LAB – 2

Name: Gandevia Keval Dharmeshbhai

Sem: VII

Roll No: CE046

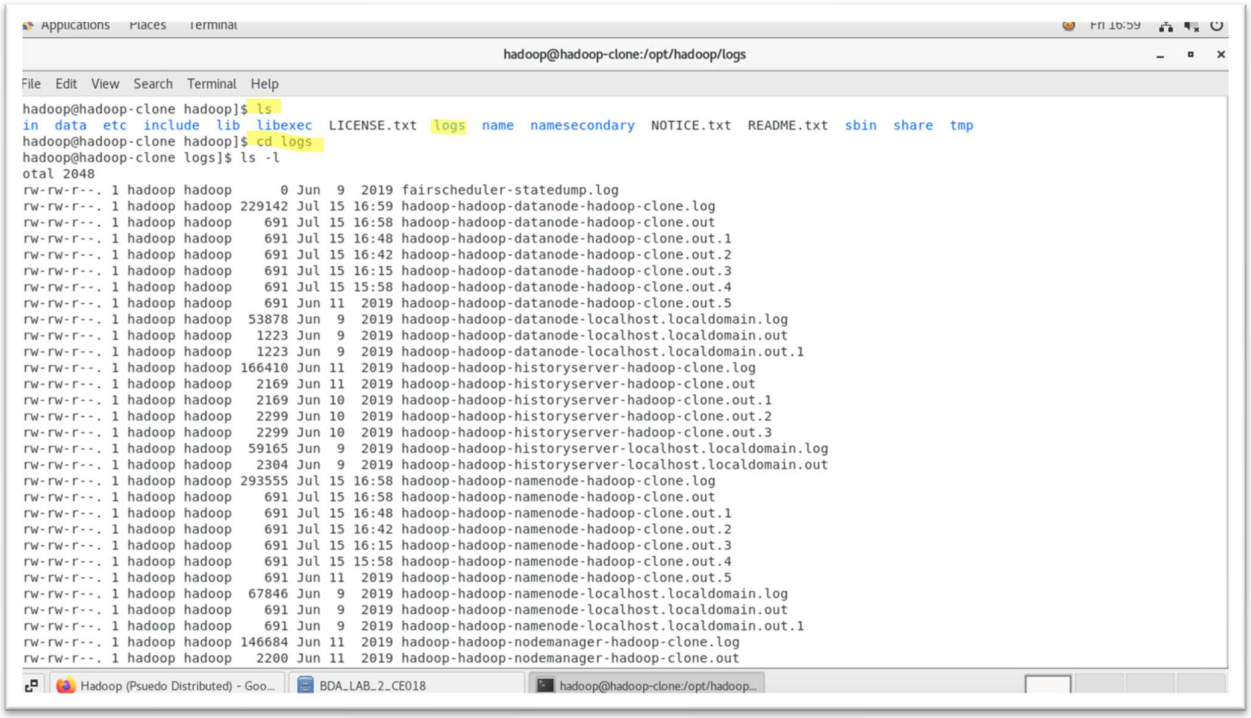
Subject: Big Data and Analytics

Aim: Configuring and experiencing Hadoop single node cluster. Start/stop script and monitoring.

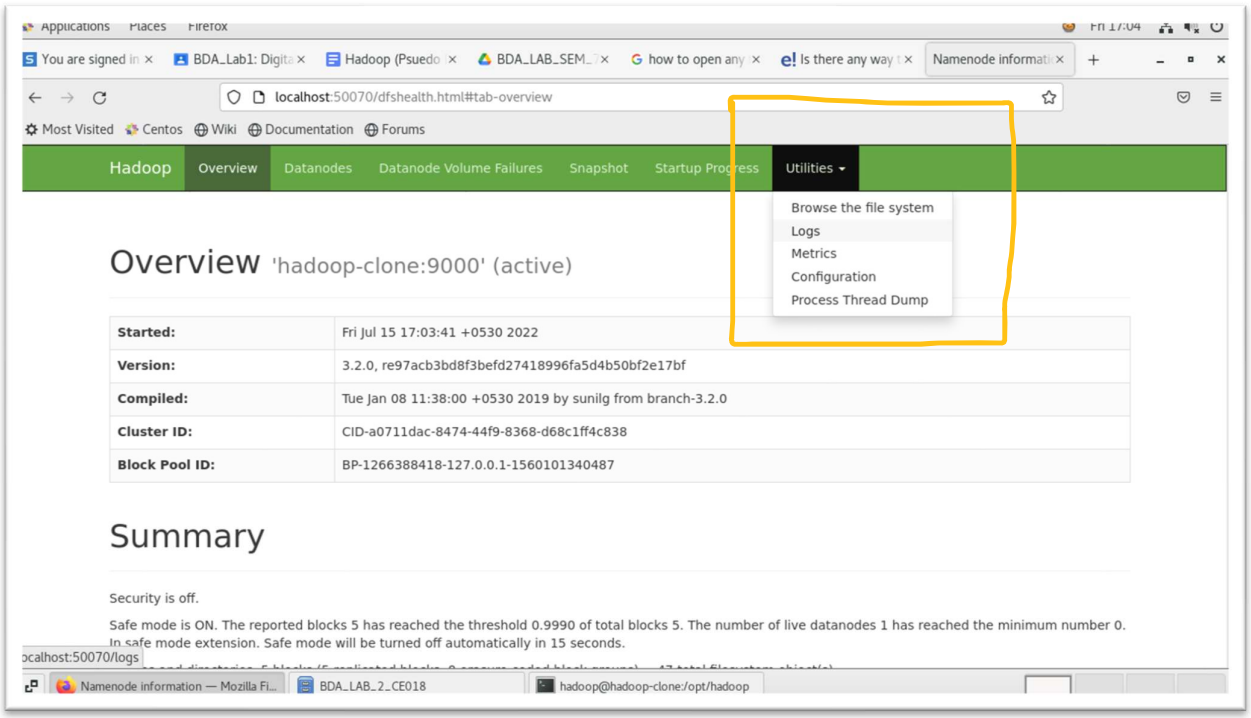
Q. 2: Locate the log directory and refer for troubleshooting.

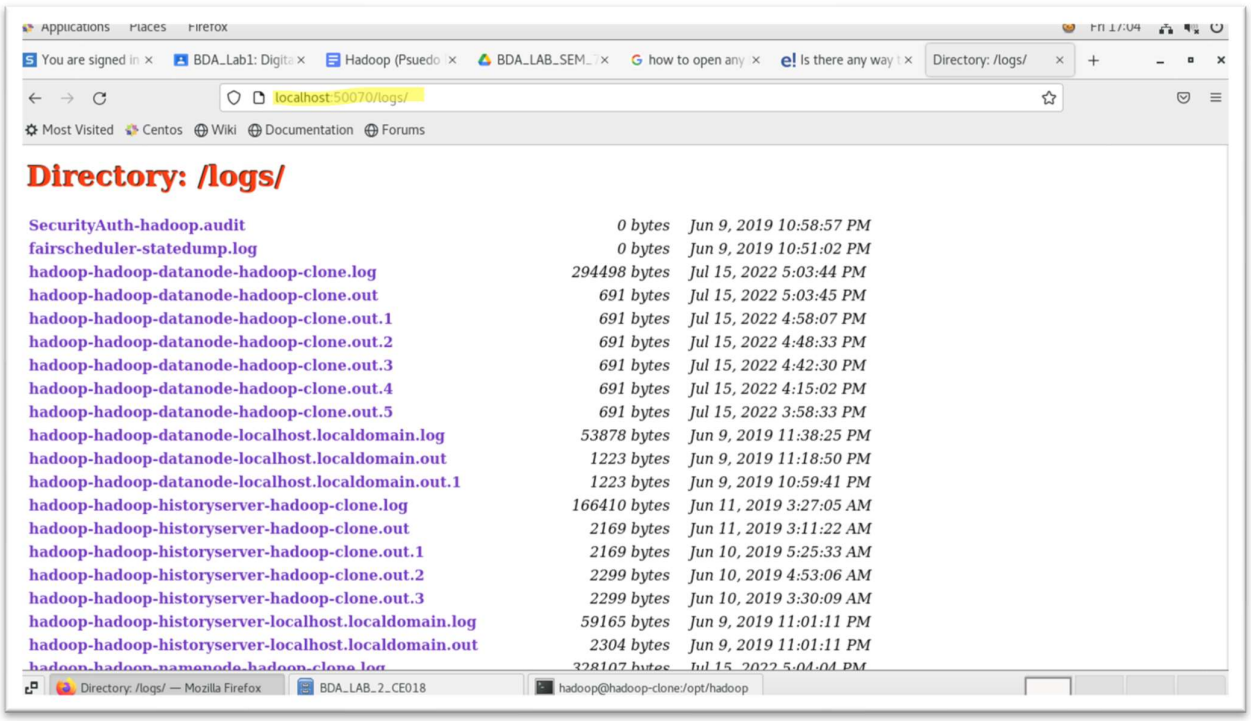
- Hadoop contains concept of logs which helps to keep track of activities that we are doing and also it keeps track of some of the configurations of the Hadoop.
- As we know, Hadoop supports processing of structured data, unstructured data, and semi-structured data, log files are the good examples of an unstructured data and processing through Hadoop will be the best use case of it.
- There are various of ways of access the log files. Some of them are listed below:

I. Through command:



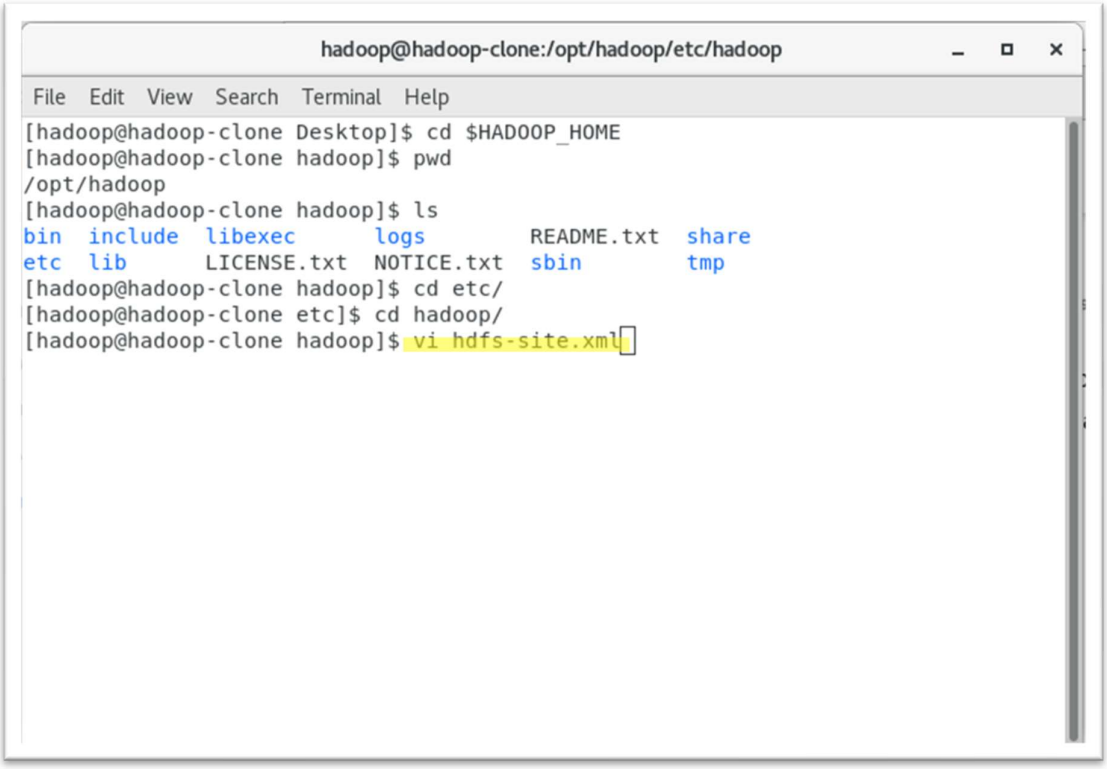
II. Through GUI:

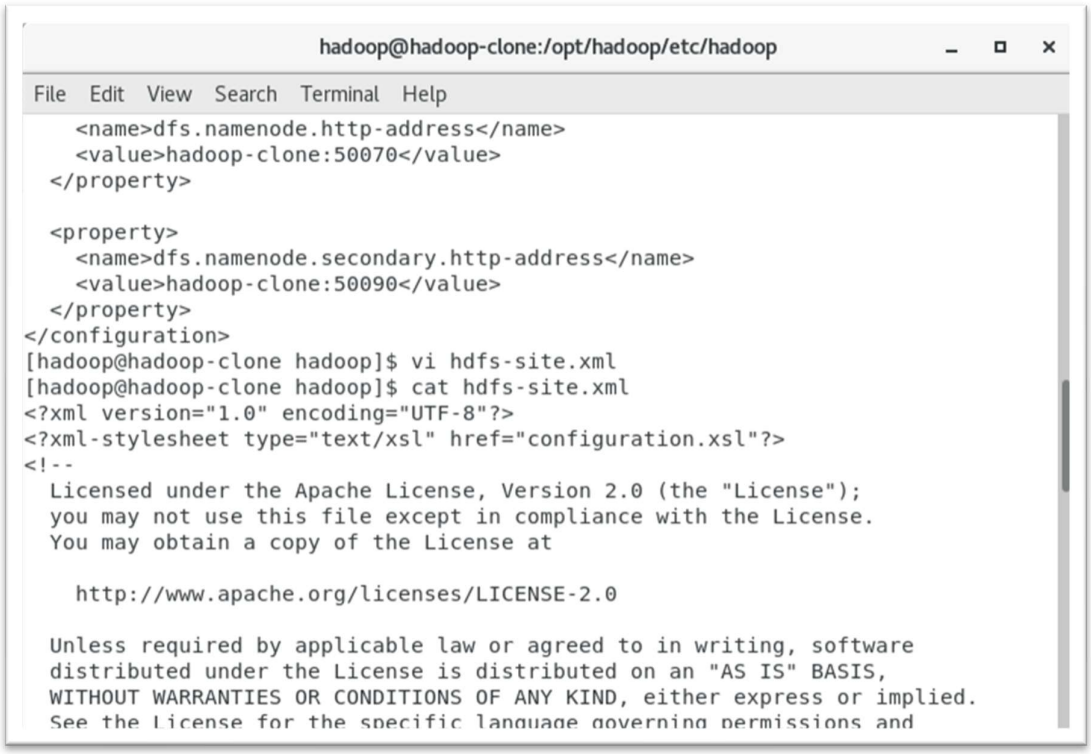


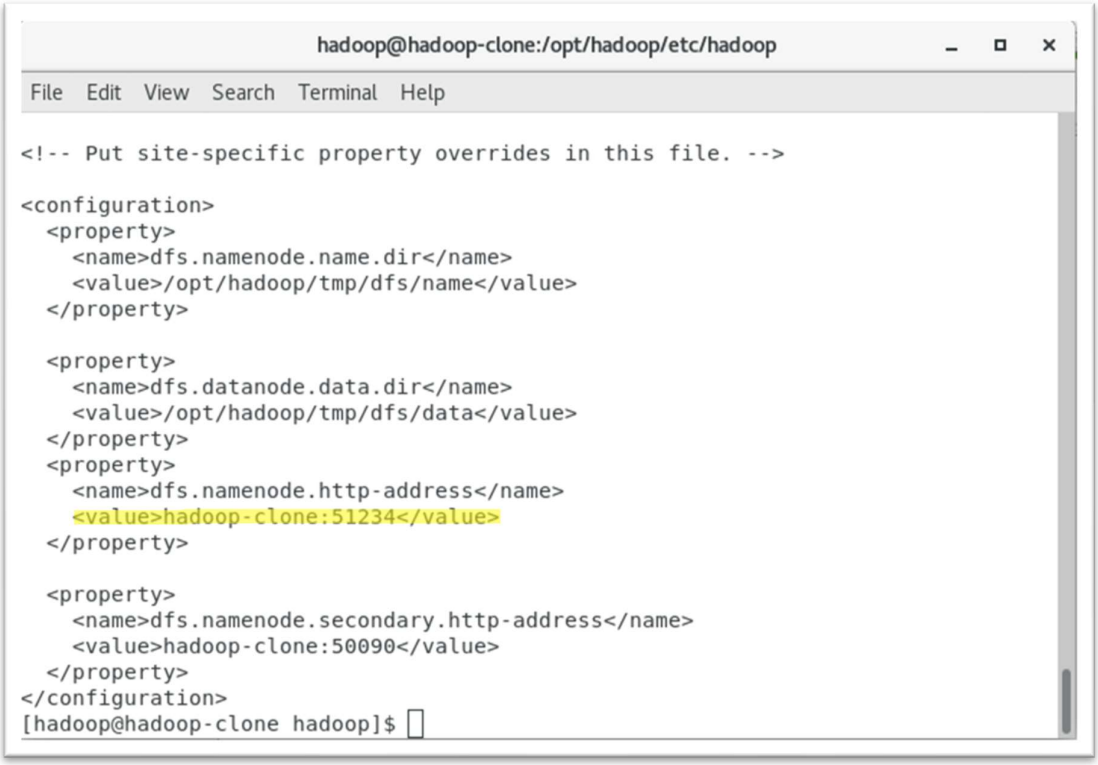


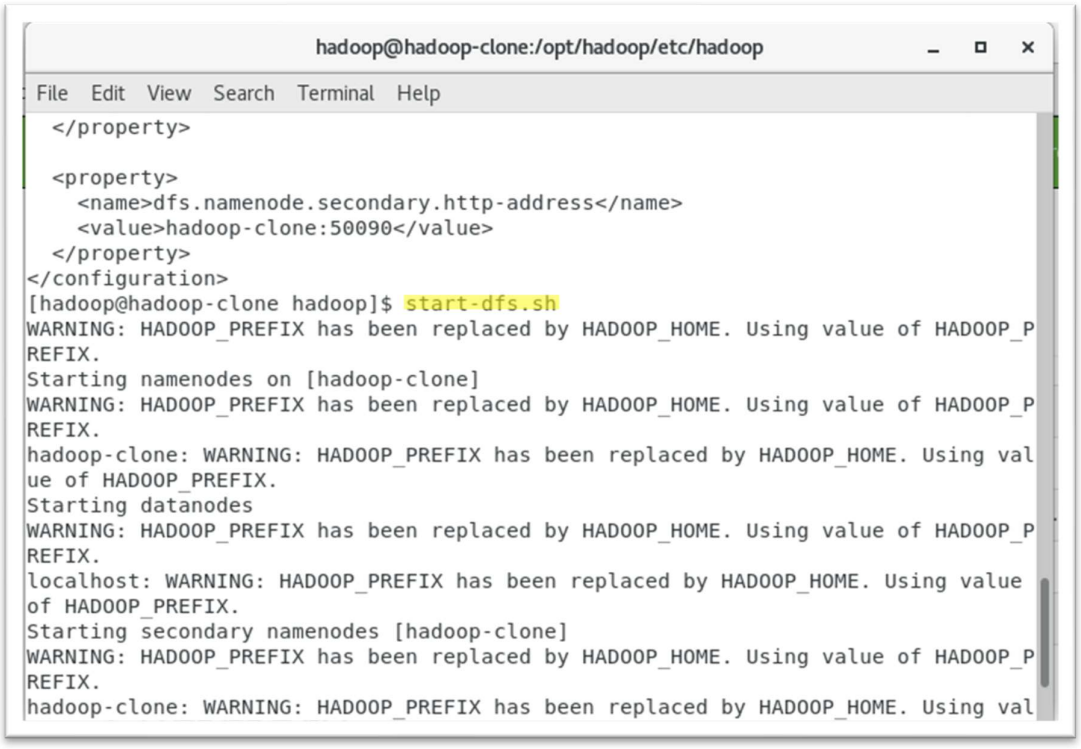
Q. 3: Change the URL of HDFS Web UI to localhost:51234 instead of localhost:50070 and the access the web UI using new URL.

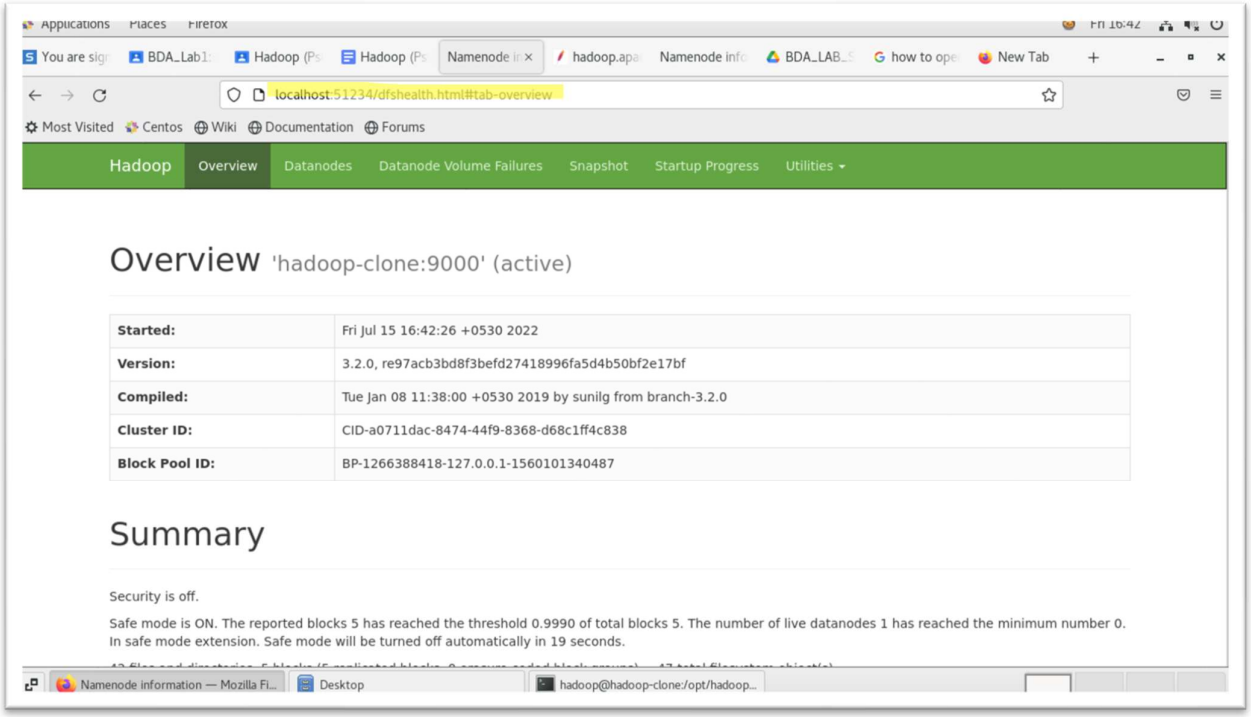
- Hadoop contains many configurations file. From the one such file is hdfs-site.xml which is used to configure the settings for HDFS daemons; the NamdNode, DataNode, and Secondary NameNode.
- In this file we can decide the actual number of replications of file, configure the permissions, block replication.
- We can also configure the port on which GUI part of Hadoop can be accessed.
- Once we do the changes in any configuration files, we need to restart the Hadoop server.


hadoop@hadoop-clone:/opt/hadoop/etc/hadoop
File Edit View Search Terminal Help
[hadoop@hadoop-clone Desktop]$ cd $HADOOP_HOME
[hadoop@hadoop-clone hadoop]$ pwd
/opt/hadoop
[hadoop@hadoop-clone hadoop]$ ls
bin include libexec logs README.txt share
etc lib LICENSE.txt NOTICE.txt sbin tmp
[hadoop@hadoop-clone hadoop]$ cd etc/
[hadoop@hadoop-clone etc]$ cd hadoop/
[hadoop@hadoop-clone hadoop]$ vi hdfs-site.xml


hadoop@hadoop-clone:/opt/hadoop/etc/hadoop
File Edit View Search Terminal Help
<name>dfs.namenode.http-address</name>
<value>hadoop-clone:50070</value>
</property>
<property>
<name>dfs.namenode.secondary.http-address</name>
<value>hadoop-clone:50090</value>
</property>
</configuration>
[hadoop@hadoop-clone hadoop]$ vi hdfs-site.xml
[hadoop@hadoop-clone hadoop]$ cat hdfs-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at
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

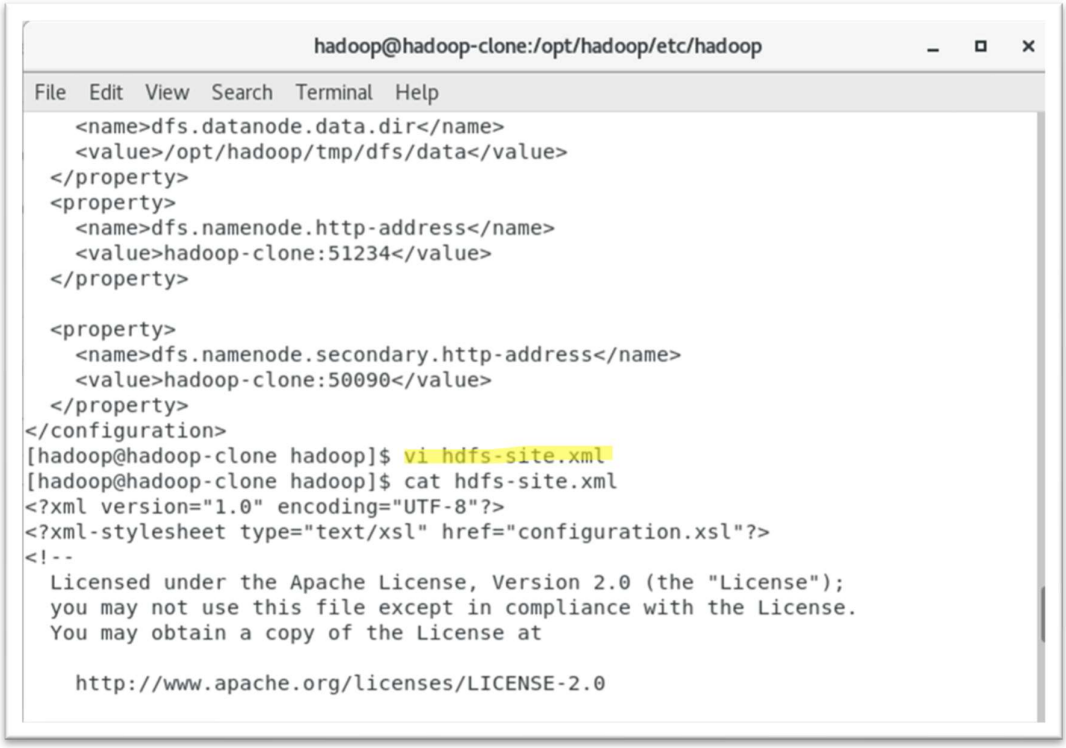

hadoop@hadoop-clone:/opt/hadoop/etc/hadoop
File Edit View Search Terminal Help
<!-- Put site-specific property overrides in this file. -->
<configuration>
<property>
<name>dfs.namenode.name.dir</name>
<value>/opt/hadoop/tmp/dfs/name</value>
</property>
<property>
<name>dfs.datanode.data.dir</name>
<value>/opt/hadoop/tmp/dfs/data</value>
</property>
<property>
<name>dfs.namenode.http-address</name>
<value>hadoop-clone:51234</value>
</property>
<property>
<name>dfs.namenode.secondary.http-address</name>
<value>hadoop-clone:50090</value>
</property>
</configuration>
[hadoop@hadoop-clone hadoop]$

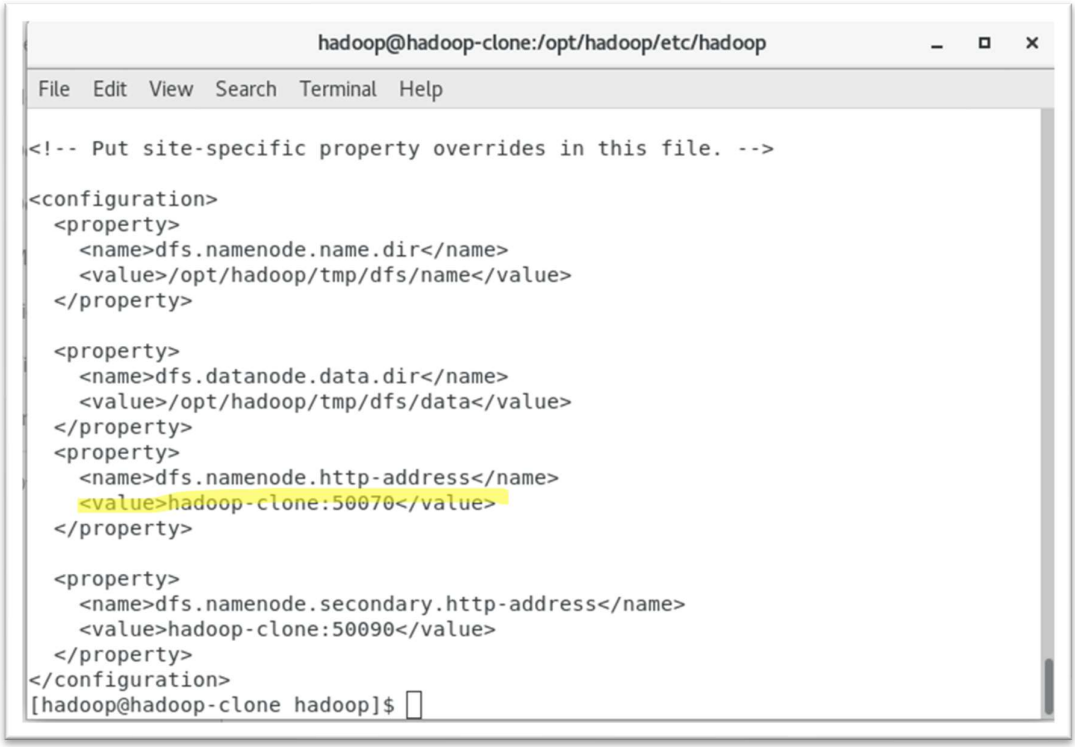

hadoop@hadoop-clone:/opt/hadoop/etc/hadoop
File Edit View Search Terminal Help
</property>
<property>
<name>dfs.namenode.secondary.http-address</name>
<value>hadoop-clone:50090</value>
</property>
</configuration>
[hadoop@hadoop-clone hadoop]$ start-dfs.sh
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Starting namenodes on [hadoop-clone]
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
hadoop-clone: WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Starting datanodes
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
localhost: WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Starting secondary namenodes [hadoop-clone]
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
hadoop-clone: WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using val

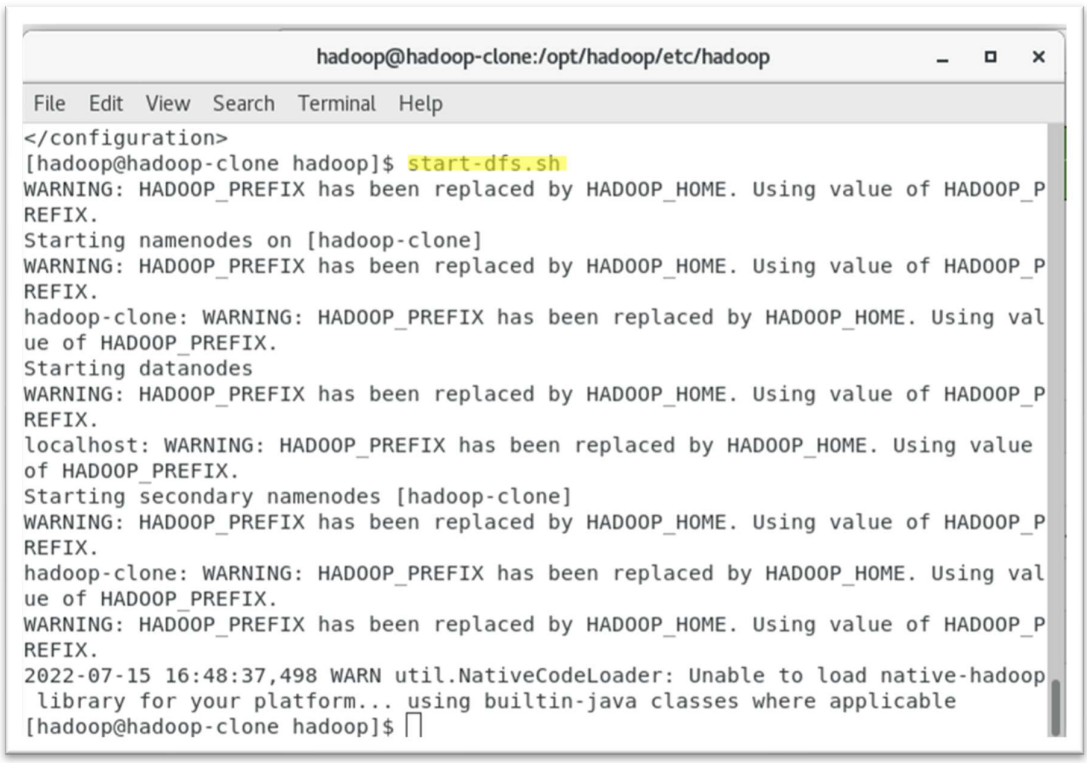


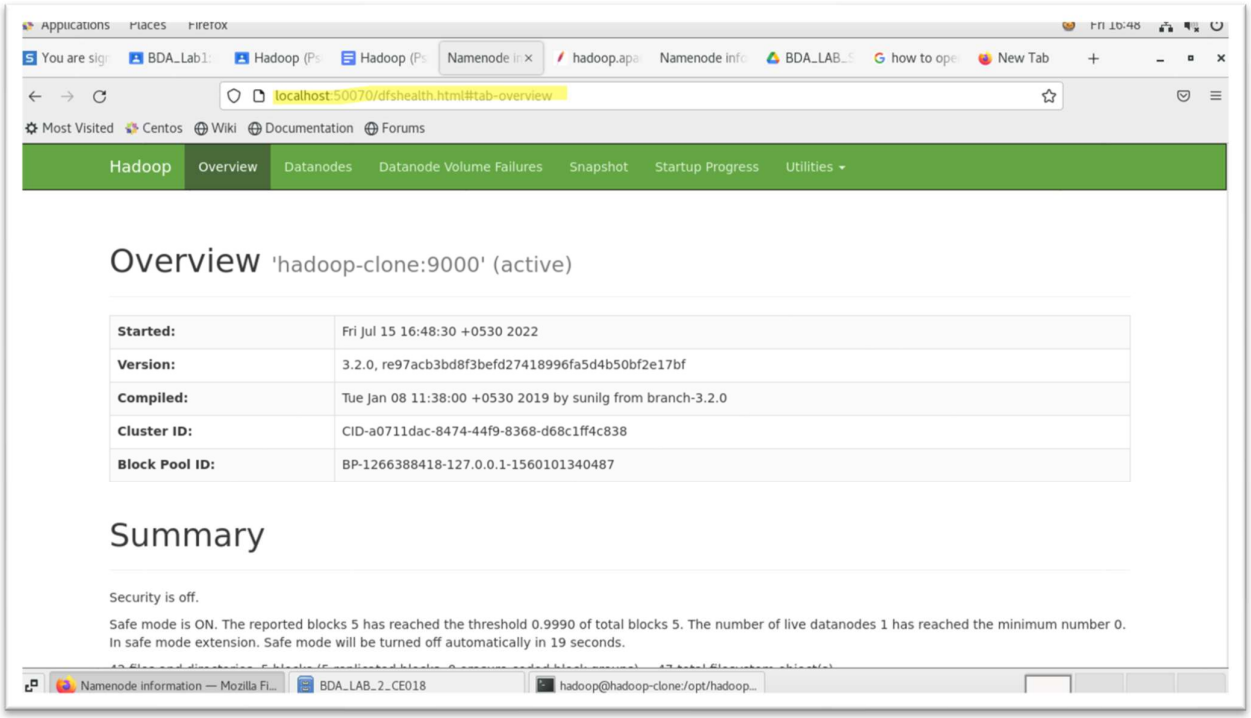
Q. 4: Undo changes done in task3.

- We can also undo the change that we have made in hdfs-site.xml.


hadoop@hadoop-clone:/opt/hadoop/etc/hadoop
File Edit View Search Terminal Help
<name>dfs.datanode.data.dir</name>
<value>/opt/hadoop/tmp/dfs/data</value>
</property>
<property>
<name>dfs.namenode.http-address</name>
<value>hadoop-clone:51234</value>
</property>
<property>
<name>dfs.namenode.secondary.http-address</name>
<value>hadoop-clone:50090</value>
</property>
</configuration>
[hadoop@hadoop-clone hadoop]$ vi hdfs-site.xml
[hadoop@hadoop-clone hadoop]$ cat hdfs-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at
http://www.apache.org/licenses/LICENSE-2.0

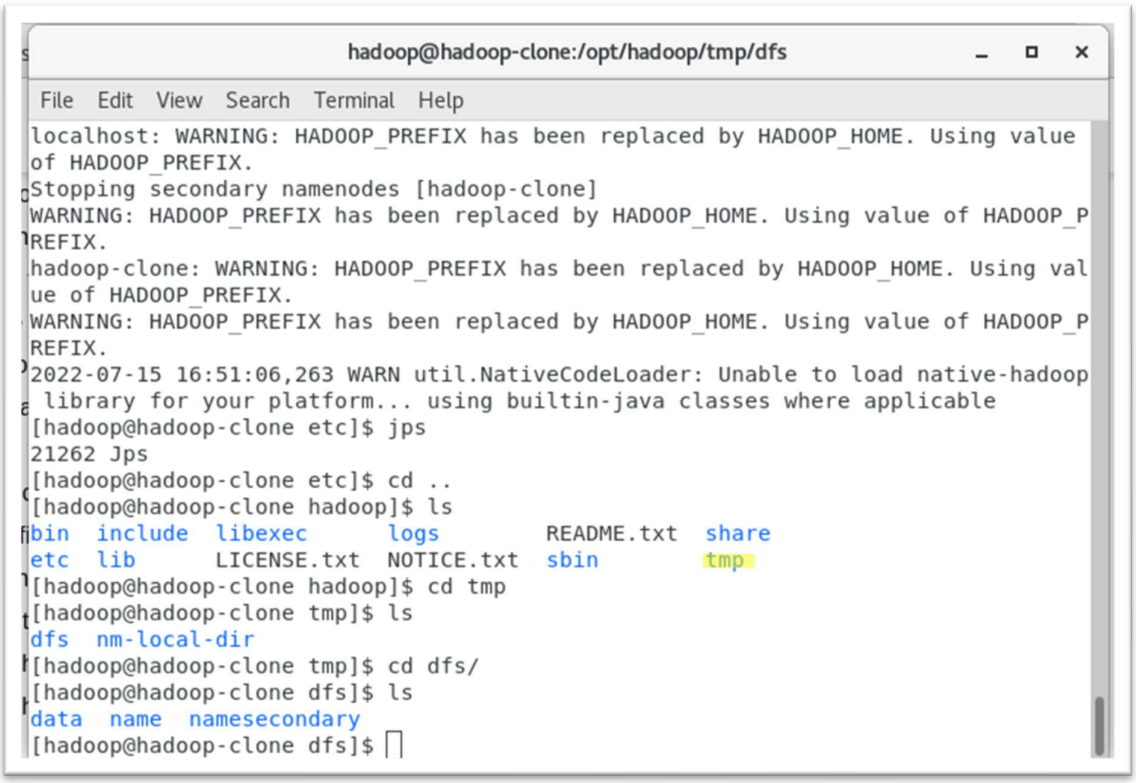

hadoop@hadoop-clone:/opt/hadoop/etc/hadoop
File Edit View Search Terminal Help
<!-- Put site-specific property overrides in this file. -->
<configuration>
<property>
<name>dfs.namenode.name.dir</name>
<value>/opt/hadoop/tmp/dfs/name</value>
</property>
<property>
<name>dfs.datanode.data.dir</name>
<value>/opt/hadoop/tmp/dfs/data</value>
</property>
<property>
<name>dfs.namenode.http-address</name>
<value>hadoop-clone:50070</value>
</property>
<property>
<name>dfs.namenode.secondary.http-address</name>
<value>hadoop-clone:50090</value>
</property>
</configuration>
[hadoop@hadoop-clone hadoop]$

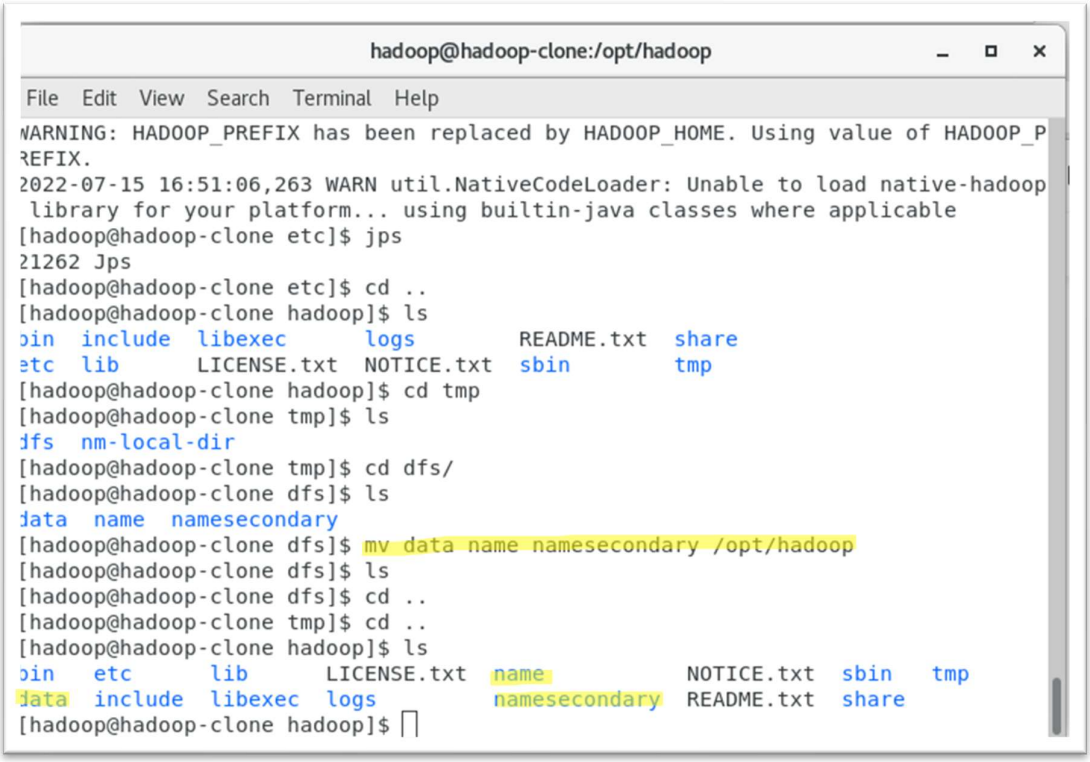

hadoop@hadoop-clone:/opt/hadoop/etc/hadoop
File Edit View Search Terminal Help
</configuration>
[hadoop@hadoop-clone hadoop]$ start-dfs.sh
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Starting namenodes on [hadoop-clone]
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
hadoop-clone: WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Starting datanodes
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
localhost: WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Starting secondary namenodes [hadoop-clone]
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
hadoop-clone: WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2022-07-15 16:48:37,498 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[hadoop@hadoop-clone hadoop]$


Applications Places Firefox
Fri 16:48
localhost:50070/dfshealth.html#tab-overview
Most Visited Centos Wiki Documentation Forums
Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities
Overview 'hadoop-clone:9000' (active)
Started: Fri Jul 15 16:48:30 +0530 2022
Version: 3.2.0, re97acb3bd8f3befd27418996fa5d4b50bf2e17bf
Compiled: Tue Jan 08 11:38:00 +0530 2019 by sunilg from branch-3.2.0
Cluster ID: CID-a0711dac-8474-44f9-8368-d68c1ff4c838
Block Pool ID: BP-1266388418-127.0.0.1-1560101340487
Summary
Security is off.
Safe mode is ON. The reported blocks 5 has reached the threshold 0.9990 of total blocks 5. The number of live datanodes 1 has reached the minimum number 0. In safe mode extension. Safe mode will be turned off automatically in 19 seconds.
Namenode information — Mozilla Fi...
BDA_LAB_2_CE018
hadoop@hadoop-clone:/opt/hadoop...

Q. 5: 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. Hence, it is advised data and metadata better be outside tmp. Do the needful to have namenode and datanode directories in /opt/Hadoop.

- ‘tmp’ directory mainly used for temporary storage during any action being performed. Ex, Intermediate data.

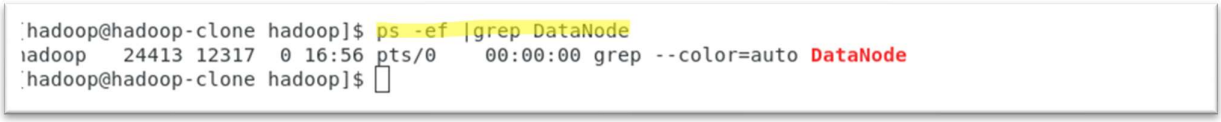




Q. 6: Analyse the output of the following command:

a) [hadoop@hadoop-clone Hadoop] $ ps -ef |grep NameNode

- Generally, 'jps' command is used to check that whether daemons are running or not. But other way is above command. It is used to verify that NameNode is up and running.

b) [hadoop@hadoop-clone Hadoop] $ ps -ef |grep DataNode

- This command used to verify that DataNoe is up and running.