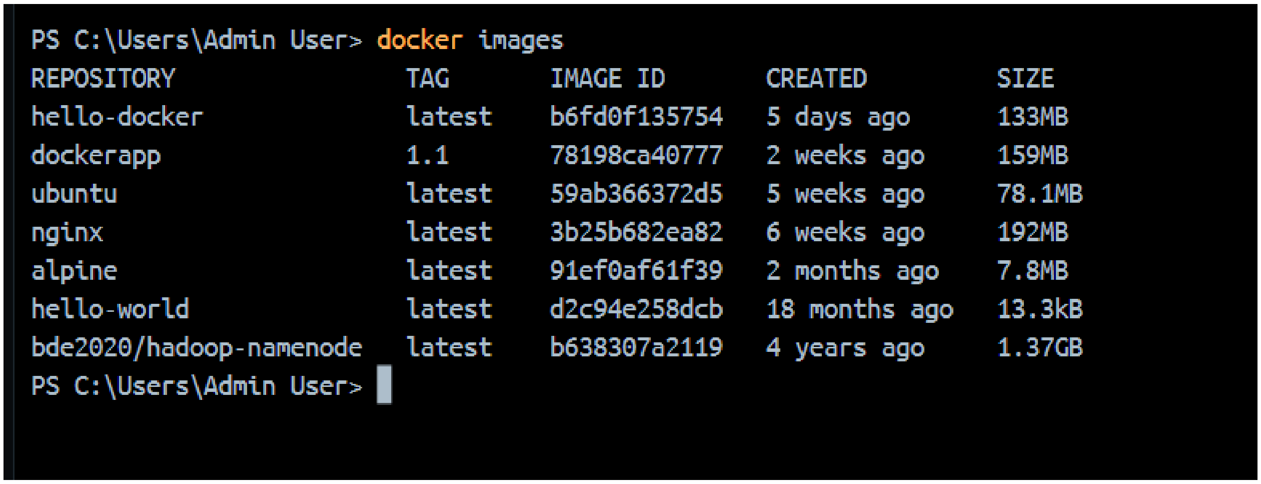
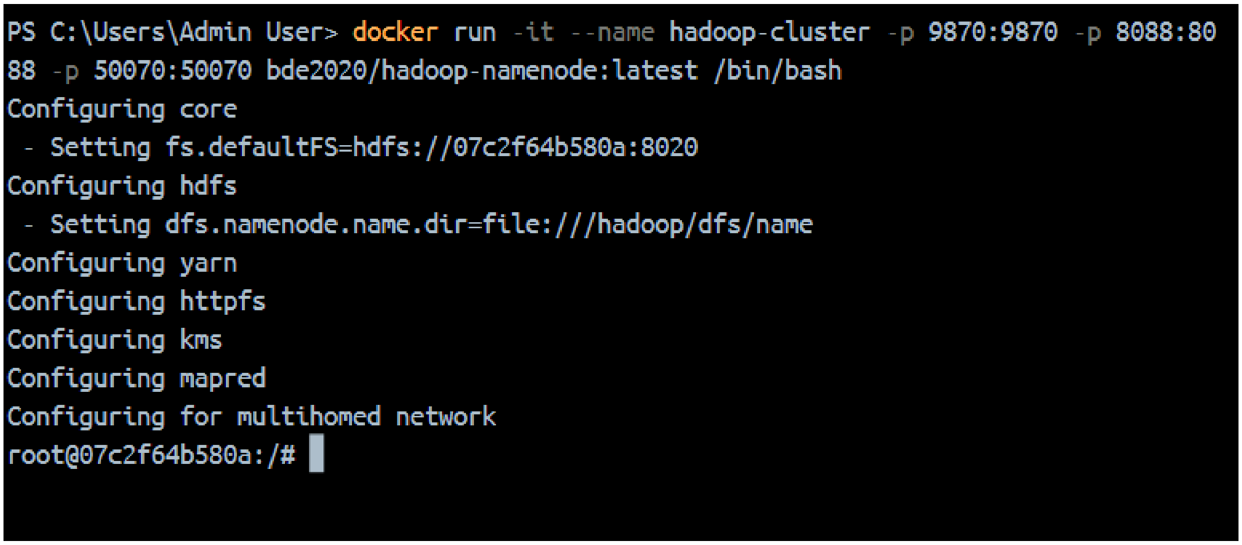
To list the docker images to confirm the Hadoop image is downloaded:



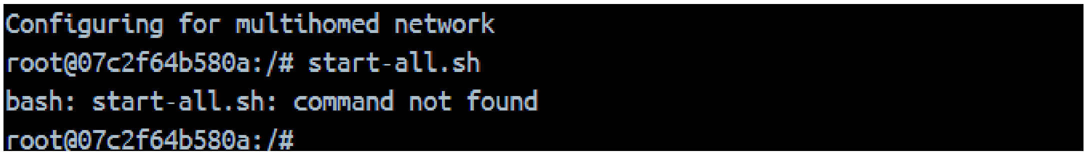
```
PS C:\Users\Admin User> docker images
REPOSITORY                 TAG       IMAGE ID       CREATED         SIZE
hello-docker               latest    b6fd0f135754   5 days ago      133MB
dockerapp                  1.1       78198ca40777   2 weeks ago     159MB
ubuntu                     latest    59ab366372d5   5 weeks ago     78.1MB
nginx                      latest    3b25b682ea82   6 weeks ago     192MB
alpine                     latest    91ef0af61f39   2 months ago    7.8MB
hello-world                latest    d2c94e258dcb   18 months ago   13.3kB
bde2020/hadoop-namenode    latest    b638307a2119   4 years ago     1.37GB
PS C:\Users\Admin User>
```

To run a Docker container for a Hadoop NameNode instance:



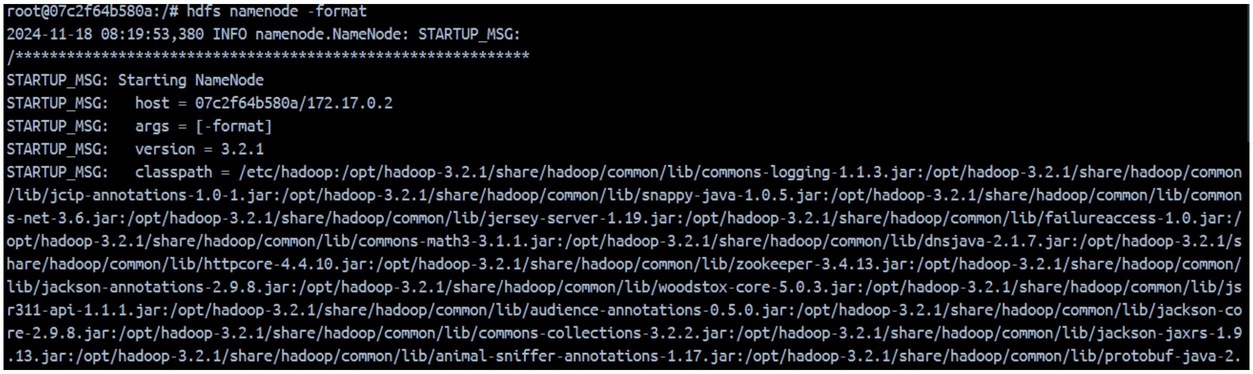
```
PS C:\Users\Admin User> docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:80
88 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash
Configuring core
 - Setting fs.defaultFS=hdfs://07c2f64b580a:8020
Configuring hdfs
 - Setting dfs.namenode.name.dir=file:///hadoop/dfs/name
Configuring yarn
Configuring httpfs
Configuring kms
Configuring mapred
Configuring for multihomed network
root@07c2f64b580a:/#
```

To start the Hadoop services once inside the container's shell at once:



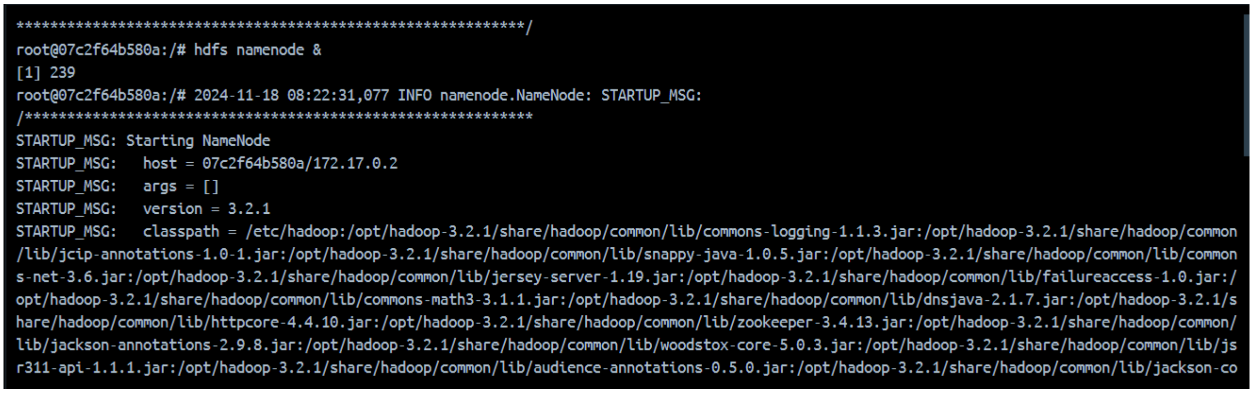
```
Configuring for multihomed network
root@07c2f64b580a:/# start-all.sh
bash: start-all.sh: command not found
root@07c2f64b580a:/#
```

To initialize and format the Hadoop Distributed File System (HDFS) NameNode:



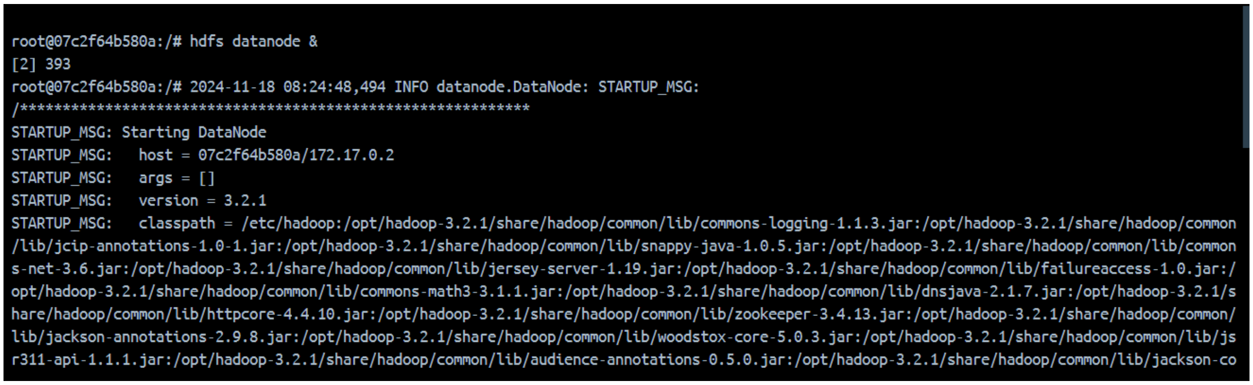
```
root@07c2f64b580a:/# hdfs namenode -format
2024-11-18 08:19:53,380 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = 07c2f64b580a/172.17.0.2
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.2.1
STARTUP_MSG:   classpath = /etc/hadoop:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-logging-1.1.3.jar:/opt/hadoop-3.2.1/share/hadoop/common
/lib/jcip-annotations-1.0-1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/snappy-java-1.0.5.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/common
s-net-3.6.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-server-1.19.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/failureaccess-1.0.jar:/
opt/hadoop-3.2.1/share/hadoop/common/lib/commons-math3-3.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/dnsjava-2.1.7.jar:/opt/hadoop-3.2.1/s
hare/hadoop/common/lib/httpcore-4.4.10.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/zookeeper-3.4.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/
lib/jackson-annotations-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/woodstox-core-5.0.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/js
r311-api-1.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/audience-annotations-0.5.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-co
re-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-collections-3.2.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-jaxrs-1.9
.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/animal-sniffer-annotations-1.17.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/protobuf-java-2.
```

To start the HDFS NameNode in the background:



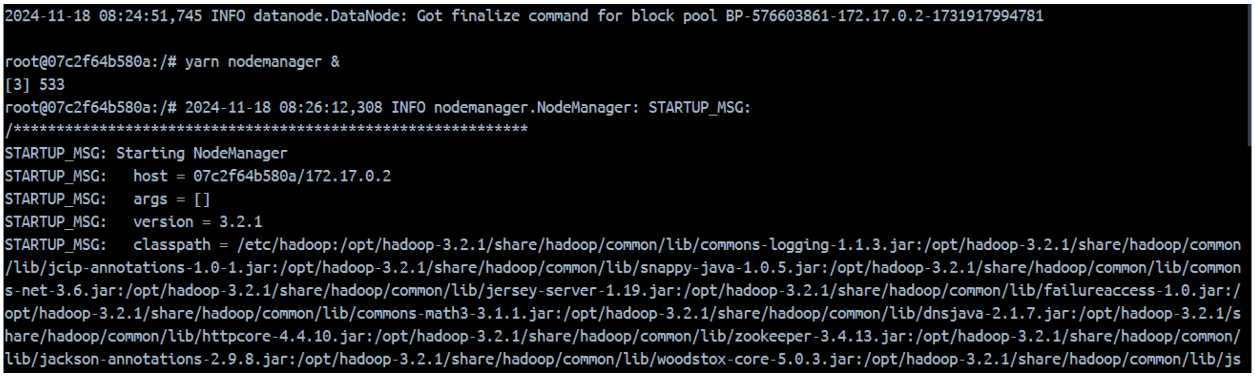
```
************************************************************/
root@07c2f64b580a:/# hdfs namenode &
[1] 239
root@07c2f64b580a:/# 2024-11-18 08:22:31,077 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = 07c2f64b580a/172.17.0.2
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 3.2.1
STARTUP_MSG:   classpath = /etc/hadoop:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-logging-1.1.3.jar:/opt/hadoop-3.2.1/share/hadoop/common
/lib/jcip-annotations-1.0-1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/snappy-java-1.0.5.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/common
s-net-3.6.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-server-1.19.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/failureaccess-1.0.jar:/
opt/hadoop-3.2.1/share/hadoop/common/lib/commons-math3-3.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/dnsjava-2.1.7.jar:/opt/hadoop-3.2.1/s
hare/hadoop/common/lib/httpcore-4.4.10.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/zookeeper-3.4.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/
lib/jackson-annotations-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/woodstox-core-5.0.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/js
r311-api-1.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/audience-annotations-0.5.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-co
```

To start the HDFS DataNode in the background:



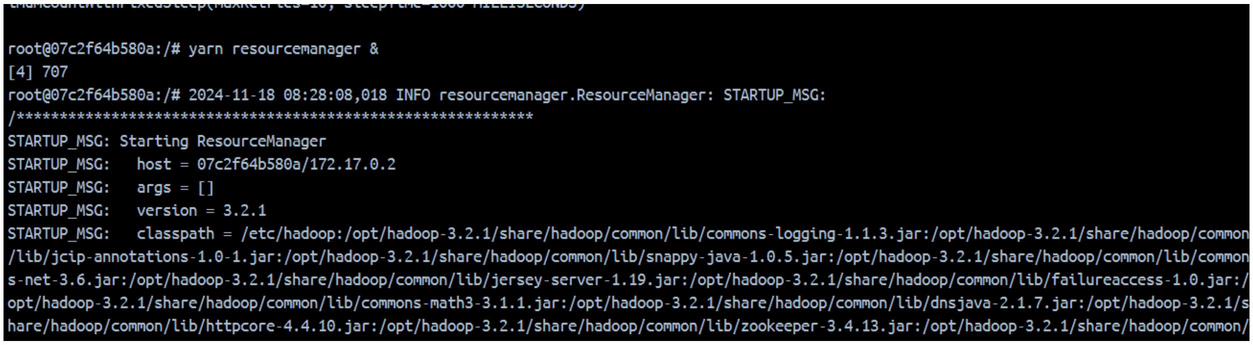
```
root@07c2f64b580a:/# hdfs datanode &
[2] 393
root@07c2f64b580a:/# 2024-11-18 08:24:48,494 INFO datanode.DataNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting DataNode
STARTUP_MSG:   host = 07c2f64b580a/172.17.0.2
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 3.2.1
STARTUP_MSG:   classpath = /etc/hadoop:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-logging-1.1.3.jar:/opt/hadoop-3.2.1/share/hadoop/common
/lib/jcip-annotations-1.0-1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/snappy-java-1.0.5.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/common
s-net-3.6.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-server-1.19.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/failureaccess-1.0.jar:/
opt/hadoop-3.2.1/share/hadoop/common/lib/commons-math3-3.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/dnsjava-2.1.7.jar:/opt/hadoop-3.2.1/s
hare/hadoop/common/lib/httpcore-4.4.10.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/zookeeper-3.4.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/
lib/jackson-annotations-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/woodstox-core-5.0.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/js
r311-api-1.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/audience-annotations-0.5.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-co
```

To start the YARN NodeManager process in the background:



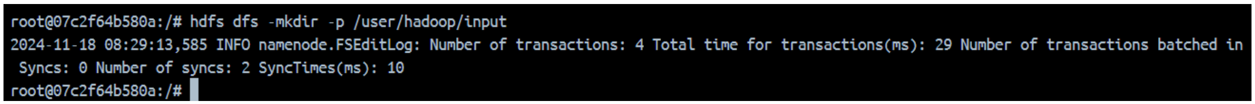
```
2024-11-18 08:24:51,745 INFO datanode.DataNode: Got finalize command for block pool BP-576603861-172.17.0.2-1731917994781

root@07c2f64b580a:/# yarn nodemanager &
[3] 533
root@07c2f64b580a:/# 2024-11-18 08:26:12,308 INFO nodemanager.NodeManager: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NodeManager
STARTUP_MSG:   host = 07c2f64b580a/172.17.0.2
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 3.2.1
STARTUP_MSG:   classpath = /etc/hadoop:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-logging-1.1.3.jar:/opt/hadoop-3.2.1/share/hadoop/common
/lib/jcip-annotations-1.0-1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/snappy-java-1.0.5.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/common
s-net-3.6.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-server-1.19.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/failureaccess-1.0.jar:/
opt/hadoop-3.2.1/share/hadoop/common/lib/commons-math3-3.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/dnsjava-2.1.7.jar:/opt/hadoop-3.2.1/s
hare/hadoop/common/lib/httpcore-4.4.10.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/zookeeper-3.4.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/
lib/jackson-annotations-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/woodstox-core-5.0.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/js
```

To start the YARN ResourceManager process in the background:

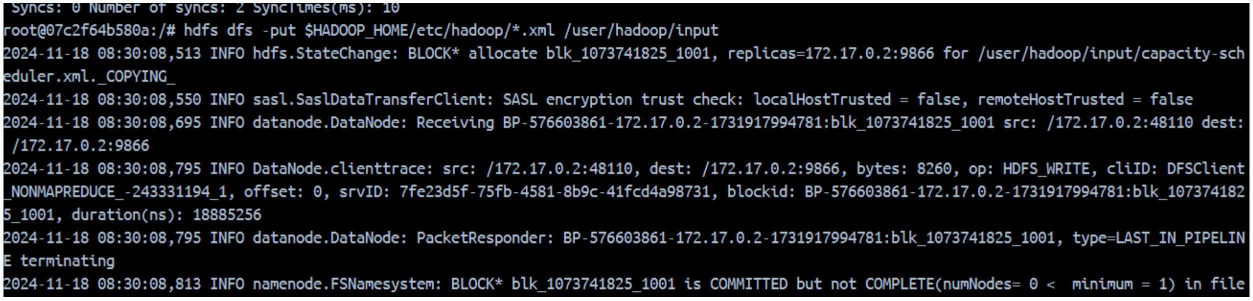


```
root@07c2f64b580a:/# yarn resourcemanager &
[4] 707
root@07c2f64b580a:/# 2024-11-18 08:28:08,018 INFO resourcemanager.ResourceManager: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting ResourceManager
STARTUP_MSG:   host = 07c2f64b580a/172.17.0.2
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 3.2.1
STARTUP_MSG:   classpath = /etc/hadoop:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-logging-1.1.3.jar:/opt/hadoop-3.2.1/share/hadoop/common
/lib/jcip-annotations-1.0-1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/snappy-java-1.0.5.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/common
s-net-3.6.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-server-1.19.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/failureaccess-1.0.jar:/
opt/hadoop-3.2.1/share/hadoop/common/lib/commons-math3-3.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/dnsjava-2.1.7.jar:/opt/hadoop-3.2.1/s
hare/hadoop/common/lib/httpcore-4.4.10.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/zookeeper-3.4.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/
```

To create an input directory and upload sample files to HDFS:

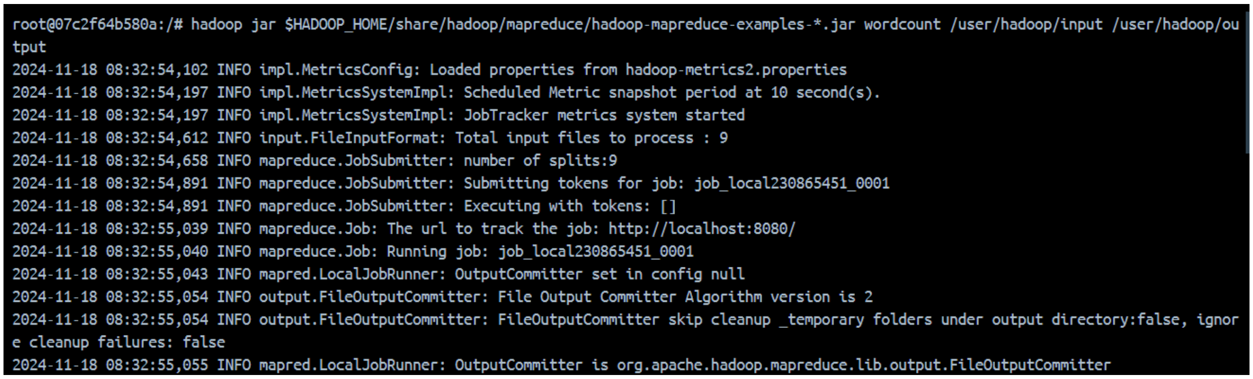


```
root@07c2f64b580a:/# hdfs dfs -mkdir -p /user/hadoop/input
2024-11-18 08:29:13,585 INFO namenode.FSEditLog: Number of transactions: 4 Total time for transactions(ms): 29 Number of transactions batched in
 Syncs: 0 Number of syncs: 2 SyncTimes(ms): 10
root@07c2f64b580a:/#
```



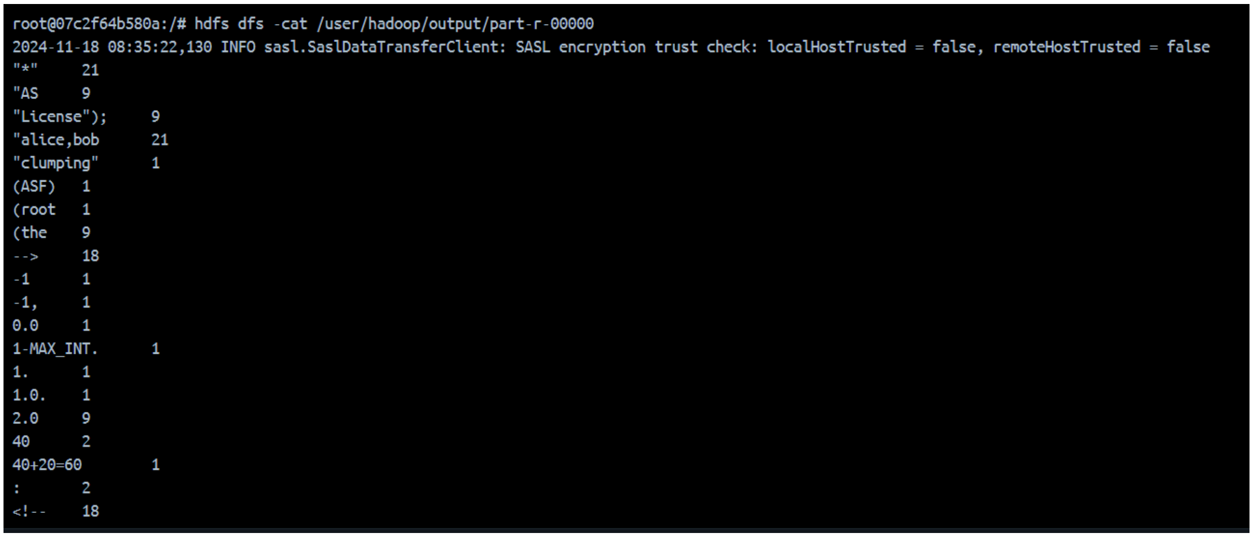
```
 Syncs: 0 Number of syncs: 2 SyncTimes(ms): 10
root@07c2f64b580a:/# hdfs dfs -put $HADOOP_HOME/etc/hadoop/*.xml /user/hadoop/input
2024-11-18 08:30:08,513 INFO hdfs.StateChange: BLOCK* allocate blk_1073741825_1001, replicas=172.17.0.2:9866 for /user/hadoop/input/capacity-sch
eduler.xml._COPYING_
2024-11-18 08:30:08,550 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 08:30:08,695 INFO datanode.DataNode: Receiving BP-576603861-172.17.0.2-1731917994781:blk_1073741825_1001 src: /172.17.0.2:48110 dest:
 /172.17.0.2:9866
2024-11-18 08:30:08,795 INFO DataNode.clienttrace: src: /172.17.0.2:48110, dest: /172.17.0.2:9866, bytes: 8260, op: HDFS_WRITE, cliID: DFSClient
_NONMAPREDUCE_-243331194_1, offset: 0, srvID: 7fe23d5f-75fb-4581-8b9c-41fcd4a98731, blockid: BP-576603861-172.17.0.2-1731917994781:blk_107374182
5_1001, duration(ns): 18885256
2024-11-18 08:30:08,795 INFO datanode.DataNode: PacketResponder: BP-576603861-172.17.0.2-1731917994781:blk_1073741825_1001, type=LAST_IN_PIPELIN
E terminating
2024-11-18 08:30:08,813 INFO namenode.FSNamesystem: BLOCK* blk_1073741825_1001 is COMMITTED but not COMPLETE(numNodes= 0 <  minimum = 1) in file
```

To create an input directory and upload sample files to HDFS:



```
root@07c2f64b580a:/# hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar wordcount /user/hadoop/input /user/hadoop/ou
tput
2024-11-18 08:32:54,102 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-11-18 08:32:54,197 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-11-18 08:32:54,197 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-11-18 08:32:54,612 INFO input.FileInputFormat: Total input files to process : 9
2024-11-18 08:32:54,658 INFO mapreduce.JobSubmitter: number of splits:9
2024-11-18 08:32:54,891 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local230865451_0001
2024-11-18 08:32:54,891 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-11-18 08:32:55,039 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-11-18 08:32:55,040 INFO mapreduce.Job: Running job: job_local230865451_0001
2024-11-18 08:32:55,043 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-11-18 08:32:55,054 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-18 08:32:55,054 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignor
e cleanup failures: false
2024-11-18 08:32:55,055 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

To display the results of the word count job:



```
root@07c2f64b580a:/# hdfs dfs -cat /user/hadoop/output/part-r-00000
2024-11-18 08:35:22,130 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
"*"     21
"AS     9
"License");     9
"alice,bob      21
"clumping"      1
(ASF)   1
(root   1
(the    9
-->     18
-1      1
-1,     1
0.0     1
1-MAX_INT.      1
1.      1
1.0.    1
2.0     9
40      2
40+20=60        1
:       2
<!--    18
```