

bde2020/hadoop-namenode

```
What's next:
  View a summary of image vulnerabilities and recommendations → docker scout quickview bde2020/hadoop-namenode
PS C:\Users\warun> docker pull bde2020/hadoop-namenode
Status: Image is up to date for bde2020/hadoop-namenode:latest
docker.io/bde2020/hadoop-namenode:latest

What's next:
  View a summary of image vulnerabilities and recommendations → docker scout quickview bde2020/hadoop-namenode
PS C:\Users\warun> |
```

bde2020/hadoop-datanode

```
C:\Users\warun>docker pull bde2020/hadoop-datanode:latest
latest: Pulling from bde2020/hadoop-datanode
5e185246c615: Download complete
4bf0ae3d5cc8: Download complete
b91d0b0b68c8: Download complete
Digest: sha256:35f899bcbe9f983825a8a3bdc135ed0e8e0eaf3b58f9b08bf257b5e86bae3b47
Status: Downloaded newer image 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\warun>|
```

docker pull bde2020/hadoop-historyserver

```
C:\Users\warun>docker pull bde2020/hadoop-historyserver
Using default tag: latest
latest: Pulling from bde2020/hadoop-historyserver
f3f6b02c1935: Download complete
78d381637ee0: Download complete
84560426d8fd: Download complete
Digest: sha256:216100a96a73717006031ff0c8b72effdc7acffca0a6c647f8820cb7eabc81fd
Status: Downloaded newer image 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\warun>|
```

docker volume create namenode

docker volume create namenode

```
PS C:\Users\warun\OneDrive\Desktop\DOCKER> docker-compose up -d
[+] Running 2/3
  - Container namenode      Starting      0.5s
  ✓ Container datanode      Created      0.1s
  ✓ Container historyserver Created      0.1s
Error response from daemon: Ports are not available: exposing port TCP 0.0.0.0:9870 -> 0.0.0.0:0: listen tcp 0.0.0.0:9870: bind: An attempt was made to access a socket in a way forbidden by its access permissions.
PS C:\Users\warun\OneDrive\Desktop\DOCKER> |
```

[illegible]

opening <http://localhost:9870>

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

Overview 'namenode:8020' (active)

Started:	Tue Nov 26 08:48:03 +0530 2024
Version:	3.2.1, rb3cbbb467e22ea829b3808f4b7b01d07e0fb3842
Compiled:	Tue Sep 10 21:26:00 +0530 2019 by rohitsharmaks from branch-3.2.1
Cluster ID:	CID-4e0f7aab-0b8c-4458-a9a0-70cc9ab0ccab
Block Pool ID:	BP-748252151-172.19.0.2-1732591082379

Summary

Security is off.

Safemode is off.

1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

Heap Memory used 150.3 MB of 240 MB Heap Memory. Max Heap Memory is 1.7 GB.

Non Heap Memory used 45.48 MB of 46.81 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	1006.85 GB
Configured Remote Capacity:	0 B

Non Heap Memory used 45.48 MB of 46.81 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	1006.85 GB
Configured Remote Capacity:	0 B
DFS Used:	24 KB (0%)
Non DFS Used:	3.43 GB
DFS Remaining:	952.21 GB (94.57%)
Block Pool Used:	24 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Tue Nov 26 08:48:03 +0530 2024
Last Checkpoint Time	Tue Nov 26 08:48:02 +0530 2024
Enabled Erasure Coding Policies	RS-6-3-1024k

NAME Node Command List

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

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

```
✓ Container historyserver Started 1.2s
PS C:\Users\warun\OneDrive\Desktop\DOCKER> 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:\Users\warun\OneDrive\Desktop\DOCKER> docker exec -it namenode hdfs dfs -put /path/to/sample.txt /input
>>
put: '/path/to/sample.txt': No such file or directory

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:\Users\warun\OneDrive\Desktop\DOCKER> 
```

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

```
PS C:\Users\warun\OneDrive\Desktop\DOCKER> docker exec -it namenode hdfs dfs -mkdir /input
mkdir: '/input': File exists

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:\Users\warun\OneDrive\Desktop\DOCKER> docker cp "C:\Users\warun\OneDrive\Desktop\DOCKER\sample.txt"
>> namenode:/sample.txt
"docker cp" requires exactly 2 arguments.
See 'docker cp --help'.

Usage: docker cp [OPTIONS] CONTAINER:SRC_PATH DEST_PATH|-
       docker cp [OPTIONS] SRC_PATH|- CONTAINER:DEST_PATH
```

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

```
PS C:\Users\warun\OneDrive\Desktop\DOCKER> docker cp "C:\Users\warun\OneDrive\Desktop\DOCKER\sample.txt" namenode:/sample.txt
>>
Successfully copied 1.54kB to namenode:/sample.txt
```

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

```
PS C:\Users\warun\OneDrive\Desktop\DOCKER> docker exec -it namenode hdfs dfs -put /sample.txt /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/
```

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

"docker cp" requires exactly 2 arguments.
See 'docker cp --help'.

Usage: docker cp [OPTIONS] CONTAINER:SRC_PATH DEST_PATH|-
       docker cp [OPTIONS] SRC_PATH|- CONTAINER:DEST_PATH

Copy files/folders between a container and the local filesystem
```

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

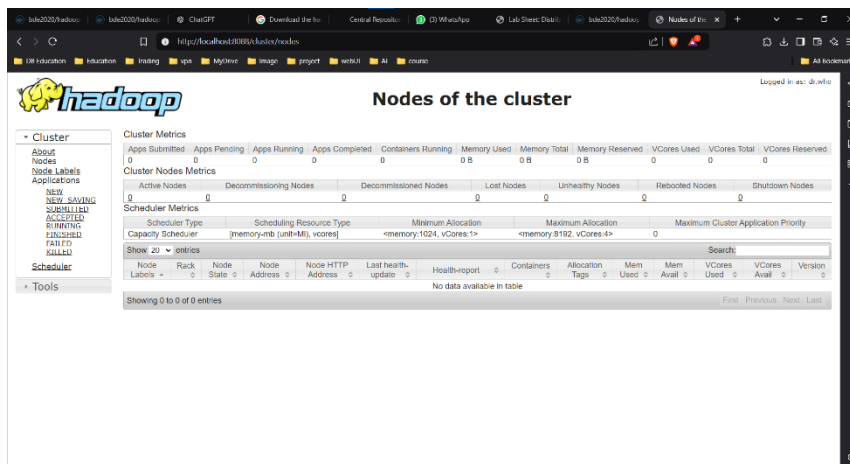
docker-compose.yaml update

```

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



`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

```