

LAB SHEET 02

Sabaragamuwa University of Sri Lanka

Faculty of Computing

Department of Software Engineering

SE6103 – Parallel and Distributed Systems

Name	: M. S Hewage
Reg. No	: 19APSE4300
Degree Program	: Software Engineering
Academic Period	: 3 RD Year 2 ND Semester

Task 1: Setting Up the Distributed Hadoop Cluster

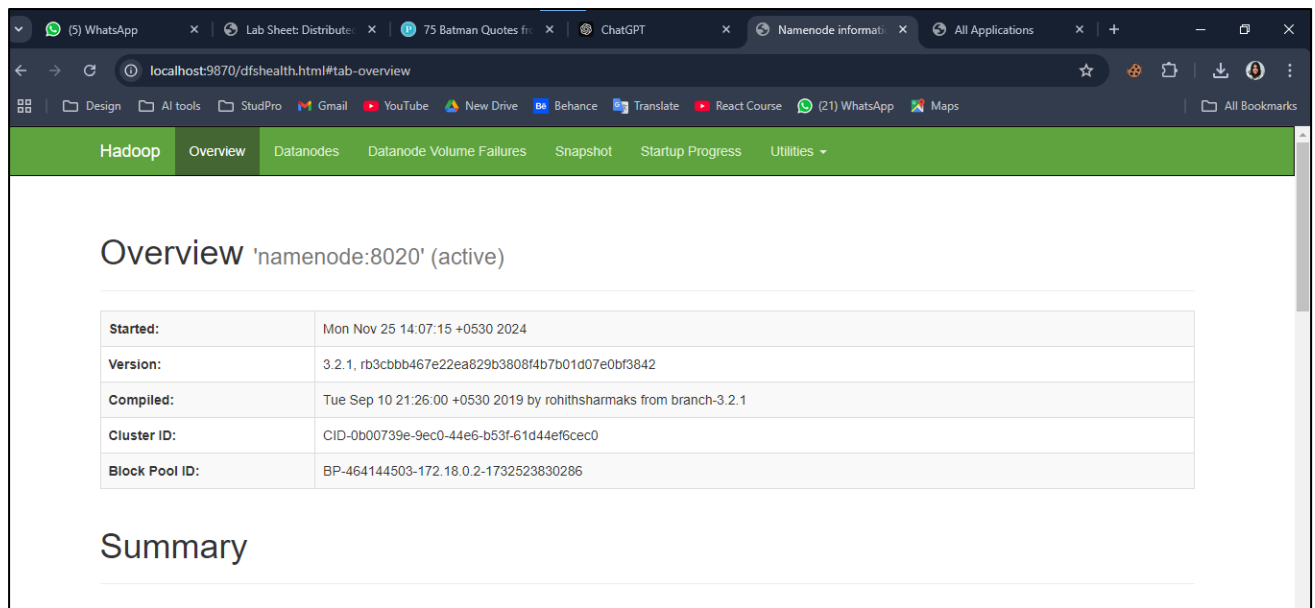
Step 1: Prepare the Docker Compose File

```
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:
33            - "8020:8020"
```

Step 2: Deploy the Cluster

[illegible]

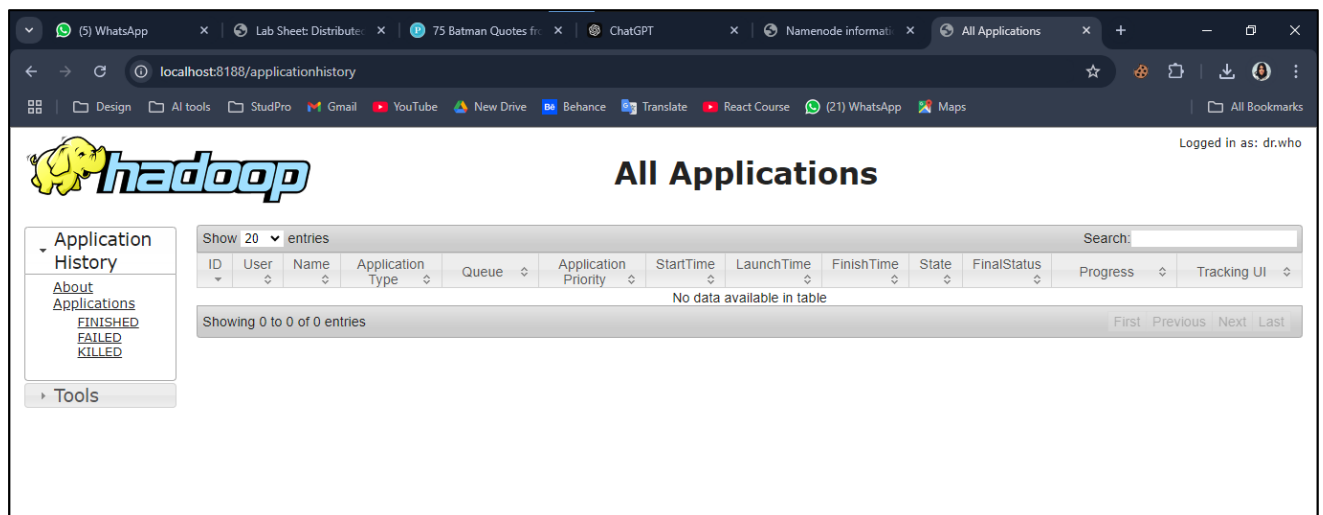
4. Step 3: Verify Cluster Status



The screenshot shows the Hadoop Overview page in a web browser. The browser's address bar displays 'localhost:9870/dfshealth.html#tab-overview'. The page has a green header with navigation tabs: 'Hadoop', 'Overview', 'Datanodes', 'Datanode Volume Failures', 'Snapshot', 'Startup Progress', and 'Utilities'. The 'Overview' tab is selected. The main content area is titled 'Overview 'namenode:8020' (active)'. Below the title is a table with the following data:

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-61d44ef6cec0
Block Pool ID:	BP-464144503-172.18.0.2-1732523830286

Below the table is a section titled 'Summary'.



The screenshot shows the Hadoop All Applications page in a web browser. The browser's address bar displays 'localhost:8188/applicationhistory'. The page has a green header with navigation tabs: 'Hadoop', 'Overview', 'Datanodes', 'Datanode Volume Failures', 'Snapshot', 'Startup Progress', and 'Utilities'. The 'All Applications' tab is selected. The main content area is titled 'All Applications'. On the left side, there is a sidebar with the Hadoop logo and a menu with the following items: 'Application History', 'About Applications', 'FINISHED', 'FAILED', 'KILLED', and 'Tools'. The main content area has a search bar and a table with the following columns: 'ID', 'User', 'Name', 'Application Type', 'Queue', 'Application Priority', 'StartTime', 'LaunchTime', 'FinishTime', 'State', 'FinalStatus', 'Progress', and 'Tracking UI'. The table is currently empty, and the text 'No data available in table' is displayed. The text 'Showing 0 to 0 of 0 entries' is also displayed. The text 'First Previous Next Last' is displayed at the bottom right of the table.

Task 2: Uploading Data to HDFS

2. 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, remoteHo
• stTrusted = 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>
```

Task 3: Running a MapReduce Job

Run the WordCount MapReduce Job

```
PS D:\Docker\Docker - Practicles_02> docker exec -it namenode hadoop jar /opt/hadoop-3.2.1/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
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 output
```

2. 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, remotelostTrusted = 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

Learn more at <https://docs.docker.com/go/debug-cli/>

- 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 attribute `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>

----- EDN -----