

# ASSIGNMENT 2

## Common configuration:

### 1. Node1 and Node2 hadoop-env.sh



```
GNU nano 6.2 hadoop-env.sh
# Specify the JVM options to be used when starting the HDFS Mover.
# These options will be appended to the options specified as HADOOP_OPTS
# and therefore may override any similar flags set in HADOOP_OPTS
#
# export HDFS_MOVER_OPTS=""

###
# Router-based HDFS Federation specific parameters
# Specify the JVM options to be used when starting the RBF Routers.
# These options will be appended to the options specified as HADOOP_OPTS
# and therefore may override any similar flags set in HADOOP_OPTS
#
# export HDFS_DFSROUTER_OPTS=""


###
# HDFS StorageContainerManager specific parameters
###
# Specify the JVM options to be used when starting the HDFS Storage Container Manager.
# These options will be appended to the options specified as HADOOP_OPTS
# and therefore may override any similar flags set in HADOOP_OPTS
#
# export HDFS_STORAGECONTAINERMANAGER_OPTS=""

###
# Advanced Users Only!
###

# When building Hadoop, one can add the class paths to the commands
# via this special env var:
# export HADOOP_ENABLE_BUILD_PATHS="true"

#
# To prevent accidents, shell commands be (superficially) locked
# to only allow certain users to execute certain subcommands.
# It uses the format of (command)_(subcommand)_USER.
#
# For example, to limit who can execute the namenode command,
# export HDFS_NAMENODE_USER=hdfs
#JAVA
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64/
export JRE_HOME=/usr/lib/jvm/java-8-openjdk-amd64/jre
#Hadoop Environment Variables
export HADOOP_HOME=/usr/local/hadoop
export HADOOP_CONF_DIR=/usr/local/hadoop/etc/hadoop
export HADOOP_LOG_DIR=$HADOOP_HOME/logs
export HDFS_NAMENODE_USER=hdfs
export HDFS_DATANODE_USER=hdfs
export HDFS_SECONDARYNAMENODE_USER=hdfs
export YARN_RESOURCEMANAGER_USER=hdfs
export YARN_NODEMANAGER_USER=hdfs
# Add Hadoop bin/ directory to PATH
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```

## 2. Node1 and node2 core-site.xml



```
GNU nano 6.2 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://manager:9000</value>
</property>
</configuration>
```

### 3. Manager yarn-site.xml



```
GNU nano 6.2 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

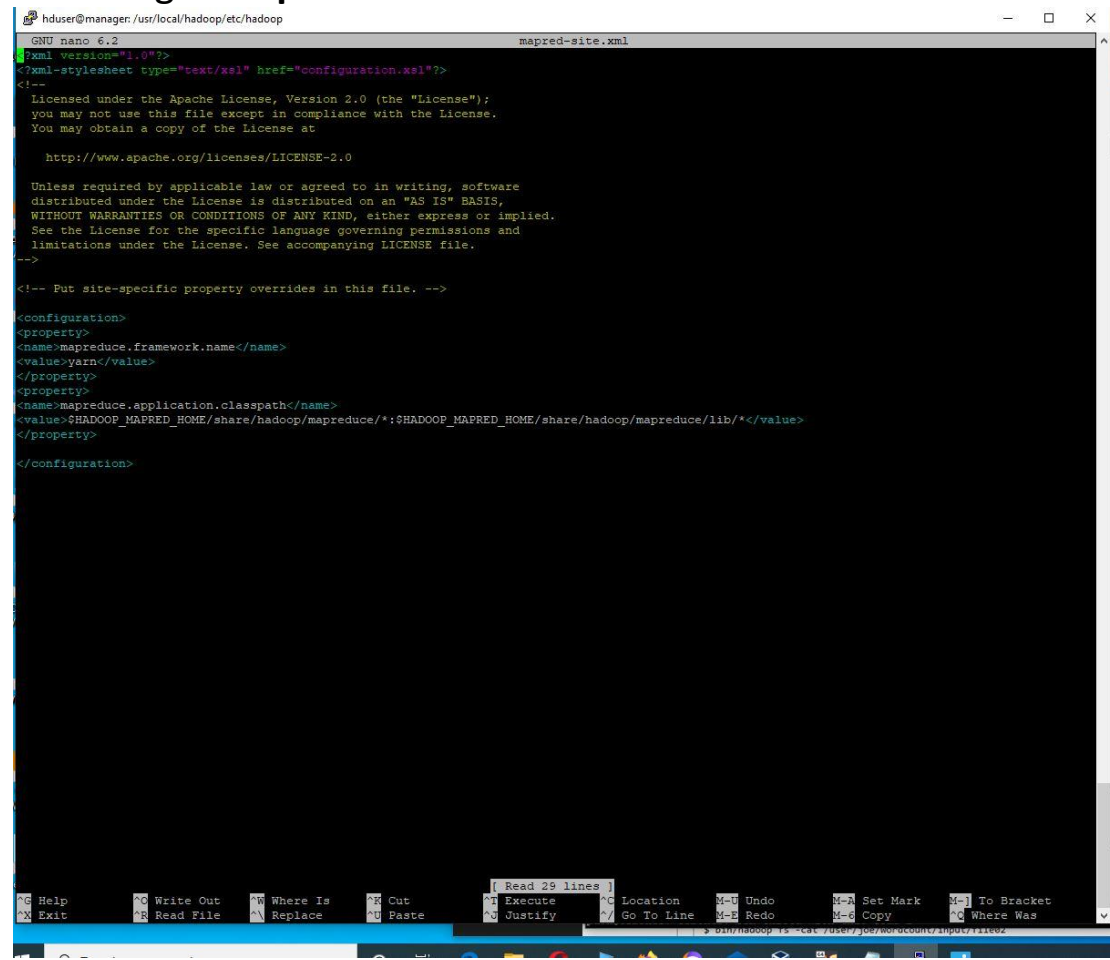
Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License.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

<property>
<name>yarn.resourcemanager.hostname</name>
<value>manager</value>
</property>
</configuration>
```

Help Write Out Where Is Cut Read 21 lines Execute Location M-U Undo M-A Set Mark M-I To Bracket  
Exit Read File Replace Paste Justify Go To Line M-E Redo M-C Copy M-W Where Was

\$ sshhdadoop15 -tcd /user/journalaccount/input/r11802

## 4. Manager mared-site.xml



```
GNU nano 6.2 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

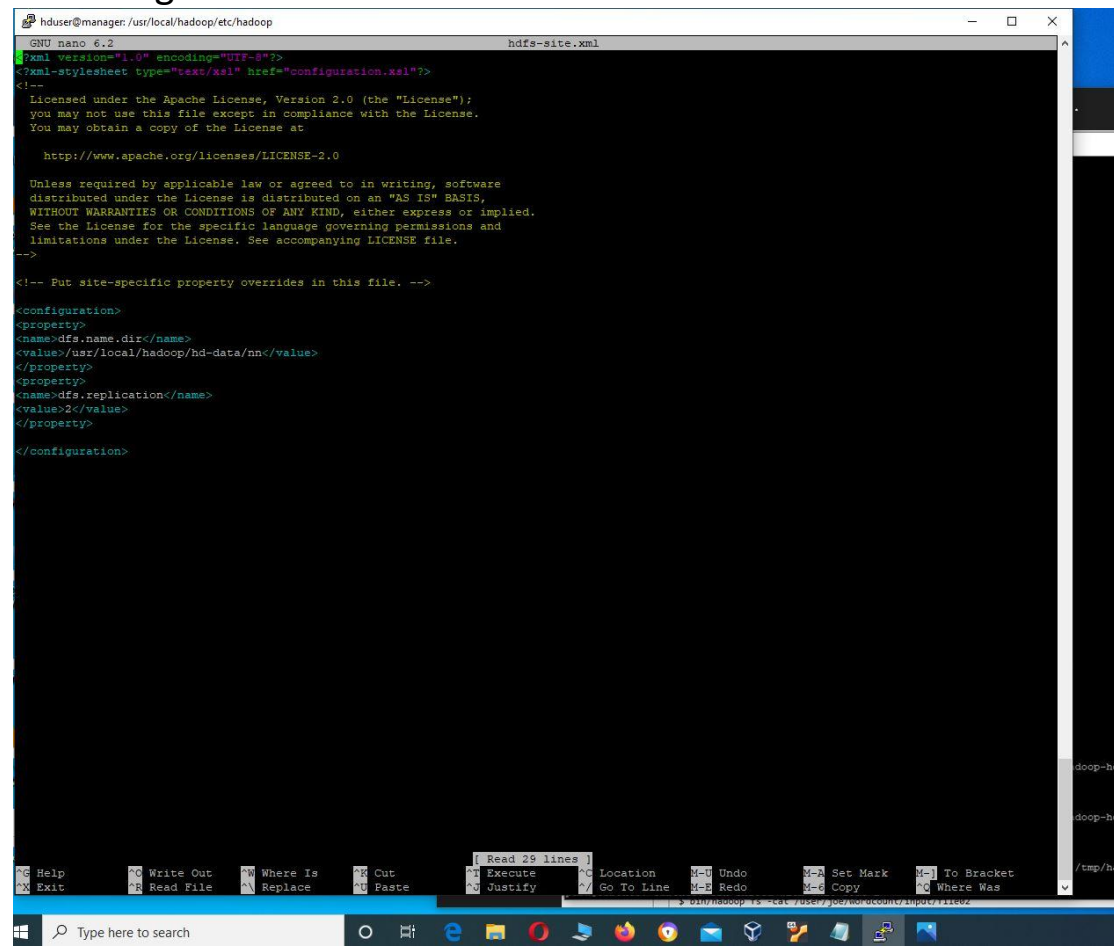
    http://www.apache.org/licenses/LICENSE-2.0

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>
<property>
<name>mapreduce.application.classpath</name>
<value>${HADOOP_MAPRED_HOME}/share/hadoop/mapreduce/*:${HADOOP_MAPRED_HOME}/share/hadoop/mapreduce/lib/*</value>
</property>
</configuration>
```

## 5. Manager hdfs-site.xml



The screenshot shows a terminal window with the nano text editor open, editing the file `hdfs-site.xml` located at `/usr/local/hadoop/etc/hadoop`. The user is `hduser` on a machine named `manager`. The file content is as follows:

```
GNU nano 6.2 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.name.dir</name>
    <value>/usr/local/hadoop/hd-data/nn</value>
  </property>
  <property>
    <name>dfs.replication</name>
    <value>2</value>
  </property>
</configuration>
```

The terminal window includes a menu bar with various options like Help, Write Out, Where Is, Cut, Paste, Read 29 lines, Execute, Justify, Location, Go To Line, Undo, Redo, Set Mark, Copy, To Bracket, and Where Was. The Windows taskbar is visible at the bottom of the screen.

## 6. Manager hadoop-env.sh

```
hduiser@manager: /usr/local/hadoop/etc/hadoop
GNU nano 6.2 hadoop-env.sh
# Specify the JVM options to be used when starting the HDFS Mover.
# These options will be appended to the options specified as HADOOP_OPTS
# and therefore may override any similar flags set in HADOOP_OPTS
#
# export HDFS_MOVER_OPTS=""

###
# Router-based HDFS Federation specific parameters
# Specify the JVM options to be used when starting the RBF Routers.
# These options will be appended to the options specified as HADOOP_OPTS
# and therefore may override any similar flags set in HADOOP_OPTS
#
# export HDFS_DFSROUTER_OPTS=""

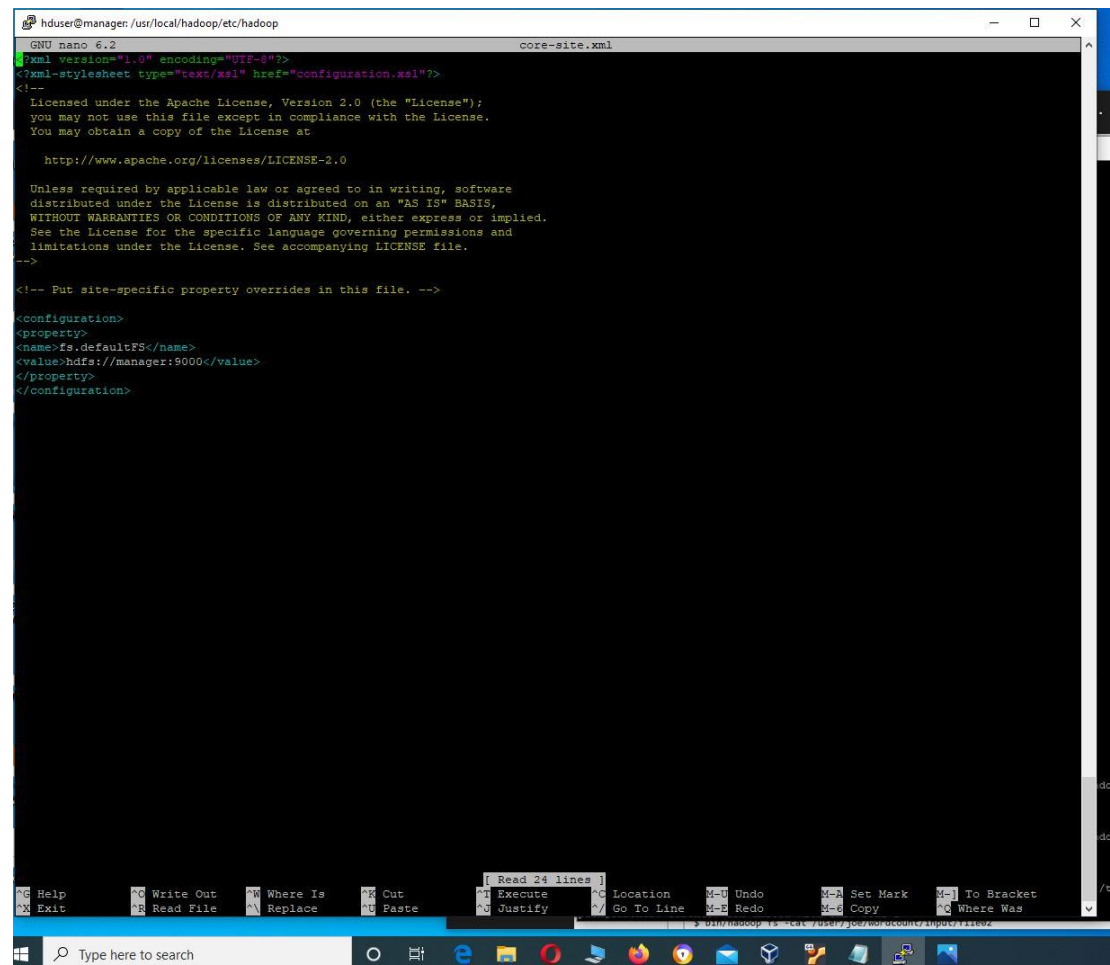
###
# HDFS StorageContainerManager specific parameters
###
# Specify the JVM options to be used when starting the HDFS Storage Container Manager.
# These options will be appended to the options specified as HADOOP_OPTS
# and therefore may override any similar flags set in HADOOP_OPTS
#
# export HDFS_STORAGECONTAINERMANAGER_OPTS=""

###
# Advanced Users Only!
###
#
# When building Hadoop, one can add the class paths to the commands
# via this special env var:
# export HADOOP_ENABLE_BUILD_PATHS="true"
#
# To prevent accidents, shell commands be (superficially) locked
# to only allow certain users to execute certain subcommands.
# It uses the format of (command)_(subcommand)_USER.
#
# For example, to limit who can execute the namenode command,
# export HDFS_NAMENODE_USER=hdfs

#JAVA
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64/
export JRE_HOME=/usr/lib/jvm/java-8-openjdk-amd64/jre
#Hadoop Environment Variables
export HADOOP_HOME=/usr/local/hadoop
export HADOOP_CONF_DIR=/usr/local/hadoop/etc/hadoop
export HADOOP_LOG_DIR=$HADOOP_HOME/logs
export HDFS_NAMENODE_USER=hduiser
export HDFS_DATANODE_USER=hduiser
export HDFS_SECONDARYNAMENODE_USER=hduiser
export YARN_RESOURCEMANAGER_USER=hduiser
export YARN_NODEMANAGER_USER=hduiser
# Add Hadoop bin/ directory to PATH
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"

Help      ^O Write Out  ^T Where Is   ^X Cut        ^Y Execute   ^G Location  ^U Undo       ^M Set Mark   ^- To Bracket
Exit      ^R Read File  ^N Replace    ^U Paste      ^J Justify   ^_ Go To Line  ^R Redo       ^C Copy       ^_ Where Was
```

## 7. Manager core-site.xml



```
GNU nano 6.2 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

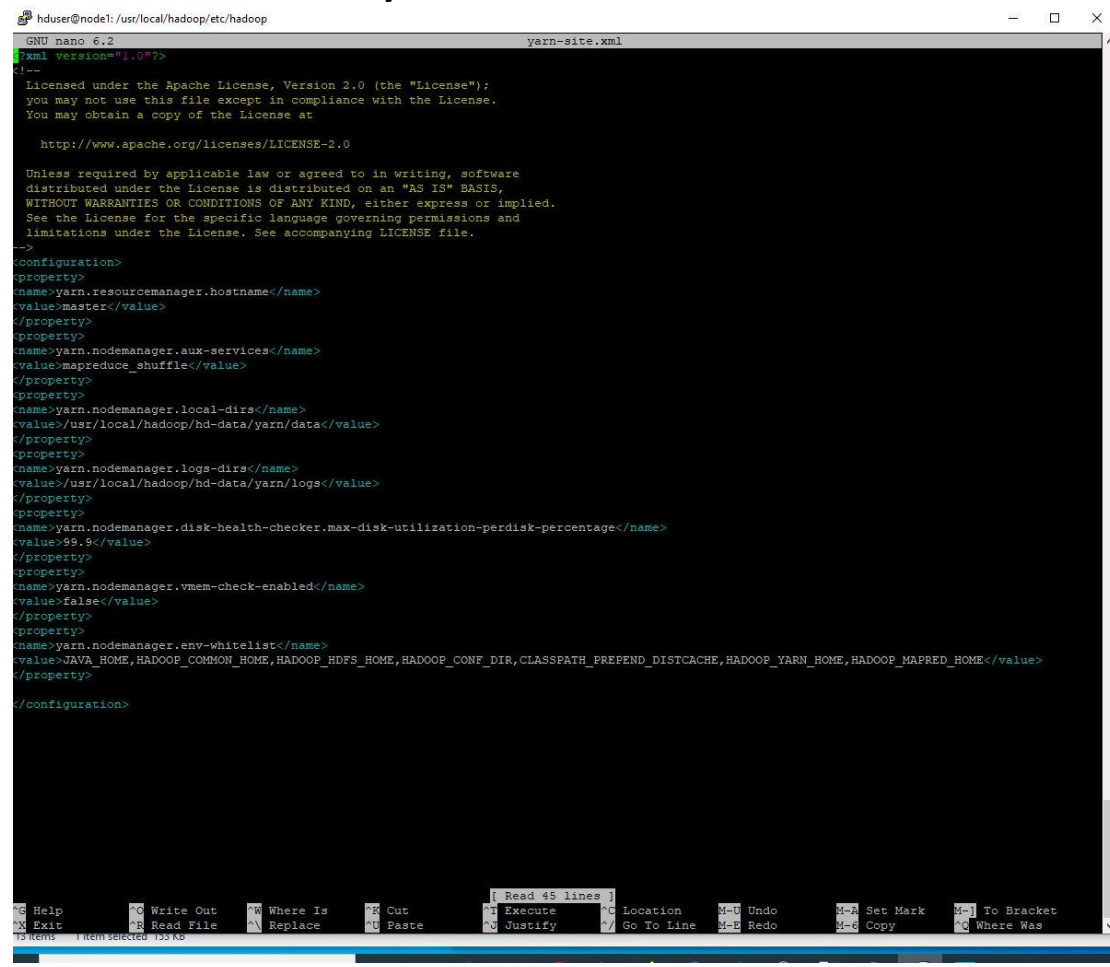
    http://www.apache.org/licenses/LICENSE-2.0

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://manager:9000</value>
</property>
</configuration>
```

## 8. Node1 and Node2 yarn-site.xml



The screenshot shows a terminal window with the title bar "hduser@node1: /usr/local/hadoop/etc/hadoop". The editor is GNU nano 6.2, editing the file "yarn-site.xml". The file content is as follows:

```

<?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.resourcemanager.hostname</name>
    <value>master</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.local-dirs</name>
    <value>/usr/local/hadoop/hd-data/yarn/data</value>
  </property>
  <property>
    <name>yarn.nodemanager.logs-dirs</name>
    <value>/usr/local/hadoop/hd-data/yarn/logs</value>
  </property>
  <property>
    <name>yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-percentage</name>
    <value>99.9</value>
  </property>
  <property>
    <name>yarn.nodemanager.vmem-check-enabled</name>
    <value>false</value>
  </property>
  <property>
    <name>yarn.nodemanager.env-whitelist</name>
    <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
  </property>
</configuration>

```

The bottom of the window shows the nano editor's status bar with various keyboard shortcuts and a message "Read 45 lines".



## 9. Node1 and Node2 mapred-site.xml



The screenshot shows a terminal window with the title bar "hduser@node1: /usr/local/hadoop/etc/hadoop". The editor is GNU nano 6.2, editing the file "mapred-site.xml". The file content is as follows:

```
<?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>
<property>
<name>mapreduce.application.classpath</name>
<value>${HADOOP_MAPRED_HOME}/share/hadoop/mapreduce/*:${HADOOP_MAPRED_HOME}/share/hadoop/mapreduce/lib/*</value>
</property>
</configuration>
```

The bottom of the window shows the nano editor's command palette with various shortcuts like Ctrl+O for Write Out, Ctrl+X for Exit, and Ctrl+R for Read File.

## Different configuration

### 1. namenode **workers**



```
hduser@managen: /usr/local/hadoop/etc/hadoop
GNU nano 6.2 workers
node1
node2
```

The screenshot shows a terminal window with the title 'hduser@managen: /usr/local/hadoop/etc/hadoop'. The window displays the GNU nano 6.2 text editor editing a file named 'workers'. The file contains two lines of text: 'node1' and 'node2'. The nano editor's status bar at the bottom shows various keyboard shortcuts and the current cursor position (1, 1).

## 2. Node1 and Node2 **workers**



## Jps

```
hduser@manager: /usr/local/hadoop/etc/hadoop
Stopping nodemanagers
Stopping resourcemanager
hduser@manager:/usr/local/hadoop/etc/hadoop$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hduser in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [manager]
Starting datanodes
Starting secondary namenodes [manager]
Starting resourcemanager
Starting nodemanagers
hduser@manager:/usr/local/hadoop/etc/hadoop$ jps
7848 ResourceManager
7115 NameNode
8155 Jps
7583 SecondaryNameNode
hduser@manager:/usr/local/hadoop/etc/hadoop$ ssh node1
Welcome to Ubuntu 22.04 LTS (GNU/Linux 6.8.0-47-generic x86_64)

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

254 updates can be applied immediately.
1 of these updates is a standard security update.
To see these additional updates run: apt list --upgradable

New release '24.04.1 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Last login: Mon Oct 21 08:58:57 2024 from 192.168.56.104
hduser@node1:~$ jps
6544 DataNode
6675 NodeManager
6815 Jps
hduser@node1:~$ exit
logout
Connection to node1 closed.
hduser@manager:/usr/local/hadoop/etc/hadoop$ ssh node2
Welcome to Ubuntu 22.04 LTS (GNU/Linux 6.8.0-47-generic x86_64)

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

253 updates can be applied immediately.
1 of these updates is a standard security update.
To see these additional updates run: apt list --upgradable

New release '24.04.1 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Last login: Mon Oct 21 09:01:24 2024 from 192.168.56.104
hduser@node2:~$ jps
4880 Jps
4707 NodeManager
4607 DataNode
hduser@node2:~$ exit
logout
Connection to node2 closed.
hduser@manager:/usr/local/hadoop/etc/hadoop$
```

# Stop

```
hduser@manager:~  
* Documentation:  https://help.ubuntu.com  
* Management:    https://landscape.canonical.com  
* Support:        https://ubuntu.com/advantage  
  
254 updates can be applied immediately.  
1 of these updates is a standard security update.  
To see these additional updates run: apt list --upgradable  
  
New release '24.04.1 LTS' available.  
Run 'do-release-upgrade' to upgrade to it.  
  
Last login: Mon Oct 21 09:16:44 2024 from 192.168.56.104  
hduser@node1:~$ nano yarn-site.xml  
hduser@node1:~$ ls  
Desktop Documents Downloads hadoop-3.2.4 hadoop.tar.gz Music Pictures Public snap Templates Videos  
hduser@node1:~$ cd /usr/local/hadoop/etc/hadoop  
hduser@node1:/usr/local/hadoop/etc/hadoop$ ls  
capacity-scheduler.xml      hadoop-policy.xml          kms-acls.xml               mapred-queues.xml.template  yarn-env.cmd  
configuration.xml           hadoop-user-functions.sh.example  kms-env.sh                mapred-site.xml            yarn-env.sh  
container-executor.cfg      hdfs-site.xml              kms-log4j.properties      shellprofile.d              yarnservice-log4j.properties  
core-site.xml               https-env.sh                kms-site.xml               ssl-client.xml.example      yarn-site.xml  
hadoop-env.cmd              https-log4j.properties        log4j.properties          ssl-server.xml.example  
hadoop-env.sh               https-signature.secret        mapred-env.cmd             user_ec_policies.xml.template  
hadoop-metrics2.properties https-site.xml               mapred-env.sh              workers  
hduser@node1:/usr/local/hadoop/etc/hadoop$ nano yarn-site.xml  
hduser@node1:/usr/local/hadoop/etc/hadoop$ jps  
6344 DataNode  
7098 Jps  
hduser@node1:/usr/local/hadoop/etc/hadoop$ ssh manager  
hduser@manager's password:  
Welcome to Ubuntu 22.04 LTS (GNU/Linux 6.8.0-47-generic x86_64)  
  
* Documentation:  https://help.ubuntu.com  
* Management:    https://landscape.canonical.com  
* Support:        https://ubuntu.com/advantage  
  
253 updates can be applied immediately.  
1 of these updates is a standard security update.  
To see these additional updates run: apt list --upgradable  
  
New release '24.04.1 LTS' available.  
Run 'do-release-upgrade' to upgrade to it.  
  
Last login: Mon Oct 21 09:26:10 2024 from 192.168.56.105  
hduser@manager:~$ cd /usr/local/hadoop/etc/hadoop  
hduser@manager:/usr/local/hadoop/etc/hadoop$ jps  
7848 ResourceManager  
7115 NameNode  
8780 Jps  
7393 SecondaryNameNode  
hduser@manager:/usr/local/hadoop/etc/hadoop$ cd  
hduser@manager:~$ stop-all.sh  
WARNING: Stopping all Apache Hadoop daemons as hduser in 10 seconds.  
WARNING: Use CTRL-C to abort.  
Stopping namenodes on [manager]  
Stopping datanodes  
Stopping secondary namenodes [manager]  
Stopping nodemanagers  
Stopping resourcemanager  
hduser@manager:~$
```