

BDA-LAB02

-Jainil Trivedi (CE166)

Aim: Configuring and experiencing Hadoop Single node cluster.

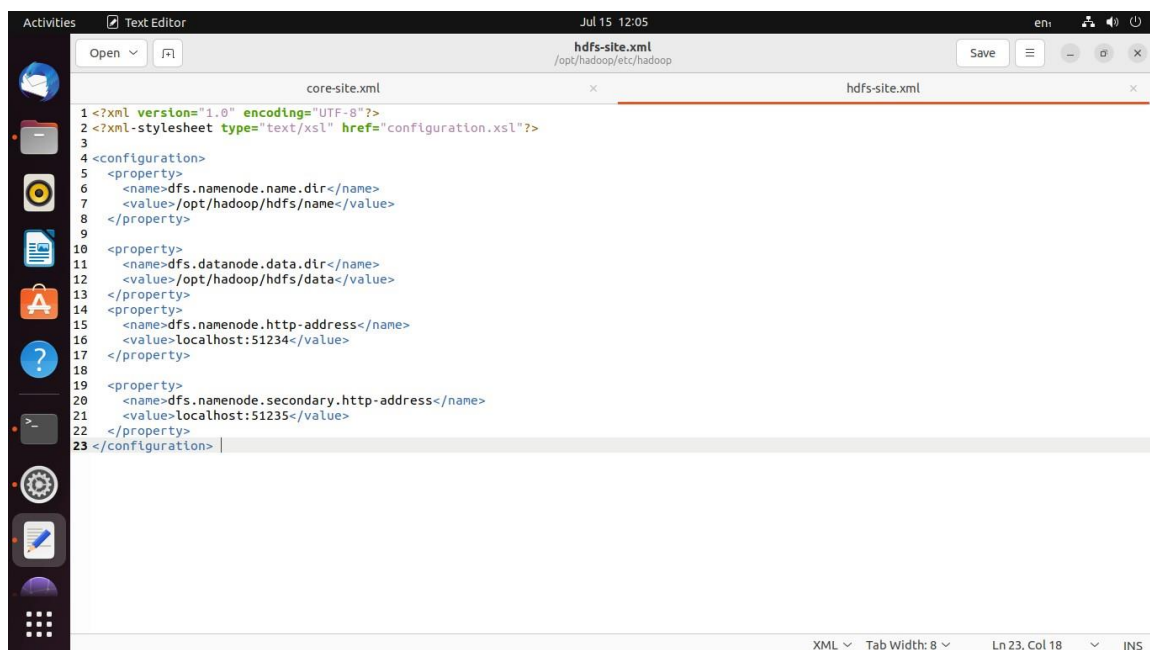
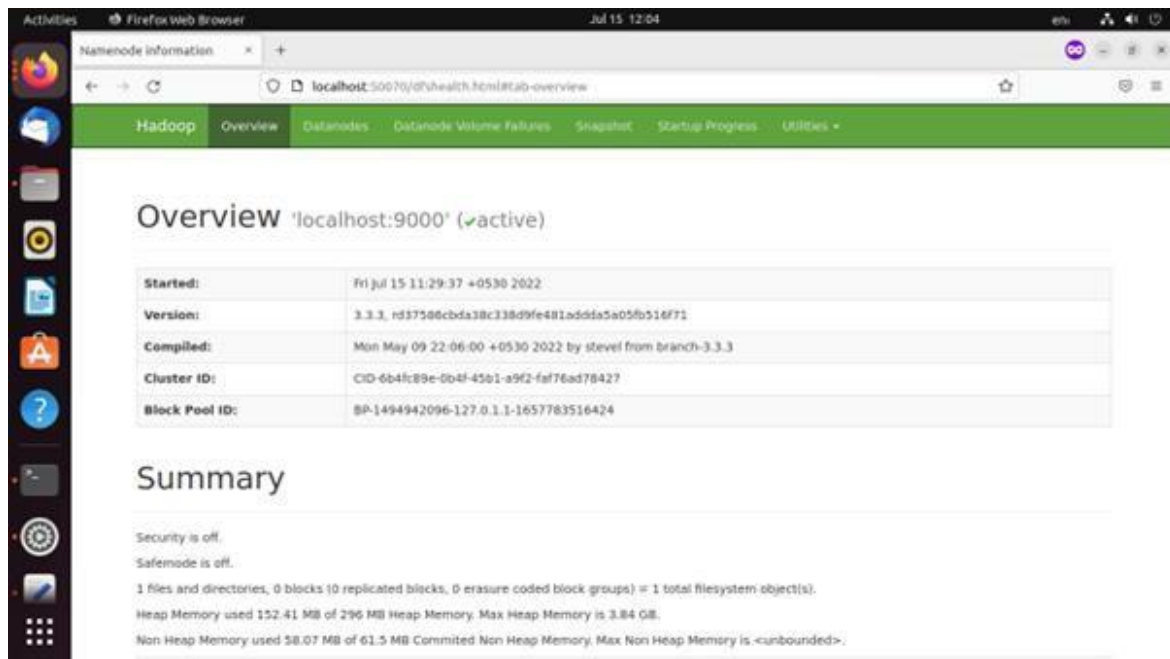
Start/stop script and monitoring.

Tasks

1. Locate the log directory and refer for troubleshooting.

```
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:/opt/hadoop$ ls
bin  hdf5  lib  LICENSE-binary  LICENSE.txt  NOTICE-binary  README.txt  share
etc  include  libexec  licenses-binary  log  NOTICE.txt  sbin  tmp
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:/opt/hadoop$ cd logs
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:/opt/hadoop/log$ ls
hadoop-hadoop-datanode-celab2-ThinkCentre-neo-50s-Gen-3.log  hadoop-hadoop-namenode-celab2-ThinkCentre-neo-50s-Gen-3.out.1
hadoop-hadoop-datanode-celab2-ThinkCentre-neo-50s-Gen-3.out  hadoop-hadoop-secondarynamenode-celab2-ThinkCentre-neo-50s-Gen-3.log
hadoop-hadoop-datanode-celab2-ThinkCentre-neo-50s-Gen-3.out.1  hadoop-hadoop-secondarynamenode-celab2-ThinkCentre-neo-50s-Gen-3.out
hadoop-hadoop-namenode-celab2-ThinkCentre-neo-50s-Gen-3.log  hadoop-hadoop-secondarynamenode-celab2-ThinkCentre-neo-50s-Gen-3.out.1
hadoop-hadoop-namenode-celab2-ThinkCentre-neo-50s-Gen-3.out  SecurityAuth-hadoop.audit
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:/opt/hadoop/log$ ls -l
total 236
-rw-rw-r-- 1 hadoop hadoop 64587 Jul 15 11:29 hadoop-hadoop-datanode-celab2-ThinkCentre-neo-50s-Gen-3.log
-rw-rw-r-- 1 hadoop hadoop 819 Jul 15 11:29 hadoop-hadoop-datanode-celab2-ThinkCentre-neo-50s-Gen-3.out
-rw-rw-r-- 1 hadoop hadoop 819 Jul 14 13:00 hadoop-hadoop-datanode-celab2-ThinkCentre-neo-50s-Gen-3.out.1
-rw-rw-r-- 1 hadoop hadoop 82222 Jul 15 11:49 hadoop-hadoop-namenode-celab2-ThinkCentre-neo-50s-Gen-3.log
-rw-rw-r-- 1 hadoop hadoop 819 Jul 15 11:29 hadoop-hadoop-namenode-celab2-ThinkCentre-neo-50s-Gen-3.out
-rw-rw-r-- 1 hadoop hadoop 819 Jul 14 13:00 hadoop-hadoop-namenode-celab2-ThinkCentre-neo-50s-Gen-3.out.1
-rw-rw-r-- 1 hadoop hadoop 65412 Jul 15 11:49 hadoop-hadoop-secondarynamenode-celab2-ThinkCentre-neo-50s-Gen-3.log
-rw-rw-r-- 1 hadoop hadoop 819 Jul 15 11:29 hadoop-hadoop-secondarynamenode-celab2-ThinkCentre-neo-50s-Gen-3.out
-rw-rw-r-- 1 hadoop hadoop 819 Jul 14 13:00 hadoop-hadoop-secondarynamenode-celab2-ThinkCentre-neo-50s-Gen-3.out.1
-rw-rw-r-- 1 hadoop hadoop 0 Jul 14 12:55 SecurityAuth-hadoop.audit
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:/opt/hadoop/log$ cat hadoop-hadoop-datanode-celab2-ThinkCentre-neo-50s-Gen-3.log
cat: hadoop-hadoop-datanode-celab2-ThinkCentre-neo-50s-Gen-3.log: No such file or directory
hadoop@celab2-ThinkCentre-neo-50s-Gen-3:/opt/hadoop/log$ cat hadoop-hadoop-datanode-celab2-ThinkCentre-neo-50s-Gen-3.log
2022-07-14 13:00:41,640 INFO org.apache.hadoop.hdfs.server.datanode.DataNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting DataNode
STARTUP_MSG: host = celab2-ThinkCentre-neo-50s-Gen-3/127.0.1.1
STARTUP_MSG: args = []
STARTUP_MSG: version = 3.3.3
STARTUP_MSG: classpath = /opt/hadoop-3.3.3/etc/hadoop:/opt/hadoop-3.3.3/share/hadoop/common/lib/httpcore-4.4.13.jar:/opt/hadoop-3.3.3/share/
hadoop/common/lib/javawebsocket-api-3.1.0.jar:/opt/hadoop-3.3.3/share/hadoop/common/lib/jetty-io-9.4.43.v20210629.jar:/opt/hadoop-3.3.3/share/
```

2. Change the URL of HDFS Web UI to localhost:51234 instead of localhost:50070 and then access the web UI using the new URL.



Activities Firefox Web Browser Jul 15 12:07 en

Problem loading page x Namenode information x

localhost:51234/dfshealth.html#tab-overview

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Overview 'localhost:9000' (✓active)

Started:	Fri Jul 15 12:06:56 +0530 2022
Version:	3.3.3, rd37586cbda38c338d9fe481addda5a05fb516f71
Compiled:	Mon May 09 22:06:00 +0530 2022 by stevel from branch-3.3.3
Cluster ID:	CID-6b4fc89e-0b4f-45b1-a9f2-faf76ad78427
Block Pool ID:	BP-1494942096-127.0.1.1-1657783516424

Summary

Security is off.
Safemode is off.

1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

Heap Memory used 145.07 MB of 246 MB Heap Memory. Max Heap Memory is 3.84 GB.

Non Heap Memory used 57.12 MB of 60.71 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

3. Undo changes done in task2.

Activities Text Editor Jul 15 12:08 en

Open hdfs-site.xml Save

core-site.xml hdfs-site.xml

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
3
4 <configuration>
5   <property>
6     <name>dfs.namenode.name.dir</name>
7     <value>/opt/hadoop/hdfs/name</value>
8   </property>
9
10  <property>
11    <name>dfs.datanode.data.dir</name>
12    <value>/opt/hadoop/hdfs/data</value>
13  </property>
14  <property>
15    <name>dfs.namenode.http-address</name>
16    <value>localhost:50070</value>
17  </property>
18
19  <property>
20    <name>dfs.namenode.secondary.http-address</name>
21    <value>localhost:50090</value>
22  </property>
23 </configuration>
```

XML Tab Width: 8 Ln 16, Col 27 INS

The screenshot shows the Hadoop NameNode Overview page in a Firefox browser. The browser's address bar shows the URL `localhost:50070/dfshealth.html#tab-overview`. The page has a green header with tabs for 'Hadoop', 'Overview', 'Datanodes', 'Datanode Volume Failures', 'Snapshot', 'Startup Progress', and 'Utilities'. The 'Overview' tab is selected, showing the title 'Overview 'localhost:9000' (✓active)'. Below the title is a table with the following information:

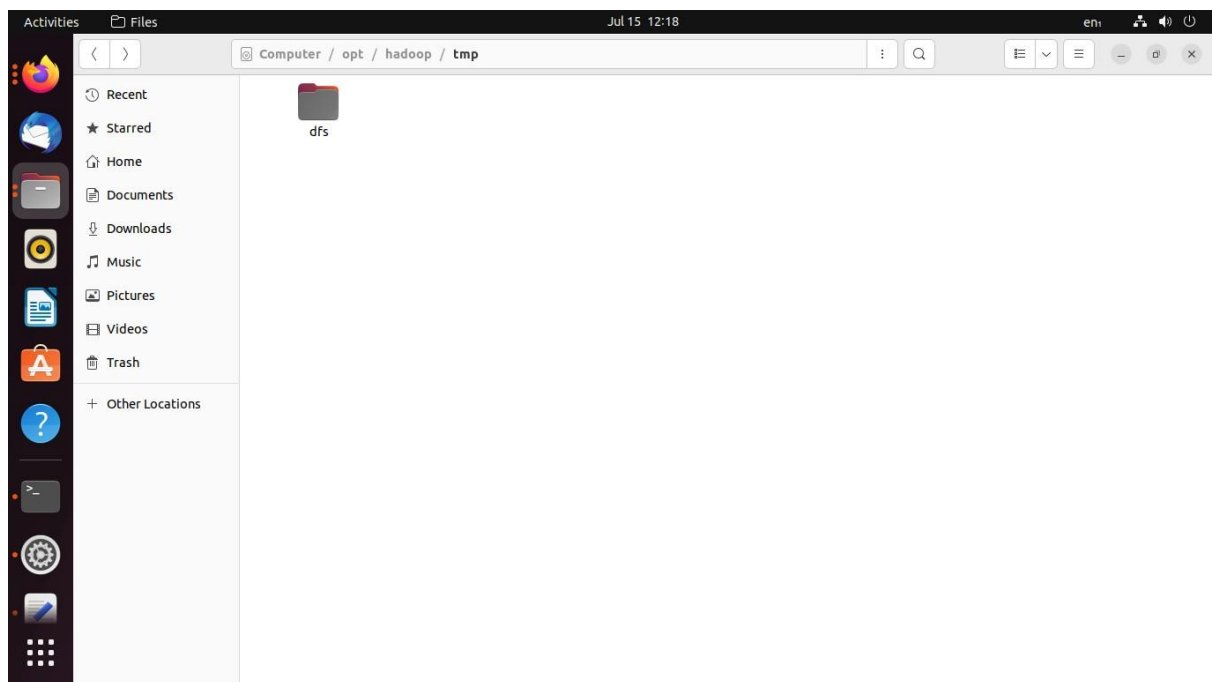
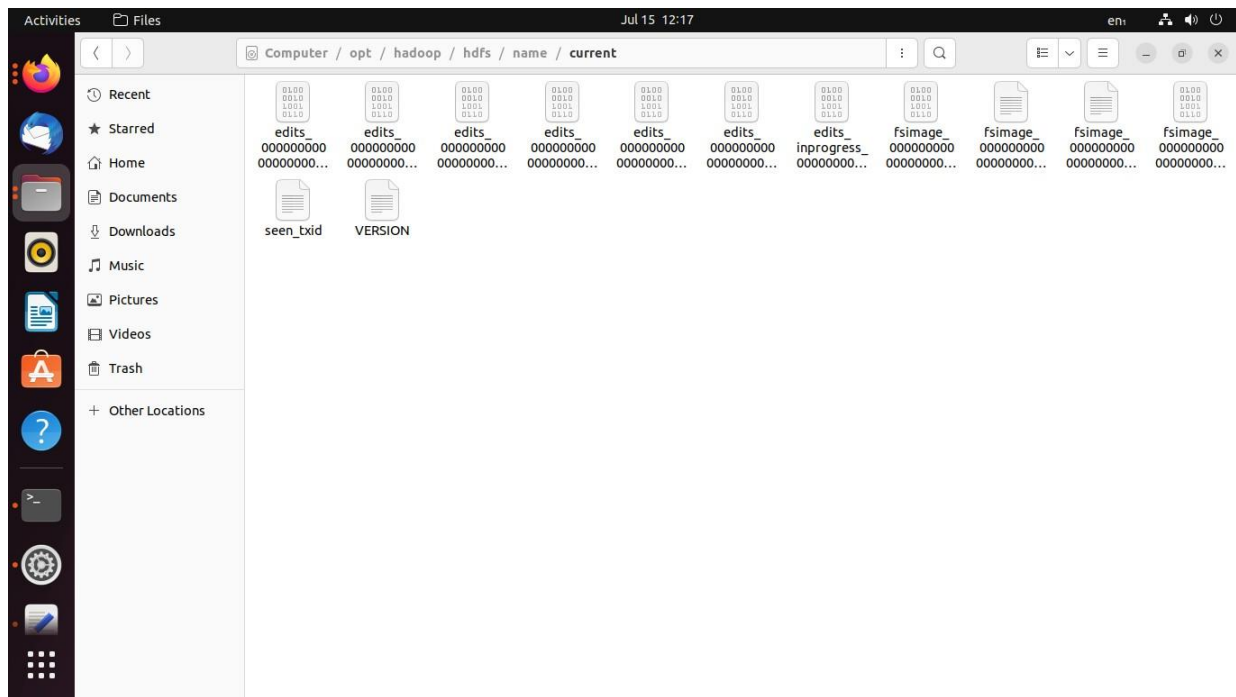
Started:	Fri Jul 15 12:09:28 +0530 2022
Version:	3.3.3, rd37586cbda38c338d9fe481addda5a05fb516f71
Compiled:	Mon May 09 22:06:00 +0530 2022 by stebel from branch-3.3.3
Cluster ID:	CID-6b4fc89e-0b4f-45b1-a9f2-faf76ad78427
Block Pool ID:	BP-1494942096-127.0.1.1-1657783516424

Below the table is a 'Summary' section with the following text:

Security is off.
Safemode is off.
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
Heap Memory used 144.22 MB of 246 MB Heap Memory. Max Heap Memory is 3.84 GB.
Non Heap Memory used 56.97 MB of 60.4 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

- Find out what is current location of namenode and datanode storage directories. Normally whenever cleanup is required, tmp folder is one to be emptied up (not to confuse with formatting file system). Hence, it is advised data and metadata better be outside tmp. Do the needful to have namenode and datanode directories in /opt/hadoop, if present within tmp.

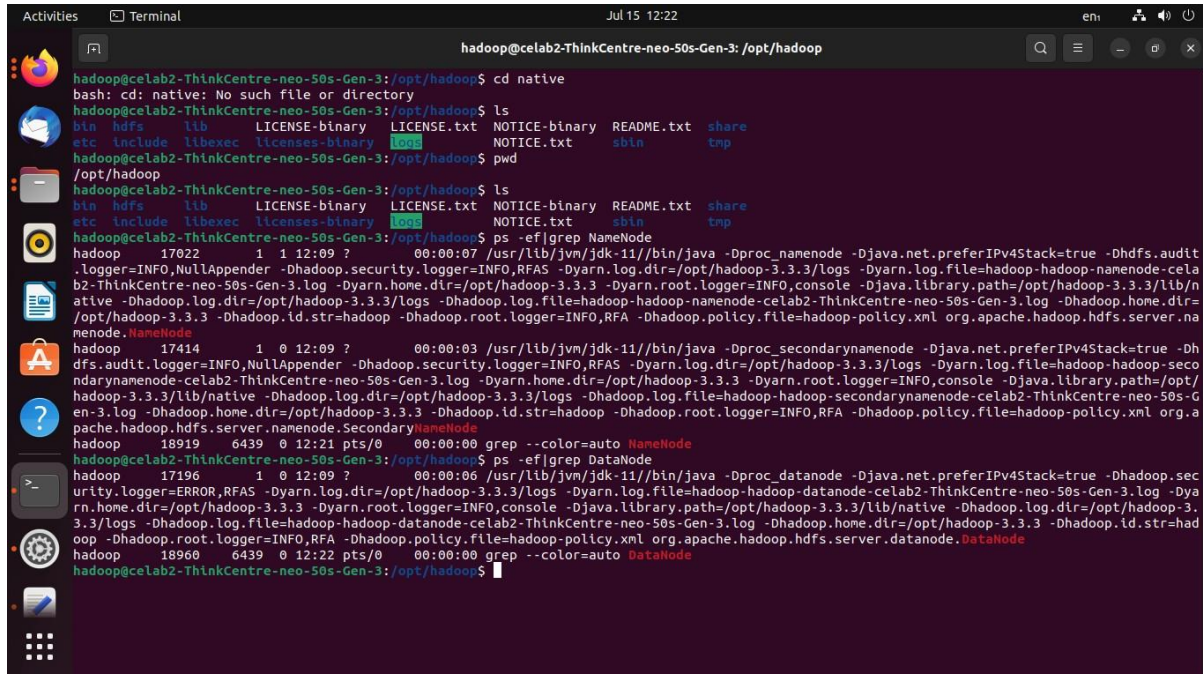
The screenshot shows a file manager window with the path `Computer / opt / hadoop / hdfs`. The window displays two folders: 'data' and 'name'. The left sidebar shows a list of locations: Recent, Starred, Home, Documents, Downloads, Music, Pictures, Videos, Trash, and Other Locations.



5. Analyze the output of following command:

```
[hadoop@hadoop-clone hadoop]$ ps -ef |grep NameNode
```

```
[hadoop@hadoop-clone hadoop]$ ps -ef |grep DataNode
```



```
hadoop@celab2-ThinkCentre-neo-50s-Gen-3: /opt/hadoop$ cd native
bash: cd: native: No such file or directory
hadoop@celab2-ThinkCentre-neo-50s-Gen-3: /opt/hadoop$ ls
bin  hdfs  lib  LICENSE-binary  LICENSE.txt  NOTICE-binary  README.txt  share
etc  include  libexec  licenses-binary  logs  NOTICE.txt  sbin  tmp
hadoop@celab2-ThinkCentre-neo-50s-Gen-3: /opt/hadoop$ pwd
/opt/hadoop
hadoop@celab2-ThinkCentre-neo-50s-Gen-3: /opt/hadoop$ ls
bin  hdfs  lib  LICENSE-binary  LICENSE.txt  NOTICE-binary  README.txt  share
etc  include  libexec  licenses-binary  logs  NOTICE.txt  sbin  tmp
hadoop@celab2-ThinkCentre-neo-50s-Gen-3: /opt/hadoop$ ps -ef|grep NameNode
hadoop 17022 1 12:09 ? 00:00:07 /usr/lib/jvm/jdk-11/bin/java -Dproc_namenode -Djava.net.preferIPv4Stack=true -Dhdfs.audit
.logger=INFO,NullAppender -Dhadoop.security.logger=INFO,RFAS -Dyarn.log.dir=/opt/hadoop-3.3.3/logs -Dyarn.log.file=hadoop-namenode-cela
b2-ThinkCentre-neo-50s-Gen-3.log -Dyarn.home.dir=/opt/hadoop-3.3.3 -Dyarn.root.logger=INFO,console -Djava.library.path=/opt/hadoop-3.3.3/lib/n
ative -Dhadoop.log.dir=/opt/hadoop-3.3.3/logs -Dhadoop.log.file=hadoop-hadoop-namenode-celab2-ThinkCentre-neo-50s-Gen-3.log -Dhadoop.home.dir=
/opt/hadoop-3.3.3 -Dhadoop.id.str=hadoop -Dhadoop.root.logger=INFO,RFA -Dhadoop.policy.file=hadoop-policy.xml org.apache.hadoop.hdfs.server.na
menode.NameNode
hadoop 17414 1 0 12:09 ? 00:00:03 /usr/lib/jvm/jdk-11/bin/java -Dproc_secondarynamenode -Djava.net.preferIPv4Stack=true -Dh
dfs.audit.logger=INFO,NullAppender -Dhadoop.security.logger=INFO,RFAS -Dyarn.log.dir=/opt/hadoop-3.3.3/logs -Dyarn.log.file=hadoop-hadoop-seco
ndarynamenode-celab2-ThinkCentre-neo-50s-Gen-3.log -Dyarn.home.dir=/opt/hadoop-3.3.3 -Dyarn.root.logger=INFO,console -Djava.library.path=/opt/
hadoop-3.3.3/lib/native -Dhadoop.log.dir=/opt/hadoop-3.3.3/logs -Dhadoop.log.file=hadoop-hadoop-secondarynamenode-celab2-ThinkCentre-neo-50s-G
en-3.log -Dhadoop.home.dir=/opt/hadoop-3.3.3 -Dhadoop.id.str=hadoop -Dhadoop.root.logger=INFO,RFA -Dhadoop.policy.file=hadoop-policy.xml org.a
pache.hadoop.hdfs.server.namenode.SecondaryNameNode
hadoop 18919 6439 0 12:21 pts/0 00:00:00 grep --color=auto NameNode
hadoop@celab2-ThinkCentre-neo-50s-Gen-3: /opt/hadoop$ ps -ef|grep DataNode
hadoop 17196 1 0 12:09 ? 00:00:06 /usr/lib/jvm/jdk-11/bin/java -Dproc_datanode -Djava.net.preferIPv4Stack=true -Dhadoop.sec
urity.logger=ERROR,RFAS -Dyarn.log.dir=/opt/hadoop-3.3.3/logs -Dyarn.log.file=hadoop-hadoop-datanode-celab2-ThinkCentre-neo-50s-Gen-3.log -Dya
rn.home.dir=/opt/hadoop-3.3.3 -Dyarn.root.logger=INFO,console -Djava.library.path=/opt/hadoop-3.3.3/lib/native -Dhadoop.log.dir=/opt/hadoop-3.
3.3/logs -Dhadoop.log.file=hadoop-hadoop-datanode-celab2-ThinkCentre-neo-50s-Gen-3.log -Dhadoop.home.dir=/opt/hadoop-3.3.3 -Dhadoop.id.str=had
oop -Dhadoop.root.logger=INFO,RFA -Dhadoop.policy.file=hadoop-policy.xml org.apache.hadoop.hdfs.server.datanode.DataNode
hadoop 18960 6439 0 12:22 pts/0 00:00:00 grep --color=auto DataNode
hadoop@celab2-ThinkCentre-neo-50s-Gen-3: /opt/hadoop$
```