EX: NO:1	Installation, Configuration and Execution of Hadoop and HDFS

AIM:

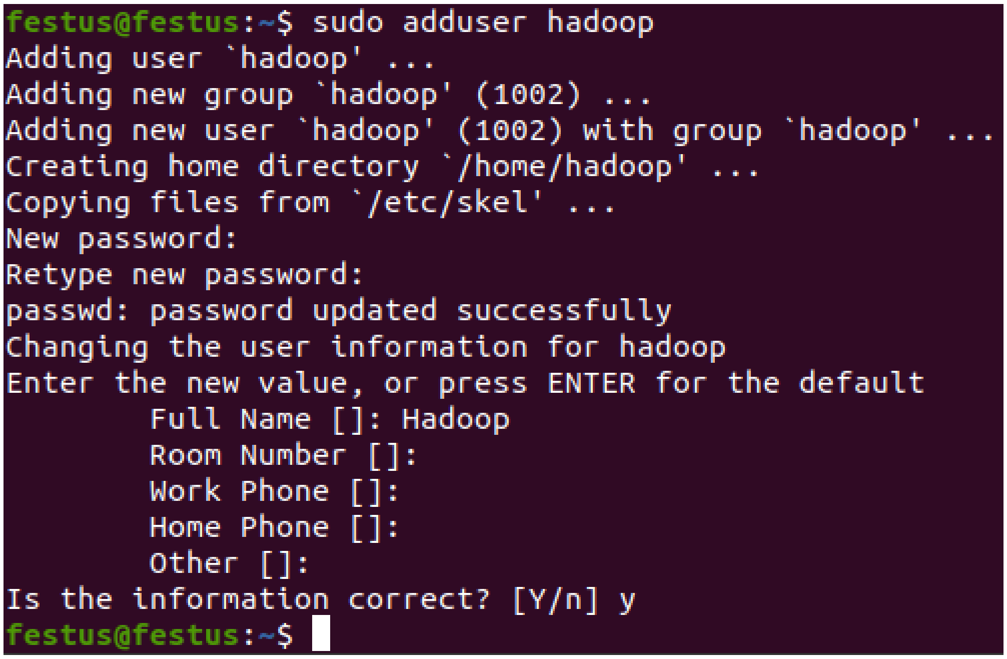
To Study how to installing Hadoop; understanding different Hadoop modes. Startup scripts, configuration files**.**

PROCEDURE

Step by Step Installing Hadoop on Ubuntu 20.04

Step 1 — Create user for Hadoop environment

sudo adduser Hadoop



```
festus@festus:~$ sudo adduser hadoop
Adding user `hadoop' ...
Adding new group `hadoop' (1002) ...
Adding new user `hadoop' (1002) with group `hadoop' ...
Creating home directory `/home/hadoop' ...
Copying files from `/etc/skel' ...
New password:
Retype new password:
passwd: password updated successfully
Changing the user information for hadoop
Enter the new value, or press ENTER for the default
        Full Name []: Hadoop
        Room Number []:
        Work Phone []:
        Home Phone []:
        Other []:
Is the information correct? [Y/n] y
festus@festus:~$ 
```

Step 2— Installing Java

The following command to update your system before initiating a new installation:

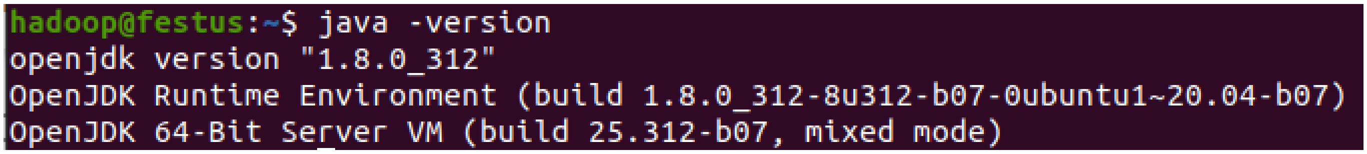
sudo apt update

Install the latest version of Java.

sudo apt install openjdk-8-jdk -y

Once installed, verify the installed version of Java with the following command:

java -version



```
hadoop@festus:~$ java -version
openjdk version "1.8.0_312"
OpenJDK Runtime Environment (build 1.8.0_312-8u312-b07-0ubuntu1~20.04-b07)
OpenJDK 64-Bit Server VM (build 25.312-b07, mixed mode)
```

Step 3: Install OpenSSH on Ubuntu

Install the OpenSSH server and client using the following command:

sudo apt install openssh-server openssh-client -y

Switch to the created user.

sudo su - hadoop

Generate public and private key pairs.

$ ssh-keygen -t rsa

Add the generated public key from id_rsa.pub to authorized_keys.

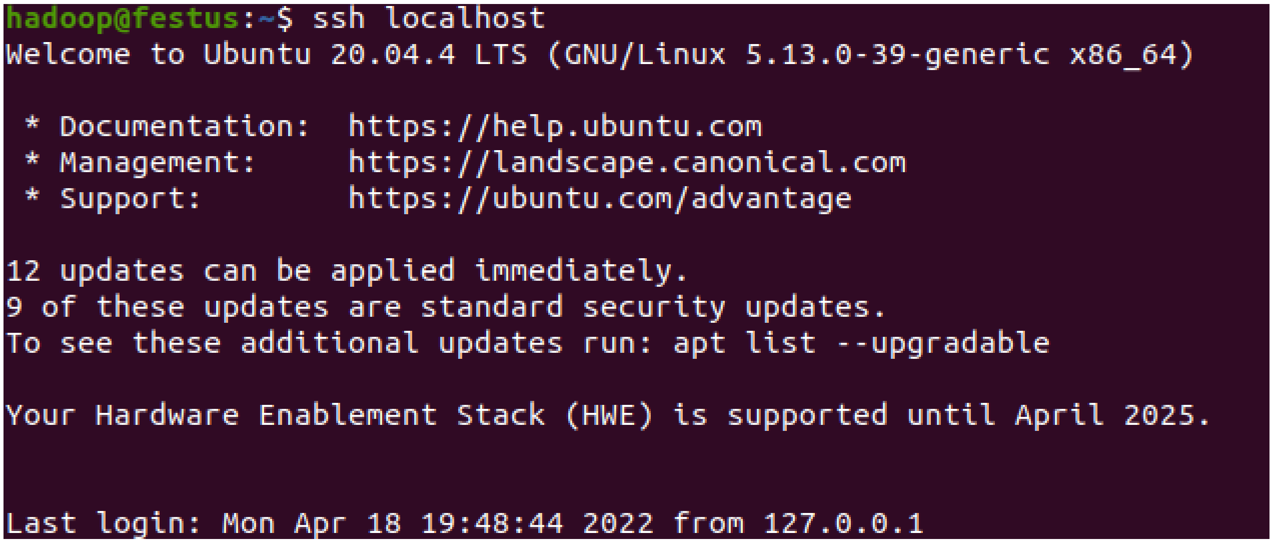
$ sudo cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys

Change the permissions of the authorized_keys file.

$ sudo chmod 640 ~/.ssh/authorized_keys

Verify if the password-less SSH is functional.

$ ssh localhost



```
hadoop@festus:~$ ssh localhost
Welcome to Ubuntu 20.04.4 LTS (GNU/Linux 5.13.0-39-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

12 updates can be applied immediately.
9 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable

Your Hardware Enablement Stack (HWE) is supported until April 2025.


Last login: Mon Apr 18 19:48:44 2022 from 127.0.0.1
```

Step 4: Install Apache Hadoop

Download the latest stable version of Hadoop.

$ wget https://downloads.apache.org/hadoop/common/hadoop-3.3.2/hadoop-3.3.2.tar.gz

Extract the downloaded file.

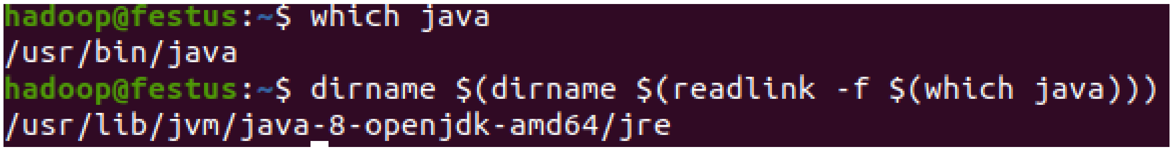
$ tar -xvzf hadoop-3.3.2.tar.gz

Rename the extracted directory as we will do by executing the below-given command:

mv hadoop-3.3.0 hadoop

Now, configure Java environment variables for setting up Hadoop. For this, we will check out the location of our “JAVA_HOME” variable:

dirname $(dirname $(readlink -f $(which java)))



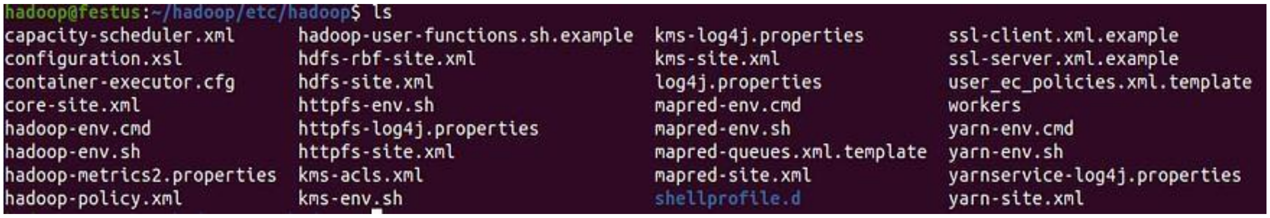
```
hadoop@festus:~$ which java
/usr/bin/java
hadoop@festus:~$ dirname $(dirname $(readlink -f $(which java)))
/usr/lib/jvm/java-8-openjdk-amd64/jre
```

Step 5: Configure Hadoop

A Hadoop environment is configured by editing a set of configuration files:

bashrc, hadoop-env.sh, core-site.xml, hdfs-site.xml, mapred-site-xml and yarn-site.xml

They can be found in the newly created hadoop folder



```
hadoop@festus:~/hadoop/etc/hadoop$ ls
capacity-scheduler.xml      hadoop-user-functions.sh.example  kms-log4j.properties        ssl-client.xml.example
configuration.xsl           hdfs-rbf-site.xml                 kms-site.xml                ssl-server.xml.example
container-executor.cfg      hdfs-site.xml                     log4j.properties            user_ec_policies.xml.template
core-site.xml               httpfs-env.sh                     mapred-env.cmd              workers
hadoop-env.cmd              httpfs-log4j.properties           mapred-env.sh               yarn-env.cmd
hadoop-env.sh               httpfs-site.xml                   mapred-queues.xml.template  yarn-env.sh
hadoop-metrics2.properties  kms-acls.xml                      mapred-site.xml             yarnservice-log4j.properties
hadoop-policy.xml           kms-env.sh                        shellprofile.d              yarn-site.xml
```

Step 5a: Configure Hadoop Environment Variables (bashrc)

Edit file ~/.bashrc to configure the Hadoop environment variables.

$ sudo nano ~/.bashrc

Add the following lines to the file. Save and close the file.

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

export HADOOP_HOME=/usr/local/hadoop

export HADOOP_INSTALL=$HADOOP_HOME

export HADOOP_MAPRED_HOME=$HADOOP_HOME

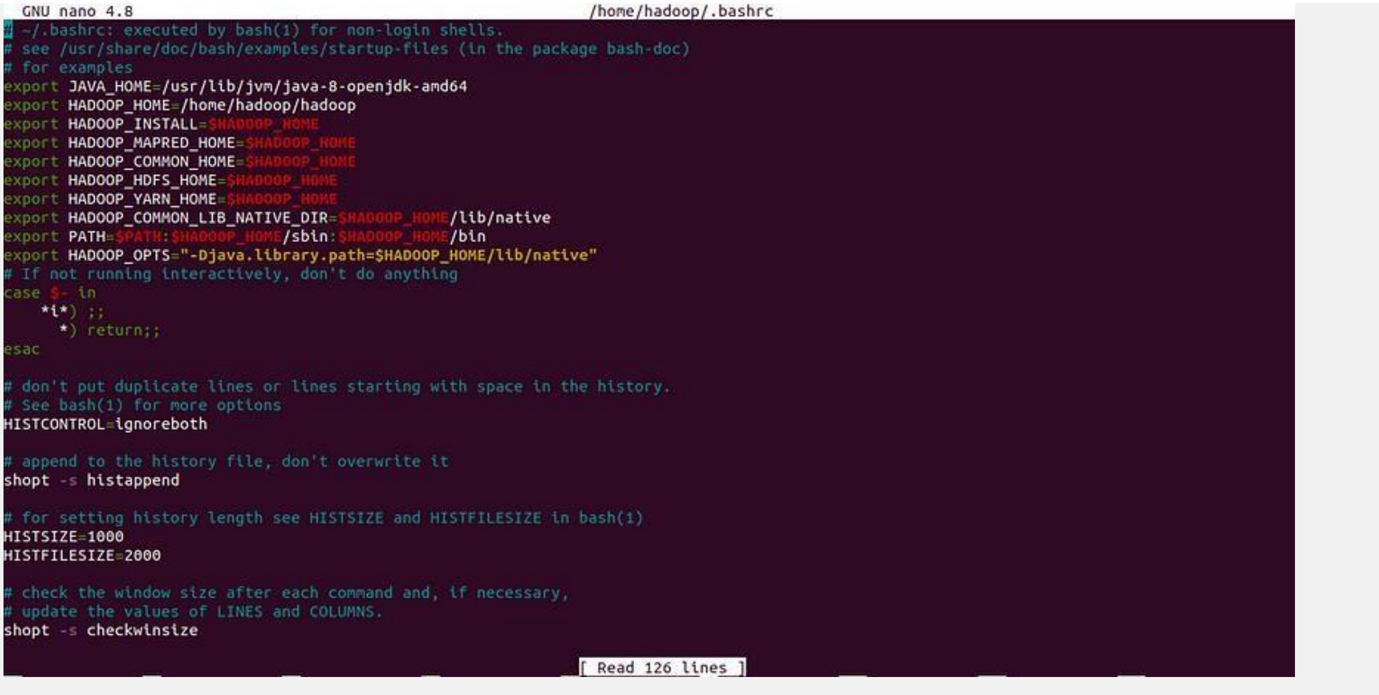
export HADOOP_COMMON_HOME=$HADOOP_HOME

export HADOOP_HDFS_HOME=$HADOOP_HOME

export YARN_HOME=$HADOOP_HOME

export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native

export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"



```
  GNU nano 4.8                                     /home/hadoop/.bashrc
# ~/.bashrc: executed by bash(1) for non-login shells.
# see /usr/share/doc/bash/examples/startup-files (in the package bash-doc)
# for examples
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export HADOOP_HOME=/home/hadoop/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export HADOOP_YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
# If not running interactively, don't do anything
case $- in
    *i*) ;;
      *) return;;
esac

# don't put duplicate lines or lines starting with space in the history.
# See bash(1) for more options
HISTCONTROL=ignoreboth

# append to the history file, don't overwrite it
shopt -s histappend

# for setting history length see HISTSIZE and HISTFILESIZE in bash(1)
HISTSIZE=1000
HISTFILESIZE=2000

# check the window size after each command and, if necessary,
# update the values of LINES and COLUMNS.
shopt -s checkwinsize

                                                     [ Read 126 lines ]
```

Activate the environment variables.

$ source ~/.bashrc

Step 5b: Edit hadoop-env.sh File

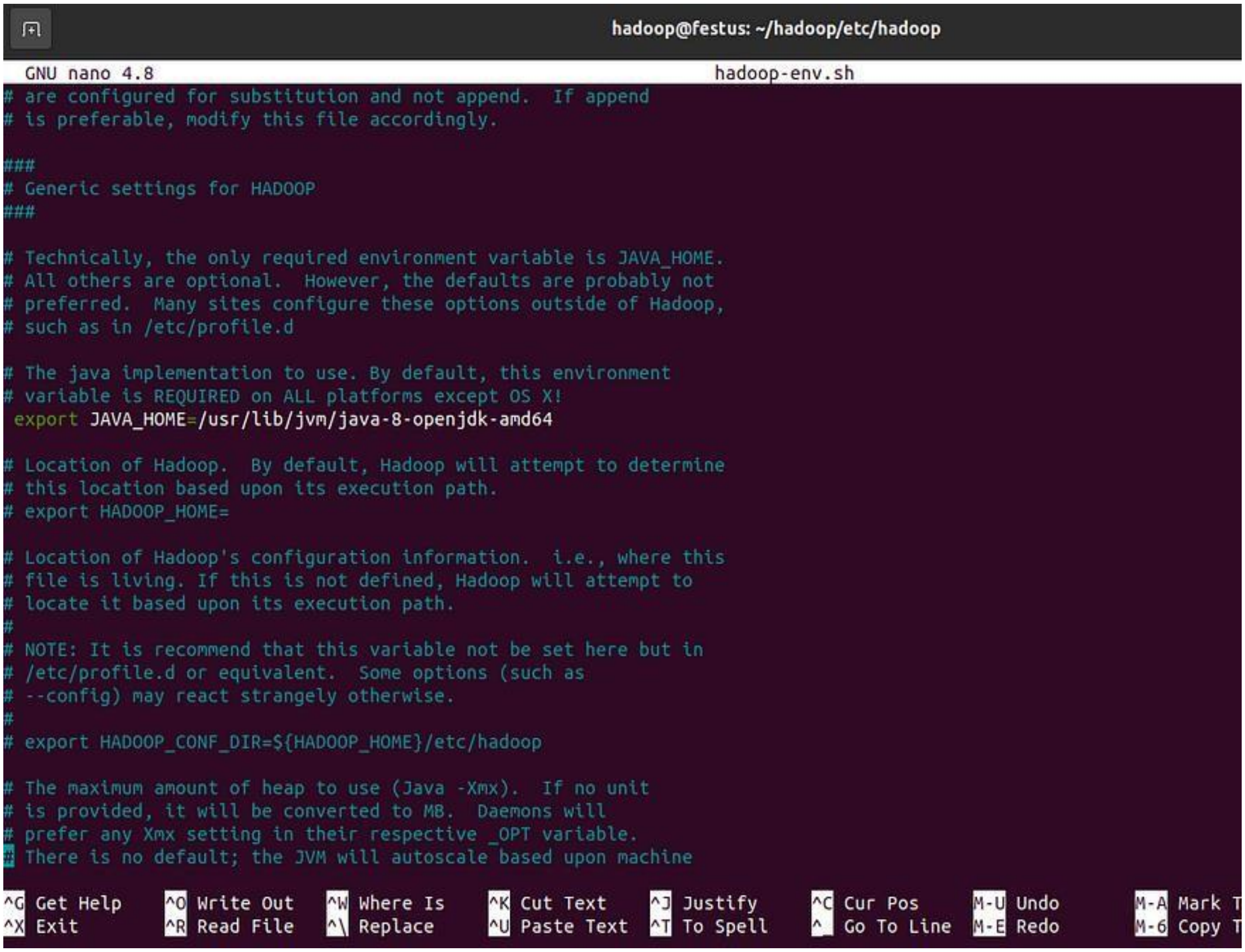
The hadoop-env.sh file serves as a master file to configure YARN, HDFS, MapReduce, and Hadoop-related project settings. When setting up a single node Hadoop cluster, you need to define which Java implementation is to be utilized. Use the previously created $HADOOP_HOME variable to access the hadoop-env.sh file:

sudo nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh

Uncomment the $JAVA_HOME variable (i.e., remove the # sign) and add the full path to the OpenJDK installation on your system.

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

The path needs to match the location of the Java installation on your system.

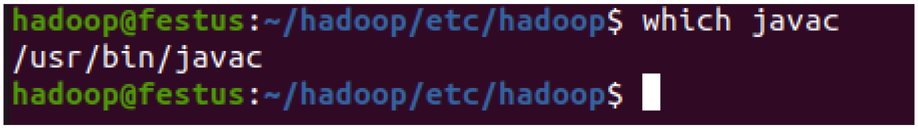


```
# are configured for substitution and not append.  If append
# is preferable, modify this file accordingly.

###
# Generic settings for HADOOP
###

# Technically, the only required environment variable is JAVA_HOME.
# All others are optional.  However, the defaults are probably not
# preferred.  Many sites configure these options outside of Hadoop,
# such as in /etc/profile.d

# The java implementation to use. By default, this environment
# variable is REQUIRED on ALL platforms except OS X!
 export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

# Location of Hadoop.  By default, Hadoop will attempt to determine
# this location based upon its execution path.
# export HADOOP_HOME=

# Location of Hadoop's configuration information.  i.e., where this
# file is living. If this is not defined, Hadoop will attempt to
# locate it based upon its execution path.
#
# NOTE: It is recommend that this variable not be set here but in
# /etc/profile.d or equivalent.  Some options (such as
# --config) may react strangely otherwise.
#
# export HADOOP_CONF_DIR=${HADOOP_HOME}/etc/hadoop

# The maximum amount of heap to use (Java -Xmx).  If no unit
# is provided, it will be converted to MB.  Daemons will
# prefer any Xmx setting in their respective _OPT variable.
# There is no default; the JVM will autoscale based upon machine
```

To locate the correct Java path, run the following command in your terminal window:

which javac

The resulting output provides the path to the Java binary directory.



```
hadoop@festus:~/hadoop/etc/hadoop$ which javac
/usr/bin/javac
hadoop@festus:~/hadoop/etc/hadoop$ 
```

Use the provided path to find the OpenJDK directory with the following command:

readlink -f /usr/bin/javac

The section of the path just before the /bin/javac directory needs to be assigned to the $JAVA_HOME variable.

Step 5c: Edit core-site.xml File

The core-site.xml file defines HDFS and Hadoop core properties.

To set up Hadoop in a pseudo-distributed mode, you need to specify the URL for your NameNode, and the temporary directory Hadoop uses for the map and reduce process.

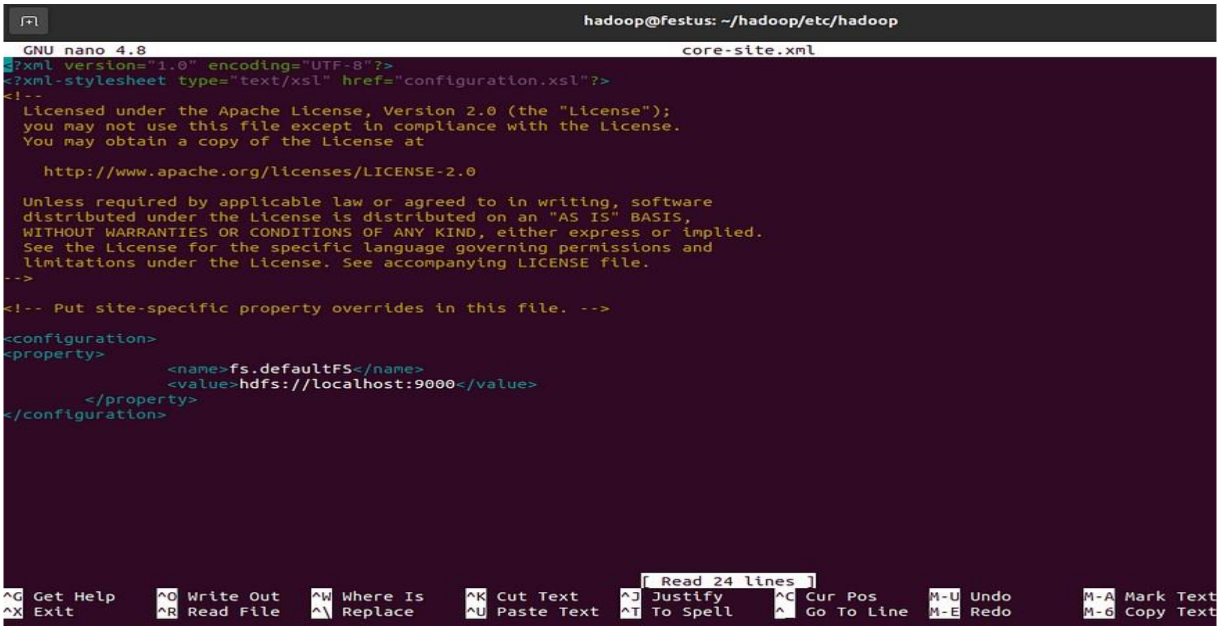
Open the core-site.xml file in a text editor:

```
sudo nano $HADOOP_HOME/etc/hadoop/core-site.xml
```

Add the following configuration to override the default values for the temporary directory and add your HDFS URL to replace the default local file system setting:

```
<configuration>
 <property>
              <name>fs.defaultFS</name>
              <value>hdfs://localhost:9000</value>
       </property>
 </configuration>
```

This example uses values specific to the local system. The data needs to be consistent throughout the configuration process.



```
hadoop@festus: ~/hadoop/etc/hadoop
  GNU nano 4.8                                         core-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
                <name>fs.defaultFS</name>
                <value>hdfs://localhost:9000</value>
        </property>
</configuration>

                                          [ Read 24 lines ]
^G Get Help    ^O Write Out   ^W Where Is    ^K Cut Text    ^J Justify     ^C Cur Pos     M-U Undo       M-A Mark Text
^X Exit        ^R Read File   ^\ Replace     ^U Paste Text  ^T To Spell    ^  Go To Line  M-E Redo       M-6 Copy Text
```

Step 5d: Edit hdfs-site.xml File

The properties in the hdfs-site.xml file govern the location for storing node metadata, fsimage file, and edit log file. Configure the file by defining the NameNode and DataNode storage directories. In this “hdfs-site.xml” file, we will change the directory path of “datanode” and “namenode”: Additionally, the default dfs.replication value of 3 needs to be changed to 1 to match the single node setup.

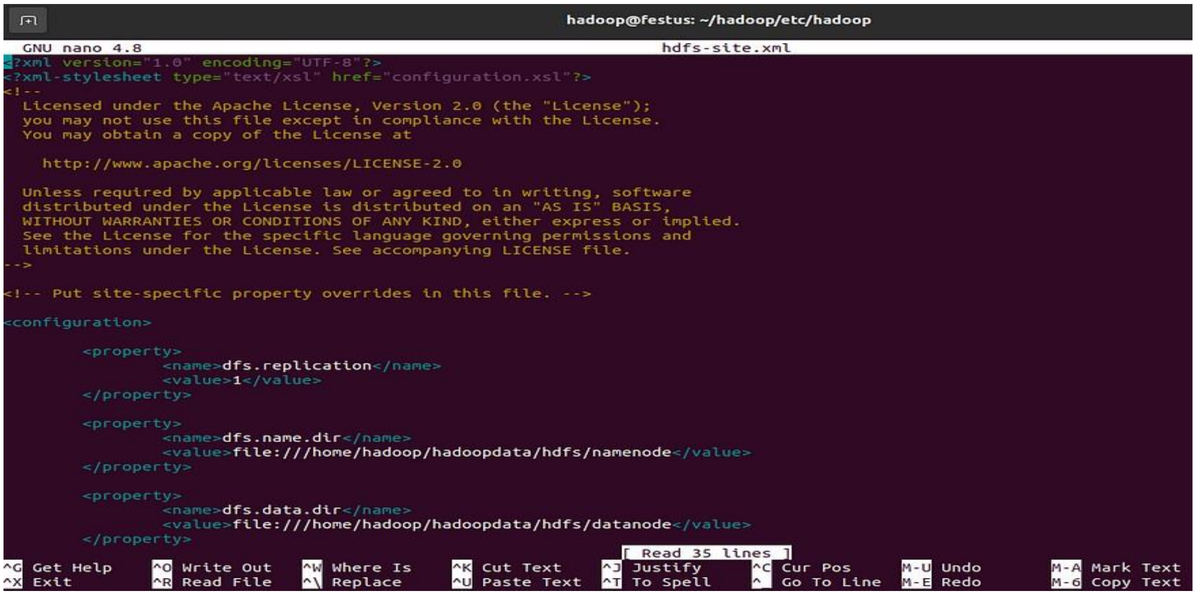
Use the following command to open the hdfs-site.xml file for editing:

```
sudo nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml
```

Add the following configuration to the file and, if needed, adjust the NameNode and DataNode directories to your custom locations:

```
<configuration>
<property>
            <name>dfs.replication</name>
            <value>1</value>
      </property>
<property>
            <name>dfs.name.dir</name>
            <value>file:///home/hadoop/hadoopdata/hdfs/namenode</value>
      </property>
<property>
            <name>dfs.data.dir</name>
            <value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>
      </property>
</configuration>
```

If necessary, create the specific directories you defined for the dfs.data.dir value.



```
  GNU nano 4.8                                               hdfs-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>

        <property>
                <name>dfs.replication</name>
                <value>1</value>
        </property>

        <property>
                <name>dfs.name.dir</name>
                <value>file:///home/hadoop/hadoopdata/hdfs/namenode</value>
        </property>

        <property>
                <name>dfs.data.dir</name>
                <value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>
        </property>
                                                            [ Read 35 lines ]
```

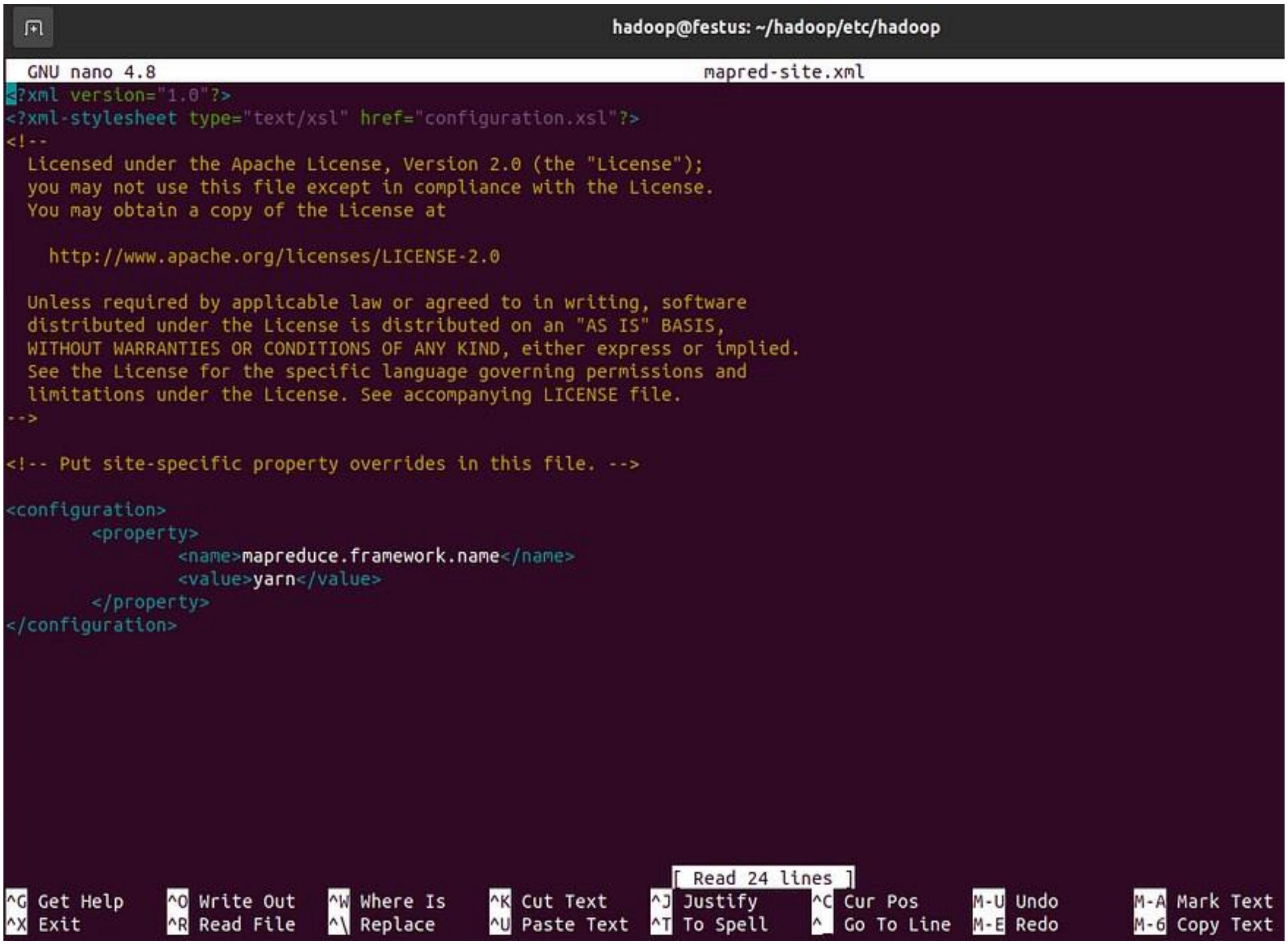
Step 5e: Edit mapred-site.xml File

Use the following command to access the mapred-site.xml file and define MapReduce values:

sudo nano $HADOOP_HOME/etc/hadoop/mapred-site.xml

Add the following configuration to change the default MapReduce framework name value to yarn:

```
<configuration>
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
</configuration>
```



hadoop@festus: ~/hadoop/etc/hadoop

GNU nano 4.8 mapred-site.xml

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
        <property>
                <name>mapreduce.framework.name</name>
                <value>yarn</value>
        </property>
</configuration>
```

[Read 24 lines]

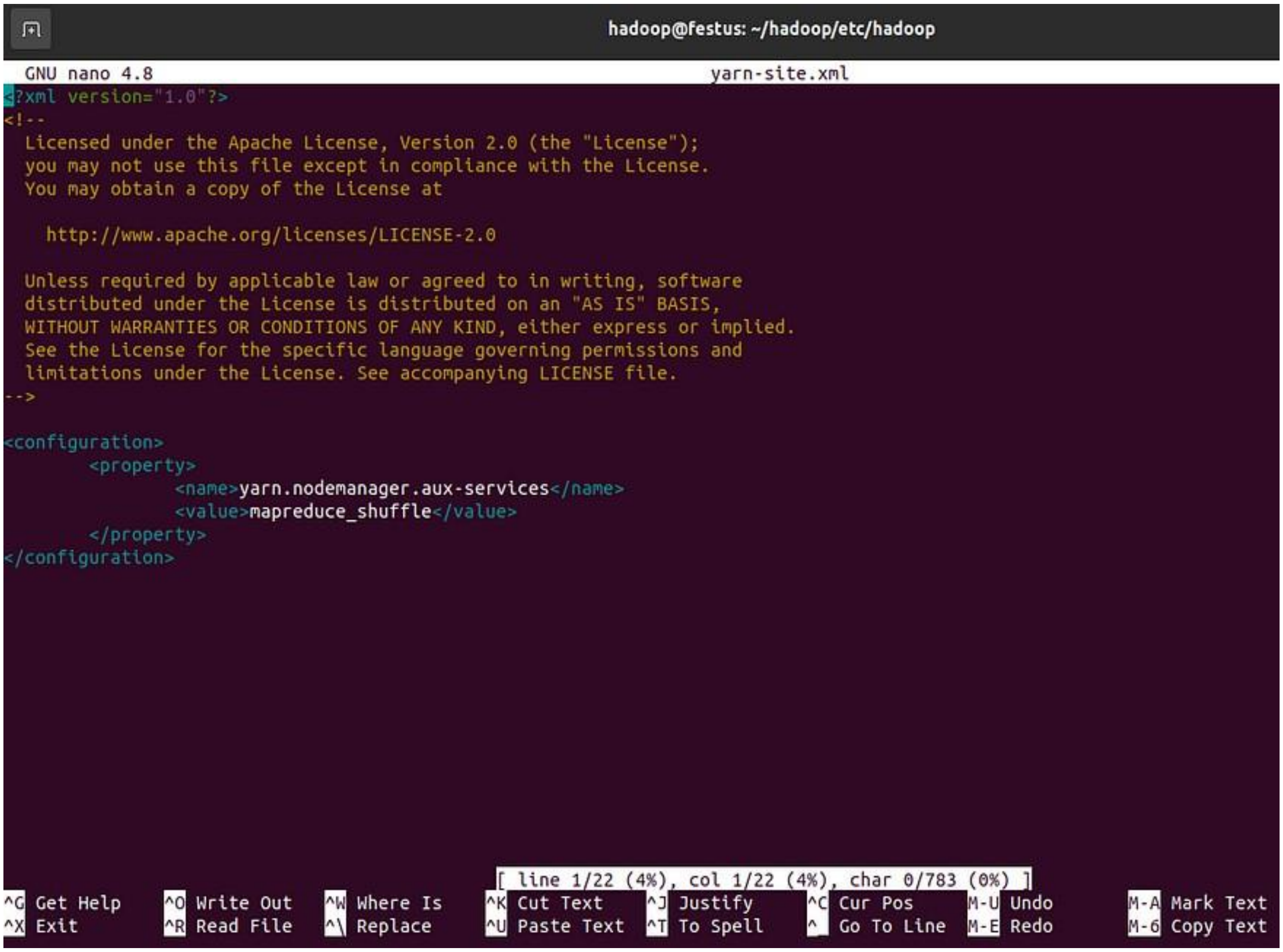
Step 5f: Edit yarn-site.xml File

The yarn-site.xml file is used to define settings relevant to YARN. It contains configurations for the Node Manager, Resource Manager, Containers, and Application Master. Open the yarn-site.xml file in a text editor:

sudo nano $HADOOP_HOME/etc/hadoop/yarn-site.xml

Append the following configuration to the file:

```
<configuration>
	<property>
		<name>yarn.nodemanager.aux-services</name>
		<value>mapreduce_shuffle</value>
	</property>
</configuration>
```



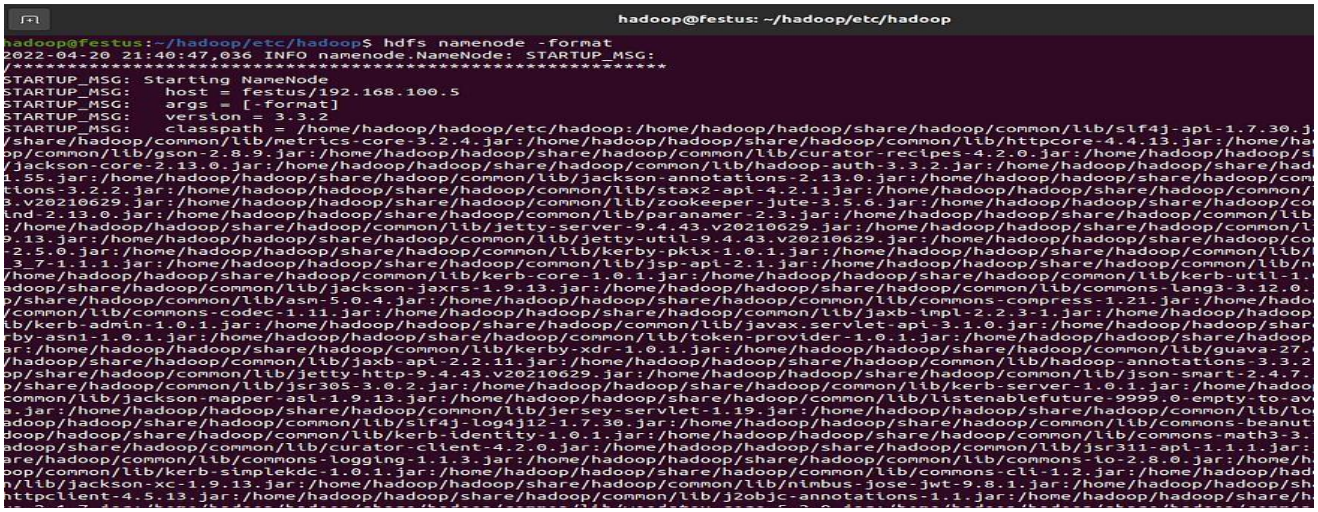
hadoop@festus: ~/hadoop/etc/hadoop

```
  GNU nano 4.8                                                   yarn-site.xml
<?xml version="1.0"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<configuration>
        <property>
                <name>yarn.nodemanager.aux-services</name>
                <value>mapreduce_shuffle</value>
        </property>
</configuration>

                                              [ line 1/22 (4%), col 1/22 (4%), char 0/783 (0%) ]
^G Get Help    ^O Write Out   ^W Where Is    ^K Cut Text    ^J Justify     ^C Cur Pos     M-U Undo       M-A Mark Text
^X Exit        ^R Read File   ^\ Replace     ^U Paste Text  ^T To Spell    ^_ Go To Line  M-E Redo       M-6 Copy Text
```

Step 5g. Format HDFS NameNode

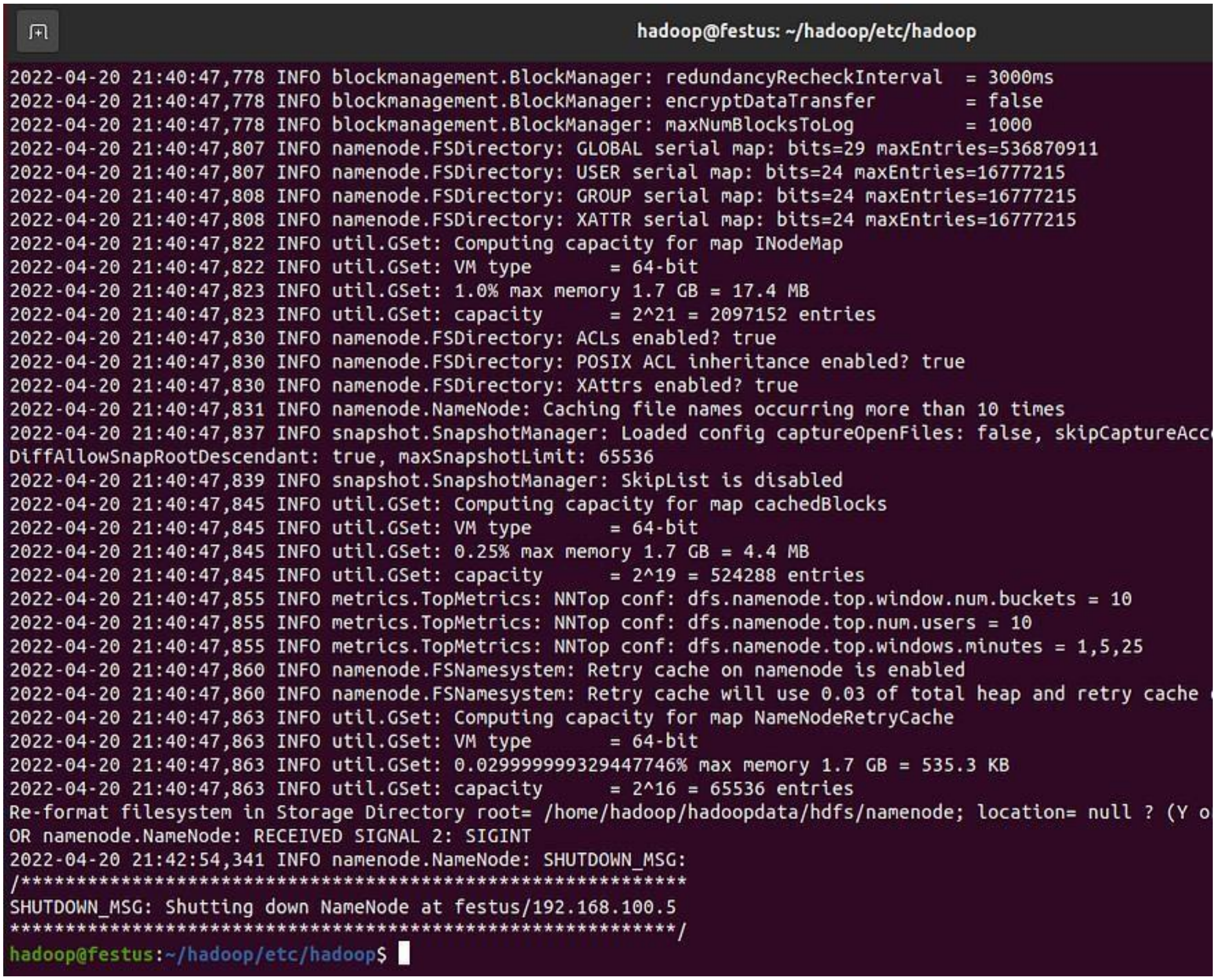
It is important to format the NameNode before starting Hadoop services for the first time:

hdfs namenode -format



```
hadoop@festus:~/hadoop/etc/hadoop$ hdfs namenode -format
2022-04-20 21:40:47,036 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = festus/192.168.100.5
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.3.2
STARTUP_MSG:   classpath = /home/hadoop/hadoop/etc/hadoop:/home/hadoop/hadoop/share/hadoop/common/lib/slf4j-api-1.7.30.j
/share/hadoop/common/lib/metrics-core-3.2.4.jar:/home/hadoop/hadoop/share/hadoop/common/lib/httpcore-4.4.13.jar:/home/ha
op/common/lib/gson-2.8.9.jar:/home/hadoop/hadoop/share/hadoop/common/lib/curator-recipes-4.2.0.jar:/home/hadoop/hadoop/s
/jackson-core-2.13.0.jar:/home/hadoop/hadoop/share/hadoop/common/lib/hadoop-auth-3.3.2.jar:/home/hadoop/hadoop/share/had
1.55.jar:/home/hadoop/hadoop/share/hadoop/common/lib/jackson-annotations-2.13.0.jar:/home/hadoop/hadoop/share/hadoop/com
tions-3.2.2.jar:/home/hadoop/hadoop/share/hadoop/common/lib/stax2-api-4.2.1.jar:/home/hadoop/hadoop/share/hadoop/common/
3.v20210629.jar:/home/hadoop/hadoop/share/hadoop/common/lib/zookeeper-jute-3.5.6.jar:/home/hadoop/hadoop/share/hadoop/co
ind-2.13.0.jar:/home/hadoop/hadoop/share/hadoop/common/lib/paranamer-2.3.jar:/home/hadoop/hadoop/share/hadoop/common/lib
:/home/hadoop/hadoop/share/hadoop/common/lib/jetty-server-9.4.43.v20210629.jar:/home/hadoop/hadoop/share/hadoop/common/l
9.13.jar:/home/hadoop/hadoop/share/hadoop/common/lib/jetty-util-9.4.43.v20210629.jar:/home/hadoop/hadoop/share/hadoop/co
-2.5.0.jar:/home/hadoop/hadoop/share/hadoop/common/lib/kerby-pkix-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/
_3_7-1.1.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/jsp-api-2.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/n
/home/hadoop/hadoop/share/hadoop/common/lib/kerb-core-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/kerb-util-1.
adoop/share/hadoop/common/lib/jackson-jaxrs-1.9.13.jar:/home/hadoop/hadoop/share/hadoop/common/lib/commons-lang3-3.12.0.
p/share/hadoop/common/lib/asm-5.0.4.jar:/home/hadoop/hadoop/share/hadoop/common/lib/commons-compress-1.21.jar:/home/hadoo
/common/lib/commons-codec-1.11.jar:/home/hadoop/hadoop/share/hadoop/common/lib/jaxb-impl-2.2.3-1.jar:/home/hadoop/hadoop
ib/kerb-admin-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/javax.servlet-api-3.1.0.jar:/home/hadoop/hadoop/shar
rby-asn1-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/token-provider-1.0.1.jar:/home/hadoop/hadoop/share/hadoop
ar:/home/hadoop/hadoop/share/hadoop/common/lib/kerby-xdr-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/guava-27.
/hadoop/share/hadoop/common/lib/jaxb-api-2.2.11.jar:/home/hadoop/hadoop/share/hadoop/common/lib/hadoop-annotations-3.3.2
op/share/hadoop/common/lib/jetty-http-9.4.43.v20210629.jar:/home/hadoop/hadoop/share/hadoop/common/lib/json-smart-2.4.7.
p/share/hadoop/common/lib/jsr305-3.0.2.jar:/home/hadoop/hadoop/share/hadoop/common/lib/kerb-server-1.0.1.jar:/home/hadoop
common/lib/jackson-mapper-asl-1.9.13.jar:/home/hadoop/hadoop/share/hadoop/common/lib/listenablefuture-9999.0-empty-to-av
a.jar:/home/hadoop/hadoop/share/hadoop/common/lib/jersey-servlet-1.19.jar:/home/hadoop/hadoop/share/hadoop/common/lib/lo
adoop/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.30.jar:/home/hadoop/hadoop/share/hadoop/common/lib/commons-beanut
doop/hadoop/share/hadoop/common/lib/kerb-identity-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/commons-math3-3.
adoop/share/hadoop/common/lib/curator-client-4.2.0.jar:/home/hadoop/hadoop/share/hadoop/common/lib/jsr311-api-1.1.1.jar:
are/hadoop/common/lib/commons-logging-1.1.3.jar:/home/hadoop/hadoop/share/hadoop/common/lib/commons-io-2.8.0.jar:/home/h
oop/common/lib/kerb-simplekdc-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/commons-cli-1.2.jar:/home/hadoop/had
n/lib/jackson-xc-1.9.13.jar:/home/hadoop/hadoop/share/hadoop/common/lib/nimbus-jose-jwt-9.8.1.jar:/home/hadoop/hadoop/sh
httpclient-4.5.13.jar:/home/hadoop/hadoop/share/hadoop/common/lib/j2objc-annotations-1.1.jar:/home/hadoop/hadoop/share/h
```

The shutdown notification signifies the end of the NameNode format process.

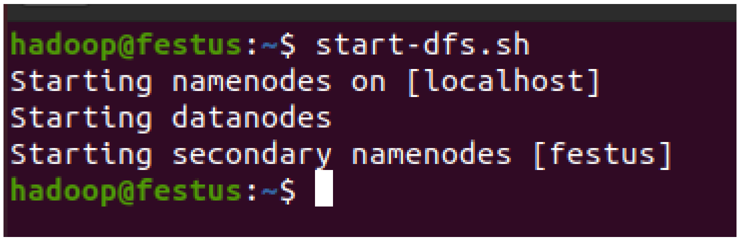


```
2022-04-20 21:40:47,778 INFO blockmanagement.BlockManager: redundancyRecheckInterval  = 3000ms
2022-04-20 21:40:47,778 INFO blockmanagement.BlockManager: encryptDataTransfer        = false
2022-04-20 21:40:47,778 INFO blockmanagement.BlockManager: maxNumBlocksToLog          = 1000
2022-04-20 21:40:47,807 INFO namenode.FSDirectory: GLOBAL serial map: bits=29 maxEntries=536870911
2022-04-20 21:40:47,807 INFO namenode.FSDirectory: USER serial map: bits=24 maxEntries=16777215
2022-04-20 21:40:47,808 INFO namenode.FSDirectory: GROUP serial map: bits=24 maxEntries=16777215
2022-04-20 21:40:47,808 INFO namenode.FSDirectory: XATTR serial map: bits=24 maxEntries=16777215
2022-04-20 21:40:47,822 INFO util.GSet: Computing capacity for map INodeMap
2022-04-20 21:40:47,822 INFO util.GSet: VM type       = 64-bit
2022-04-20 21:40:47,823 INFO util.GSet: 1.0% max memory 1.7 GB = 17.4 MB
2022-04-20 21:40:47,823 INFO util.GSet: capacity      = 2^21 = 2097152 entries
2022-04-20 21:40:47,830 INFO namenode.FSDirectory: ACLs enabled? true
2022-04-20 21:40:47,830 INFO namenode.FSDirectory: POSIX ACL inheritance enabled? true
2022-04-20 21:40:47,830 INFO namenode.FSDirectory: XAttrs enabled? true
2022-04-20 21:40:47,831 INFO namenode.NameNode: Caching file names occurring more than 10 times
2022-04-20 21:40:47,837 INFO snapshot.SnapshotManager: Loaded config captureOpenFiles: false, skipCaptureAcc
DiffAllowSnapRootDescendant: true, maxSnapshotLimit: 65536
2022-04-20 21:40:47,839 INFO snapshot.SnapshotManager: SkipList is disabled
2022-04-20 21:40:47,845 INFO util.GSet: Computing capacity for map cachedBlocks
2022-04-20 21:40:47,845 INFO util.GSet: VM type       = 64-bit
2022-04-20 21:40:47,845 INFO util.GSet: 0.25% max memory 1.7 GB = 4.4 MB
2022-04-20 21:40:47,845 INFO util.GSet: capacity      = 2^19 = 524288 entries
2022-04-20 21:40:47,855 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
2022-04-20 21:40:47,855 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2022-04-20 21:40:47,855 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2022-04-20 21:40:47,860 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2022-04-20 21:40:47,860 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache 
2022-04-20 21:40:47,863 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2022-04-20 21:40:47,863 INFO util.GSet: VM type       = 64-bit
2022-04-20 21:40:47,863 INFO util.GSet: 0.029999999329447746% max memory 1.7 GB = 535.3 KB
2022-04-20 21:40:47,863 INFO util.GSet: capacity      = 2^16 = 65536 entries
Re-format filesystem in Storage Directory root= /home/hadoop/hadoopdata/hdfs/namenode; location= null ? (Y o
OR namenode.NameNode: RECEIVED SIGNAL 2: SIGINT
2022-04-20 21:42:54,341 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at festus/192.168.100.5
************************************************************/
hadoop@festus:~/hadoop/etc/hadoop$ 
```

Step 6: Start Hadoop Cluster

Start the NameNode and DataNode.

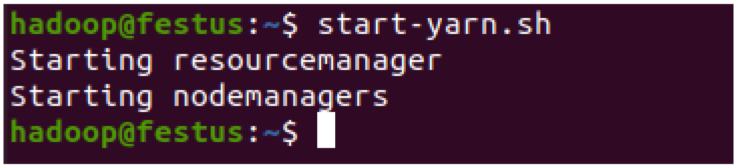
$ start-dfs.sh



```
hadoop@festus:~$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [festus]
hadoop@festus:~$ 
```

Start the YARN resource and node managers.

$ start-yarn.sh

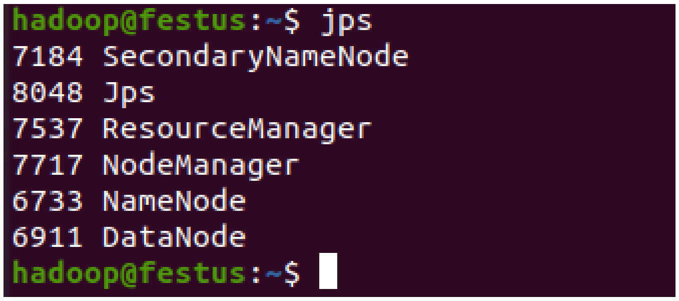


```
hadoop@festus:~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
hadoop@festus:~$ 
```

Verify all the running components.

$ jps

The system takes a few moments to initiate the necessary nodes. If everything is working as intended, the resulting list of running Java processes contains all the HDFS and YARN daemons.



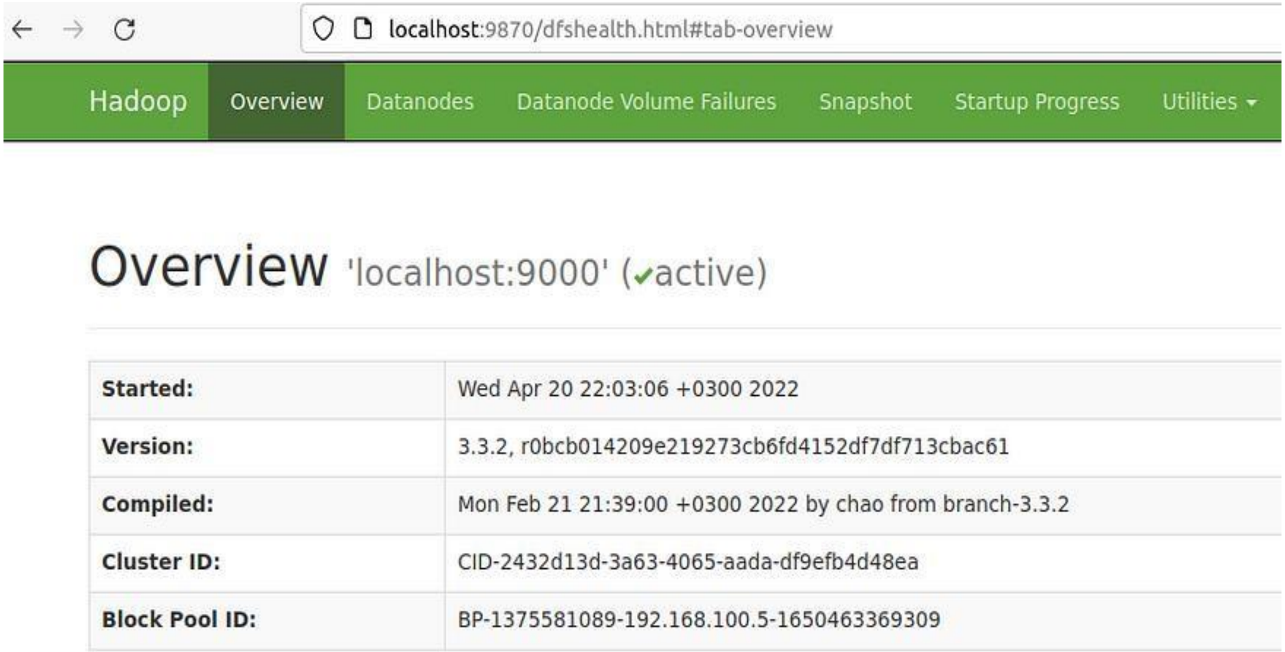
```
hadoop@festus:~$ jps
7184 SecondaryNameNode
8048 Jps
7537 ResourceManager
7717 NodeManager
6733 NameNode
6911 DataNode
hadoop@festus:~$ 
```

Step 7: Access Hadoop UI from Browser

Use your preferred browser and navigate to your localhost URL or IP. The default port number 9870 gives you access to the Hadoop NameNode UI:

http://localhost:9870

The NameNode user interface provides a comprehensive overview of the entire cluster

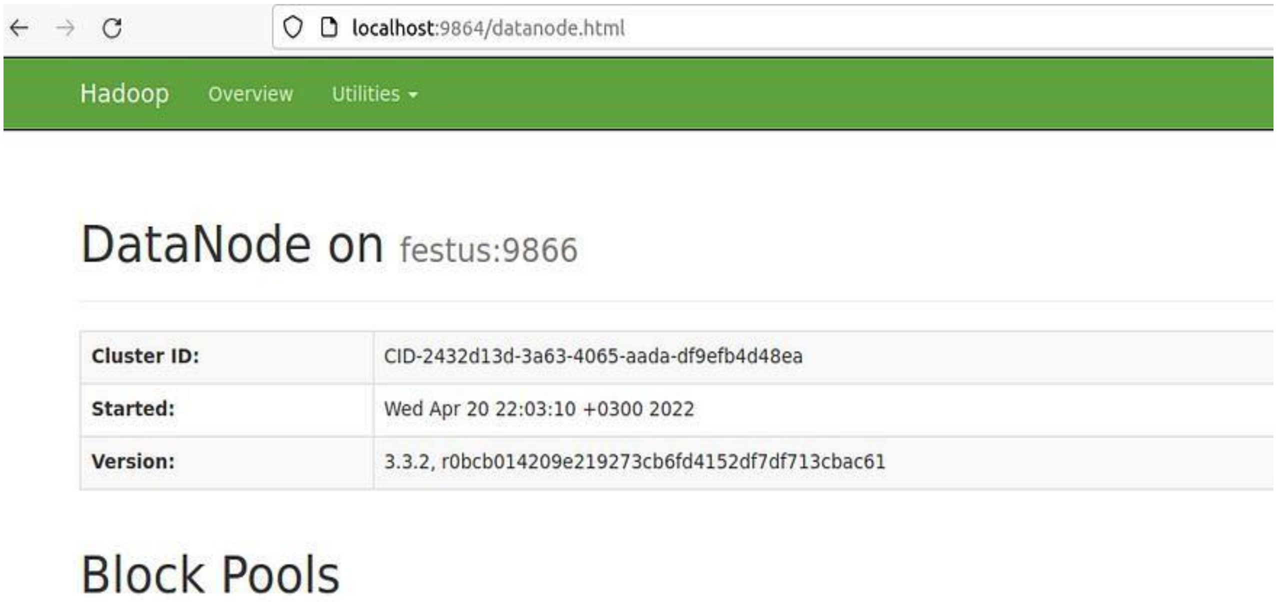


Overview 'localhost:9000' (✓active)

Started:	Wed Apr 20 22:03:06 +0300 2022
Version:	3.3.2, r0bcb014209e219273cb6fd4152df7df713cbac61
Compiled:	Mon Feb 21 21:39:00 +0300 2022 by chao from branch-3.3.2
Cluster ID:	CID-2432d13d-3a63-4065-aada-df9efb4d48ea
Block Pool ID:	BP-1375581089-192.168.100.5-1650463369309

The default port 9864 is used to access individual DataNodes directly from your browser:

http://localhost:9864



DataNode on festus:9866

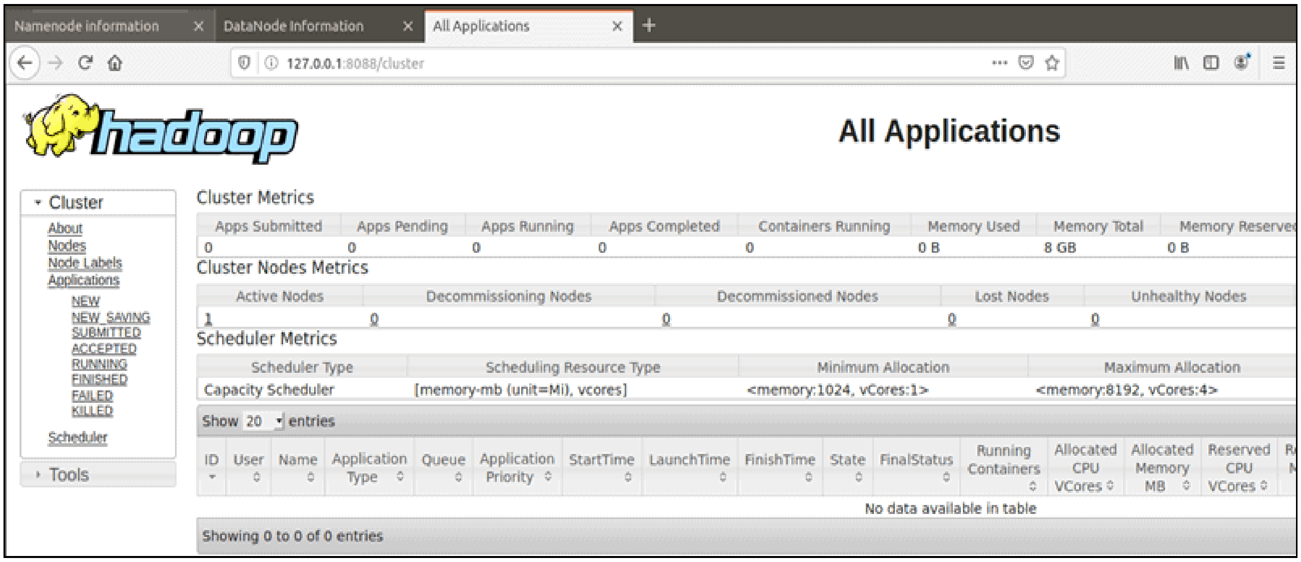
Cluster ID:	CID-2432d13d-3a63-4065-aada-df9efb4d48ea
Started:	Wed Apr 20 22:03:10 +0300 2022
Version:	3.3.2, r0bcb014209e219273cb6fd4152df7df713cbac61

Block Pools

The YARN Resource Manager is accessible on port 8088:

http://localhost:8088

The Resource Manager is an invaluable tool that allows you to monitor all running processes in your Hadoop cluster.



All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserve
0	0	0	0	0	0 B	8 GB	0 B

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes
1	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reserved CPU VCores	R M
No data available in table															

Showing 0 to 0 of 0 entries

Result