Hadoop Cluster Commands

https://www.youtube.com/watch?v=_zR0v4XIFNs

```
## on all nodes
sudo apt update

sudo apt install ssh
sudo apt install pdsh

nano .bashrc
## add to the end of .bashrc file of all nodes
export PDSH_RCMD_TYPE=ssh

## create keygen on all three nodes and skip any input req using enter
ssh-keygen -t rsa -P ""

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

## login to localhost and enter yes in req
ssh localhost
sudo apt install openjdk-8-jdk

sudo apt install openjdk-11-jdk

java -version

sudo wget -P ~
```
https://mirrors.sonic.net/apache/hadoop/common/hadoop-3.2.1/hadoop-3.2.1.tar.gz

Hadoop 3.3.1
```
Alternative : sudo wget -P ~
https://mirror.olnevhost.net/pub/apache/hadoop/common/hadoop-3.3.1/hadoop-3.3.1.tar.gz


tar xzf hadoop-3.3.1.tar.gz
mv hadoop-3.3.1 hadoop

tar xzf hadoop-3.2.1.tar.gz
mv hadoop-3.2.1 hadoop
nano ~/hadoop/etc/hadoop/hadoop-env.sh
```

Goes at the top of Hadoop-env.sh file

export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64/

sudo mv hadoop /usr/local/hadoop
sudo nano /etc/environment

put the following path below the existing path in the /etc/environment

PATH="/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/usr/local/hadoop/bin:/usr/local/hadoop/sbin"JAVA_HOME="/usr/lib/jvm/java-11-openjdk-amd64/jre"

enter 5 commands below on all the nodes and enter password twice and press enter for
everyother option

```
[hadoop-master@hadoop-master:~$ tar xzf hadoop-3.3.1.tar.gz
[hadoop-master@hadoop-master:~$ mv hadoop-3.3.1 hadoop
[hadoop-master@hadoop-master:~$ nano ~/hadoop/etc/hadoop/hadoop-env.sh
[hadoop-master@hadoop-master:~$ sudo mv hadoop /usr/local/hadoop
[hadoop-master@hadoop-master:~$ sudo nano /etc/environment
hadoop-master@hadoop-master:~$ sudo adduser hadoopuser
Adding user `hadoopuser' ...
Adding new group `hadoopuser' (1001) ...
Adding new user `hadoopuser' (1001) with group `hadoopuser' ...
Creating home directory `/home/hadoopuser' ...
Copying files from `/etc/skel' ...
[Enter new UNIX password:
[Retype new UNIX password:
 passwd: password updated successfully
Changing the user information for hadoopuser
Enter the new value, or press ENTER for the default
[        Full Name []:
[        Room Number []:
[        Work Phone []:
         Home Phone []:
[        Other []:
[Is the information correct? [Y/n]
hadoop-master@hadoop-master:~$ []
```

sudo adduser hadoopuser
sudo usermod -aG hadoopuser hadoopuser
sudo chown hadoopuser:root -R /usr/local/hadoop/
sudo chmod g+rwx -R /usr/local/hadoop/
sudo adduser hadoopuser sudo

```
# to find the ip addresses of each node then paste that info in /etc/hosts
ip addr

10.0.0.4 hadoop-master
10.0.0.5 hadoop-slave1
10.0.0.6 hadoop-slave2



sudo nano /etc/hosts

10.0.0.4 hadoop-master
10.0.0.5 hadoop-slave1
10.0.0.6 hadoop-slave2


sudo nano /etc/hostname
sudo reboot



# login to hadoopuser on master-node only
su - hadoopuser
# for master-node only and press enter for
ssh-keygen -t rsa

# enter following one by one only on master and enter yes and enter password
ssh-copy-id hadoopuser@hadoop-master
ssh-copy-id hadoopuser@hadoop-slave1
ssh-copy-id hadoopuser@hadoop-slave2


# run below command block individually on slave 1 and slave 2
sudo adduser hadoopuser
sudo usermod -aG hadoopuser hadoopuser
sudo chown hadoopuser:root -R /usr/local/hadoop/
sudo chmod g+rwx -R /usr/local/hadoop/
sudo adduser hadoopuser sudo

sudo nano /usr/local/hadoop/etc/hadoop/core-site.xml


<property>
<name>fs.defaultFS</name>
```

```
<value>hdfs://hadoop-master:9000</value>
</property>
```

```
sudo nano /usr/local/hadoop/etc/hadoop/hdfs-site.xml
```

```
<property>
<name>dfs.namenode.name.dir</name><value>/usr/local/hadoop/data/nameNode</value>
</property>
<property>
<name>dfs.datanode.data.dir</name><value>/usr/local/hadoop/data/dataNode</value>
</property>
<property>
<name>dfs.replication</name>
<value>2</value>
</property>
```

```
# only on hadoop-master
sudo nano /usr/local/hadoop/etc/hadoop/workers
```

```
hadoop-slave1
hadoop-slave2
```

```
scp /usr/local/hadoop/etc/hadoop/* hadoop-slave1:/usr/local/hadoop/etc/hadoop/
scp /usr/local/hadoop/etc/hadoop/* hadoop-slave2:/usr/local/hadoop/etc/hadoop/
```

```
## only on master node
source /etc/environment
cd /usr/local/hadoop/
bin/hdfs namenode -format
sbin/start-dfs.sh
```

```
## now go to azure cloud web portal and in the overview section of hadoop-master change dns
name to hadoop-master-1765
```

go in networking of hadoop-master and add inbound port rule. Change port name and destination port ranges to 9870

go in networking of hadoop-master and add inbound port rule. Change port name and destination port ranges to 8088

go to hadoop-master copy dns name and paste it in the search bar
Enter : url:9870

only on master node

```
export HADOOP_HOME="/usr/local/hadoop"
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export HADOOP_HDFS_HOME=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_YARN_HOME=$HADOOP_HOME
```

sudo nano /usr/local/hadoop/etc/hadoop/yarn-site.xml

```
<property>
<name>yarn.resourcemanager.hostname</name>
<value>hadoop-master</value>
</property>
```

sbin/start-yarn.sh

sbin/start-all.sh

scp /usr/local/hadoop/etc/hadoop/* hadoop-slave1:/usr/local/hadoop/etc/hadoop/
scp /usr/local/hadoop/etc/hadoop/* hadoop-slave2:/usr/local/hadoop/etc/hadoop/

sbin/stop-all.sh

sbin/start-all.sh

hadoop fs -put /Users/wajeeh/Desktop/Big\ Data\ Assignments/A3/Big-Data-Assignment-3/data.txt /user

hadoop fs -put /Users/wajeeh/Desktop/Big\ Data\ Assignments/A3/Big-Data-Assignment-3/data.txt /user

scp data.txt hadoopuser@hadoop-master:"/Users/wajeeh/Desktop/Big\ Data\ Assignments/A3/Big-Data-Assignment-3/data.txt"

"/Users/wajeeh/Desktop/Big\ Data\ Assignments/A3/Big-Data-Assignment-3/data.txt"

hadoop fs -mkdir /user

cat data.txt |python3 /Users/wajeeh/Desktop/Big Data Assignments/A3/Big-Data-Assignment-3/mapper.py |python3 /Users/wajeeh/Desktop/Big Data Assignments/A3/Big-Data-Assignment-3/reducer.py^C

hadoop fs -put data.txt /user
hadoop jar /home/talal/hadoop-3.3.4/share/hadoop/tools/lib/hadoop-streaming-3.3.4.jar -file /home/talal/mapper1.py -mapper 'python3 mapper1.py' -file /home/talal/reducer1.py -reducer 'python3 reducer1.py' -input /user/file.txt -output /user/output4
hadoop fs -rm -r /user/output5

uploading mapper.py sample.txt and data.txt and reducer.py from local machine to hadoop-master at default /home/hadoop-master

scp Desktop/Big\ Data\ Assignments/A3/Big-Data-Assignment-3/sample.txt Desktop/Big\ Data\ Assignments/A3/Big-Data-Assignment-3/mapper.py Desktop/Big\ Data\ Assignments/A3/Big-Data-Assignment-3/reducer.py hadoop-master@20.110.56.176:~

scp Desktop/Big\ Data\ Assignments/A3/Big-Data-Assignment-3/sample.txt Desktop/Big\ Data\ Assignments/A3/Big-Data-Assignment-3/mapper.py Desktop/Big\ Data\ Assignments/A3/Big-Data-Assignment-3/reducer.py hadoop-slave1@20.110.153.135:~

scp Desktop/Big\ Data\ Assignments/A3/Big-Data-Assignment-3/sample.txt Desktop/Big\ Data\ Assignments/A3/Big-Data-Assignment-3/mapper.py Desktop/Big\ Data\ Assignments/A3/Big-Data-Assignment-3/reducer.py hadoop-slave2@20.110.159.120:~

scp Desktop/Big\ Data\ Assignments/A3/Big-Data-Assignment-3/sample.txt
hadoop-master@20.110.56.176:~

uploading mapper.py sample.txt and reducer.py from default /home/hadoop-master to
hadoop-master at default /home/hadoopuser to make it accessible

scp ~/mapper.py hadoopuser@hadoop-master:~
scp ~/reducer.py hadoopuser@hadoop-master:~
scp ~/sample.txt hadoopuser@hadoop-master:~

scp ~/mapper.py hadoopuser@hadoop-slave1:~
scp ~/reducer.py hadoopuser@hadoop-slave1:~
scp ~/sample.txt hadoopuser@hadoop-slave1:~

scp ~/mapper.py hadoopuser@hadoop-slave2:~
scp ~/reducer.py hadoopuser@hadoop-slave2:~
scp ~/sample.txt hadoopuser@hadoop-slave2:~

/path/to/hadoop-streaming.jar
find / -name "hadoop-streaming*.jar" 2>/dev/null

In out case it was as follows
/usr/local/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.3.1.jar

run mapreduce

hadoop jar /usr/local/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.3.1.jar -file mapper.py -mapper mapper.py
-file reducer.py -reducer reducer.py -input /user/sample.txt -output /user/result

# copy the output.txt created in hadoopuser to home directory of hadoop-master
sudo scp hadoopuser@hadoop-master:output.txt /home/hadoop-master

copy the output.txt from home directory of hadoop-master to local machine

scp hadoop-master@20.110.56.176:output.txt Desktop/Big\ Data\
Assignments/A3/Big-Data-Assignment-3