**LAB – 2**

Name: Gandevia Keval Dharmeshbhai

Sem: VII

Roll No: CE046

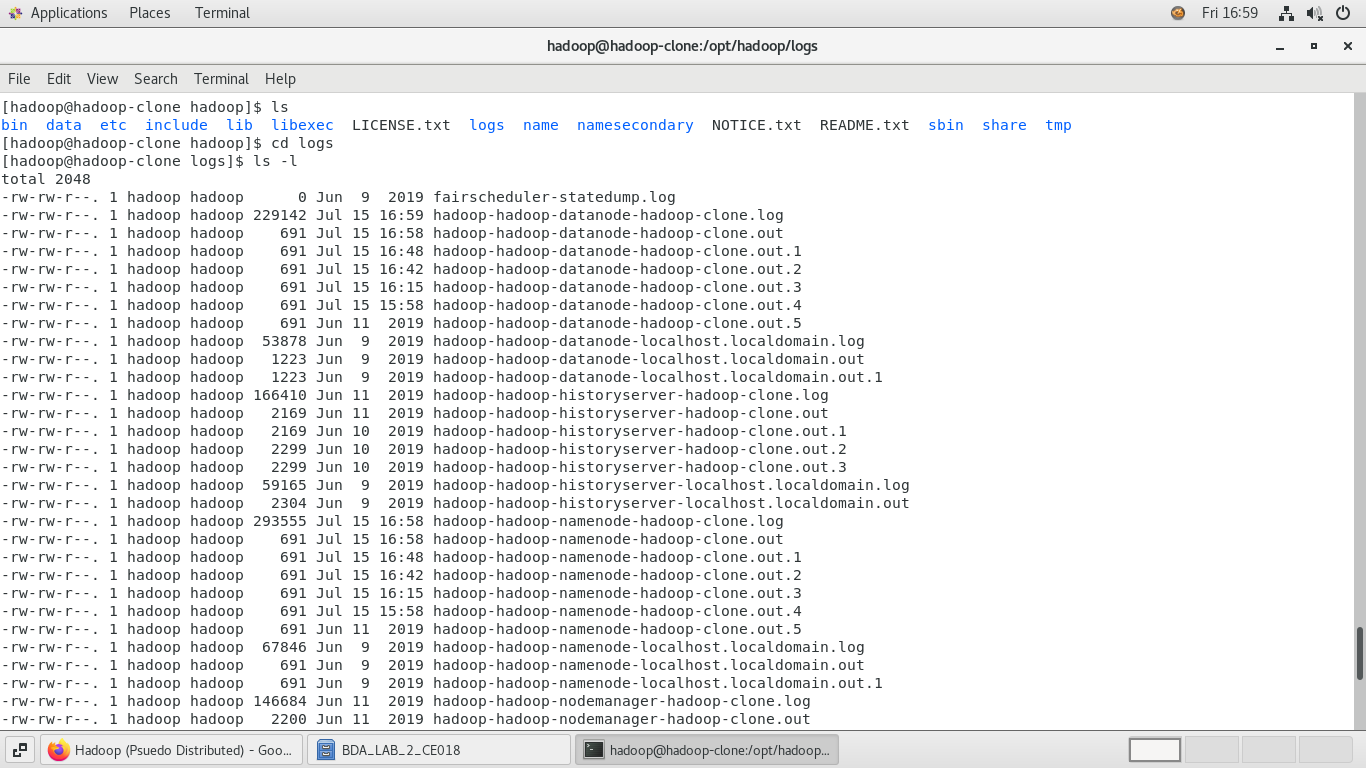
Subject: Big Data and Analytics

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

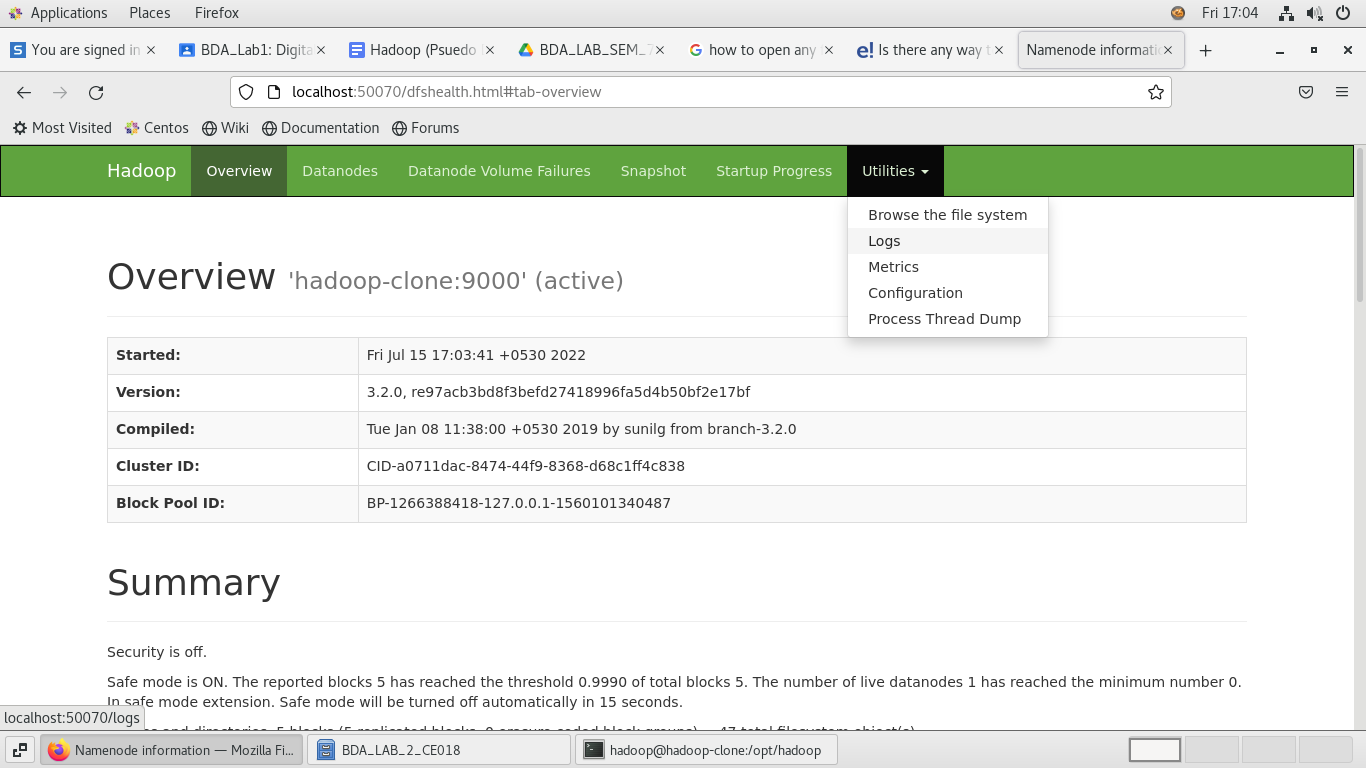
**Q. 2: Locate the log directory and refer for troubleshooting.**

* 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:

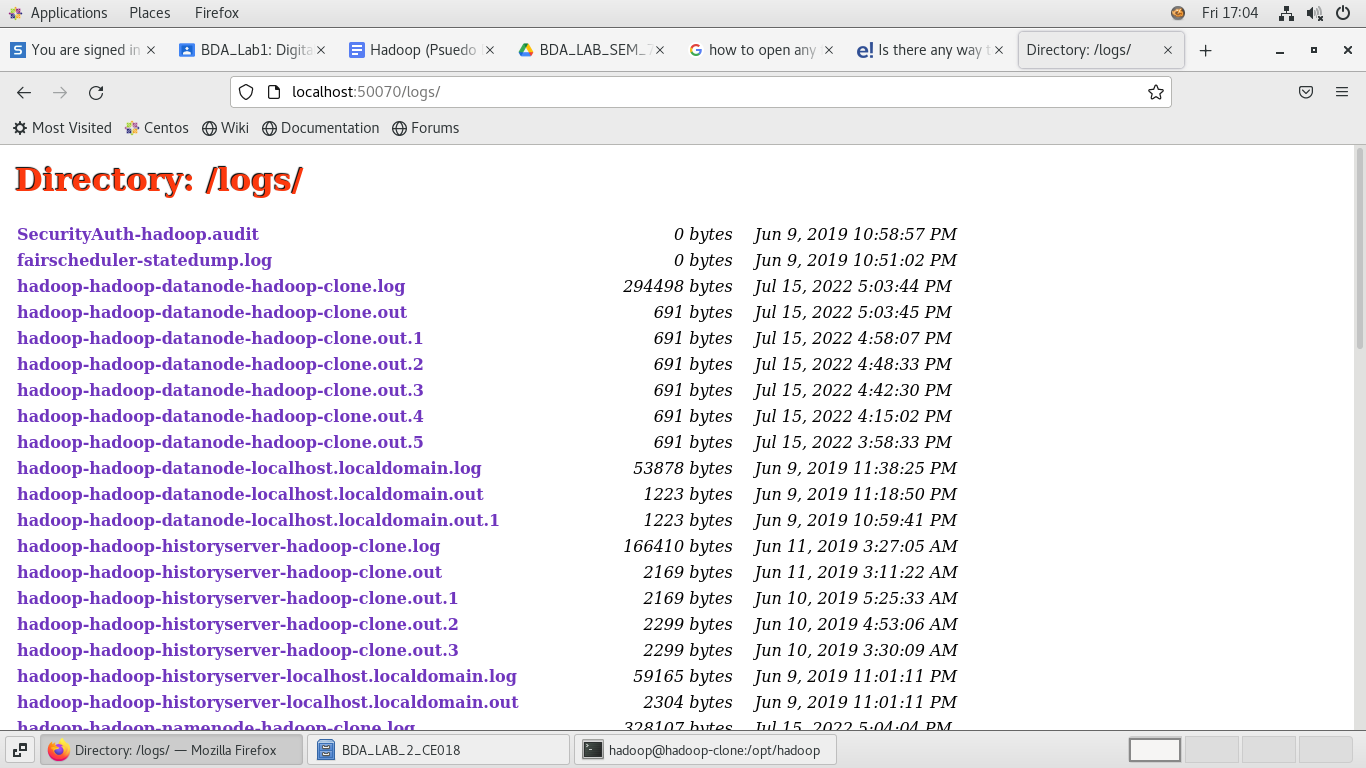
1. **Through command:**





1. **Through GUI:**

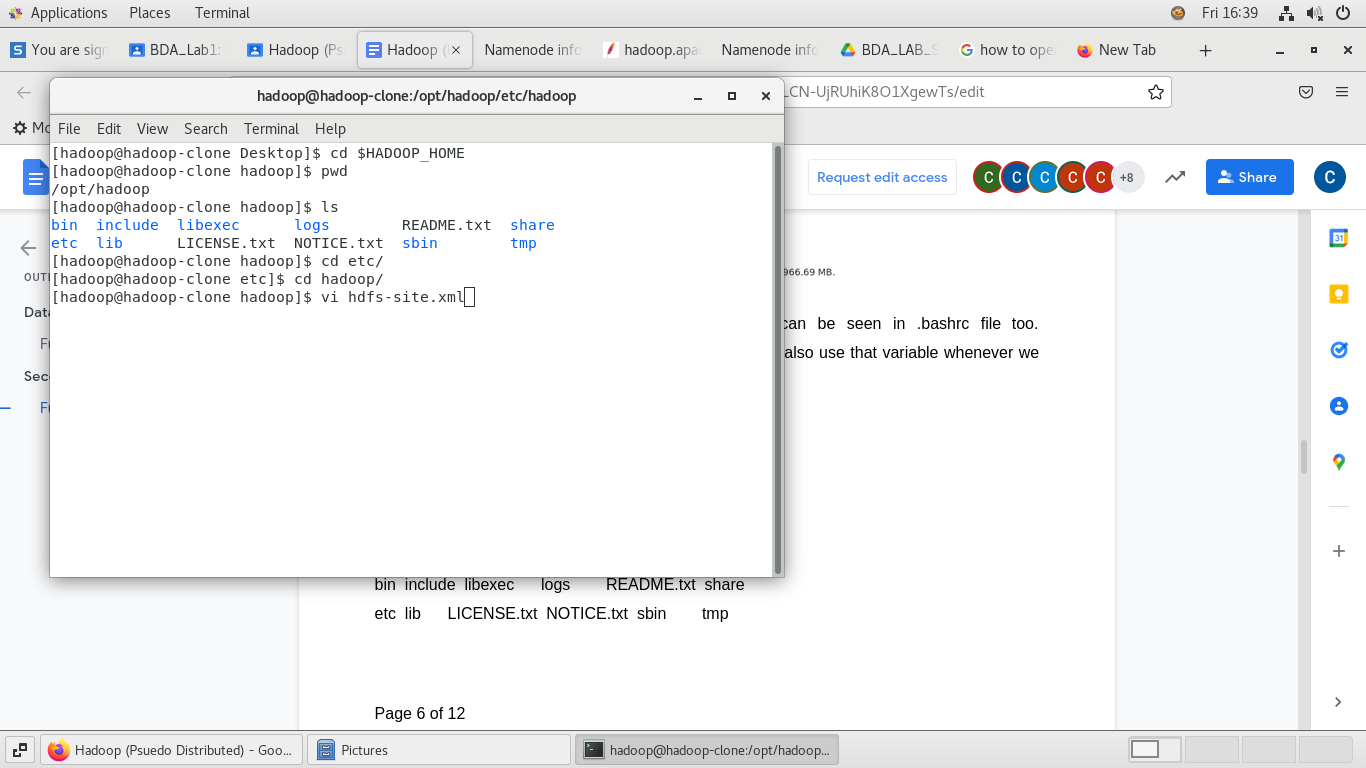
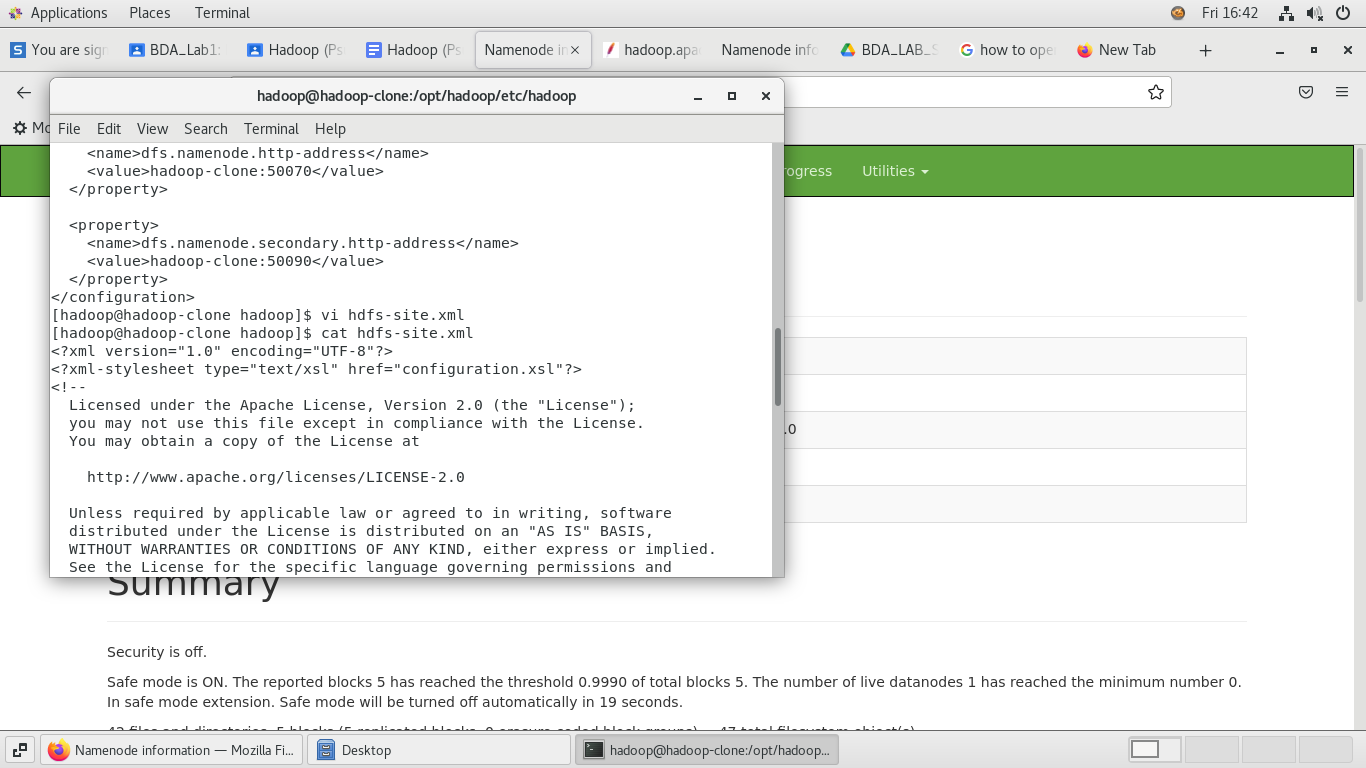




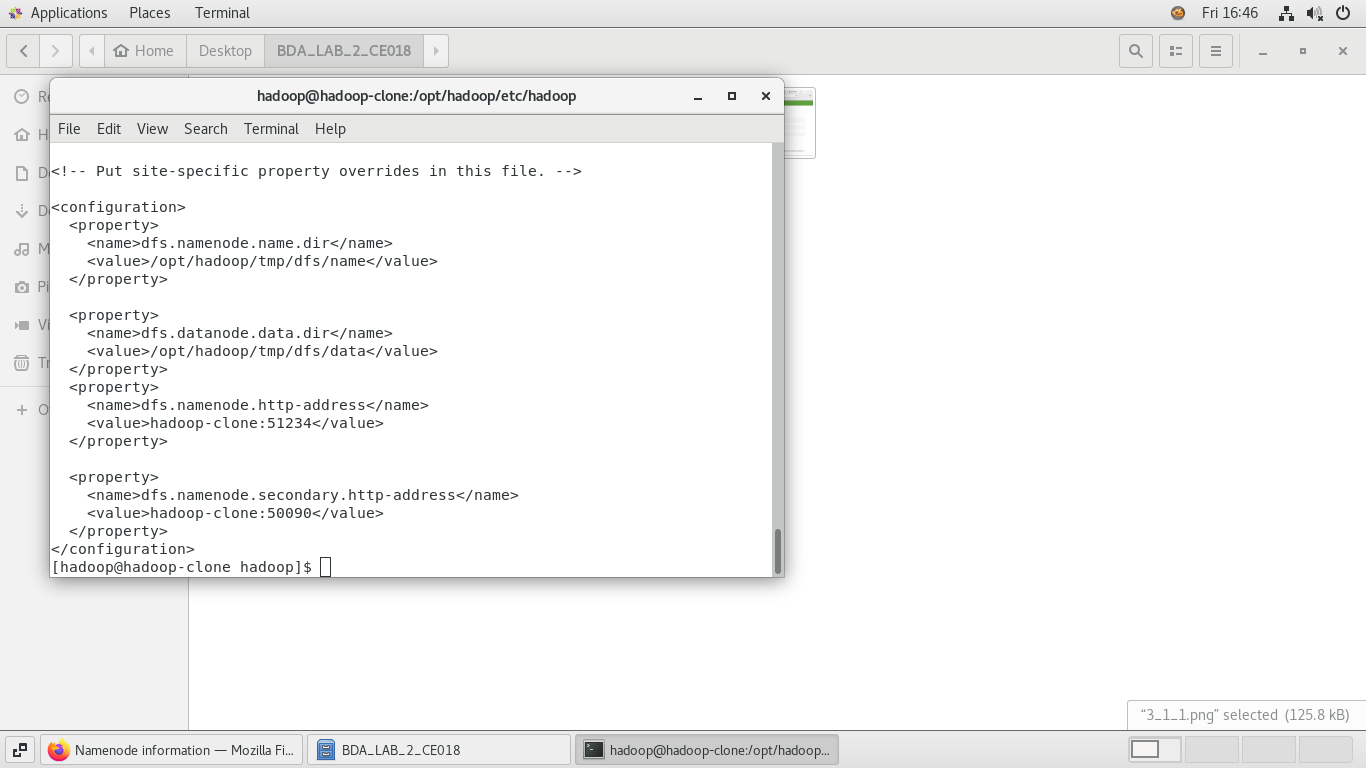
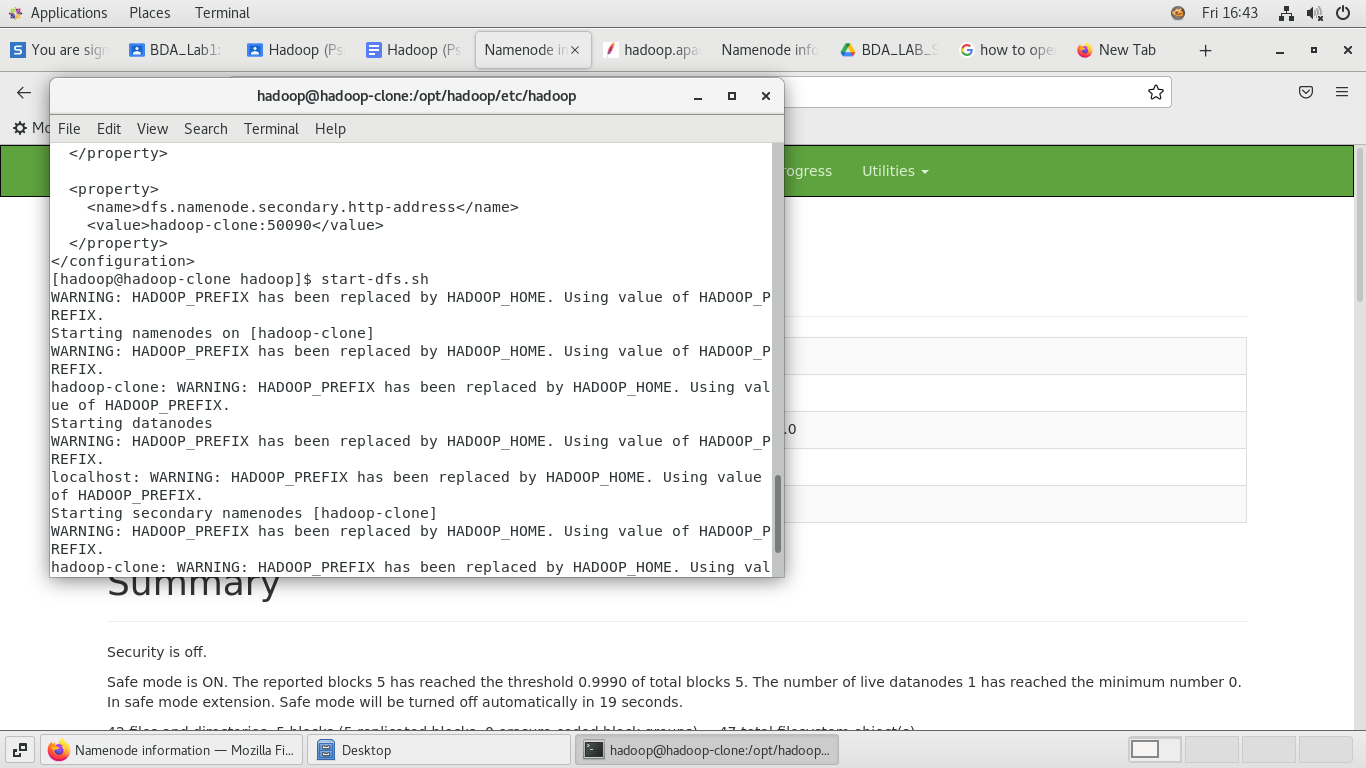


**Q. 3: Change the URL of HDFS Web UI to localhost:51234 instead of localhost:50070 and the access the web UI using new URL.**

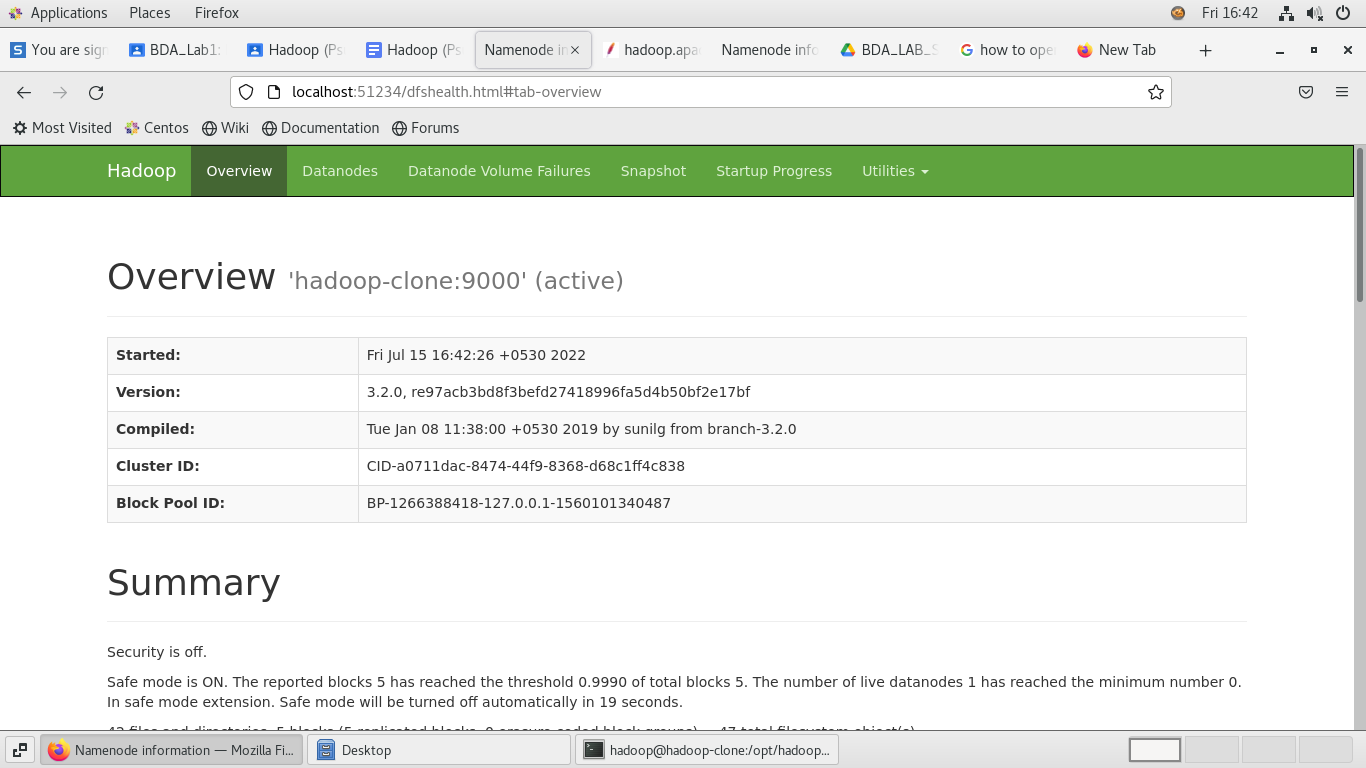
* Hadoop contains many configurations file. From the one such file is hdfs-site.xml which is used to configure the settings for HDFS daemons; the NamdNode, DataNode, and Secondary NameNode.
* In this file we can decide the actual number of replications of file, configure the permissions, block replication.
* 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.







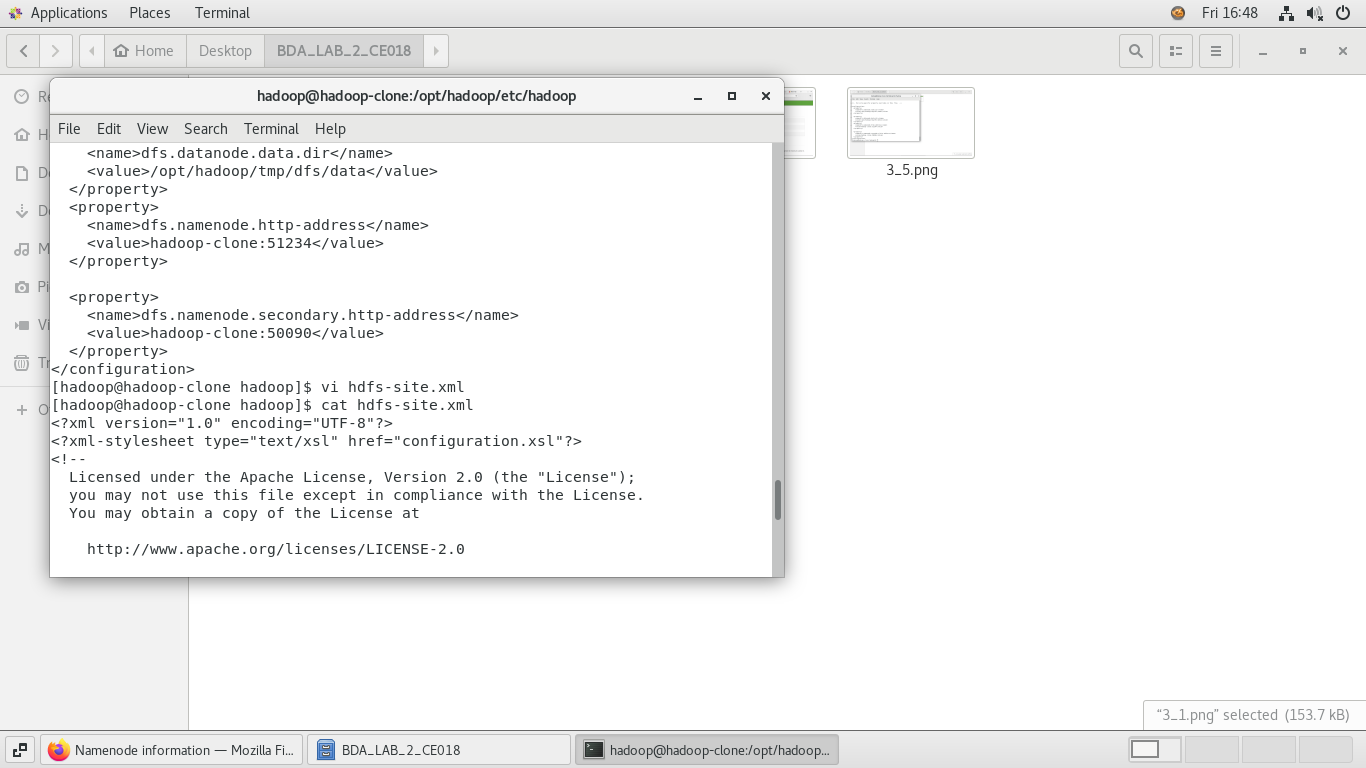
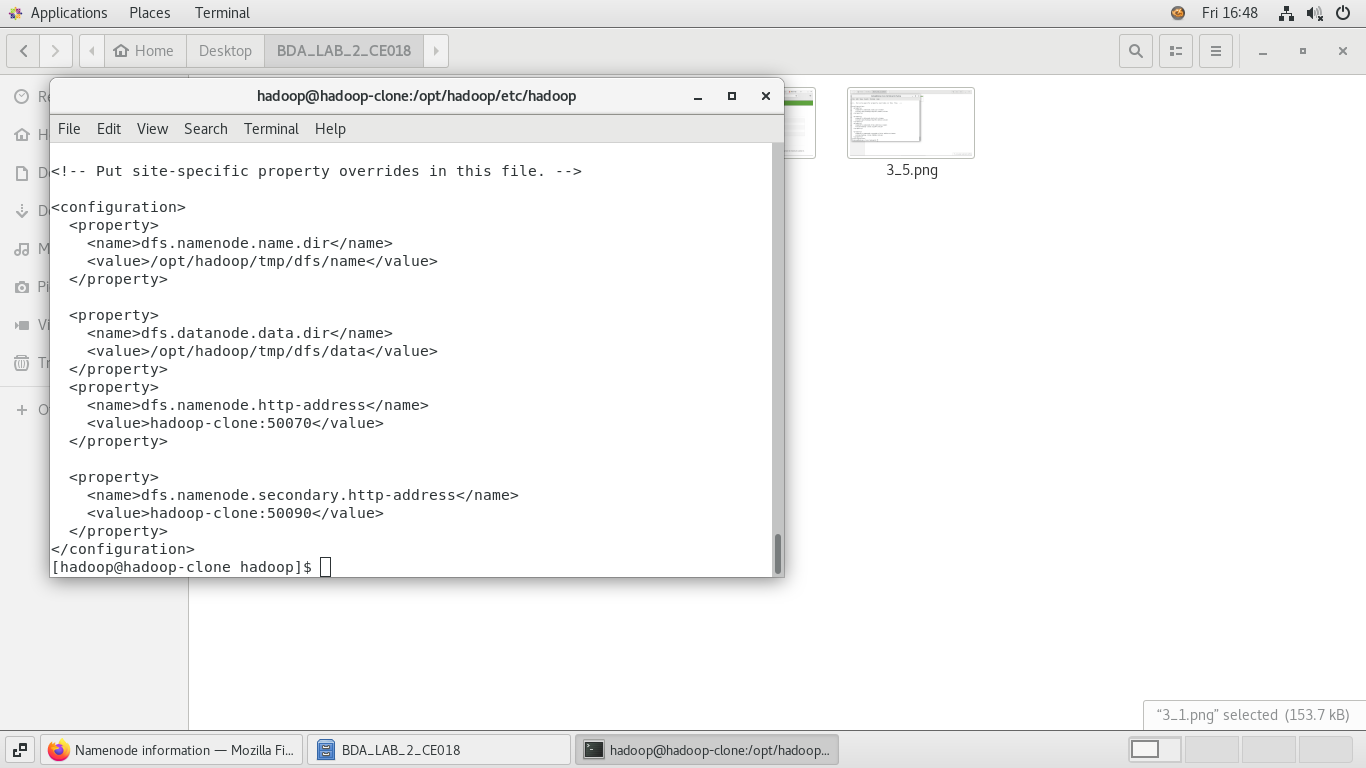




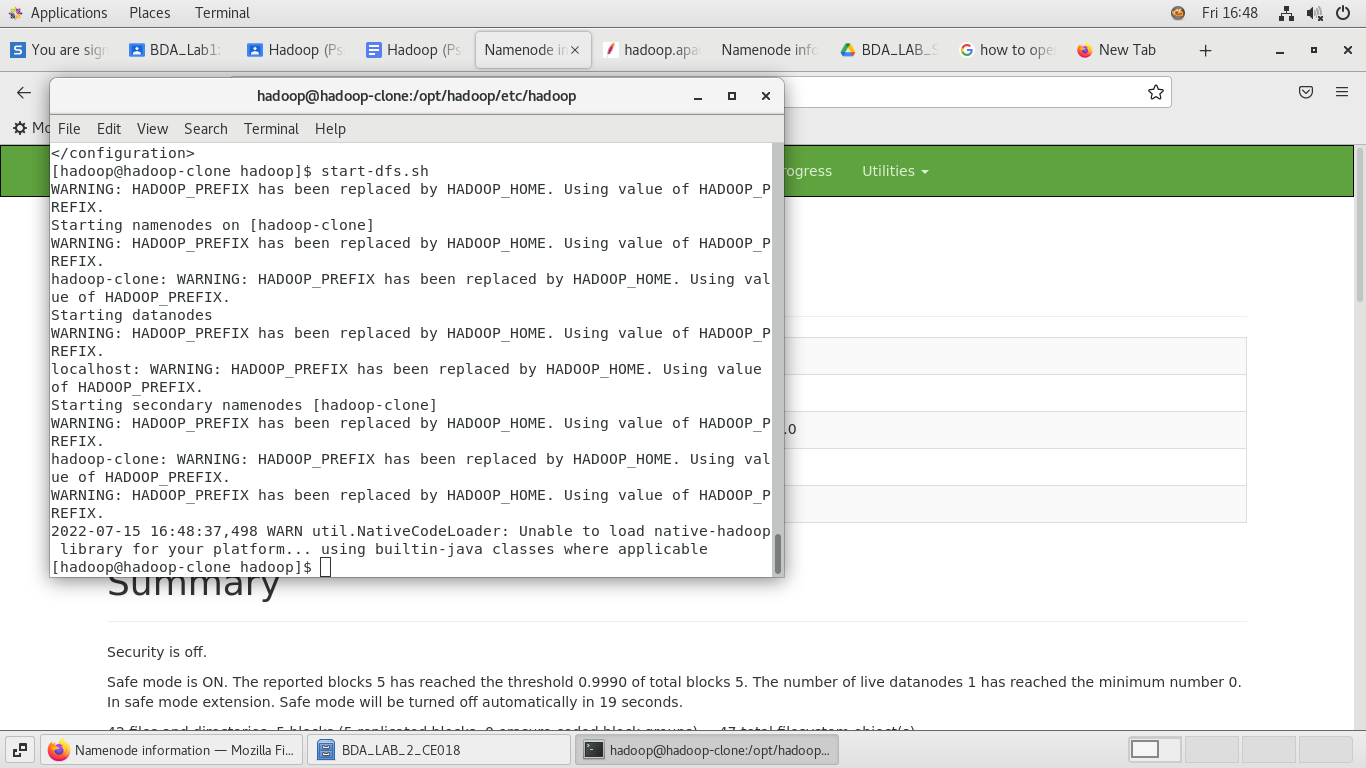
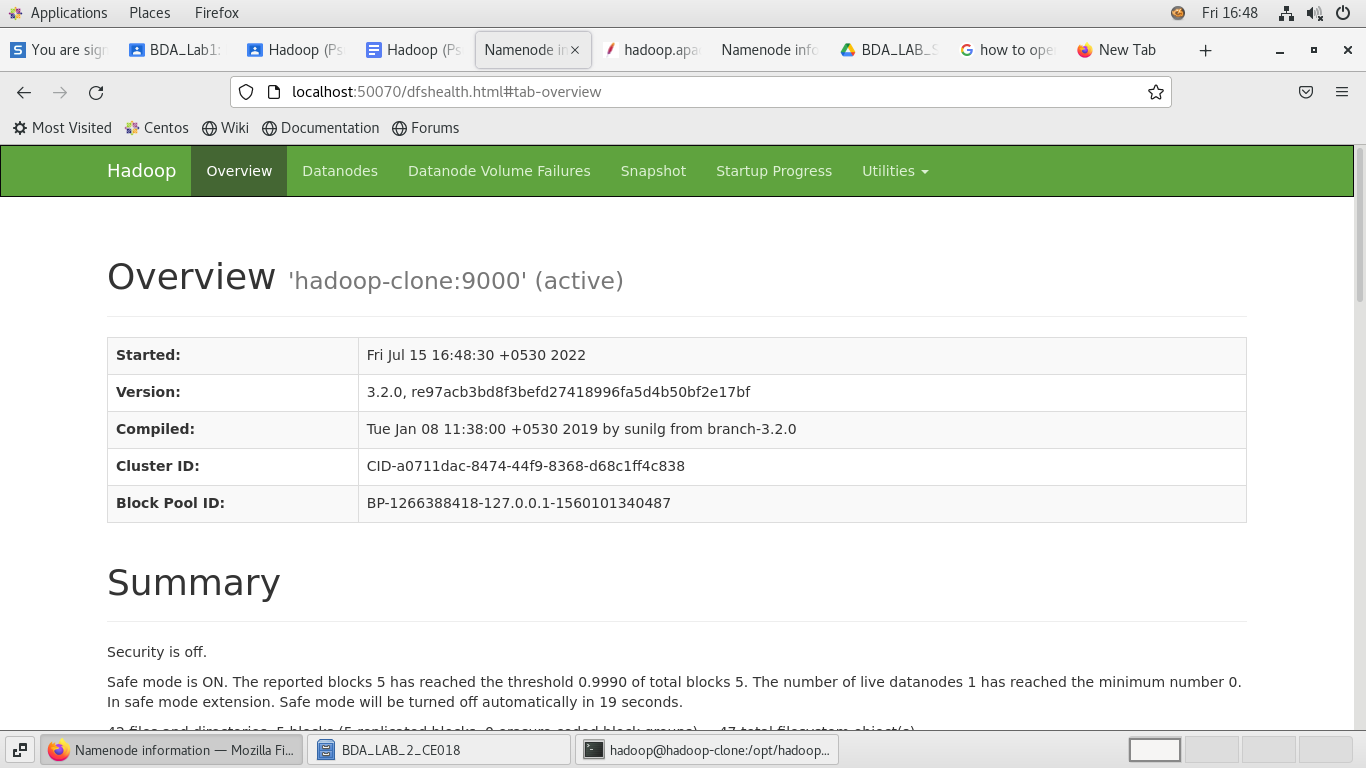


**Q. 4: Undo changes done in task3.**

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

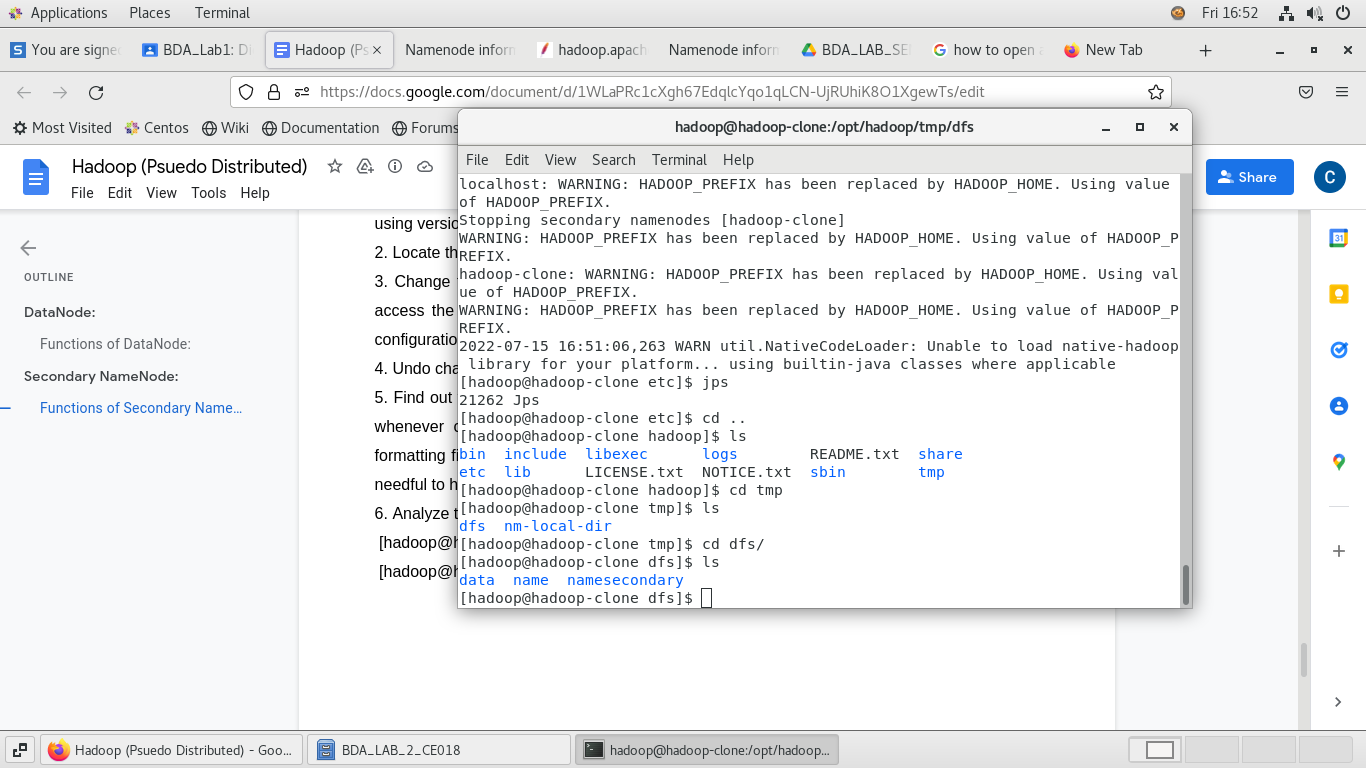




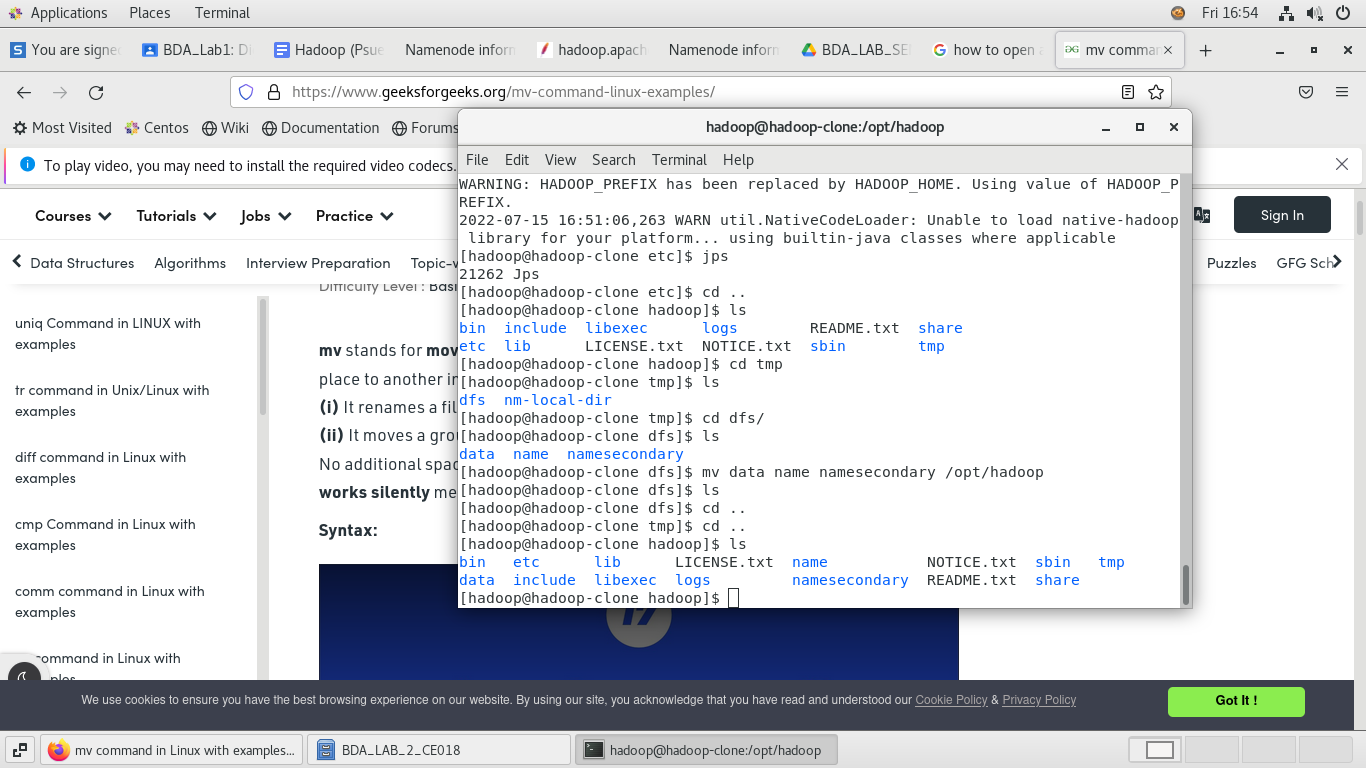




**Q. 5: Find out what is current location of namenode and datanode storage directories. Normally, whenever cleanup is required, tmp folder is one to be emptied up. Hence, it is advised data and metadata better be outside tmp. Do the needful to have namenode and datanode directories in /opt/Hadoop.**

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

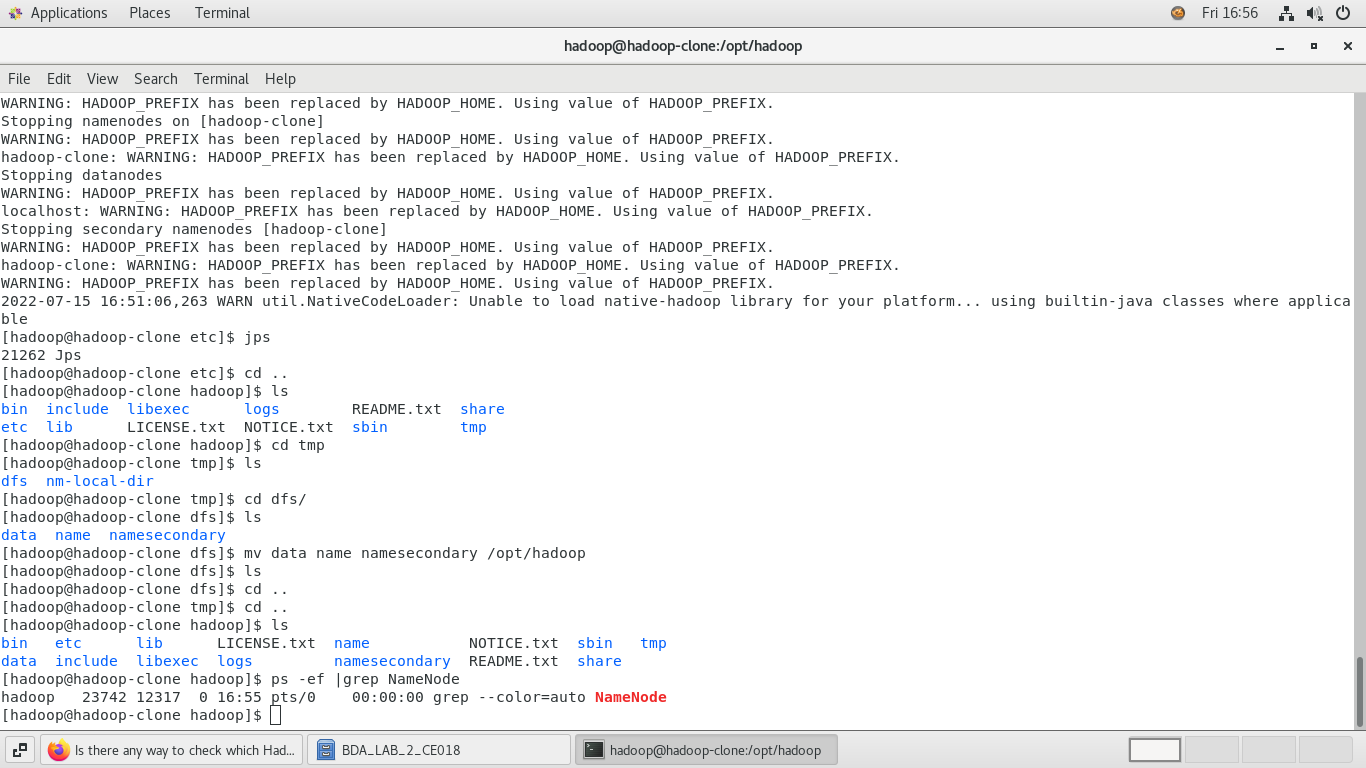






**Q. 6: Analyse the output of the following command:**

1. **[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.





1. **[hadoop@hadoop-clone Hadoop] $ ps -ef |grep DataNode**
   * This command used to verify that DataNoe is up and running.



