

docker pull bde2020/hadoop-datanode:latest

docker pull bde2020/hadoop-historyserver

```
Command Prompt
Microsoft Windows [Version 10.0.19045.5131]
(c) Microsoft Corporation. All rights reserved.

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

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

C:\Users\user>
```

docker volume create namenode

docker volume create namenode

```
Command Prompt
Microsoft Windows [Version 10.0.19045.5131]
(c) Microsoft Corporation. All rights reserved.

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

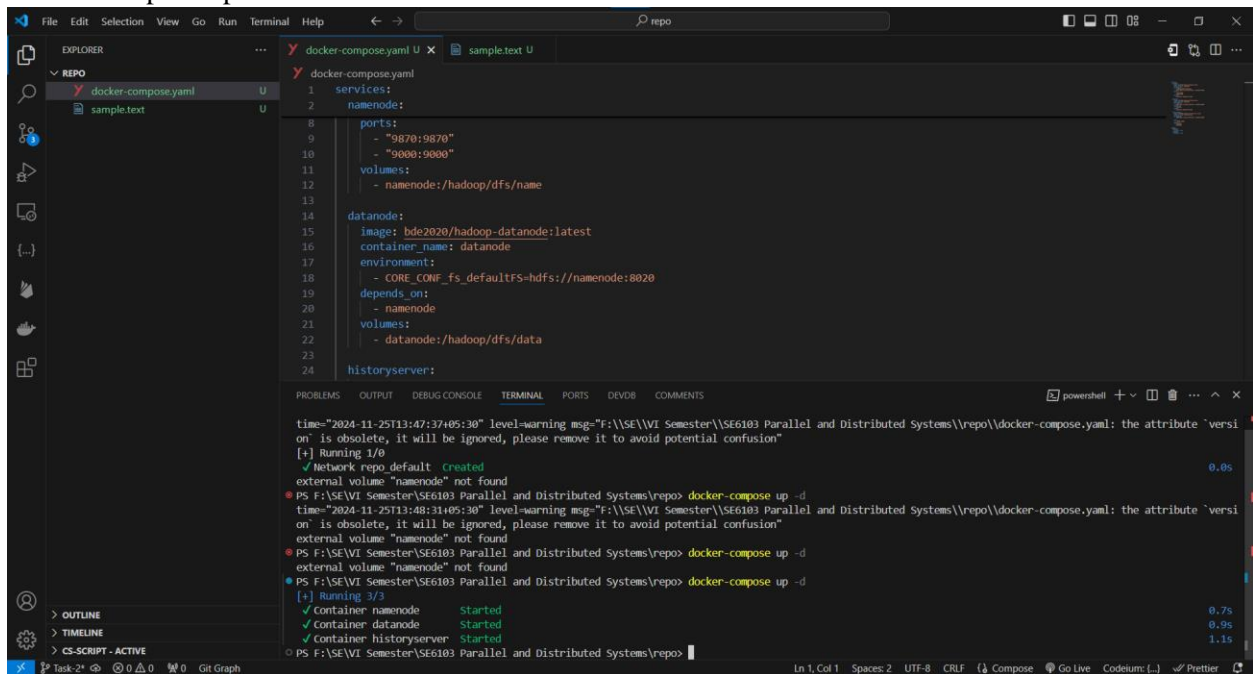
C:\Users\user>docker pull bde2020/hadoop-historyserver
Using default tag: latest
latest: Pulling from bde2020/hadoop-historyserver
f3f6b02c1935: Download complete
84560426d8fd: Download complete
78d381637ee0: Download complete
Digest: sha256:216100a96a73717006031ff0c8b72effdc7acffca0a6c647f8820cb7eabc81fd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

C:\Users\user>docker volume create namenode
namenode

C:\Users\user>docker volume create datanode
datanode

C:\Users\user>
```

docker-compose up -d



The screenshot shows a Visual Studio Code editor with a Docker Compose file named `docker-compose.yml` open. The file defines three services: `namenode`, `datanode`, and `historyserver`. The `namenode` service is configured with ports 9870 and 9000, and a volume `namenode` mapped to `/hadoop/dfs/name`. The `datanode` service uses the `bde2020/hadoop-datanode:latest` image, depends on `namenode`, and has a volume `datanode` mapped to `/hadoop/dfs/data`. The `historyserver` service is also defined. The terminal output shows the command `docker-compose up -d` being executed, with warnings about the `version` attribute being obsolete. The output indicates that the containers `namenode`, `datanode`, and `historyserver` were successfully started.

```
1 services:
2   namenode:
3
4   ports:
5     - "9870:9870"
6     - "9000:9000"
7   volumes:
8     - namenode:/hadoop/dfs/name
9
10  datanode:
11    image: bde2020/hadoop-datanode:latest
12    container_name: datanode
13    environment:
14      - CORE_CONF_fs_defaultFS=hdfs://namenode:8020
15    depends_on:
16      - namenode
17    volumes:
18      - datanode:/hadoop/dfs/data
19
20  historyserver:
```

time="2024-11-25T13:47:37+05:30" level=warning msg="F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo\docker-compose.yml: the attribute 'version' is obsolete, it will be ignored, please remove it to avoid potential confusion"

[+] Running 1/0

✓ Network repo default Created 0.0s

external volume "namenode" not found

PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo> docker-compose up -d

time="2024-11-25T13:48:31+05:30" level=warning msg="F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo\docker-compose.yml: the attribute 'version' is obsolete, it will be ignored, please remove it to avoid potential confusion"

external volume "namenode" not found

PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo> docker-compose up -d

external volume "namenode" not found

PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo> docker-compose up -d

[+] Running 3/3

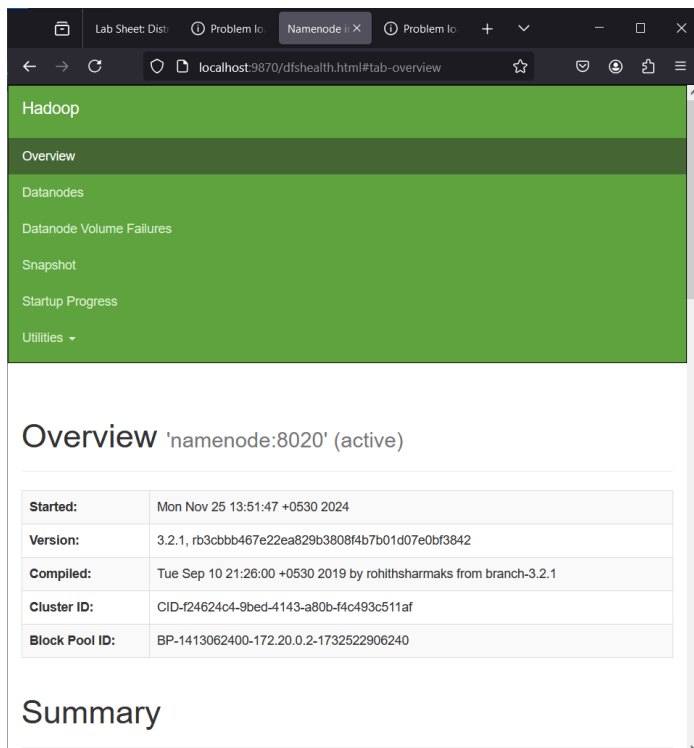
✓ Container namenode Started 0.7s

✓ Container datanode Started 0.9s

✓ Container historyserver Started 1.1s

PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo>

opening <http://localhost:9870>



The screenshot shows the Hadoop Namenode web interface. The top navigation bar includes links for Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The main content area displays the Overview for the 'namenode:8020' (active) instance. A table provides details about the instance, including its start time, version, compile date, cluster ID, and block pool ID.

Property	Value
Started:	Mon Nov 25 13:51:47 +0530 2024
Version:	3.2.1, rb3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled:	Tue Sep 10 21:26:00 +0530 2019 by rohitsharmaks from branch-3.2.1
Cluster ID:	CID-f24624c4-9bed-4143-a80b-f4c493c511af
Block Pool ID:	BP-1413062400-172.20.0.2-1732522906240

Summary

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

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

The screenshot shows a VS Code terminal window with the following content:

```
Semester/SE6103 Parallel and Distributed Systems/repo/sample.txt"

File "f:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo\sample.txt", line 1
    This sample text for hadoop
    ^^^^^^
SyntaxError: invalid syntax
PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo
> docker exec -it namenode hdfs dfs -mkdir -p /input
PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo
> docker exec -it namenode hdfs dfs -put /path/to/sample.txt /input
put: `/path/to/sample.txt': No such file or directory
PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo
>
```

The left sidebar shows the Explorer view with a file named "sample.txt" selected. The bottom status bar shows "Task-2*", "0 0 0", "Git Graph", "Go Live", and "Codeium: {...}".

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

The screenshot shows a VS Code terminal window with the following content:

```
PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo> docker exec -it namenode hdfs dfs -put "F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo\sample.txt" /input
put: No FileSystem for scheme "F"
PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo> docker exec -it namenode hdfs dfs -ls /input
PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo> docker cp "F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo\sample.txt" namenode:/sample.txt
Successfully copied 2.05kB to namenode:/sample.txt
PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo> docker exec -it namenode hdfs dfs -put /sample.txt /input
2024-11-25 08:38:08,107 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo> docker exec -it namenode hdfs dfs -ls /input
Found 1 items
-rw-r--r-- 3 root supergroup 27 2024-11-25 08:38 /input/sample.txt
PS F:\SE\VI Semester\SE6103 Parallel and Distributed Systems\repo>
```

The left sidebar shows the Explorer view with a file named "sample.txt" selected. The bottom status bar shows "Task-2*", "0 0 0", "Git Graph", "Go Live", and "Codeium: {...}".

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

`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`