

# **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

```
Welcome  docker-compose.yml x
docker-compose.yml
1  version: '3'
2  services:
3
4      namenode:
5          image: bde2020/hadoop-namenode:latest
6          container_name: namenode
7          environment:
8              - CLUSTER_NAME=ShopSmartCluster
9              - CORE_CONF_fs_defaultFS=hdfs://namenode:8020
10         ports:
11             - "9870:9870"
12             - "9000:9000"
13         volumes:
14             - namenode-data:/hadoop/dfs/namenode
15
16     datanode:
17         image: bde2020/hadoop-datanode:latest
18         container_name: datanode
19         environment:
20             - CORE_CONF_fs_defaultFS=hdfs://namenode:8020
21         volumes:
22             - datanode-data:/hadoop/dfs/datanode
23         depends_on:
24             - namenode
25
26     historyserver:
27         image: bde2020/hadoop-historyserver:latest # Replace with a valid image if necessary
28         container_name: historyserver
29         depends_on:
30             - namenode
31             - datanode
32         ports:
```

### 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 obs...
[+] Running 2/2
X historyserver Error pull access denied for bde2020/yarn-historyserver, repository does not exist or may r... 4.1s
X datanode Error context canceled 4.1s
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 obs...
olete, it will be ignored, please remove it to avoid potential confusion"
[+] Running 8/2
✓ historyserver Pulled 7.5s
✓ datanode Pulled 7.5s
[+] Running 4/4
[+] Running 4/5er-practicles_02_default Created 0.2s
[+] Running 5/5er-practicles_02_default Created 0.2s
✓ Network docker-practicles_02_default Created 0.2s
[+] Running 5/6er-practicles_02_namenode-data" Created 0.0s
✓ Network docker-practicles_02_default Created 0.2s
[+] Running 6/6er-practicles_02_namenode-data" Created 0.0s
✓ Network docker-practicles_02_default Created 0.2s
[+] Running 5/6er-practicles_02_namenode-data" Created 0.0s
✓ Network docker-practicles_02_default Created 0.2s
[+] Running 5/6er-practicles_02_namenode-data" Created 0.0s
✓ Network docker-practicles_02_default Created 0.2s
[+] Running 5/6er-practicles_02_namenode-data" Created 0.0s
✓ Network docker-practicles_02_default Created 0.2s
[+] Running 5/6er-practicles_02_namenode-data" Created 0.0s
✓ Network docker-practicles_02_default Created 0.2s
[+] Running 5/6er-practicles_02_namenode-data" Created 0.0s
✓ Network docker-practicles_02_default Created 0.2s
[+] Running 5/6er-practicles_02_namenode-data" Created 0.0s
✓ Network docker-practicles_02_default Created 0.2s
```

## Verify Cluster Status

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot


Startup Progress

Utilities

### Overview 'namenode:8020' (active)

Started:	Mon Nov 25 14:07:15 +0530 2024
Version:	3.2.1, rb3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled:	Tue Sep 10 21:26:00 +0530 2019 by rohithsharmaks from branch-3.2.1
Cluster ID:	CID-0b00739e-9ec0-44e6-b53f-61d44ef5cec0
Block Pool ID:	BP-464144503-172.18.0.2-1732523830286

### Summary



## All Applications

Logged in as: dr.who

Application History

About Applications

FINISHED

FAILED

KILLED

Tools

Show 20 entries

Search:

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Progress	Tracking UI
No data available in table												

Showing 0 to 0 of 0 entries

First Previous Next Last

## 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, remoteHostTrusted = false

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 /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>
```

```
● 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

```
PS D:\Docker\Docker - Practicles_02> docker exec -it namenode hadoop jar /opt/hadoop-3.12.1/share/hadoop/mapreduce/hadoop-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 output directory:false, ignore cleanup failures: false
2024-11-25 09:45:50,303 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
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 output
```

## 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--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 → `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
If 1
Legend, 1
Mr. 1
Wayne. 1
a 1
an 1
and 1
become 1
can't 1
devote 1
else 1
entirely.. 1
ideal, 1
if 2
just 1
make 1
man, 1
more 1
```

## Clean Up the Cluster

```
● PS D:\Docker\Docker - Practicles_02> docker-compose down
time="2024-11-25T15:36:38+05:30" level=warning msg="D:\\Docker\\Docker - Practicles_02\\docker-compose.yml: the attribut
e `version` is obsolete, it will be ignored, please remove it to avoid potential confusion"
[+] Running 4/4
✓ Container historyserver Removed 11.7s
✓ Container datanode Removed 10.5s
✓ Container namenode Removed 12.0s
✓ Network docker-practicles_02_default Removed 0.4s
○ PS D:\Docker\Docker - Practicles_02>
```