

Intermediate

```
root@p930026096:~/docker-hadoop-master# docker-compose up -d
resourcemanager is up-to-date
nodemanager is up-to-date
datanode is up-to-date
namenode is up-to-date
historyserver is up-to-date
```

Start the docker container. Here the docker is ready.

The screenshot shows a web browser window titled 'Namenode information — Mozilla Firefox'. The address bar shows 'localhost:9870/dfshealth.html#tab-overview'. The page has a green navigation bar with tabs: 'Hadoop', 'Overview', 'Datanodes', 'Datanode Volume Failures', 'Snapshot', 'Startup Progress', and 'Utilities'. The 'Overview' tab is selected, showing 'Overview 'namenode:9000' (active)'. Below this is a table with the following information:

Started:	Thu Oct 21 00:13:00 +0000 2021
Version:	3.2.1, rb3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled:	Tue Sep 10 15:56:00 +0000 2019 by rohitsharmaks from branch-3.2.1
Cluster ID:	CID-8a930d7f-79c4-423a-b707-ed4154e10203
Block Pool ID:	BP-862717634-172.18.0.5-1634745176705

Below the table, there is a 'Summary' section which is partially visible.

Namenode is running by checking the port 9870.

```
root@p930026096:~/docker-hadoop-master# docker container ls
CONTAINER ID   IMAGE                                     PORTS          NAMES
82ad8e40aa5d   bde2020/hadoop-namenode:2.0.0-hadoop3.2.1-java8   0.0.0.0:9000->9000/tcp, :::9000->9000/tcp, 0.0.0.0:9870->9870/tcp, :::9870->9870/tcp   namenode
744cfcfd981    bde2020/hadoop-nodemanager:2.0.0-hadoop3.2.1-java8   8042/tcp      nodemanager
36f319e4d32f   bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8     9864/tcp      datanode
e22b3f43542e   bde2020/hadoop-historyserver:2.0.0-hadoop3.2.1-java8   8188/tcp      historyserver
87d4e242f9a6   bde2020/hadoop-resourcemanager:2.0.0-hadoop3.2.1-java8   8088/tcp      resourcemanager
```

Find out the ID of my own namenode.

```
root@p930026096:~/docker-hadoop-master# docker cp ~/Downloads/books/. 82ad8e40aa5d:/input/
```

Copy the downloaded .txt file from books to the docker Hadoop cluster.

```
root@p930026096:~/docker-hadoop-master# docker exec -it namenode bash
```

Execute the namenode bash.

```
root@82ad8e40aa5d:/# hdfs dfs -put ./input/* input-royma
```

Put all the .txt file under the directory, inpyt-royma. Here you can check whether all the .txt files had been copied to the right path by checking the file information on the port 9870.

```
root@7fd5d0089d8a:/# hadoop jar wc.jar org.apache.hadoop.examples.WordCount input-royma output-book
```

Run the WordCount.java file to do the map reduce.

```
pes.xml'.
2021-10-22 12:04:55,010 INFO impl.YarnClientImpl: Submitted application application_1634900922307_0002
2021-10-22 12:04:55,047 INFO mapreduce.Job: The url to track the job: http://resourcemanager:8088/proxy/application_1634900922307_0002/
2021-10-22 12:04:55,047 INFO mapreduce.Job: Running job: job_1634900922307_0002
2021-10-22 12:05:01,183 INFO mapreduce.Job: Job job_1634900922307_0002 running in uber mode : false
2021-10-22 12:05:01,184 INFO mapreduce.Job: map 0% reduce 0%
2021-10-22 12:05:09,259 INFO mapreduce.Job: map 1% reduce 0%
2021-10-22 12:05:10,264 INFO mapreduce.Job: map 2% reduce 0%
2021-10-22 12:05:13,279 INFO mapreduce.Job: map 3% reduce 0%
2021-10-22 12:05:15,288 INFO mapreduce.Job: map 4% reduce 0%
2021-10-22 12:05:18,303 INFO mapreduce.Job: map 5% reduce 0%
2021-10-22 12:05:20,312 INFO mapreduce.Job: map 6% reduce 0%
2021-10-22 12:05:23,324 INFO mapreduce.Job: map 7% reduce 0%
2021-10-22 12:05:25,333 INFO mapreduce.Job: map 8% reduce 0%
2021-10-22 12:05:28,348 INFO mapreduce.Job: map 9% reduce 0%
2021-10-22 12:05:30,356 INFO mapreduce.Job: map 10% reduce 0%
2021-10-22 12:05:32,365 INFO mapreduce.Job: map 11% reduce 0%
2021-10-22 12:05:34,374 INFO mapreduce.Job: map 12% reduce 0%
2021-10-22 12:05:36,383 INFO mapreduce.Job: map 13% reduce 0%
2021-10-22 12:05:38,391 INFO mapreduce.Job: map 14% reduce 0%
2021-10-22 12:05:41,405 INFO mapreduce.Job: map 15% reduce 0%
2021-10-22 12:05:43,414 INFO mapreduce.Job: map 16% reduce 0%
```

```

2021-10-22 12:08:30,207 INFO mapreduce.Job: map 100% reduce 0%
2021-10-22 12:08:36,222 INFO mapreduce.Job: map 100% reduce 100%
2021-10-22 12:08:37,230 INFO mapreduce.Job: Job job_1634900922307_0002 completed
successfully
2021-10-22 12:08:37,355 INFO mapreduce.Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=4736533
    FILE: Number of bytes written=50239736
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=111020101
    HDFS: Number of bytes written=6869854
    HDFS: Number of read operations=452
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Launched map tasks=149
    Launched reduce tasks=1
    Rack-local map tasks=149
    Total time spent by all maps in occupied slots (ms)=1327660
    Total time spent by all reduces in occupied slots (ms)=44664
    Total time spent by all map tasks (ms)=331915
    Total time spent by all reduce tasks (ms)=5583

```

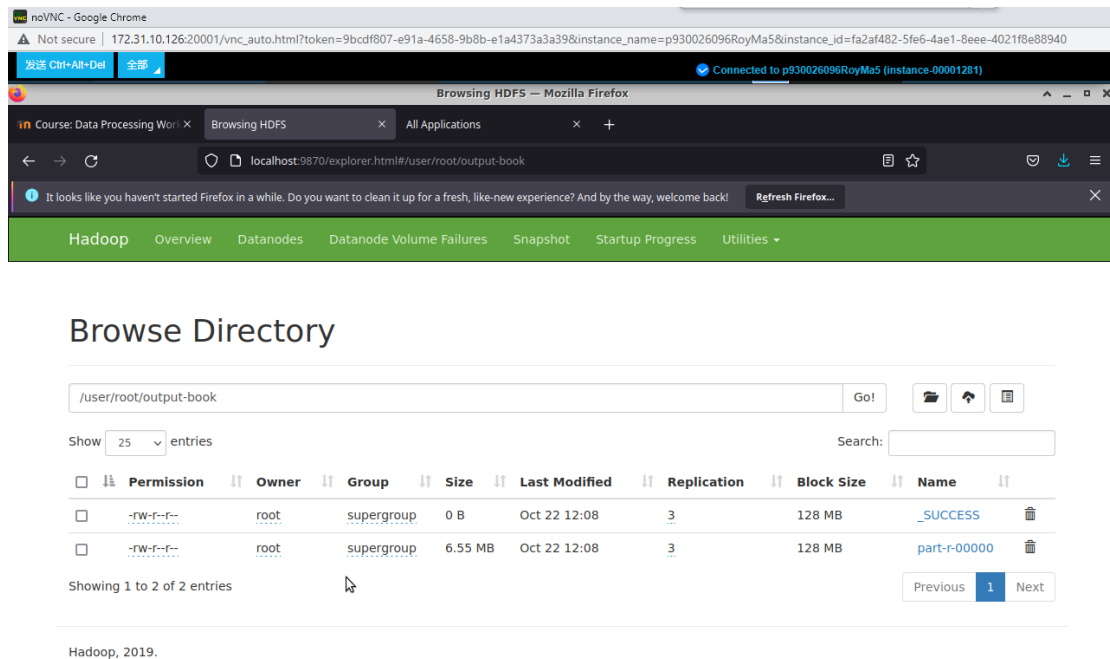
It takes about 3-5 minutes, since the data is huge. The message continuously shows the progress of map reduce.

```

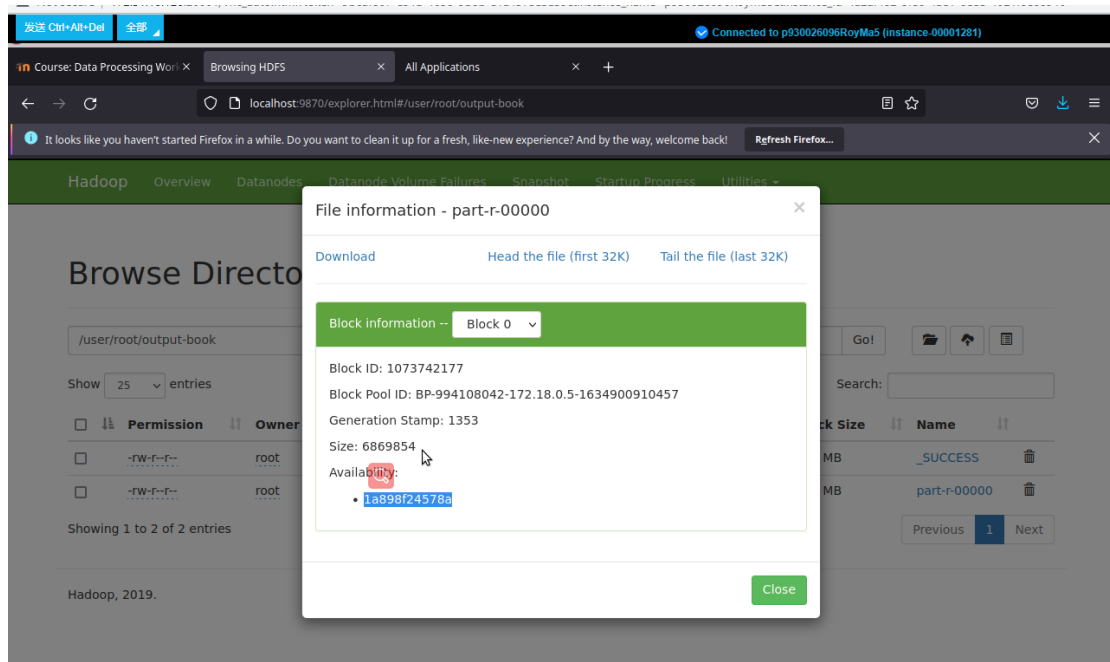
  Total megabyte-milliseconds taken by all map tasks=1359523840
  Total megabyte-milliseconds taken by all reduce tasks=45735936
Map-Reduce Framework
  Map input records=2196338
  Map output records=19110138
  Map output bytes=184173719
  Map output materialized bytes=11105220
  Input split bytes=17343
  Combine input records=19110138
  Combine output records=2354509
  Reduce input groups=550427
  Reduce shuffle bytes=11105220
  Reduce input records=2354509
  Reduce output records=550427
  Spilled Records=4709018
  Shuffled Maps =149
  Failed Shuffles=0
  Merged Map outputs=149
  GC time elapsed (ms)=9119
  CPU time spent (ms)=217110
  Physical memory (bytes) snapshot=54525767680
  Virtual memory (bytes) snapshot=769437564928
  Total committed heap usage (bytes)=65790279680
  Peak Map Physical memory (bytes)=607965184
  Peak Map Virtual memory (bytes)=5112524800

```

Finish map reduce.



By checking port 9870, output files are under output-book directory.



Also download to your host computer.

Finally, use the “exit” and “docker-compose down” commands to leave the container and shut down the cluster safely as the last two steps in basic.