ASSIGNMENT 01

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
Academic Period	: 3rd Year 2nd Semester
Due Date	: 11/18/2024
Degree program	: Software Engineering

Parallel and Distributed System Practical

Step 1:

1.Check the Docker Version

```
C:\ProgramData\Microsoft\Windows\Start Menu>docker --version
Docker version 27.2.0, build 3ab4256
```

2.Choose and pull Hadoop Docker Image

```
C:\ProgramData\Microsoft\Windows\Start Menu>docker pull bde2020/hadoop-namenode:latest
latest: Pulling from bde2020/hadoop-namenode
3192219afd04: Pull complete
7127a1d8cced: Pull complete
883a89599900: Pull complete
77920a3e82af: Pull complete
92329e81aec4: Pull complete
f373218fec59: Pull complete
aa53513fe997: Pull complete
8b1800105b98: Pull complete
c3a84a3e49c8: Pull complete
a65640a64a76: Pull complete
a29cc756d786: Pull complete
abf352b16046: Pull complete
dddd5a449e99: Pull complete
Digest: sha256:fdf74110805132d646cf6f12635efc0919e1fb2ac5bd376c5366272fc261301e
Status: Downloaded newer image for bde2020/hadoop-namenode:latest
docker.io/bde2020/hadoop-namenode:latest

What's next:
    View a summary of image vulnerabilities and recommendations → docker scout quickview bde2020/hadoop-namenode:latest

C:\ProgramData\Microsoft\Windows\Start Menu>
C:\ProgramData\Microsoft\Windows\Start Menu>
```

3.Verify the Downloads

```
C:\ProgramData\Microsoft\Windows\Start Menu>
C:\ProgramData\Microsoft\Windows\Start Menu>docker images
REPOSITORY                  TAG       IMAGE ID       CREATED        SIZE
bde2020/hadoop-namenode     latest    b638307a2119   4 years ago    1.37GB

C:\ProgramData\Microsoft\Windows\Start Menu>
```

Step 2: Start the Hadoop Container

1. Run the Container:

```
C:\ProgramData\Microsoft\Windows\Start Menu>docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:
8088 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash
Configuring core
 - Setting fs.defaultFS=hdfs://644cd7f22964:8020
Configuring hdfs
 - Setting dfs.namenode.name.dir=file:///hadoop/dfs/name
Configuring yarn
Configuring httpfs
Configuring kms
Configuring mapred
Configuring for multihomed network
root@644cd7f22964:/#
```

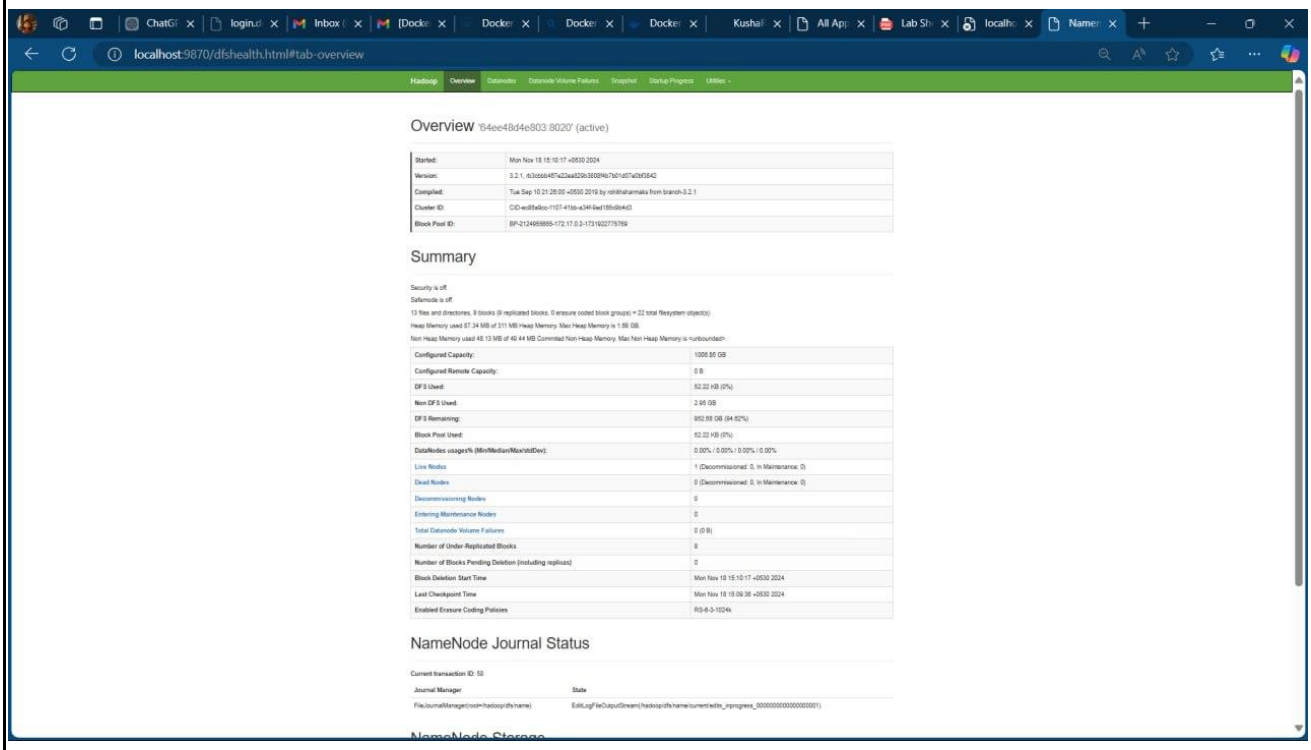
2. Start Hadoop Services:

```
C:\ProgramData\Microsoft\Windows\Start Menu>docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:
8088 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash
Configuring core
 - Setting fs.defaultFS=hdfs://644cd7f22964:8020
Configuring hdfs
 - Setting dfs.namenode.name.dir=file:///hadoop/dfs/name
Configuring yarn
Configuring httpfs
Configuring kms
Configuring mapred
Configuring for multihomed network
root@644cd7f22964:/#
```

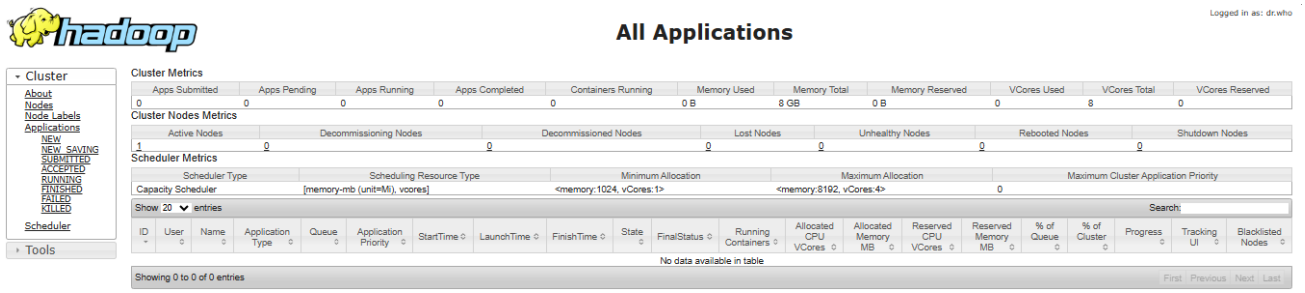
```
root@644cd7f22964:/# /opt/hadoop-3.2.1/bin/hdfs --daemon start namenode
root@644cd7f22964:/# /opt/hadoop-3.2.1/bin/hdfs --daemon start datanode
root@644cd7f22964:/# /opt/hadoop-3.2.1/bin/yarn --daemon start resourcemanager
root@644cd7f22964:/# /opt/hadoop-3.2.1/bin/yarn --daemon start nodemanager
root@644cd7f22964:/#
```

Step 3: Access Hadoop Web Interfaces

HDFS Web Interface



YARN Web Interface:



Step 4: Running a Sample MapReduce Job

1. Upload Sample Data to HDFS:

```
root@64ee48d4e803:/# jps
258 DataNode
2628 NameNode
340 ResourceManager
2729 Jps
425 NodeManager
root@64ee48d4e803:/# hdfs dfs -mkdir -p /user/hadoop/input
root@64ee48d4e803:/# hdfs dfs -put $HADOOP_HOME/etc/hadoop/*.xml /user/hadoop/input
2024-11-18 09:41:10,651 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 09:41:11,704 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 09:41:12,205 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 09:41:12,688 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 09:41:12,770 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 09:41:13,276 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 09:41:13,797 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 09:41:13,876 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 09:41:14,353 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
root@64ee48d4e803:/#
```

2. Run the Word Count Job:

```
2024-11-18 14:53:14,987 INFO namenode.FSDirectory: USER serial map: bits=24 maxEntries=16777215
2024-11-18 14:53:14,987 INFO namenode.FSDirectory: GROUP serial map: bits=24 maxEntries=16777215
2024-11-18 14:53:14,987 INFO namenode.FSDirectory: XATTR serial map: bits=24 maxEntries=16777215
2024-11-18 14:53:15,011 INFO util.GSet: Computing capacity for map INodeMap
2024-11-18 14:53:15,011 INFO util.GSet: VM type       = 64-bit
2024-11-18 14:53:15,012 INFO util.GSet: 1.0% max memory 1.7 GB = 17.4 MB
2024-11-18 14:53:15,012 INFO util.GSet: capacity      = 2^21 = 2097152 entries
2024-11-18 14:53:15,050 INFO namenode.FSDirectory: ACLs enabled? false
2024-11-18 14:53:15,050 INFO namenode.FSDirectory: POSIX ACL inheritance enabled? true
2024-11-18 14:53:15,050 INFO namenode.FSDirectory: XAttrs enabled? true
2024-11-18 14:53:15,051 INFO namenode.NameNode: Caching file names occurring more than 10 times
2024-11-18 14:53:15,062 INFO snapshot.SnapshotManager: Loaded config captureOpenFiles: false, skipCapt
ureAccessTimeOnlyChange: false, snapshotDiffAllowSnapRootDescendant: true, maxSnapshotLimit: 65536
2024-11-18 14:53:15,067 INFO snapshot.SnapshotManager: SkipList is disabled
2024-11-18 14:53:15,075 INFO util.GSet: Computing capacity for map cachedBlocks
2024-11-18 14:53:15,075 INFO util.GSet: VM type       = 64-bit
2024-11-18 14:53:15,075 INFO util.GSet: 0.25% max memory 1.7 GB = 4.4 MB
2024-11-18 14:53:15,076 INFO util.GSet: capacity      = 2^19 = 524288 entries
2024-11-18 14:53:15,090 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
2024-11-18 14:53:15,090 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2024-11-18 14:53:15,090 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2024-11-18 14:53:15,098 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2024-11-18 14:53:15,099 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry
cache entry expiry time is 600000 millis
2024-11-18 14:53:15,103 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2024-11-18 14:53:15,103 INFO util.GSet: VM type       = 64-bit
2024-11-18 14:53:15,104 INFO util.GSet: 0.029999999329447746% max memory 1.7 GB = 535.3 KB
2024-11-18 14:53:15,104 INFO util.GSet: capacity      = 2^16 = 65536 entries
2024-11-18 14:53:15,156 INFO namenode.FSImage: Allocated new BlockPoolId: BP-865134611-172.17.0.2-1731
941595143
```

3. Check the Output:

```
zero    2
root@64ee48d4e803:/# hdfs dfs -cat /user/hadoop/output/part-r-00000
2024-11-18 09:56:44,644 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
"*"     21
"AS     9
"License");     9
"alice,bob      21
"clumping"      1
(ASF)   1
(root   1
(the    9
-->     18
-1      1
-1,     1
0.0     1
1-MAX_INT.      1
1.      1
1.0.    1
2.0     9
40      2
40+20=60        1
:       2
<!--    18
</configuration>        9
</description>  33
</name> 2
</property>     57
<?xml   8
<?xml-stylesheet        4
<configuration> 9
<description>   31
<description>ACL        25
<description>Default    1
<name>default.key.acl.DECRYPT_EEK</name>        1
<name>default.key.acl.GENERATE_EEK</name>       1
<name>default.key.acl.MANAGEMENT</name> 1
<name>default.key.acl.READ</name>       1
<name>hadoop.kms.acl.CREATE</name>      1
<name>hadoop.kms.acl.DECRYPT_EEK</name> 1
<name>hadoop.kms.acl.DELETE</name>      1
```

Step 5: Exiting the Container

1. Stop the Container:

```
root@64ee48d4e803:/#  docker stop hadoop-cluster
bash: docker: command not found
root@64ee48d4e803:/#  stop hadoop-cluster
bash: stop: command not found
root@64ee48d4e803:/# exit
```

2. Restart the container

```
C:\ProgramData\Microsoft\Windows\Start Menu>docker restart hadoop-cluster
hadoop-cluster

C:\ProgramData\Microsoft\Windows\Start Menu>
```