LAB SHEET 02

Sabaragamuwa University of Sri Lanka
Faculty of Computing
Department of Software Engineering
SE6103 – Parallel and Distributed Systems

Name : Ariyawansha R.S.K

Reg. No : 19APSE4297

Degree Program : Software Engineering

Academic Period : 3rd Year 2nd Semester

Task 1: Setting Up the Distributed Hadoop Cluster

Step 1: Prepare the Docker Compose File

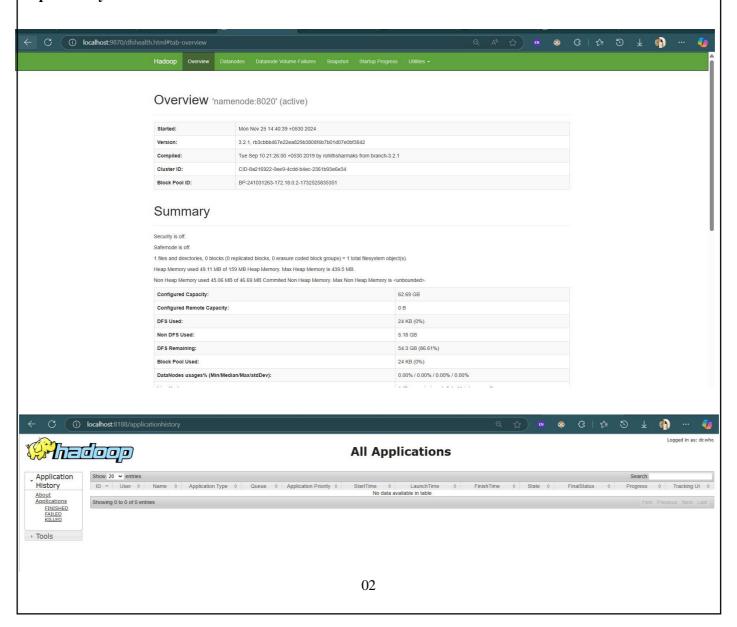
```
docker-compose.yml ×
docker-compose.yml
          image: bde2020/hadoop-namenode:latest
          container_name: namenode
           - CLUSTER_NAME=ShopSmartCluster
              CORE_CONF_fs_defaultFS=hdfs://namenode:8020
              "9870:9870"
              "9000:9000"
            - namenode-data:/hadoop/dfs/namenode
         image: bde2020/hadoop-datanode:latest
         container_name: datanode
            - CORE_CONF_fs_defaultFS=hdfs://namenode:8020
             - datanode-data:/hadoop/dfs/datanode

    namenode

          image: bde2020/hadoop-historyserver:latest # Replace with a valid image if necessary
          container_name: historyserver
 29
            - namenode
          - datanode
            - "8188:8188"
           - CORE_CONF_fs_defaultFS=hdfs://namenode:8020
```

Step 2: Deploy the Cluster

Step 3: Verify Cluster Status



Task 2: Uploading Data to HDFS

D:\SLSU\Level 06\Parallel and Distributed Systems\practical>docker exec -it namenode hdfs dfs -mkdir -p /input

```
D:\SLSU\Level 06\Parallel and Distributed Systems\practical>docker exec -it namenode hdfs dfs -put HelloWorld.txt /input
2024-11-25 09:30:18,572 INFO sasl.SaslDataTransferClient: SA^Z encryption trust check: localHostTrusted = false, remoteHostTrusted = false
```

Task 3: Running a MapReduce Job

Run the Word Count Map Reduce Job

```
D:\SisUltevel 06\Parallel and Distributed Systems\practicalDooker exec -it namenode hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-x amples-3.2.1.js mordcount /input /nem.output
amples-3.2.1.js mordcount /input /nem.output
3024-11-25 09.409.41, 393 INFO impl. MetricsSystemInpl: Scheduled Metric snapshot period at 10 second(s).
2024-11-25 09.409.41, 945 INFO impl. MetricsSystemInpl: Scheduled Metric snapshot period at 10 second(s).
2024-11-25 09.409.42, 366 INFO input.FileInputFormat: Total input files to process: 1
2024-11-25 09.409.42, 366 INFO input.FileInputFormat: Total input files to process: 1
2024-11-25 09.409.42, 414 INFO mapreduce. JobsUmmitter: submitting tokens for job: job_local143016881_0001
2024-11-25 09.409.42, 42, 414 INFO mapreduce. JobsUmmitter: submitting tokens for job: job_local143016881_0001
2024-11-25 09.409.42, 42, 414 INFO mapreduce. JobsUmmitter: Submitting with tokens: []
2024-11-25 09.409.40, 42, 404 INFO mapreduce. JobsUmmitter: Submitting with tokens: []
2024-11-25 09.409.40, 42, 404 INFO mapreduce. Jobs Interpretation into a submitter Su
```

View Job Output

```
D:\SLSU\Level 06\Parallel and Distributed Systems\practical>docker exec -it namenode hdfs dfs -ls /new_output
Found 2 items
-rw-r--r- 3 root supergroup 0 2024-11-25 09:49 /new_output/_SUCCESS
-rw-r--r- 3 root supergroup 93 2024-11-25 09:49 /new_output/part-r-00000
```

```
D:\SLSU\Level 06\Parallel and Distributed Systems\practical>docker exec -it namenode hdfs dfs -cat /new_output/part-r-00000

2024-11-25 10:05:01,909 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false

Hello 3

This 1

again. 1

document. 1

everyone123. 1

first 1

is 1

my 1

text 1

world. 1

D:\SLSU\Level 06\Parallel and Distributed Systems\practical>
```

Task 4: Analyze and Clean Up

```
D:\SLSU\Level 06\Parallel and Distributed Systems\practical>docker-compose down

[+] Running 4/4

<Container historyserver

<Container datanode

<Container namenode

<Container
```