ASSIGNMENT 2

Sabaragamuwa University of Sri Lanka

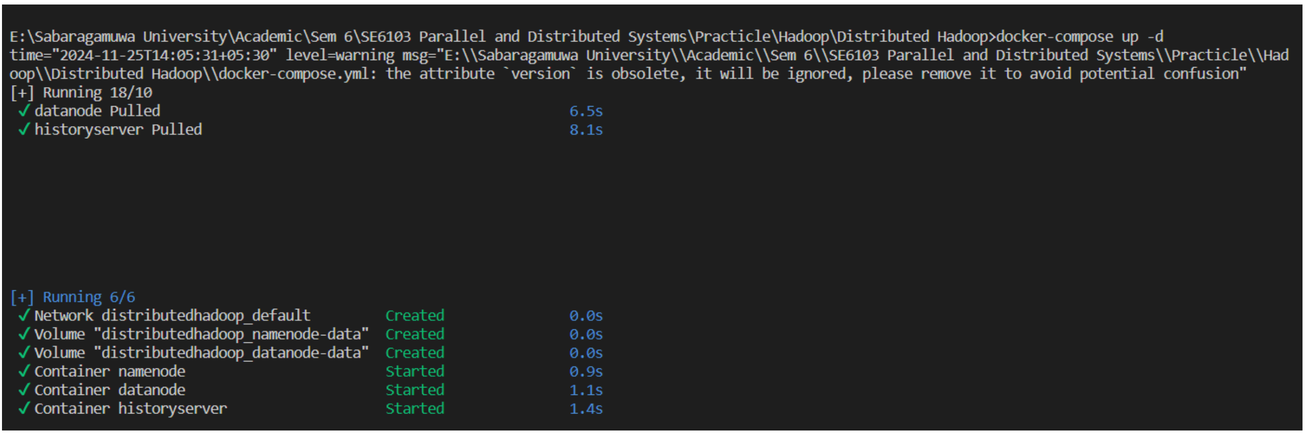
Faculty of Computing

Department of Software Engineering

SE6103 Parallel and Distributed Systems

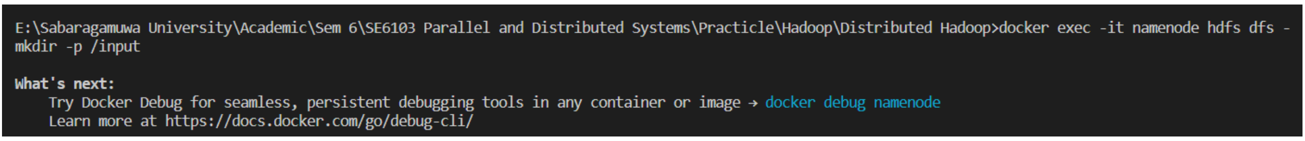
Name	: K.M.Andarawewa
Reg. No	: 19APSE4269
Academic Period	: 3rd Year 2nd Semester
Due Date	: 09/12/2024

1) Deploy the Cluster

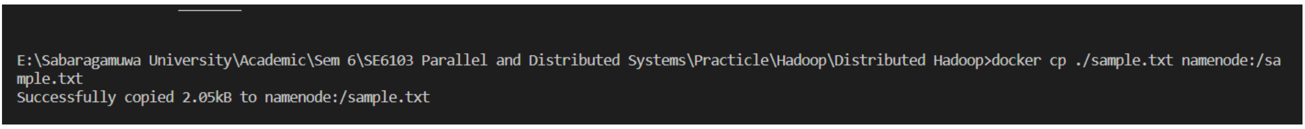


```
E:\Sabaragamuwa University\Academic\Sem 6\SE6103 Parallel and Distributed Systems\Practicle\Hadoop\Distributed Hadoop>docker-compose up -d
time="2024-11-25T14:05:31+05:30" level=warning msg="E:\\Sabaragamuwa University\\Academic\\Sem 6\\SE6103 Parallel and Distributed Systems\\Practicle\\Hadoop\\Distributed Hadoop\\docker-compose.yml: the attribute `version` is obsolete, it will be ignored, please remove it to avoid potential confusion"
[+] Running 18/10
 ✓ datanode Pulled                                              6.5s
 ✓ historyserver Pulled                                         8.1s







[+] Running 6/6
 ✓ Network distributedhadoop_default        Created             0.0s
 ✓ Volume "distributedhadoop_namenode-data" Created             0.0s
 ✓ Volume "distributedhadoop_datanode-data" Created             0.0s
 ✓ Container namenode                       Started             0.9s
 ✓ Container datanode                       Started             1.1s
 ✓ Container historyserver                  Started             1.4s
```

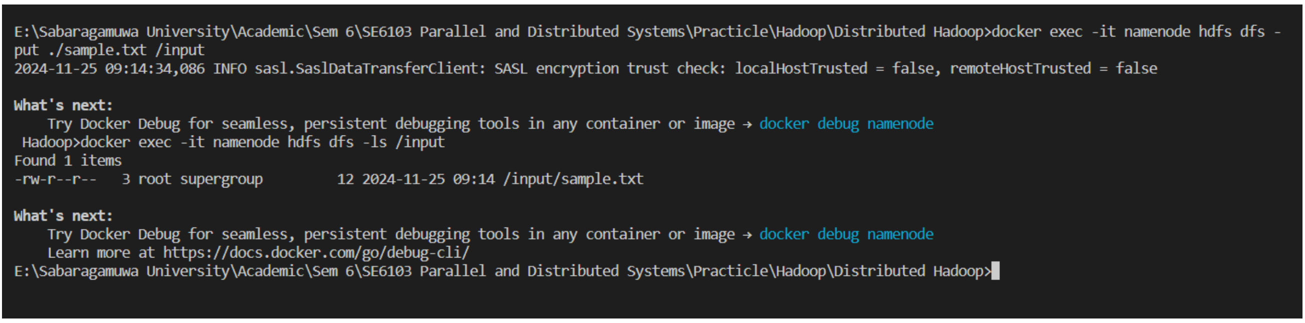
2) upload the file to HDFS:



```
E:\Sabaragamuwa University\Academic\Sem 6\SE6103 Parallel and Distributed Systems\Practicle\Hadoop\Distributed Hadoop>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/
```

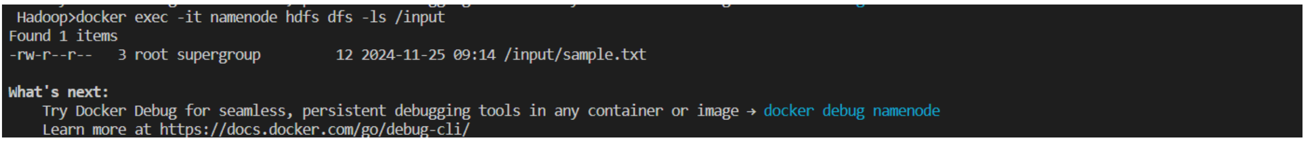


```
E:\Sabaragamuwa University\Academic\Sem 6\SE6103 Parallel and Distributed Systems\Practicle\Hadoop\Distributed Hadoop>docker cp ./sample.txt namenode:/sample.txt
Successfully copied 2.05kB to namenode:/sample.txt
```



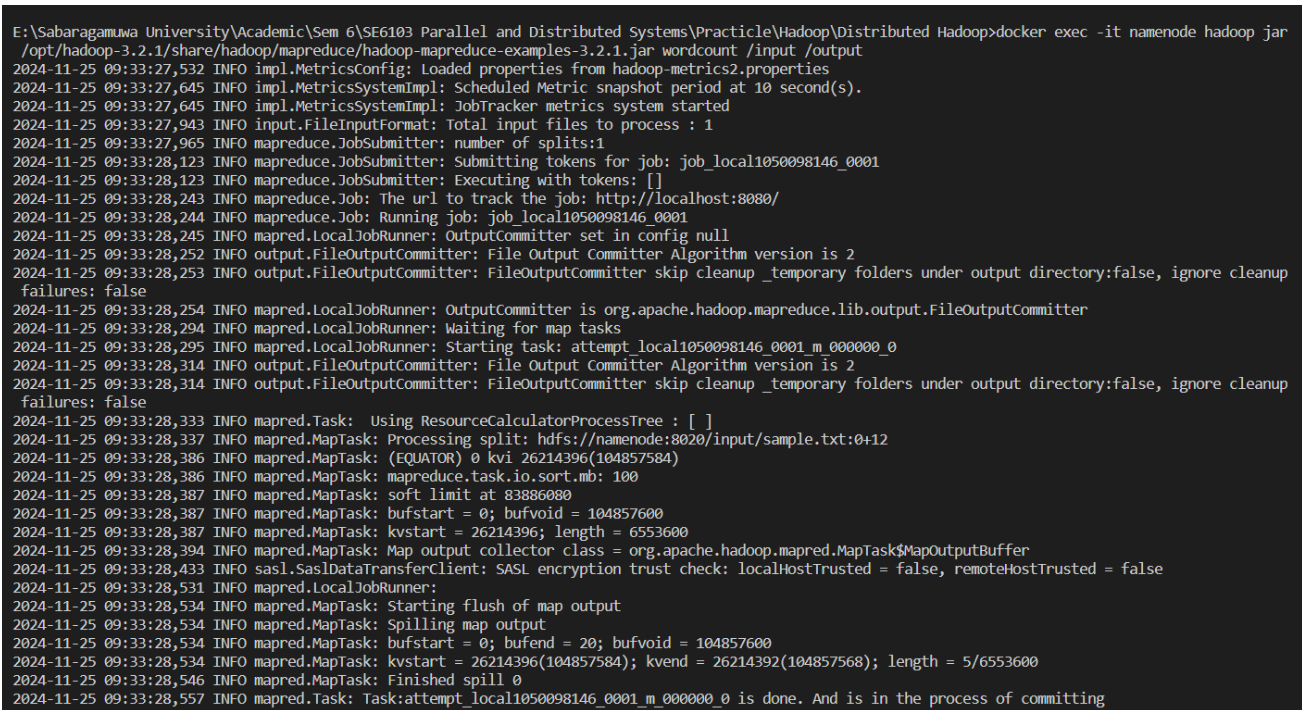
```
E:\Sabaragamuwa University\Academic\Sem 6\SE6103 Parallel and Distributed Systems\Practicle\Hadoop\Distributed Hadoop>docker exec -it namenode hdfs dfs -put ./sample.txt /input
2024-11-25 09:14:34,086 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false

What's next:
    Try Docker Debug for seamless, persistent debugging tools in any container or image → docker debug namenode
 Hadoop>docker exec -it namenode hdfs dfs -ls /input
Found 1 items
-rw-r--r--   3 root supergroup         12 2024-11-25 09:14 /input/sample.txt

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/
E:\Sabaragamuwa University\Academic\Sem 6\SE6103 Parallel and Distributed Systems\Practicle\Hadoop\Distributed Hadoop>
```

3) Verify the upload



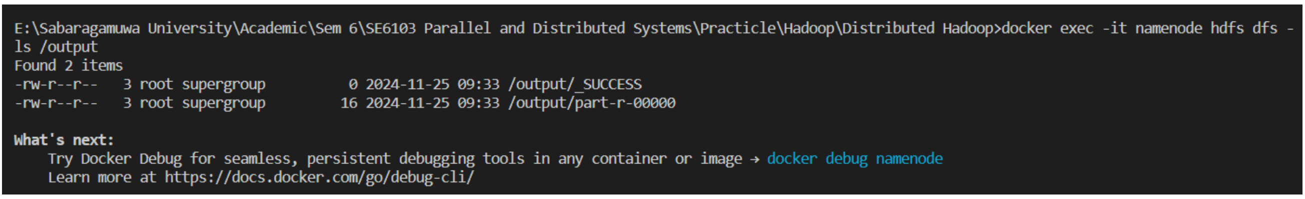
```
 Hadoop>docker exec -it namenode hdfs dfs -ls /input
Found 1 items
-rw-r--r--   3 root supergroup         12 2024-11-25 09:14 /input/sample.txt

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/
```

4) Running a MapReduce Job



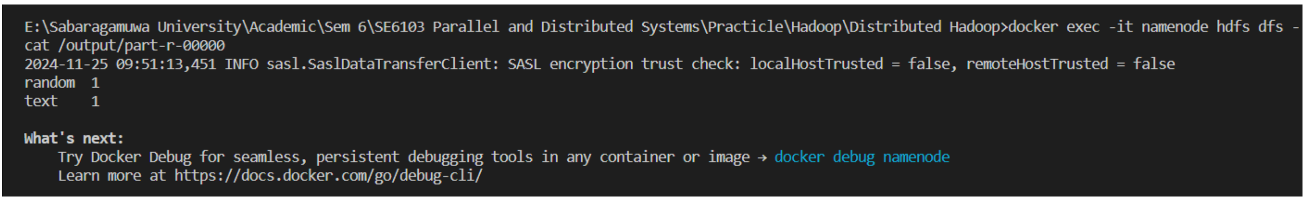
```
E:\Sabaragamuwa University\Academic\Sem 6\SE6103 Parallel and Distributed Systems\Practicle\Hadoop\Distributed Hadoop>docker exec -it namenode hadoop jar
 /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.1.jar wordcount /input /output
2024-11-25 09:33:27,532 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-11-25 09:33:27,645 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-11-25 09:33:27,645 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-11-25 09:33:27,943 INFO input.FileInputFormat: Total input files to process : 1
2024-11-25 09:33:27,965 INFO mapreduce.JobSubmitter: number of splits:1
2024-11-25 09:33:28,123 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1050098146_0001
2024-11-25 09:33:28,123 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-11-25 09:33:28,243 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-11-25 09:33:28,244 INFO mapreduce.Job: Running job: job_local1050098146_0001
2024-11-25 09:33:28,245 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-11-25 09:33:28,252 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-25 09:33:28,253 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup
 failures: false
2024-11-25 09:33:28,254 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-11-25 09:33:28,294 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-11-25 09:33:28,295 INFO mapred.LocalJobRunner: Starting task: attempt_local1050098146_0001_m_000000_0
2024-11-25 09:33:28,314 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-25 09:33:28,314 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup
 failures: false
2024-11-25 09:33:28,333 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
2024-11-25 09:33:28,337 INFO mapred.MapTask: Processing split: hdfs://namenode:8020/input/sample.txt:0+12
2024-11-25 09:33:28,386 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2024-11-25 09:33:28,386 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2024-11-25 09:33:28,387 INFO mapred.MapTask: soft limit at 83886080
2024-11-25 09:33:28,387 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2024-11-25 09:33:28,387 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2024-11-25 09:33:28,394 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2024-11-25 09:33:28,433 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-25 09:33:28,531 INFO mapred.LocalJobRunner:
2024-11-25 09:33:28,534 INFO mapred.MapTask: Starting flush of map output
2024-11-25 09:33:28,534 INFO mapred.MapTask: Spilling map output
2024-11-25 09:33:28,534 INFO mapred.MapTask: bufstart = 0; bufend = 20; bufvoid = 104857600
2024-11-25 09:33:28,534 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214392(104857568); length = 5/6553600
2024-11-25 09:33:28,546 INFO mapred.MapTask: Finished spill 0
2024-11-25 09:33:28,557 INFO mapred.Task: Task:attempt_local1050098146_0001_m_000000_0 is done. And is in the process of committing
```

5) view the output files in HDFS



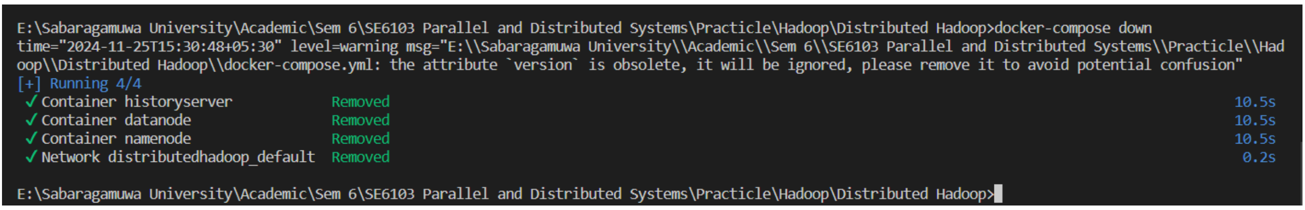
```
E:\Sabaragamuwa University\Academic\Sem 6\SE6103 Parallel and Distributed Systems\Practicle\Hadoop\Distributed Hadoop>docker exec -it namenode hdfs dfs -
ls /output
Found 2 items
-rw-r--r--   3 root supergroup          0 2024-11-25 09:33 /output/_SUCCESS
-rw-r--r--   3 root supergroup         16 2024-11-25 09:33 /output/part-r-00000

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/
```

6) Download the results:



```
E:\Sabaragamuwa University\Academic\Sem 6\SE6103 Parallel and Distributed Systems\Practicle\Hadoop\Distributed Hadoop>docker exec -it namenode hdfs dfs -
cat /output/part-r-00000
2024-11-25 09:51:13,451 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
random  1
text    1

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/
```

7) Clean Up the Cluster



```
E:\Sabaragamuwa University\Academic\Sem 6\SE6103 Parallel and Distributed Systems\Practicle\Hadoop\Distributed Hadoop>docker-compose down
time="2024-11-25T15:30:48+05:30" level=warning msg="E:\\Sabaragamuwa University\\Academic\\Sem 6\\SE6103 Parallel and Distributed Systems\\Practicle\\Had
oop\\Distributed Hadoop\\docker-compose.yml: the attribute `version` is obsolete, it will be ignored, please remove it to avoid potential confusion"
[+] Running 4/4
 ✓Container historyserver           Removed                                                                                    10.5s
 ✓Container datanode                Removed                                                                                    10.5s
 ✓Container namenode                Removed                                                                                    10.5s
 ✓Network distributedhadoop_default  Removed                                                                                     0.2s

E:\Sabaragamuwa University\Academic\Sem 6\SE6103 Parallel and Distributed Systems\Practicle\Hadoop\Distributed Hadoop>
```