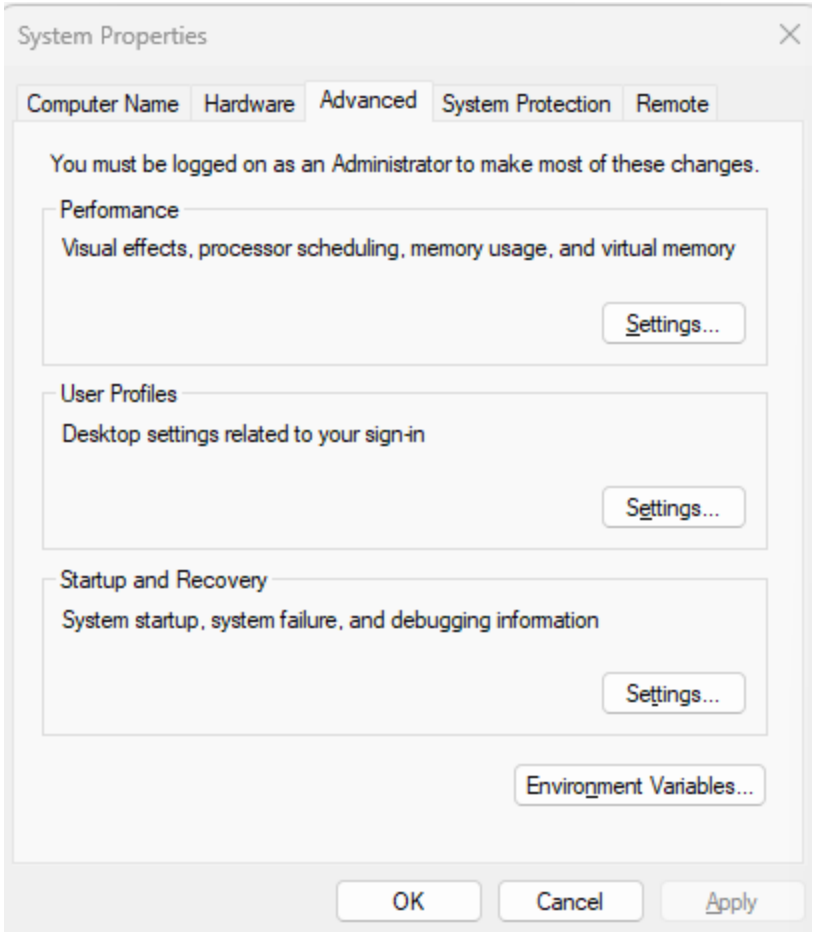
Practical 1:

Aim: Installation of Hadoop

Step 1: Download Binary File for Windows https://hadoop.apache.org/releases.html

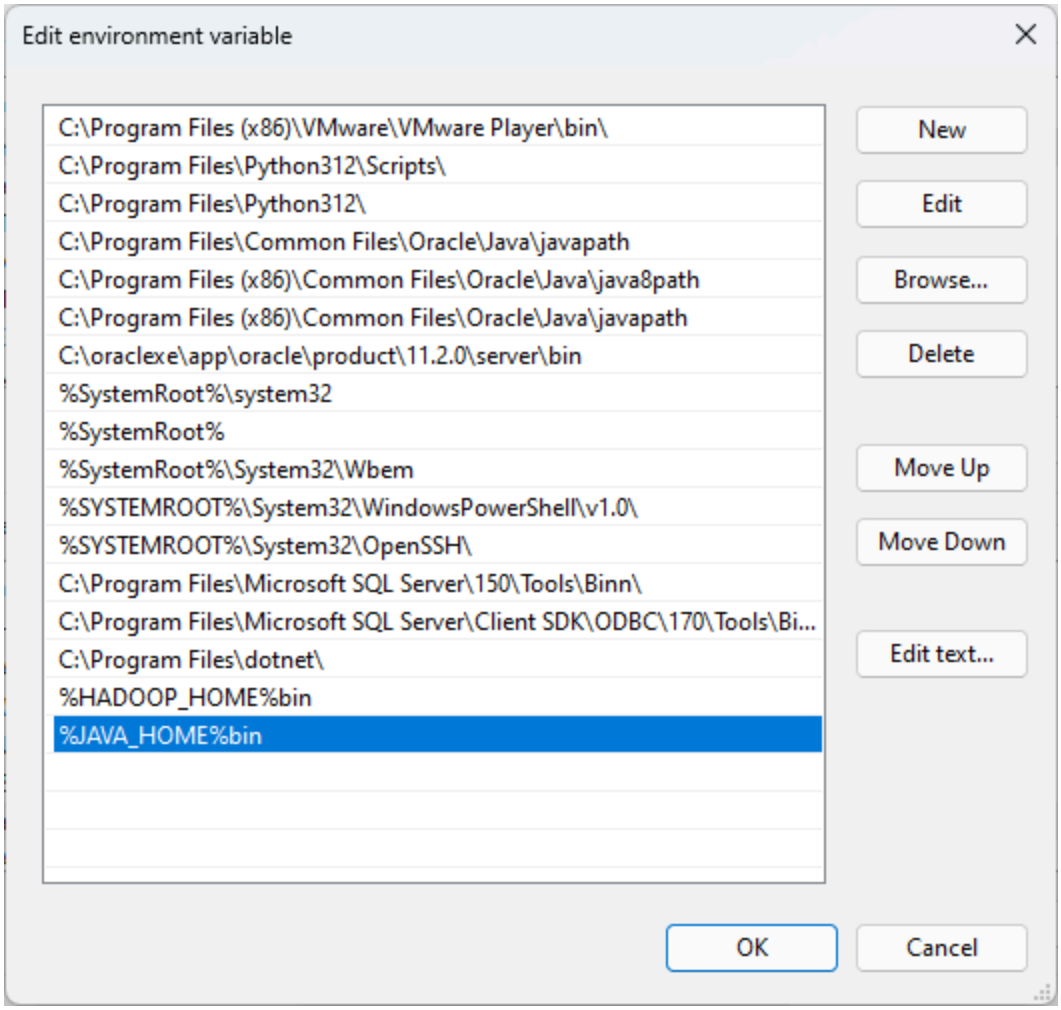
Step 2: Extract the files in C drive .
Step 3: Edit Environment Variables.



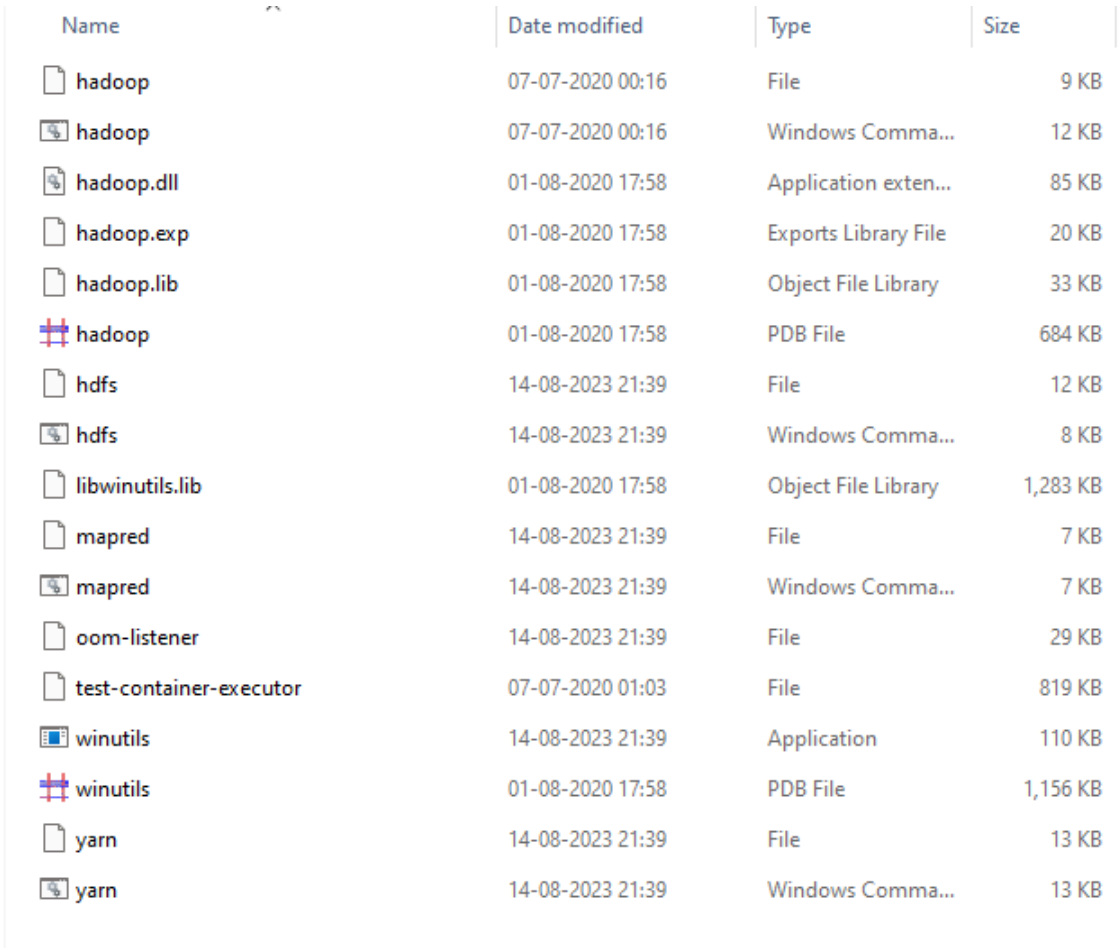
Step 4: Under System Variables click “New” and set “Variable name” as JAVA_HOME and “Variable value” as the path of your JAVA JDK.

Step 5: Similarly add “HADOOP_HOME” variable and download the bin folder from the below link

https://drive.google.com/drive/folders/1iURNbow2IglhAhSy3sfY5xxVfAg33NBW



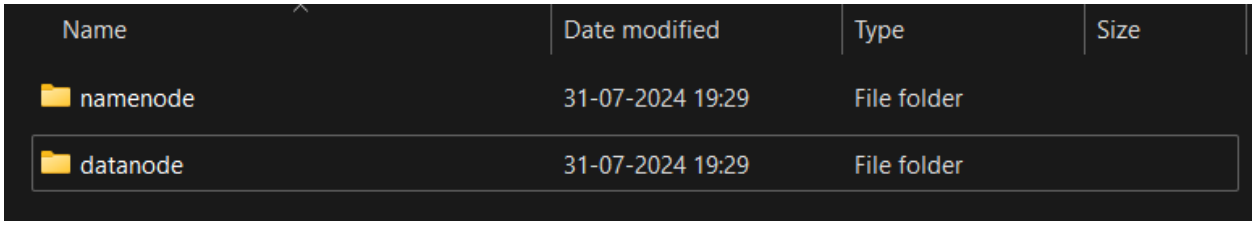
Step 6: Extract the bin archive and replace the bin folder in Hadoop folder with the bin folder in this archive.



Name	Date modified	Type	Size
hadoop	07-07-2020 00:16	File	9 KB
hadoop	07-07-2020 00:16	Windows Comma...	12 KB
hadoop.dll	01-08-2020 17:58	Application exten...	85 KB
hadoop.exp	01-08-2020 17:58	Exports Library File	20 KB
hadoop.lib	01-08-2020 17:58	Object File Library	33 KB
hadoop	01-08-2020 17:58	PDB File	684 KB
hdfs	14-08-2023 21:39	File	12 KB
hdfs	14-08-2023 21:39	Windows Comma...	8 KB
libwinutils.lib	01-08-2020 17:58	Object File Library	1,283 KB
mapred	14-08-2023 21:39	File	7 KB
mapred	14-08-2023 21:39	Windows Comma...	7 KB
oom-listener	14-08-2023 21:39	File	29 KB
test-container-executor	07-07-2020 01:03	File	819 KB
winutils	14-08-2023 21:39	Application	110 KB
winutils	01-08-2020 17:58	PDB File	1,156 KB
yarn	14-08-2023 21:39	File	13 KB
yarn	14-08-2023 21:39	Windows Comma...	13 KB

Step 7: Check if “winutils” is working. If you get any dll error then download that dll and paste in the Windows -> System32 folder.

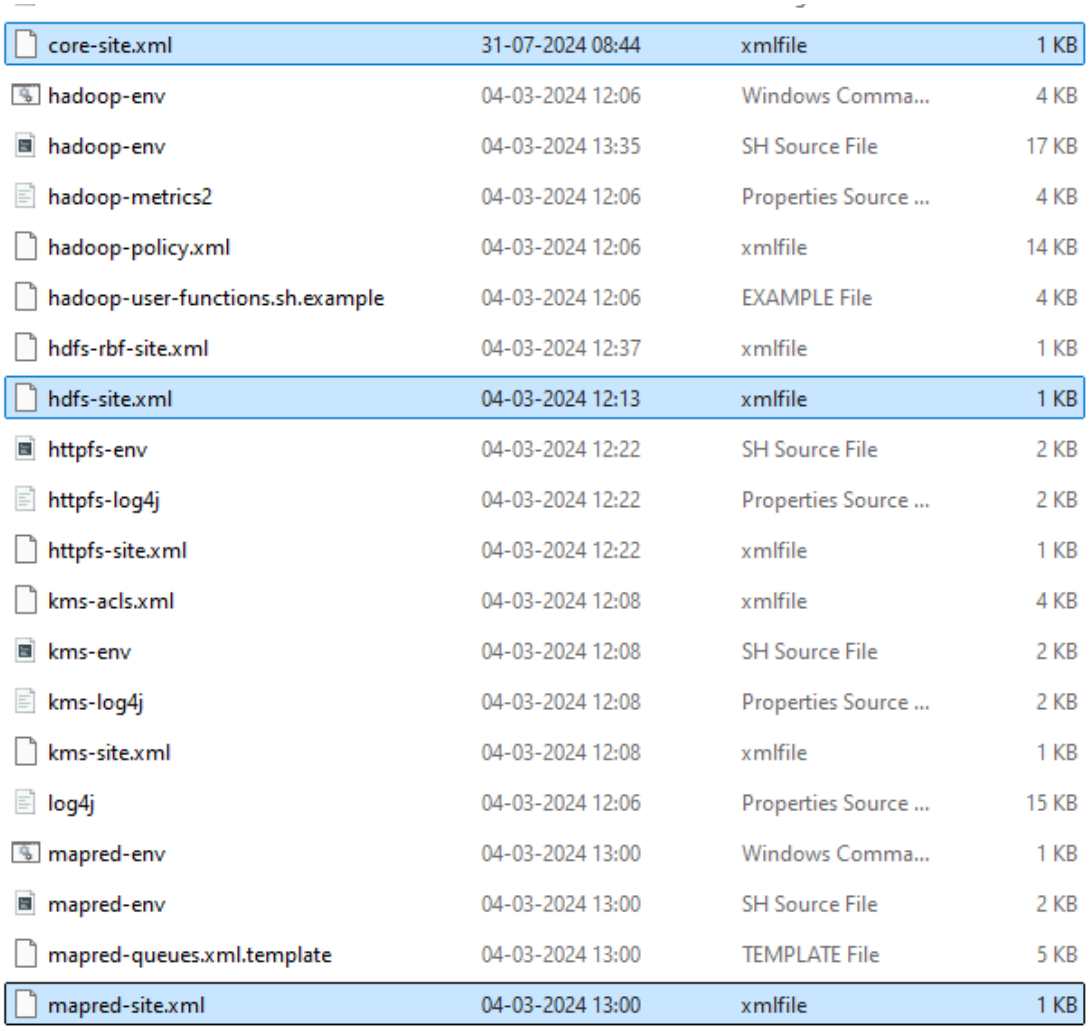
Step 8: Create a data folder in the hadoop home directory and add the folders datanode and namenode to it.



Name	Date modified	Type	Size
namenode	31-07-2024 19:29	File folder	
datanode	31-07-2024 19:29	File folder	

9: Add the following path to “Path” under “System Variables” in “Edit Environment Variables”
C:\hadoop-3.4.0\sbin

10: Make the changes to the following files as given, in “etc/hadoop” folder of hadoop home.



core-site.xml	31-07-2024 08:44	xmlfile	1 KB
hadoop-env	04-03-2024 12:06	Windows Comma...	4 KB
hadoop-env	04-03-2024 13:35	SH Source File	17 KB
hadoop-metrics2	04-03-2024 12:06	Properties Source ...	4 KB
hadoop-policy.xml	04-03-2024 12:06	xmlfile	14 KB
hadoop-user-functions.sh.example	04-03-2024 12:06	EXAMPLE File	4 KB
hdfs-rbf-site.xml	04-03-2024 12:37	xmlfile	1 KB
hdfs-site.xml	04-03-2024 12:13	xmlfile	1 KB
httpfs-env	04-03-2024 12:22	SH Source File	2 KB
httpfs-log4j	04-03-2024 12:22	Properties Source ...	2 KB
httpfs-site.xml	04-03-2024 12:22	xmlfile	1 KB
kms-acls.xml	04-03-2024 12:08	xmlfile	4 KB
kms-env	04-03-2024 12:08	SH Source File	2 KB
kms-log4j	04-03-2024 12:08	Properties Source ...	2 KB
kms-site.xml	04-03-2024 12:08	xmlfile	1 KB
log4j	04-03-2024 12:06	Properties Source ...	15 KB
mapred-env	04-03-2024 13:00	Windows Comma...	1 KB
mapred-env	04-03-2024 13:00	SH Source File	2 KB
mapred-queues.xml.template	04-03-2024 13:00	TEMPLATE File	5 KB
mapred-site.xml	04-03-2024 13:00	xmlfile	1 KB

core-site.xml

```
<configuration>
        <property>
                <name>fs.default.name</name>
                 <value>hdfs://localhost:9000</value>
        </property>
</configuration>
```

mapred-site.xml

```
<configuration>
        <property>
                <name>mapred.framework.name</name>
                <value>yarn</value>
```

```
        </property>
    </configuration>
```

hdfs-site.xml

```
    <configuration>
        <property>
            <name>dfs.replication</name>
            <value>1</value>
        </property>

        <property>
            <name>dfs.namenode.name.dir</name>
            <value>C:\hadoop-3.4.0\data\namenode</value>
        </property>

        <property>
            <name>dfs.datanode.data.dir</name>
            <value>C:\hadoop-3.4.0\data\datanode</value>
        </property>
    </configuration>
```

yarn-site.xml

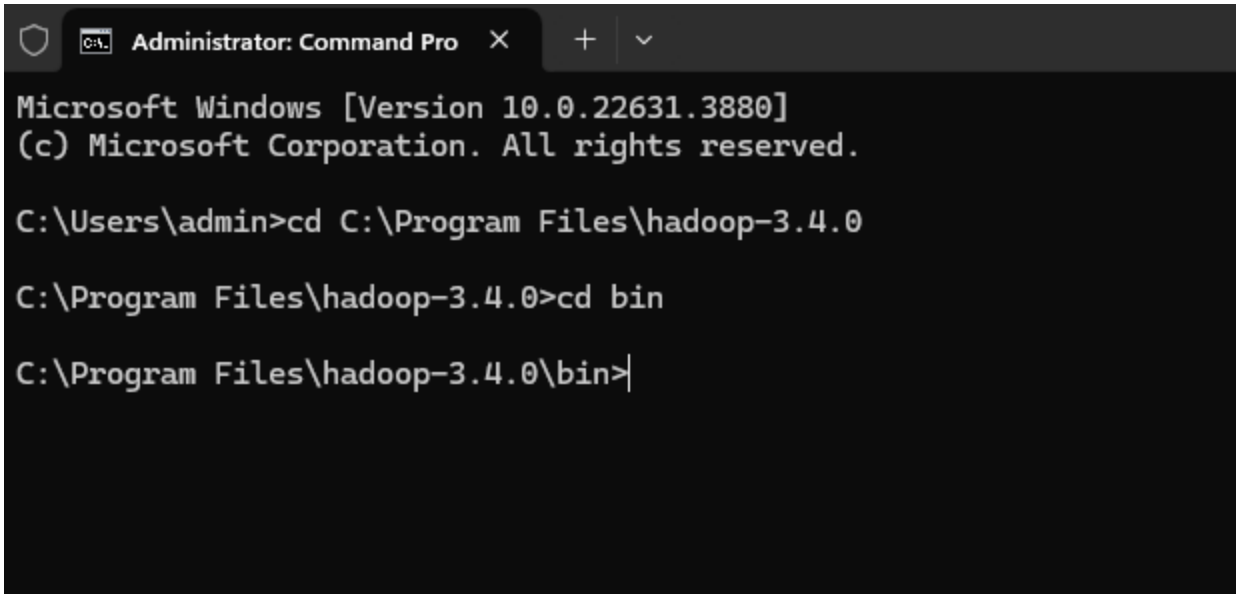
```
    <configuration>
        <property>
            <name>yarn.nodemanager.aux-services</name>
            <value>mapreduce_shuffle</value>
        </property>
        <property>
            <name>yarn.nodemanager.auxservice.mapreduce.shuffle.class</name>
            <value>org.apache.hadoop.mapred.shuffleHandler</value>
        </property>
    </configuration>
```

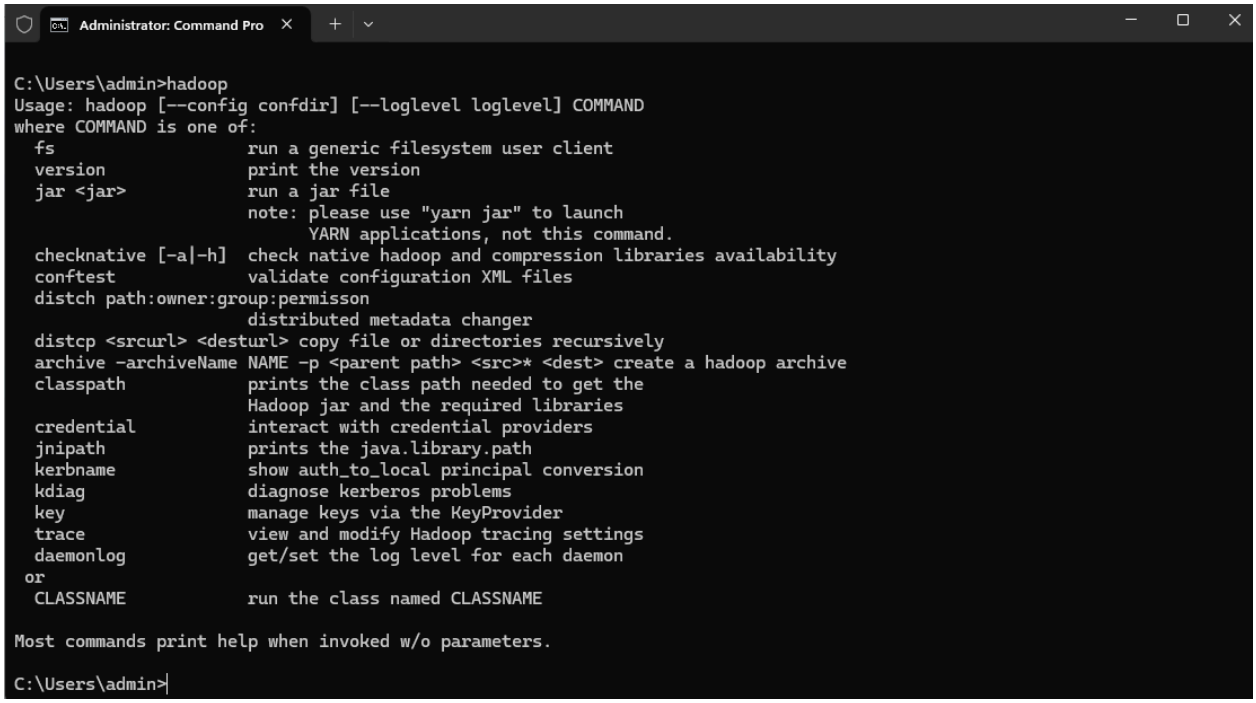
Step 11: Go to hadoop-env.cmd file in /etc/hadoop folder and replace the set JAVA_HOME=%JAVA_HOME% line with the following:

```
set JAVA_HOME=C:\Progra~1\Java\jdk-21
```

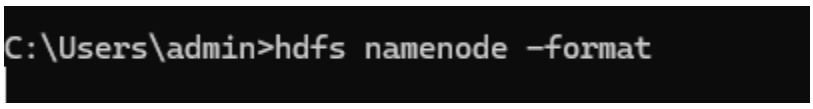
Step 12: Restart your PC for the changes to take effect.

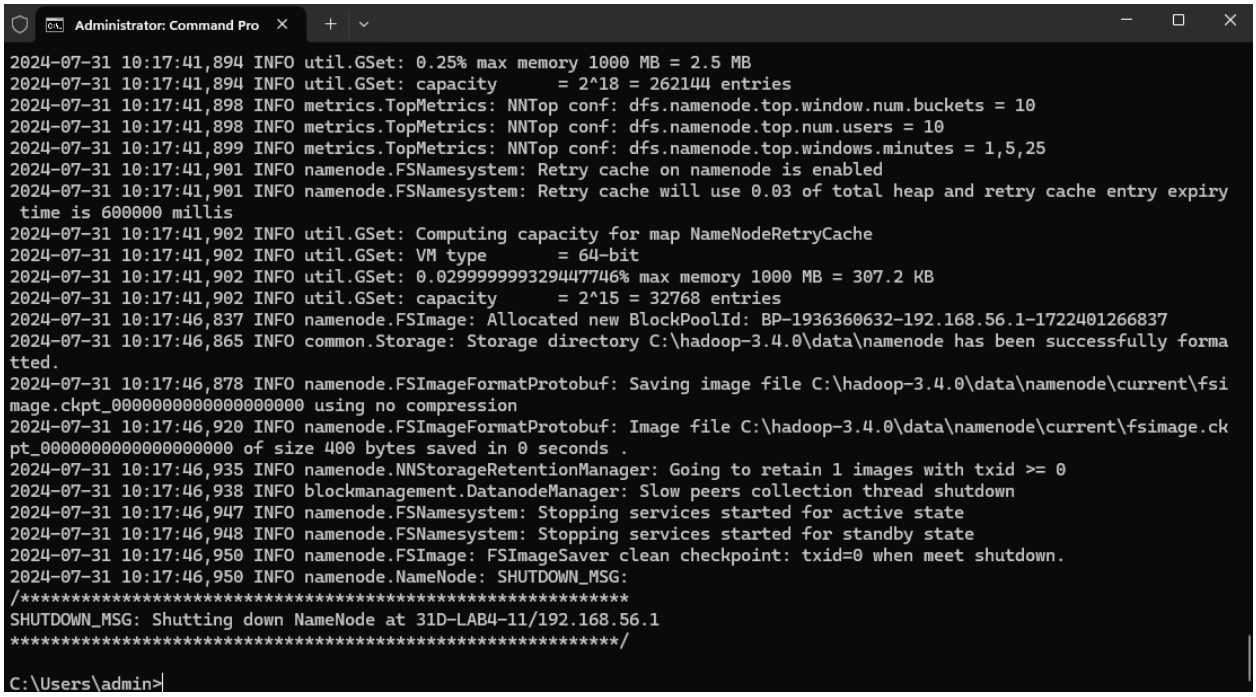
Step 13: Go to Admin Command prompt and type “hadoop” to see if the server is recognized.



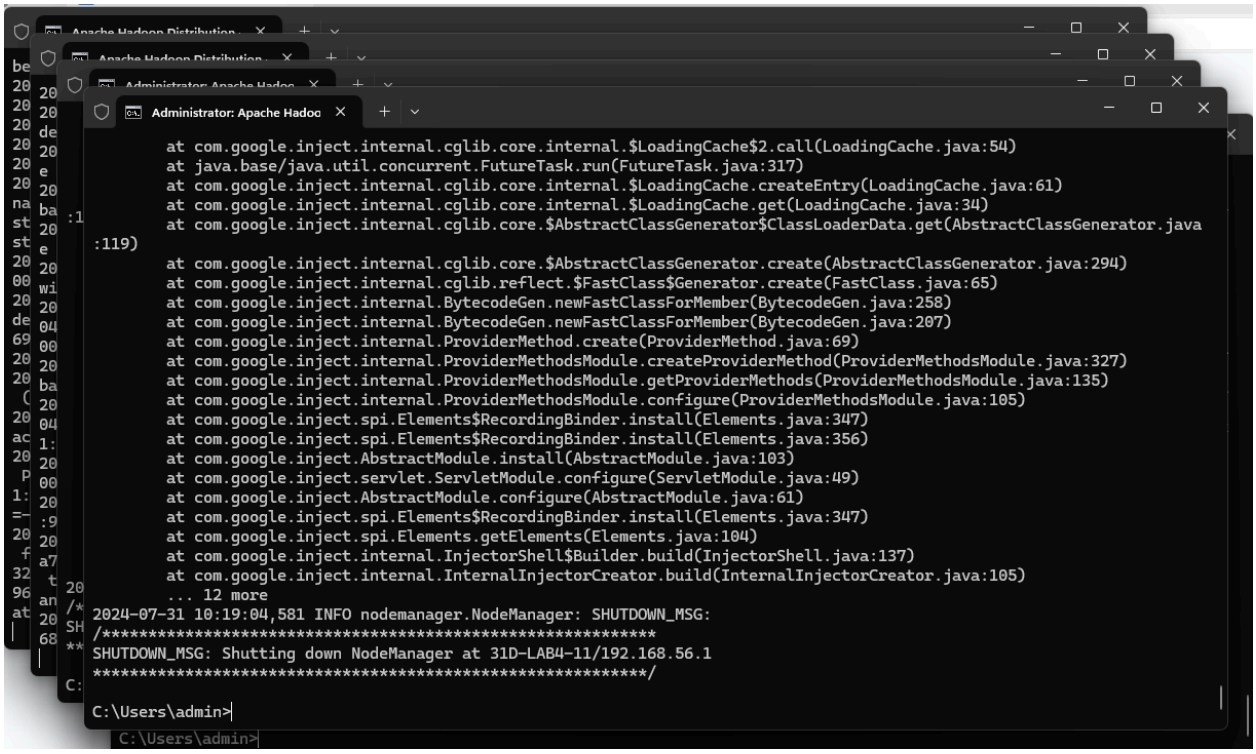


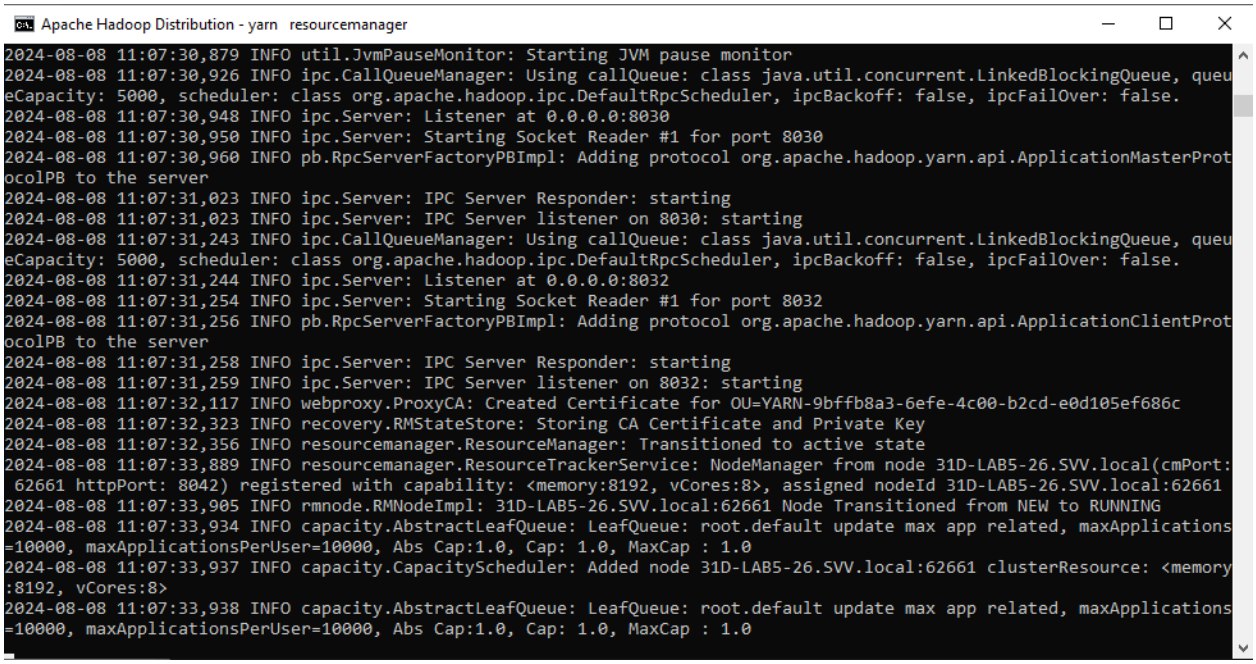
Step 14: Type “hdfs namenode -format” to format the namenode.

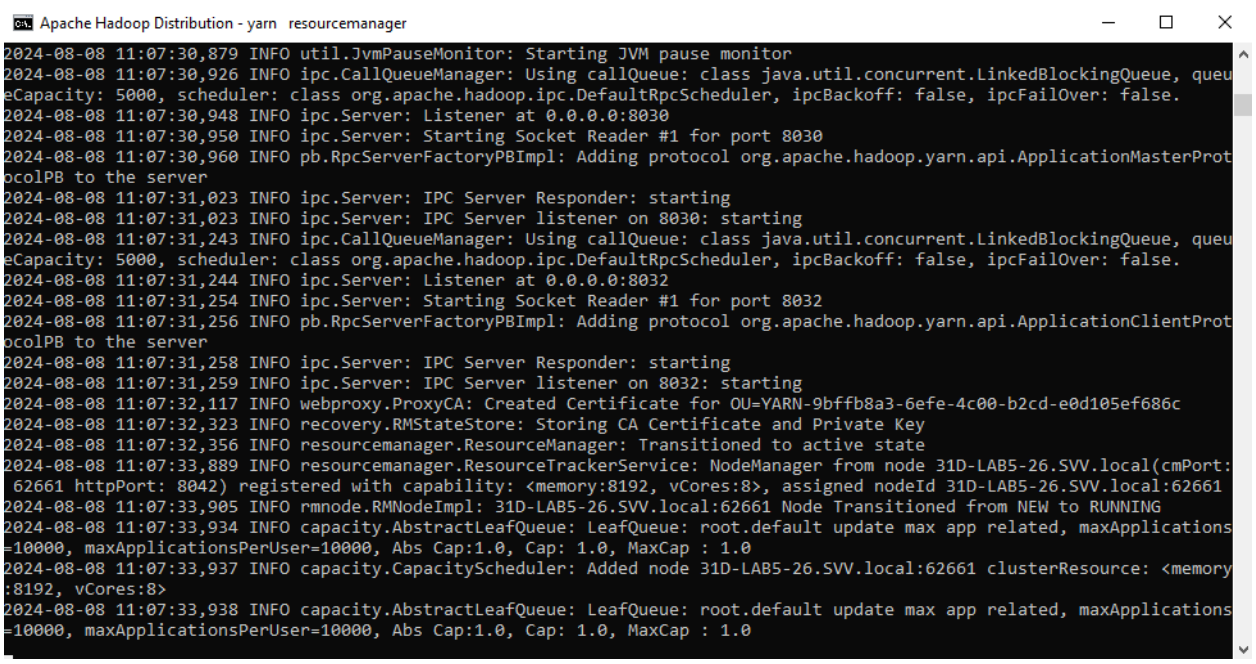




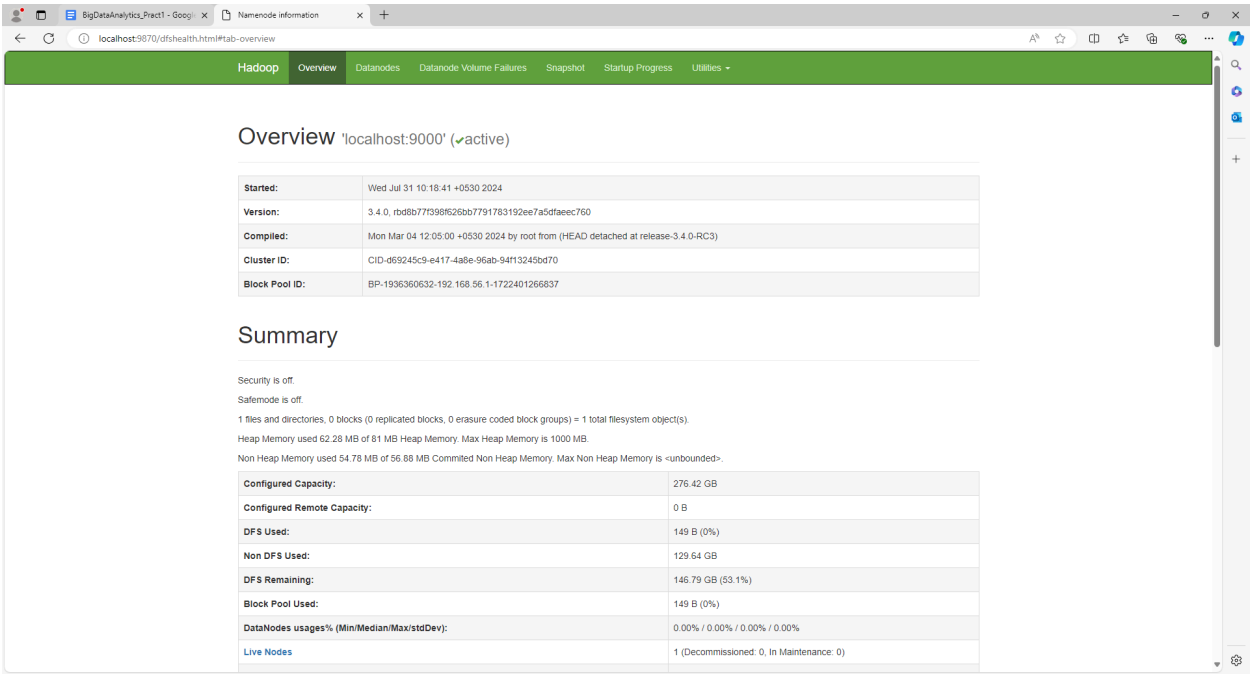
Step 15: Type start-all.cmd to start all hadoop processes (make sure you have added set JAVA_HOME=C:\Progra~1\Java\jdk-22)



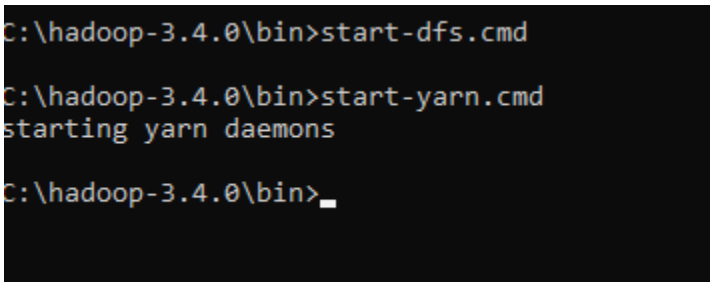




Step 16: Go to your browser and type localhost:9870 to view Hadoop Page.



Step 17: Now, go to cmd and type start-yarn.cmd:



Step 18: Now, go to localhost:8088 and observe accordingly: