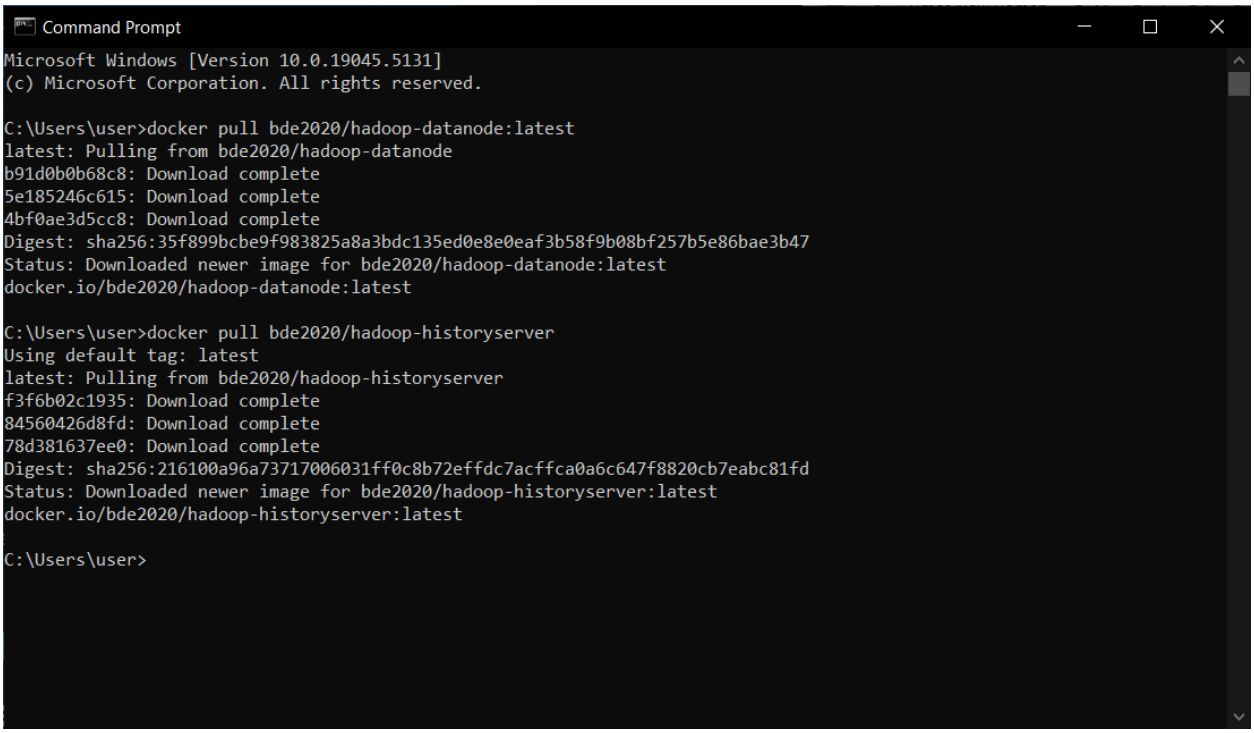
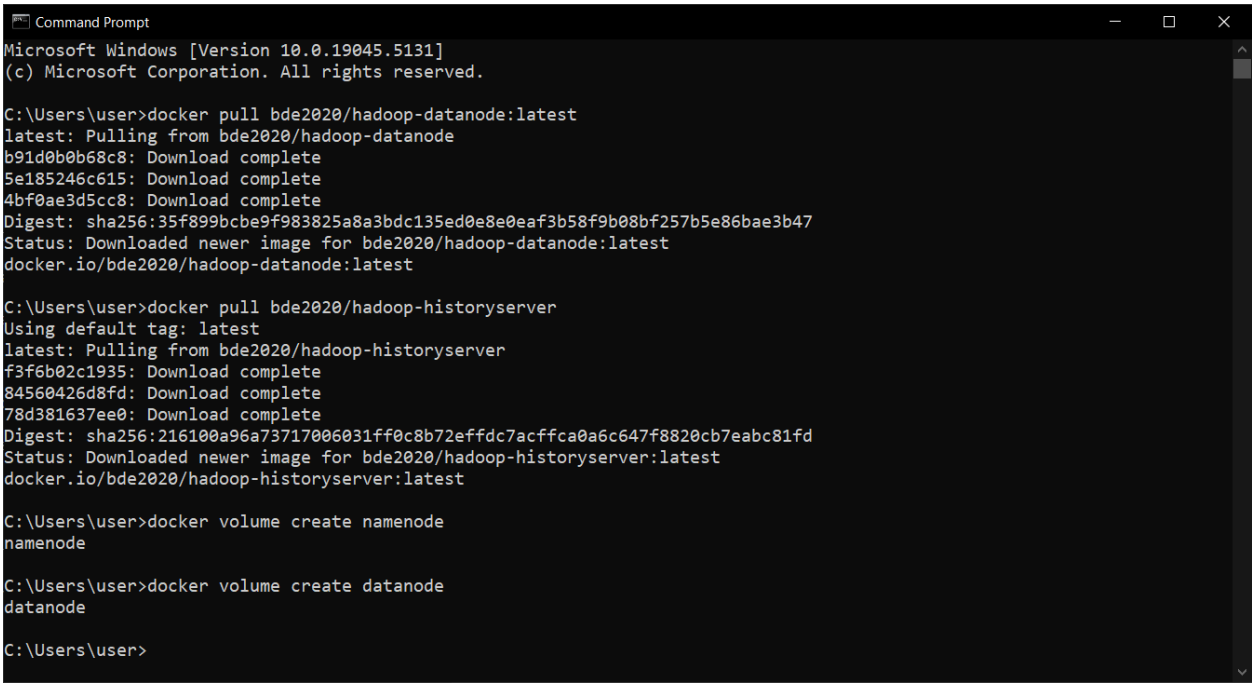
docker pull bde2020/hadoop-datanode:latest

docker pull bde2020/hadoop-historyserver

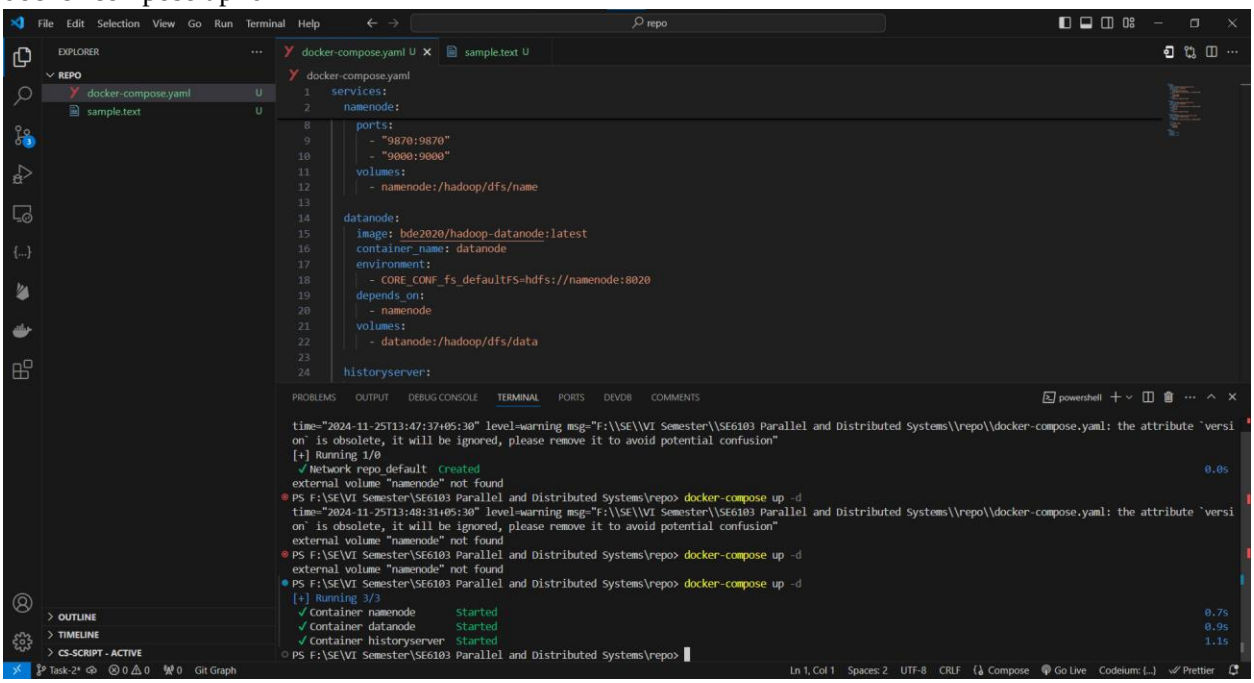


docker volume create namenode

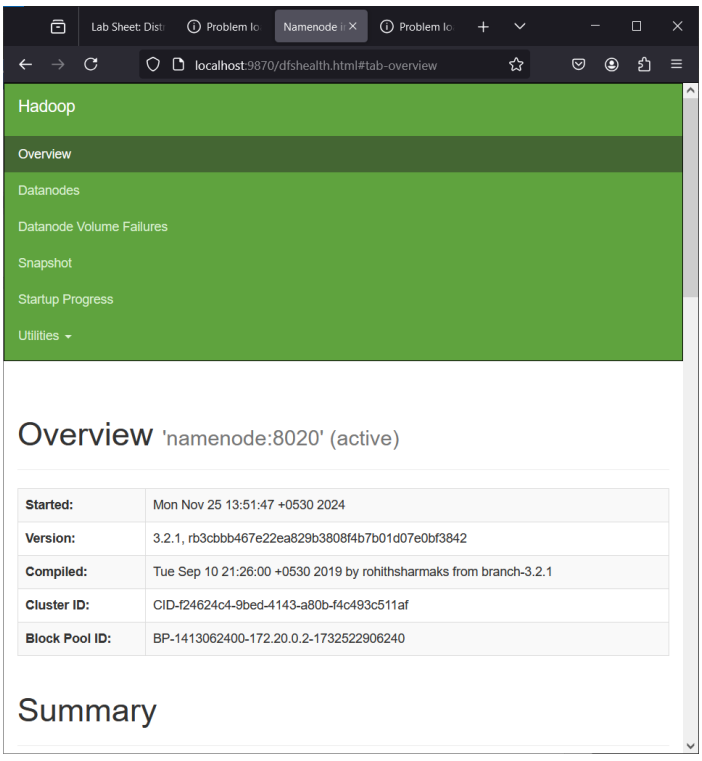
docker volume create namenode



docker-compose up -d

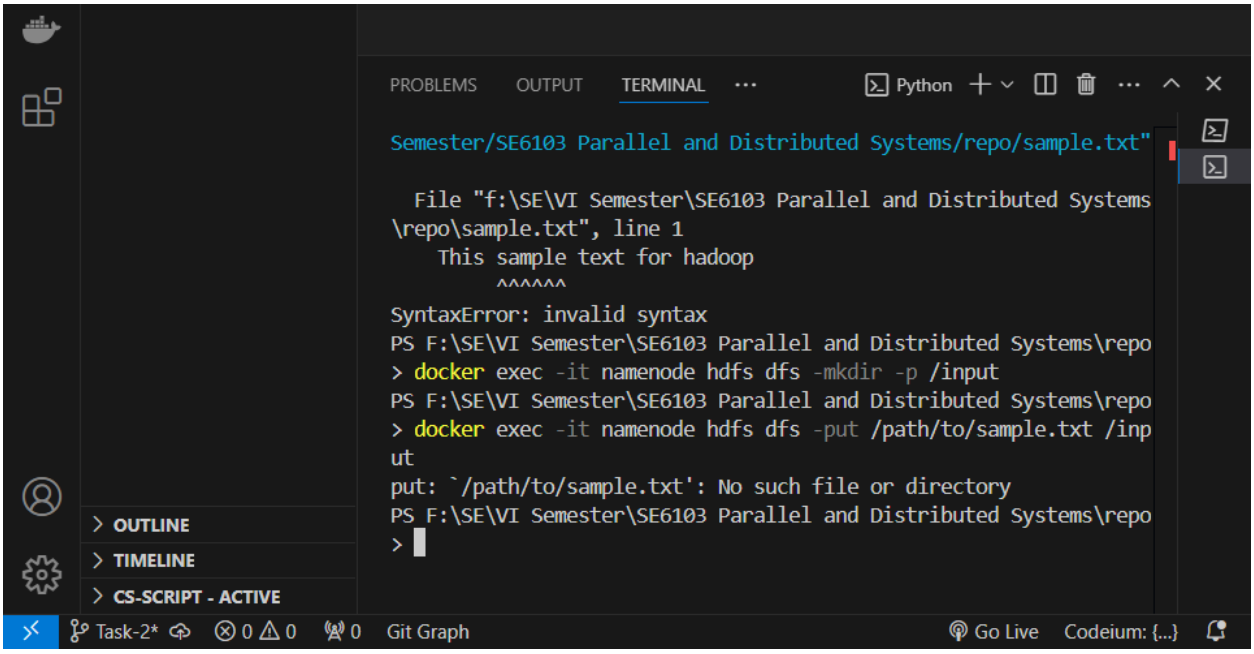


opening http://localhost:9870



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

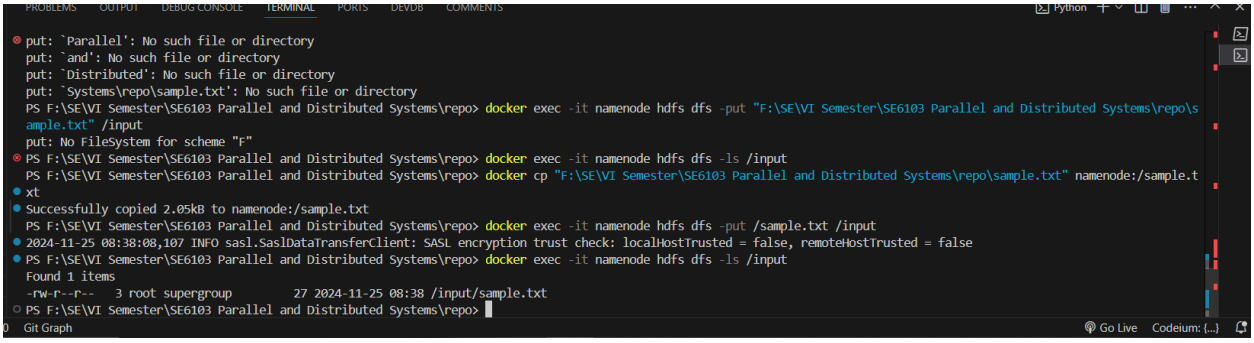
docker exec -it namenode hdfs dfs -put /path/to/sample.txt /input



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



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
```

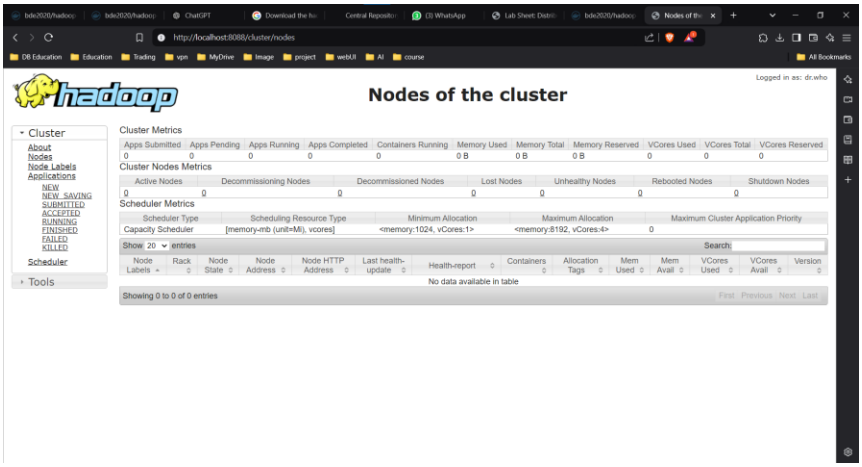
http://localhost:8088 run you should add the

docker-compose.yaml update

```yaml
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