

LAB - 7

Name	Keval D Gandevia
Roll Number	CE046
ID	19CEUEG017
Subject	Big Data Analytics

Aim: SSH, User Equivalence, Clusters / Cloud / Horizontal Scaling. Hadoop cluster set up.

❖ **SSH:**

- ✓ SSH command is used for connecting the other machines through IP address.

❖ **Checking my IP address through 'ifconfig' command:**

```

hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop/lab7$ ifconfig
enp3s0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST> mtu 1500
    inet 192.168.26.46 netmask 255.255.255.0 broadcast 192.168.26.255
    inet6 fe80::e307:aab5:a8e1:a137 prefixlen 64 scopeid 0x20<link>
    ether f4:6b:8c:8c:ea:c0 txqueuelen 1000 (Ethernet)
    RX packets 83836 bytes 98499824 (98.4 MB)
    RX errors 0 dropped 93 overruns 0 frame 0
    TX packets 44695 bytes 5016581 (5.0 MB)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0

lo: flags=73<UP,LOOPBACK,RUNNING> mtu 65536
    inet 127.0.0.1 netmask 255.0.0.0
    inet6 ::1 prefixlen 128 scopeid 0x10<host>
    loop txqueuelen 1000 (Local Loopback)
    RX packets 3165 bytes 315287 (315.2 KB)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 3165 bytes 315287 (315.2 KB)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0

hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop/lab7$

```

❖ Connecting to another machine:

- ✓ Now, we have access to that machine. Hence, we can run any command that we want.

```

hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop/lab7$ ssh 192.168.32.46
The authenticity of host '192.168.32.46 (192.168.32.46)' can't be established.
ED25519 key fingerprint is SHA256:aISKCr0D0je2s4RDwAu+CuijMXHxpAjXKgXk9Zc5EaE.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added '192.168.32.46' (ED25519) to the list of known hosts.
hadoop@192.168.32.46's password:
Last login: Thu Sep  1 15:36:11 2022
[hadoop@hadoop-clone-46 ~]$

```

```

hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop/lab7$ ssh 192.168.32.46
The authenticity of host '192.168.32.46 (192.168.32.46)' can't be established.
ED25519 key fingerprint is SHA256:aISKCr0D0je2s4RDwAu+CuijMXHxpAjXKgXk9Zc5EaE.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added '192.168.32.46' (ED25519) to the list of known hosts.
hadoop@192.168.32.46's password:
Last login: Thu Sep  1 15:36:11 2022
[hadoop@hadoop-clone-46 ~]$ pwd
/home/hadoop
[hadoop@hadoop-clone-46 ~]$ ls
commandlab3.txt  Desktop  Downloads  Pictures  Public  Templates  Untitled.ipynb
derby.log       Documents Music     pig_1560125128936.log  simple.txt  units      Videos
[hadoop@hadoop-clone-46 ~]$

```

❖ Checking the status of the firewall. And, it is active right now.

```
hadoop@cclab2-ThinkCentre-neo-50s-Gen-3:~/Desktop/lab7$ ssh 192.168.32.11
The authenticity of host '192.168.32.11 (192.168.32.11)' can't be established.
ED25519 key fingerprint is SHA256:aISKCr0D0je2s4RDwAU+Cu1jMXHxpAjXKgXk9Zc5EaE.
This host key is known by the following other names/addresses:
  ~/.ssh/known_hosts:3: [hashed name]
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added '192.168.32.11' (ED25519) to the list of known hosts.
hadoop@192.168.32.11's password:
Permission denied, please try again.
hadoop@192.168.32.11's password:
Last login: Sat Sep 10 12:01:03 2022 from 192.168.26.23
[hadoop@hadoop-clone-11 ~]$ systemctl firewallstatus
Unknown operation 'firewallstatus'.
[hadoop@hadoop-clone-11 ~]$ systemctl firewall status firewalld
Unknown operation 'firewall'.
[hadoop@hadoop-clone-11 ~]$ systemctl firewall status firewalld.service
Unknown operation 'firewall'.
[hadoop@hadoop-clone-11 ~]$ systemctl status firewalld.service
● firewalld.service - firewalld - dynamic firewall daemon
   Loaded: loaded (/usr/lib/systemd/system/firewalld.service; enabled; vendor preset: enabled)
   Active: active (running) since Sat 2022-09-10 11:38:50 IST; 23min ago
     Docs: man:firewalld(1)
    Main PID: 899 (firewalld)
       Tasks: 2
      CGroup: /system.slice/firewalld.service
              └─899 /usr/bin/python2 -Es /usr/sbin/firewalld --nofork --nopid
[hadoop@hadoop-clone-11 ~]$ systemctl status firewalld.service
● firewalld.service - firewalld - dynamic firewall daemon
   Loaded: loaded (/usr/lib/systemd/system/firewalld.service; enabled; vendor preset: enabled)
   Active: inactive (dead) since Sat 2022-09-10 12:04:03 IST; 1min 3s ago
     Docs: man:firewalld(1)
    Process: 899 ExecStart=/usr/sbin/firewalld --nofork --nopid $FIREWALLD_ARGS (code=exited, status=0/SUCCESS)
    Main PID: 899 (code=exited, status=0/SUCCESS)
[hadoop@hadoop-clone-11 ~]$
```

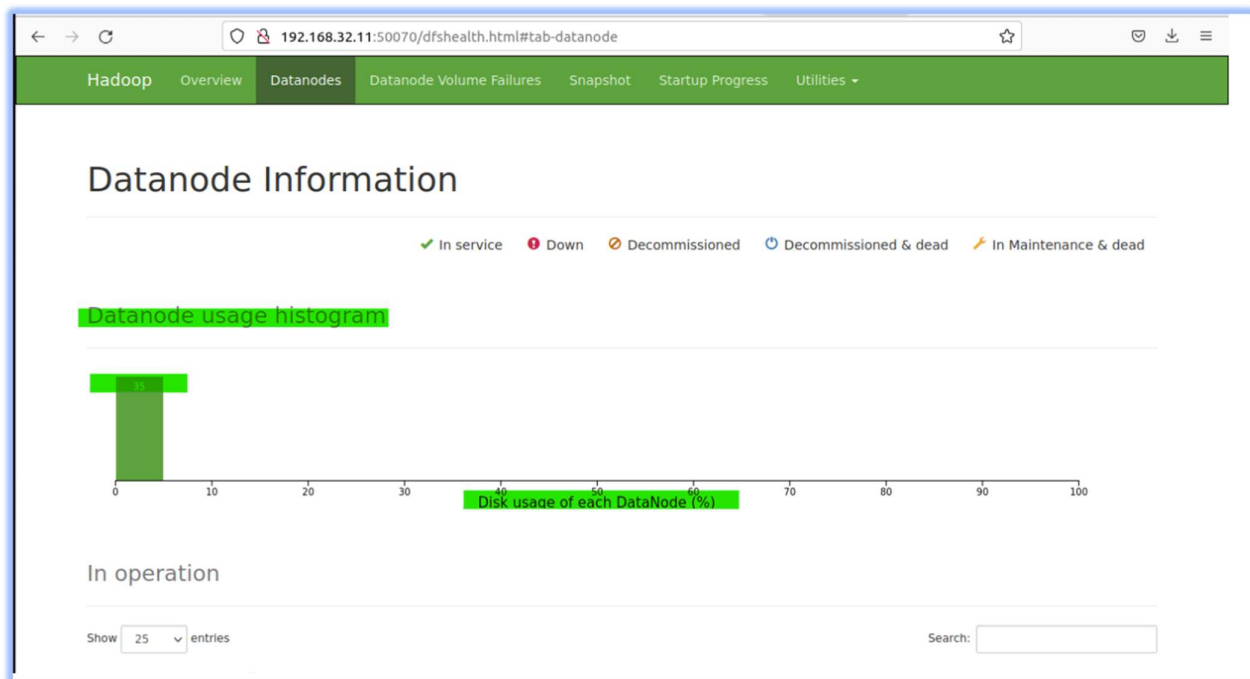
❖ Purpose of the Hadoop cluster:

- ✓ Hadoop cluster is a collection of the various commodity hardware. This hardware components work together as a single unit.
- ✓ In the Hadoop cluster, there are lots of nodes that contain the masters and slaves.
- ✓ The NameNode and ResourceManager work as master node and data node, NodeManager works as a slave.
- ✓ It is designed to store and analyze the large amount of structured and unstructured data in distributed computing environment.

❖ Configuring the Hadoop cluster:

```
[hadoop@hadoop-clone-46 ~]$ ls
1 bin boot data dev etc home lib lib64 lost+found media mnt opt proc root run sbin srv sys tmp usr var
[hadoop@hadoop-clone-46 ~]$ cd /home
[hadoop@hadoop-clone-46 home]$ ls
hadoop user1
[hadoop@hadoop-clone-46 home]$ cd user1
-bash: cd: user1: Permission denied
[hadoop@hadoop-clone-46 home]$
```

```
[hadoop@hadoop-clone-11 ~]$ jps
30320 Jps
28520 NameNode
14937 SecondaryNameNode
29659 DataNode
[hadoop@hadoop-clone-11 ~]$
```



← → ↻ 192.168.32.11:50070/dfshealth.html#tab-datanode ☆ 📄

In operation

Show 25 entries Search:

Node	Http Address	Last contact	Last Block Report	Capacity	Blocks	Block pool used	Version
✓hadoop-clone-11:9866 (192.168.32.11:9866)	http://hadoop-clone-11:9864	0s	2m	309.81 GB <div><div></div></div>	0	32 KB (0%)	3.2.0
✓hadoop-clone-12:9866 (192.168.32.12:9866)	http://hadoop-clone-12:9864	0s	4m	309.81 GB <div><div></div></div>	0	32 KB (0%)	3.2.0
✓hadoop-clone-14:9866 (192.168.32.14:9866)	http://hadoop-clone-14:9864	0s	4m	309.81 GB <div><div></div></div>	0	32 KB (0%)	3.2.0
✓hadoop-clone-15:9866 (192.168.32.15:9866)	http://hadoop-clone-15:9864	0s	4m	309.81 GB <div><div></div></div>	0	32 KB (0%)	3.2.0
✓hadoop-clone-16:9866 (192.168.32.16:9866)	http://hadoop-clone-16:9864	0s	4m	309.81 GB <div><div></div></div>	0	32 KB (0%)	3.2.0
✓hadoop-clone-17:9866 (192.168.32.17:9866)	http://hadoop-clone-17:9864	0s	4m	309.81 GB <div><div></div></div>	0	32 KB (0%)	3.2.0
✓hadoop-clone-18:9866 (192.168.32.18:9866)	http://hadoop-clone-18:9864	0s	4m	309.81 GB <div><div></div></div>	0	32 KB (0%)	3.2.0
✓hadoop-clone-19:9866 (192.168.32.19:9866)	http://hadoop-clone-19:9864	0s	4m	309.81 GB <div><div></div></div>	0	32 KB (0%)	3.2.0
✓hadoop-clone-20:9866	http://hadoop-	0s	4m	309.81 GB <div><div></div></div>	0	32 KB (0%)	3.2.0

```
[hadoop@hadoop-clone-46 ~]$ ls
1 bin boot data dev etc home lib lib64 lost+found media mnt opt proc root run sbin srv sys tmp usr var
[hadoop@hadoop-clone-46 ~]$ cd /home
[hadoop@hadoop-clone-46 ~]$ ls
hadoop user1
[hadoop@hadoop-clone-46 ~]$ cd user1
-bash: cd: user1: Permission denied
[hadoop@hadoop-clone-46 ~]$ jps
22449 Jps
[hadoop@hadoop-clone-46 ~]$ jps
25811 Jps
24359 DataNode
[hadoop@hadoop-clone-46 ~]$ exit
logout
Connection to 192.168.32.46 closed.
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:~/Desktop/lab7$
```

❖ Application of Hadoop cluster: (Sports)

- ✓ On such scenario where cluster can be used is sports.
- ✓ When any league auction comes, then each team have to decide which player they can have in their team or which player will perform best for their team.
- ✓ When there is little performance data available to train our model on, we can have advantage of clustering. In such case, we can find the similar players using some characteristics.