Assignment 01

Parallel Distributed System

SE-5101

Name: M.A.K.R.Senadheera

Index: 19APSE4310

Semester: 3 nd Year 2 nd Semester

Parallel and Distributed System Practical

Step 1:

1.Check The Docker Version

```
C:\Users\rashm>docker --version
Docker version 27.2.0, build 3ab4256
```

2.Choose and pull Hadoop Docker Image

```
C:\Users\rashm>docker pull bde2020/hadoop-namenode:latest
latest: Pulling from bde2020/hadoop-namenode
dddd5a449e99: Download complete
3192219afd04: Download complete
7127a1d8cced: Download complete
883a89599900: Download complete
77920a3e82af: Download complete
92329e81aec4: Download complete
f373218fec59: Download complete
aa53513fe997: Download complete
8b1800105b98: Download complete
c3a84a3e49c8: Download complete
a65640a64a76: Download complete
a29cc756d786: Download complete
abf352b16046: Download 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:\Users\rashm>
```

3.Verify The Downloads

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

Step 2: Start the Hadoop Container

1. **Run the Container:**

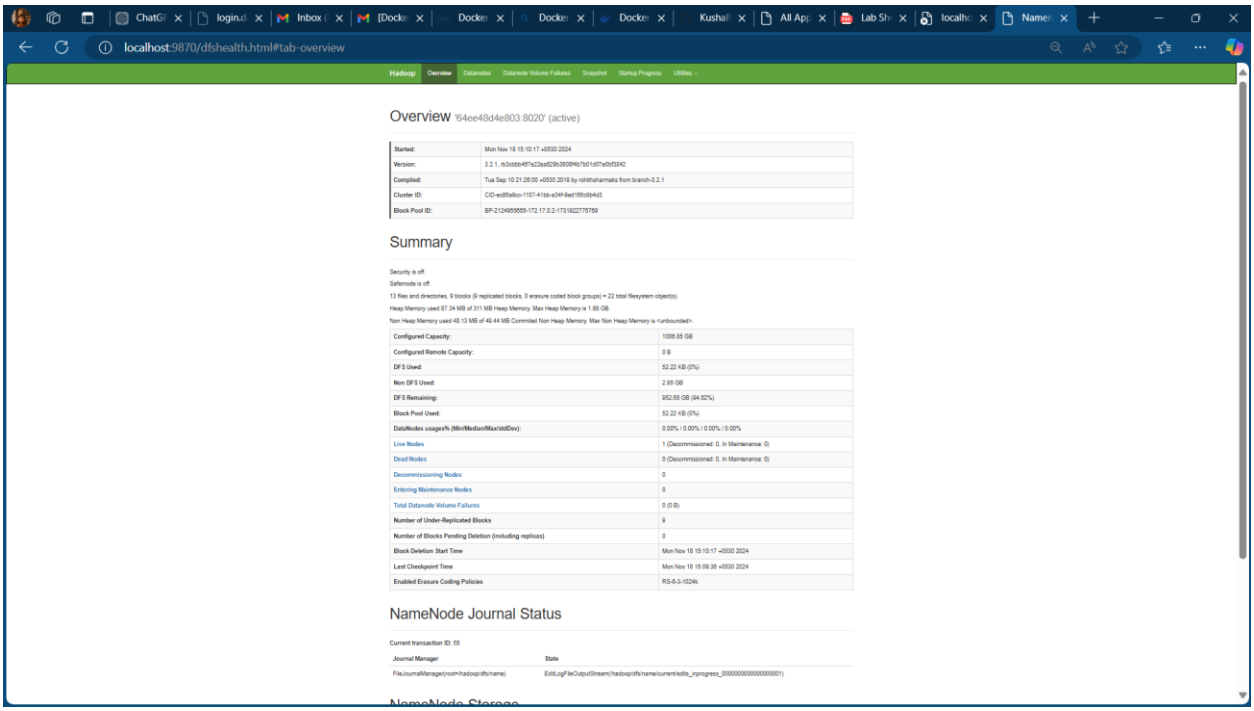
```
C:\Users\rashm> docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:80
88 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash
Configuring core
 - Setting fs.defaultFS=hdfs://64ee48d4e803:8020
Configuring hdfs
 - Setting dfs.namenode.name.dir=file:///hadoop/dfs/name
Configuring yarn
Configuring httpfs
Configuring kms
Configuring mapred
Configuring for multihomed network
root@64ee48d4e803:/#
```

2. **Start Hadoop Services:**

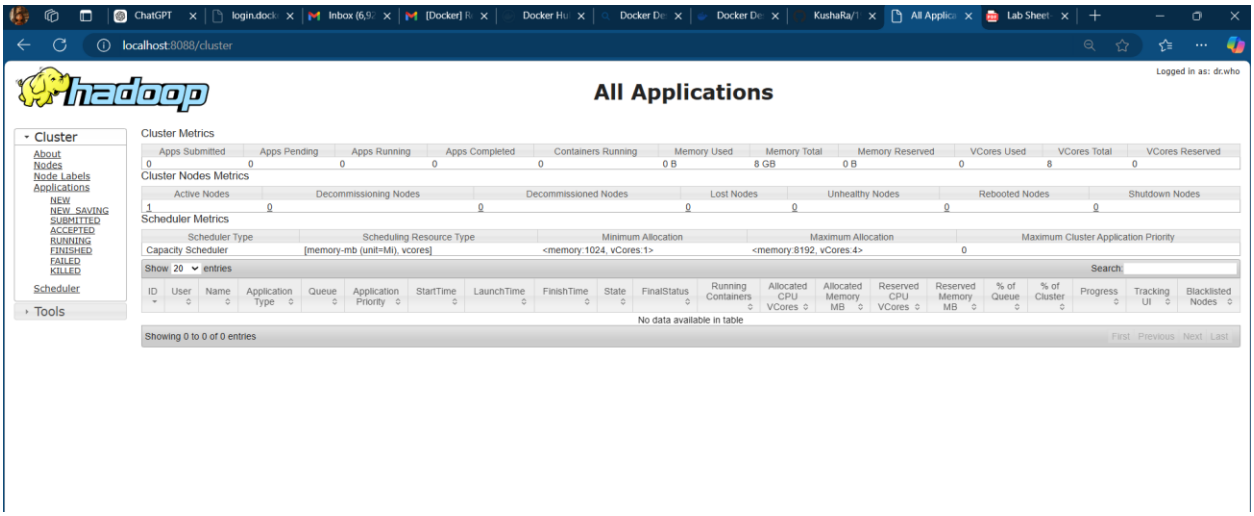
```
C:\Users\rashm> docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:80
88 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash
Configuring core
 - Setting fs.defaultFS=hdfs://64ee48d4e803:8020
Configuring hdfs
 - Setting dfs.namenode.name.dir=file:///hadoop/dfs/name
Configuring yarn
Configuring httpfs
Configuring kms
Configuring mapred
Configuring for multihomed network
root@64ee48d4e803:/# /opt/hadoop-3.2.1/bin/hdfs --daemon start namenode
root@64ee48d4e803:/# /opt/hadoop-3.2.1/bin/hdfs --daemon start datanode
root@64ee48d4e803:/# /opt/hadoop-3.2.1/bin/yarn --daemon start resourcemanag
er
root@64ee48d4e803:/# /opt/hadoop-3.2.1/bin/yarn --daemon start nodemanager
```

Step 3: Access Hadoop Web Interfaces

HDFS Web Interface



● YARN Web Interface:



Step 4: Running a Sample MapReduce Job

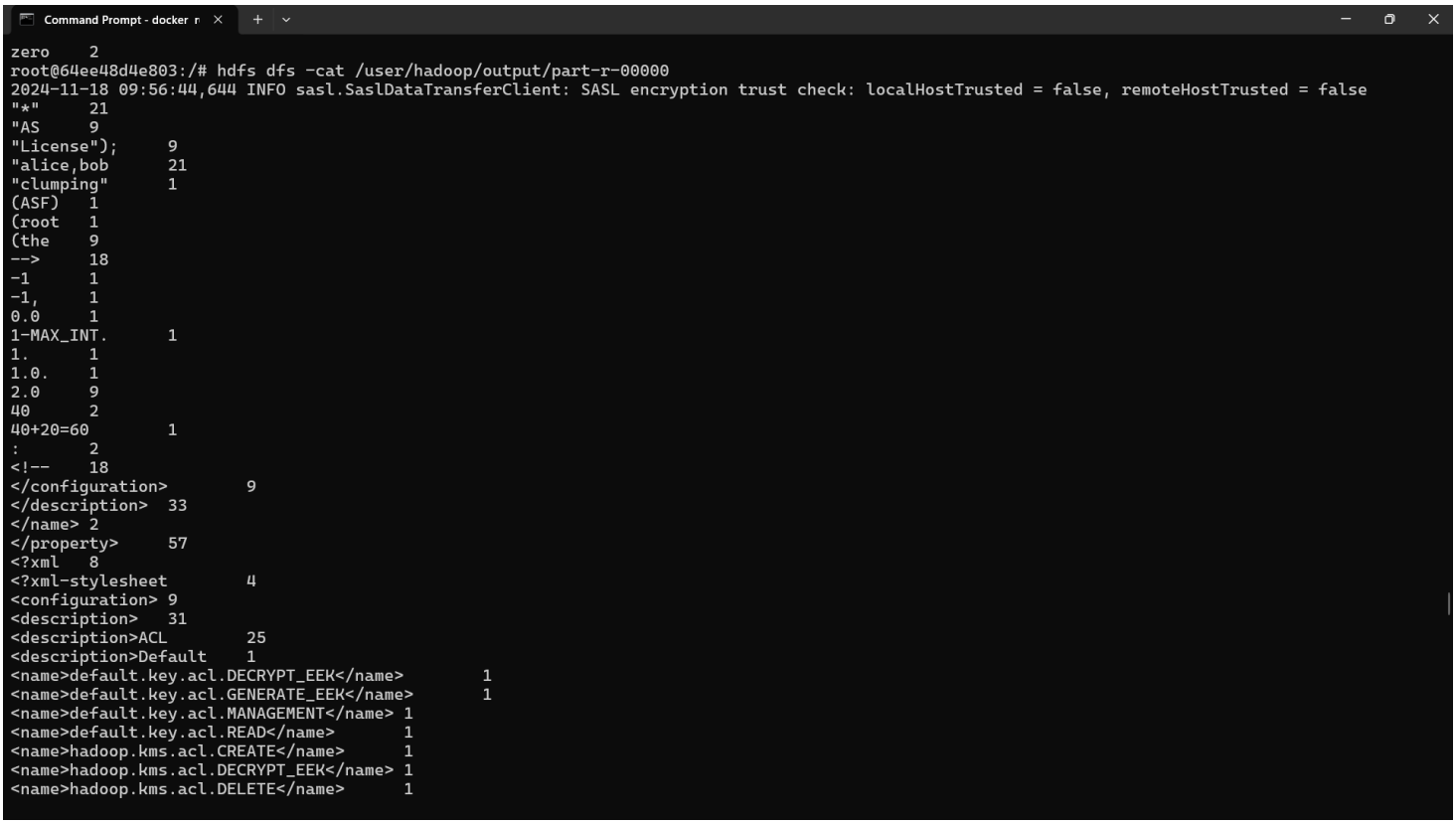
1. Upload Sample Data to HDFS:

```
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:

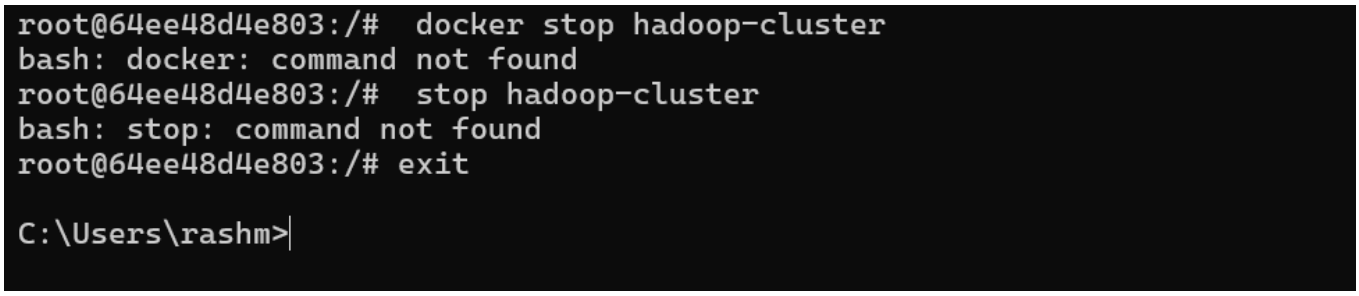
```
root@64ee48d4e803:/# $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jarwordcount /user/hadoop/input /user/hadoop/output
bash: /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jarwordcount: No such file or directory
root@64ee48d4e803:/# hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar wordcount /user/hadoop/input /user/hadoop/output
2024-11-18 09:50:39,126 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-11-18 09:50:39,319 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-11-18 09:50:39,320 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-11-18 09:50:39,992 INFO input.FileInputFormat: Total input files to process : 9
2024-11-18 09:50:40,050 INFO mapreduce.JobSubmitter: number of splits:9
2024-11-18 09:50:40,395 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local893345318_0001
2024-11-18 09:50:40,395 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-11-18 09:50:40,619 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-11-18 09:50:40,622 INFO mapreduce.Job: Running job: job_local893345318_0001
2024-11-18 09:50:40,623 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-11-18 09:50:40,648 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-18 09:50:40,649 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-11-18 09:50:40,656 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-11-18 09:50:40,756 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-11-18 09:50:40,757 INFO mapred.LocalJobRunner: Starting task: attempt_local893345318_0001_m_000000_0
2024-11-18 09:50:40,805 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-18 09:50:40,805 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-11-18 09:50:40,847 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
2024-11-18 09:50:40,861 INFO mapred.MapTask: Processing split: hdfs://64ee48d4e803:8020/user/hadoop/input/hadoop-policy.xml:0+11392
2024-11-18 09:50:40,949 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2024-11-18 09:50:40,949 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2024-11-18 09:50:40,951 INFO mapred.MapTask: soft limit at 83886080
2024-11-18 09:50:40,951 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2024-11-18 09:50:40,951 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2024-11-18 09:50:40,964 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2024-11-18 09:50:41,043 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 09:50:41,265 INFO mapred.LocalJobRunner:
2024-11-18 09:50:41,272 INFO mapred.MapTask: Starting flush of map output
2024-11-18 09:50:41,272 INFO mapred.MapTask: Spilling map output
2024-11-18 09:50:41,272 INFO mapred.MapTask: bufstart = 0; bufend = 15783; bufvoid = 104857600
2024-11-18 09:50:41,272 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26209160(104836640); length = 5237/6553600
2024-11-18 09:50:41,322 INFO mapred.MapTask: Finished spill 0
2024-11-18 09:50:41,344 INFO mapred.Task: Task:attempt_local893345318_0001_m_000000_0 is done. And is in the process of committing
2024-11-18 09:50:41,360 INFO mapred.LocalJobRunner: map
```

3. Check the Output:



Step 5: Exiting the Container

1. Stop the Container:



2. Restart the container