Assignment 1.

File_Creation

```
acts@osboxes:~$ chown -R acts:acts /usr/local/hadoop
chown: changing ownership of '/usr/local/hadoop': Operation not permitted
acts@osboxes:~$ sudo chown -R acts:acts /usr/local/hadoop
acts@osboxes:~$ sudo chmod 755 -R /usr/local/hadoop
acts@osboxes:~$ mkdir -p /usr/local/hadoop/hd_store/tmp
acts@osboxes:~$ mkdir -p /usr/local/hadoop/hd_store/namenode
acts@osboxes:~$ mkdir -p /usr/local/hadoop/hd_store/datanode
acts@osboxes:~$ cd $HADOOP_HOME/etc/hadoop
acts@osboxes:/usr/local/hadoop/etc/hadoop$ nano hadoop-env.sh
acts@osboxes:/usr/local/hadoop/etc/hadoop$ nano core-site.xml
acts@osboxes:/usr/local/hadoop/etc/hadoop$ nano yarn-site.xml
acts@osboxes:/usr/local/hadoop/etc/hadoop$ nano hdfs-site.xml
acts@osboxes:/usr/local/hadoop/etc/hadoop$ nano mapred-site.xml
acts@osboxes:/usr/local/hadoop/etc/hadoop$ nano mapred-site.xml
acts@osboxes:/usr/local/hadoop/etc/hadoop$ nano workers
acts@osboxes:/usr/local/hadoop/etc/hadoop$ cd $HADOOP_HOME/sbin
```

Hadoop_file

```bash
# HDFS Mover specific parameters
###
# 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=$HADOOP_HOME/logs
export HDFS_NAMENODE_USER=acts
export HDFS_DATANODE_USER=acts
export HDFS_SECONDARYNAMENODE_USER=acts
export YARN_RESOURCEMANAGER_USER=acts
export YARN_NODEMANAGER_USER=acts
export YARN_NODEMANAGER_USER=acts
# Add Hadoop bin/ directory to PATH
export PATH=$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"
```

Hdfs-site-xml

```
acts@osboxes: /usr/local/hadoop/etc/hadoop

  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.replication</name>
<value>1</value>
</property>
<property>
<name>dfs.name.dir</name>
<value>/usr/local/hadoop/hd_store/namenode</value>
</property>
<property>
<name>dfs.data.dir</name>
<value>/usr/local/hadoop/hd_store/datanode</value>
</property>
</configuration>
```

Mapred-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>yarn.app.mapreduce.am.env</name>
<value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
</property>
<property>
<name>mapreduce.map.env</name>
<value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
</property>
<property>
<name>mapreduce.reduce.env</name>
<value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
</property>
</configuration>
```

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>hadoop.tmp.dir</name>
<value>/usr/local/hadoop/hd_store/tmp</value>
</property>
<property>
<name>fs.defaultFS</name>
<value>hdfs://localhost:9000</value>
</property>
</configuration>
```

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>

<!-- Site specific YARN configuration properties -->
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>

</configuration>
```

File Copy

```
acts@osboxes:~$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x   - acts supergroup          0 2024-10-18 06:34 /input
acts@osboxes:~$ hdfs dfs -copyFromLocal word.txt /input/
acts@osboxes:~$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x   - acts supergroup          0 2024-10-18 06:36 /input
acts@osboxes:~$ jps
52835 Jps
18325 DataNode
18118 NameNode
18630 SecondaryNameNode
acts@osboxes:~$ cd /usr/local/hadoop/etc/hadoop
acts@osboxes:/usr/local/hadoop/etc/hadoop$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x   - acts supergroup          0 2024-10-18 06:36 /input
acts@osboxes:/usr/local/hadoop/etc/hadoop$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x   - acts supergroup          0 2024-10-18 06:36 /input
acts@osboxes:/usr/local/hadoop/etc/hadoop$ cd ~
acts@osboxes:~$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x   - acts supergroup          0 2024-10-18 06:36 /input
acts@osboxes:~$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x   - acts supergroup          0 2024-10-18 06:36 /input
acts@osboxes:~$ hdfs dfs -ls /input
Found 1 items
-rw-r--r--   1 acts supergroup         21 2024-10-18 06:36 /input/word.txt
acts@osboxes:~$ m5sum word.txt
Command 'm5sum' not found, did you mean:
  command 'md5sum' from deb coreutils (8.32-4.1ubuntu1.2)
  command 'mdsum' from deb ucommon-utils (7.0.0-20ubuntu2)
Try: apt install <deb name>
acts@osboxes:~$ md5sum word.txt
1065a3d3c7fbc470c4b59446f7bc25d7  word.txt
acts@osboxes:~$ nano word.txt
acts@osboxes:~$ nano word.txt
acts@osboxes:~$ 
```

Display_node jps

```
/*****************************************************
SHUTDOWN_MSG: Shutting down NameNode at osboxes/127.0.1.1
*****************************************************/
acts@osboxes:/usr/local/hadoop/etc/hadoop$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [osboxes]
osboxes: Warning: Permanently added 'osboxes' (ED25519) to the list of known hosts.
acts@osboxes:/usr/local/hadoop/etc/hadoop$ jps
18325 DataNode
18118 NameNode
18630 SecondaryNameNode
19082 Jps
acts@osboxes:/usr/local/hadoop/etc/hadoop$ start-dfs.sh
Starting namenodes on [localhost]
localhost: namenode is running as process 18118.  Stop it first and ensure /tmp/hadoop-acts-namenode.pid file is empty
pdsh@osboxes: localhost: ssh exited with exit code 1
Starting datanodes
localhost: datanode is running as process 18325.  Stop it first and ensure /tmp/hadoop-acts-datanode.pid file is empty
pdsh@osboxes: localhost: ssh exited with exit code 1
Starting secondary namenodes [osboxes]
osboxes: secondarynamenode is running as process 18630.  Stop it first and ensure /tmp/hadoop-acts-secondarynamenode.pi
efore retry.
pdsh@osboxes: osboxes: ssh exited with exit code 1
acts@osboxes:/usr/local/hadoop/etc/hadoop$ jps
18325 DataNode
18118 NameNode
18630 SecondaryNameNode
22072 Jps
acts@osboxes:/usr/local/hadoop/etc/hadoop$ ls
capacity-scheduler.xml          hdfs-site.xml           log4j.properties            worker
configuration.xsl               httpfs-env.sh           mapred-env.cmd              workers
container-executor.cfg          httpfs-log4j.properties mapred-env.sh               yarn-env.cmd
core-site.xml                   httpfs-signature.secret mapred-queues.xml.template  yarn-env.sh
hadoop-env.cmd                  httpfs-site.xml         mapred-site.xml             yarnservice-log4j.properties
hadoop-env.sh                   kms-acls.xml            shellprofile.d              yarn-site.xml
hadoop-metrics2.properties      kms-env.sh              ssl-client.xml.example
hadoop-policy.xml               kms-log4j.properties    ssl-server.xml.example
hadoop-user-functions.sh.example kms-site.xml           user_ec_policies.xml.template
acts@osboxes:/usr/local/hadoop/etc/hadoop$ cd ~
```