1) **Single node cluster**

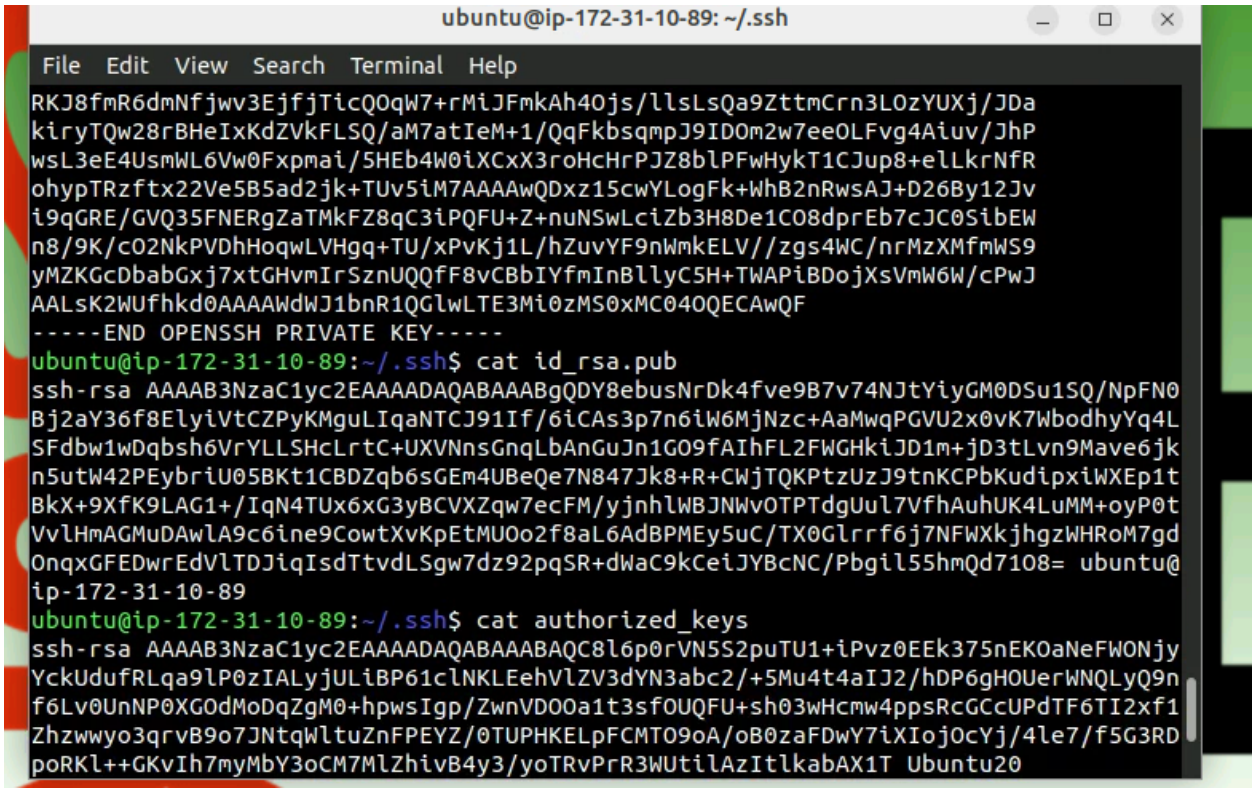
Sudo apt update

```
Hello python
ubuntu@ip-172-31-10-89:~$ sudo apk update
sudo: apk: command not found
ubuntu@ip-172-31-10-89:~$ sudo apt update
Hit:1 http://ap-south-1.ec2.archive.ubuntu.com/ubuntu jammy InRelease
Get:2 http://ap-south-1.ec2.archive.ubuntu.com/ubuntu jammy-updates InRelease [1
19 kB]
```

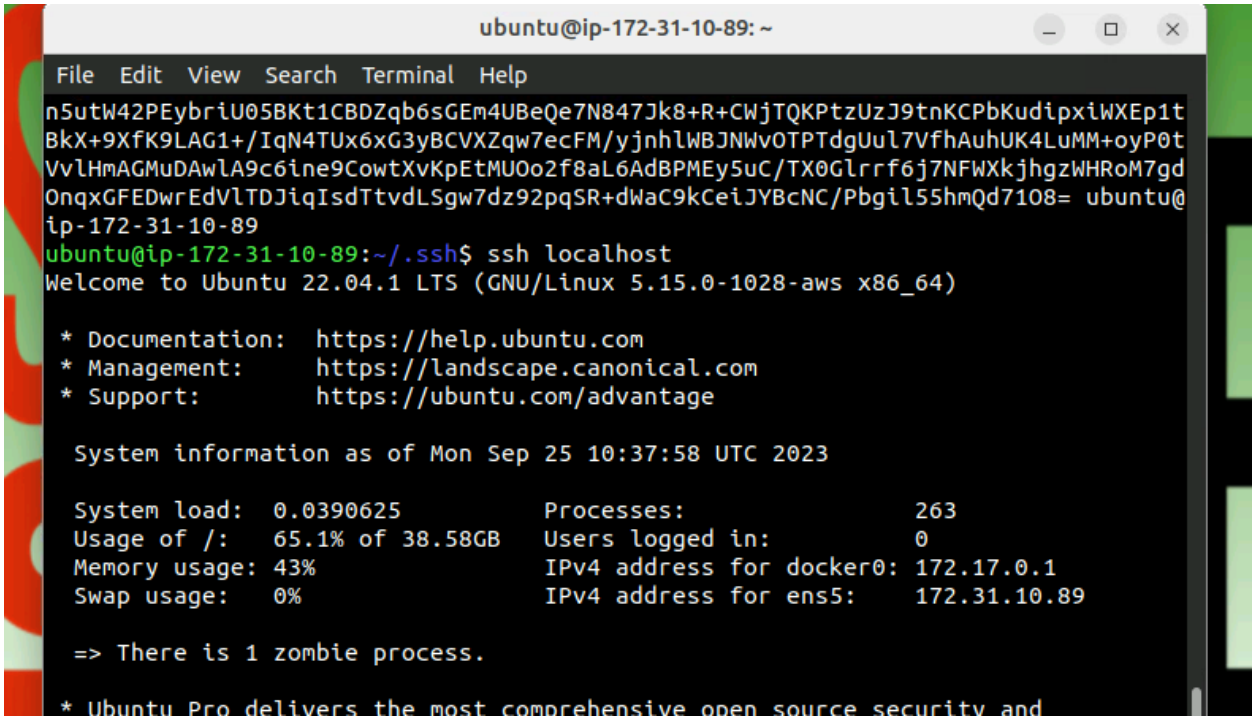
ssh -keygen

cat id_rsa.pub >> authorizedkeys

ssh localhost



Checking localhost



Downloading hadoop

```
ubuntu@ip-172-31-10-89:~$ wget https://archive.apache.org/dist/hadoop/core/hadoop-2.5.0/hadoop-2.5.0.tar.gz
```

Extracting tar files

```
ubuntu@ip-172-31-10-89:~$ tar -xvzf hadoop-2.5.0.tar.gz hadoop-2.5.0/
```

Confirm the files

```
ubuntu@ip-172-31-10-89:~$ ls
 Desktop                          Templates            'python git'
 Documents                        Videos               'python git 1'
 Downloads                        aws                  'python git 2'
 HR_Hiring_details_transaction.csv eclipse              sample.txt
 Music                            eclipse-workspace     snap
 Pictures                         hadoop-2.5.0.tar.gz   thinclient_drives
 Public                           kafka_2.13-3.5.1.tgz
 PycharmProjects                  pyt
```

Moving downloaded file to hadoop

```
ubuntu@ip-172-31-10-89:~$ sudo mv hadoop-2.5.0 /usr/local/hadoop
```

Checking the path

```
ubuntu@ip-172-31-10-89:~$ cd /usr/local/hadoop/
```

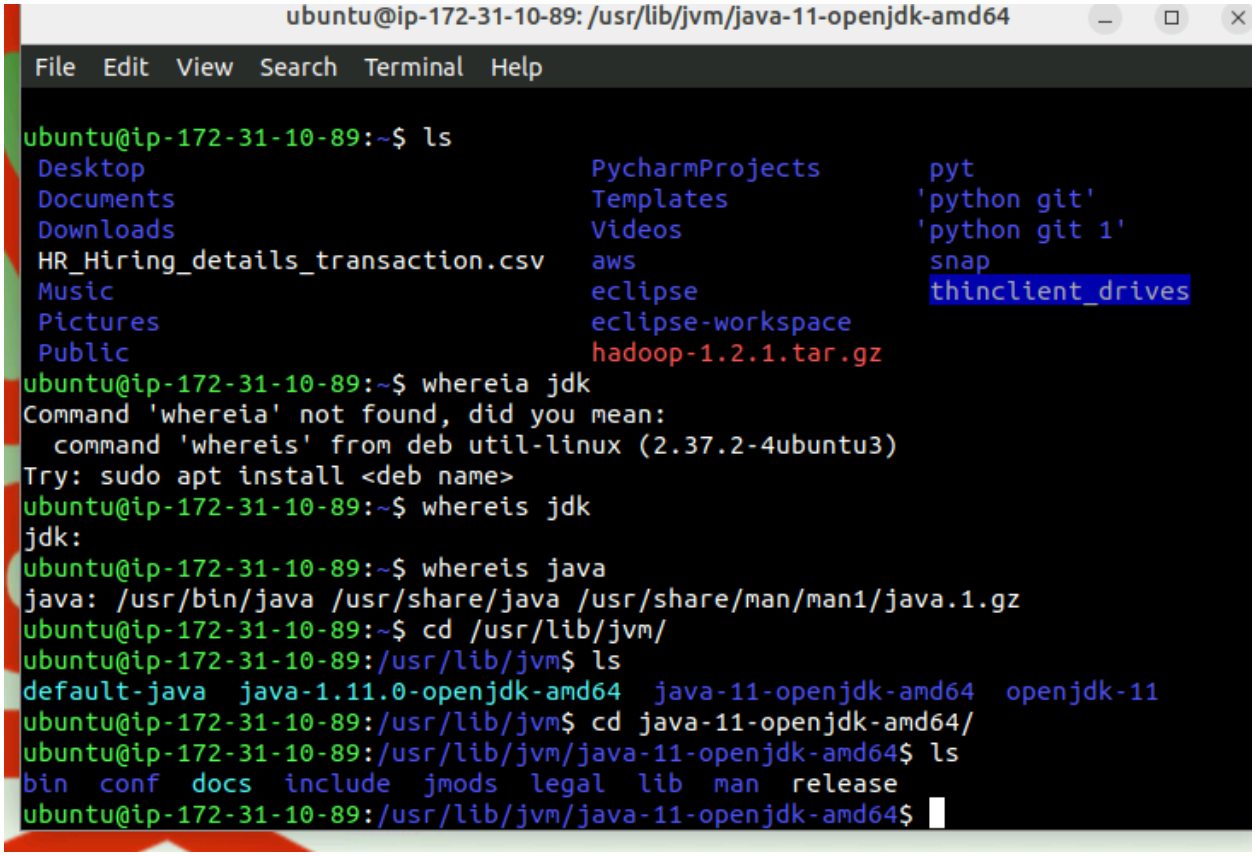
Configure hadoop.env.sh

```
# The directory where pid files are stored. /tmp by default.
# NOTE: this should be set to a directory that can only be written to by
#       the user that will run the hadoop daemons.  Otherwise there is the
#       potential for a symlink attack.
export HADOOP_PID_DIR=${HADOOP_PID_DIR}
export HADOOP_SECURE_DN_PID_DIR=${HADOOP_PID_DIR}

# A string representing this instance of hadoop. $USER by default.
export HADOOP_IDENT_STRING=$USER

export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
export HADOOP_OPTS=-Djava.net.preferIPV4Stack=true
```

Installing Checking Java version:



Configure all files:

core-site.xml

<property>

<name>fs.default.name</name>

<value>hdfs://localhost:9000</value>

</property>

<property>

<name>hadoop.tmp.dir</name>

<value>/usr/local/hadoop/tmp</value>

</property>

hdfs-site.xml

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

mapred.xml

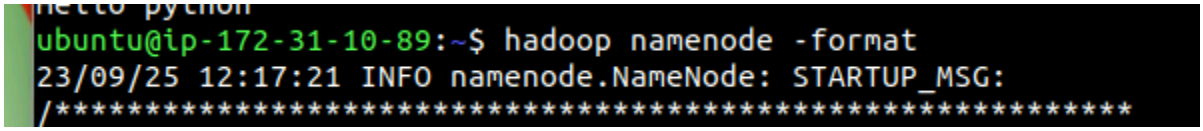
<property>

<name>mapred.job.tracker</name>

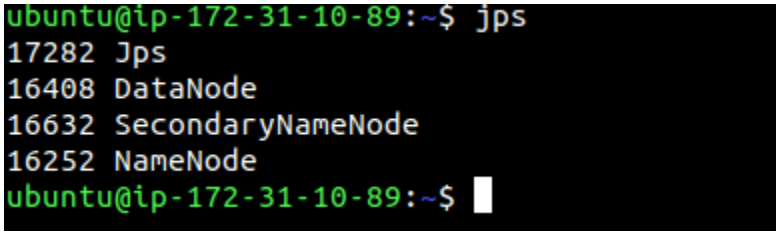
<value>hdfs://localhost:9001</value>

</property>

Starting namenode



Checking demons after (start-all.sh)



Checking localhost(127.0.0.01:50070)

NameNode 'localhost:9000' (active)

Started:	Fri Oct 06 04:58:02 UTC 2023
Version:	2.5.0, 1616291
Compiled:	2014-08-06T17:31Z by jenkins from branch-2.5.0
Cluster ID:	CID-c97b2ee0-cab4-4224-a725-d6ba0d823e88
Block Pool ID:	BP-669782240-172.31.10.89-1696511311332

Browse the filesystem
NameNode Logs

Cluster Summary

Security is *OFF*
1 files and directories, 0 blocks = 1 total.
Heap Memory used 41.60 MB is 37% of Commited Heap Memory 110 MB. Max Heap Memory is 1000 MB.
Non Heap Memory used 32.06 MB is 91% of Commited Non Heap Memory 35.19 MB. Max Non Heap Memory is -1 B.

Configured Capacity	: 38.58 GB
DFS Used	: 28 KB
Non DFS Used	: 29.99 GB