LAB SHEET 01

Sabaragamuwa University of Sri Lanka

Faculty of Computing

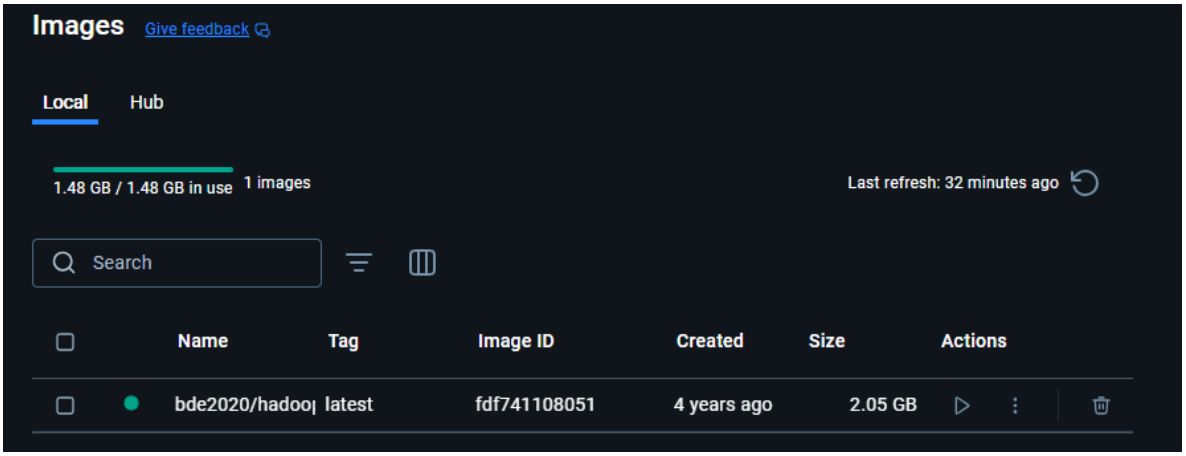
Department of Software Engineering

SE6103 – Parallel and Distributed Systems

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

Step 1: Pull the Hadoop Docker Image

1. Choose a Hadoop Docker Image:



2. Verify the Download:

```
PS C:\Users\HP> docker images
REPOSITORY                TAG       IMAGE ID       CREATED       SIZE
bde2020/hadoop-namenode   latest    fdf741108051   4 years ago   2.05GB
PS C:\Users\HP>
```

Step 2: Start the Hadoop Container

1. Run the Container:

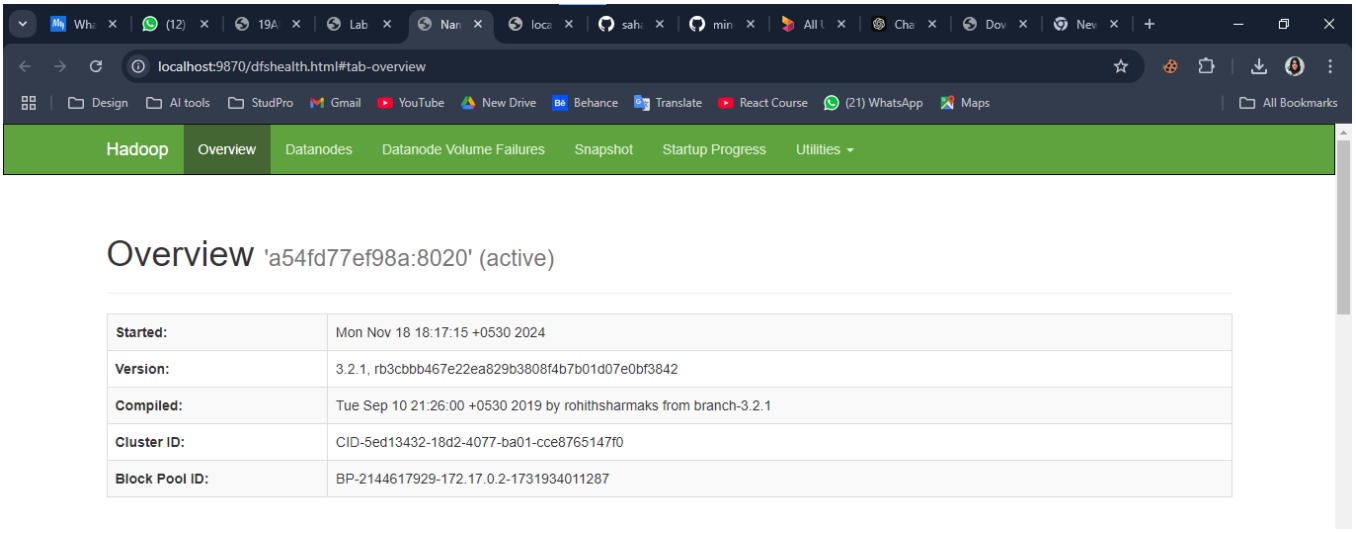
```
PS C:\Users\HP> docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:8088 -p 50070:50070 b
de2020/hadoop-namenode:latest /bin/bash
Configuring core
 - Setting fs.defaultFS=hdfs://a54fd77ef98a:8020
Configuring hdfs
 - Setting dfs.namenode.name.dir=file:///hadoop/dfs/name
Configuring yarn
Configuring httpfs
Configuring kms
Configuring mapred
Configuring for multihomed network
root@a54fd77ef98a:/#
```

2. Start Hadoop Services:

HDFS Web Interface:

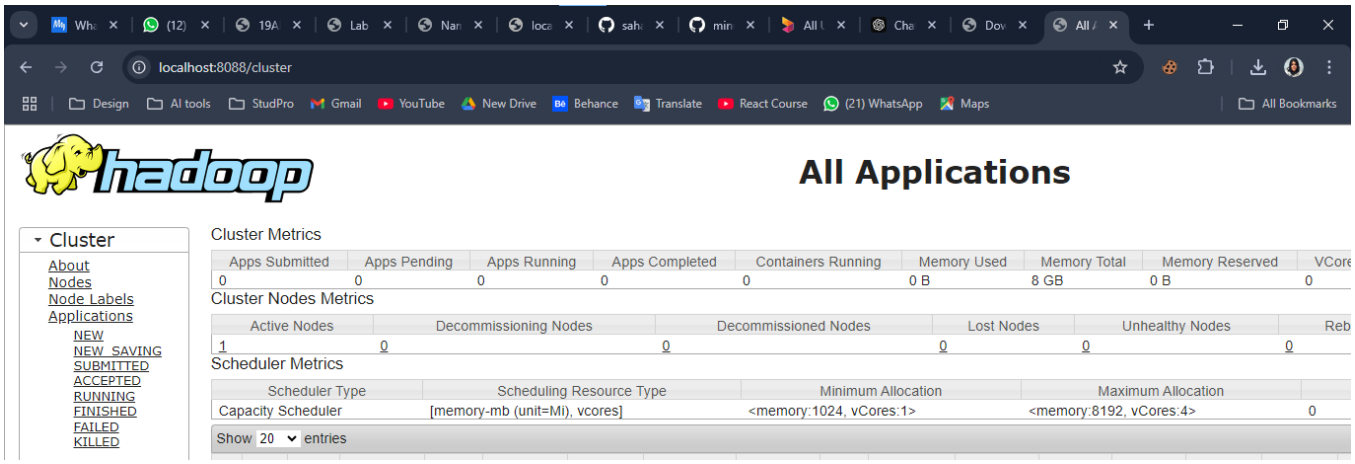
```
2024-11-18 12:46:49,529 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = a54fd77ef98a/172.17.0.2
STARTUP_MSG:   args = [-format]
```

```
SHUTDOWN_MSG: Shutting down NameNode at a54fd77ef98a/172.17.0.2
************************************************************/
root@a54fd77ef98a:/# /opt/hadoop-3.2.1/bin/hdfs --daemon start namenode
root@a54fd77ef98a:/# /opt/hadoop-3.2.1/bin/hdfs --daemon start datanode
root@a54fd77ef98a:/#
```



YARN Web Interface:

```
root@a54fd77ef98a:/# /opt/hadoop-3.2.1/bin/yarn --daemon start resourcemanager
root@a54fd77ef98a:/# /opt/hadoop-3.2.1/bin/yarn --daemon start nodemanager
root@a54fd77ef98a:/#
```



Step 4: Running a Sample MapReduce Job

1. Upload Sample Data to HDFS:

```
root@a54fd77ef98a:/# hdfs dfs -mkdir -p /user/hadoop/input
root@a54fd77ef98a:/# hdfs dfs -put $HADOOP_HOME/etc/hadoop/*.xml /user/hadoop/input
2024-11-18 13:02:22,799 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remot
eHostTrusted = false
2024-11-18 13:02:23,561 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remot
eHostTrusted = false
2024-11-18 13:02:23,610 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remot
eHostTrusted = false
2024-11-18 13:02:23,657 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remot
eHostTrusted = false
2024-11-18 13:02:23,704 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remot
eHostTrusted = false
2024-11-18 13:02:23,761 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remot
eHostTrusted = false
```

2. Run the WordCount Job:

```
root@a54fd77ef98a:/# hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.1.jar wordcount /use
r/hadoop/input /user/hadoop/output
2024-11-18 13:08:28,292 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-11-18 13:08:28,381 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-11-18 13:08:28,381 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-11-18 13:08:28,846 INFO input.FileInputFormat: Total input files to process : 9
2024-11-18 13:08:28,901 INFO mapreduce.JobSubmitter: number of splits:9
2024-11-18 13:08:29,072 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local397008345_0001
2024-11-18 13:08:29,073 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-11-18 13:08:29,260 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-11-18 13:08:29,262 INFO mapreduce.Job: Running job: job_local397008345_0001
2024-11-18 13:08:29,264 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-11-18 13:08:29,283 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-18 13:08:29,283 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under out
put directory:false, ignore cleanup failures: false
```

3. heck the Output:

```
<value>default</value>  1
<value>false</value>    2
<value>org.apache.hadoop.yarn.util.resource.DefaultResourceCalculator</value>  1
A       22
ACL     37
ACL,    2
ACLs    1
ANY     9
ASF     1
AdminOperationsProtocol.        1
Any     2
Apache  10
ApplicationClientProtocol,      1
ApplicationHistoryProtocol,     1
ApplicationMaster       1
ApplicationMasterProtocol,      2
ApplicationMasters      3
BASIS,  9
But     1
CONDITIONS      9
```

Step 5: Exiting the Container

1. Stop the Container:

```
root@a54fd77ef98a:/# exit
exit
PS C:\Users\HP> docker stop hadoop-cluster
hadoop-cluster
PS C:\Users\HP>
```

2. Restart the Container:

```
PS C:\Users\HP> docker start -i hadoop-cluster
Configuring core
 - Setting fs.defaultFS=hdfs://a54fd77ef98a:8020
Configuring hdfs
 - Setting dfs.namenode.name.dir=file:///hadoop/dfs/name
Configuring yarn
Configuring httpfs
Configuring kms
Configuring mapred
Configuring for multihomed network
root@a54fd77ef98a:/#
```