

ASSIGNMENT

Sabaragamuwa University of Sri Lanka

Faculty of Computing

Department of Software Engineering

SE6103 - Parallel and Distributed Systems

Name	: Sabeeb A.I.M.
Reg. No	: 19APSE4289
Academic Period	: 03 rd Year 02 nd Semester
Degree Program	: Software Engineering
Due Date	: 19/11/2024

1. To check the docker version:

docker --version

```
Terminal

Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

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

2. To pull the Hadoop docker image:

docker pull bde2020/hadoop-namenode:latest

```
PS C:\Users\SHABEEB> docker pull bde2020/hadoop-namenode:latest
latest: Pulling from bde2020/hadoop-namenode
Digest: sha256:fd74110805132d646cf6f12635efc0919e1fb2ac5bd376c5366272fc261301e
Status: Image is up to date for bde2020/hadoop-namenode:latest
docker.io/bde2020/hadoop-namenode:latest
```

3. To list the docker images to confirm the Hadoop image is downloaded:

docker images

```
PS C:\Users\SHABEEB> docker images
REPOSITORY          TAG         IMAGE ID      CREATED        SIZE
dockerapp            1.1         fdeef914e82f  3 weeks ago   185MB
nginx                latest      3b25b682ea82  6 weeks ago   192MB
hello-world          latest      d2c94e258dcb  18 months ago 13.3kB
bde2020/hadoop-namenode latest      b638307a2119  4 years ago   1.37GB
PS C:\Users\SHABEEB>
```

4. To run the Hadoop docker image:

docker run bde2020/hadoop-namenode:latest

```
PS C:\Users\SHABEEB> docker run bde2020/hadoop-namenode:latest
Configuring core
- Setting fs.defaultFS=hdfs://9548cfae69ef:8020
Configuring hdfs
- Setting dfs.namenode.name.dir=file:///hadoop/dfs/name
Configuring yarn
Configuring httpfs
Configuring kms
Configuring mapred
Configuring for multihomed network
Cluster name not specified
PS C:\Users\SHABEEB>
```

5. To run a Docker container for a Hadoop NameNode instance:

docker run -it --name hadoop-cluster202 -p 9872:9871 -p 8090:8089 -p 60070:50070
bde2020/hadoop-namenode:latest /bin/bash

```
PS C:\Users\SHABEEB> docker run -it --name hadoop-cluster999 -p 9872:9871 -p 8090:8089 -p 60070:50070 bde2020/hadoop-namenode:latest /bin/bash
Configuring core
- Setting fs.defaultFS=hdfs://8af718d1f9f1:8020
Configuring hdfs
- Setting dfs.namenode.name.dir=file:///hadoop/dfs/name
Configuring yarn
Configuring httpfs
Configuring kms
Configuring mapred
Configuring for multihomed network
root@8af718d1f9f1:/#
```

6. To start the Hadoop services once inside the container's shell at once:

start-all.sh

```
root@8af718d1f9f1:/# start-all.sh
bash: start-all.sh: command not found
root@8af718d1f9f1:/#

RAM 1.29 GB CPU -- % Disk -- GB avail. of -- GB
```

Note: If above code produces an error as "ERROR: namenode is not COMMAND nor fully qualified CLASSNAME" continue from below codes.

7. To initialize and format the Hadoop Distributed File System (HDFS) NameNode:

hdfs namenode -format

```
Terminal
root@8af718d1f9f1:/# hdfs namenode -format
2024-11-18 17:02:19,169 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = 8af718d1f9f1/172.17.0.2
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.2.1
STARTUP_MSG: classpath = /etc/hadoop:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-io-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/co
mmon/lib/jsr311-api-1.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/accessors-smart-1.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/guava-
27.0-jre.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-xdr-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-util-1.0.1.jar:/opt/h
adoop-3.2.1/share/hadoop/common/lib/gson-2.2.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-beanutils-1.9.3.jar:/opt/hadoop-3.2.1/share
/hadoop/common/lib/kerby-pkix-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/zookeeper-3.4.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib
/commons-text-1.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-identity-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/curator-client
-2.13.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/dnsjava-2.1.7.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsr305-3.0.0.jar:/opt/hadoop-3
.2.1/share/hadoop/common/lib/checker-qual-2.5.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/log4j-1.2.17.jar:/opt/hadoop-3.2.1/share/hadoop/co
mmon/lib/kerb-simplekdc-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/error_prone_annotations-2.2.0.jar:/opt/hadoop-3.2.1/share/hadoop/com
mon/lib/javax.servlet-api-3.1.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-servlet-1.19.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/
htrace-core4-4.1.0-incubating.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-client-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/j2ob
jc-annotations-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-annotations-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-a
sn1-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/json-smart-2.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-core-1.0.1.jar:/opt/h
adoop-3.2.1/share/hadoop/common/lib/re2j-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/asm-5.0.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/li
b/commons-compress-1.18.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsp-api-2.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-xml-9.3.24
.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/slf4j-api-1.7.25.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/stax2-api-3.1.4.jar:/opt
/hadoop-3.2.1/share/hadoop/common/lib/jackson-jaxrs-1.9.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsch-0.1.54.jar:/opt/hadoop-3.2.1/share
/hadoop/common/lib/commons-io-2.5.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/audience-annotations-0.5.0.jar:/opt/hadoop-3.2.1/share/hadoop/co
```

8. To start the HDFS NameNode in the background:

hdfs namenode &

```
Terminal
*****
root@8af718d1f9f1:/# hdfs namenode &
[1] 170
root@8af718d1f9f1:/# 2024-11-18 17:03:15,892 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = 8af718d1f9f1/172.17.0.2
STARTUP_MSG: args = []
STARTUP_MSG: version = 3.2.1
STARTUP_MSG: classpath = /etc/hadoop:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-io-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/co
mmon/lib/jsr311-api-1.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/accessors-smart-1.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/guava-
27.0-jre.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-xdr-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-util-1.0.1.jar:/opt/h
adoop-3.2.1/share/hadoop/common/lib/gson-2.2.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-beanutils-1.9.3.jar:/opt/hadoop-3.2.1/share
/hadoop/common/lib/kerby-pkix-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/zookeeper-3.4.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib
/commons-text-1.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-identity-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/curator-client
-2.13.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/dnsjava-2.1.7.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsr305-3.0.0.jar:/opt/hadoop-3
.2.1/share/hadoop/common/lib/checker-qual-2.5.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/log4j-1.2.17.jar:/opt/hadoop-3.2.1/share/hadoop/co
mmon/lib/kerb-simplekdc-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/error_prone_annotations-2.2.0.jar:/opt/hadoop-3.2.1/share/hadoop/com
mon/lib/javax.servlet-api-3.1.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-servlet-1.19.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/
htrace-core4-4.1.0-incubating.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-client-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/j2ob
jc-annotations-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-annotations-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-a
sn1-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/json-smart-2.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-core-1.0.1.jar:/opt/h
adoop-3.2.1/share/hadoop/common/lib/re2j-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/asm-5.0.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/li
b/commons-compress-1.18.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsp-api-2.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-xml-9.3.24
.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/slf4j-api-1.7.25.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/stax2-api-3.1.4.jar:/opt
RAM 1.60 GB CPU 0.13% Disk -- GB avail. of -- GB
BETA > Terminal
```

9. To start the HDFS DataNode in the background:

hdfs datanode &

```
Terminal

root@8af718d1f9f1:/# hdfs datanode &
[2] 295
root@8af718d1f9f1:/# 2024-11-18 17:04:13,746 INFO datanode.DataNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting DataNode
STARTUP_MSG: host = 8af718d1f9f1/172.17.0.2
STARTUP_MSG: args = []
STARTUP_MSG: version = 3.2.1
STARTUP_MSG: classpath = /etc/hadoop:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-io-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/co
mmon/lib/jsr311-api-1.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/accessors-smart-1.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/guava-
27.0-jre.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-xdr-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-util-1.0.1.jar:/opt/h
adoop-3.2.1/share/hadoop/common/lib/gson-2.2.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-beanutils-1.9.3.jar:/opt/hadoop-3.2.1/share
/hadoop/common/lib/kerby-pkix-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/zookeeper-3.4.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib
/commons-text-1.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-identity-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/curator-client
-2.13.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/dnsjava-2.1.7.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsr305-3.0.0.jar:/opt/hadoop-3
.2.1/share/hadoop/common/lib/checker-qual-2.5.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/log4j-1.2.17.jar:/opt/hadoop-3.2.1/share/hadoop/co
mmon/lib/kerb-simplekdc-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/error_prone_annotations-2.2.0.jar:/opt/hadoop-3.2.1/share/hadoop/com
mon/lib/javax.servlet-api-3.1.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-servlet-1.19.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/
htrace-core4-4.1.0-incubating.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-client-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/j2ob
jc-annotations-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-annotations-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-a
sn1-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/json-smart-2.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-core-1.0.1.jar:/opt/ha
doo-3.2.1/share/hadoop/common/lib/re2j-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/asm-5.0.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/li
b/commons-compress-1.18.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsp-api-2.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-xml-9.3.24
.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/slf4j-api-1.7.25.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/stax2-api-3.1.4.jar:/opt
/hadoop-3.2.1/share/hadoop/common/lib/jackson-jaxrs-1.9.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsch-0.1.54.jar:/opt/hadoop-3.2.1/share
```

10. To start the YARN NodeManager process in the background:

yarn nodemanager &

```
Terminal

root@8af718d1f9f1:/# yarn nodemanager &
[3] 406
root@8af718d1f9f1:/# 2024-11-18 17:04:59,720 INFO nodemanager.NodeManager: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NodeManager
STARTUP_MSG: host = 8af718d1f9f1/172.17.0.2
STARTUP_MSG: args = []
STARTUP_MSG: version = 3.2.1
STARTUP_MSG: classpath = /etc/hadoop:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-io-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/co
mmon/lib/jsr311-api-1.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/accessors-smart-1.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/guava-
27.0-jre.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-xdr-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-util-1.0.1.jar:/opt/h
adoop-3.2.1/share/hadoop/common/lib/gson-2.2.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-beanutils-1.9.3.jar:/opt/hadoop-3.2.1/share
/hadoop/common/lib/kerby-pkix-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/zookeeper-3.4.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib
/commons-text-1.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-identity-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/curator-client
-2.13.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/dnsjava-2.1.7.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsr305-3.0.0.jar:/opt/hadoop-3
.2.1/share/hadoop/common/lib/checker-qual-2.5.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/log4j-1.2.17.jar:/opt/hadoop-3.2.1/share/hadoop/co
mmon/lib/kerb-simplekdc-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/error_prone_annotations-2.2.0.jar:/opt/hadoop-3.2.1/share/hadoop/com
mon/lib/javax.servlet-api-3.1.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-servlet-1.19.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/
htrace-core4-4.1.0-incubating.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-client-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/j2ob
jc-annotations-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-annotations-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-a
sn1-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/json-smart-2.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-core-1.0.1.jar:/opt/ha
doo-3.2.1/share/hadoop/common/lib/re2j-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/asm-5.0.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/li
b/commons-compress-1.18.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsp-api-2.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-xml-9.3.24
.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/slf4j-api-1.7.25.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/stax2-api-3.1.4.jar:/opt
/hadoop-3.2.1/share/hadoop/common/lib/jackson-jaxrs-1.9.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsch-0.1.54.jar:/opt/hadoop-3.2.1/share

RAM 2.16 GB CPU 1.25% Disk -- GB avail. of -- GB BETA >_ Terminal
```

11. To start the YARN ResourceManager process in the background:

yarn resourcemanager &

```
Terminal
root@8af718d1f9f1:/# yarn resourcemanager &
[5] 922
root@8af718d1f9f1:/# 2024-11-18 17:08:02,162 INFO resourcemanager.ResourceManager: STARTUP_MSG:
/*****
STARTUP_MSG: Starting ResourceManager
STARTUP_MSG: host = 8af718d1f9f1/172.17.0.2
STARTUP_MSG: args = []
STARTUP_MSG: version = 3.2.1
STARTUP_MSG: classpath = /etc/hadoop:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-io-9.3.24.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/co
mmon/lib/jsr311-api-1.1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/accessors-smart-1.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/guava-
27.0-jre.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-xdr-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-util-1.0.1.jar:/opt/h
adoop-3.2.1/share/hadoop/common/lib/gson-2.2.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/commons-beanutils-1.9.3.jar:/opt/hadoop-3.2.1/share
/hadoop/common/lib/kerby-pkix-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/zookeeper-3.4.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib
/commons-text-1.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-identity-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/curator-client
-2.13.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/dnsjava-2.1.7.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsr305-3.0.0.jar:/opt/hadoop-3
.2.1/share/hadoop/common/lib/checker-qual-2.5.2.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/log4j-1.2.17.jar:/opt/hadoop-3.2.1/share/hadoop/co
mmon/lib/kerb-simplekdc-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/error-prone-annotations-2.2.0.jar:/opt/hadoop-3.2.1/share/hadoop/com
mon/lib/javax-servlet-api-3.1.0.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jersey-servlet-1.19.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib
/htrace-core4-4.1.0-incubating.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-client-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/j2ob
jc-annotations-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jackson-annotations-2.9.8.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerby-a
sni-1.0.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/json-smart-2.3.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/kerb-core-1.0.1.jar:/opt/h
adoop-3.2.1/share/hadoop/common/lib/re2j-1.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/asn-5.0.4.jar:/opt/hadoop-3.2.1/share/hadoop/common/li
b/commons-compress-1.18.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsp-api-2.1.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jetty-xml-9.3.24
.v20180605.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/slf4j-api-1.7.25.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/stax2-api-3.1.4.jar:/opt
/hadoop-3.2.1/share/hadoop/common/lib/jackson-jaxrs-1.9.13.jar:/opt/hadoop-3.2.1/share/hadoop/common/lib/jsch-0.1.54.jar:/opt/hadoop-3.2.1/share
RAM 2.49 GB CPU 0.50% Disk -- GB avail. of -- GB BETA Terminal
```

12. To create an input directory and upload sample files to HDFS:

hdfs dfs -mkdir -p /user/hadoop/input

```
[5]+  Exit 255                  yarn resourcemanager
root@8af718d1f9f1:/# hdfs dfs -mkdir -p /user/hadoop/input
2024-11-18 17:08:46,516 INFO namenode.FSEditLog: Number of transactions: 4 Total time for transactions(ms): 21 Number of transactions batched in
Syncs: 0 Number of syncs: 2 SyncTimes(ms): 15
root@8af718d1f9f1:/#
RAM 2.54 GB CPU 0.13% Disk -- GB avail. of -- GB BETA Terminal
```

hdfs dfs -put \$HADOOP_HOME/etc/hadoop/*.xml /user/hadoop/input

```
root@8af718d1f9f1:/# hdfs dfs -put $HADOOP_HOME/etc/hadoop/*.xml /user/hadoop/input
2024-11-18 17:09:15,131 INFO hdfs.StateChange: BLOCK* allocate blk_1073741825_1001, replicas=172.17.0.2:9866 for /user/hadoop/input/capacity-sch
eduler.xml._COPYING_
2024-11-18 17:09:15,186 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 17:09:15,426 INFO datanode.DataNode: Receiving BP-618151755-172.17.0.2-1731949349024:blk_1073741825_1001 src: /172.17.0.2:35410 dest:
/172.17.0.2:9866
2024-11-18 17:09:15,561 INFO DataNode.clienttrace: src: /172.17.0.2:35410, dest: /172.17.0.2:9866, bytes: 8260, op: HDFS_WRITE, cliID: DFSClient
_NONMAPREDUCE_616880655_1, offset: 0, srvID: 7dc0bd10-42d8-412c-a90e-67c614cc31b8, blockid: BP-618151755-172.17.0.2-1731949349024:blk_1073741825
_1001, duration(ns): 82098977
2024-11-18 17:09:15,562 INFO datanode.DataNode: PacketResponder: BP-618151755-172.17.0.2-1731949349024:blk_1073741825_1001, type=LAST_IN_PIPELIN
E terminating
2024-11-18 17:09:15,577 INFO namenode.FSNamesystem: BLOCK* blk_1073741825_1001 is COMMITTED but not COMPLETE(numNodes= 0 < minimum = 1) in file
/user/hadoop/input/capacity-scheduler.xml._COPYING_
2024-11-18 17:09:15,988 INFO hdfs.StateChange: DIR* completeFile: /user/hadoop/input/capacity-scheduler.xml._COPYING_ is closed by DFSClient_NON
MAPREDUCE_616880655_1
2024-11-18 17:09:16,034 INFO hdfs.StateChange: BLOCK* allocate blk_1073741826_1002, replicas=172.17.0.2:9866 for /user/hadoop/input/core-site.xml
_COPYING_
2024-11-18 17:09:16,041 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2024-11-18 17:09:16,044 INFO datanode.DataNode: Receiving BP-618151755-172.17.0.2-1731949349024:blk_1073741826_1002 src: /172.17.0.2:35422 dest:
/172.17.0.2:9866
2024-11-18 17:09:16,053 INFO DataNode.clienttrace: src: /172.17.0.2:35422, dest: /172.17.0.2:9866, bytes: 860, op: HDFS_WRITE, cliID: DFSClient
_NONMAPREDUCE_616880655_1, offset: 0, srvID: 7dc0bd10-42d8-412c-a90e-67c614cc31b8, blockid: BP-618151755-172.17.0.2-1731949349024:blk_1073741826
_1002, duration(ns): 4569079
2024-11-18 17:09:16,053 INFO datanode.DataNode: PacketResponder: BP-618151755-172.17.0.2-1731949349024:blk_1073741826_1002, type=LAST_IN_PIPELIN
E terminating
RAM 2.55 GB CPU 0.25% Disk -- GB avail. of -- GB BETA Terminal
```

13. To run the word count job:

```
hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar  
wordcount /user/hadoop/input /user/hadoop/output
```

```
Terminal
root@8af718d1f9f1:/# hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar wordcount /user/hadoop/input /user/hadoop/output
2024-11-18 17:10:26,812 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-11-18 17:10:26,887 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-11-18 17:10:26,887 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-11-18 17:10:27,285 INFO input.FileInputFormat: Total input files to process : 9
2024-11-18 17:10:27,325 INFO mapreduce.JobSubmitter: number of splits:9
2024-11-18 17:10:27,563 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1391096320_0001
2024-11-18 17:10:27,563 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-11-18 17:10:27,741 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-11-18 17:10:27,742 INFO mapreduce.Job: Running job: job_local1391096320_0001
2024-11-18 17:10:27,745 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-11-18 17:10:27,757 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-18 17:10:27,757 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-11-18 17:10:27,758 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2024-11-18 17:10:27,772 INFO namenode.FSEditLog: Number of transactions: 61 Total time for transactions(ms): 29 Number of transactions batched in Syncs: 10 Number of syncs: 49 SyncTimes(ms): 115
2024-11-18 17:10:27,834 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-11-18 17:10:27,835 INFO mapred.LocalJobRunner: Starting task: attempt_local1391096320_0001_m_0000000_0
2024-11-18 17:10:27,871 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-11-18 17:10:27,871 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-11-18 17:10:27,905 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2024-11-18 17:10:27,911 INFO mapred.MapTask: Processing split: hdfs://8af718d1f9f1:8020/user/hadoop/input/hadoop-policy.xml:0+11392
RAM 2.56 GB CPU 0.38% Disk -- GB avail. of -- GB BETA >_ Terminal
```

14. To display the results of the word count job:

```
hdfs dfs -cat /user/hadoop/output/part-r-00000
```

```
Terminal
root@8af718d1f9f1:/# hdfs dfs -cat /user/hadoop/output/part-r-00000
2024-11-18 17:11:43,663 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
"AS" 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
```