LAB SHEET 02

Sabaragamuwa University of Sri Lanka

Faculty of Computing

Department of Software Engineering

SE6103 – Parallel and Distributed Systems

Name	: K.O.K.S.Dayarathna
Reg. No	: 19APSE4299
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

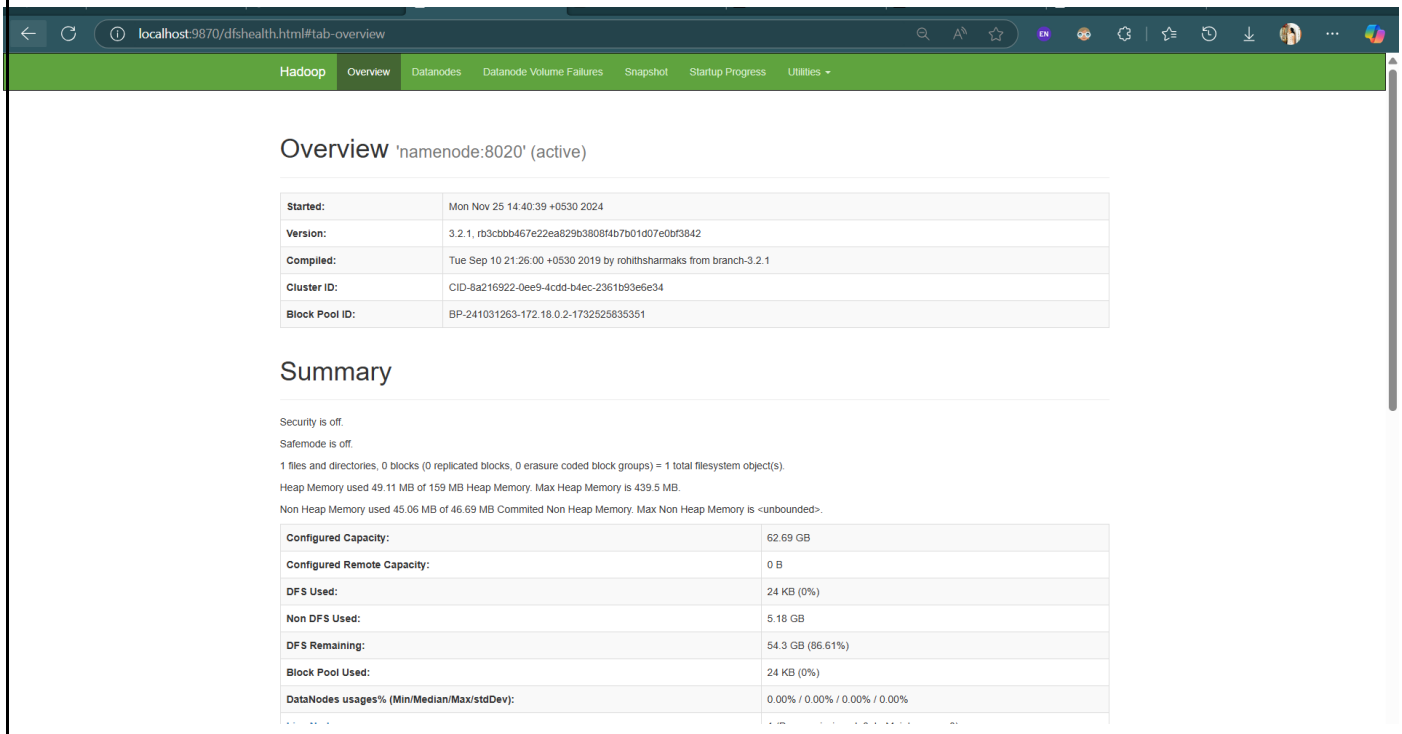
docker-compose.yml

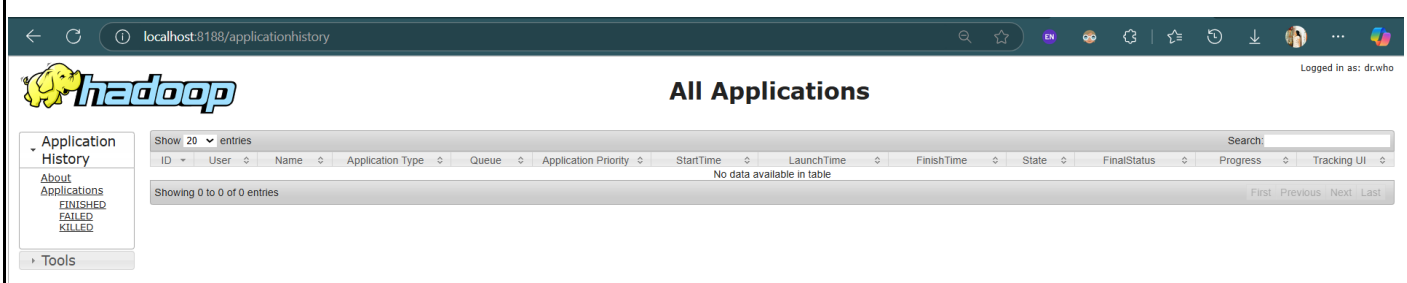
```yaml
version: '3'
services:

  namenode:
    image: bde2020/hadoop-namenode:latest
    container_name: namenode
    environment:
      - CLUSTER_NAME=ShopSmartCluster
      - CORE_CONF_fs_defaultFS=hdfs://namenode:8020
    ports:
      - "9870:9870"
      - "9000:9000"
    volumes:
      - namenode-data:/hadoop/dfs/namenode

  datanode:
    image: bde2020/hadoop-datanode:latest
    container_name: datanode
    environment:
      - CORE_CONF_fs_defaultFS=hdfs://namenode:8020
    volumes:
      - datanode-data:/hadoop/dfs/datanode
    depends_on:
      - namenode

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

volumes:
  namenode-data:
  datanode-data:
```

Step 2: Deploy the Cluster

```
D:\SLSU\Level 06\Parallel and Distributed Systems\practical>docker-compose up -d
[+] Running 18/18
 ✔datanode 8 layers [##########]      0B/0B      Pulled                                  9.8s
   ✔77920a3e82af Already exists                                                        0.0s
   ✔92329e81aec4 Already exists                                                        0.0s
   ✔f373218fec59 Already exists                                                        0.0s
   ✔c3a84a3e49c8 Already exists                                                        0.0s
   ✔a65640a64a76 Already exists                                                        0.0s
   ✔4bf0ae3d5cc8 Pull complete                                                         2.7s
   ✔b91d0b0b68c8 Pull complete                                                         3.0s
   ✔5e185246c615 Pull complete                                                         3.1s
 ✔historyserver 8 layers [##########]      0B/0B      Pulled                           8.5s
   ✔3192219afd04 Already exists                                                        0.0s
   ✔7127a1d8cced Already exists                                                        0.0s
   ✔883a89599900 Already exists                                                        0.0s
   ✔aa53513fe997 Already exists                                                        0.0s
   ✔8b1800105b98 Already exists                                                        0.0s
   ✔78d381637ee0 Pull complete                                                         1.8s
   ✔84560426d8fd Pull complete                                                         1.7s
   ✔f3f6b02c1935 Pull complete                                                         1.2s
[+] Building 0.0s (0/0)                                                  docker:default
[+] Running 6/6
 ✔Network practical_default          Created                                          0.1s
 ✔Volume "practical_namenode-data"   Created                                          0.0s
 ✔Volume "practical_datanode-data"   Created                                          0.0s
 ✔Container namenode                 Started                                          0.1s
 ✔Container datanode                 Started                                          0.1s
 ✔Container historyserver            Started                                          0.1s
```

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 WordCount MapReduce Job

```
D:\SLSU\Level 06\Parallel and Distributed Systems\practical>docker exec -it namenode hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-ex
amples-3.2.1.jar wordcount /input /new_output
2024-11-25 09:49:41,830 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-11-25 09:49:41,945 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-11-25 09:49:41,946 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-11-25 09:49:42,366 INFO input.FileInputFormat: Total input files to process : 1
2024-11-25 09:49:42,404 INFO mapreduce.JobSubmitter: number of splits:1
2024-11-25 09:49:42,614 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local143016881_0001
2024-11-25 09:49:42,614 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-11-25 09:49:42,767 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-11-25 09:49:42,769 INFO mapreduce.Job: Running job: job_local143016881_0001
2024-11-25 09:49:42,774 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-11-25 09:49:42,786 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-25 09:49:42,786 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup fa
ilures: false
2024-11-25 09:49:42,787 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-11-25 09:49:42,840 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-11-25 09:49:42,841 INFO mapred.LocalJobRunner: Starting task: attempt_local143016881_0001_m_000000_0
2024-11-25 09:49:42,889 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-25 09:49:42,889 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup fa
ilures: false
2024-11-25 09:49:42,921 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
2024-11-25 09:49:42,926 INFO mapred.MapTask: Processing split: hdfs://namenode:8020/input/HelloWorld.txt:0+82
2024-11-25 09:49:43,033 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2024-11-25 09:49:43,033 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2024-11-25 09:49:43,033 INFO mapred.MapTask: soft limit at 83886080
2024-11-25 09:49:43,033 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2024-11-25 09:49:43,033 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2024-11-25 09:49:43,050 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2024-11-25 09:49:43,141 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-25 09:49:43,279 INFO mapred.LocalJobRunner:
2024-11-25 09:49:43,283 INFO mapred.MapTask: Starting flush of map output
2024-11-25 09:49:43,283 INFO mapred.MapTask: Spilling map output
2024-11-25 09:49:43,283 INFO mapred.MapTask: bufstart = 0; bufend = 135; bufvoid = 104857600
2024-11-25 09:49:43,283 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214348(104857392); length = 49/6553600
```

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
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   Removed                                  10.3s
 ✔Container datanode        Removed                                  10.2s
 ✔Container namenode        Removed                                  10.2s
 ✔Network practical_default Removed                                   0.1s
```