LAB SHEET 2

(Single-Node Hadoop Cluster with Docker)

Department of Software Engineering Faculty of Computing Sabaragamuwa university of Sri Lanka

Parallel and Distributed System-SE6103

Name: Rasika Wedaarachchi

Reg. No: 19APSE4331

Academic Period: 3rd Year 2nd Semester

Setting up the Distributed Hadoop Cluster

Prepare the Docker Compose File

```
docker-compose.yml x

docker-compose.yml

version: '3'
services:

namenode:
image: bde2020/hadoop-namenode:latest

container_name: namenode
environment:

- CLUSTER_NAWE-ShopSmartCluster
- CORE_CONF_fs_defaultFS-hdfs://namenode:8020
ports:

- "9270:9870"

volumes:

datanode:
image: bde2020/hadoop-datanode:latest
container_name: datanode
environment:
- CRE_CONF_fs_defaultFS-hdfs://namenode:8020
volumes:

- namenode-data:/hadoop/dfs/namenode

datanode:
- image: bde2020/hadoop-datanode:latest
container_name: datanode
environment:
- CRE_CONF_fs_defaultFS-hdfs://namenode:8020
volumes:
- image: bde2020/hadoop-historyserver:latest # Replace with a valid image if necessary
container_name: historyserver
image: bde2020/hadoop-historyserver
limage: bde2020/hadoop-historyserver
```

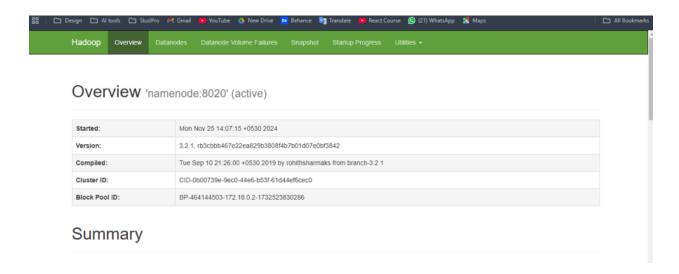
Deploy the Cluster

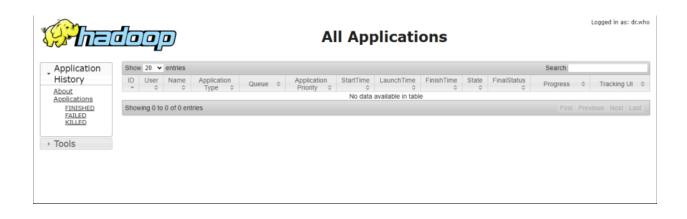
```
PS D:\Docker\Docker - Practicles_02> docker-compose up -d time="2024-11-25T14:04:25+05:30" level=warning msg="D:\\Docker\\Docker - Practicles_02\\docker-compose.yml: the attribute `version` is obsine time="2024-11-25T14:04:25+05:30" level=warning msg="D:\\Docker\\Docker\\Docker - Practicles_02\\docker-compose.yml: the attribute `version` is obsine time="2024-11-25T14:04:25+05:30" level=warning msg="D:\\Docker\\Docker\\Docker\\Docker\\Docker - Practicles_02\\docker-compose.yml: the attribute `version` is obsine time="2024-11-25T14:04:25+05:30" level=warning msg="D:\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\\Docker\Docker\Docker\\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Docker\Dock
  X historyserver Error pull access denied for bde2020/yarn-historyserver, repository does not exist or may r...
X datanode Error context canceled
Error response from daemon: pull access denied for bde2020/yarn-historyserver, repository does not exist or may require 'docker login'
PS D:\Docker\Docker - Practicles_02> docker-compose up -d
time='2024-11-25T14:06:53+05:30" level_warning msg='D:\\Docker\\Docker - Practicles_02\\docker-compose.yml: the attribute `version` is obsolete, it will be ignored, please remove it to avoid potential confusion"
[+] Running 8/2
  √ historyserver Pulled

√ datanode Pulled

 [+] Running 4/5er-practicles_02 default
         Running 5/5er-practicles 02 default
   ✓ Network docker-practicles_02_default
                                                                                                                        Created
[+] Running 5/6er-practicles_02_namenode-data"
                                                                                                                       Created
                                                                                                                                                                                                                                                                                                                                             0.05
   ✓ Network docker-practicles_02_default
        Running 6/6er-practicles_02_namenode-data"
                                                                                                                       Created
  ✓ Network docker-practicles_02_default
                                                                                                                       Created
[+] Running 5/6er-practicles_02_namenode-data"
   ✓ Network docker-practicles_02_default
[+] Running 5/6er-practicles_02_namenode-data"
     / Network docker-practicles_02_default
[+] Running 5/6er-practicles_02_namenode-data"
      Network docker-practicles_02_default
 [+] Running 5/6er-practicles_02_namenode-data"
  ✓ Network docker-practicles_02_default
 [+] Running 5/6er-practicles_02_namenode-data"
```

Verify Cluster Status





Uploading Data to HDFS

Upload Data to HDFS

```
PS D:\Docker\Docker - Practicles_02> docker cp "D:\Docker\Docker - Practicles_02\sample.txt" namenode:/sample.txt
Successfully copied 2.05kB to namenode:/sample.txt
PS D:\Docker\Docker - Practicles 02> docker cp "D:\Docker\Docker - Practicles 02\sample.txt" namenode:/sample.txt
Successfully copied 2.05kB to namenode:/sample.txt
PS D:\Docker\Docker - Practicles_02> docker exec -it namenode hdfs dfs -put /sample.txt /input
2024-11-25 09:00:48,315 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHol
stTrusted = false
     \textit{Try Docker Debug for seamless, persistent debugging tools in any container or image $\rightarrow$ docker debug namenode } \\
    Learn more at https://docs.docker.com/go/debug-cli/
PS D:\Docker\Docker - Practicles_02> docker exec -it namenode hdfs dfs -ls /input
Found 1 items
-rw-r--r-- 3 root supergroup
                                      174 2024-11-25 09:00 /input/sample.txt
    Try Docker Debug for seamless, persistent debugging tools in any container or image \rightarrow docker debug namenode
    Learn more at https://docs.docker.com/go/debug-cli/
PS D:\Docker\Docker - Practicles 02>
PS D:\Docker\Docker - Practicles_02> docker exec -it namenode hdfs dfs -ls /input
 Found 1 items
 -rw-r--r-- 3 root supergroup
                                        174 2024-11-25 09:00 /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/
PS D:\Docker\Docker - Practicles_02>
```

Running a MapReduce Job

Run the WordCount MapReduce Job

```
-mapreduce-examples-3.2.1.jar wordcount /input /new_output
 2024-11-25 09:45:49,376 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
 2024-11-25 09:45:49,471 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
 2024-11-25 09:45:49,471 INFO impl.MetricsSystemImpl: JobTracker metrics system started
 2024-11-25 09:45:49,889 INFO input.FileInputFormat: Total input files to process : 1
 2024-11-25 09:45:49,924 INFO mapreduce.JobSubmitter: number of splits:1
 2024-11-25 09:45:50,109 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1561072138_0001
 2024-11-25 09:45:50,110 INFO mapreduce.JobSubmitter: Executing with tokens: []
 2024-11-25 09:45:50,284 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
 2024-11-25 09:45:50,286 INFO mapreduce.Job: Running job: job_local1561072138_0001
 2024-11-25 09:45:50,288 INFO mapred.LocalJobRunner: OutputCommitter set in config null
 2024-11-25 09:45:50,302 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
 2024-11-25 09:45:50,302 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under outpu
 t directory:false, ignore cleanup failures: false
 2024-11-25 09:45:50,303 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutput
 Committer
 2024-11-25 09:45:50,371 INFO mapred.LocalJobRunner: Waiting for map tasks
 2024-11-25 09:45:50,372 INFO mapred.LocalJobRunner: Starting task: attempt local1561072138 0001 m 000000 0
 2024-11-25 09:45:50,426 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
 2024-11-25 09:45:50,426 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under outpu
```

Monitor the Job

```
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 D:\Docker\Docker - Practicles_02> docker exec -it namenode hdfs dfs -ls /new_output
Found 2 items

-rw-r--- 3 root supergroup
0 2024-11-25 09:45 /new_output/_SUCCESS
-rw-r--r- 3 root supergroup
205 2024-11-25 09:45 /new_output/part-r-00000
```

```
What's next:
    Try Docker Debug for seamless, persistent debugging tools in any container or image \rightarrow docker debug namenode Learn more at https://docs.docker.com/go/debug-cli/
PS D:\Docker\Docker - Practicles_02> docker exec -it namenode hdfs dfs -cat /new_output/part-r-00000
2024-11-25 10:03:21,729 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remote
ostTrusted = false
Τf
Legend, 1
Mr.
Wayne.
an
and
become 1
can't
devote
else
entirely...
ideal, 1
if
just
make
man,
```

Clean Up the Cluster