ROLLNO:19UITE005 REG NO:920419205010

NAME:M.DHARIKA

SUBJECT:IT8711-FOSS and cloud computing laboratory

EXERCISE NO:9

PROCEDURAL STEPS

Step 1: Download the following Packages

1. hadoop-3.2.2.tar.gz

   Link: https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.2.2/hadoop-3.2.2.tar.gz
2. 7zip to unzip the tar.gz file

   Link: https://www.7-zip.org/download.html
3. Java 8 (JDK-8U333 & JRE-8U333)

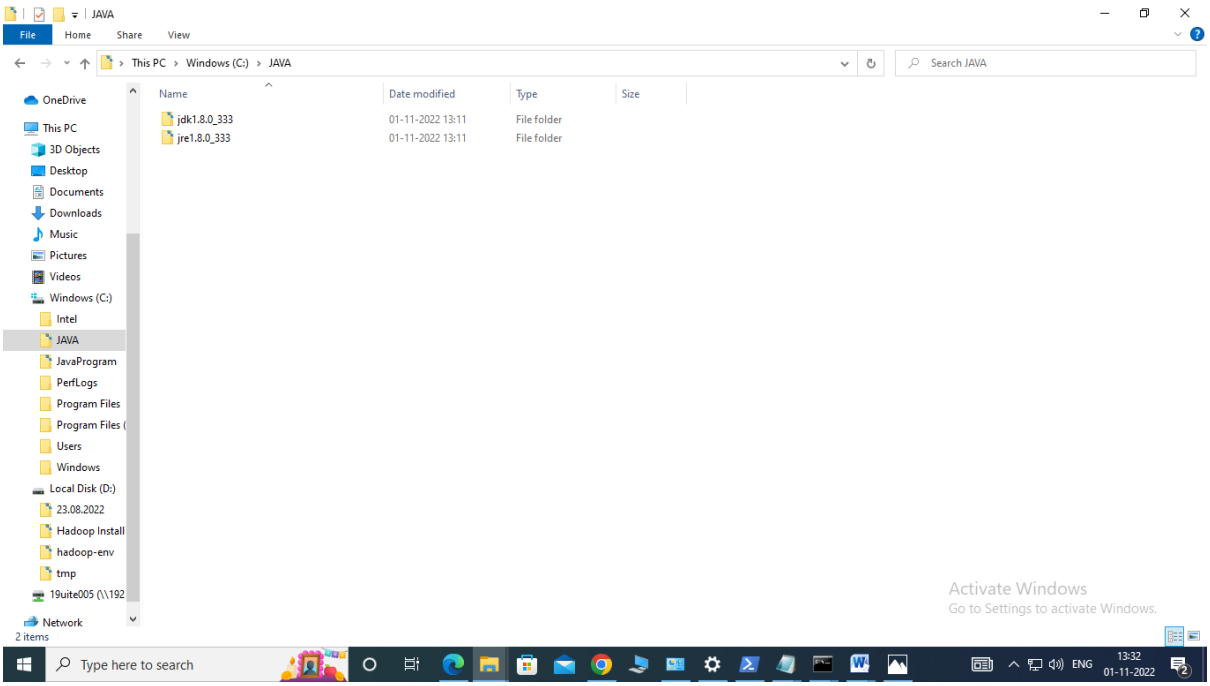
   Link: https://www.oracle.com/java/technologies/javase/javase8u211-later-archive-downloads.html
4. Hadoop dll Files

   Link: https://github.com/cdarlint/winutils/archive/refs/heads/master.zip
5. hadoop-hdfs-3.2.2.jar

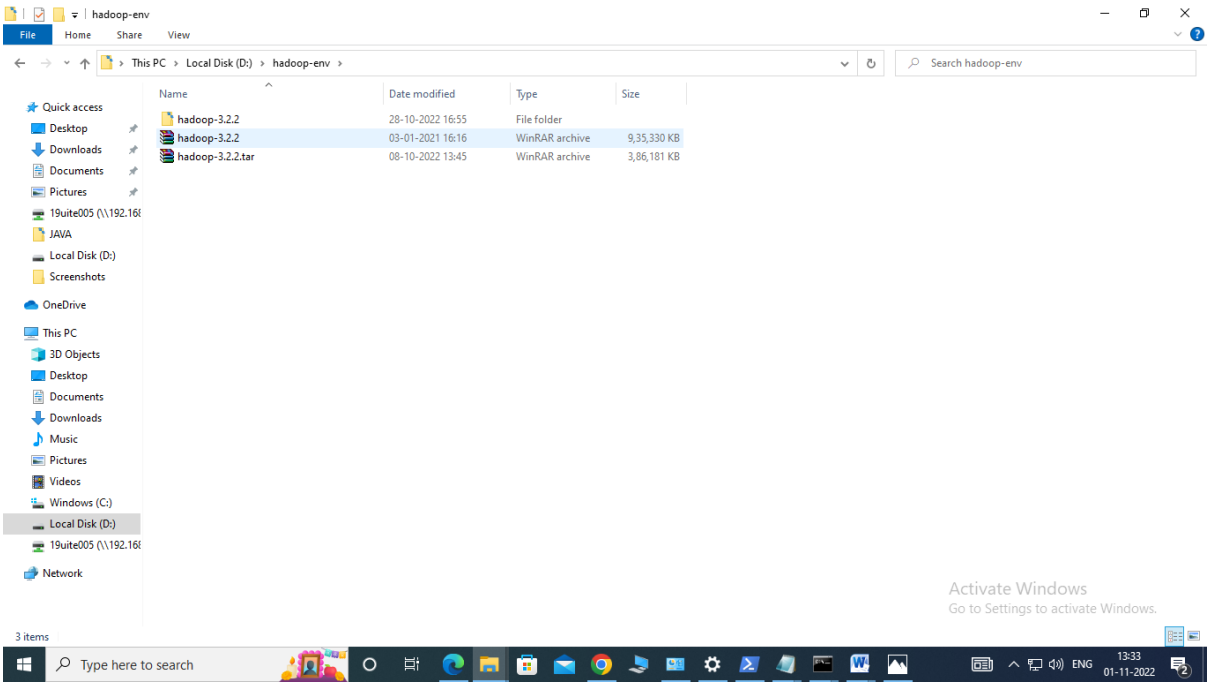
   Link: https://jar-download.com/artifacts/org.apache.hadoop/hadoop-hdfs/3.2.2/source-code

Step 2: Install / Extract the following Packages:

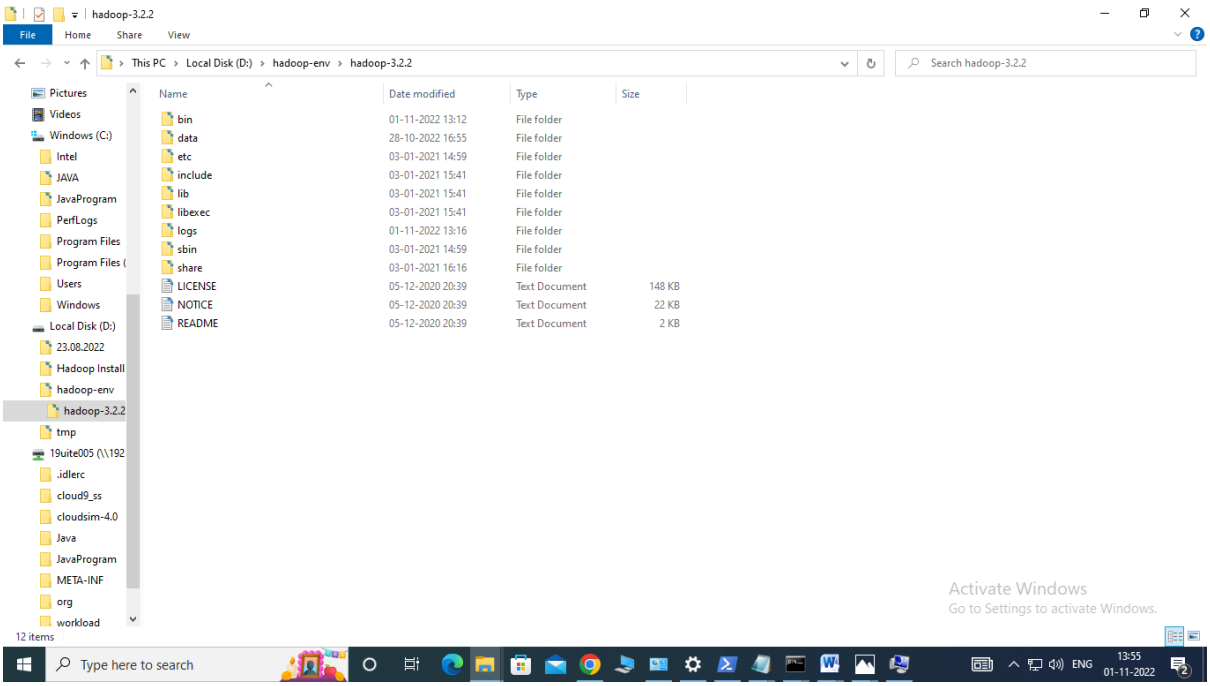
1. Install Java 8 in the location **c:\JAVA**. Also create two folders namely **"jdk1.8.0_333"** and **"jre1.8.0_333"**.



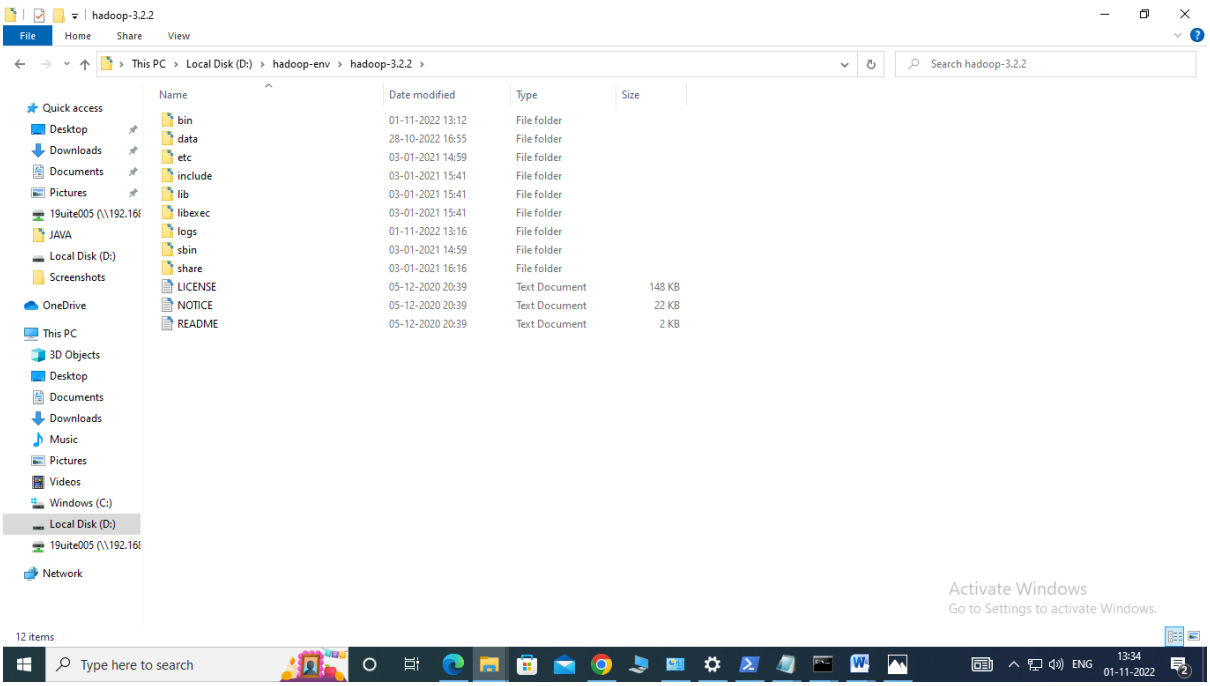
2. Create a folder **“hadoop-env”** in **D:**



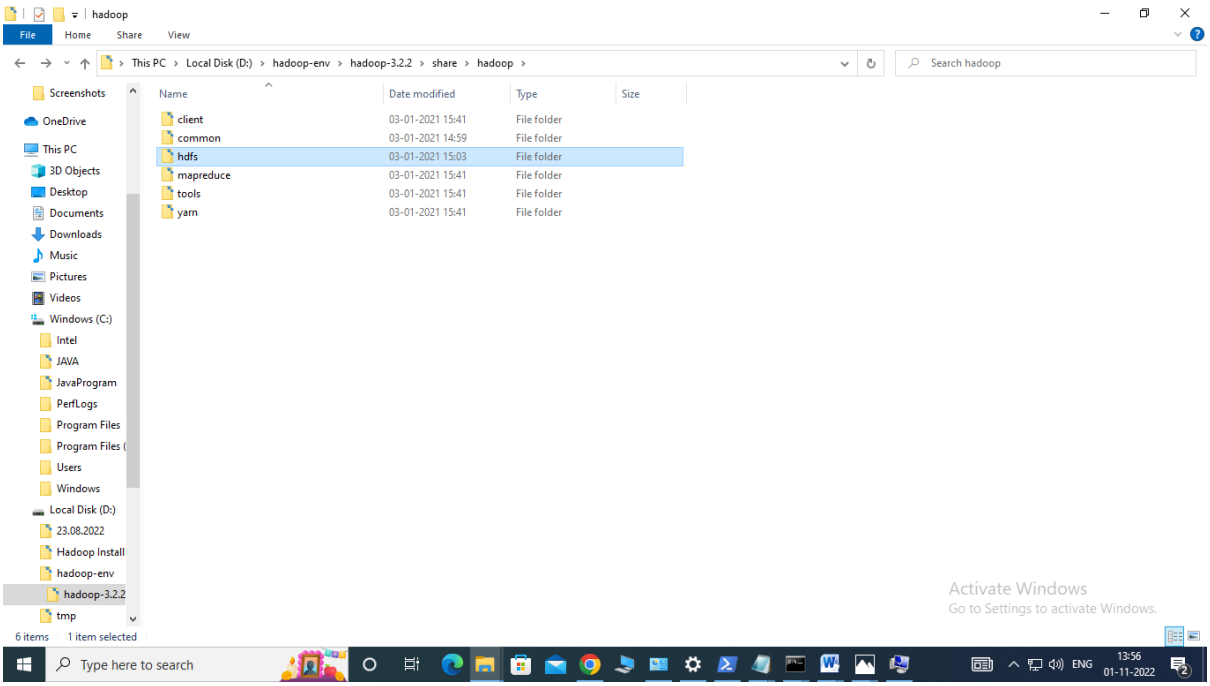
3. Paste the downloaded “hadoop-3.2.2.tar.gz” into “d:\ hadoop-env”



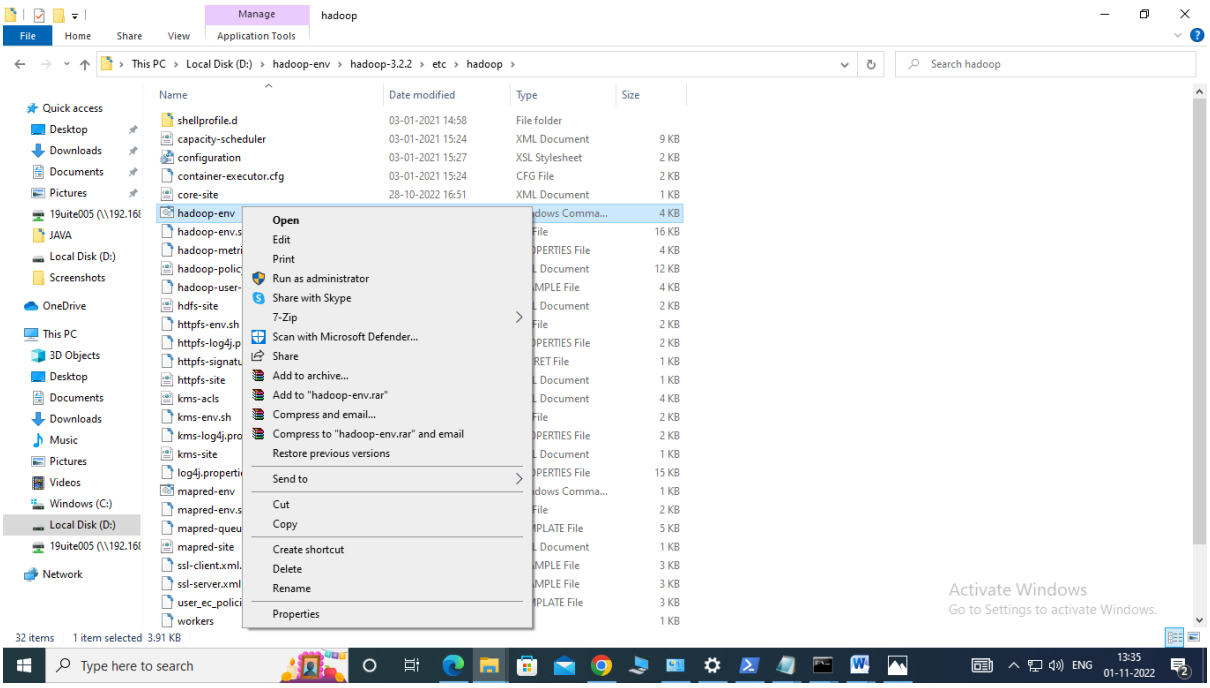
4. Unzip the “hadoop-3.2.2.tar.gz” using 7zip (do the unzip function two times)
5. Copy the downloaded Hadoop dll files to the location “D:\hadoop-env\hadoop-3.2.2\bin”



6. Copy the downloaded hadoop-hdfs-3.2.2.jar to “D:\hadoop-env\hadoop-3.2.2\share\hadoop\hdfs”

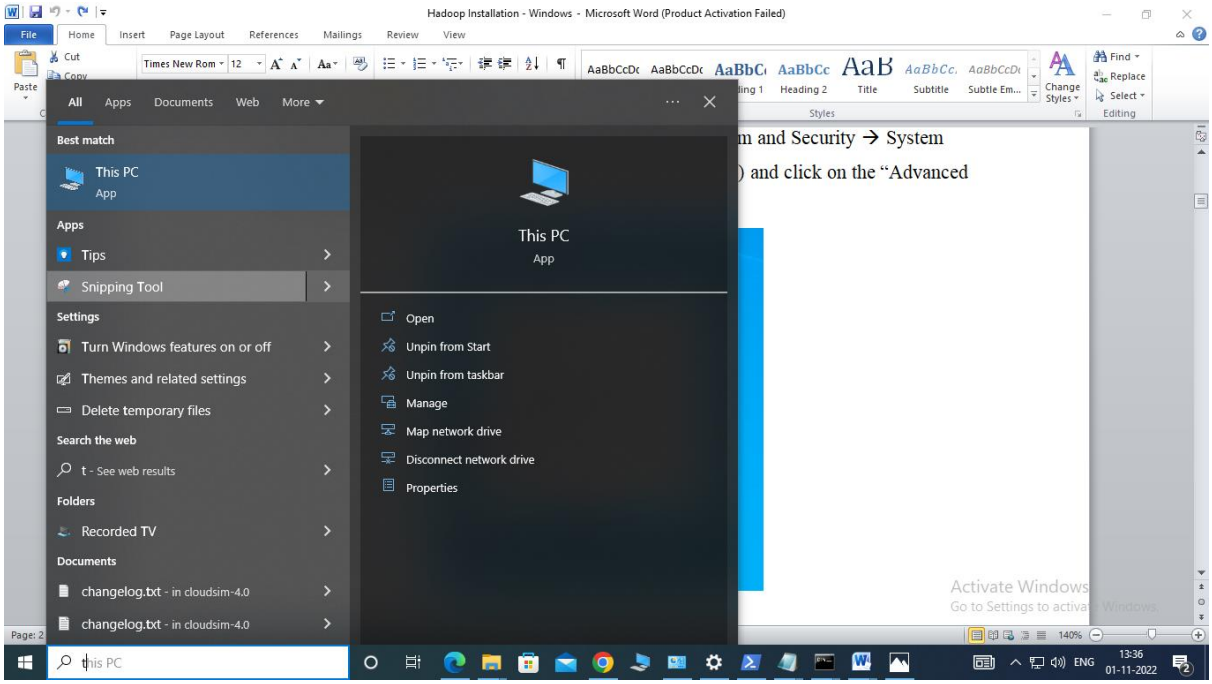


7. Edit the JAVA_HOME = C:\JAVA\jdk1.8.0_333 in the path "**D:\hadoop-env\hadoop-3.2.2\etc\hadoop\hadoop-env.cmd**

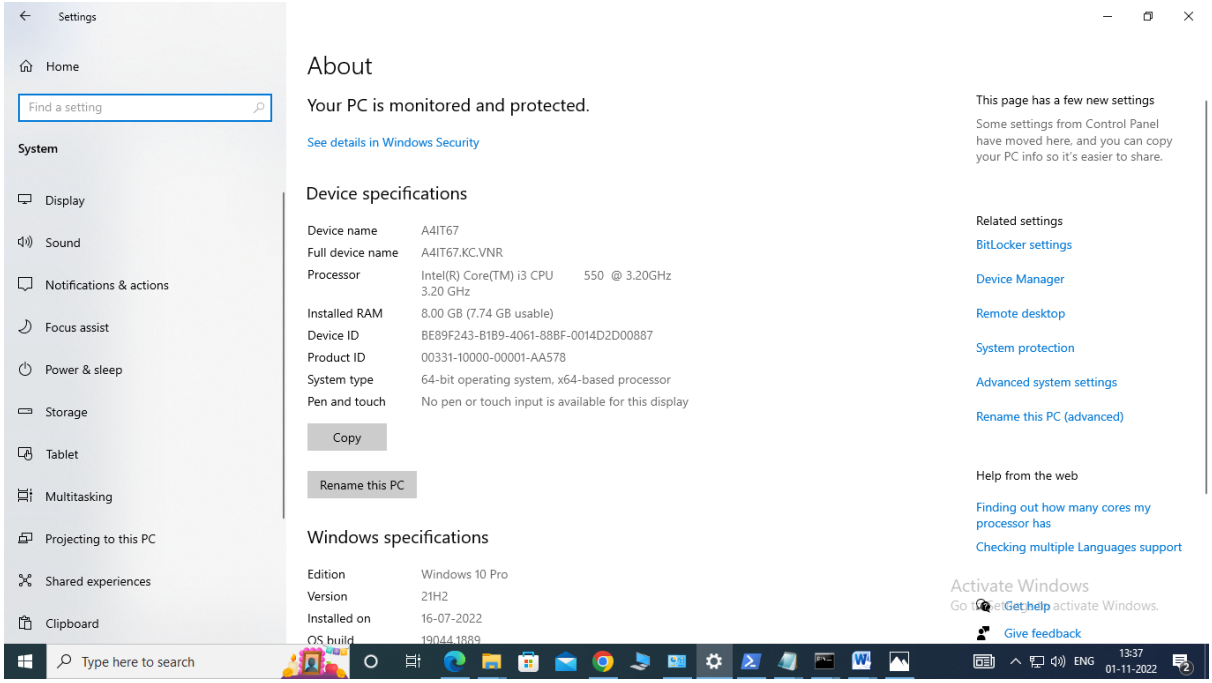


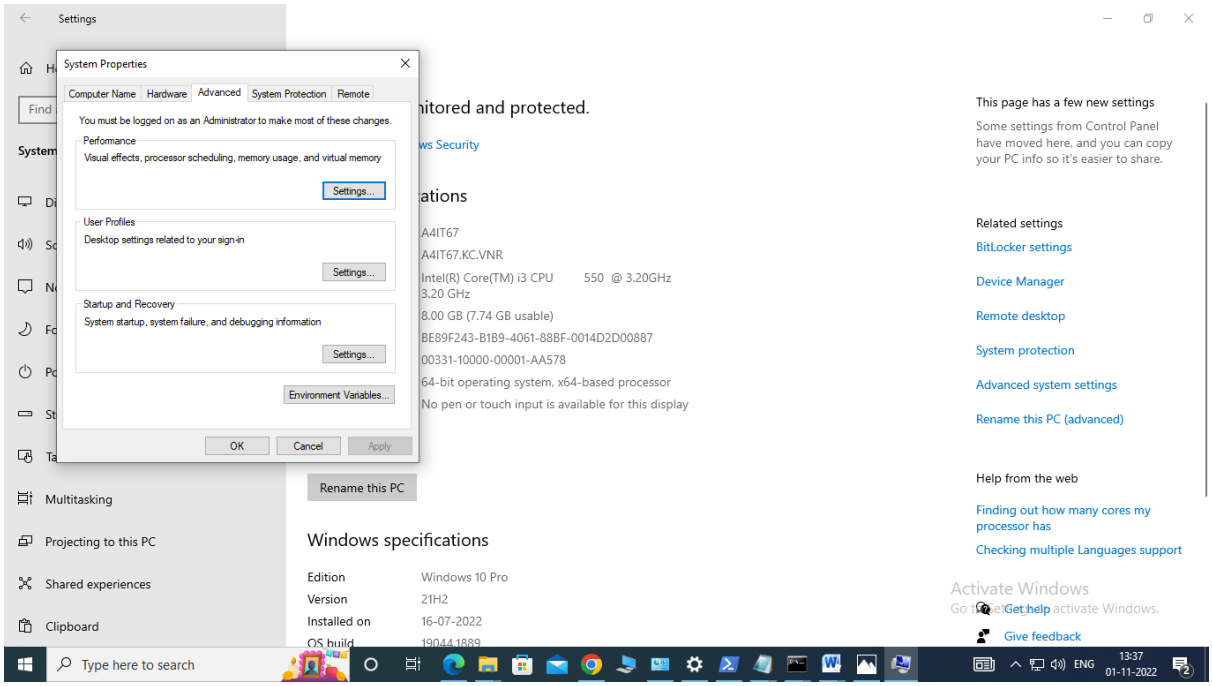
Step 3: Set the path for Java and Hadoop in system environment variables

To edit environment variables, go to Control Panel → System and Security → System (or) right-click "This PC" → Properties (My Computer icon) and click on the "Advanced system settings" link.

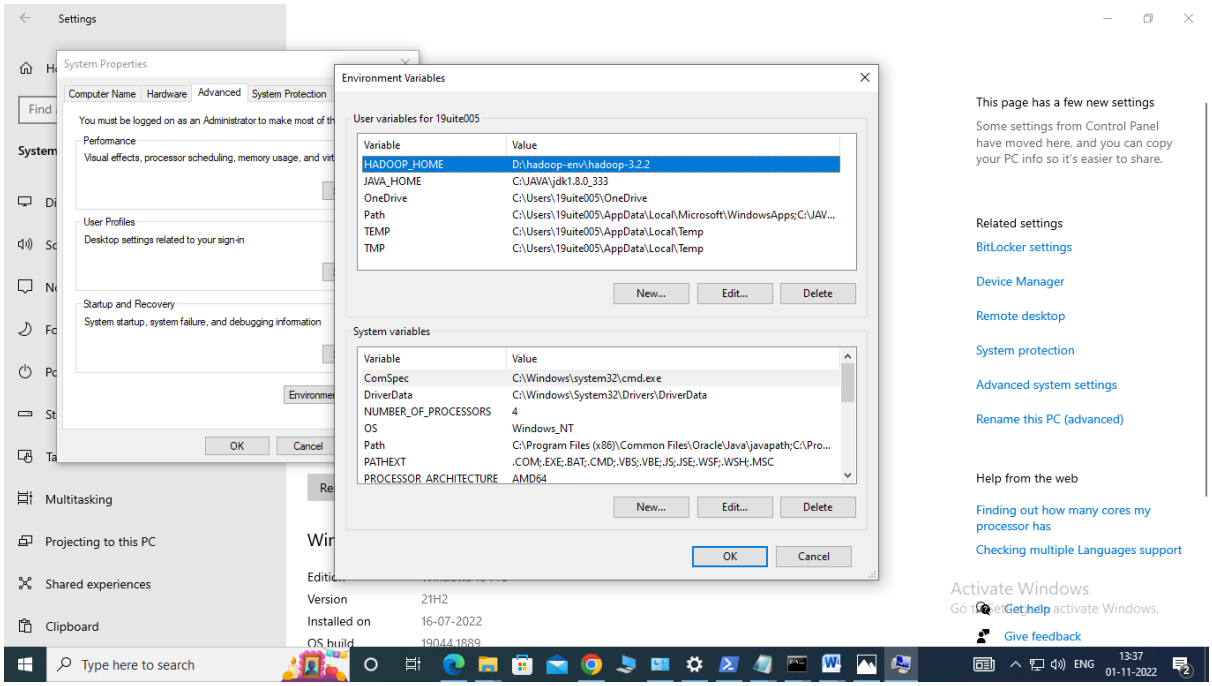


When the "Advanced system settings" dialog appears, go to the "Advanced" tab and click on the "Environment variables" button located on the bottom of the dialog.

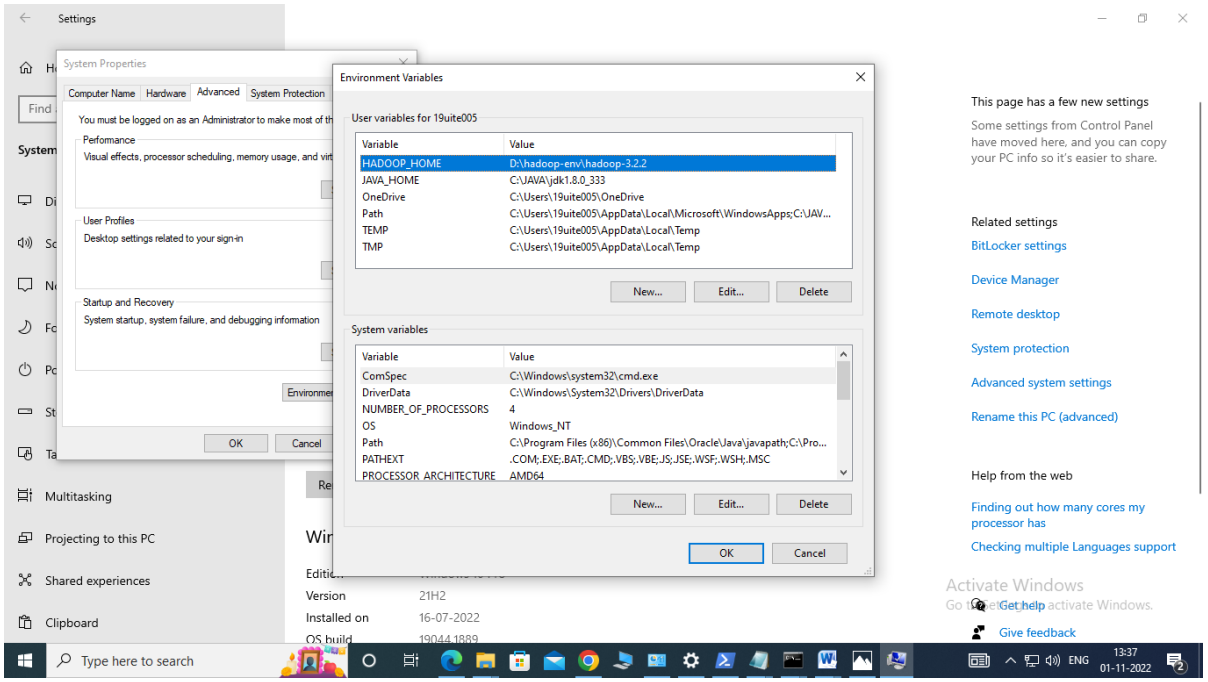


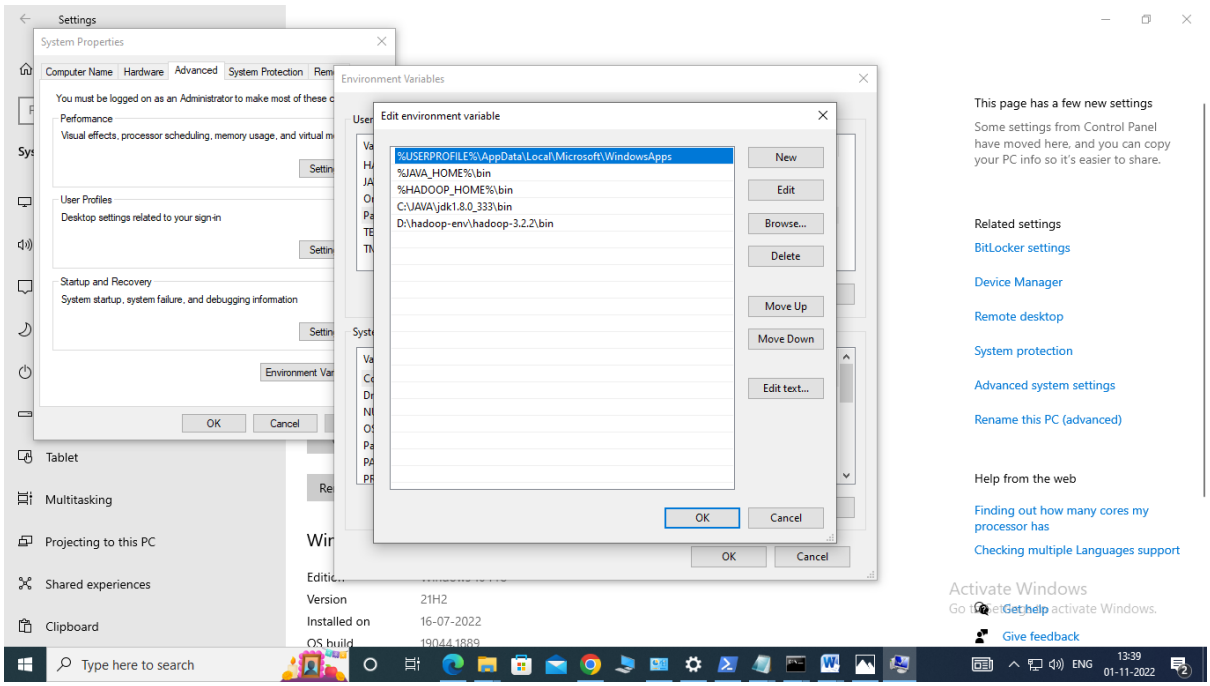


In the “Environment Variables” dialog, press the “New” button to add a new variable.



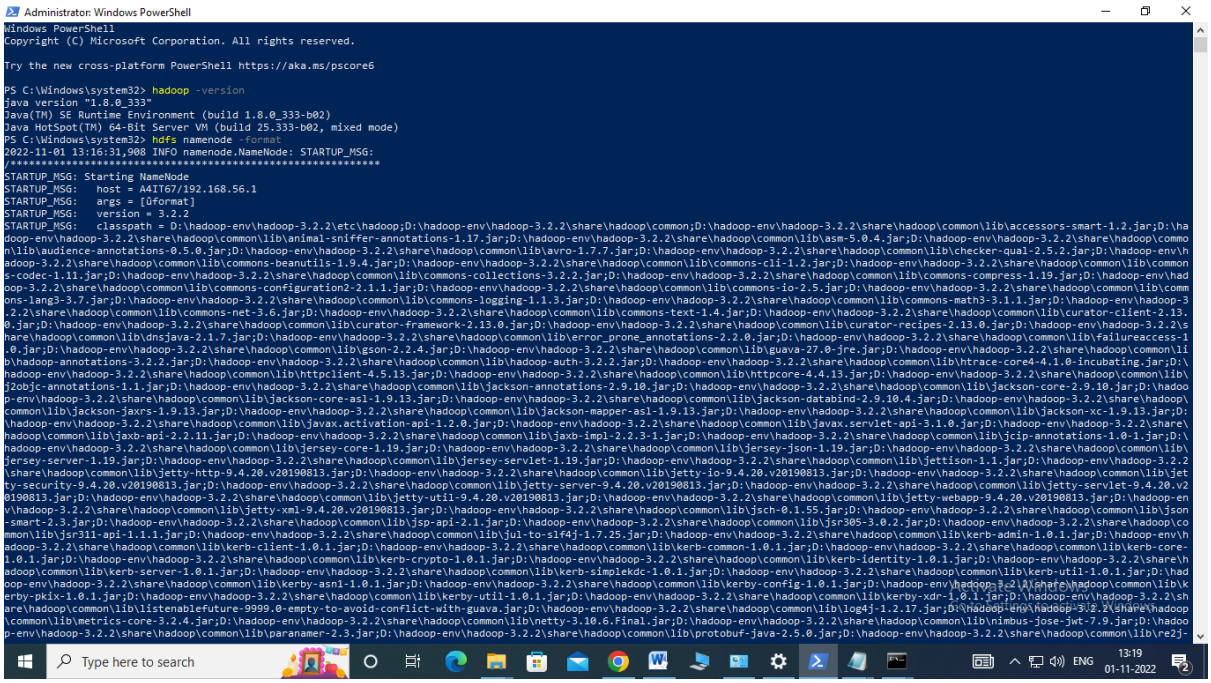
Now, Edit the PATH variable to add the Java and Hadoop binaries paths as shown in the following screenshots.





Step 4: Open Command Prompt as Administrator and run the following command:

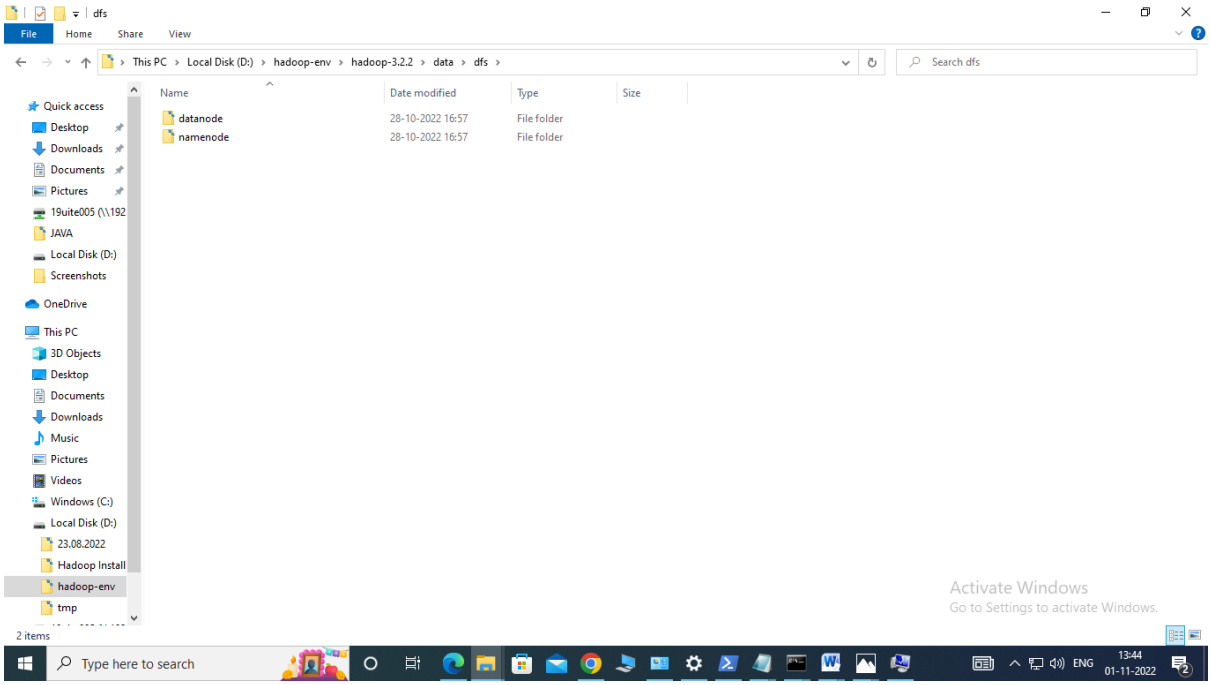
hadoop –version



Step 4: Create the following folders

D:\hadoop-env\hadoop-3.2.2**data\dfs\namenode**

D:\hadoop-env\hadoop-3.2.2**data\dfs\datanode**



Step 5: Configuring Hadoop cluster

There are **four files** to configure Hadoop cluster:

Location of the File: **"D:\hadoop-env\hadoop-3.2.2\etc\hadoop"**

File Name: "hdfs-site.xml"

Add the following properties within the <configuration></configuration> element:

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

*<value>**file:///D:/hadoop-env/hadoop-3.2.2/data/dfs/namenode**</value>*

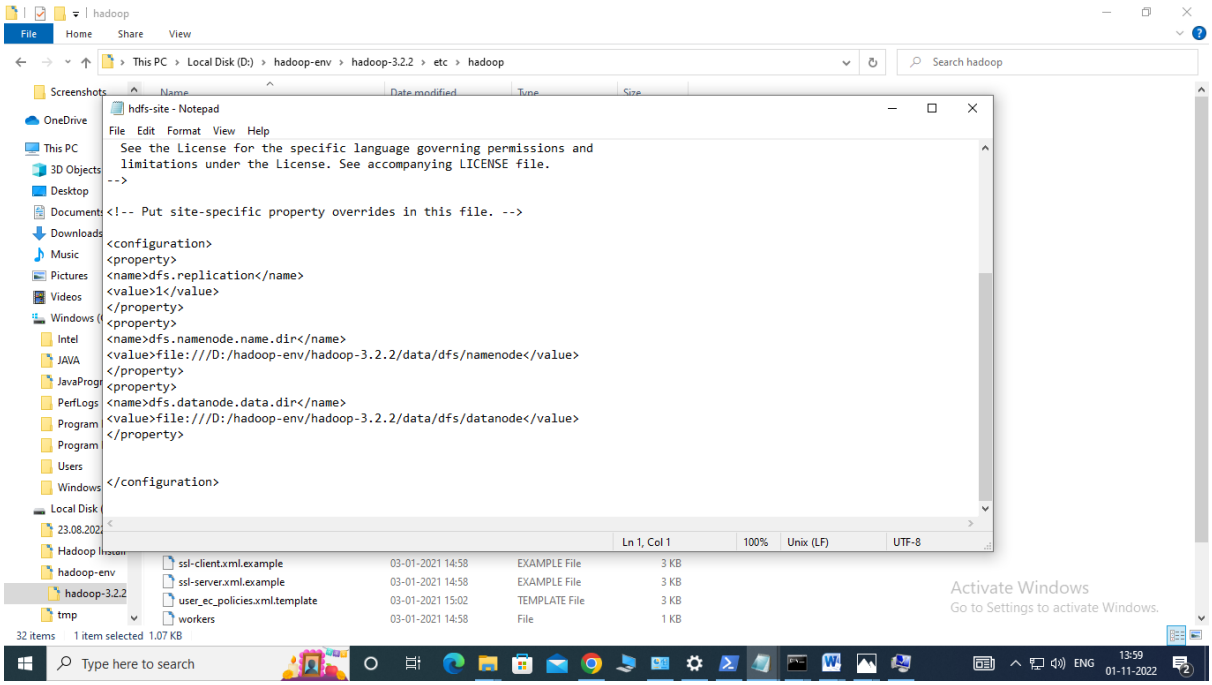
</property>

<property>

<name>dfs.datanode.data.dir</name>

*<value>**file:///D:/hadoop-env/hadoop-3.2.2/data/dfs/datanode**</value>*

</property>



File Name: “core-site.xml”

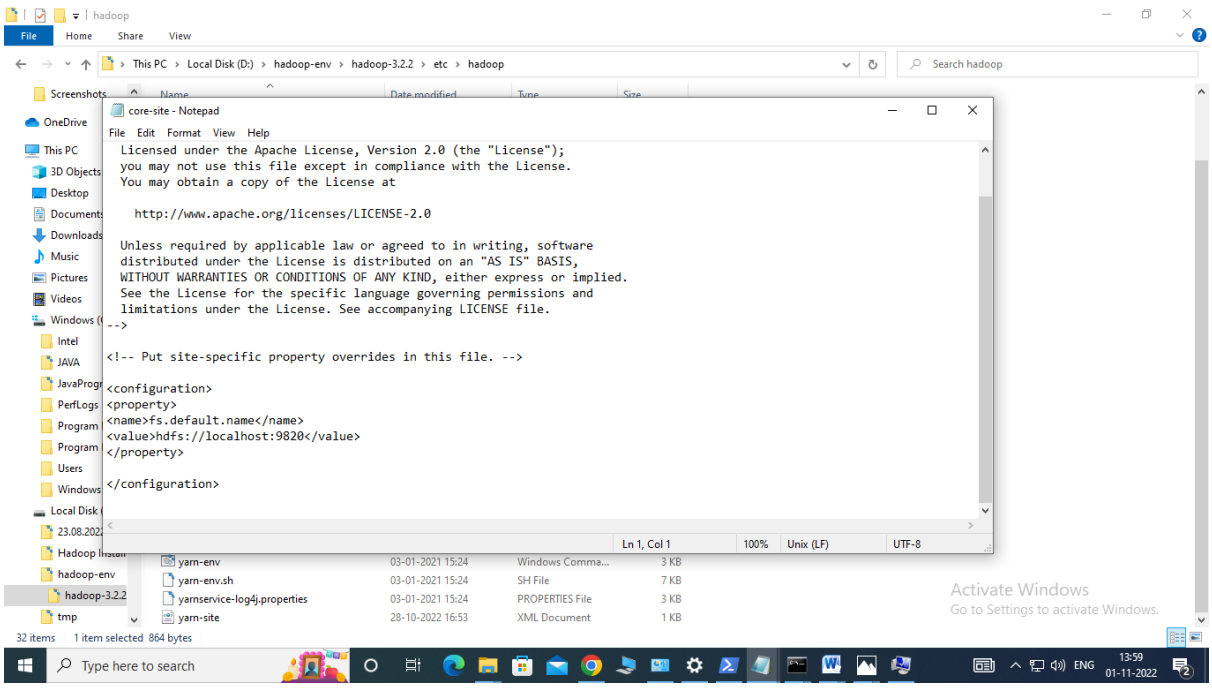
Add the following properties within the <configuration></configuration> element:

<property>

<name>fs.default.name</name>

*<value>**hdfs://localhost:9820***</value>*

</property>



File Name: “mapred-site.xml”

Add the following properties within the <configuration></configuration> element:

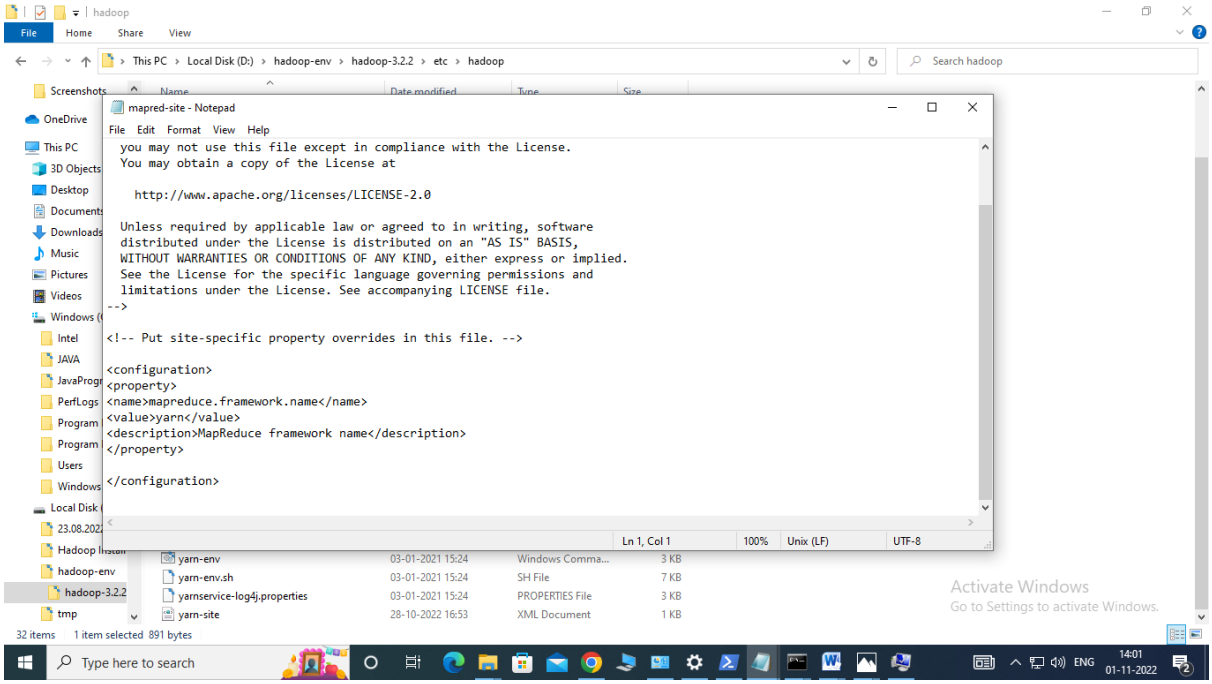
<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

<description>MapReduce framework name</description>

</property>



File Name: "yarn-site.xml"

Add the following properties within the <configuration></configuration> element:

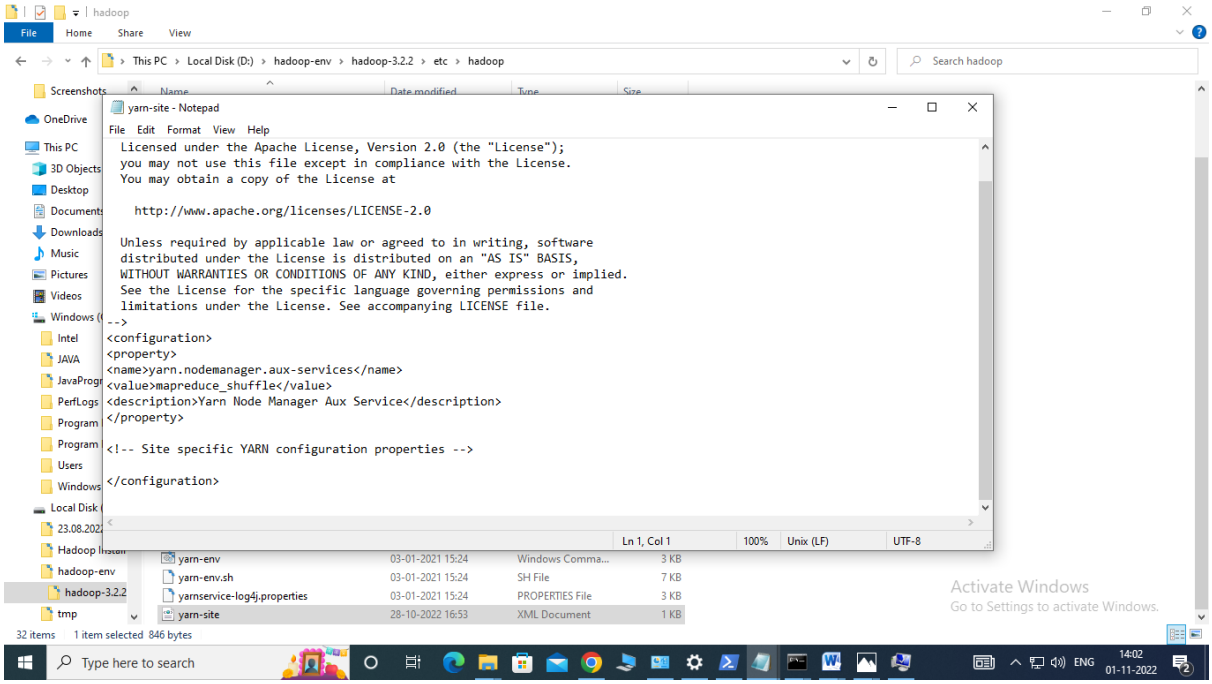
<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce_shuffle</value>

<description>Yarn Node Manager Aux Service</description>

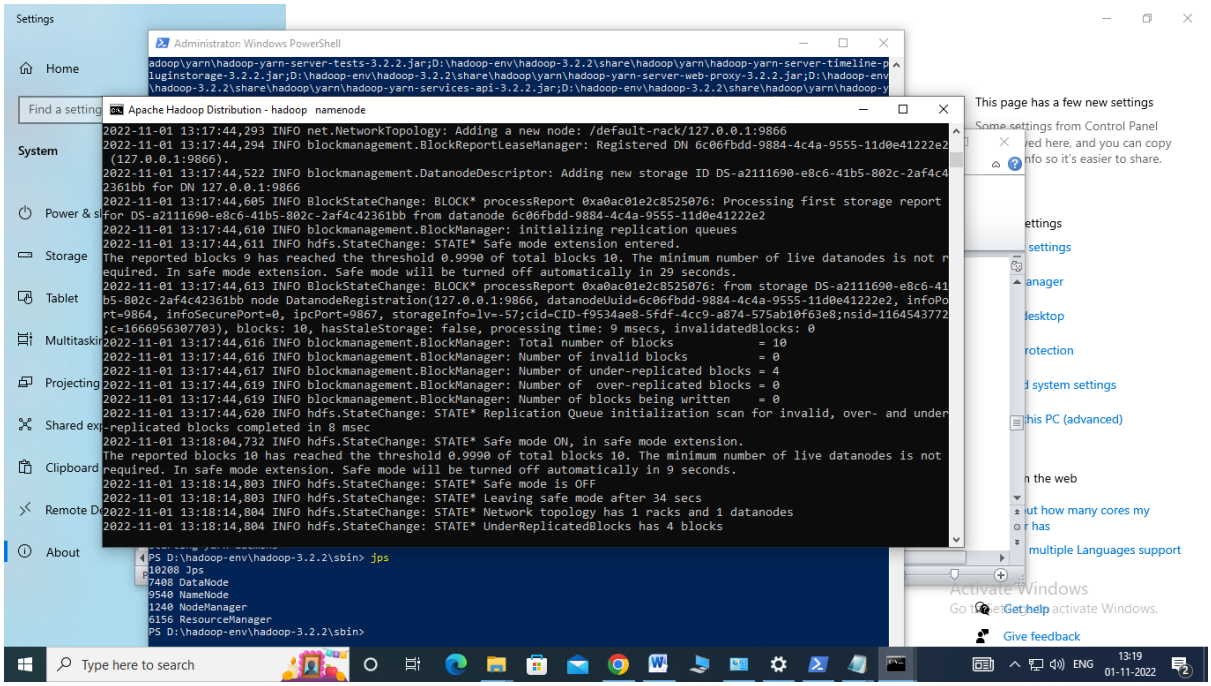
</property>



Step 6: Format the namenode

After finishing the configuration, format the name using the following command

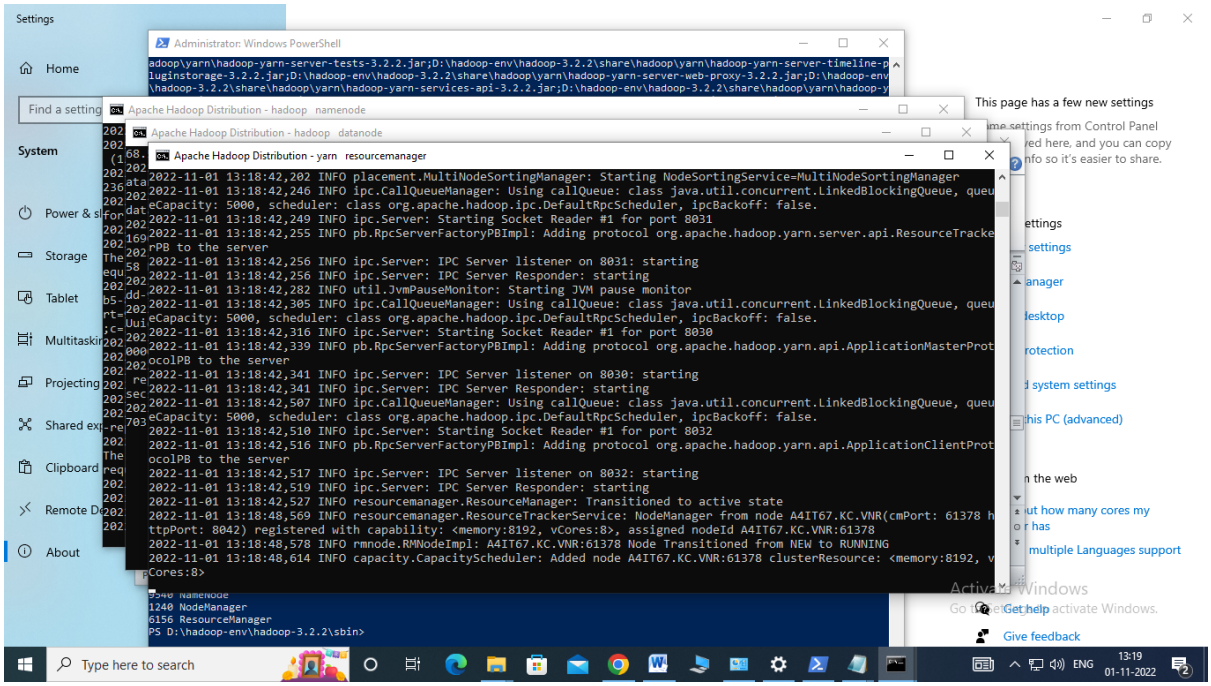
hdfs namenode –format



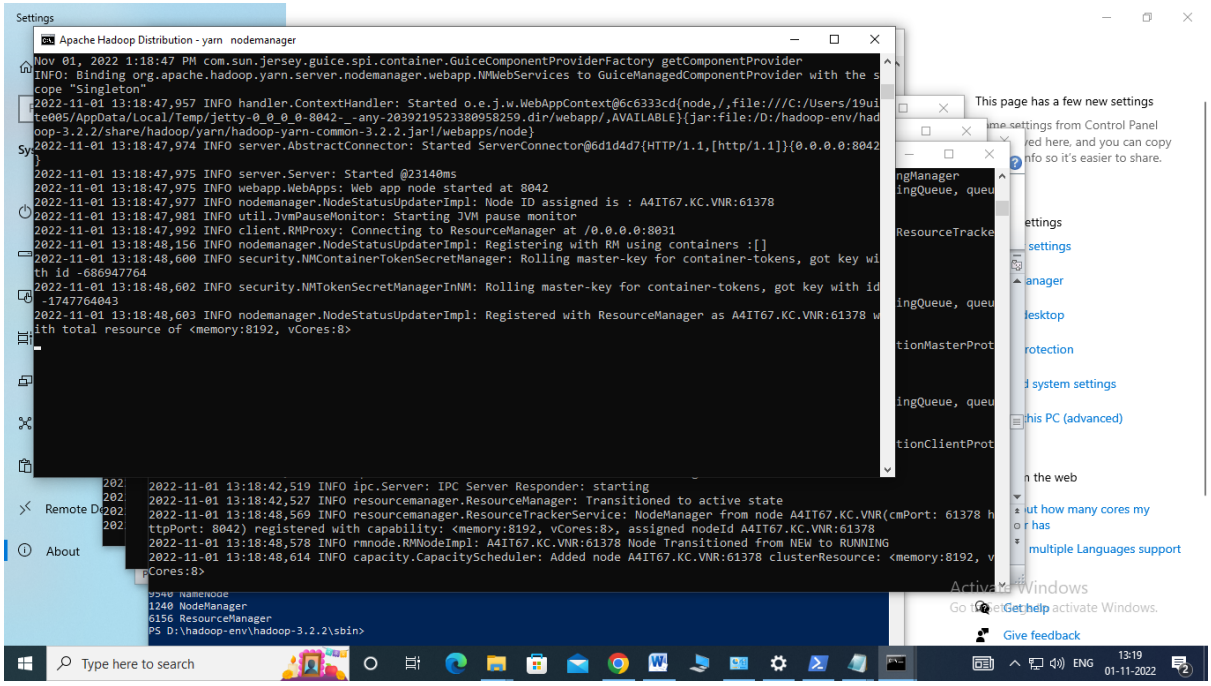
Step 7: Start the Hadoop nodes

Now navigate to the location “D:\hadoop-env\hadoop-3.2.2\sbin” in powershell and then run the following command to start the Hadoop nodes:

.\start-dfs.cmd

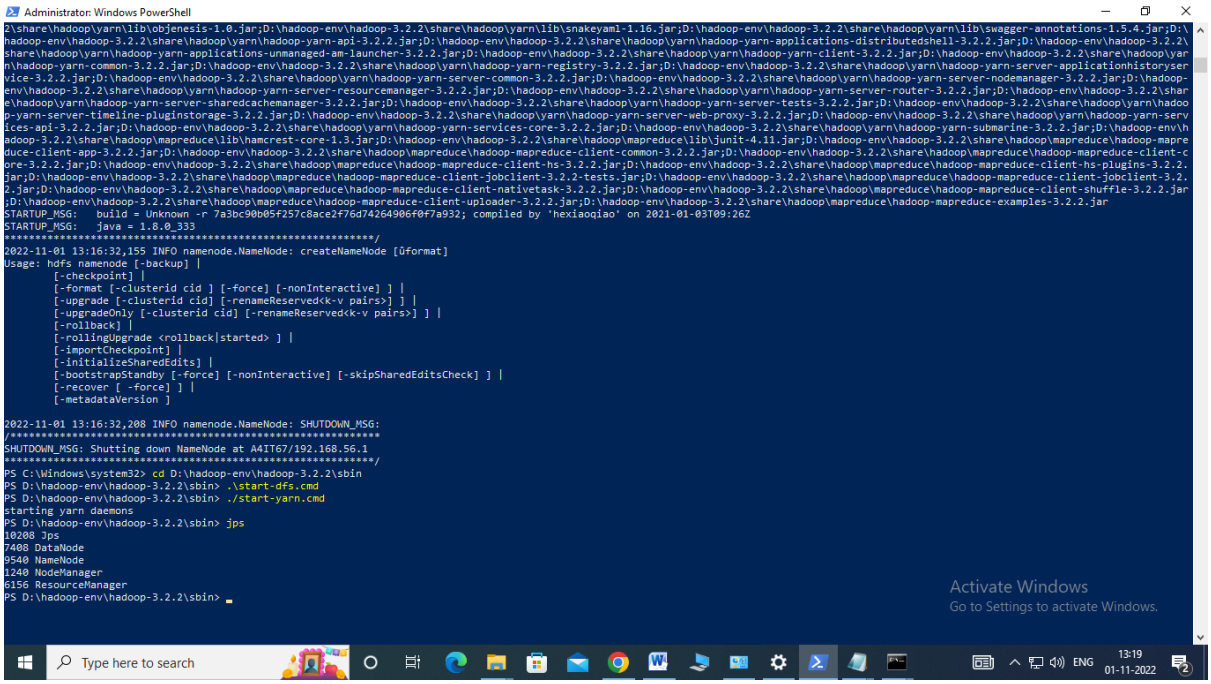


Two command prompt windows will open (one for the **namenode** and one for the **datanode**) as follows:



Next, Start the Hadoop Yarn service using the following command:

./start-yarn.cmd



To make sure that all services started successfully, Run the following command:

jps

It should display the following services:

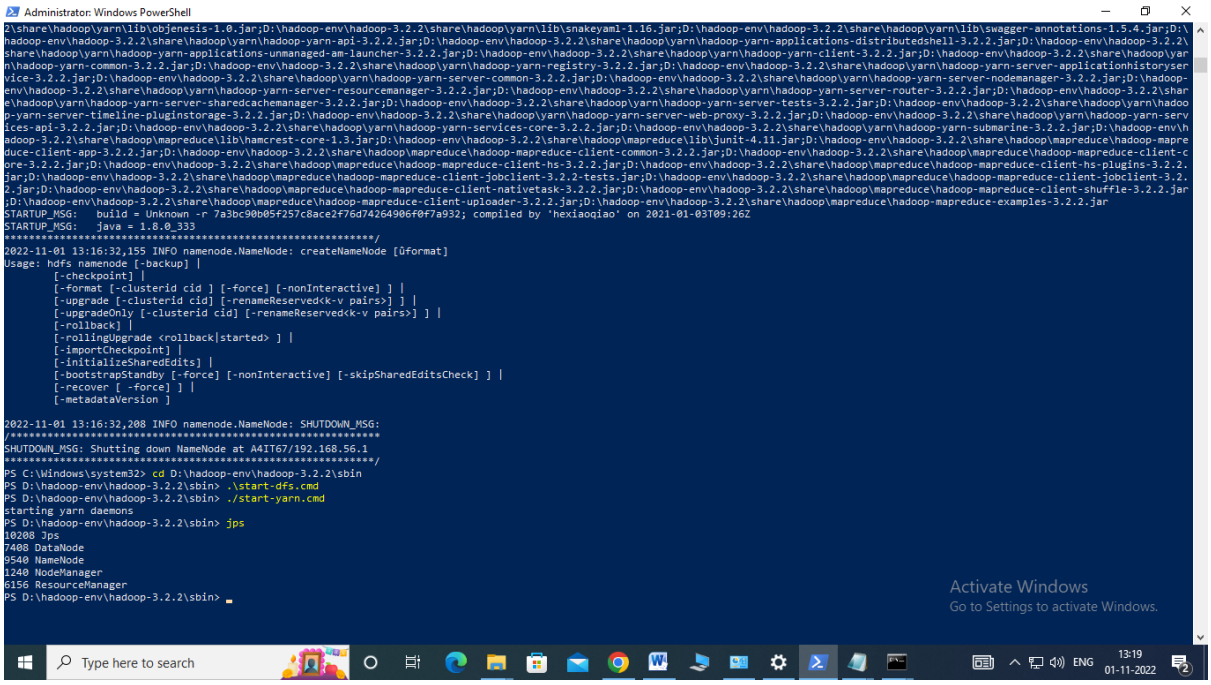
14560 DataNode

4960 ResourceManager

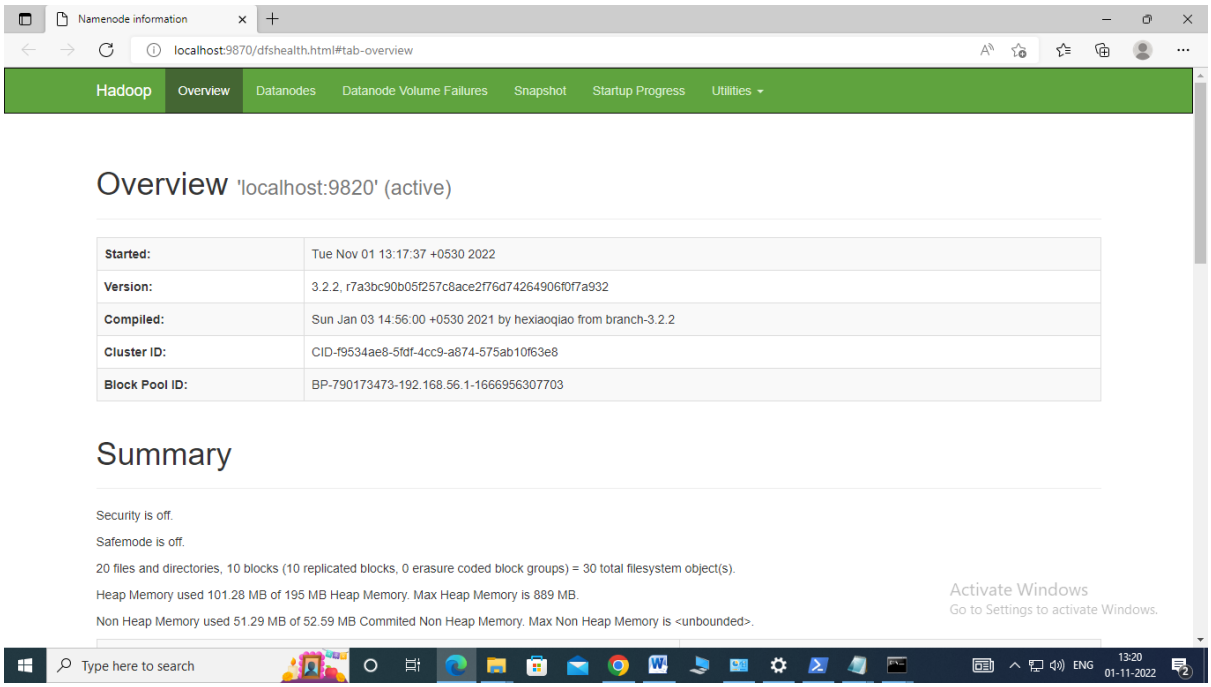
5936 NameNode

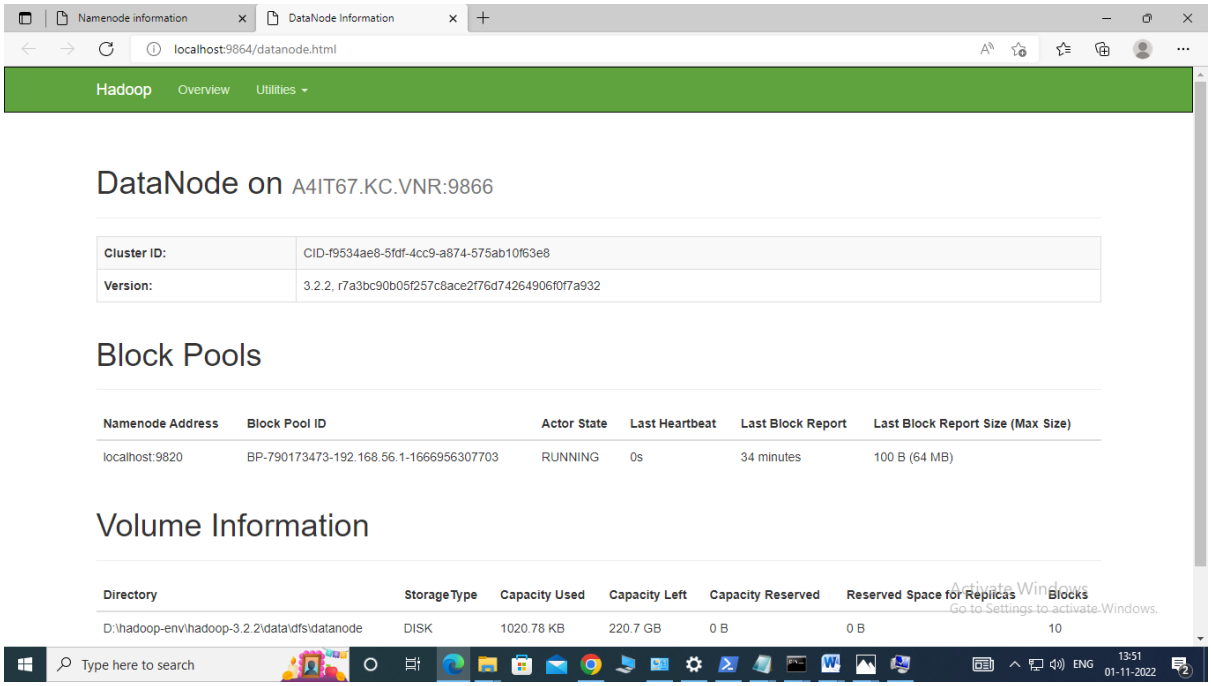
768 NodeManager

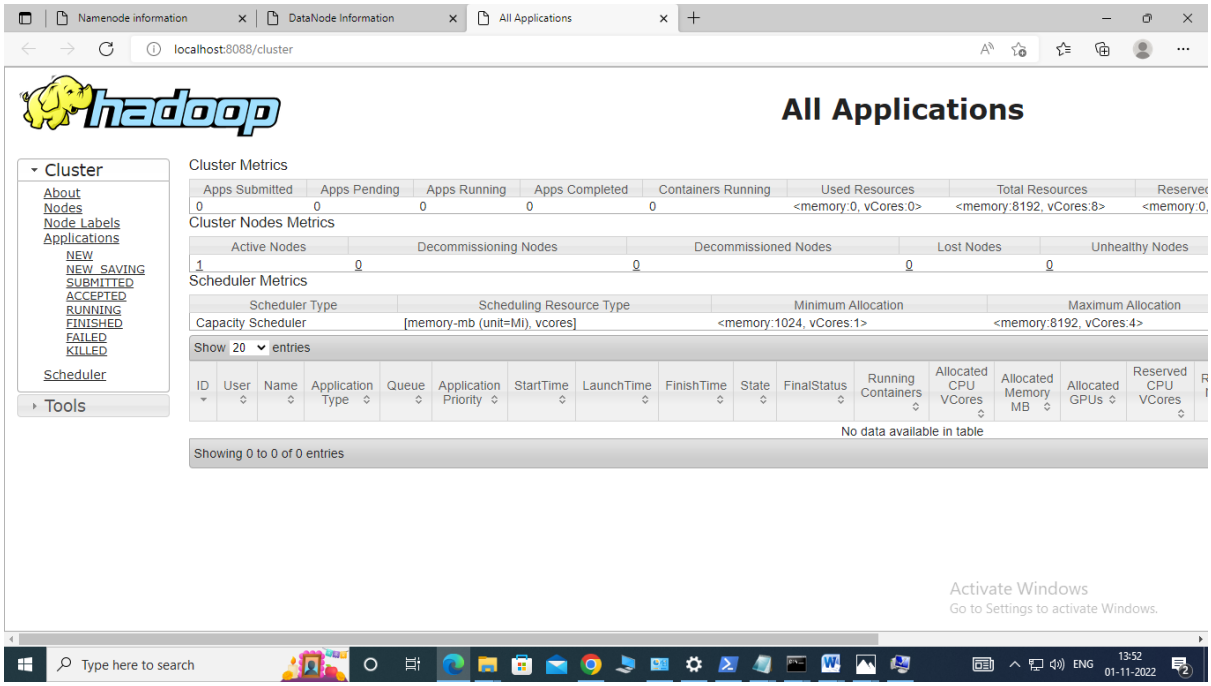
14636 Jps



OUTPUT


Namenode information
localhost:9870/dfshealth.html#tab-overview
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Overview 'localhost:9820' (active)
Started: Tue Nov 01 13:17:37 +0530 2022
Version: 3.2.2, r7a3bc90b05f257c8ace2f76d74264906f0f7a932
Compiled: Sun Jan 03 14:56:00 +0530 2021 by hexiaoqiao from branch-3.2.2
Cluster ID: CID-f9534ae8-5fdf-4cc9-a874-575ab10f63e8
Block Pool ID: BP-790173473-192.168.56.1-1666956307703
Summary
Security is off.
Safemode is off.
20 files and directories, 10 blocks (10 replicated blocks, 0 erasure coded block groups) = 30 total filesystem object(s).
Heap Memory used 101.28 MB of 195 MB Heap Memory. Max Heap Memory is 889 MB.
Non Heap Memory used 51.29 MB of 52.59 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.
Activate Windows
Go to Settings to activate Windows.
Type here to search
ENG
13:20
01-11-2022


Namenode information
DataNode Information
localhost:9864/datanode.html
Hadoop
Overview
Utilities
DataNode on A4IT67.KC.VNR:9866
Cluster ID: CID-f9534ae8-5fdf-4cc9-a874-575ab10f63e8
Version: 3.2.2, r7a3bc90b05f257c8ace2f76d74264906f0f7a932
Block Pools
Namenode Address
Block Pool ID
Actor State
Last Heartbeat
Last Block Report
Last Block Report Size (Max Size)
localhost:9820
BP-790173473-192.168.56.1-1666956307703
RUNNING
0s
34 minutes
100 B (64 MB)
Volume Information
Directory
StorageType
Capacity Used
Capacity Left
Capacity Reserved
Reserved Space for Replicas
Blocks
D:\hadoop-env\hadoop-3.2.2\data\dfs\datanode
DISK
1020.78 KB
220.7 GB
0 B
0 B
10
Activate Windows
Go to Settings to activate Windows.
Type here to search
ENG
13:51
01-11-2022


Namenode information
DataNode Information
All Applications
localhost:8088/cluster
All Applications
Cluster
About
Nodes
Node Labels
Applications
NEW
NEW_SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED
Scheduler
Tools
Cluster Metrics
Apps Submitted
Apps Pending
Apps Running
Apps Completed
Containers Running
Used Resources
Total Resources
Reserved
0
0
0
0
0
<memory:0, vCores:0>
<memory:8192, vCores:8>
<memory:0,
Cluster Nodes Metrics
Active Nodes
Decommissioning Nodes
Decommissioned Nodes
Lost Nodes
Unhealthy Nodes
1
0
0
0
0
Scheduler Metrics
Scheduler Type
Scheduling Resource Type
Minimum Allocation
Maximum Allocation
Capacity Scheduler
[memory-mb (unit=Mi), vcores]
<memory:1024, vCores:1>
<memory:8192, vCores:4>
Show 20 entries
ID
User
Name
Application Type
Queue
Application Priority
StartTime
LaunchTime
FinishTime
State
FinalStatus
Running Containers
Allocated CPU VCores
Allocated Memory MB
Allocated GPUs
Reserved CPU VCores
No data available in table
Showing 0 to 0 of 0 entries
Activate Windows
Go to Settings to activate Windows.
Type here to search
ENG
13:52
01-11-2022