Lab Report

Course Title: Big Data And IOT

Course Code: CSE413

Report No: 01

Report Title: Hadoop Installation On My_NativeOS[Linux]

Submitted To:

Name: Husne Mubarak

Designation: Lecturer

Dept. of Computer Science and Engineering

Daffodil International University

Submitted By:

Name: Md Shamsuzzaman

ID: 211-15-4031

Section: 58_E(E1)

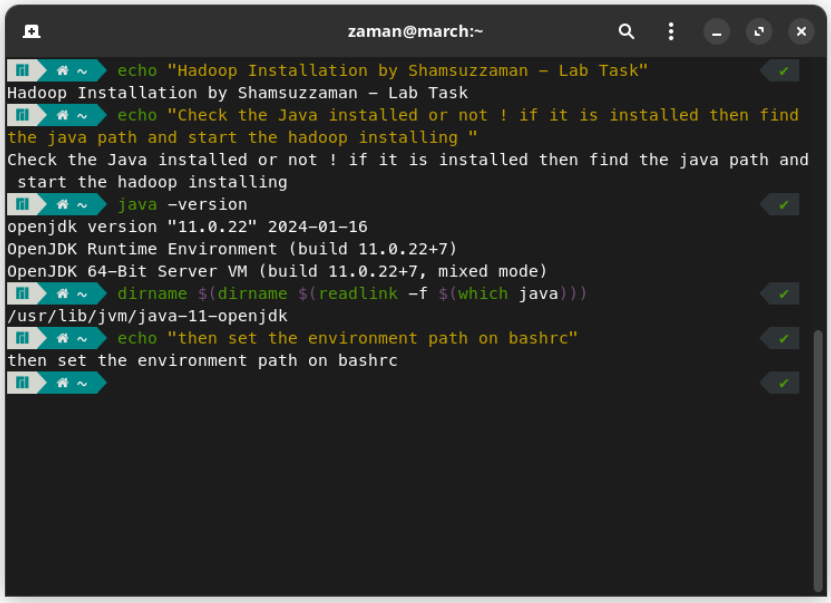
Department Of CSE

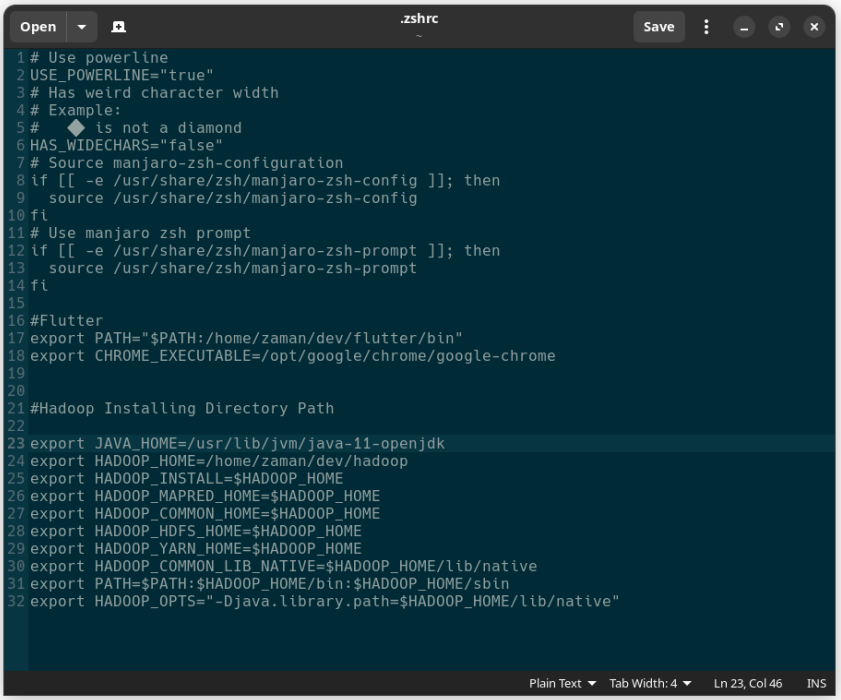
Daffodil International University

Date of Submission: 24 - 04 – 2024

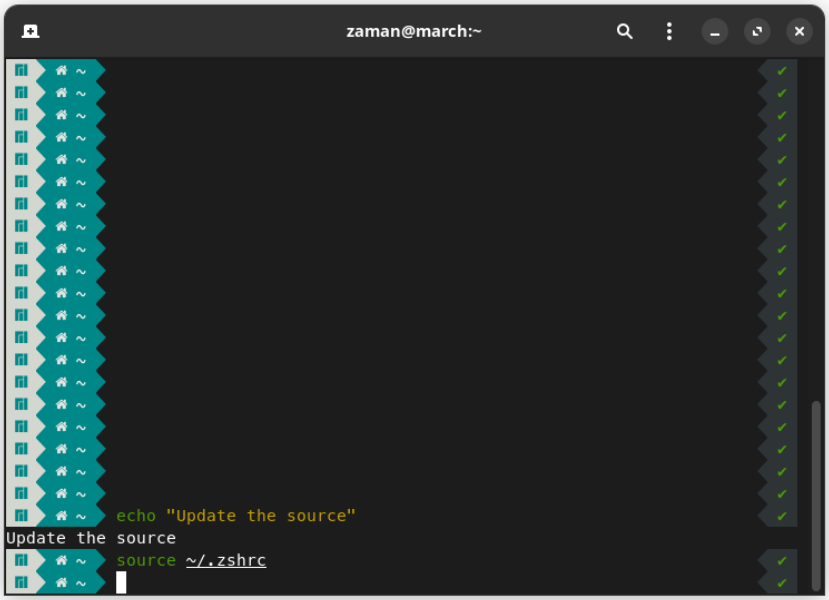
Title: Installation of Linux, JDK and Hadoop in Virtual Box/Linux

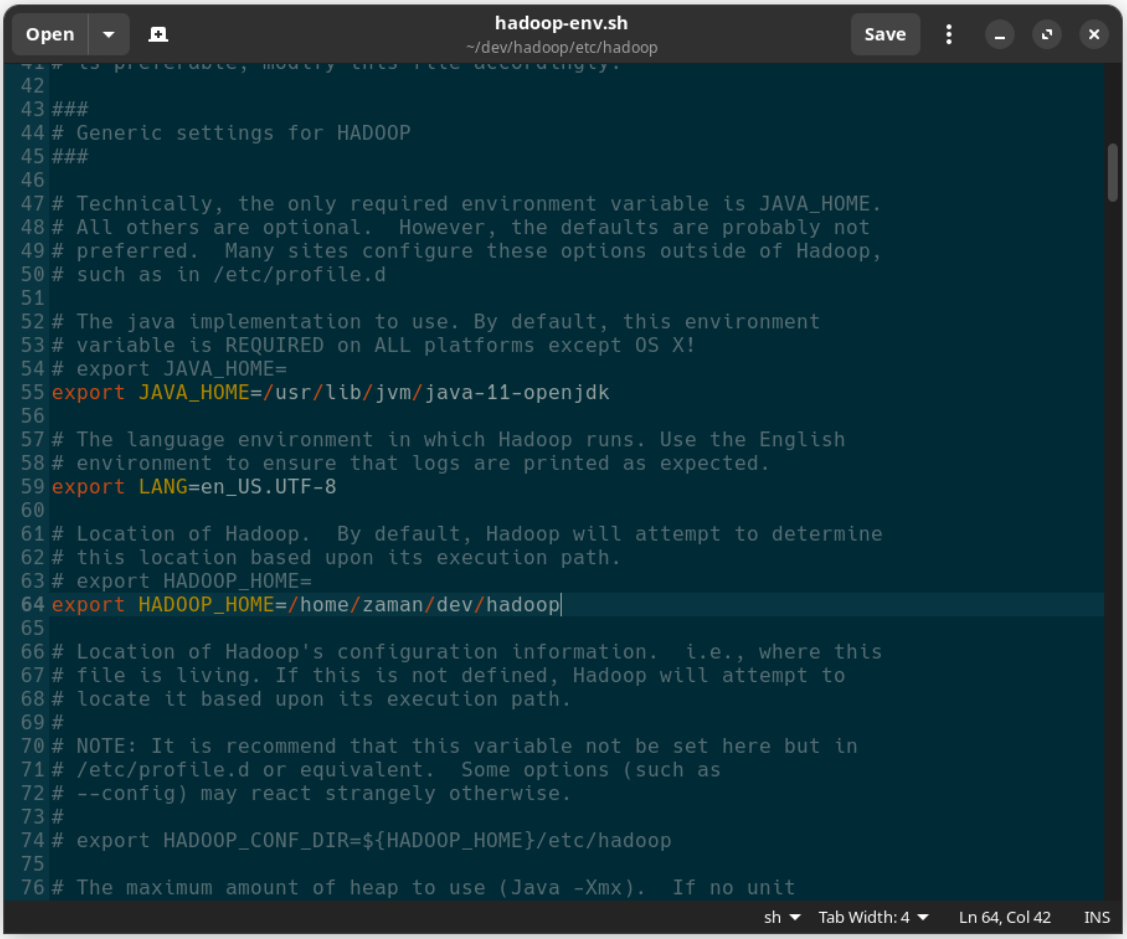
Description: At First install JavaJDK or If it is installed then check the version and path.





Edit .Zshrc like above.


zaman@march:~
echo "Update the source"
Update the source
source ~/.zshrc


Open
hadoop-env.sh
~/dev/hadoop/etc/hadoop
Save
42
43 ###
44 # Generic settings for HADOOP
45 ###
46
47 # Technically, the only required environment variable is JAVA_HOME.
48 # All others are optional. However, the defaults are probably not
49 # preferred. Many sites configure these options outside of Hadoop,
50 # such as in /etc/profile.d
51
52 # The java implementation to use. By default, this environment
53 # variable is REQUIRED on ALL platforms except OS X!
54 # export JAVA_HOME=
55 export JAVA_HOME=/usr/lib/jvm/java-11-openjdk
56
57 # The language environment in which Hadoop runs. Use the English
58 # environment to ensure that logs are printed as expected.
59 export LANG=en_US.UTF-8
60
61 # Location of Hadoop. By default, Hadoop will attempt to determine
62 # this location based upon its execution path.
63 # export HADOOP_HOME=
64 export HADOOP_HOME=/home/zaman/dev/hadoop
65
66 # Location of Hadoop's configuration information. i.e., where this
67 # file is living. If this is not defined, Hadoop will attempt to
68 # locate it based upon its execution path.
69 #
70 # NOTE: It is recommend that this variable not be set here but in
71 # /etc/profile.d or equivalent. Some options (such as
72 # --config) may react strangely otherwise.
73 #
74 # export HADOOP_CONF_DIR=${HADOOP_HOME}/etc/hadoop
75
76 # The maximum amount of heap to use (Java -Xmx). If no unit
sh
Tab Width: 4
Ln 64, Col 42
INS

Open | core-site.xml | ~/dev/hadoop/etc/hadoop | Save

```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>
```

XML | Tab Width: 4 | Ln 23, Col 12 | INS

Open | hdfs-site.xml | ~/dev/hadoop/etc/hadoop | Save

```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>3</value>
</property><property>
  <name>dfs.name.dir</name>
  <value>/home/zaman/dev/hadoop/data/namenode</value>
</property><property>
  <name>dfs.data.dir</name>
  <value>/home/zaman/dev/hadoop/data/datanode</value>
</property>

</configuration>
```

XML | Tab Width: 4 | Ln 29, Col 12 | INS

Open | mapred-site.xml ~/dev/hadoop/etc/hadoop | Save

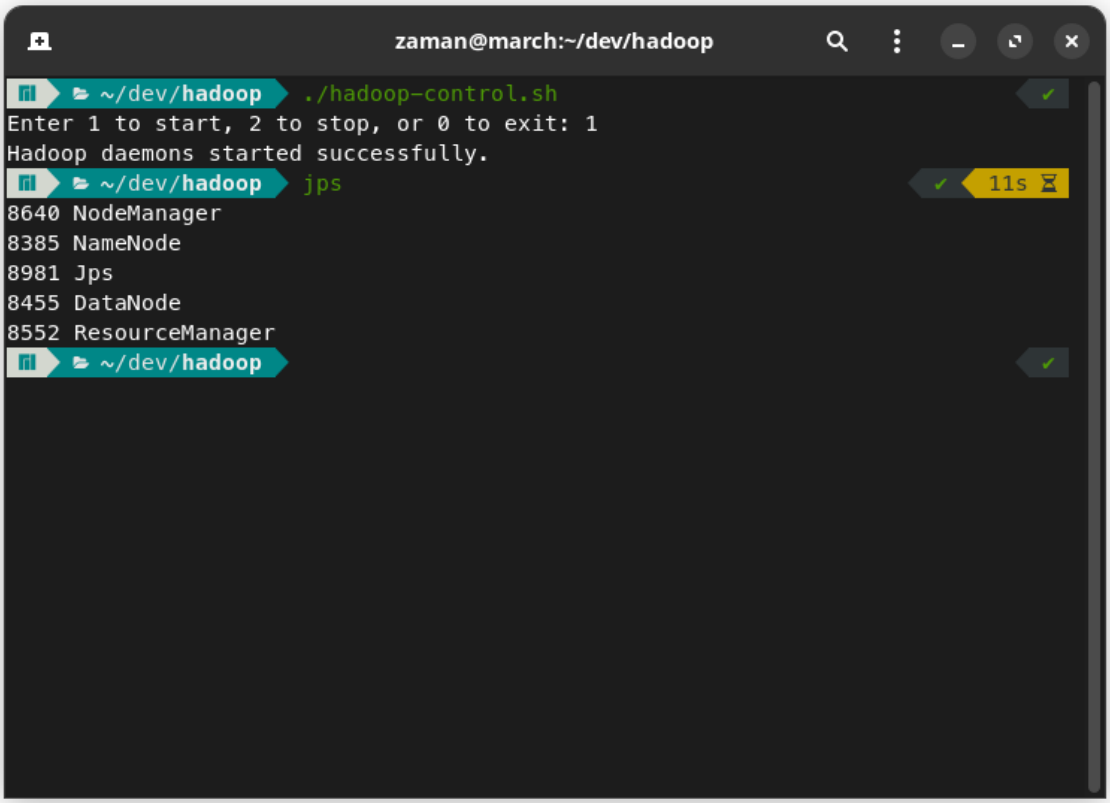
```xml
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    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>
```

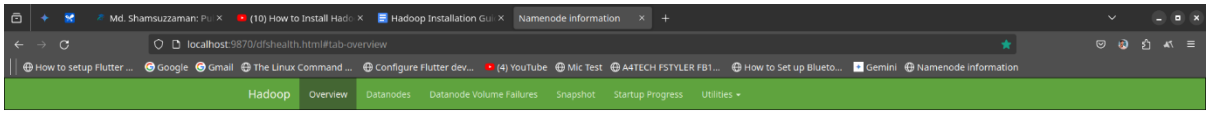
Loading file "~/dev/hadoop/etc/hadoop/mapred-site.xml"... XML Tab Width: 4 Ln 24, Col 1 INS

Open | yarn-site.xml ~/dev/hadoop/etc/hadoop | Save

```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><property>
  <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
  <value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>

<!-- Site specific YARN configuration properties -->

</configuration>
```

Loading file "~/dev/hadoop/etc/hadoop/yarn-site.xml"... XML Tab Width: 4 Ln 23, Col 1 INS





Overview 'localhost:9000' (active)

Started:	Fri May 10 21:13:17 +0600 2024
Version:	3.4.0, rbd8b77f398f626bb7791783192ee7a5dfaeec760
Compiled:	Mon Mar 04 12:35:00 +0600 2024 by root from (HEAD detached at release-3.4.0-RC3)
Cluster ID:	CID-6c27212e-e371-44db-94d3-7a208da14a8e
Block Pool ID:	BP-2109161109-127.0.1.1-1714735074023

Summary

Security is off.

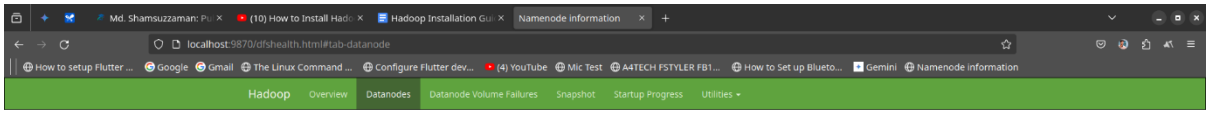
Safemode is off.

1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

Heap Memory used 91.12 MB of 212 MB Heap Memory. Max Heap Memory is 1.87 GB.

Non Heap Memory used 62.34 MB of 65.82 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.

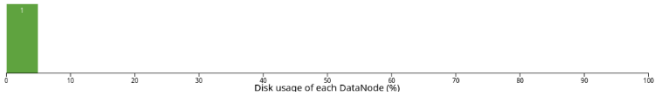
Configured Capacity:	224.71 GB
Configured Remote Capacity:	0 B
DFS Used:	32 KB (0%)
Non DFS Used:	67.88 GB
DFS Remaining:	145.35 GB (64.68%)
Block Pool Used:	32 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)



Datanode Information

In service | Down | Decommissioning | Decommissioned | Decommissioned & dead
Entering Maintenance | In Maintenance | In Maintenance & dead

Datanode usage histogram



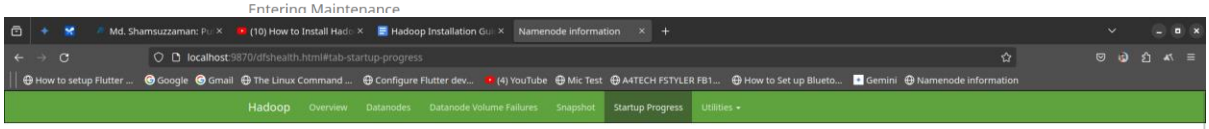
In operation

DataNode State: All | Show 25 entries | Search:

Node	Http Address	Last contact	Last Block Report	Used	Non DFS Used	Capacity	Blocks	Block pool used	Block pool usage StdDev	Version
/default-rack/march:9866 (127.0.0.1:9866)	http://march:9864	0s	1m	32 KB	67.89 GB	224.71 GB	0	32 KB (0%)	0%	3.4.0

Showing 1 to 1 of 1 entries | Previous | 1 | Next

Entering Maintenance



Startup Progress

Elapsed Time: 1 sec, Percent Complete: 100%

Phase	Completion	Elapsed Time
Loading fsimage /home/zaman/dev/hadoop/data/namenode/current/fsimage_0000000000000000000 400 B	**100%**	**0 sec**
erasure coding policies (0/0)	100%	
inodes (1/1)	100%	
delegation tokens (0/0)	100%	
cache pools (0/0)	100%	
Loading edits	**100%**	**0 sec**
/home/zaman/dev/hadoop/data/namenode/current/edits_0000000000000000001-0000000000000000001 1 MB (1/1)	100%	
/home/zaman/dev/hadoop/data/namenode/current/edits_0000000000000000002-0000000000000000002 1 MB (1/1)	100%	
/home/zaman/dev/hadoop/data/namenode/current/edits_0000000000000000003-0000000000000000003 1 MB (1/1)	100%	
/home/zaman/dev/hadoop/data/namenode/current/edits_0000000000000000004-0000000000000000004 1 MB (1/1)	100%	
Saving checkpoint	**100%**	**0 sec**
erasure coding policies /home/zaman/dev/hadoop/data/namenode/current/fsimage.ckpt_0000000000000000004 (0/0)	100%	
inodes /home/zaman/dev/hadoop/data/namenode/current/fsimage.ckpt_0000000000000000004 (0/0)	100%	
delegation tokens /home/zaman/dev/hadoop/data/namenode/current/fsimage.ckpt_0000000000000000004 (0/0)	100%	
cache pools /home/zaman/dev/hadoop/data/namenode/current/fsimage.ckpt_0000000000000000004 (0/0)	100%	
Safe mode	**100%**	**0 sec**
awaiting reported blocks (0/0)	100%	

All My command:

Find Java Location Command
dirname $(dirname $(readlink -f $(which java)))
Output:
/usr/lib/jvm/java-11-openjdk
For bashrc
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk
export HADOOP_HOME=/home/zaman/dev/hadoop
export HADOOP_INSTALL=$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=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"

Hadoop Configuration
For core-site.xml
```
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://localhost:9000</value>
</property>
```

For hdfs-site.xml or https-site.xml
```
<property>
  <name>dfs.replication</name>
  <value>3</value>
</property><property>
  <name>dfs.name.dir</name>
  <value>/home/zaman/dev/hadoop/data/namenode</value>
</property><property>
  <name>dfs.data.dir</name>
  <value>/home/zaman/dev/hadoop/data/datanode</value>
</property>
```

For mapred-site.xml
```
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
```

For yarn-site.xml
```
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property><property>
  <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
  <value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
```

```
hdfs namenode -format

start-dfs.sh
```

SSH Key Configure
```
ssh-keygen -t rsa
# replace id_rsa as authorized keys
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
# add read and write access
chmod 640 ~/.ssh/authorized_keys

sudo systemctl enable sshd
sudo systemctl start sshd
```

```
#!/bin/bash

# Start individual daemons
sbin/hadoop-daemon.sh start namenode
sbin/hadoop-daemon.sh start datanode
sbin/yarn-daemon.sh start resourcemanager
sbin/yarn-daemon.sh start nodemanager
# Add additional daemons if needed based on your configuration

echo "Hadoop daemons started successfully."
```

```
#!/bin/bash

# Stop individual daemons
sbin/hadoop-daemon.sh stop namenode
sbin/hadoop-daemon.sh stop datanode
sbin/yarn-daemon.sh stop resourcemanager
sbin/yarn-daemon.sh stop nodemanager
# Add additional daemons if needed based on your configuration
```

```
echo "Hadoop daemons stopped successfully."
```

```
#!/bin/bash

function start_daemons {
  # Start daemons using recommended 'hdfs' and 'yarn' commands
  hdfs --daemon start namenode
  hdfs --daemon start datanode
  yarn --daemon start resourcemanager
  yarn --daemon start nodemanager
  # Add additional daemons if needed
  echo "Hadoop daemons started successfully."
}


function stop_daemons {
  # Stop daemons using recommended 'hdfs' and 'yarn' commands
  hdfs --daemon stop namenode
  hdfs --daemon stop datanode
  yarn --daemon stop resourcemanager
  yarn --daemon stop nodemanager
  # Add additional daemons if needed
  echo "Hadoop daemons stopped successfully."
}

# Main loop for user interaction
while true; do
  read -p "Enter 1 to start, 2 to stop, or 0 to exit: " choice

  case $choice in
                1)
                start_daemons
                break
                ;;
                2)
                stop_daemons
                break
                ;;
                0)
                echo "Exiting..."
                exit 0
                ;;
                *)
                echo "Invalid choice. Please enter 1, 2, or 0."
                ;;
  esac
done

# Make the script executable (optional)
#chmod +x hadoop-control.sh
```