**NAME: SANGEETHA.N**

**ROLL NO. : 19UITE013**

**EX. NAME : Installation of Single Node Hadoop Cluster**

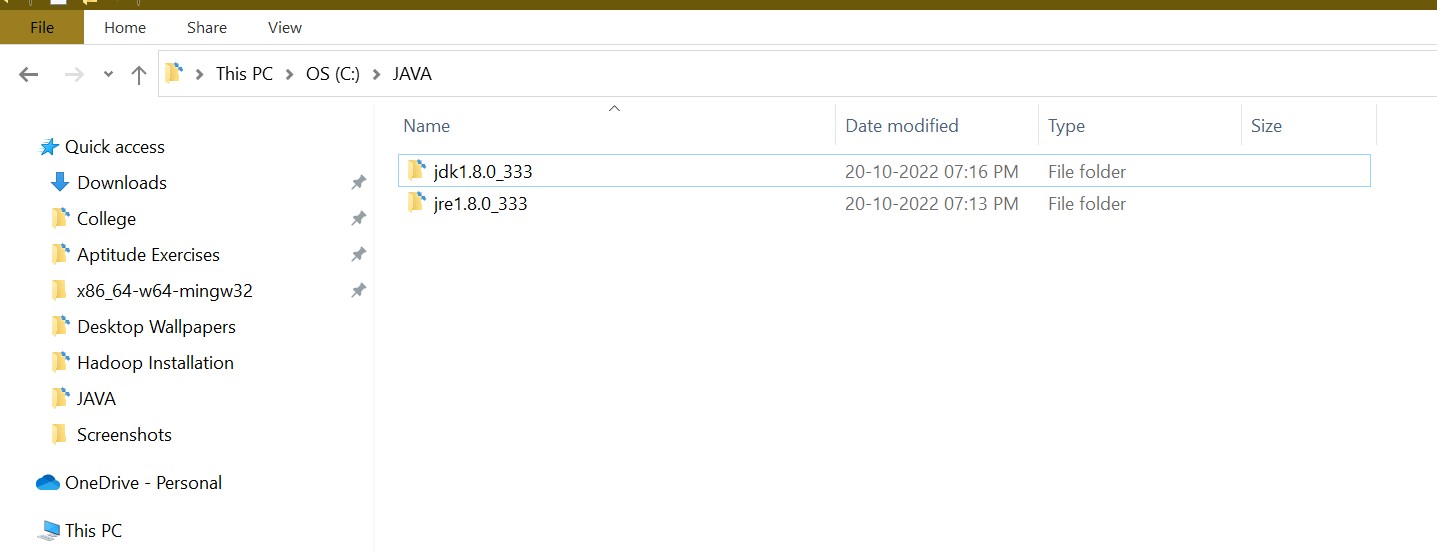
**DATE :15/10/2022**

# PROCEDURAL STEPS

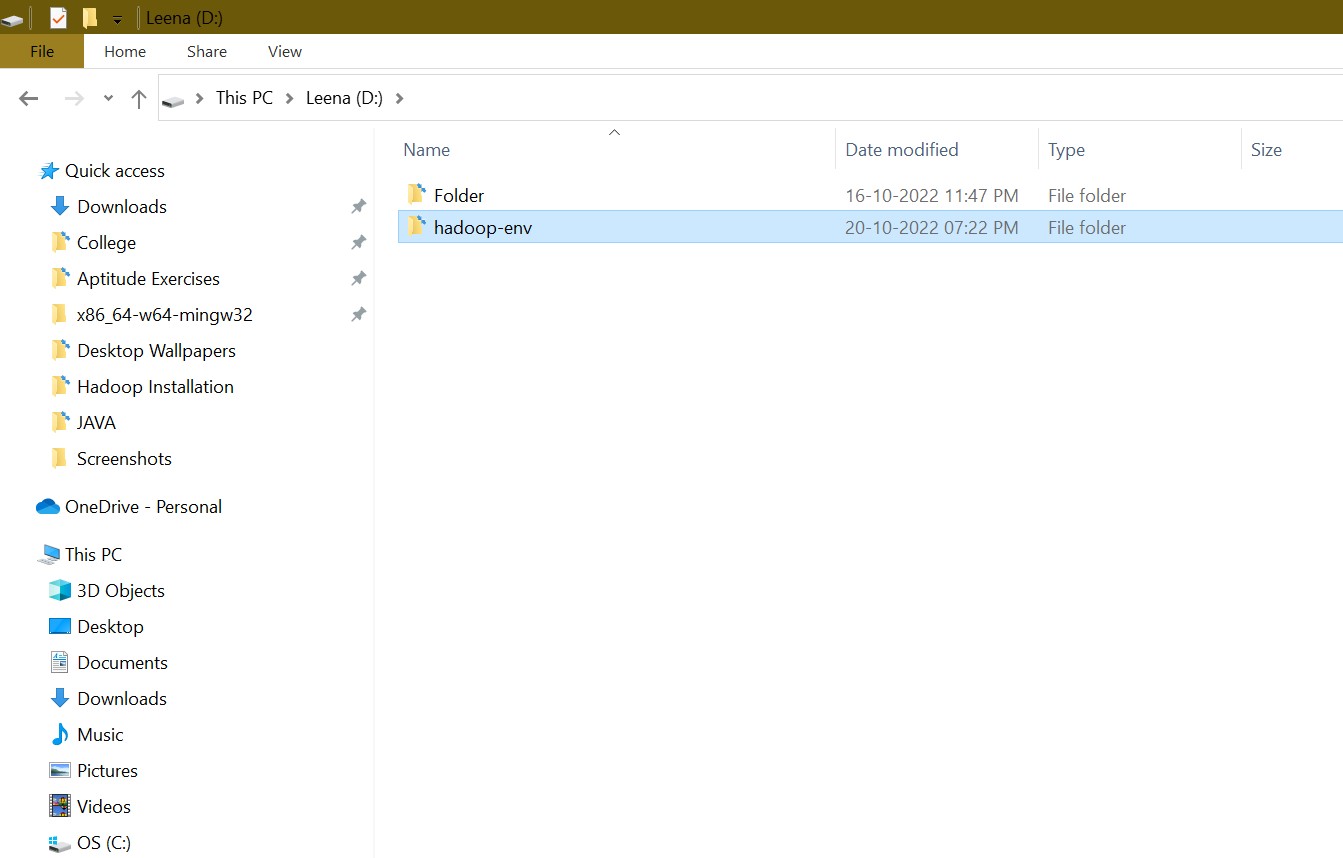
**Step 1:** Install / Extract the following Packages:

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

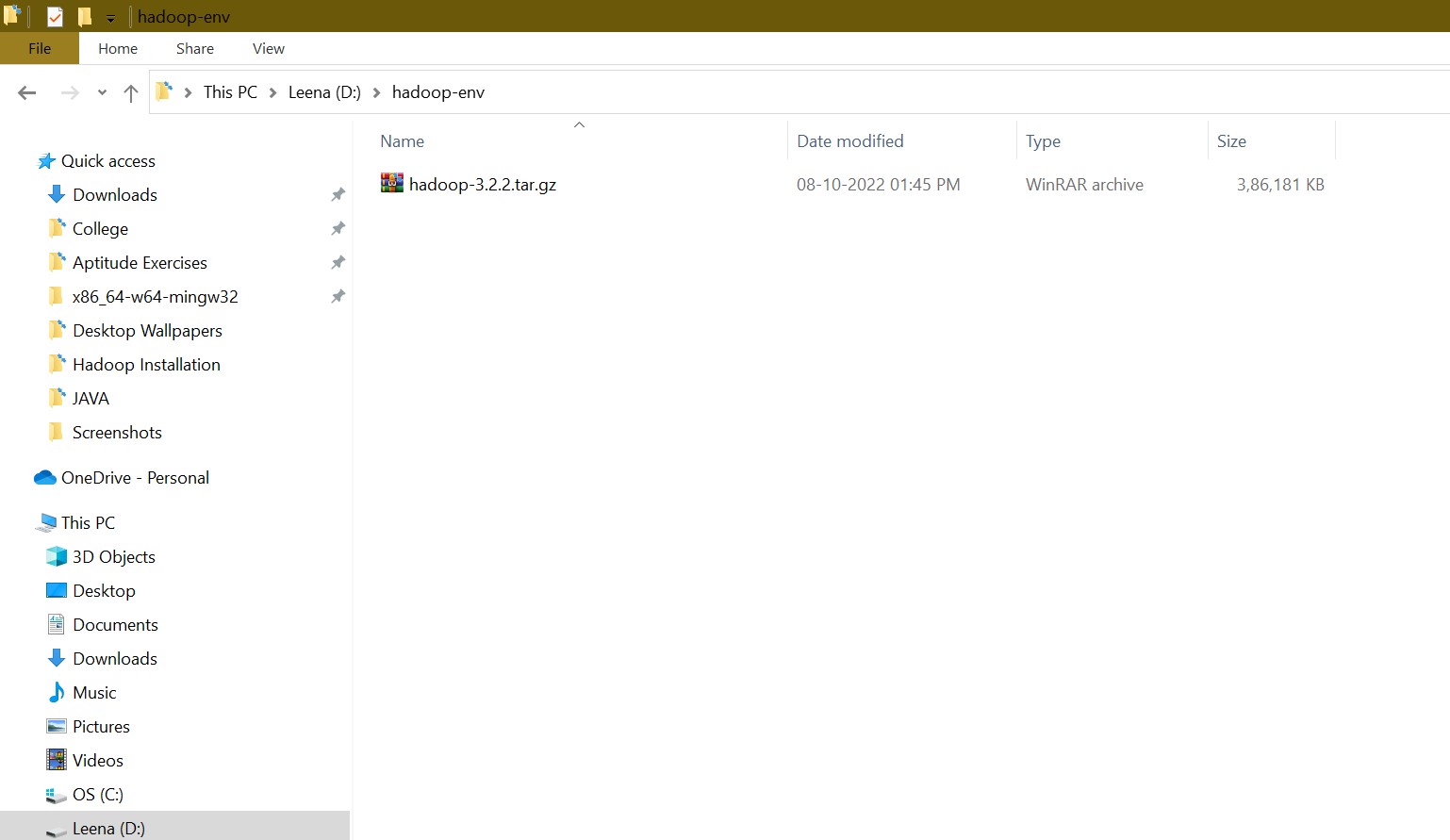
# “jre1.8.0\_333”.



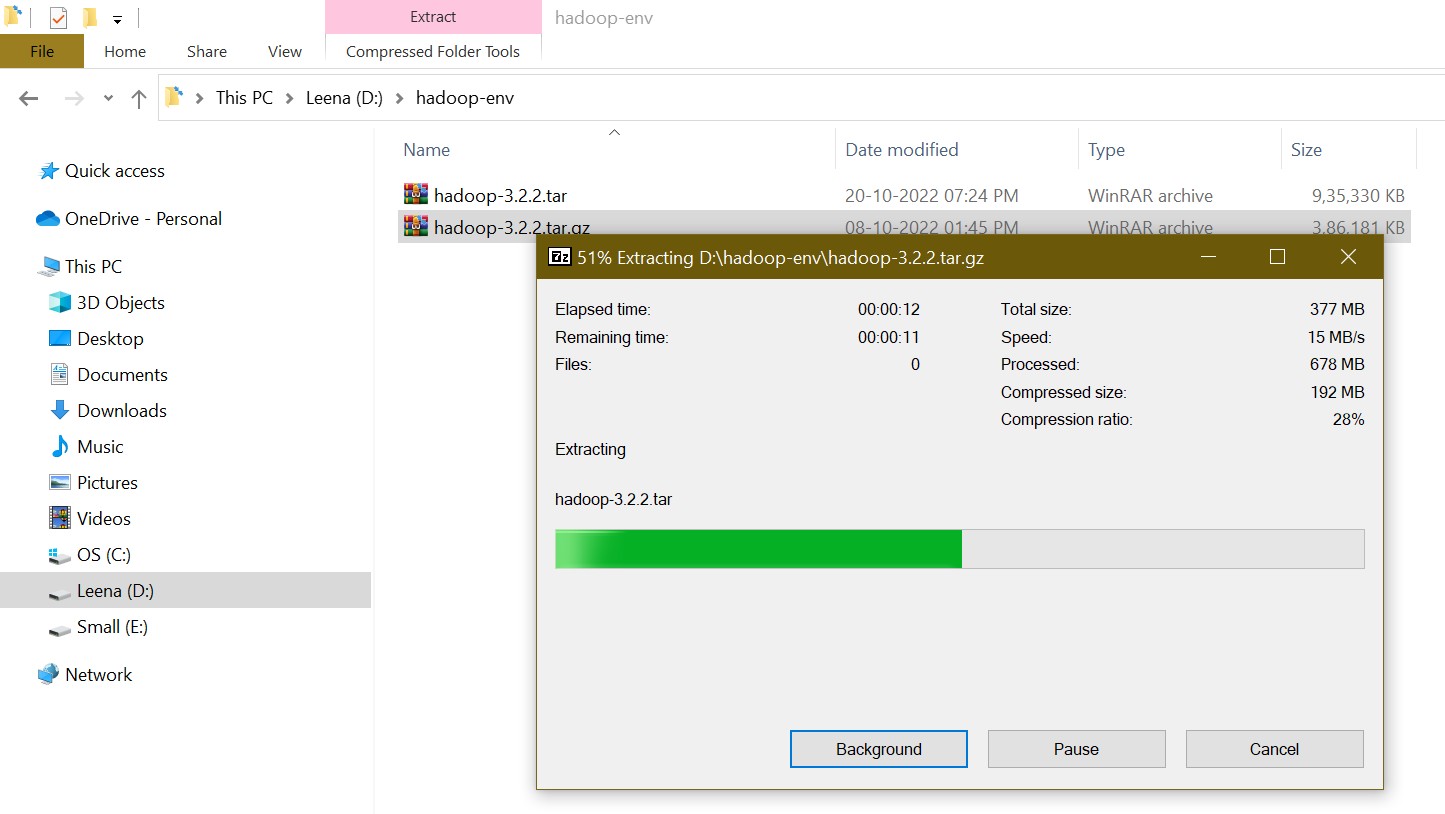
1. Create a folder **“hadoop-env”** in **D:\**

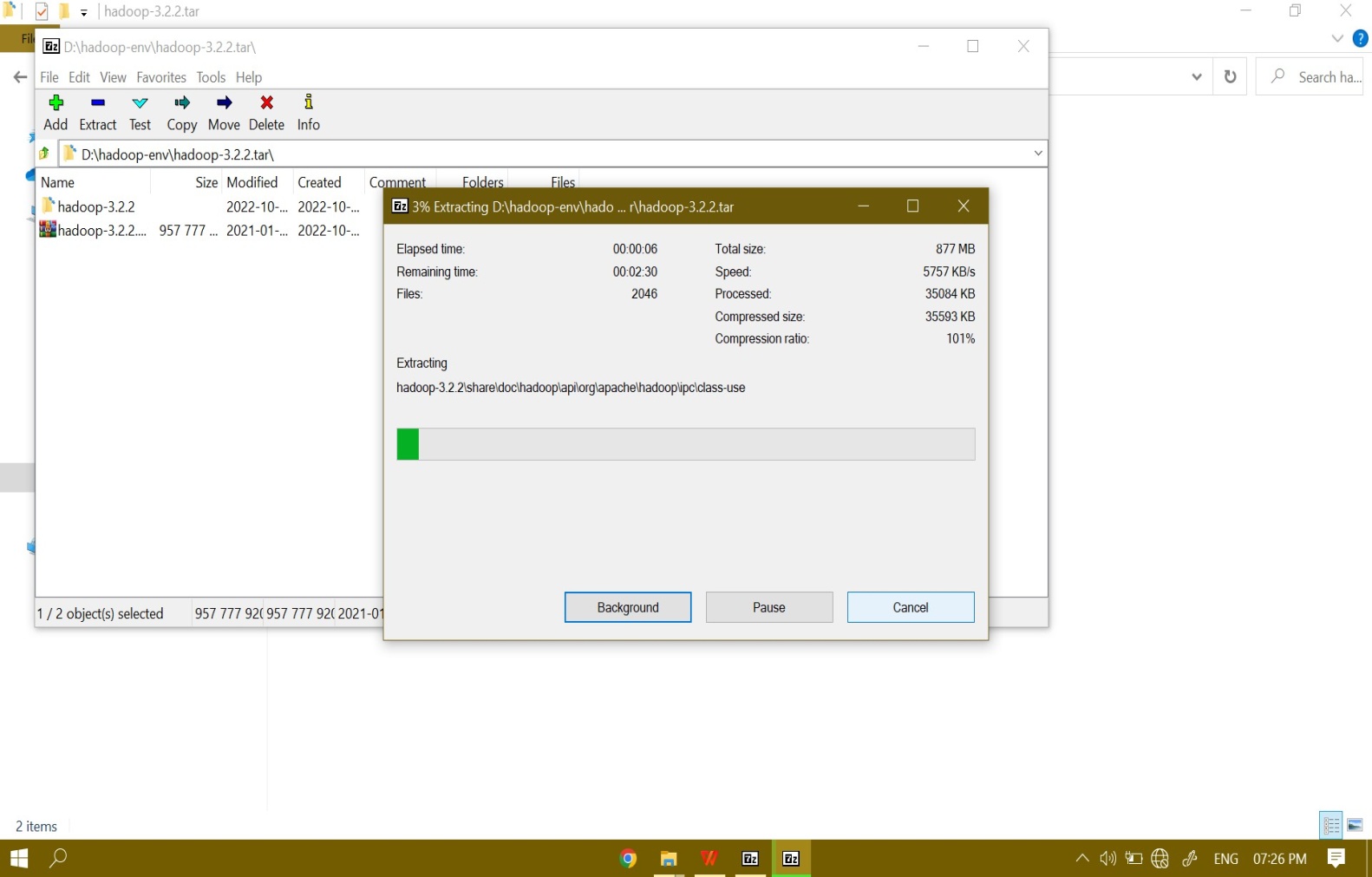


1. Paste the downloaded “hadoop-3.2.2.tar.gz” into “d:\ hadoop-env”

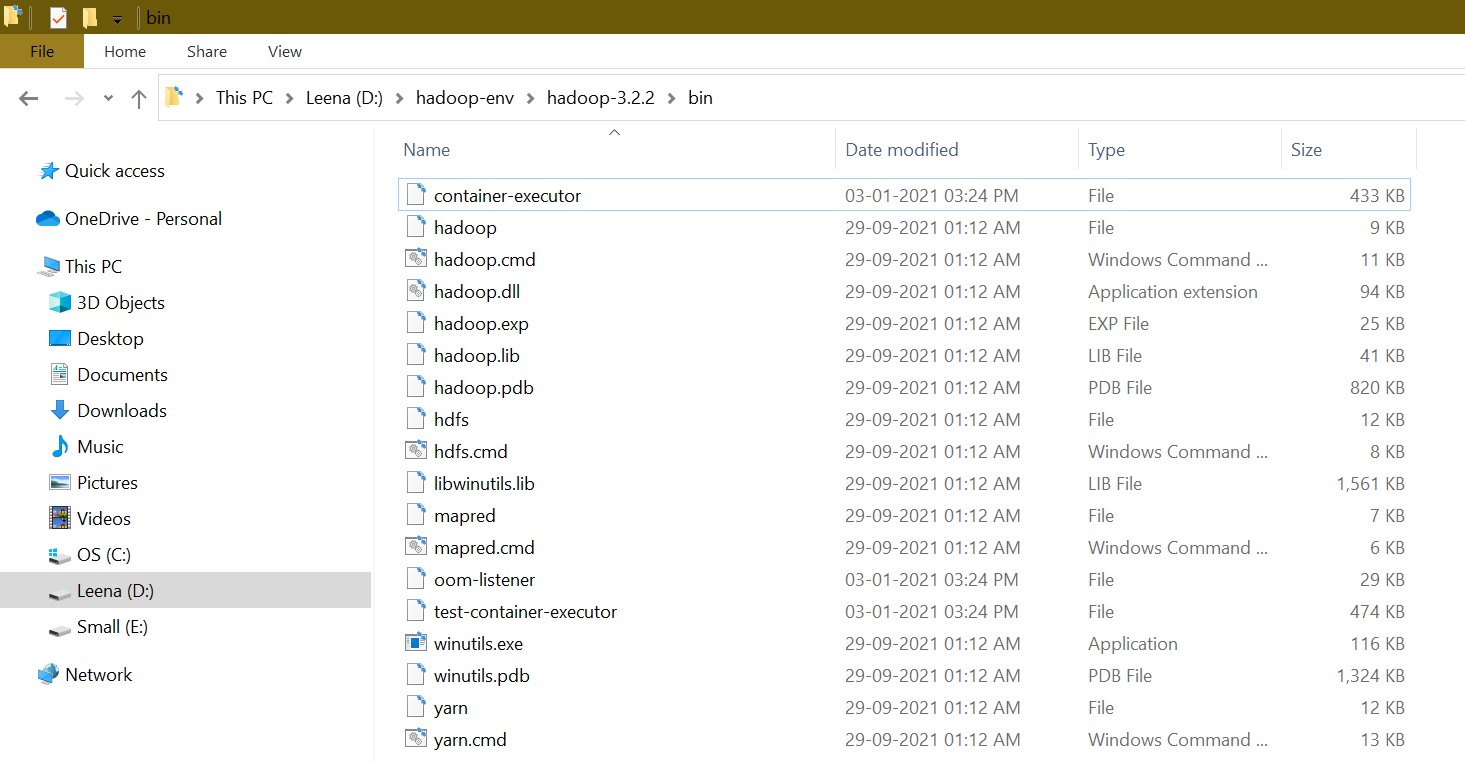


1. Unzip the “hadoop-3.2.2.tar.gz” using 7zip (do the unzip function two times)

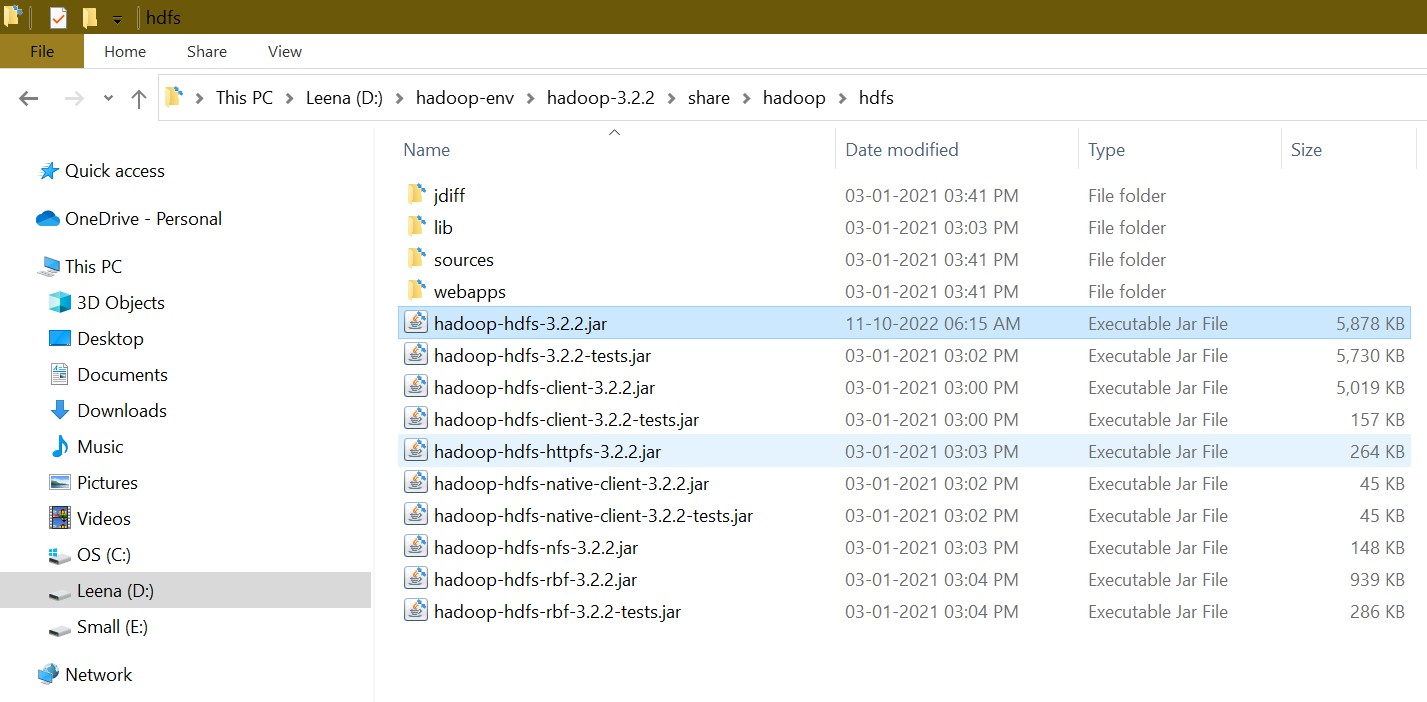




1. Copy the downloaded Hadoop dll files to the location “D:\hadoop-env\hadoop-3.2.2\bin”

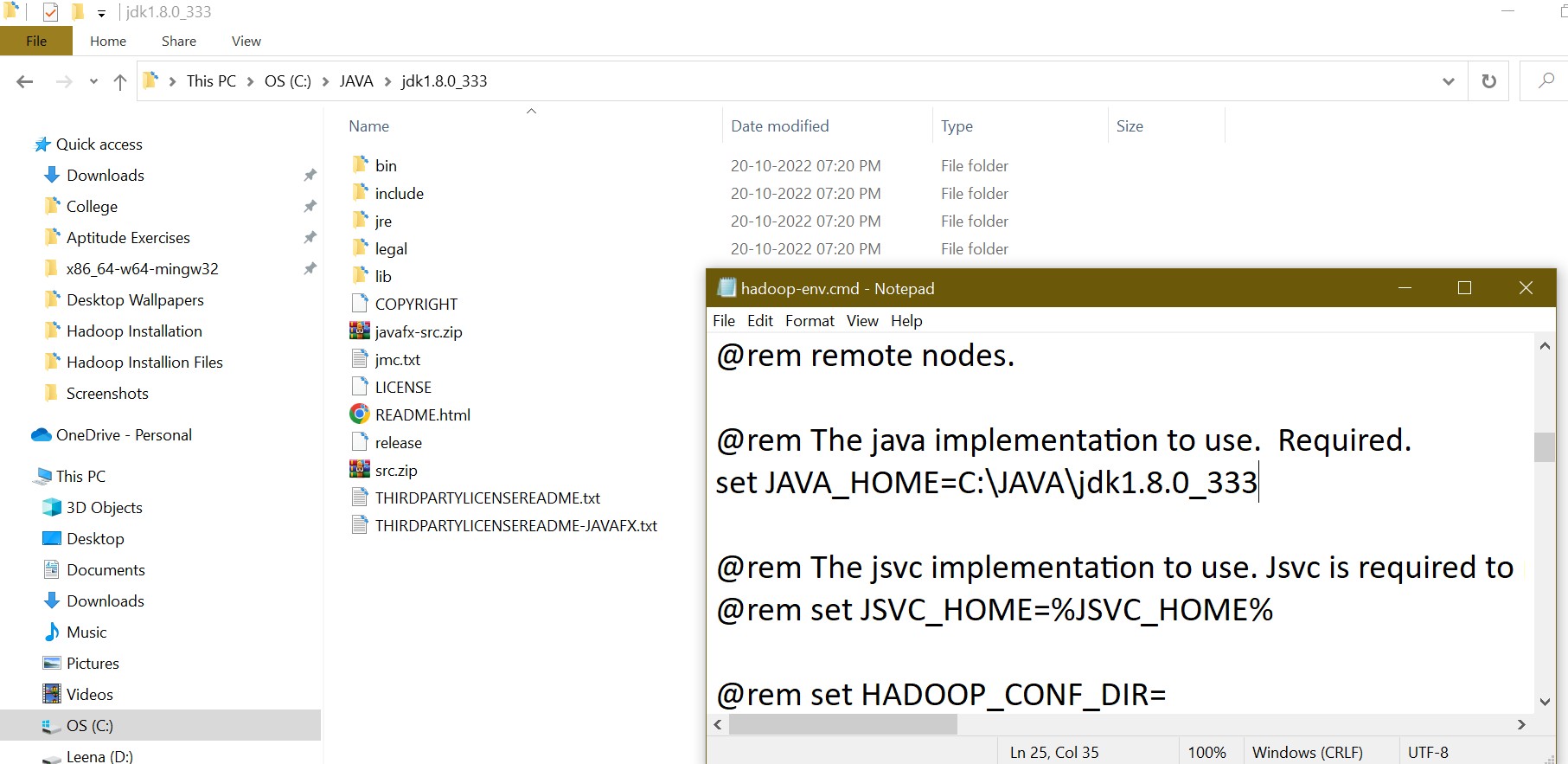


1. Copy the downloaded hadoop-hdfs-3.2.2.jar to “D:\hadoop-env\hadoop-3.2.2\share\hadoop\hdfs”



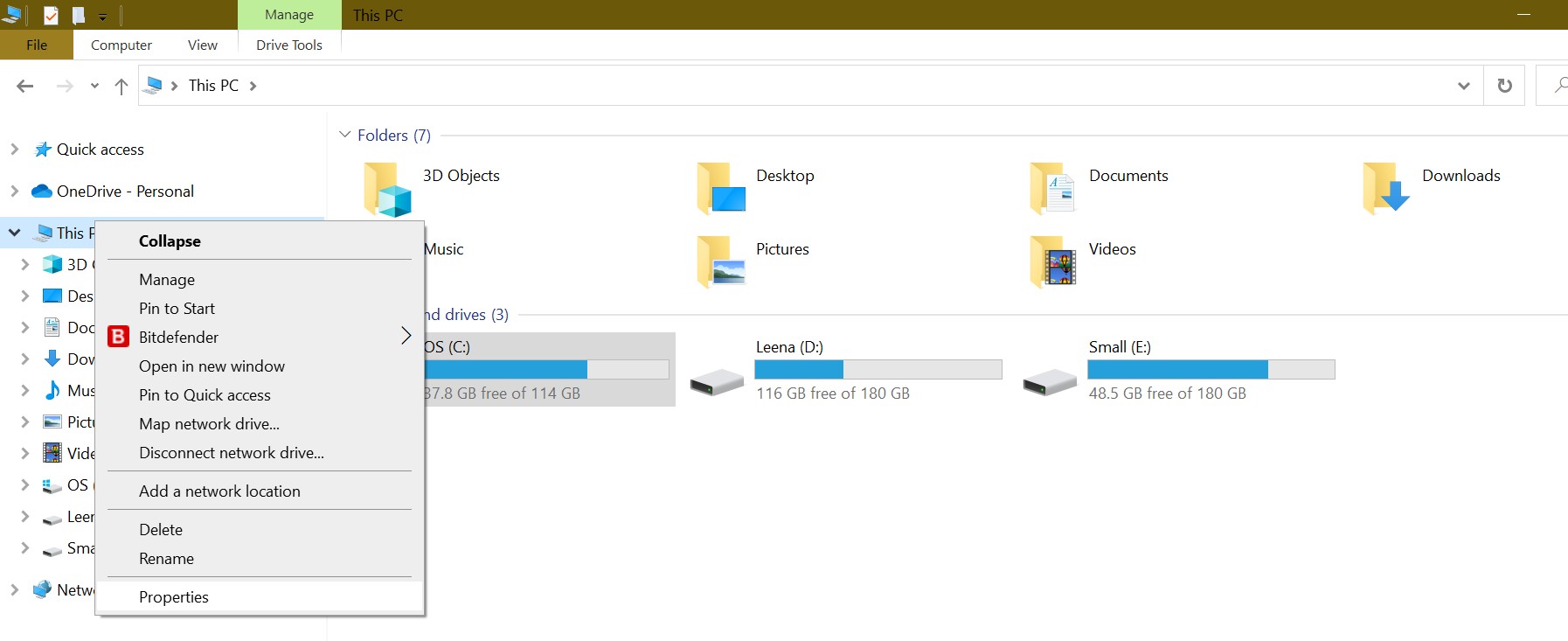
1. Edit the JAVA\_HOME = C:\JAVA\jdk1.8.0\_333 in the path “**D:\hadoop-env\hadoop- 3.2.2\etc\hadoop\hadoop-env.cmd**



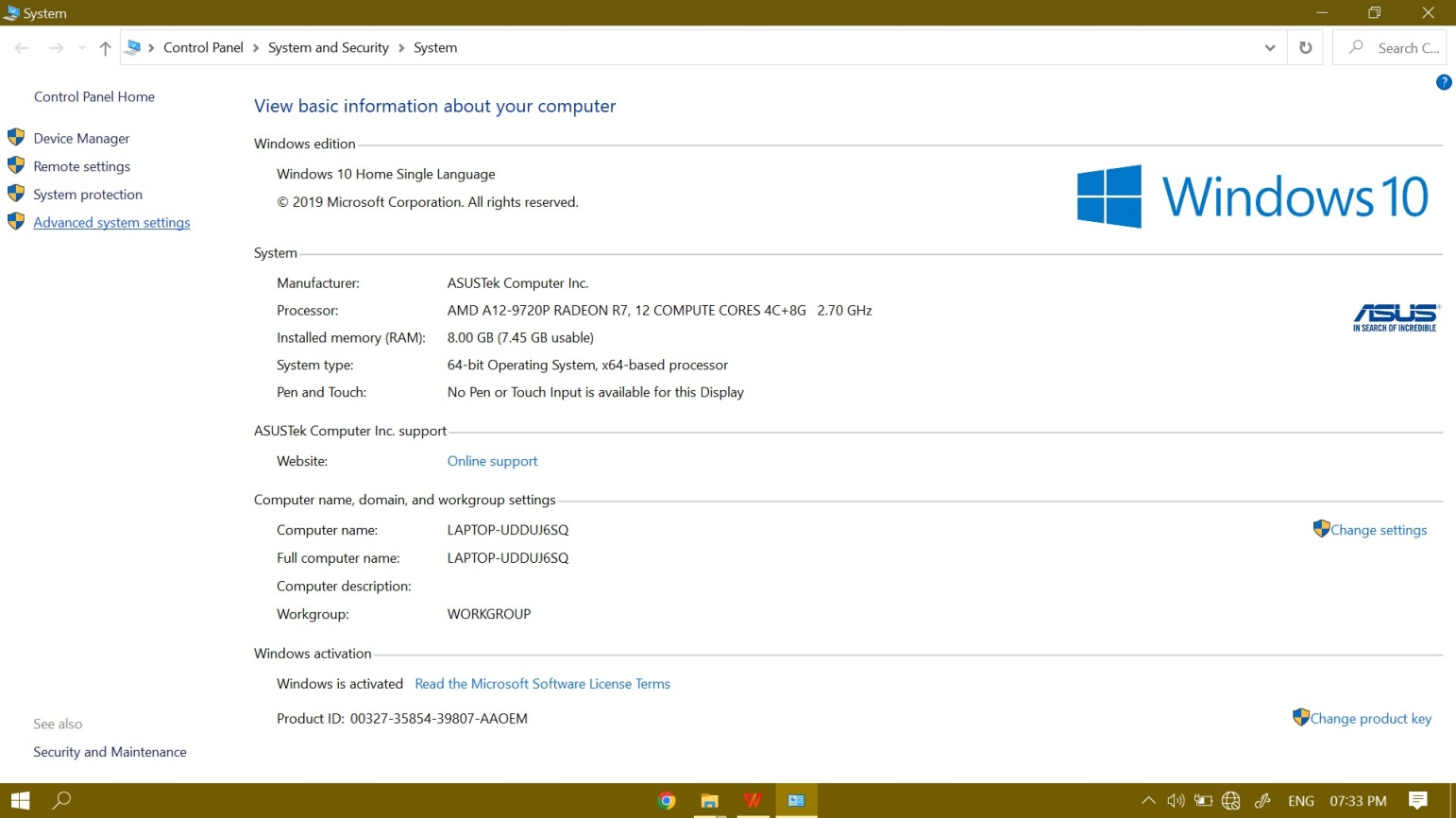


**Step 2:** Set the path for Java and Hadoop in system environment variables

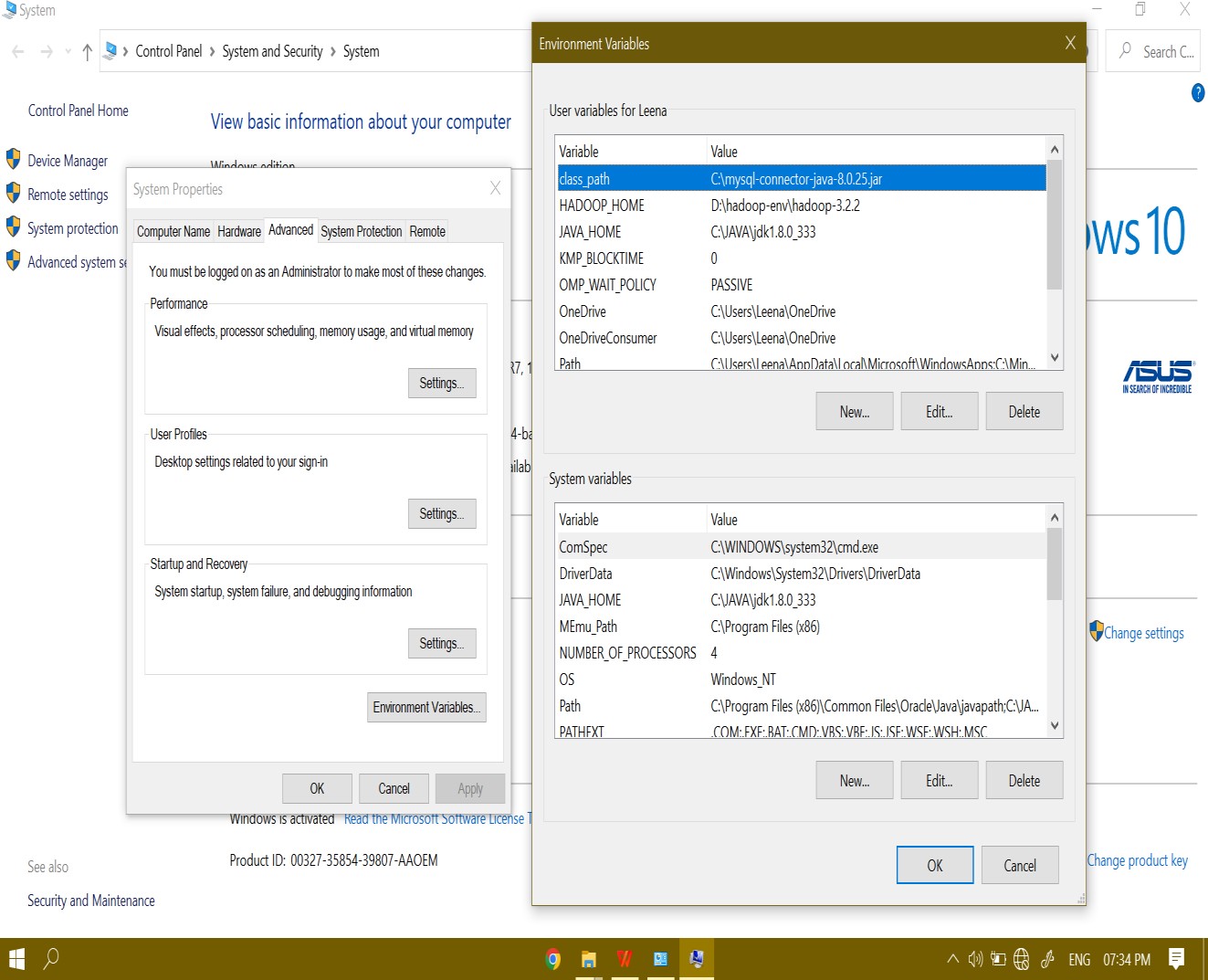
To edit environment variables, go to Control Panel  System and Security  System (or) right-click “This PC”  Properties (My Computer icon) and click on the “Advanced system settings” link.



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.



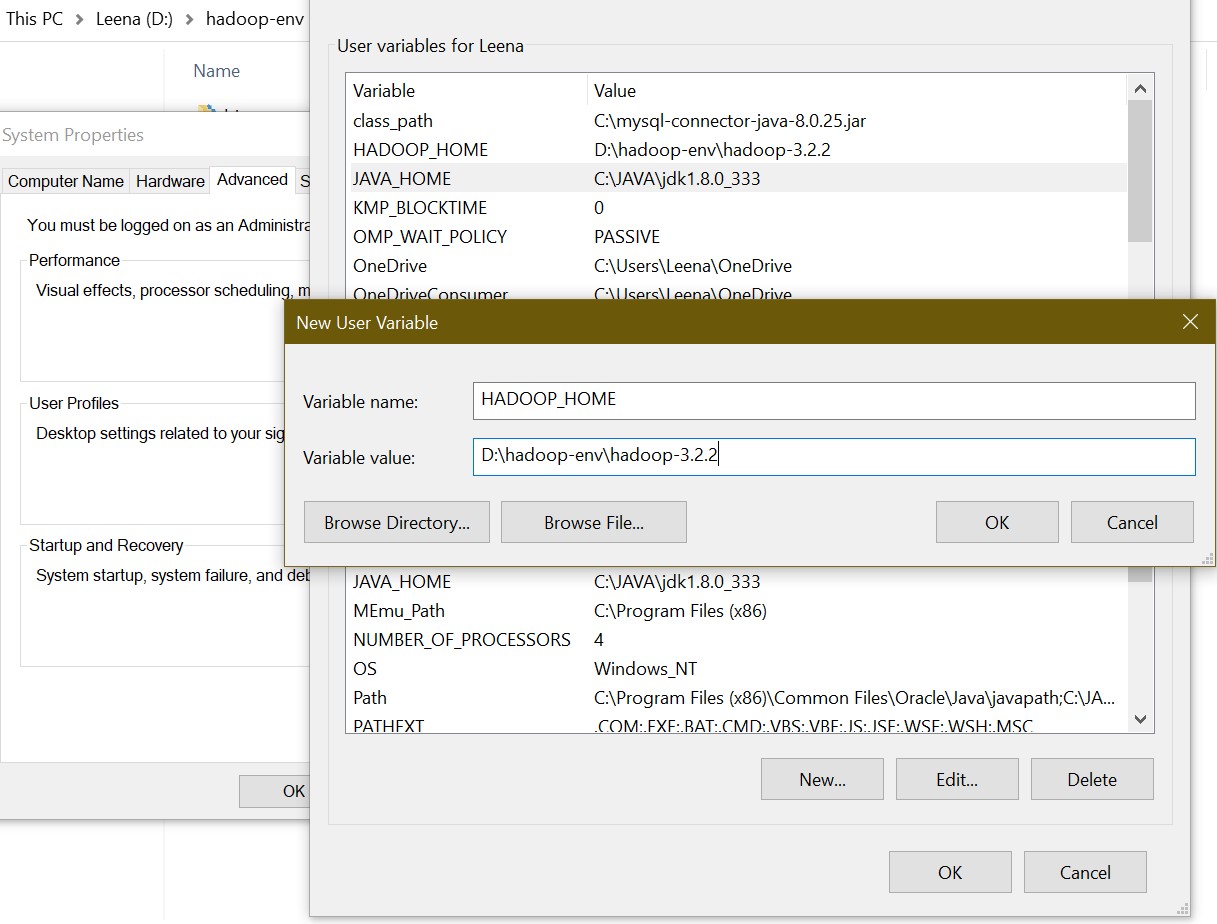
In the “Environment Variables” dialog, press the “New” button to add a new variable.

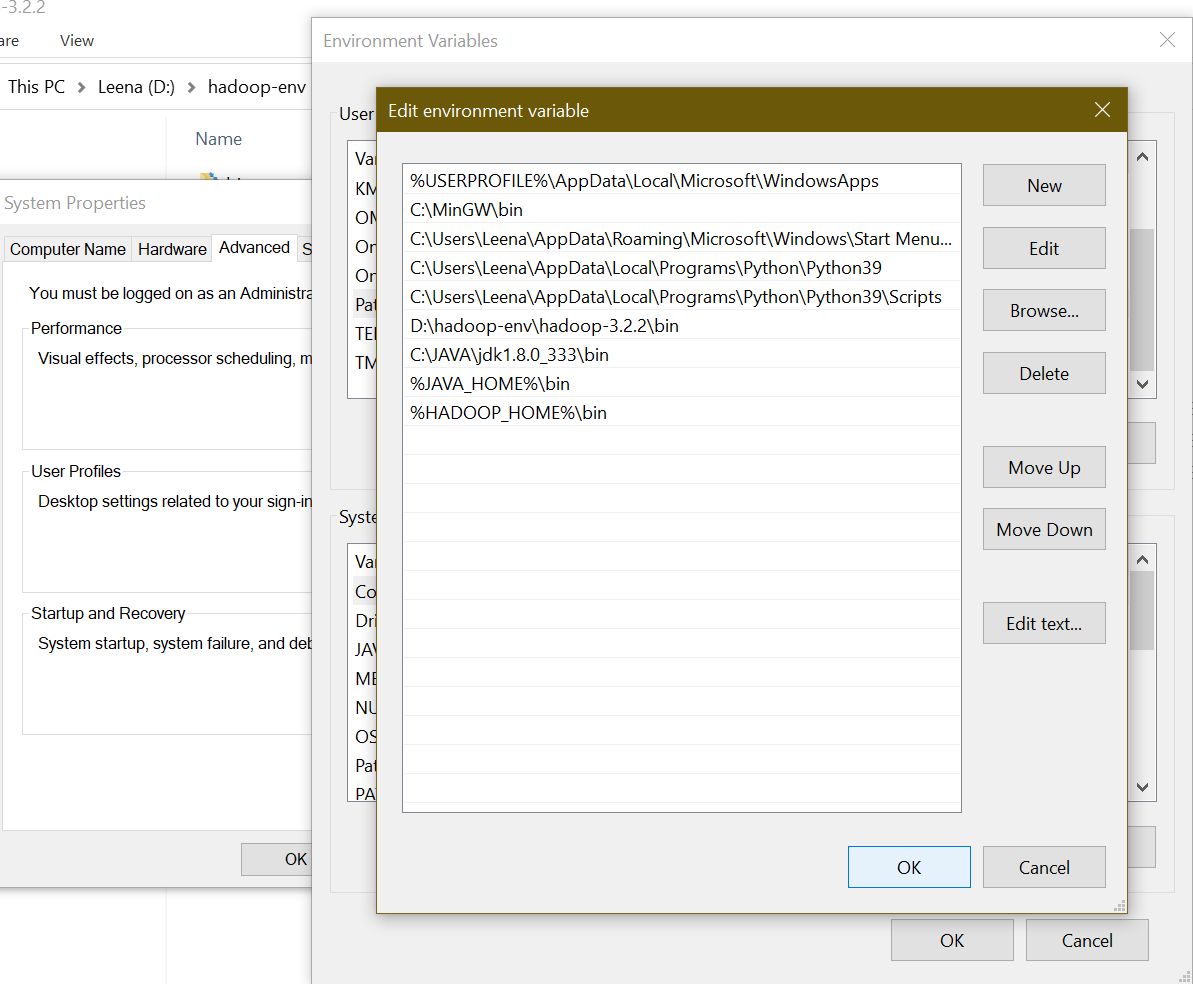


Now, Edit the PATH variable to add the Java and Hadoop binaries paths as shown in the following screenshots.

# JAVA\_HOME: JDK installation folder path

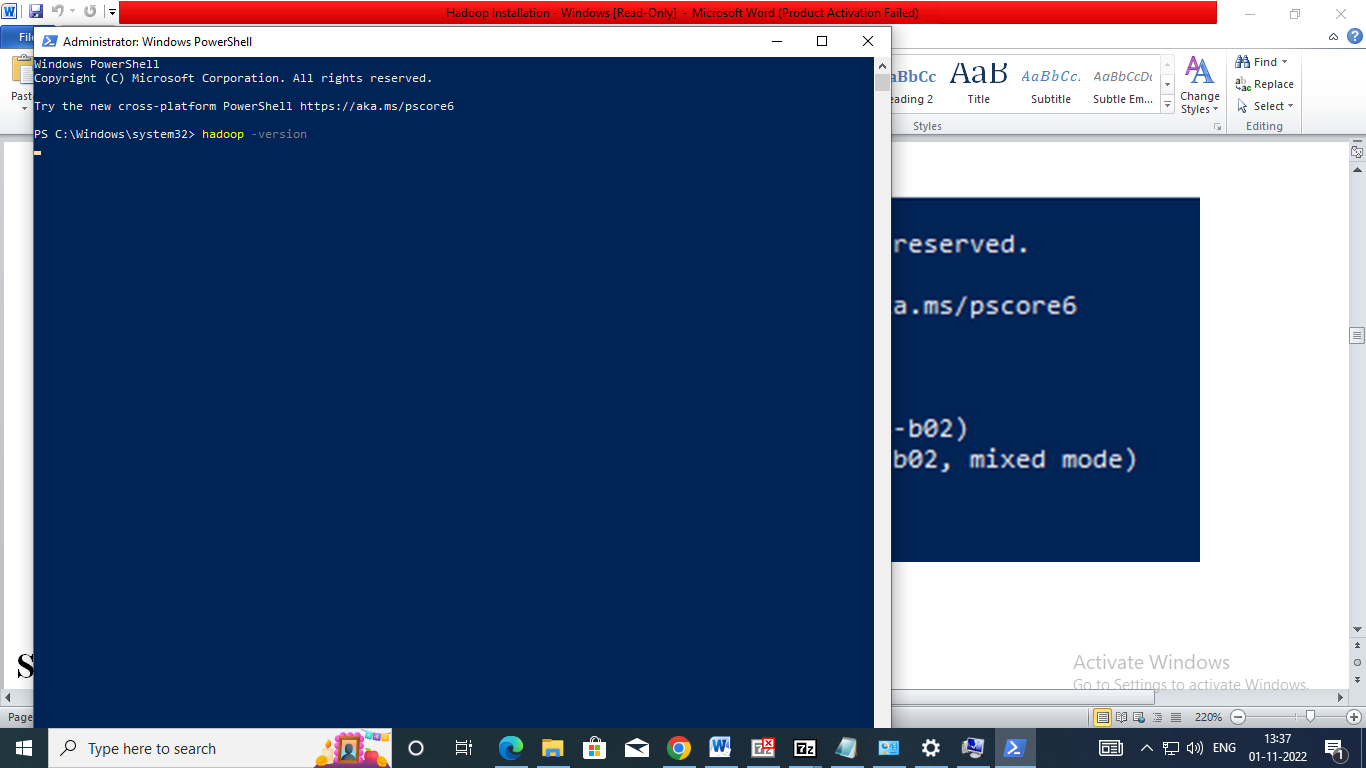
**HADOOP\_HOME: Hadoop installation folder path**





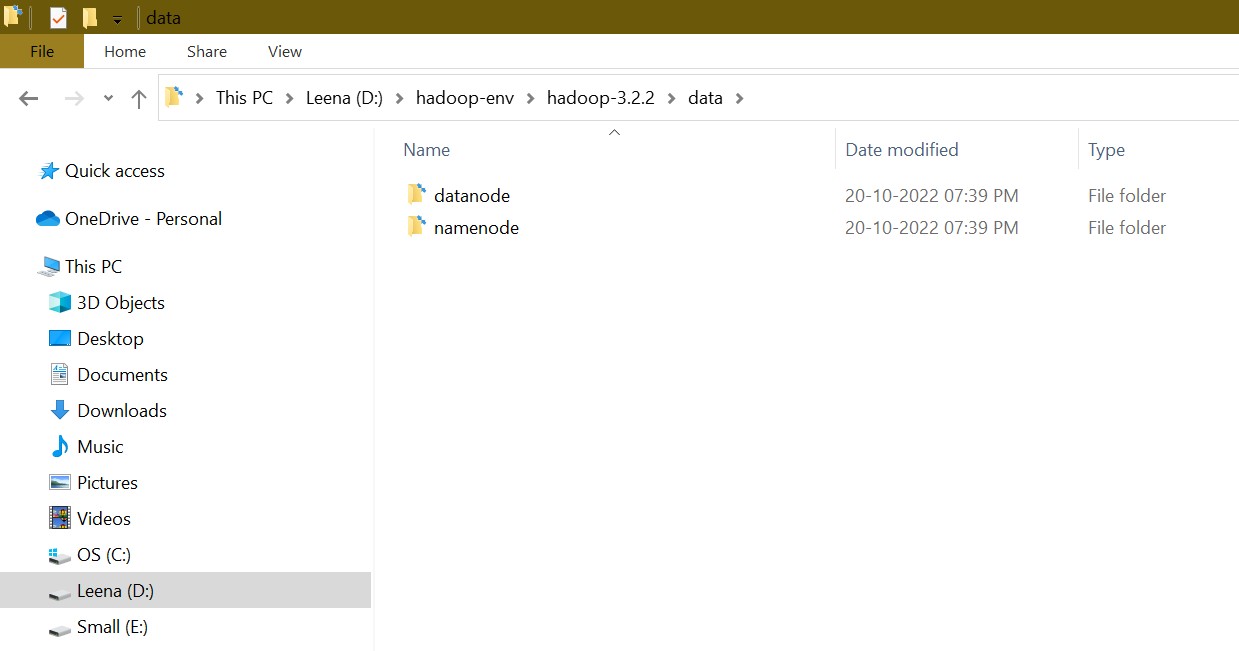
**Step 3:** Open Command Prompt as Administrator and run the following command:

# hadoop –version

****

**Step 4:** Create the following folders

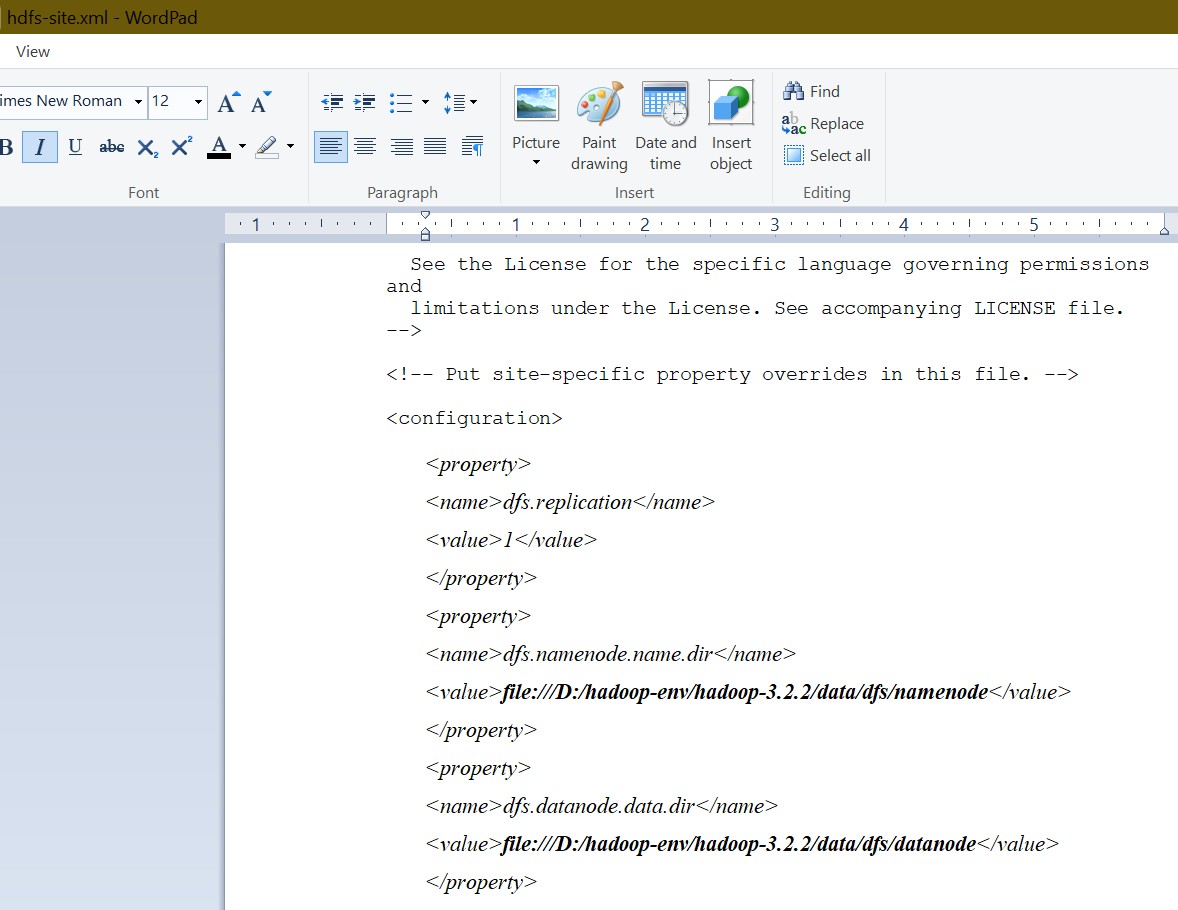
D:\hadoop-env\hadoop-3.2.2\**data\dfs\namenode** D:\hadoop-env\hadoop-3.2.2\**data\dfs\datanode**



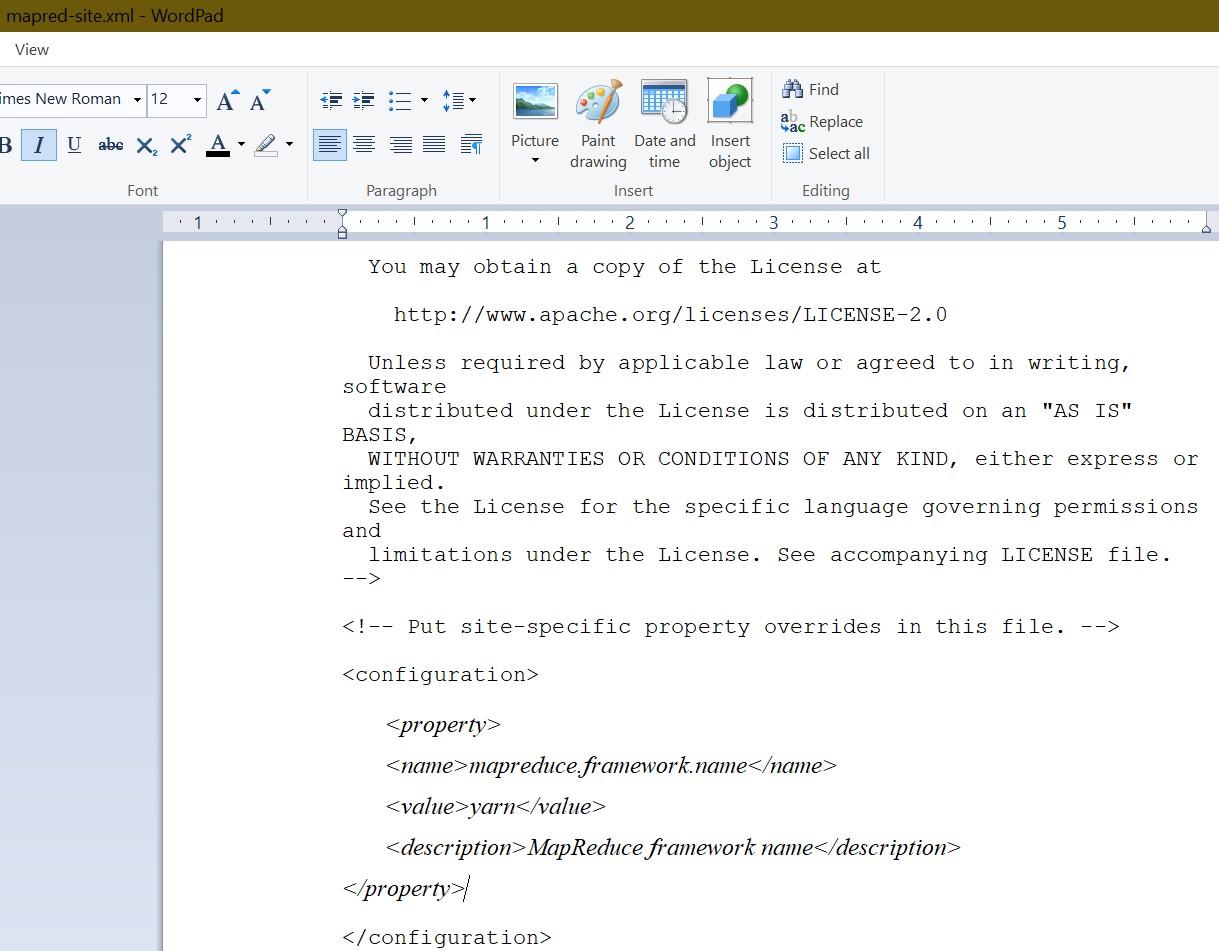
**Step 5:** Configuring Hadoop cluster

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

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





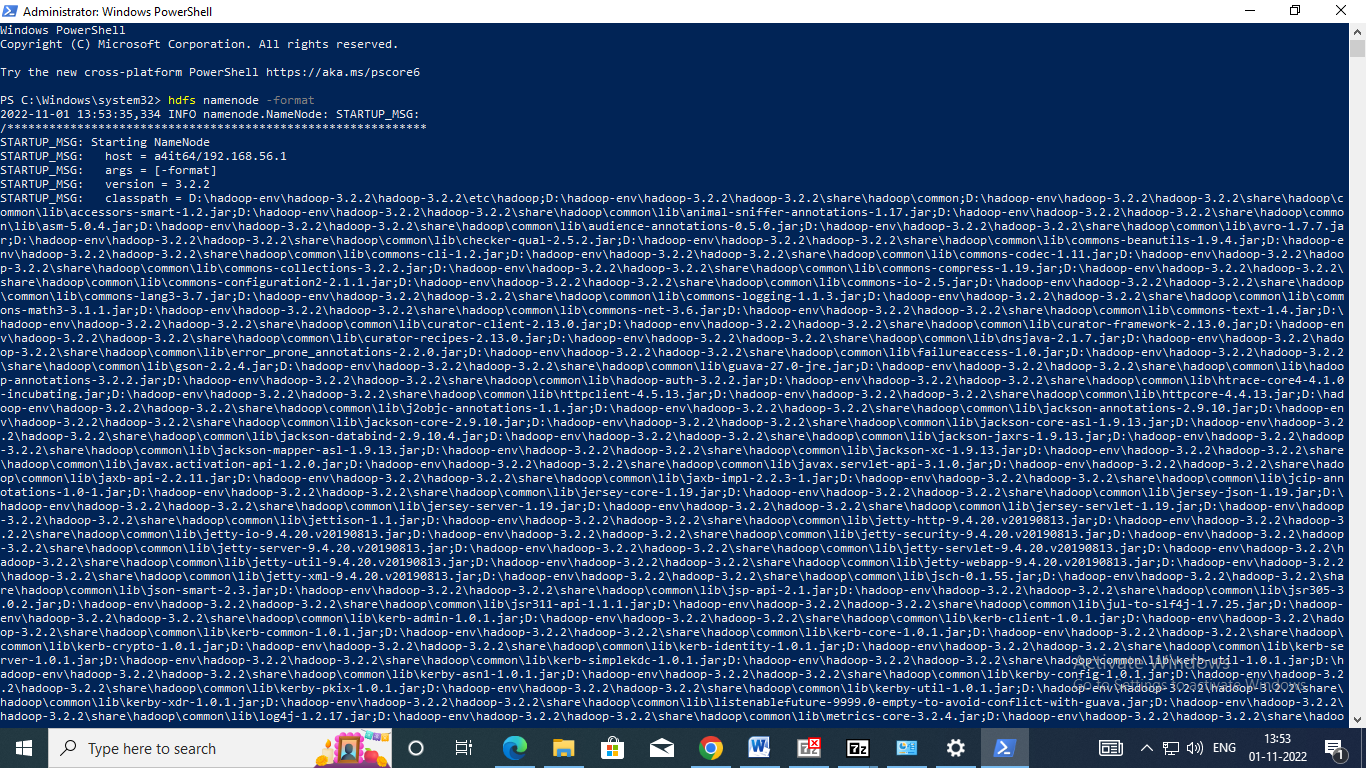


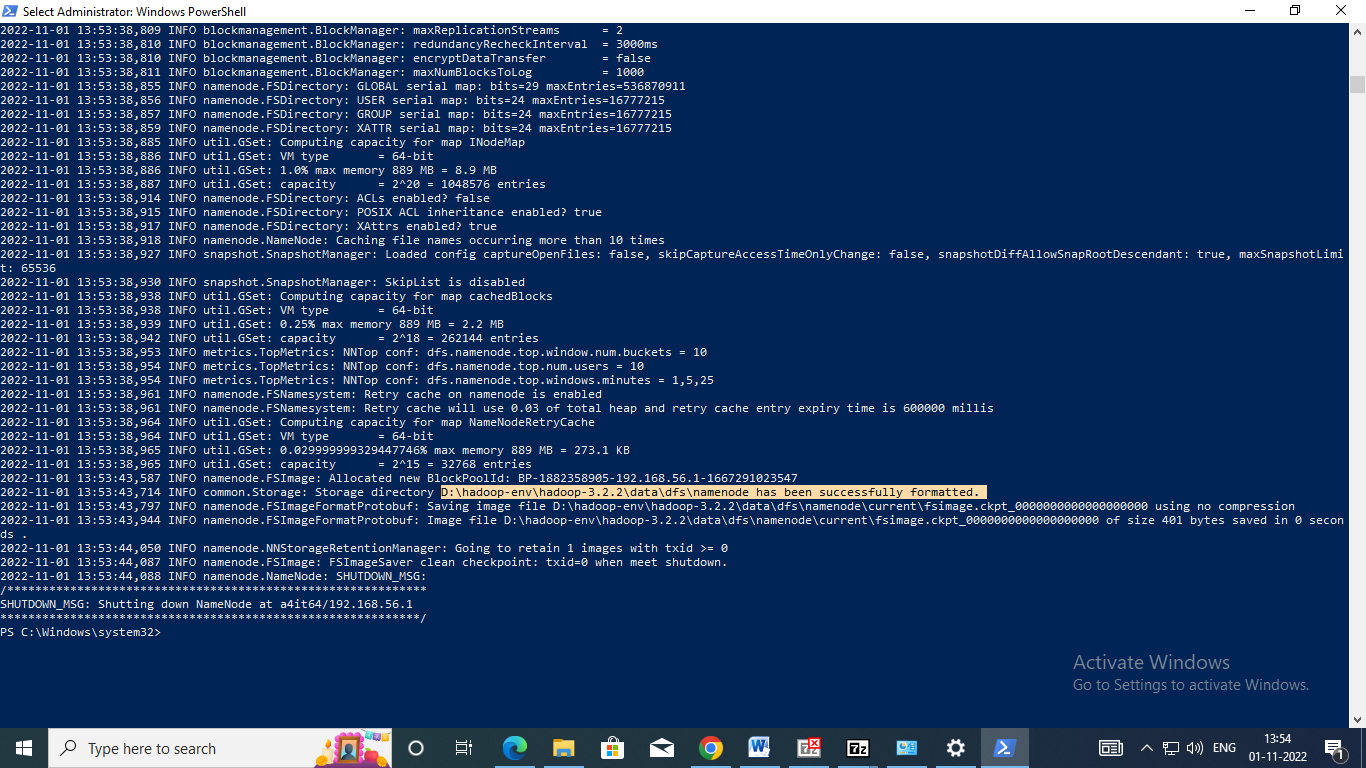


**Step 6:** Format the namenode

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

# hdfs namenode –format

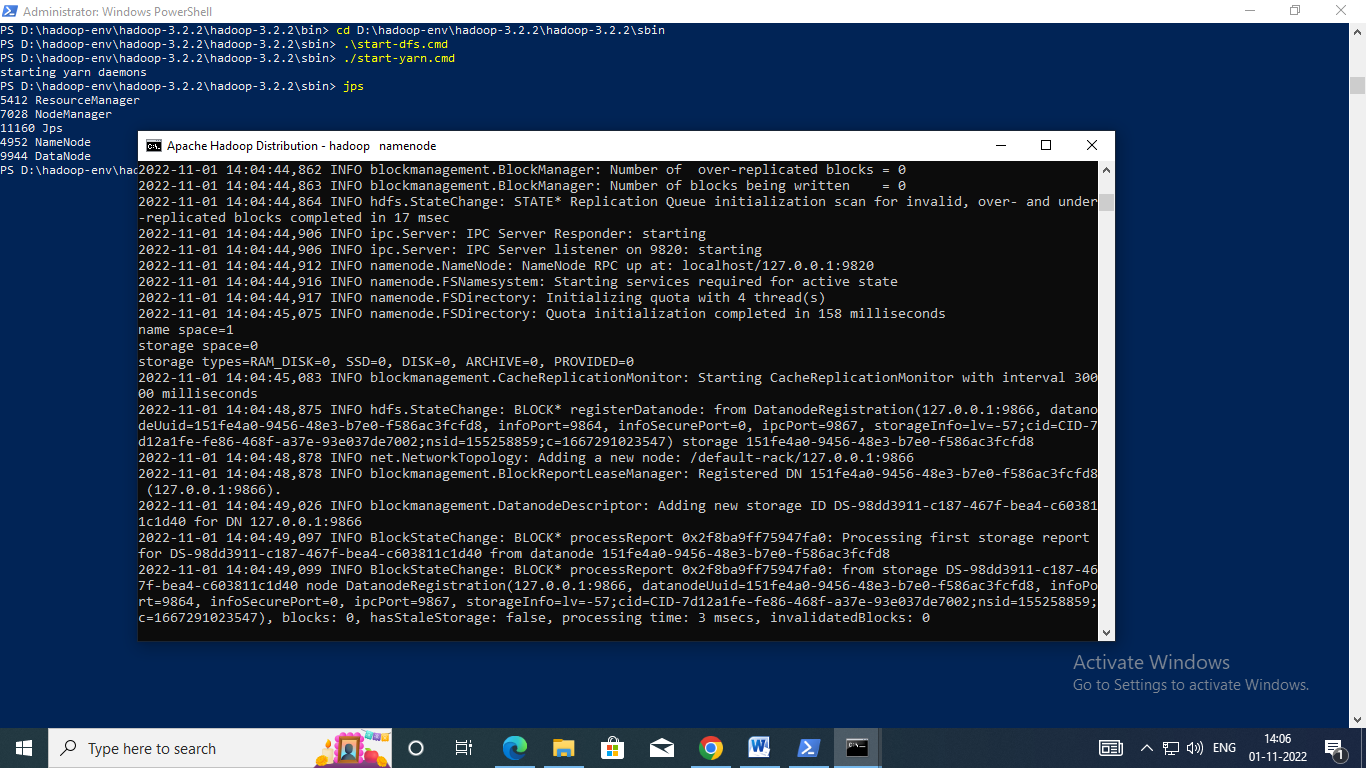


****

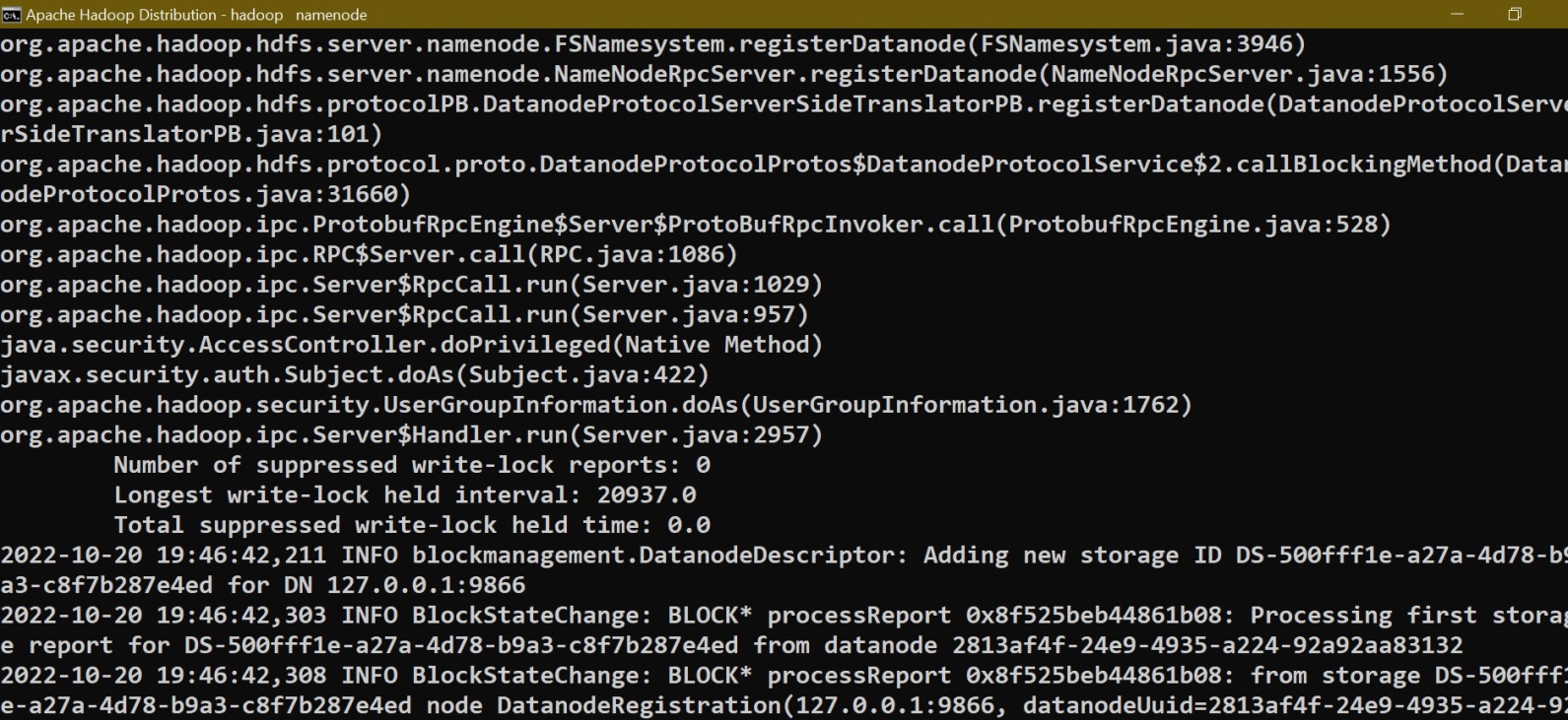
**Step 7:** Start the Hadoop nodes

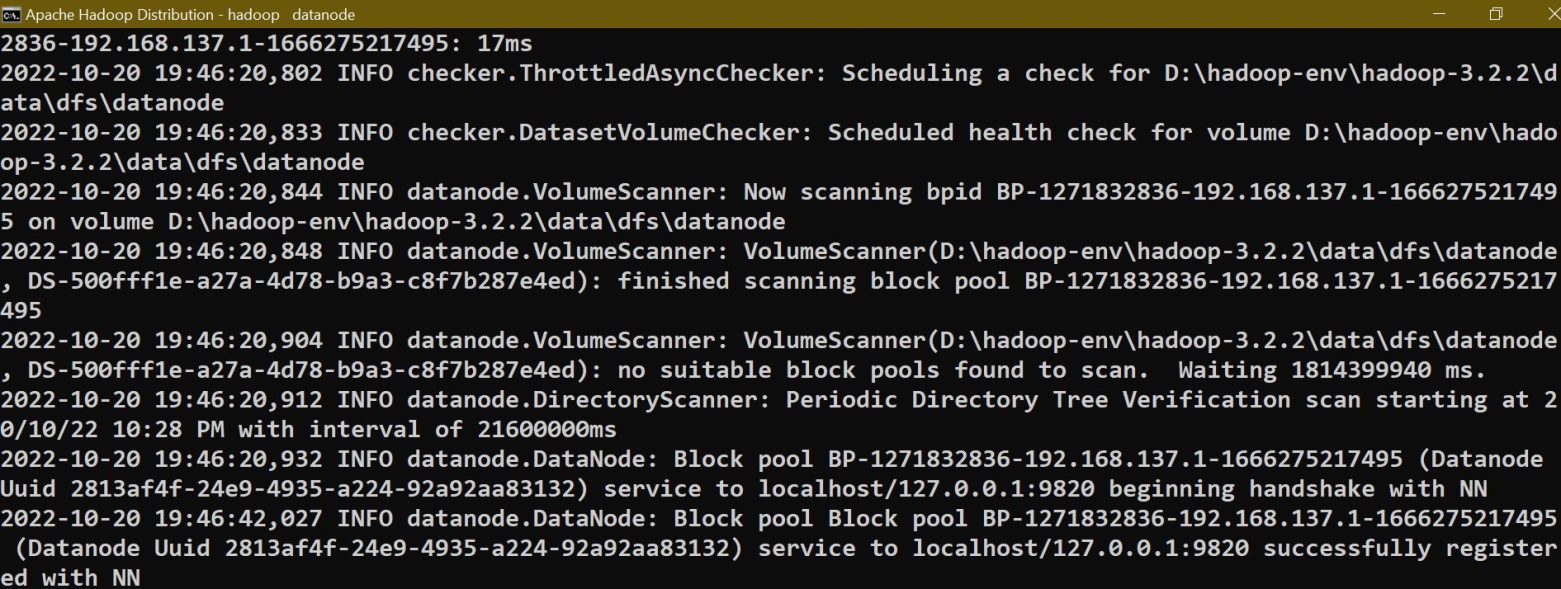
Now navigate to the location “D:\hadoop-env\hadoop-3.2.2\sbin” in powershell and then run the following command to start the Hadoop nodes:

# .\start-dfs.cmd

****

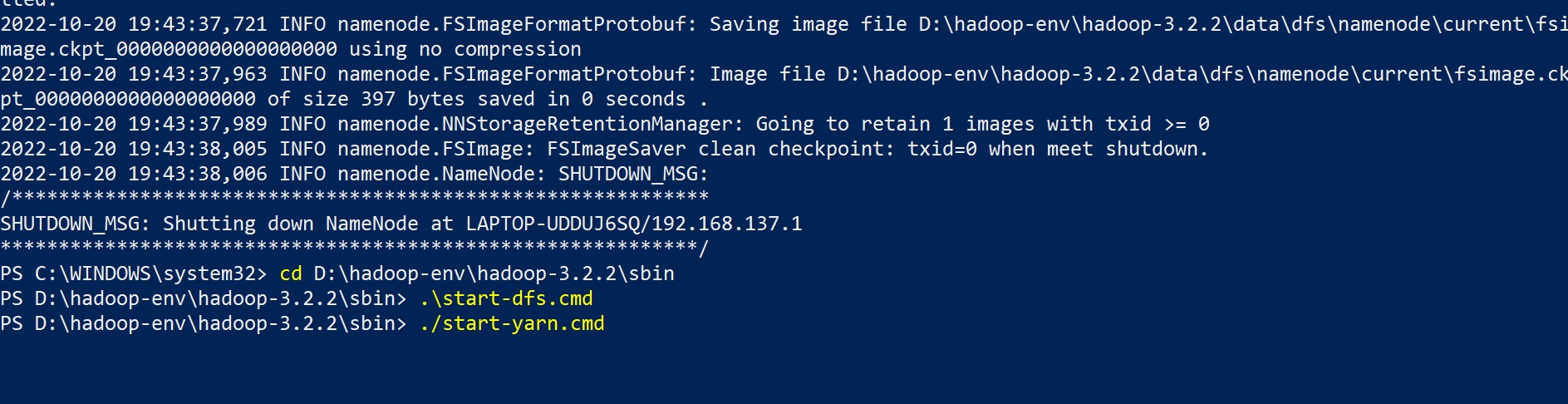
Two command prompt windows will open (one for the **namenode** and one for the **datanode**) as follows:



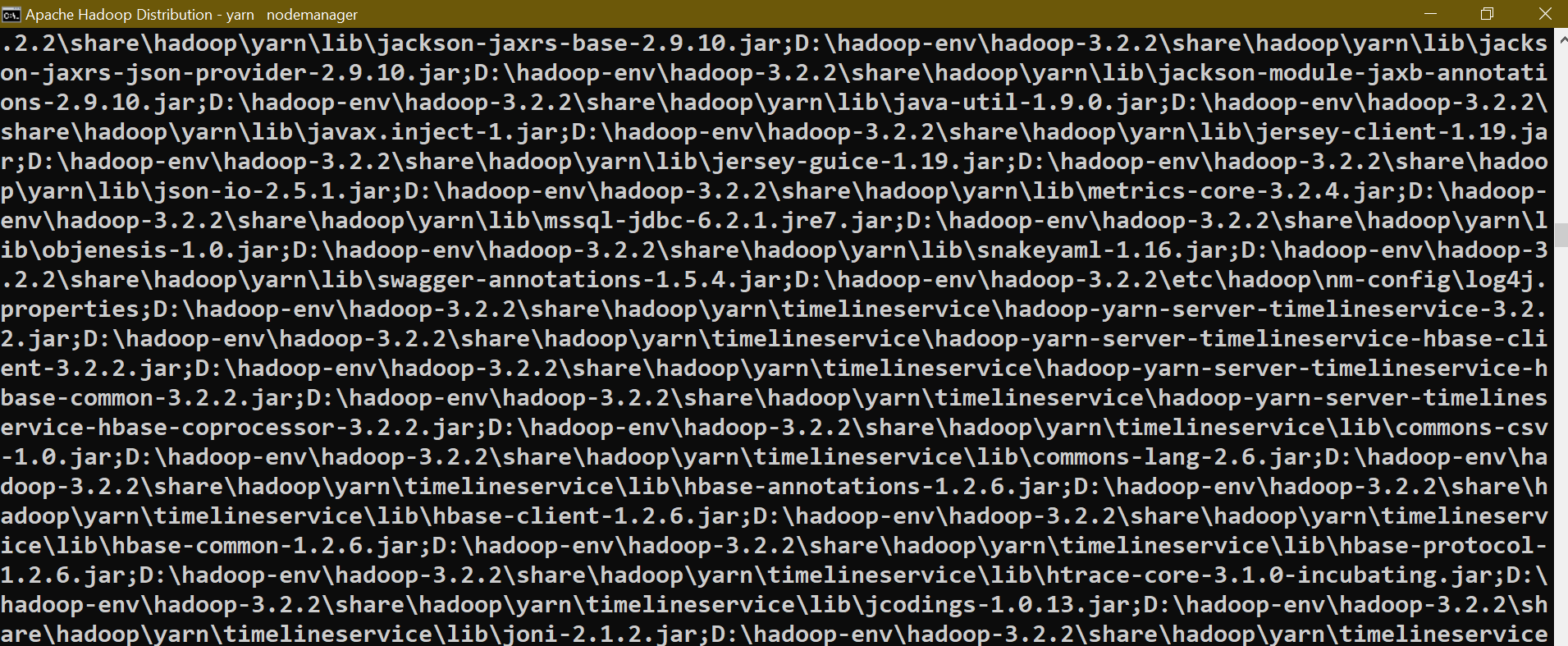


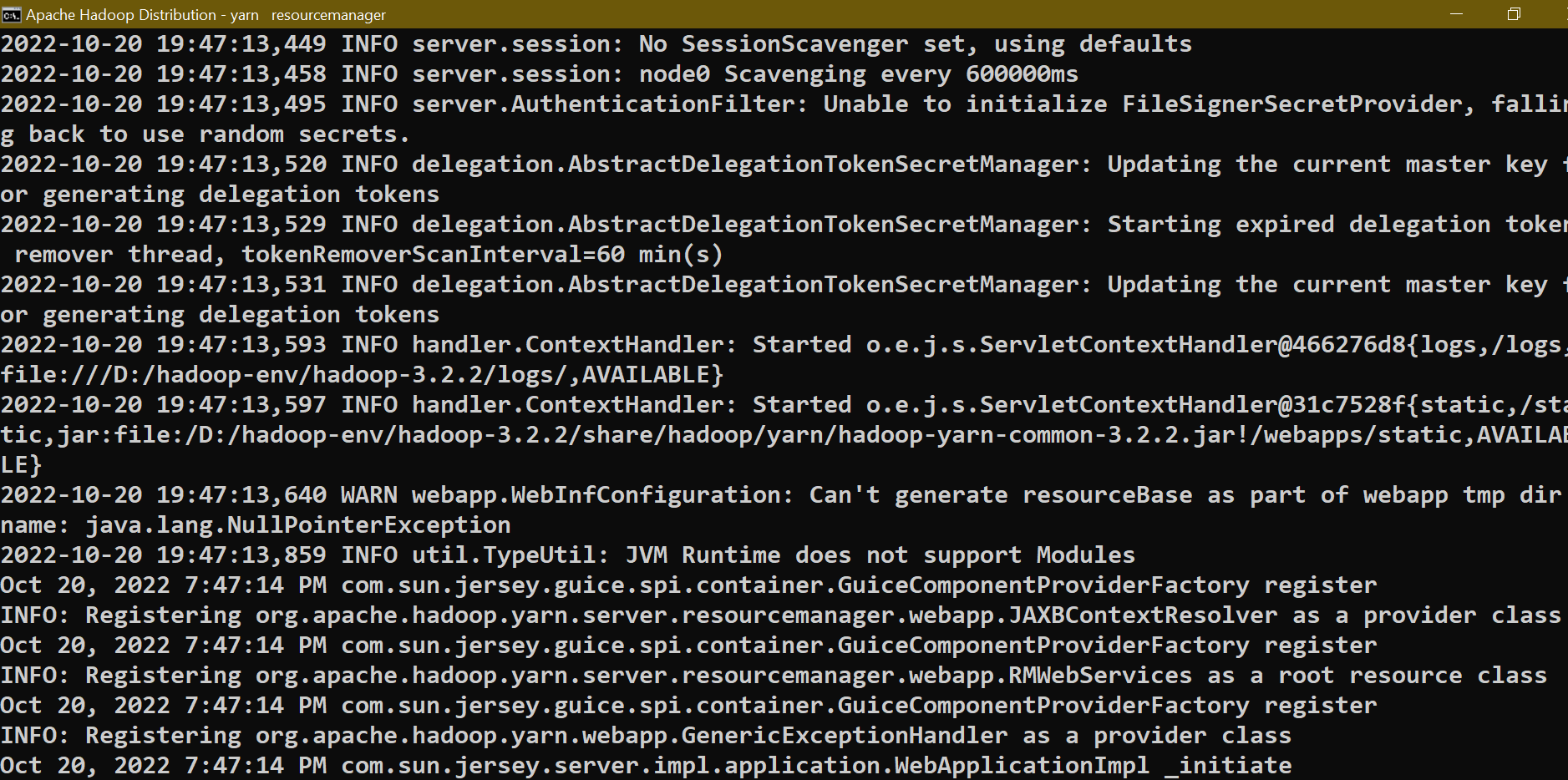
Next, Start the Hadoop Yarn service using the following command:

# ./start-yarn.cmd



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





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

# jps

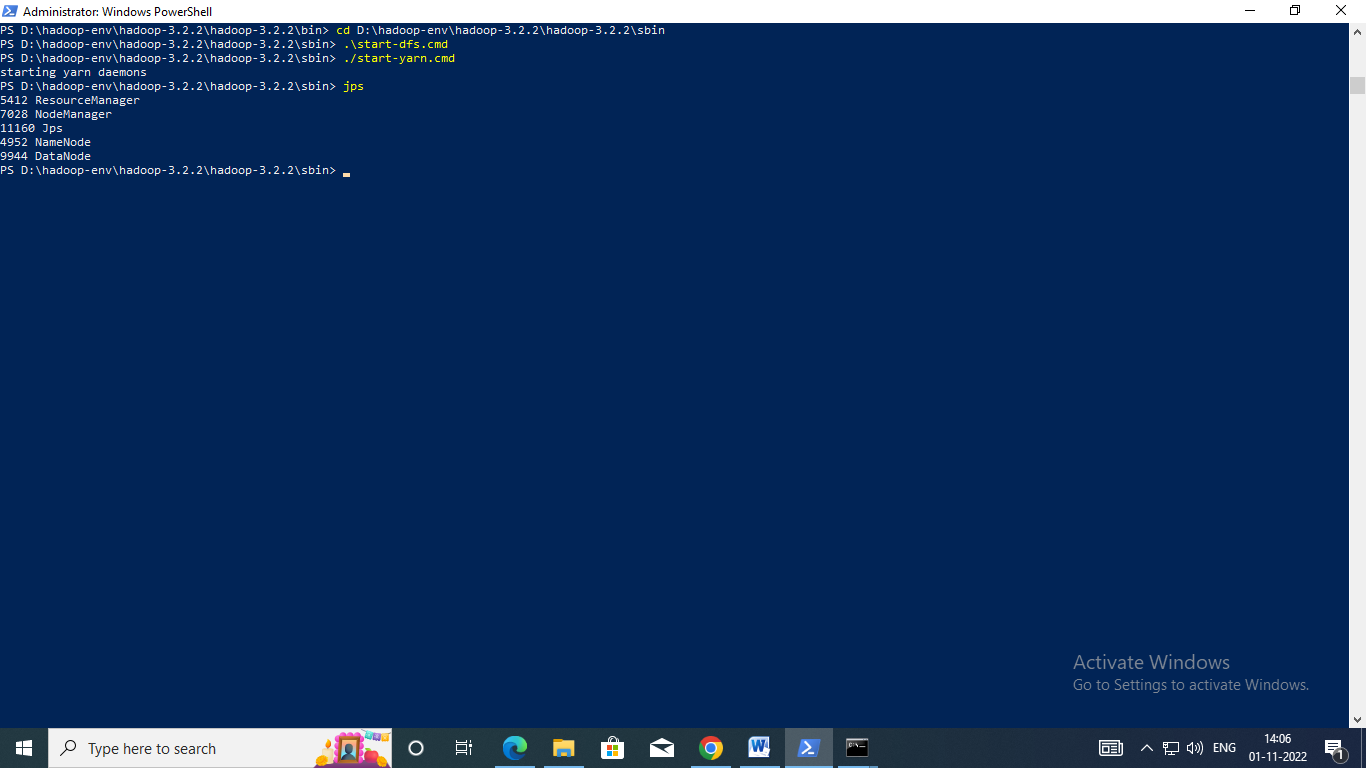
It should display the following services: 9944 DataNode

4960 ResourceManager

4952 NameNode

7028 NodeManager

11160 Jps

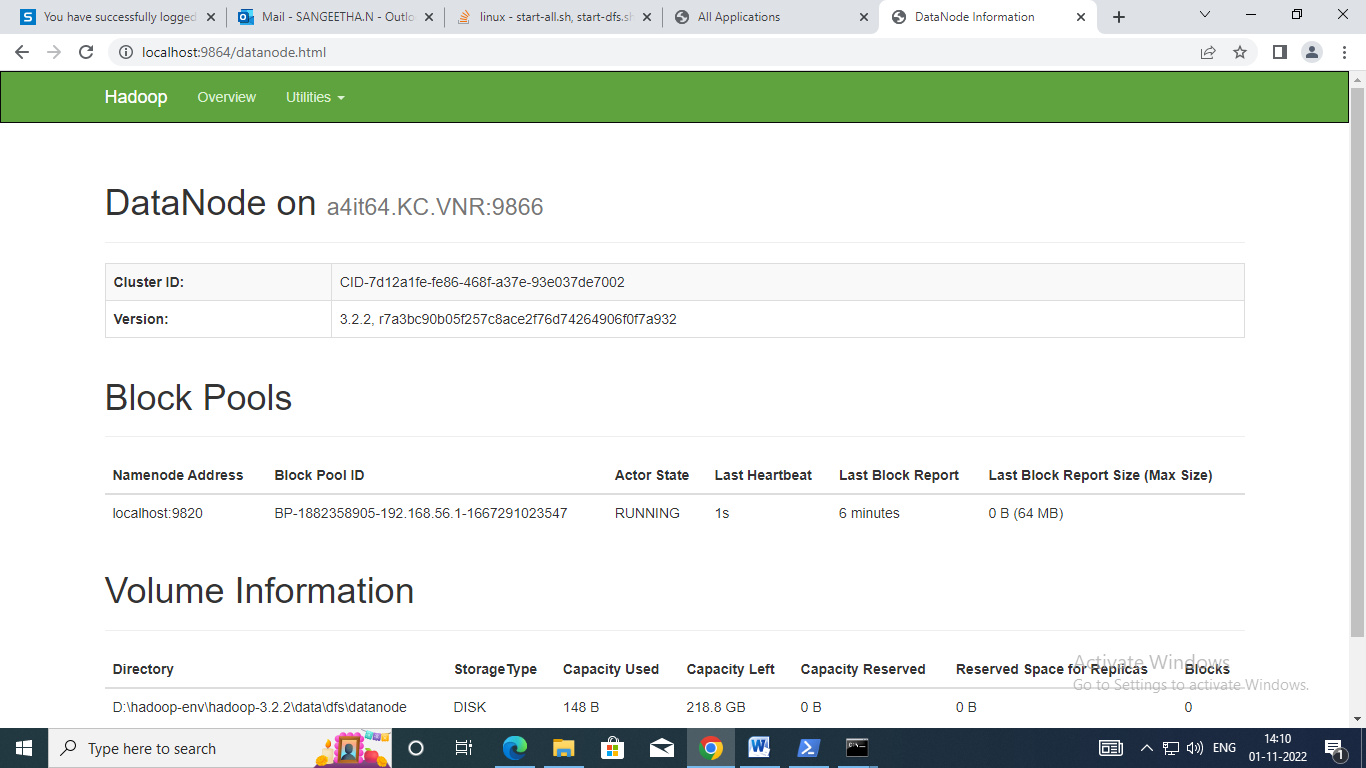


# Output

Hadoop Web UI http://localhost:9870/dfshealth.html



http://localhost:9864/datanode.html



http://localhost:8088/cluster

