Name : R. Venkatesh Prabhu Roll No : 19UITE001

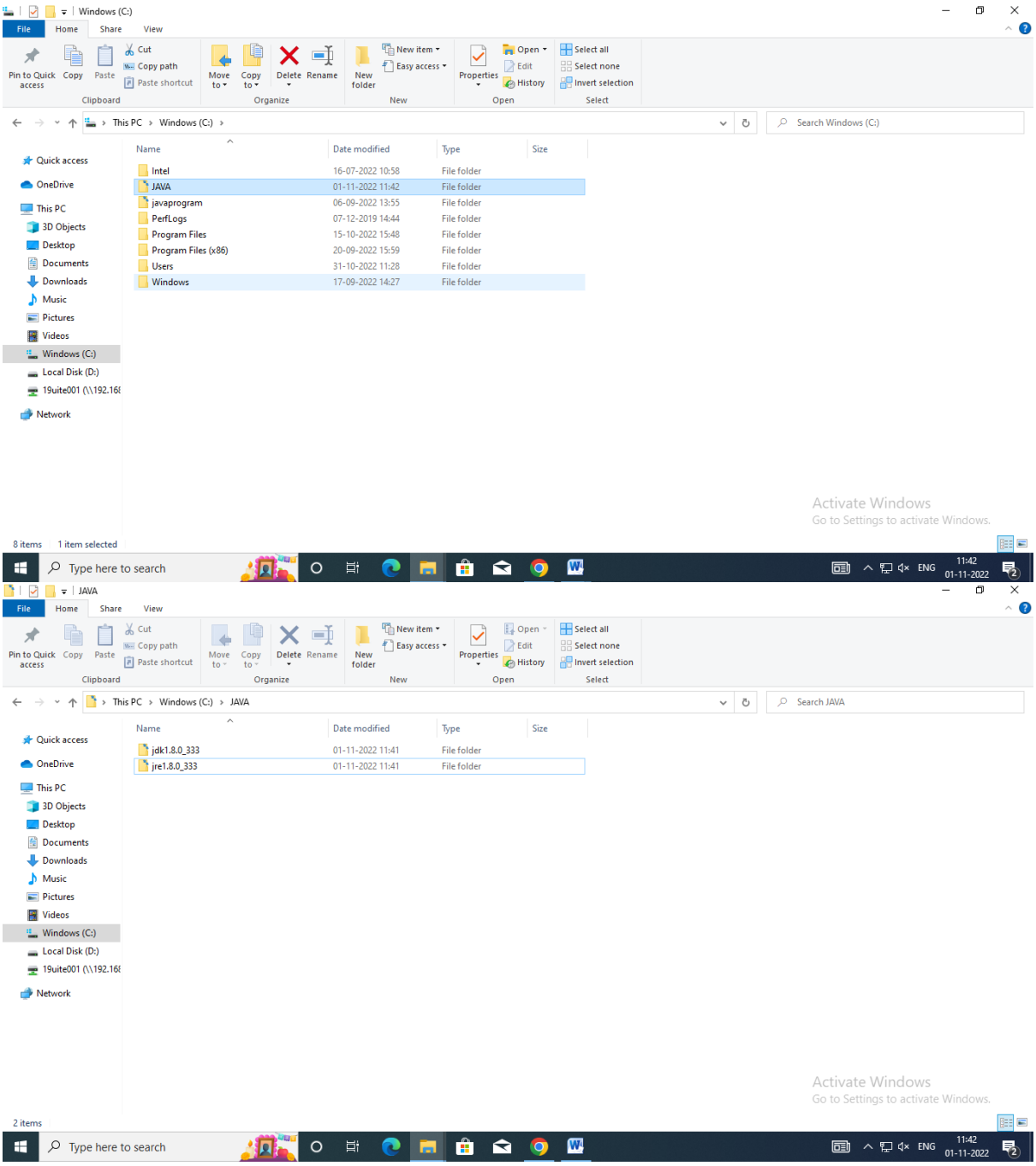
EXERCISE: 9

Installation of Single Node Hadoop Cluster

PROCEDURAL STEPS

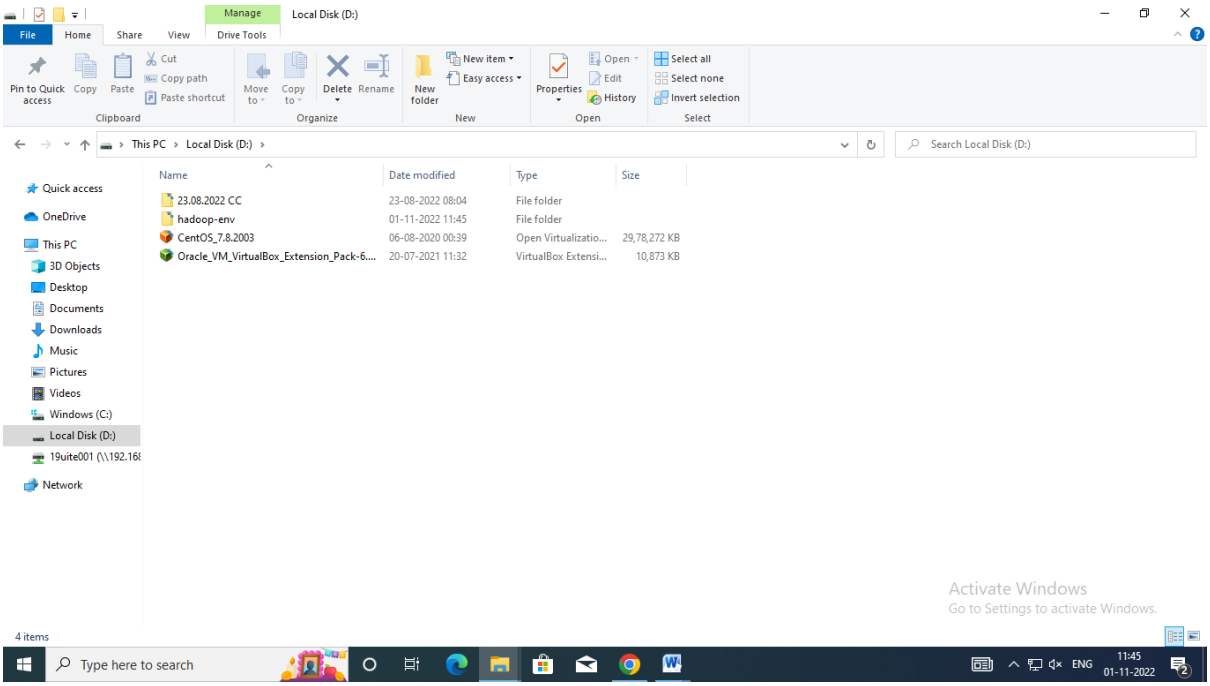
Step 1: Install / Extract the following Packages:

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



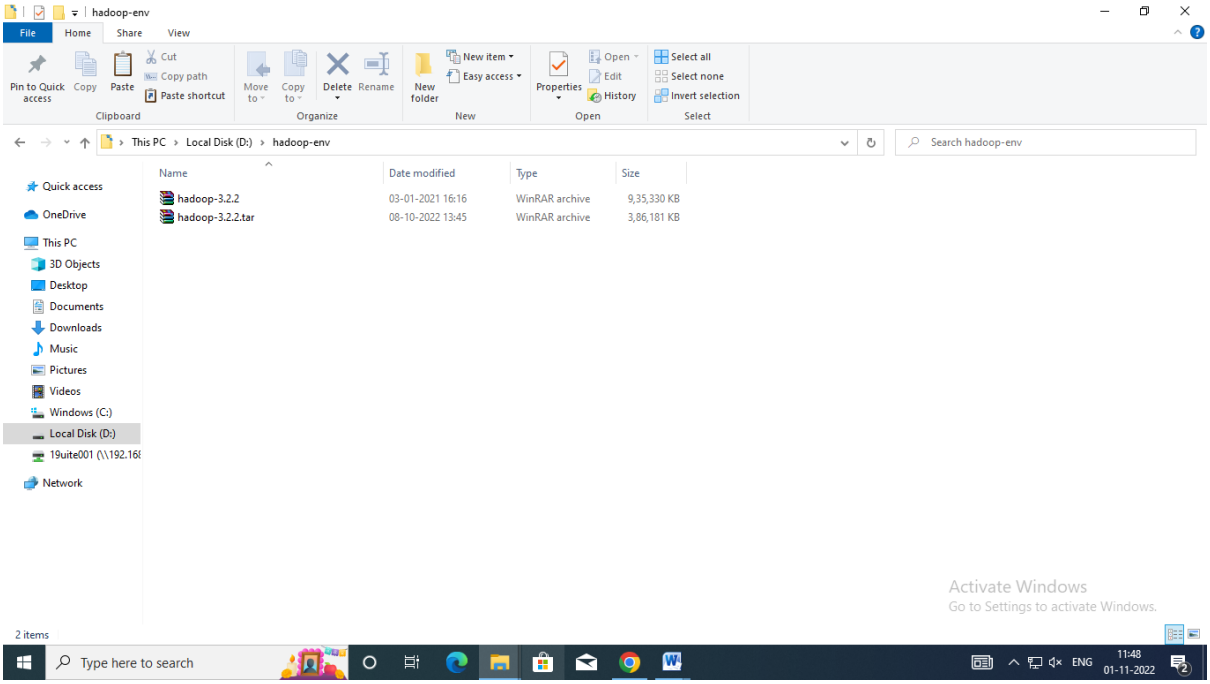
Name : R. Venkatesh Prabhu Roll No : 19UITE001

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

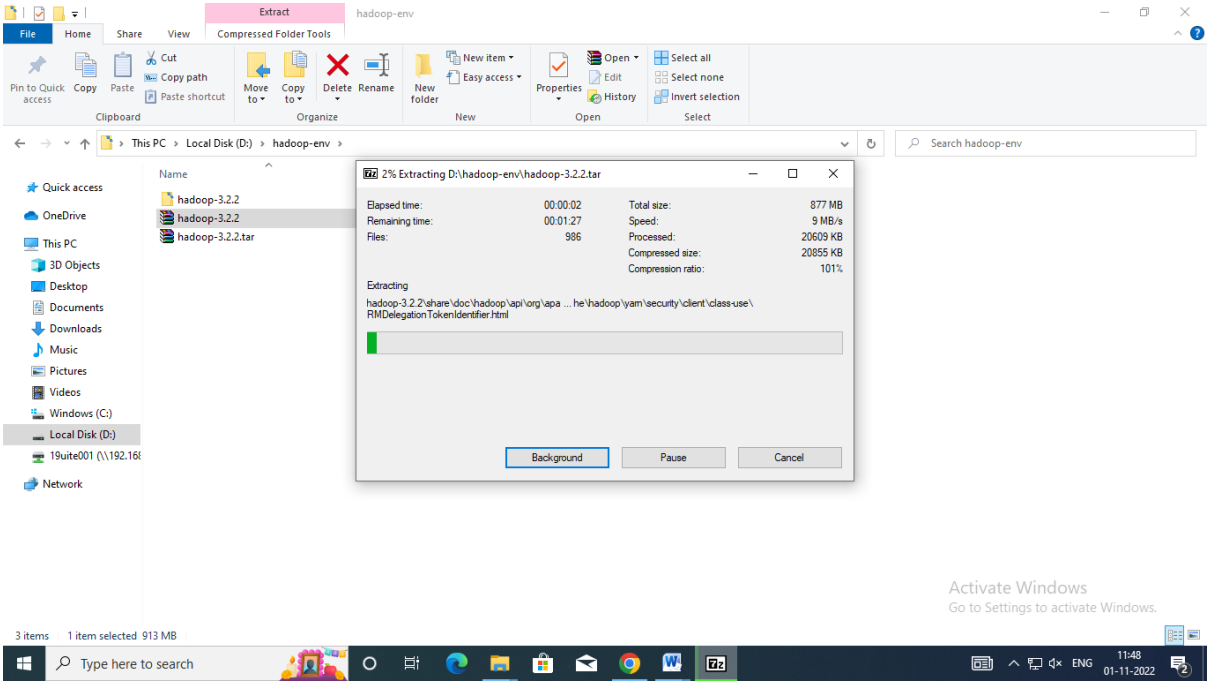


Name : R. Venkatesh Prabhu Roll No : 19UITE001

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

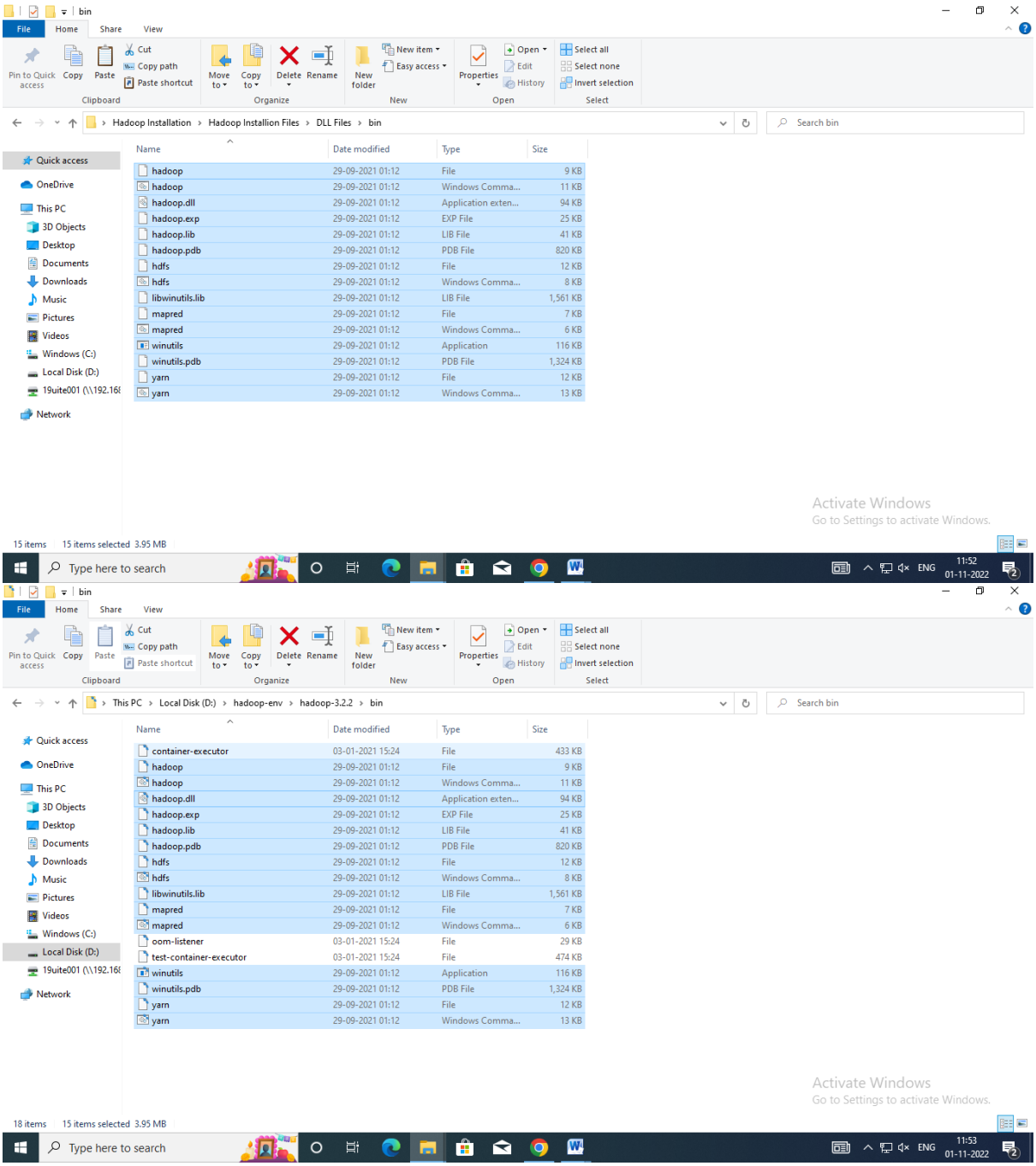


4. Unzip the “hadoop-3.2.2.tar.gz” using 7zip (do the unzip function two times)



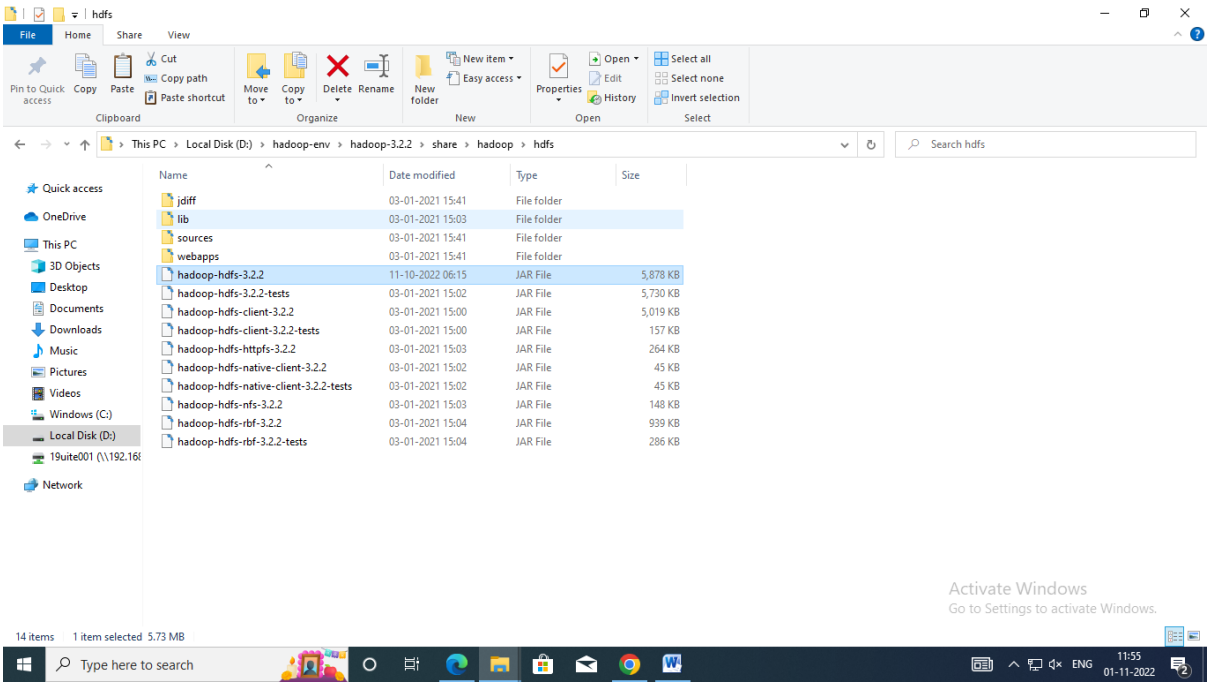
Name : R. Venkatesh Prabhu Roll No : 19UITE001

5. Copy the downloaded Hadoop dll files to the location “D:\hadoop-env\hadoop-3.2.2\bin”

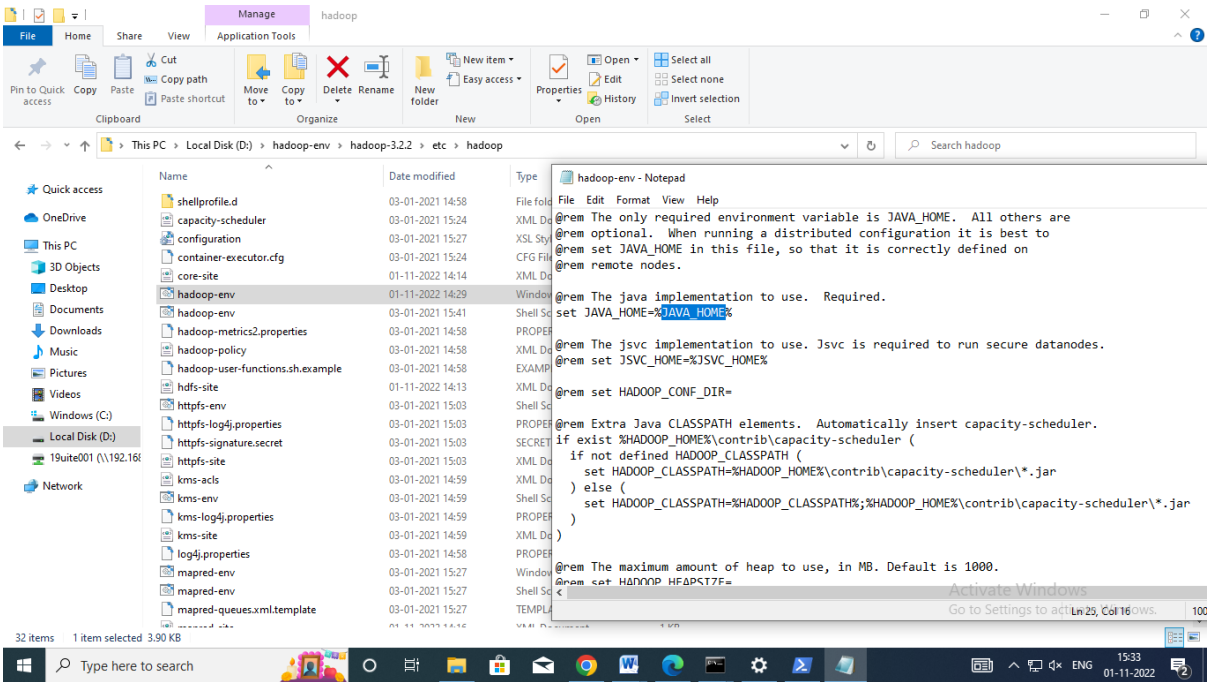


Name : R. Venkatesh Prabhu Roll No : 19UITE001

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**



Name : R. Venkatesh Prabhu Roll No : 19UITE001

Step 2: Set the path for Java and Hadoop in system environment variables

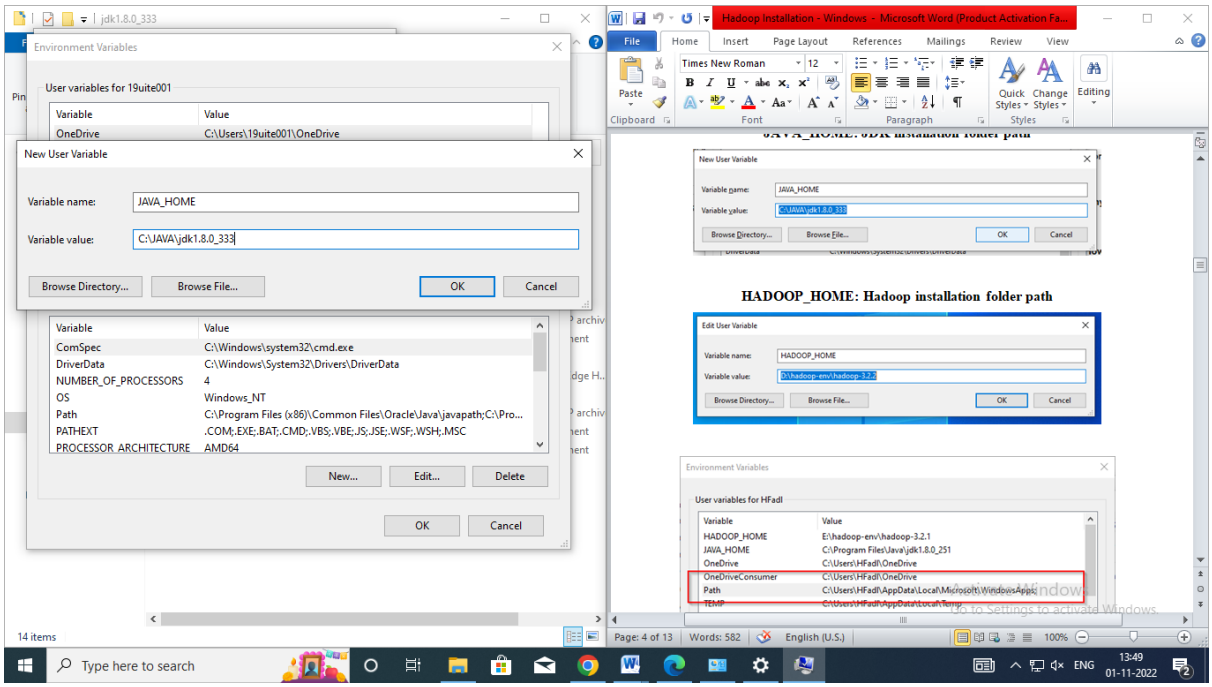
To edit environment variables, go to Control Panel → System and Security → System (or) right-click “This PC” → Properties (My Computer icon) and click on the “Advanced system settings” link.

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.

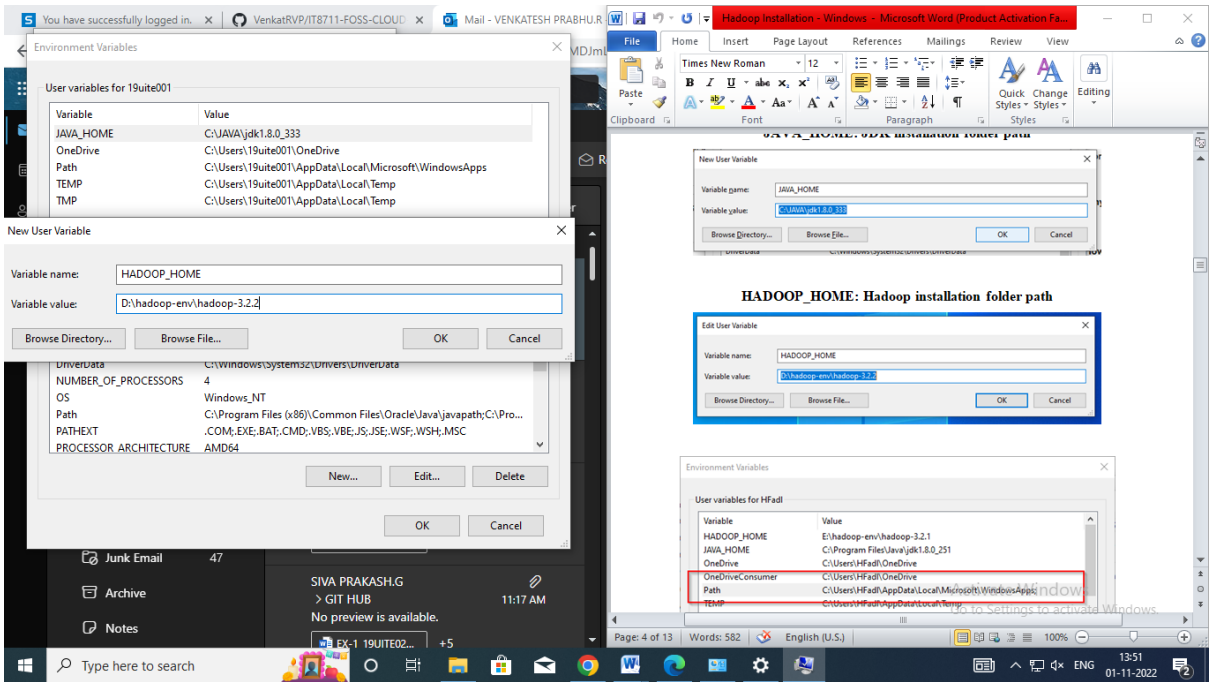
Now, Edit the PATH variable to add the Java and Hadoop binaries paths as shown in the following screenshots.

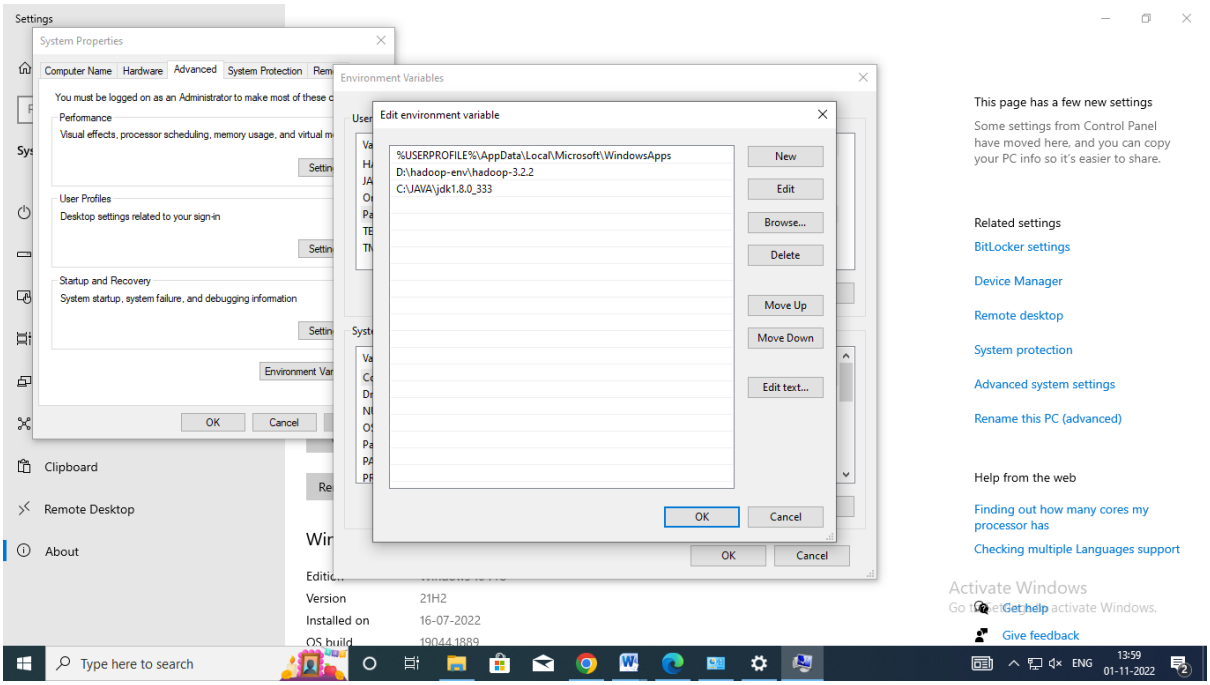
JAVA_HOME: JDK installation folder path



Name : R. Venkatesh Prabhu Roll No : 19UITE001

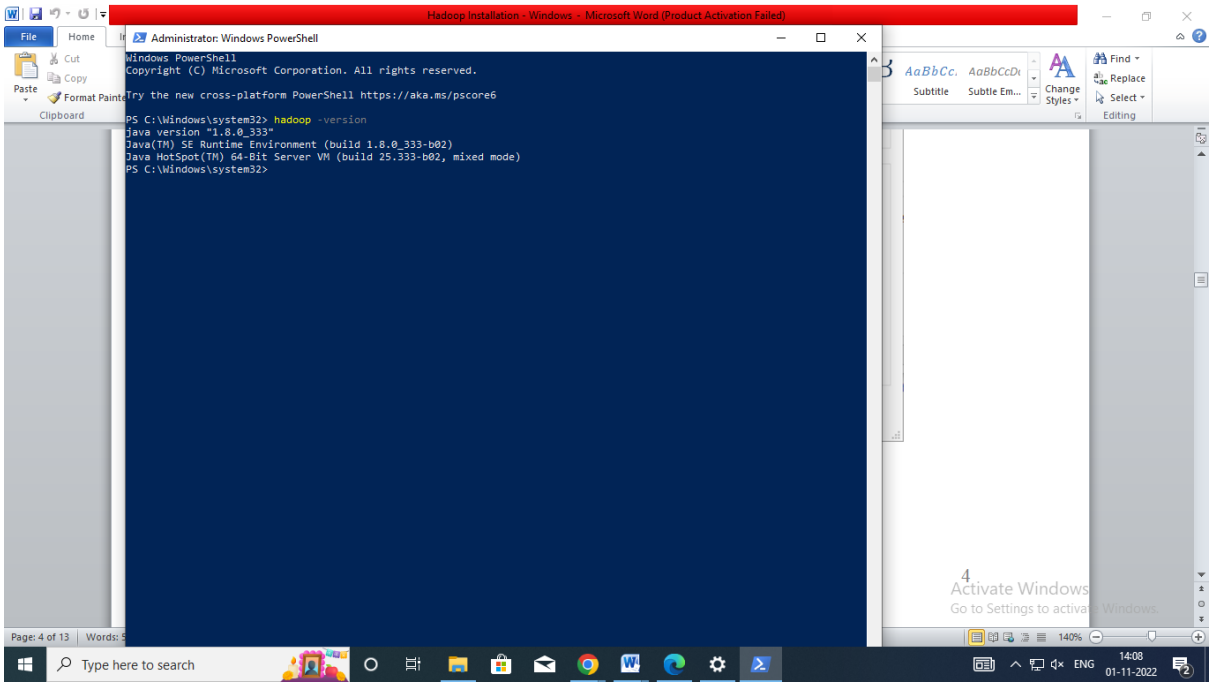
HADOOP_HOME: Hadoop installation folder path





Step 3: Open Command Prompt as Administrator and run the following command:

hadoop –version

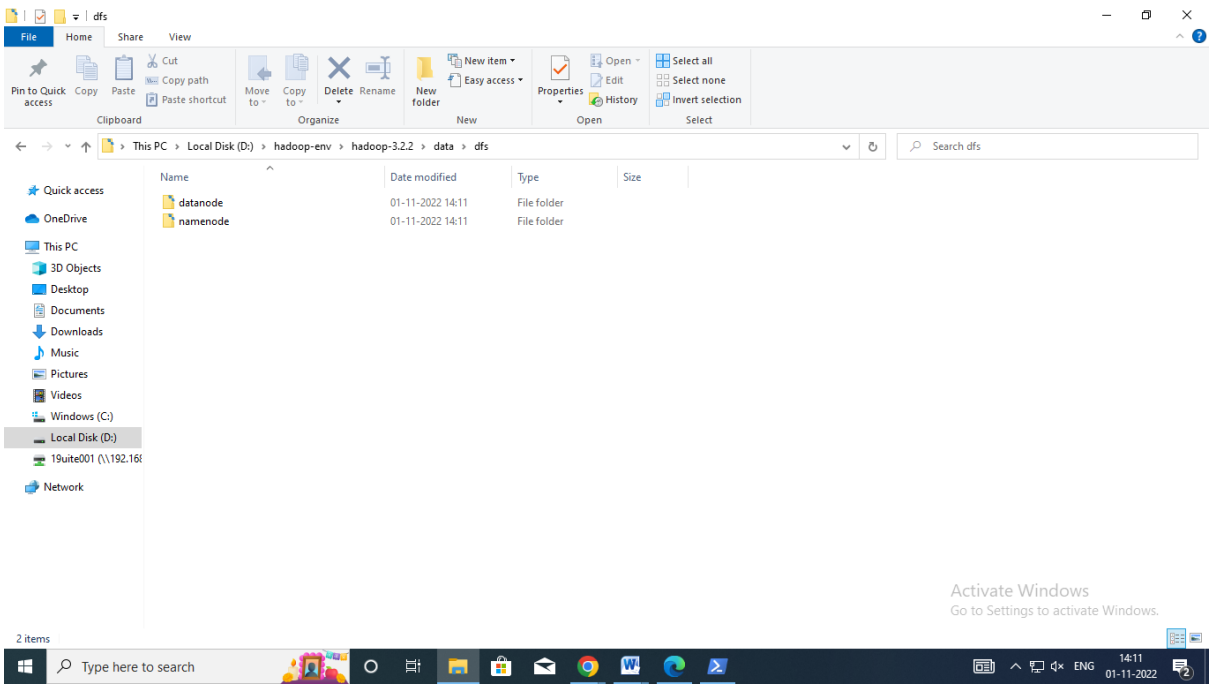


Name : R. Venkatesh Prabhu Roll No : 19UITE001

Step 4: Create the following folders

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

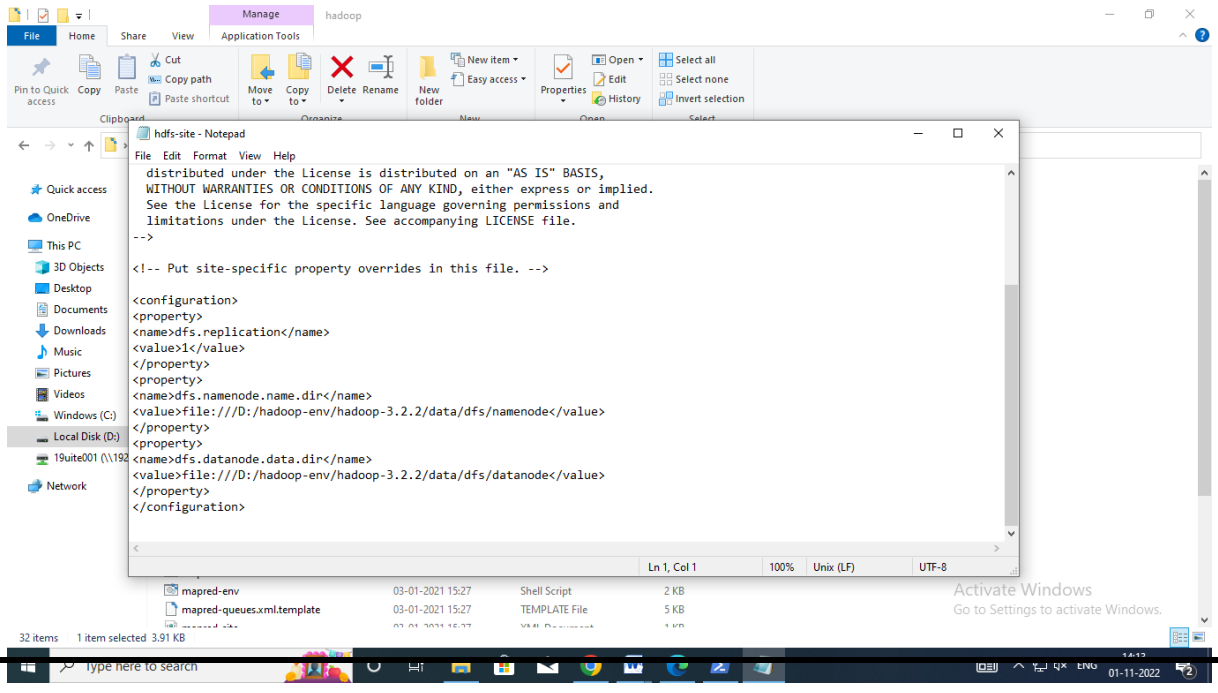
D:\hadoop-env\hadoop-3.2.2**data\dfs\datanode**



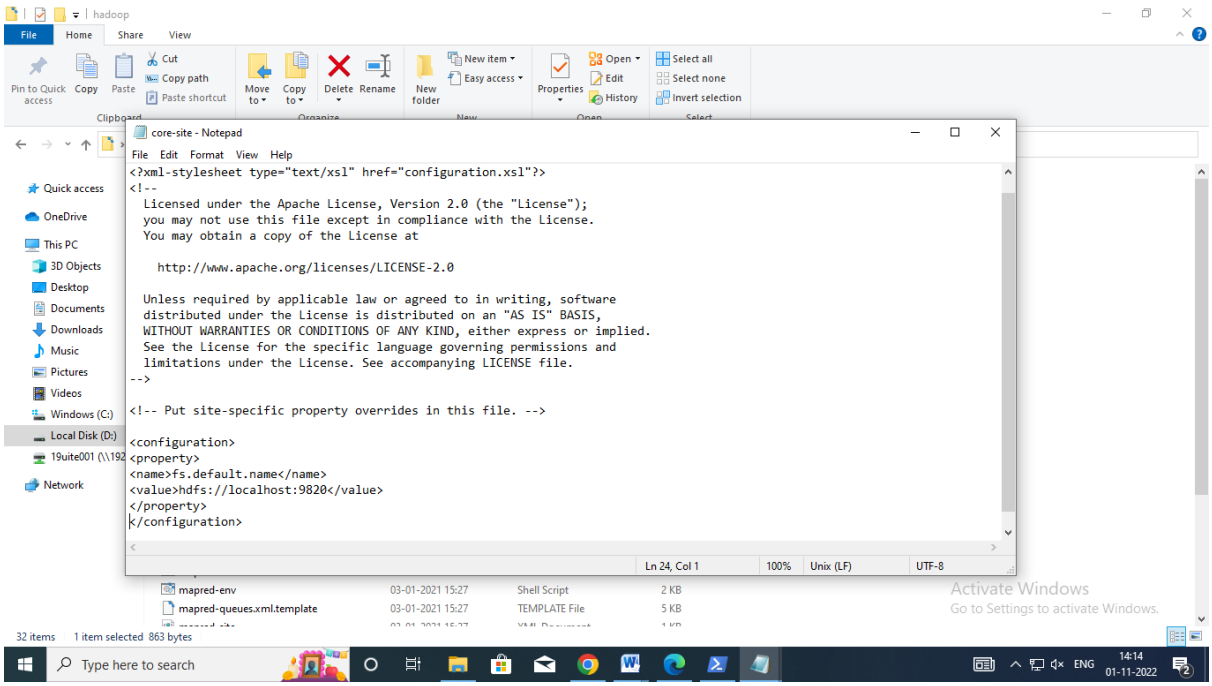
Step 5: Configuring Hadoop cluster

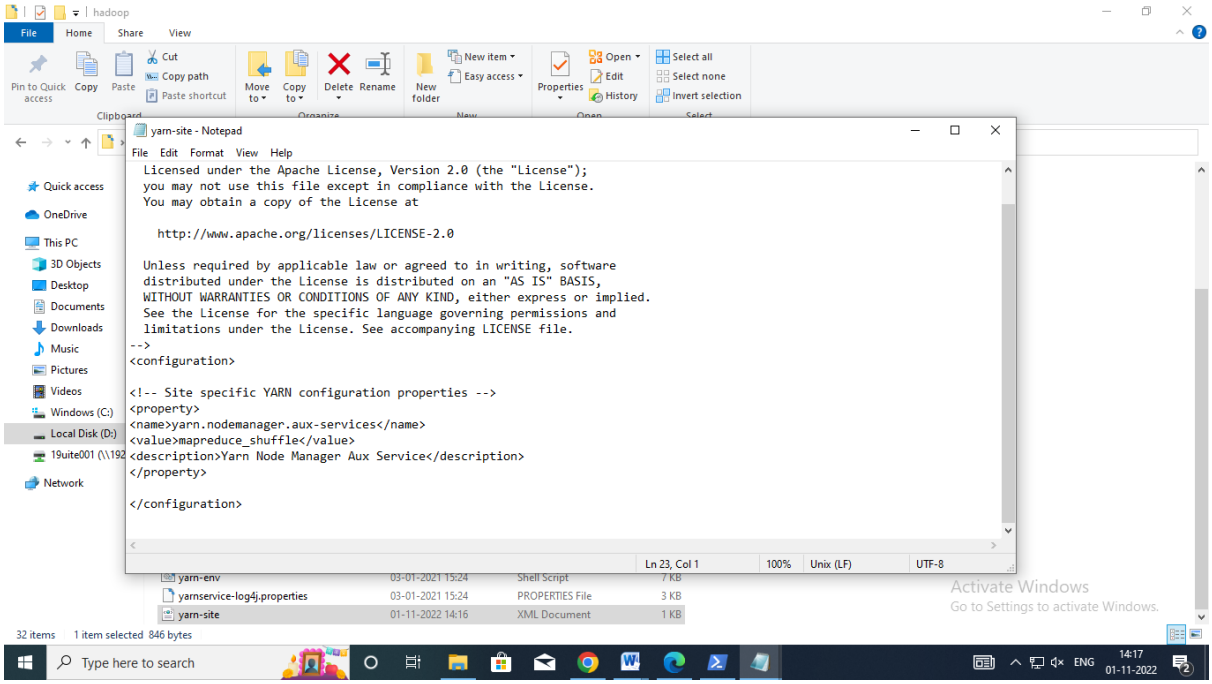
There are **four files** to configure Hadoop cluster:

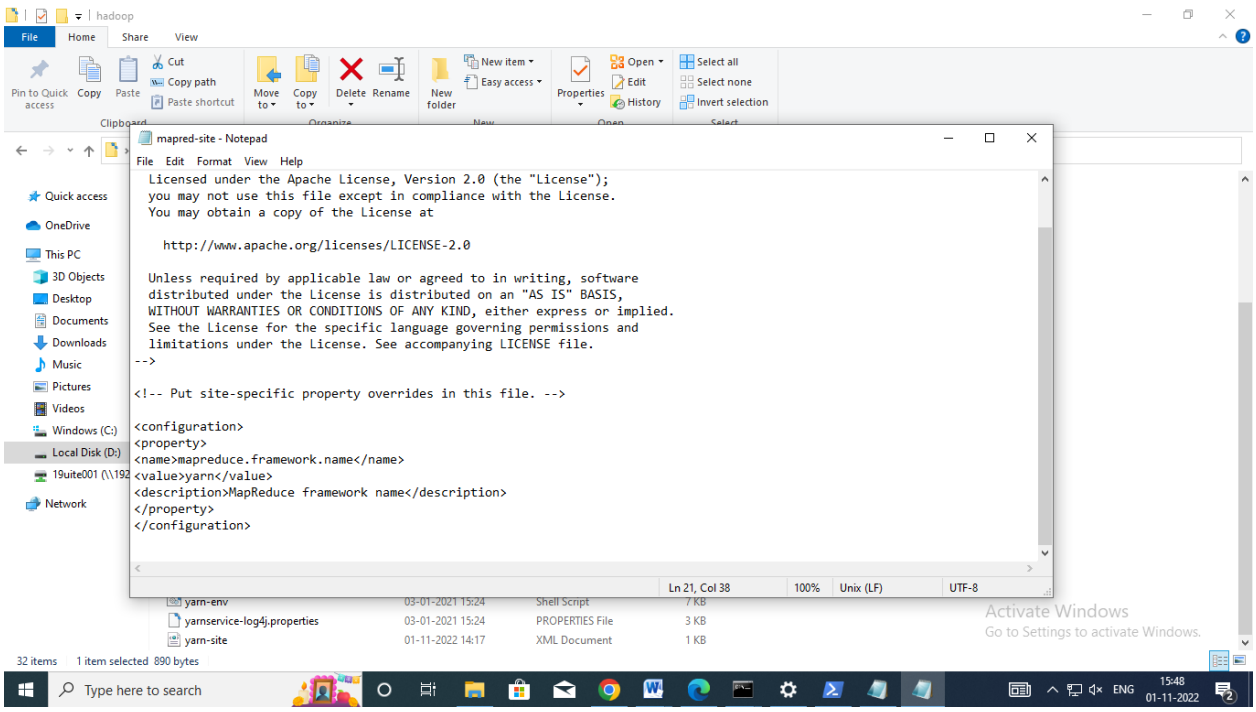
Location of the File: **"D:\hadoop-env\hadoop-3.2.2\etc\hadoop"**



Name : R. Venkatesh Prabhu Roll No : 19UITE001





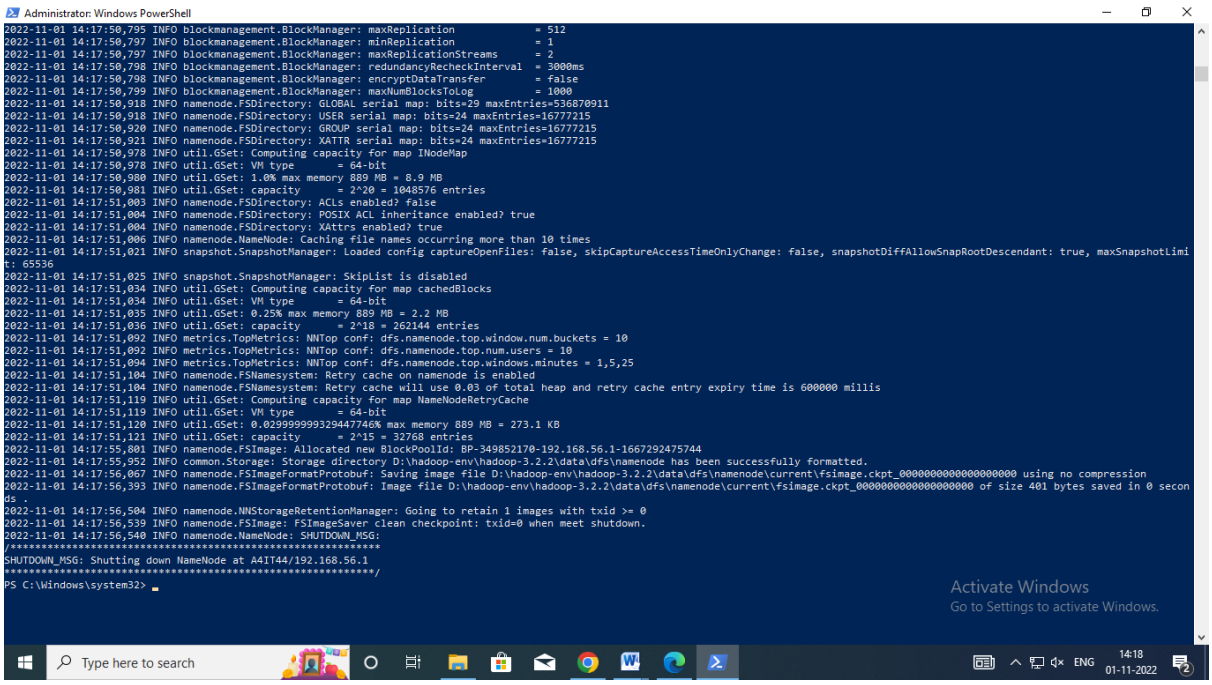


Step 6: Format the namenode

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

hdfs namenode –format

Name : R. Venkatesh Prabhu Roll No : 19UITE001

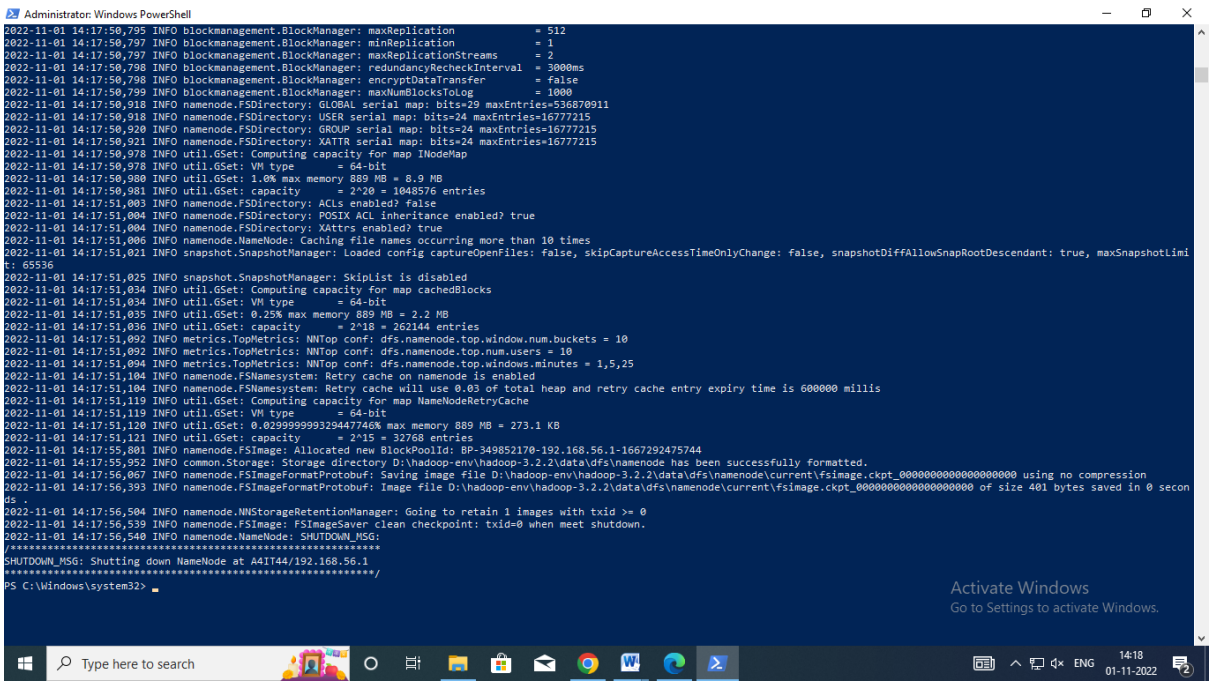


Name : R. Venkatesh Prabhu Roll No : 19UITE001

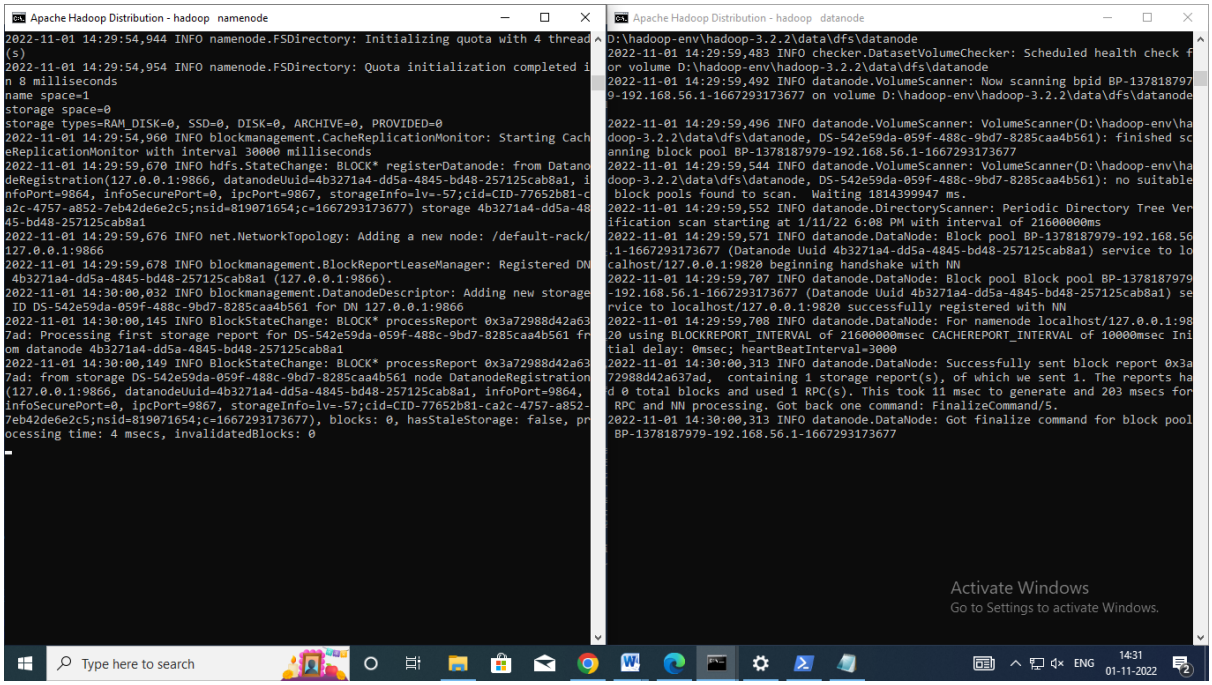
Step 7: Start the Hadoop nodes

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



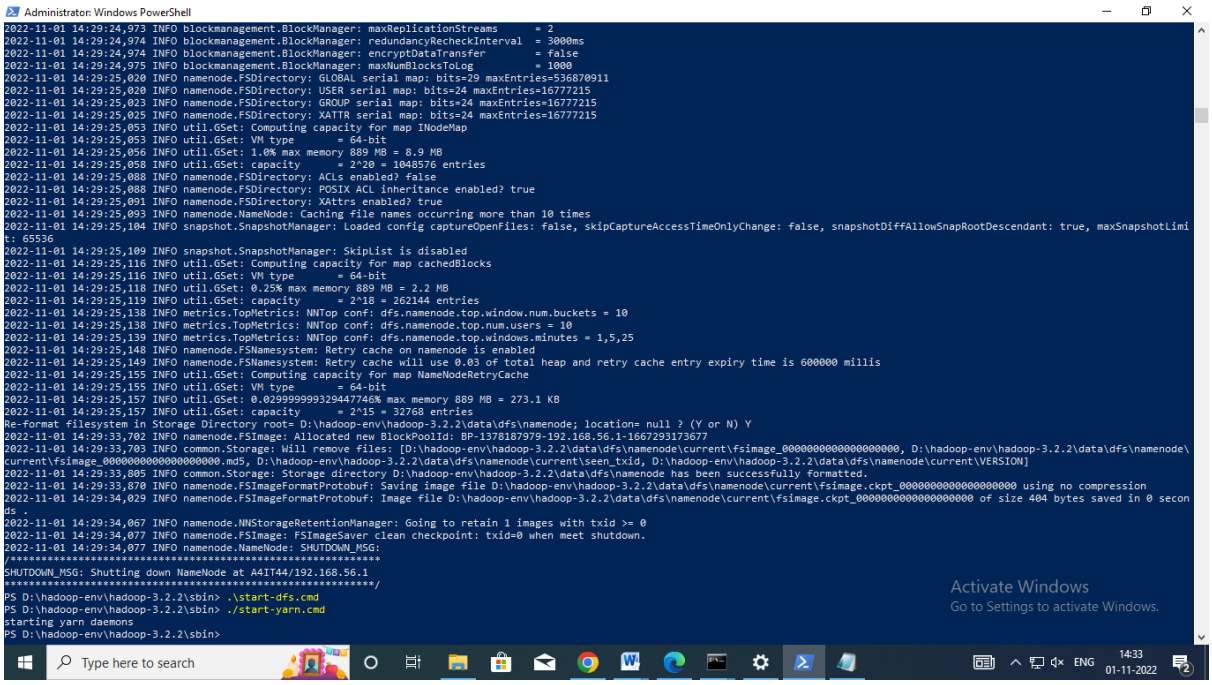
Two command prompt windows will open (one for the **namenode** and one for the **datanode**) as follows:



Name : R. Venkatesh Prabhu Roll No : 19UITE001

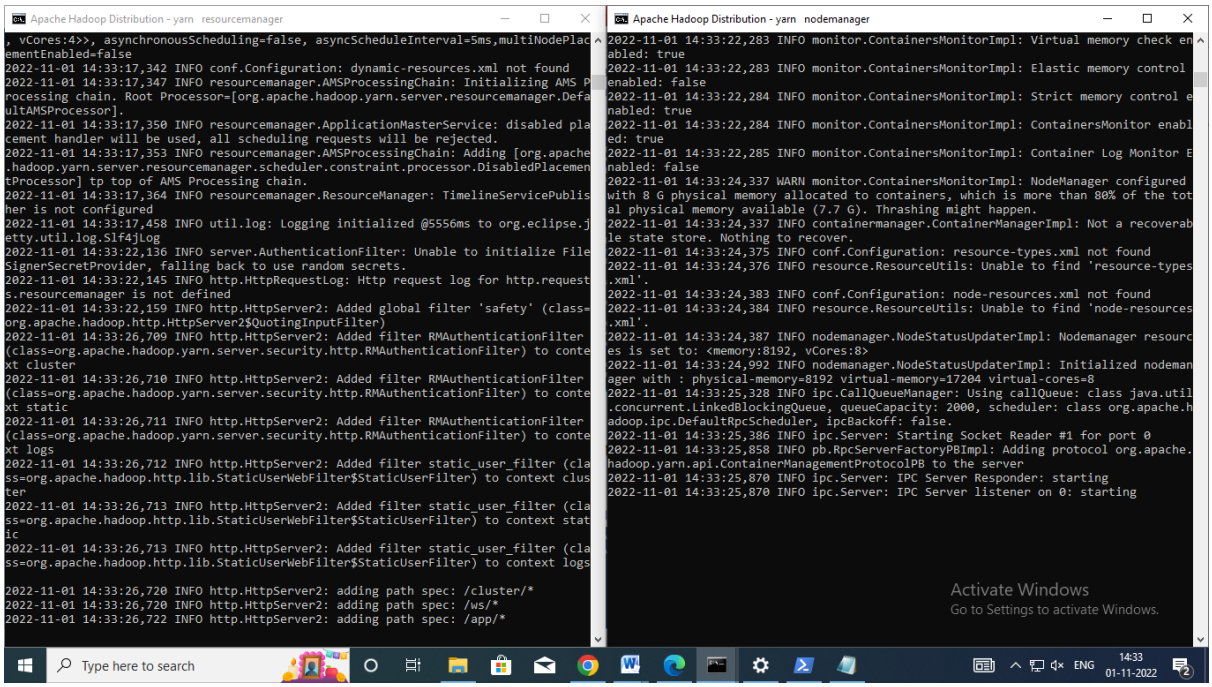
Next, Start the Hadoop Yarn service using the following command:

./start-yarn.cmd



Name : R. Venkatesh Prabhu Roll No : 19UITE001

Two command prompt windows will open (one for the **resource manager** and one for the **node manager**) as follows:



Name : R. Venkatesh Prabhu Roll No : 19UITE001

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

jps

It should display the following services:

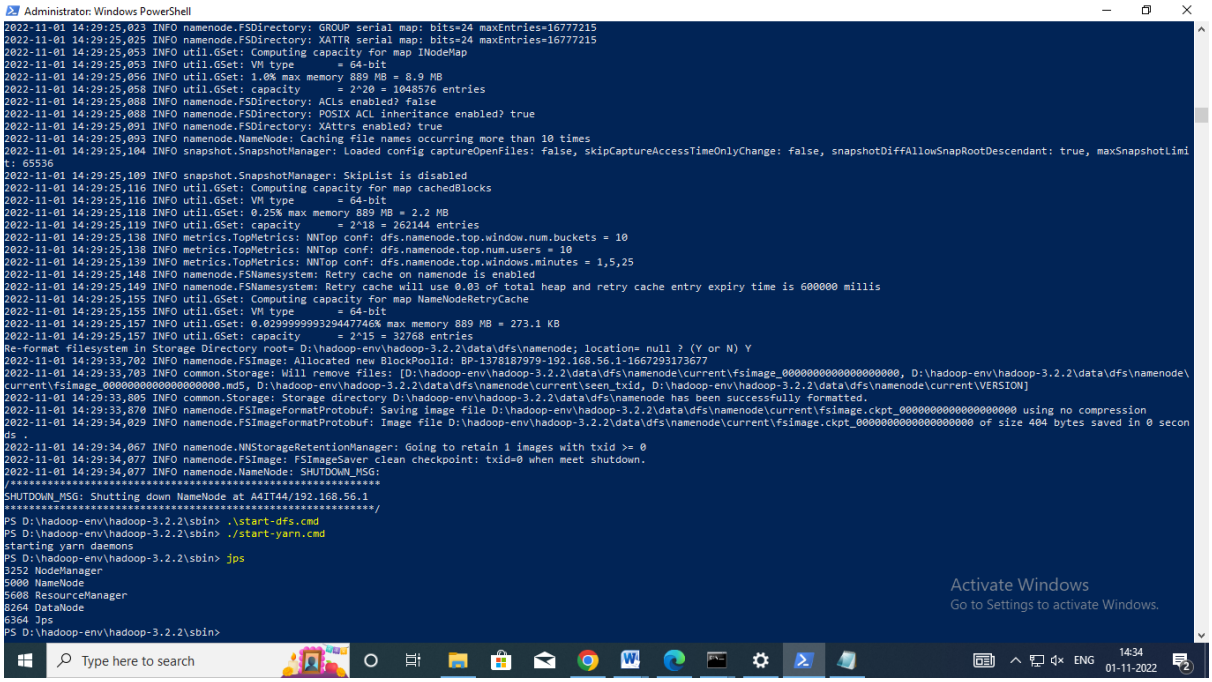
14560 DataNode

4960 ResourceManager

5936 NameNode

768 NodeManager

14636 Jps

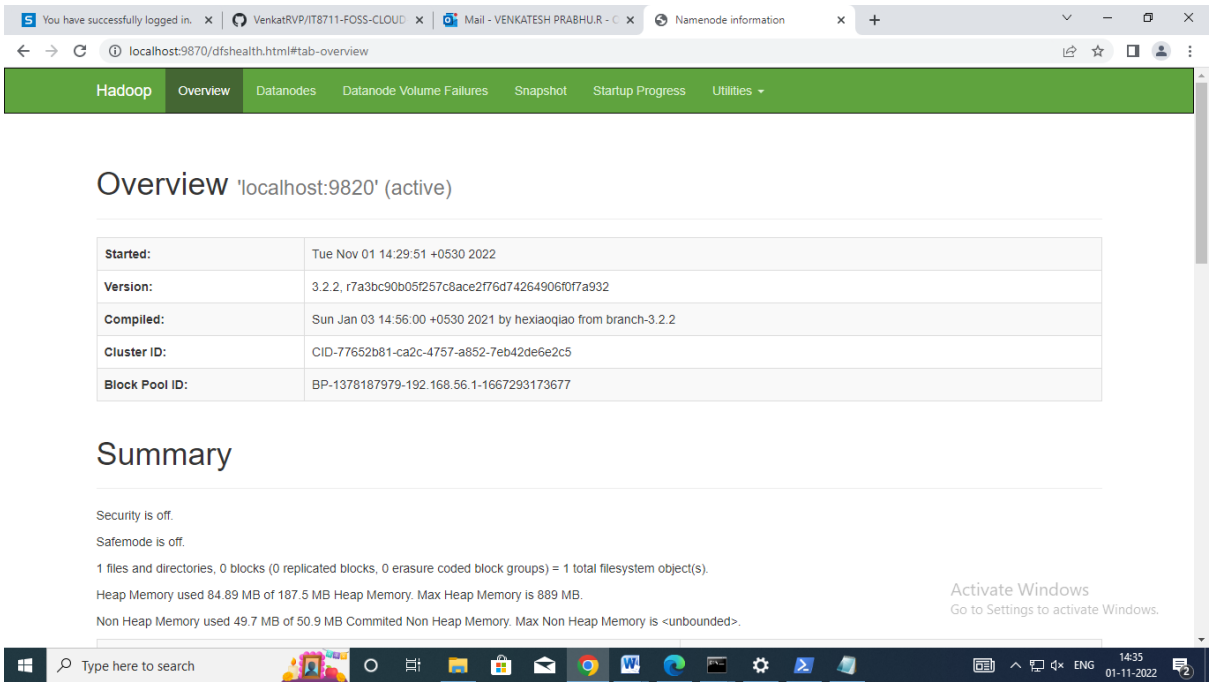


Name : R. Venkatesh Prabhu Roll No : 19UITE001

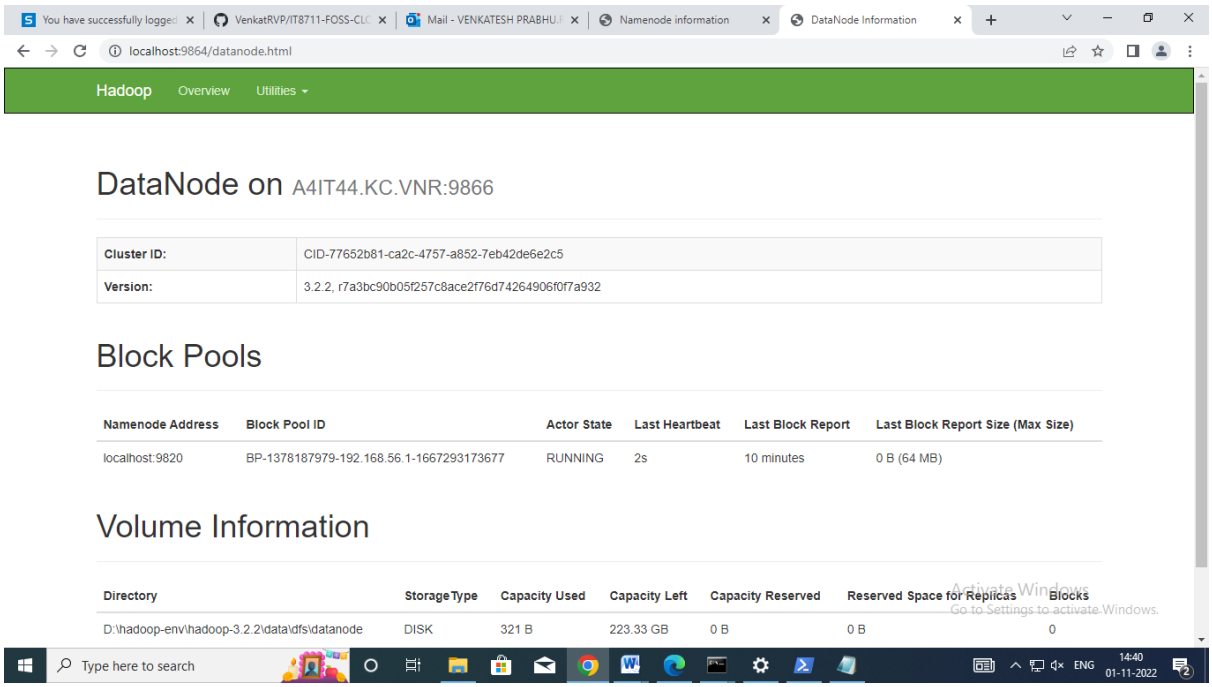
Output

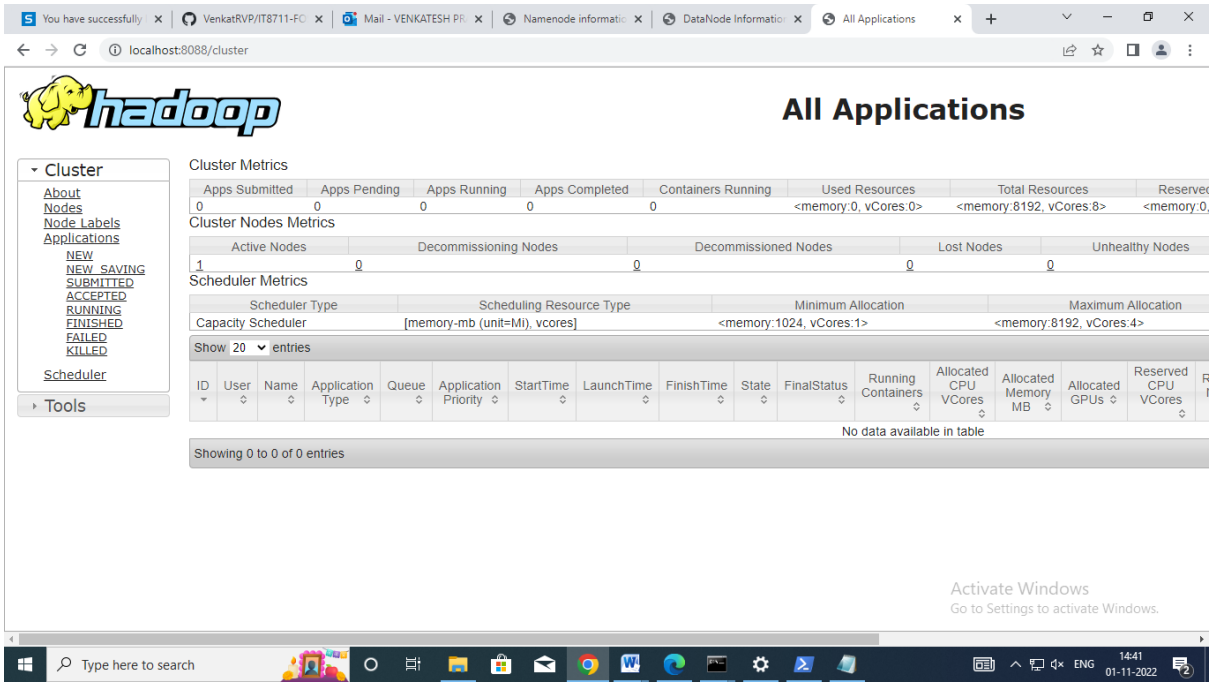
Hadoop Web UI

http://localhost:9870/dfshealth.html



http://localhost:9864/datanode.html



Name : R. Venkatesh Prabhu Roll No : 19UITE001

http://localhost:8088/cluster