Install hadoop in linux (ubentu):

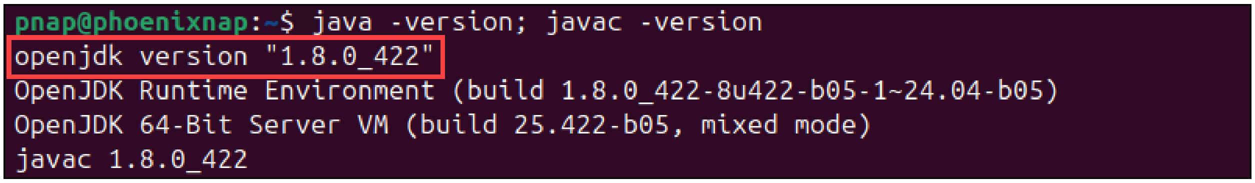
# Install JDK on Ubuntu:

$ sudo apt update

$ sudo apt install openjdk-8-jdk -y

- Once the installation process is complete, verify the current Java version:

$ java -version; javac -version

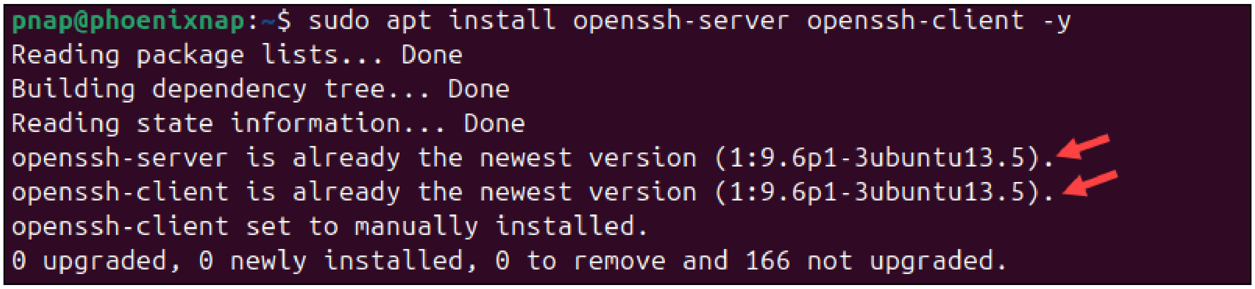


```
pnap@phoenixnap:~$ java -version; javac -version
openjdk version "1.8.0_422"
OpenJDK Runtime Environment (build 1.8.0_422-8u422-b05-1~24.04-b05)
OpenJDK 64-Bit Server VM (build 25.422-b05, mixed mode)
javac 1.8.0_422
```

# Install OpenSSH on Ubuntu

$ sudo apt install openssh-server openssh-client -y

- In the example below, the output confirms that the latest version is already installed.



```
pnap@phoenixnap:~$ sudo apt install openssh-server openssh-client -y
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
openssh-server is already the newest version (1:9.6p1-3ubuntu13.5).
openssh-client is already the newest version (1:9.6p1-3ubuntu13.5).
openssh-client set to manually installed.
0 upgraded, 0 newly installed, 0 to remove and 166 not upgraded.
```

# Create Hadoop User

$ sudo adduser hdoop

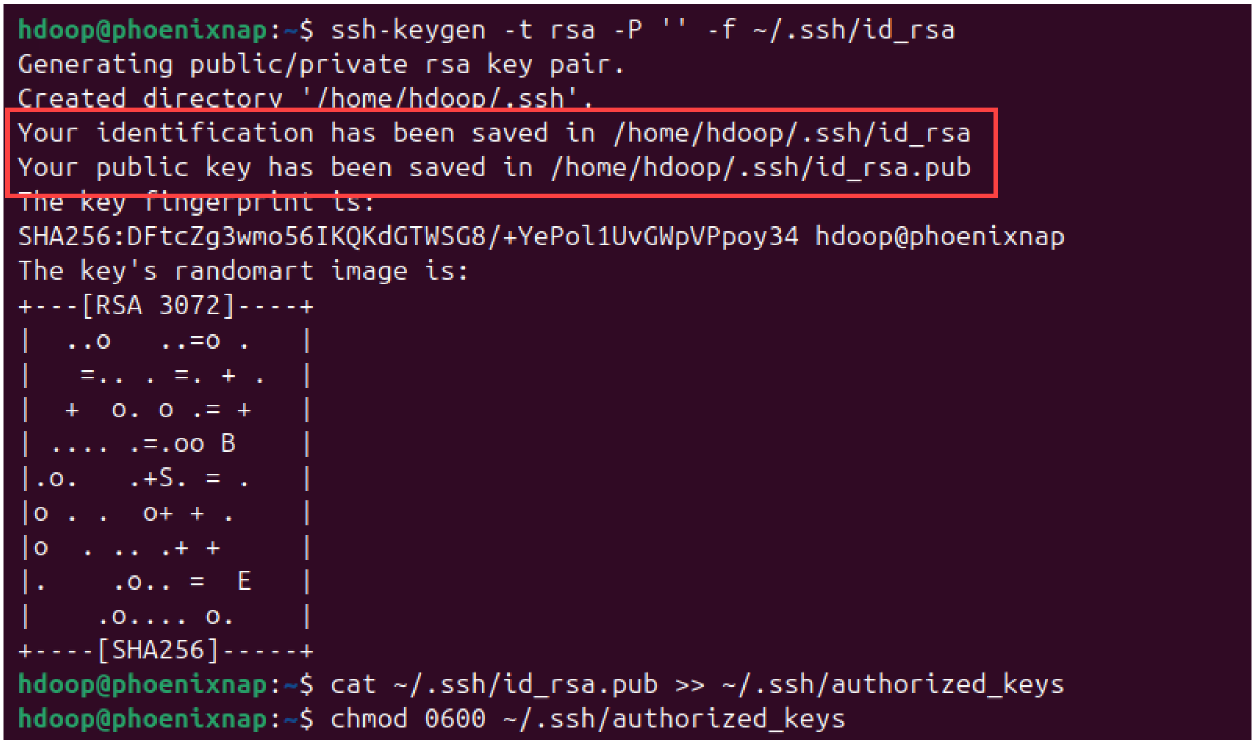
- The username, in this example, is **hdoop**. You are free to use any username and password you see fit.

- Switch to the newly created user and enter the corresponding password:

$ su – hdoop

# **Enable Passwordless SSH for Hadoop User**

$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa



```
hdoop@phoenixnap:~$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
Generating public/private rsa key pair.
Created directory '/home/hdoop/.ssh'.
Your identification has been saved in /home/hdoop/.ssh/id_rsa
Your public key has been saved in /home/hdoop/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:DFtcZg3wmo56IKQKdGTWSG8/+YePol1UvGWpVPpoy34 hdoop@phoenixnap
The key's randomart image is:
+---[RSA 3072]----+
|  ..o   ..=o .   |
|   =.. . =. + .  |
|  +  o. o .= +   |
| .... .=.oo B    |
|.o.   .+S. = .   |
|o . .  o+ + .    |
|o  . .. .+ +     |
|.    .o.. =  E   |
|    .o.... o.    |
+----[SHA256]-----+
hdoop@phoenixnap:~$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
hdoop@phoenixnap:~$ chmod 0600 ~/.ssh/authorized_keys
```

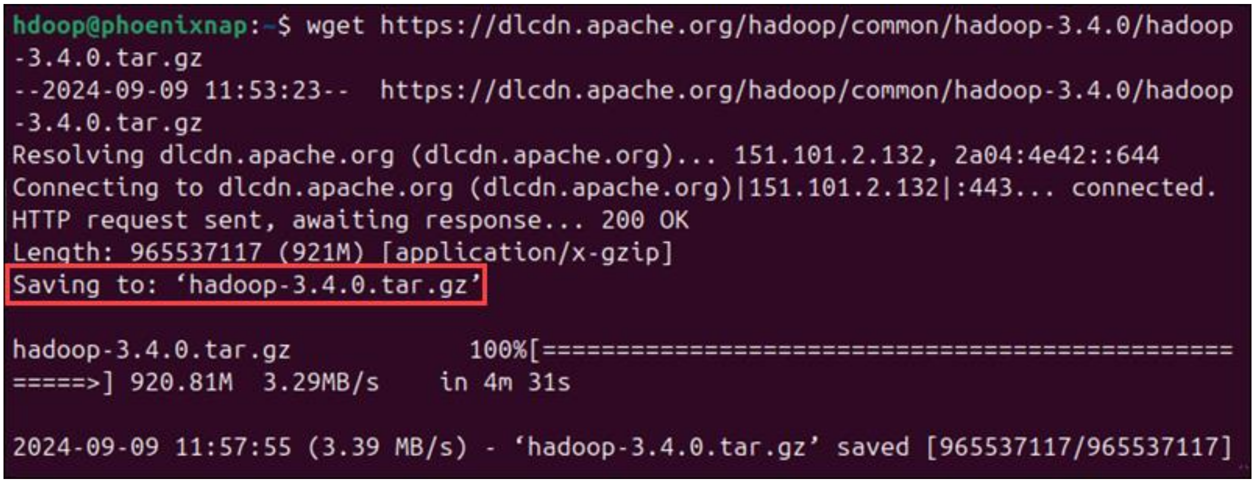
$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys

$ chmod 0600 ~/.ssh/authorized_keys

$ ssh localhost

# **Download and Install Hadoop on Ubuntu**

$ wget https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz



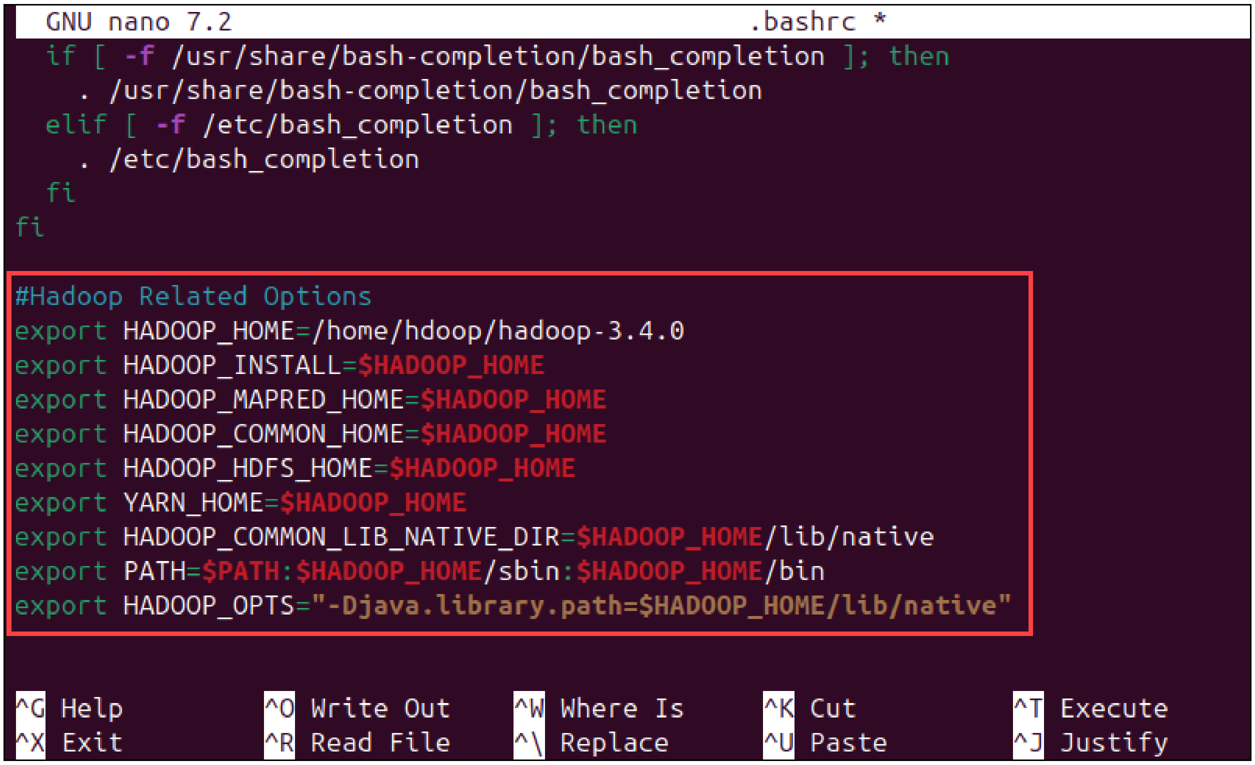
```
hdoop@phoenixnap:~$ wget https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop
-3.4.0.tar.gz
--2024-09-09 11:53:23--  https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop
-3.4.0.tar.gz
Resolving dlcdn.apache.org (dlcdn.apache.org)... 151.101.2.132, 2a04:4e42::644
Connecting to dlcdn.apache.org (dlcdn.apache.org)|151.101.2.132|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 965537117 (921M) [application/x-gzip]
Saving to: ‘hadoop-3.4.0.tar.gz’

hadoop-3.4.0.tar.gz          100%[==============================================
=====>] 920.81M  3.29MB/s    in 4m 31s

2024-09-09 11:57:55 (3.39 MB/s) - ‘hadoop-3.4.0.tar.gz’ saved [965537117/965537117]
```

$ tar xzf hadoop-3.4.0.tar.gz

$nano .bashrc

-insert this :

```
#Hadoop Related Options
export HADOOP_HOME=/home/hdoop/hadoop-3.4.0
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```



```
  GNU nano 7.2                                        .bashrc *
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi

#Hadoop Related Options
export HADOOP_HOME=/home/hdoop/hadoop-3.4.0
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```

-click Ctrl+s and Ctrl+x to go out the editor

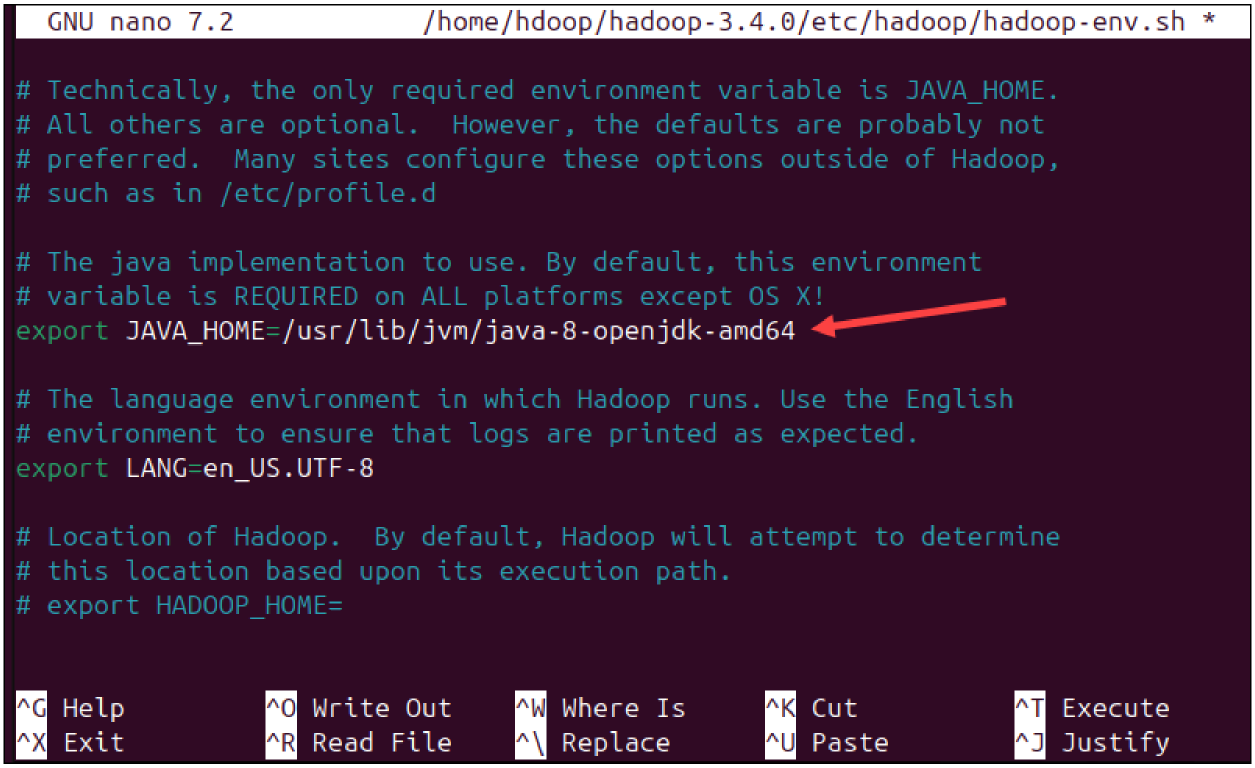
$source ~/.bashrc

#**Edit hadoop-env.sh File**

$nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh

- Uncomment the **$JAVA_HOME** variable (i.e., remove the **#** sign) and add the full path to the OpenJDK installation on your system. If you have installed the same version as presented in the first part of this tutorial, add the following line:

$export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64



```
  GNU nano 7.2              /home/hdoop/hadoop-3.4.0/etc/hadoop/hadoop-env.sh *

# Technically, the only required environment variable is JAVA_HOME.
# All others are optional.  However, the defaults are probably not
# preferred.  Many sites configure these options outside of Hadoop,
# such as in /etc/profile.d

# The java implementation to use. By default, this environment
# variable is REQUIRED on ALL platforms except OS X!
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

# The language environment in which Hadoop runs. Use the English
# environment to ensure that logs are printed as expected.
export LANG=en_US.UTF-8

# Location of Hadoop.  By default, Hadoop will attempt to determine
# this location based upon its execution path.
# export HADOOP_HOME=
```

-click Ctrl+s and Ctrl+x to go out the editor

#Edit core-site.xml File

Note: keep this number , we will use it alot

$hostname -I | awk '{print $1}'

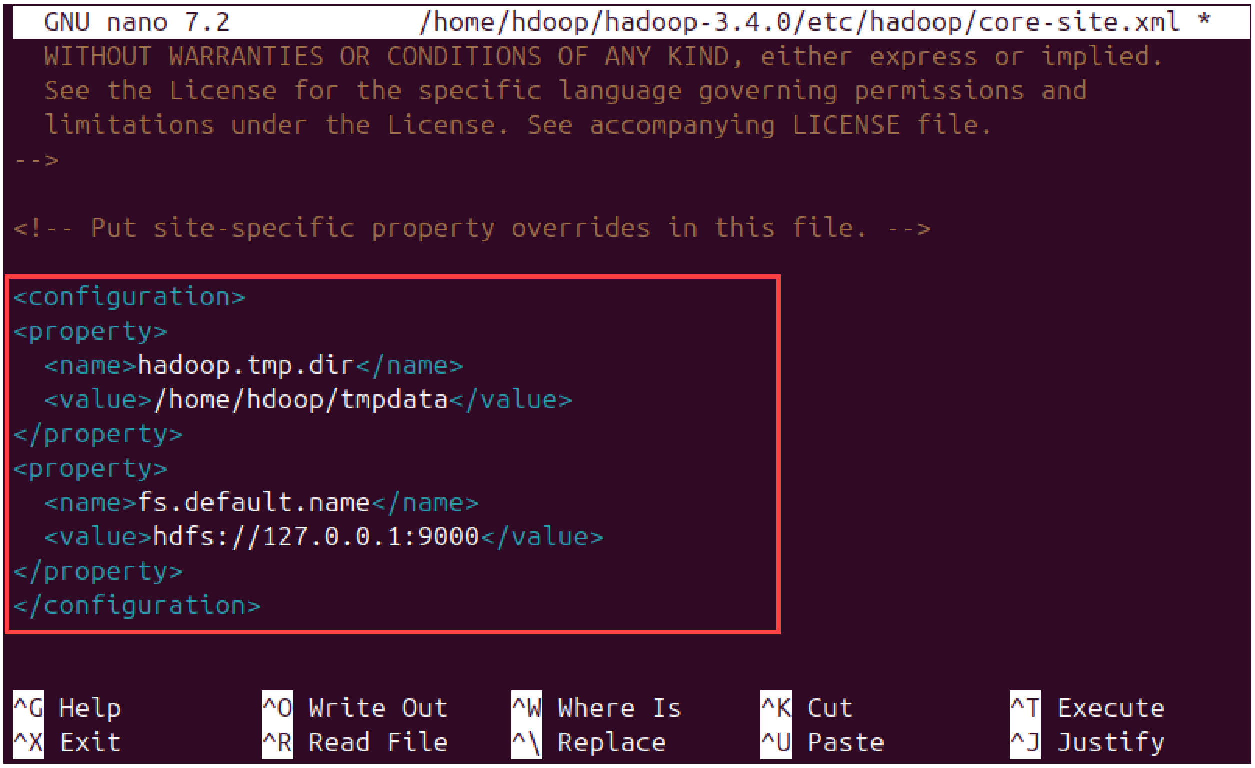
-copy the number address

$nano $HADOOP_HOME/etc/hadoop/core-site.xml

-Add the following configuration to override the default values for the temporary directory and add your HDFS URL to replace the default local file system setting:

```
<configuration>
<property>
 <name>hadoop.tmp.dir</name>
 <value>/home/hdoop/tmpdata</value>
```

```
</property>
<property>
 <name>fs.default.name</name>
 <value>hdfs://put the number here:9000</value>
</property>
</configuration>
```



```
  GNU nano 7.2              /home/hdoop/hadoop-3.4.0/etc/hadoop/core-site.xml *
  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>/home/hdoop/tmpdata</value>
</property>
<property>
  <name>fs.default.name</name>
  <value>hdfs://127.0.0.1:9000</value>
</property>
</configuration>

^G Help       ^O Write Out  ^W Where Is   ^K Cut        ^T Execute
^X Exit       ^R Read File  ^\ Replace    ^U Paste      ^J Justify
```

#Edit hdfs-site.xml File

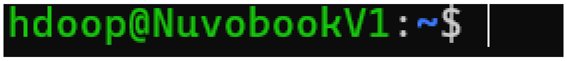
$ sudo nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml

-Add the following configuration to the file and, if needed, adjust the NameNode and DataNode directories to your custom locations:

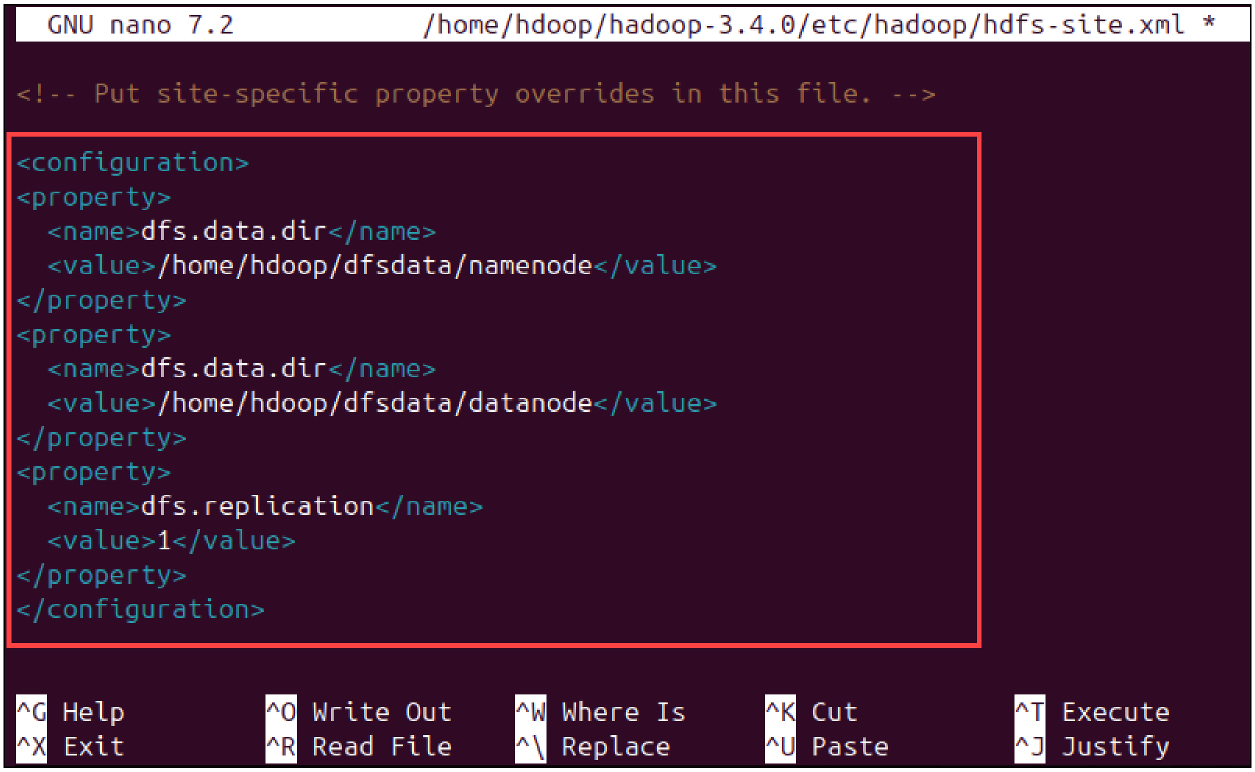
```
<configuration>
<property>
```

```
 <name>dfs.data.dir</name>
 <value>/home/hdoop/dfsdata/namenode</value>
</property>
<property>
 <name>dfs.data.dir</name>
 <value>/home/hdoop/dfsdata/datanode</value>
</property>
<property>
 <name>dfs.replication</name>
 <value>1</value>
</property>
</configuration>
```

Note: -make sure you replace “hdoop” with your user, you can find it in the command line,like here :



```
hdoop@NuvobookV1:~$
```



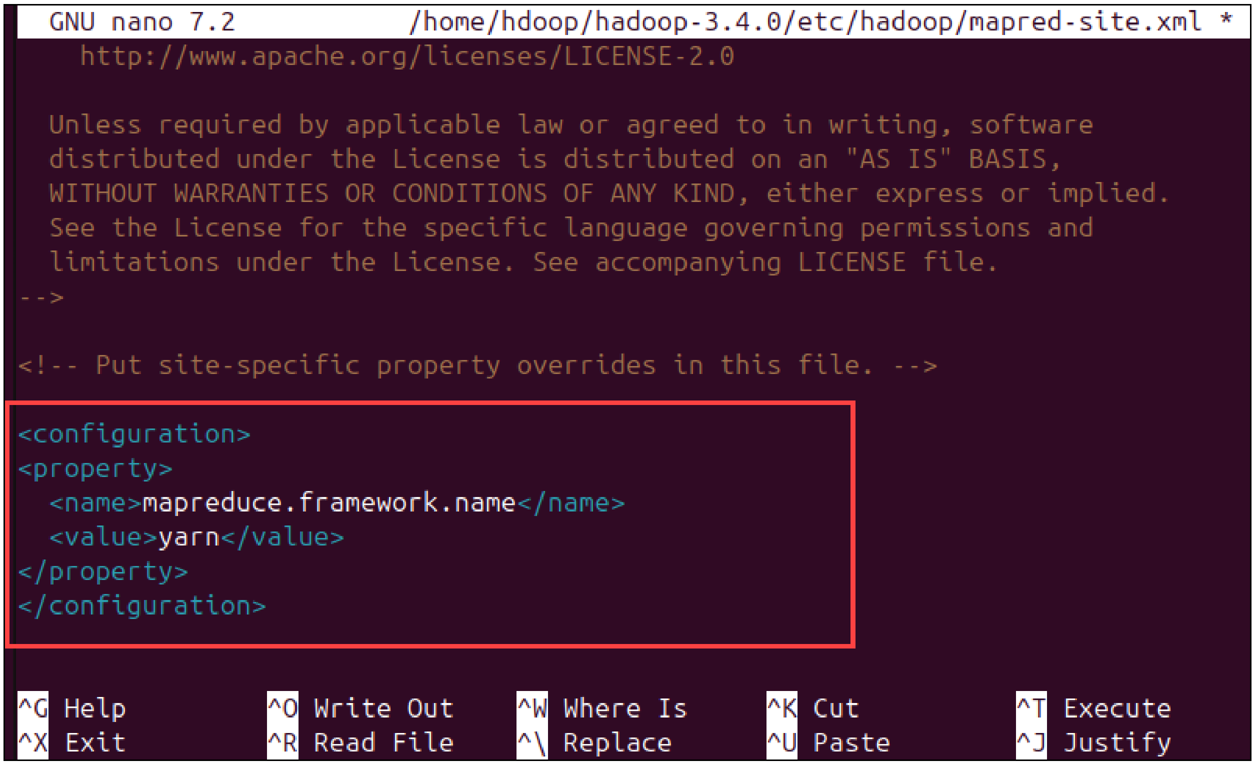
```
<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
  <name>dfs.data.dir</name>
  <value>/home/hdoop/dfsdata/namenode</value>
</property>
<property>
  <name>dfs.data.dir</name>
  <value>/home/hdoop/dfsdata/datanode</value>
</property>
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
</configuration>
```

#Edit mapred-site.xml File

$sudo nano $HADOOP_HOME/etc/hadoop/mapred-site.xml

-Add the following configuration to change the default MapReduce framework name value to **yarn**:

```
<configuration>
<property>
 <name>mapreduce.framework.name</name>
 <value>yarn</value>
</property>
</configuration>
```



```
  GNU nano 7.2            /home/hdoop/hadoop-3.4.0/etc/hadoop/mapred-site.xml *
    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>

^G Help        ^O Write Out   ^W Where Is    ^K Cut         ^T Execute
^X Exit        ^R Read File   ^\ Replace     ^U Paste       ^J Justify
```

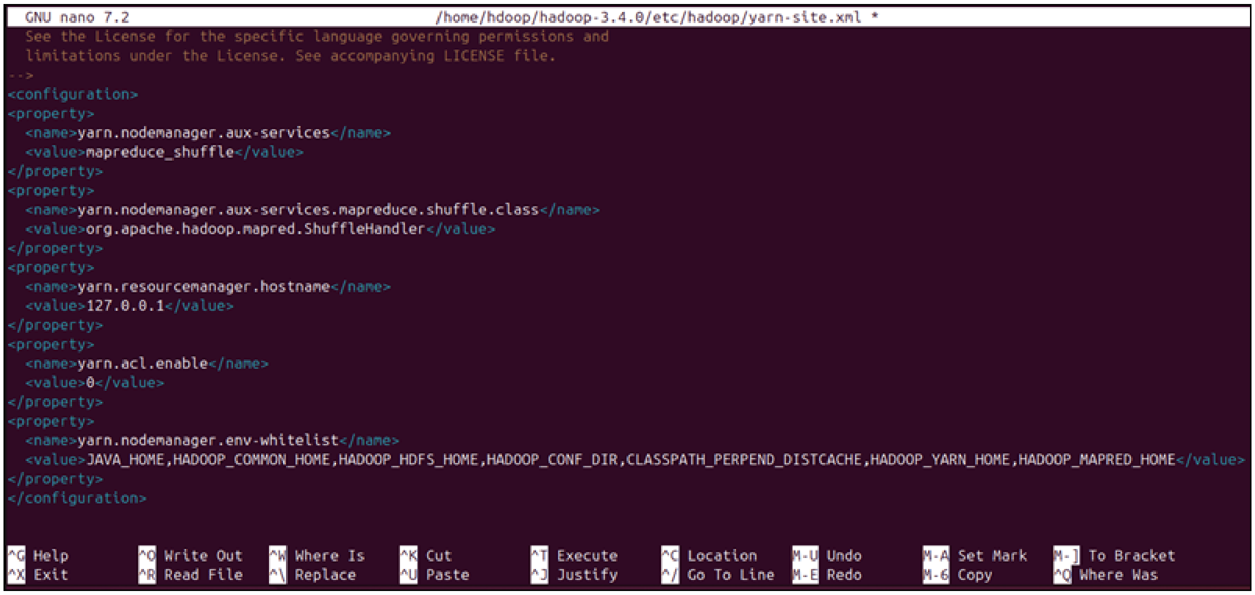
Edit yarn-site.xml File

$nano $HADOOP_HOME/etc/hadoop/yarn-site.xml

-Append the following configuration to the file:

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

```
</property>
<property>
 <name>yarn.resourcemanager.hostname</name>
 <value>put the same number that you puut before here</value>
</property>
<property>
 <name>yarn.acl.enable</name>
 <value>0</value>
</property>
<property>
 <name>yarn.nodemanager.env-whitelist</name>

<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
</property>
</configuration>
```

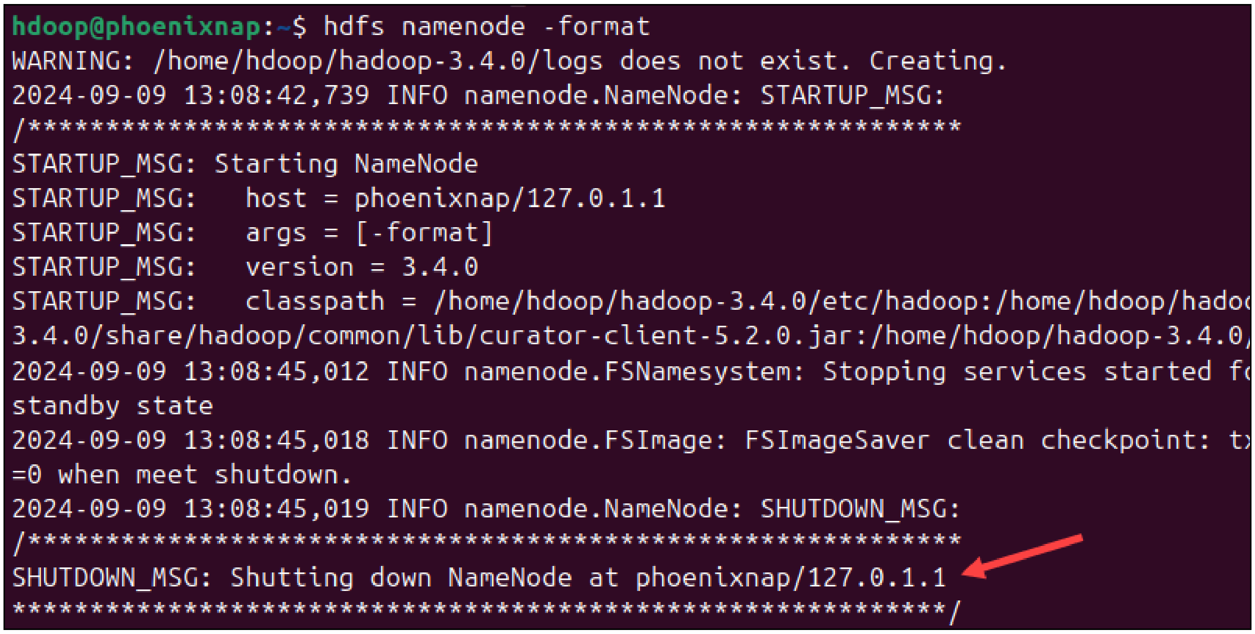


```
  GNU nano 7.2                    /home/hdoop/hadoop-3.4.0/etc/hadoop/yarn-site.xml *
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->
<configuration>
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
<property>
  <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
  <value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>127.0.0.1</value>
</property>
<property>
  <name>yarn.acl.enable</name>
  <value>0</value>
</property>
<property>
  <name>yarn.nodemanager.env-whitelist</name>
  <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
</property>
</configuration>
```

Format HDFS NameNode

$hdfs namenode -format

-The shutdown notification signifies the end of the NameNode format process.



```
hdoop@phoenixnap:~$ hdfs namenode -format
WARNING: /home/hdoop/hadoop-3.4.0/logs does not exist. Creating.
2024-09-09 13:08:42,739 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = phoenixnap/127.0.1.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.4.0
STARTUP_MSG:   classpath = /home/hdoop/hadoop-3.4.0/etc/hadoop:/home/hdoop/hado
3.4.0/share/hadoop/common/lib/curator-client-5.2.0.jar:/home/hdoop/hadoop-3.4.0/
2024-09-09 13:08:45,012 INFO namenode.FSNamesystem: Stopping services started fo
standby state
2024-09-09 13:08:45,018 INFO namenode.FSImage: FSImageSaver clean checkpoint: tx
=0 when meet shutdown.
2024-09-09 13:08:45,019 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at phoenixnap/127.0.1.1
************************************************************/
```

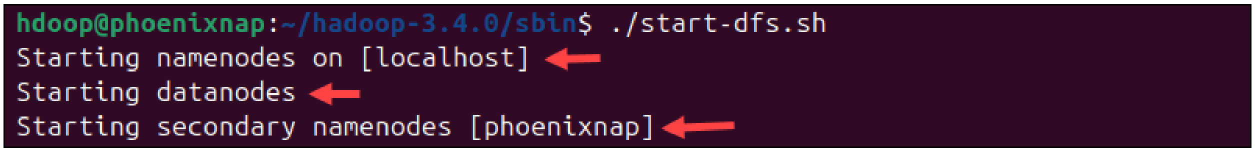
Start Hadoop Cluster

$cd

$cd *hadoop-3.4.0/sbin*

- execute the following command to start the NameNode and DataNode:

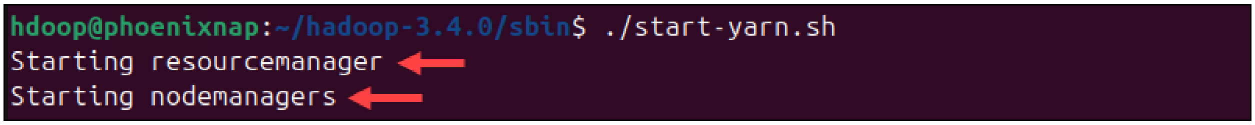
$./start-dfs.sh



```
hdoop@phoenixnap:~/hadoop-3.4.0/sbin$ ./start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [phoenixnap]
```

- Once the *namenode*, *datanodes*, and *secondary namenode* are up and running, start the YARN resource and *nodemanagers* by typing:

$./start-yarn.sh

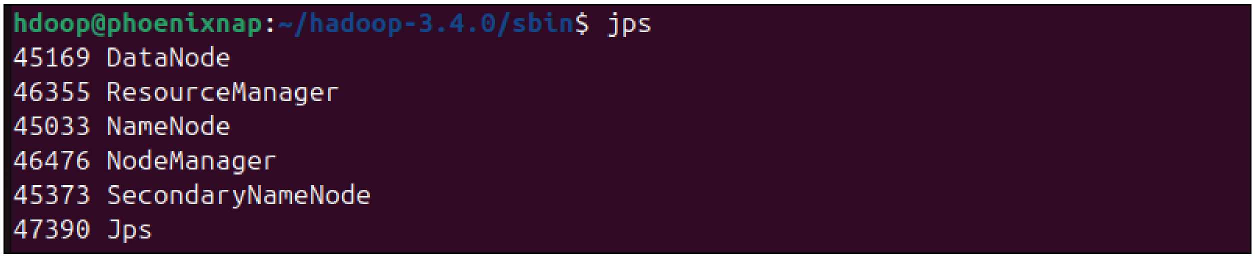


```
hdoop@phoenixnap:~/hadoop-3.4.0/sbin$ ./start-yarn.sh
Starting resourcemanager
Starting nodemanagers
```

-Run the following command to check if all the daemons are active and running as Java processes:

$jps

-If everything works as intended, the resulting list of running Java processes contains all the HDFS and YARN daemons.

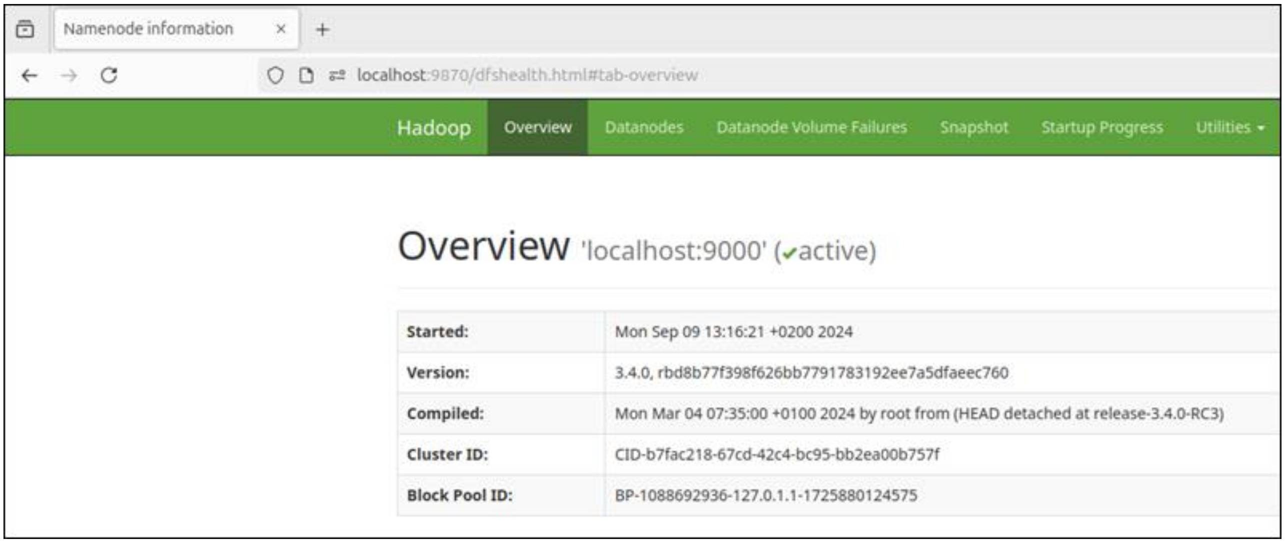


```
hdoop@phoenixnap:~/hadoop-3.4.0/sbin$ jps
45169 DataNode
46355 ResourceManager
45033 NameNode
46476 NodeManager
45373 SecondaryNameNode
47390 Jps
```

Access Hadoop from Browser

#nameNode

http://localhost:9870

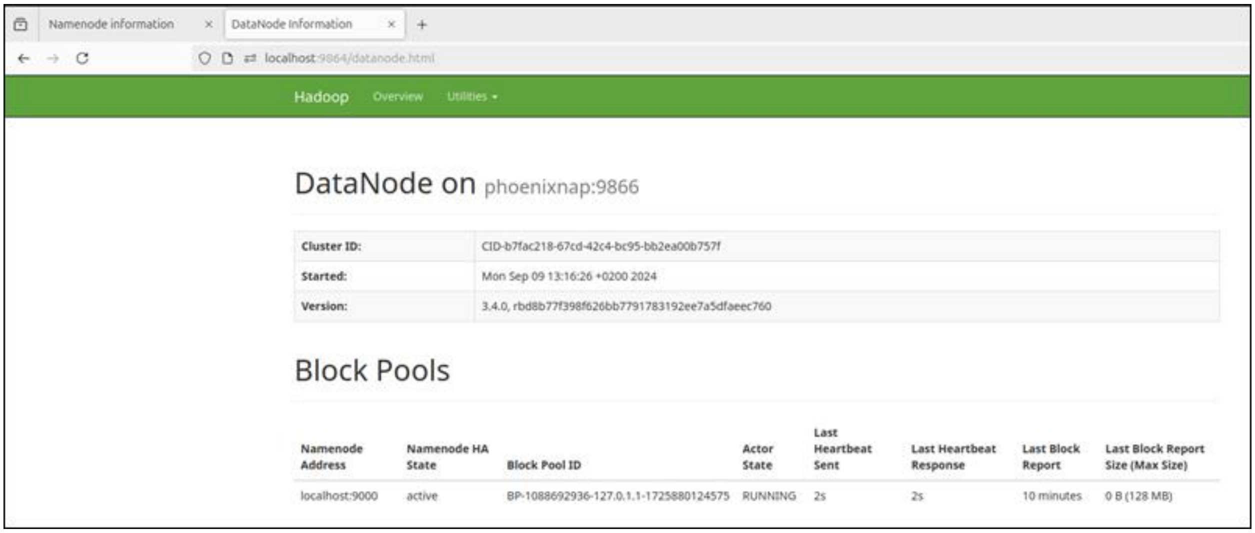


Overview 'localhost:9000' (✓active)

Started:	Mon Sep 09 13:16:21 +0200 2024
Version:	3.4.0, rbd8b77f398f626bb7791783192ee7a5dfaeec760
Compiled:	Mon Mar 04 07:35:00 +0100 2024 by root from (HEAD detached at release-3.4.0-RC3)
Cluster ID:	CID-b7fac218-67cd-42c4-bc95-bb2ea00b757f
Block Pool ID:	BP-1088692936-127.0.1.1-1725880124575

#dataNode:

http://localhost:9864



DataNode on phoenixnap:9866

Cluster ID:	CID-b7fac218-67cd-42c4-bc95-bb2ea00b757f
Started:	Mon Sep 09 13:16:26 +0200 2024
Version:	3.4.0, rbd8b77f398f626bb7791783192ee7a5dfaeec760

Block Pools

Namenode Address	Namenode HA State	Block Pool ID	Actor State	Last Heartbeat Sent	Last Heartbeat Response	Last Block Report	Last Block Report Size (Max Size)
localhost:9000	active	BP-1088692936-127.0.1.1-1725880124575	RUNNING	2s	2s	10 minutes	0 B (128 MB)

#resorce manager:

http:// put the same number here:8088