

docker pull bde2020/hadoop-datanode:latest

docker pull bde2020/hadoop-historyserver

```
C:\Users\LENOVO>docker pull bde2020/hadoop-datanode:latest
latest: Pulling from bde2020/hadoop-datanode
Digest: sha256:35f899bcbe9f983825a8a3bdc135ed0e8e0eaf3b58f9b08bf257b5e86bae3b47
Status: Image is up to date for bde2020/hadoop-datanode:latest
docker.io/bde2020/hadoop-datanode:latest

What's next:
  View a summary of image vulnerabilities and recommendations → docker scout quickview bde2020/hadoop-datanode:latest

C:\Users\LENOVO>docker pull bde2020/hadoop-historyserver
Using default tag: latest
latest: Pulling from bde2020/hadoop-historyserver
Digest: sha256:216100a96a73717006031ff0c8b72effdc7acffca0a6c647f8820cb7eabc81fd
Status: Image is up to date for bde2020/hadoop-historyserver:latest
docker.io/bde2020/hadoop-historyserver:latest

What's next:
  View a summary of image vulnerabilities and recommendations → docker scout quickview bde2020/hadoop-historyserver

C:\Users\LENOVO>
```

docker volume create namenode

docker volume create datanode

```
C:\Users\LENOVO>docker volume create namenode
namenode

C:\Users\LENOVO>docker volume create datanode
datanode
```

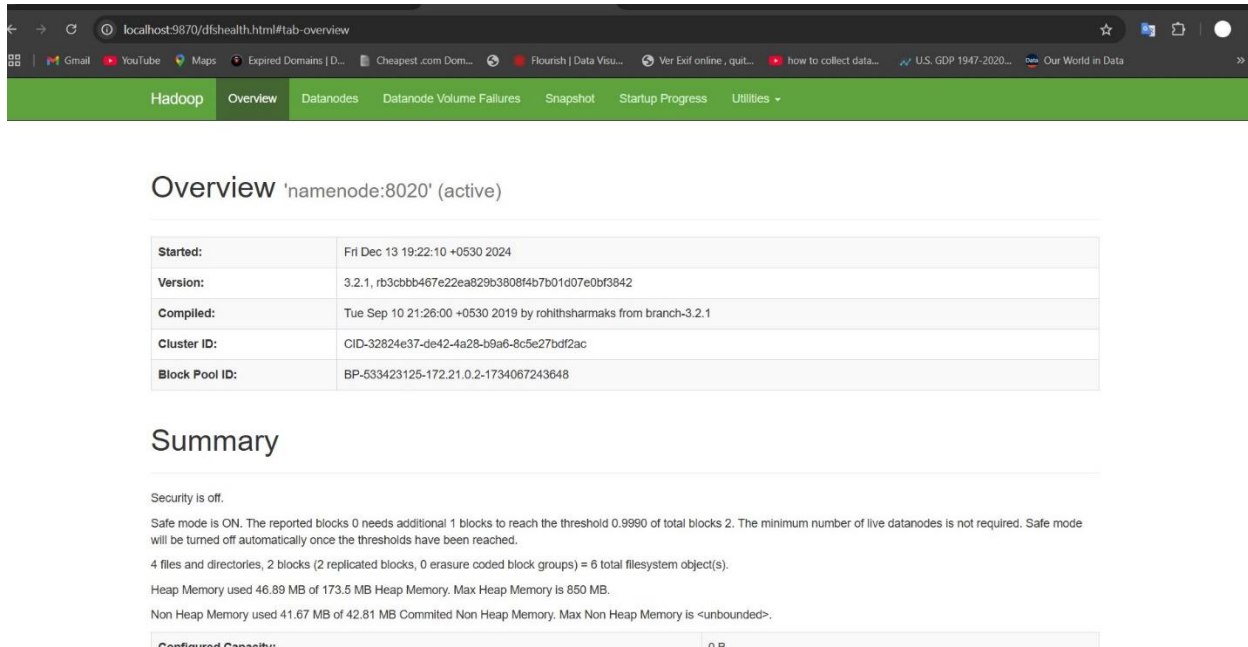
docker-compose up -d

```
11     - "9870:9870"
12     - "9000:9000"
13     volumes:
14     - namenode-data:/hadoop/dfs/name
15

PROBLEMS  DEBUG CONSOLE  OUTPUT  TERMINAL  PORTS
[+] Running 3/3
✔ Container namenode      Started          48.0s
✔ Container datanode      Started          55.7s
✔ Container historyserver Started          67.4s
• PS C:\sabaragumwa 6th\Single_Node_Hadoop_Exercise> docker exec -it namenode hdfs dfs -mkdir -p /input

What's next:
  Try Docker Debug for seamless, persistent debugging tools in any container or image → docker debug namenode
  Learn more at https://docs.docker.com/go/debug-cli/
○ PS C:\sabaragumwa 6th\Single_Node_Hadoop_Exercise>
```

opening <http://localhost:9870>



Overview 'namenode:8020' (active)

Started:	Fri Dec 13 19:22:10 +0530 2024
Version:	3.2.1, rb3cbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled:	Tue Sep 10 21:26:00 +0530 2019 by rohithsharmaks from branch-3.2.1
Cluster ID:	CID-32824e37-de42-4a28-b9a6-8c5e27bdf2ac
Block Pool ID:	BP-533423125-172.21.0.2-1734067243648

Summary

Security is off.

Safe mode is ON. The reported blocks 0 needs additional 1 blocks to reach the threshold 0.9990 of total blocks 2. The minimum number of live datanodes is not required. Safe mode will be turned off automatically once the thresholds have been reached.

4 files and directories, 2 blocks (2 replicated blocks, 0 erasure coded block groups) = 6 total filesystem object(s).

Heap Memory used 46.89 MB of 173.5 MB Heap Memory. Max Heap Memory is 850 MB.

Non Heap Memory used 41.67 MB of 42.81 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity: 0 B

```
docker exec -it namenode hdfs dfs -mkdir -p /input
```

```
docker cp "F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo\sample.txt" namenode:/sample.txt
```

```
docker exec -it namenode hdfs dfs -put /sample.txt /input
```



```
11 - "9870:9870"
12 - "9000:9000"
13 volumes:
14 - namenode-data:/hadoop/dfs/name
15

PROBLEMS  DEBUG CONSOLE  OUTPUT  TERMINAL  PORTS

PS C:\sabaragumawa 6th\Single_Node_Hadoop_Exercise> docker-compose up -d
time="2024-12-13T19:20:02+05:30" level=warning msg="C:\sabaragumawa 6th\Single_Node_Hadoop_Exercise\docker-compose.yml: the attribute 'version' is obsolete, it will be ignored, please remove it to avoid potential confusion"
[+] Running 3/3
✔ Container namenode Started 48.0s
✔ Container datanode Started 55.7s
✔ Container historyserver Started 67.4s

PS C:\sabaragumawa 6th\Single_Node_Hadoop_Exercise> docker exec -it namenode hdfs dfs -mkdir -p /input

What's next:
Try Docker Debug for seamless, persistent debugging tools in any container or image → docker debug namenode
Learn more at https://docs.docker.com/go/debug-cli/
PS C:\sabaragumawa 6th\Single_Node_Hadoop_Exercise>
```

Download the .jar file

<https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-mapreduce-examples/2.7.1/>

```
docker cp C:\Users\user\Downloads\hadoop-mapreduce-examples-2.7.1.jar namenode:/root/
```

`docker exec -it namenode hadoop jar /root/hadoop-mapreduce-examples-2.7.1.jar wordcount /input /output`

```
no such directory
PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo> docker cp C:\Users\user\Downloads\hadoop-mapreduce-examples-2.7.1.jar namenode:/root/
Successfully copied 274kB to namenode:/root/
PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo> docker exec -it namenode hadoop jar /root/hadoop-mapreduce-examples-2.7.1.jar wordcount /input /output
2024-11-25 08:46:36,962 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-11-25 08:46:37,010 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-11-25 08:46:37,010 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-11-25 08:46:37,203 INFO input.FileInputFormat: Total input files to process : 1
2024-11-25 08:46:37,222 INFO mapreduce.JobSubmitter: number of splits:1
2024-11-25 08:46:37,301 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1620145309_0001
2024-11-25 08:46:37,301 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-11-25 08:46:37,379 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-11-25 08:46:37,379 INFO mapreduce.Job: Running job: job_local1620145309_0001
2024-11-25 08:46:37,380 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-11-25 08:46:37,385 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-25 08:46:37,385 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-11-25 08:46:37,386 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-11-25 08:46:37,416 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-11-25 08:46:37,417 INFO mapred.LocalJobRunner: Starting task: attempt_local1620145309_0001_m_000000_0
2024-11-25 08:46:37,433 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-25 08:46:37,433 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-11-25 08:46:37,447 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2024-11-25 08:46:37,449 INFO mapred.MapTask: Processing split: hdfs://namenode:8020/input/sample.txt:0+27
2024-11-25 08:46:37,483 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2024-11-25 08:46:37,483 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2024-11-25 08:46:37,483 INFO mapred.MapTask: soft limit at 83886080
2024-11-25 08:46:37,483 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2024-11-25 08:46:37,483 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2024-11-25 08:46:37,487 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2024-11-25 08:46:37,516 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2024-11-25 08:46:37,583 INFO mapred.LocalJobRunner:
2024-11-25 08:46:37,585 INFO mapred.MapTask: Starting flush of map output
2024-11-25 08:46:37,585 INFO mapred.MapTask: Spilling map output
2024-11-25 08:46:37,585 INFO mapred.MapTask: bufstart = 0; bufend = 48; bufvoid = 104857600
```

`http://localhost:8088` run you should add the

`docker-compose.yaml` update

```
resourcemanager:
  image: bde2020/hadoop-resourcemanager:latest
  container_name: resourcemanager
  environment:
    - CORE_CONF_fs_defaultFS=hdfs://namenode:8020
  ports:
    - "8088:8088"
```

Then we can run 8088 port

Nodes of the cluster

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved
0	0	0	0	0	0 B	0 B	0 B	0	0	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
0	0	0	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
Capacity Scheduler	[memory-mb (unit=M), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0

Showing 0 to 0 of 0 entries

No data available in table

docker exec -it namenode hdfs dfs -ls /output

```
PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo> docker exec -it namenode hdfs dfs -ls /output2
Found 2 items
-rw-r--r-- 3 root supergroup 0 2024-11-25 08:58 /output2/_SUCCESS
-rw-r--r-- 3 root supergroup 38 2024-11-25 08:58 /output2/part-r-00000
```

docker exec -it namenode hdfs dfs -ls /output

docker exec -it namenode hdfs dfs -cat /output/part-r-00000

```
PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo> docker exec -it namenode hdfs dfs -ls /output2
Found 2 items
-rw-r--r-- 3 root supergroup 0 2024-11-25 08:58 /output2/_SUCCESS
-rw-r--r-- 3 root supergroup 38 2024-11-25 08:58 /output2/part-r-00000
PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo> docker exec -it namenode hdfs dfs -cat /output/part-r-00000
2024-11-25 09:03:28,371 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
This 1
for 1
hadoop 1
sample 1
text 1
```

docker-compose down