# LAB - 04

## HADOOP MULTI NODE CLUSTER SETUP

**STEP - 1: Know the IP address of your machine:**

```
eno1: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>
        inet 192.168.28.11  netmask 255.255.255.0
.255
        inet6 fe80::2a98:d6d9:d4e0:a9a3  prefixlen
k>
        ether 9c:7b:ef:43:63:74  txqueuelen 1000
        RX packets 960  bytes 778276 (778.2 KB)
        RX errors 0  dropped 0  overruns 0  frame
```

**STEP - 2: Check the status of Firewall (whether active or inactive):**

```
○ ufw.service - Uncomplicated firewall
     Loaded: loaded (/lib/systemd/system/ufw.service; enabled; vendor preset>
     Active: inactive (dead) since Wed 2025-03-05 16:03:13 IST; 1s ago
       Docs: man:ufw(8)
    Process: 622 ExecStart=/lib/ufw/ufw-init start quiet (code=exited, statu>
    Process: 4187 ExecStop=/lib/ufw/ufw-init stop (code=exited, status=0/SUC>
   Main PID: 622 (code=exited, status=0/SUCCESS)
        CPU: 325ms

Warning: some journal files were not opened due to insufficient permissions.
~
~
~
~
~
~
~
~
~
~
~
~
lines 1-10/10 (END)
```

**STEP - 3:Verify the connection with the remote machine (slave machine on which you want to disable the firewall):**

```
hadoop@hadoop-clone-11:~$ ssh 192.168.28.12
The authenticity of host '192.168.28.12 (192.168.28.12)' can't be established.
ED25519 key fingerprint is SHA256:AxVWZz6+Io+JzX7rXeFV5uUNvijNO6WQBJgmD7mnSlY.
This host key is known by the following other names/addresses:
    ~/.ssh/known_hosts:1: [hashed name]
    ~/.ssh/known_hosts:7: [hashed name]
    ~/.ssh/known_hosts:8: [hashed name]
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added '192.168.28.12' (ED25519) to the list of known hosts.
Welcome to Ubuntu 22.04.4 LTS (GNU/Linux 6.8.0-52-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/pro

Expanded Security Maintenance for Applications is not enabled.

195 updates can be applied immediately.
55 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable

22 additional security updates can be applied with ESM Apps.
Learn more about enabling ESM Apps service at https://ubuntu.com/esm

New release '24.04.2 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Last login: Wed Mar  5 16:08:11 2025 from 192.168.28.14
```

● **Check for configuration settings in remote machine:**

This is checked by editing the SSH configuration file (/etc/ssh/sshd_config) on the remote machine and looking for the following line:

```
  GNU nano 6.2                    /etc/ssh/sshd_config

# This is the sshd server system-wide configuration file.  See
# sshd_config(5) for more information.

# This sshd was compiled with PATH=/usr/local/sbin:/usr/local/bin:/usr/sbin>

# The strategy used for options in the default sshd_config shipped with
# OpenSSH is to specify options with their default value where
# possible, but leave them commented.  Uncommented options override the
# default value.

Include /etc/ssh/sshd_config.d/*.conf

#Port 22
#AddressFamily any
#ListenAddress 0.0.0.0
#ListenAddress ::

#HostKey /etc/ssh/ssh_host_rsa_key
                [ File '/etc/ssh/sshd_config' is unwritable ]
^G Help         ^O Write Out    ^W Where Is     ^K Cut          ^T Execute
^X Exit         ^R Read File    ^\ Replace      ^U Paste        ^J Justify
```

**STEP - 4: Update and verify the IP address to the master node's IP address:**
My machine is the slave machine here.

```
  GNU nano 6.2      /opt/hadoop/etc/hadoop/hdfs-site.xml *

  <property>
    <name>dfs.datanode.data.dir</name>
    <value>/opt/hadoop/dfs/data</value>
  </property>
  <property>
    <name>dfs.namenode.http-address</name>
    <value>192.168.28.14:50070</value>

  </property>

  <property>
    <name>dfs.namenode.secondary.http-address</name>
   <value>192.168.28.14:50090</value>
  </property>
</configuration>
```

```
  GNU nano 6.2      /opt/hadoop/etc/hadoop/core-site.xml
  limitations under the License. See accompanying LICENSE file.
-->

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

<configuration>
<property>
    <name>fs.defaultFS</name>
        <value>hdfs://192.168.28.14:9000/</value>
  </property>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>/opt/hadoop/tmp</value>
  </property>
</configuration>
```

```
GNU nano 6.2     /opt/hadoop/etc/hadoop/mapred-site.xml *
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>

  <property>
    <name>mapreduce.jobhistory.address</name>
    <value>192.168.28.14:10020</value>
  </property>

  <property>
    <name>mapreduce.jobhistory.webapp.address</name>
    <value>192.168.28.14:19888</value>
  </property>
<property>
  <name>yarn.app.mapreduce.am.env</name>
  <value>HADOOP_MAPRED_HOME=${HADOOP_HOME}</value>
```

**STEP - 5: Every Machine should have the mapping of IP addresses and Hostname, so that we can refer to the remote machines by either of them.**
Editing hosts file in /etc/ folder on all nodes, specify the IP address of each system followed by their host names.

```
hadoop@hadoop-clone-12:~$ cat /etc/hosts
127.0.0.1       localhost
127.0.1.1       celab4-HP-ProDesk-400-G7-Microtower-PC
192.168.28.16   hadoop-clone-16
# The following lines are desirable for IPv6 capable hosts
::1     ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters
# Added by Docker Desktop
# To allow the same kube context to work on the host and the container:
127.0.0.1       kubernetes.docker.internal
# End of section
192.168.28.11 hadoop-clone-11
192.168.28.12 hadoop-clone-12
192.168.28.13 hadoop-clone-13
192.168.28.14 hadoop-clone-14
192.168.28.15 hadoop-clone-15
192.168.28.16 hadoop-clone-16
192.168.28.17 hadoop-clone-17
192.168.28.18 hadoop-clone-18
192.168.28.19 hadoop-clone-19
192.168.28.20 hadoop-clone-20
192.168.28.21 hadoop-clone-21
192.168.28.22 hadoop-clone-22
192.168.28.23 hadoop-clone-23
192.168.28.24 hadoop-clone-24
192.168.28.25 hadoop-clone-25
192.168.28.26 hadoop-clone-26
192.168.28.27 hadoop-clone-27
192.168.28.28 hadoop-clone-28
192.168.28.29 hadoop-clone-29
192.168.28.30 hadoop-clone-30
192.168.28.31 hadoop-clone-31
192.168.28.32 hadoop-clone-32
192.168.28.33 hadoop-clone-33
192.168.28.34 hadoop-clone-34
192.168.28.35 hadoop-clone-35
192.168.28.36 hadoop-clone-36
192.168.28.37 hadoop-clone-37
192.168.28.38 hadoop-clone-38
192.168.28.39 hadoop-clone-39
192.168.28.40 hadoop-clone-40
192.168.28.41 hadoop-clone-41
192.168.28.42 hadoop-clone-42
192.168.28.43 hadoop-clone-43
192.168.28.44 hadoop-clone-44
192.168.28.45 hadoop-clone-45
192.168.28.46 hadoop-clone-46
192.168.28.47 hadoop-clone-47
192.168.28.48 hadoop-clone-48
192.168.28.49 hadoop-clone-49
192.168.28.50 hadoop-clone-50
192.168.28.51 hadoop-clone-51
192.168.28.52 hadoop-clone-52
```

Open slaves /opt/... Save

```
hadoop-clone-11
hadoop-clone-12
hadoop-clone-13
hadoop-clone-14
hadoop-clone-15
hadoop-clone-16
hadoop-clone-17
hadoop-clone-18
hadoop-clone-19
hadoop-clone-20
hadoop-clone-21
hadoop-clone-22
hadoop-clone-23
hadoop-clone-24
hadoop-clone-25
hadoop-clone-26
hadoop-clone-27
hadoop-clone-28
hadoop-clone-29
hadoop-clone-30
hadoop-clone-31
hadoop-clone-32
hadoop-clone-33
hadoop-clone-34
hadoop-clone-35
hadoop-clone-36
hadoop-clone-37
hadoop-clone-38
hadoop-clone-39
hadoop-clone-40
hadoop-clone-41
hadoop-clone-42
hadoop-clone-43
hadoop-clone-44
hadoop-clone-45
hadoop-clone-46
hadoop-clone-47
hadoop-clone-48
hadoop-clone-49
hadoop-clone-50
hadoop-clone-51
hadoop-clone-52
hadoop-clone-53
hadoop-clone-54
hadoop-clone-55
hadoop-clone-56
```

**STEP - 6: Setup ssh in every node such that they can communicate with one another without any prompt for password.**

● Generate the RSA Key Pair

```
hadoop@hadoop-clone-11:~$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hadoop/.ssh/id_rsa):
/home/hadoop/.ssh/id_rsa already exists.
Overwrite (y/n)? y
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/hadoop/.ssh/id_rsa
Your public key has been saved in /home/hadoop/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:uxsiSOKNlGmEwqEIoTX9CF9dQj3IiCuC7uRQEjtjX68 hadoop@hadoop-clone-11
The key's randomart image is:
+---[RSA 3072]----+
|ooo. . =o+.      |
|O.o.o o +.o      |
|B= o =     .     |
|X.= +..          |
|+@o.. . S        |
|==oo   . .       |
|=o... o o        |
| o   E . o       |
|        o.       |
+----[SHA256]-----+
```

**View the Private Key (~/.ssh/id_rsa):**
To view the contents of the private key file, use **cat.**

```
hadoop@hadoop-clone-11:~$ cat ~/.ssh/id_rsa
-----BEGIN OPENSSH PRIVATE KEY-----
b3BlbnNzaC1rZXktdjEAAAAABG5vbmUAAAAEbm9uZQAAAAAAAAABAAABlwAAAAdzc2gtcn
NhAAAAAwEAAQAAAYEAmOl8vlgx9Kkxxpxa0dRWMfyKMdvdd6/I64tuiZFPhzXZI9Vu8ug7
/nqrEvmt00BKwcDPvTH4TYdK7i3mhjODRF3NqELlZxRgLM9WDjpwESkZo5cSiNdmH9qCJU
QxvBtjqq2ApTdtP0/mH5yI7DvY+8TMB5tHa27MF9vlzAPohIygq+6NDXyBqDhNg+L3Cgsf
Z1NajzxOoCTgdWfGHEpDuG4GGMXm6EZMSBUCs8fDZP+5UToJfbcVOmZBRTH+EZzgP6Pbi2
lJSoL0UN0mTAZdKlqg/Xglw8WWd/asnYqpERAmWykhn/ptQKFA5v/F4WvMdqNxyx0RXhs7
6zQ111c85dhCrIDJ/yjQSyNF7zK27WRCo9Ss7czOzpxxf7Y6J9SChvdD0Lnpnxdv HqiIIW
vS57XRgmveRoTbHYmK4KLOVqQPtnU4ZHcEpIxpiJ5U9urNcXvSbmaOTlnmhi1V6nIkUP7i
yxwY7qLrdRZcdWjAMMorHQthySg3mT7VEozTROlfAAAFkJgpsC+YKbAvAAAAB3NzaC1yc2
EAAAGBAJjpfL5YMfSpMcacWtHUVjH8ijHb3XevyOuLbomRT4c12SPVbvLoO/56qxL5rdNA
SsHAz70x+E2HSu4t5oYzg0RdzahC5WcUYCzPVg46cBEpGaOXEojXZh/agiVEMbwbY6qtgK
U3bT9P5h+ciOw72PvEzAebR2tuzBfb5cwD6ISMoKvujQ18gag4TYPi9woLH2dTWo88TqAk
4HVnxhxKQ7huBhjF5uhGTEgVArPHw2T/uVE6CX23FTpmQUUx/hGc4D+j24tpSUqC9FDdJk
wGXSpaoP14JcPFlnf2rJ2KqREQJlspIZ/6bUChQOb/xeFrzHajccsdEV4bO+s0NddXPOXY
QqyAyf8o0EsjRe8ytu1kQqPUrO3Mzs6ccX+2OifUgob3Q9C56Z8Xbx6oiCFr0ue10YJr3k
aE2x2JiuCizlakD7Z1OGR3BKSMaYieVPbqzXF70m5mjk5Z5oYtVepyJFD+4sscGO6i63UW
XHVowDDKKx0LYckoN5k+1RKM00TpXwAAAAMBAAEAAAGAN0g2mqRNDzx6K+lMJk8jiHfaSH
NKKqKpO4sRv/eCboonwj5AGd6fDf4wSjBtSRIV1N2N3UMrF3BhxSFgFcWa4AUfvITO5gjm
Slz9KjCzYmjfBZcxErr5B2waoDl615x5WHPYriKPQxPySRxxHiL26/L2potbczo/3CJafn
stPXqP/gjFAwqv327Xi2ZrVLy3skXe2Cj/7gLXCWsmowSwXgWc+6jUtcXwfyDldLS0bhfL
2lNmqgYeTRo537fWuGxj0Uz/OFgQgV2T2Wh2fhw6reaWkji0zsFWMkkDhS8e9op1iSe7HY
snRq4eR6wTaygrgnvpyxDPbEo1xjsEbDZpglHO3vmIZrRl8oEJDS3aq6TKsezz+umbuKei
Z3zogSzN1DiQ+ZjFJhXwW5LRBVSEJryXkxJ/0kXdsFqY4QQx4PjY7uOgFbt9IV5vFx0Oz9
eRXp0E4H8ea9o5GTMNGzt8BpKaj+TrOntu0nTcSIB+MlInwBKUNvsBFsieCfgK/efZAAAA
wQC1e9uxYajxJURAX8ktwryJgTDWhJMRLPeW9kPTpuvrF9DD9FkHbIqZm88AFLoTcaA0T0
jkJfit4SLMqe9hi24Pnhy+2D7FSVVFyKntvKuIB4BSTaQ8aFQb2BGwiUosYK+pfbfJ5RgP
lx5hIRr/YRfqcabiNXq2cXg5XL7lbgpwkF3LH7uFxJyihLwA40mPylnUfypYIyiiJi3ykj
Gf8dqBOxKZIsg2js+DIIB87XHSWYi4FcRYlHG9ehba0/p//98AAADBAL/KmPPob5HjSyV4
EUXiloyNusgrY2ugSOiE8tOumLoPcHE3Y0efmPtEJ0WTwC3ZSNc8jcTPq6mY2CWynF/scy
pjfwEDNxVOXvVzijyDfkTBaoDY33SZwZdOd9iHpjlnOuTU4BBs+jey2Is4oso64CYcjTkV
epAlHJYGRr5T5daz0OVvfjmkTOswKBNuxUnnY/9MtnNouUw9a3UxB4/lho8QDFc+Xg64qN
QrpBKFx75gA56zZGF6PQnpKQxWYdCttwAAAMEAzBrAlikUnNlz3AlmfB/HyCr96mjpDWnt
4kJ2gBsFXE8G/D02Cp+MeG7Y4WtiQcmfGOPnJx1svORuMXf0ZVvkM34LIPh03lv7wnI+Sp
7X1TdtGnqO9CqMAtxBcBXtWCcHJqIF22P6lTWHdcgnExKoWkmS3IzO6ynbQax+gqRG4cnQ
cygFflIW/BxeRJGTkQv6So3Dd/PorEszWI3QfL1XqWkQWW7JV9mxhffVos3OtIZHK5kfYo
5LGZ2iA0iiG6GZAAAAFmhhZG9vcEBoYWRvb3AtY2xvbmUtMTEBAgME
-----END OPENSSH PRIVATE KEY-----
```

**To view the contents of the public key file:**

**cat ~/.ssh/id_rsa.pub**

The public key will look something like this:

```
hadoop@hadoop-clone-11:~$ cat ~/.ssh/id_rsa.pub
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAABgQCY6Xy+WDH0qTHGnFrR1FYx/Iox2913r
8jri26JkU+HNdkj1W7y6Dv+eqsS+a3TQErBwM+9MfhNh0ruLeaGM4NEXc2oQuVnFGAsz1
YOOnARKRmjlxKI12Yf2oIlRDG8G2OqrYClN20/T+YfnIjsO9j7xMwHm0drbswX2+XMA+i
EjKCr7o0NfIGoOE2D4vcKCx9nU1qPPE6gJOB1Z8YcSkO4bgYYxeboRkxIFQKzx8Nk/7lR
Ogl9txU6ZkFFMf4RnOA/o9uLaUlKgvRQ3SZMBl0qWqD9eCXDxZZ39qydiqkRECZbKSGf+
m1AoUDm/8Xha8x2o3HLHRFeGzvrNDXXVzzl2EKsgMn/KNBLI0XvMrbtZEKj1KztzM7OnH
F/tjon1IKG90PQuemfF28eqIgha9LntdGCa95GhNsdiYrgos5WpA+2dThkdwSkjGmInlT
26s1xe9JuZo5OWeaGLVXqciRQ/uLLHBjuout1Flx1aMAwyisdC2HJKDeZPtUSjNNE6V8=
 hadoop@hadoop-clone-11
```

- **This is how [http://hadoop-master:50070/](http://hadoop-master:50070/) looks like on master machine:**

hadoop-clone-14:9864/datanode.html

Hadoop Overview Utilities

## DataNode on hadoop-clone-14:9866

| | |
|---|---|
| **Cluster ID:** | CID-29264843-a53b-4a15-afbd-70f79750860c |
| **Started:** | Wed Mar 05 16:43:13 +0530 2025 |
| **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) |
|---|---|---|---|---|---|---|---|
| hadoop-clone-14:9000 | active | BP-793875283-192.168.28.14-1727668994178 | RUNNING | 0s | 0s | 21 minutes | 18 B (128 MB) |

## Volume Information

| Directory | StorageType | Capacity Used | Capacity Left | Capacity Reserved | Reserved Space for Replicas | Blocks |
|---|---|---|---|---|---|---|
| /opt/hadoop/dfs/data | DISK | 56 KB | 32.49 GB | 0 B | 0 B | 2 |

Hadoop, 2024.

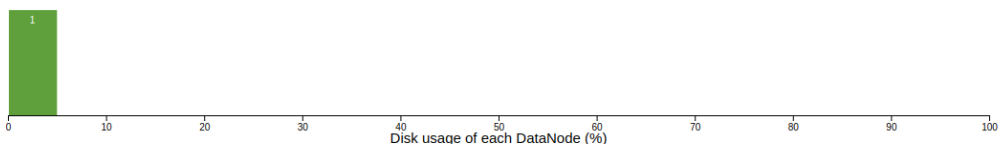
192.168.28.14:50070/dfshealth.html#tab-datanode

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## 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/hadoop-clone-14:9866 (192.168.28.14:9866) | http://hadoop-clone-14:9864 | 0s | 12m | 56 KB | 52.88 GB | 89.99 GB | 2 | 56 KB (0%) | 0% | 3.4.0 |

**Summarised learning:**

The lab work outlines the process for setting up a multi-node Hadoop cluster, including configuring master and slave nodes to communicate via SSH, setting up hostnames and IP address mappings, and ensuring all nodes are synchronized for Hadoop operations. Key steps involve disabling firewalls on each node, generating SSH key pairs for passwordless communication, editing configuration files (like /etc/hosts and Hadoop's core-site.xml, etc.) and ensuring correct slave node listings in the slaves file. Finally, it includes running scripts to automate the process, starting the Hadoop DFS on all nodes, and confirming the cluster's operational status through the NameNode web interface.