

Practical 7

Aim: Install, configure and run Hadoop and HDFS and explore HDFS.

Hadoop for windows 10

(Students for your reference, I am providing the link of the video that you have to watch and follow to install Hadoop.

<https://m.youtube.com/watch?v=g7Qpnmi0Q-s>)

- Pre requisite to install Hadoop is java
- Hadoop supports Java version 8
- Steps:
 - Install java
 - Install Hadoop
 - Configuration files

Step 1: Download Java from oracle website

Search for java SE development kit 8 download

<https://www.oracle.com/java/technologies/downloads/#java8-windows>

The screenshot shows a Microsoft Edge browser window with three tabs open, all titled "Java Archive Downloads - Java SE". The main content area displays the Oracle Java SE Development Kit 8u241 download page. At the top, there are links for "Products", "Industries", "Resources", "Customers", "Partners", "Developers", "Events", a search bar, and account-related buttons. Below this, a table lists Java archive downloads for Windows x64, showing file sizes and download links. The table has columns for "Product / File Description", "File Size", and "Download". The first row shows "server-jre-8u251-windows-x64.tar.gz" with a size of 49 MB. Other rows include "jdk-8u241-linux-arm32-vfp-hfllt.tar.gz" (72.94 MB), "jdk-8u241-linux-arm64-vfp-hfllt.tar.gz" (69.83 MB), "jdk-8u241-linux-i586.rpm" (171.28 MB), and "jdk-8u241-linux-i586.tar.gz" (186.1 MB). The bottom of the screen shows the Windows taskbar with a search bar, pinned icons for File Explorer, Mail, Edge, Google Chrome, and Word, and system status icons.

| Product / File Description | File Size | Download |
|--|-----------|--|
| server-jre-8u251-windows-x64.tar.gz | 49 MB | server-jre-8u251-windows-x64.tar.gz |
| jdk-8u241-linux-arm32-vfp-hfllt.tar.gz | 72.94 MB | jdk-8u241-linux-arm32-vfp-hfllt.tar.gz |
| jdk-8u241-linux-arm64-vfp-hfllt.tar.gz | 69.83 MB | jdk-8u241-linux-arm64-vfp-hfllt.tar.gz |
| jdk-8u241-linux-i586.rpm | 171.28 MB | jdk-8u241-linux-i586.rpm |
| jdk-8u241-linux-i586.tar.gz | 186.1 MB | jdk-8u241-linux-i586.tar.gz |

This screenshot shows the Oracle Java Archive Downloads page for Java SE Runtime Environment 8u241. The page lists download links for different operating systems:

| Platform | File Size | Download Link |
|-------------------------------------|-----------|--|
| macOS x64 | 254.06 MB | jdk-8u241-macosx-x64.dmg |
| Solaris SPARC 64-bit (SVR4 package) | 133.01 MB | jdk-8u241-solaris-sparcv9.tar.Z |
| Solaris SPARC 64-bit | 94.24 MB | jdk-8u241-solaris-sparcv9.tar.gz |
| Solaris x64 (SVR4 package) | 133.8 MB | jdk-8u241-solaris-x64.tar.Z |
| Solaris x64 | 92.01 MB | jdk-8u241-solaris-x64.tar.gz |
| Windows x86 | 200.86 MB | jdk-8u241-windows-i586.exe |
| Windows x64 | 210.92 MB | jdk-8u241-windows-x64.exe |

Java SE Runtime Environment 8u241

This software is licensed under the [Oracle Technology Network License Agreement for Oracle Java SE](#)

This screenshot shows the same Java Archive Downloads page as above, but with a modal dialog box overlaid. The dialog requires the user to accept the Oracle Technology Network License Agreement for Oracle Java SE before proceeding.

The dialog contains the following text:

You must accept the [Oracle Technology Network License Agreement for Oracle Java SE](#) to download this software.

I reviewed and accept the Oracle Technology Network License Agreement for Oracle Java SE
Required

You will be redirected to the login screen in order to download the file.

[Download jdk-8u241-windows-x64.exe](#)

Java SE Runtime Environment 8u241

This screenshot shows the Java Archive Downloads page again, with the modal dialog still present. The dialog has been partially closed, showing the underlying download links for the Windows x64 platform.

| Platform | File Size | Download Link |
|-------------|-----------|---|
| Windows x64 | 210.92 MB | jdk-8u241-windows-x64.exe |

Oracle registration

Provide your email address and password if already registered, else register.

Jdk is downloaded

Step 2: Download Hadoop for the local system

<https://hadoop.apache.org/releases.html>

The screenshot shows a web browser window with the URL <https://hadoop.apache.org/releases.html>. The page displays a list of Apache Hadoop releases:

- Release 2.10.0 available** (2019 Oct 29)
This is the first stable release of Apache Hadoop 2.10 line. It contains 362 bug fixes, improvements and enhancements since 2.9.0.
Users are encouraged to read the [overview of major changes](#) since 2.9.0. For details of 362 bug fixes, improvements, and other enhancements since the previous 2.9.0 release, please check [release notes](#) and [changelog](#).
- Release 3.1.3 available** (2019 Oct 21)
This is the third stable release of Apache Hadoop 3.1 line. It contains 246 bug fixes, improvements and enhancements since 3.1.2.
Users are encouraged to read the [overview of major changes](#) since 3.1.2. For details of the bug fixes, improvements, and other enhancements since the previous 3.1.2 release, please check [release notes](#) and [changelog](#).
- Release 3.2.1 available** (2019 Sep 22)
This is the second stable release of Apache Hadoop 3.2 line. It contains 493 bug fixes, improvements and enhancements since 3.2.0.

Release 3.1.3 available

The screenshot shows a web browser window with the URL <https://hadoop.apache.org/release/3.1.3.html>. The page displays the Apache Hadoop 3.1.3 release information:

Release 3.1.3 available (2019 Oct 21)

This is the third stable release of Apache Hadoop 3.1 line. It contains 246 bug fixes, improvements and enhancements since 3.1.2.

Users are encouraged to read the [overview of major changes](#) since 3.1.2. For details of the bug fixes, improvements, and other enhancements since the previous 3.1.2 release, please check [release notes](#) and [changelog](#).

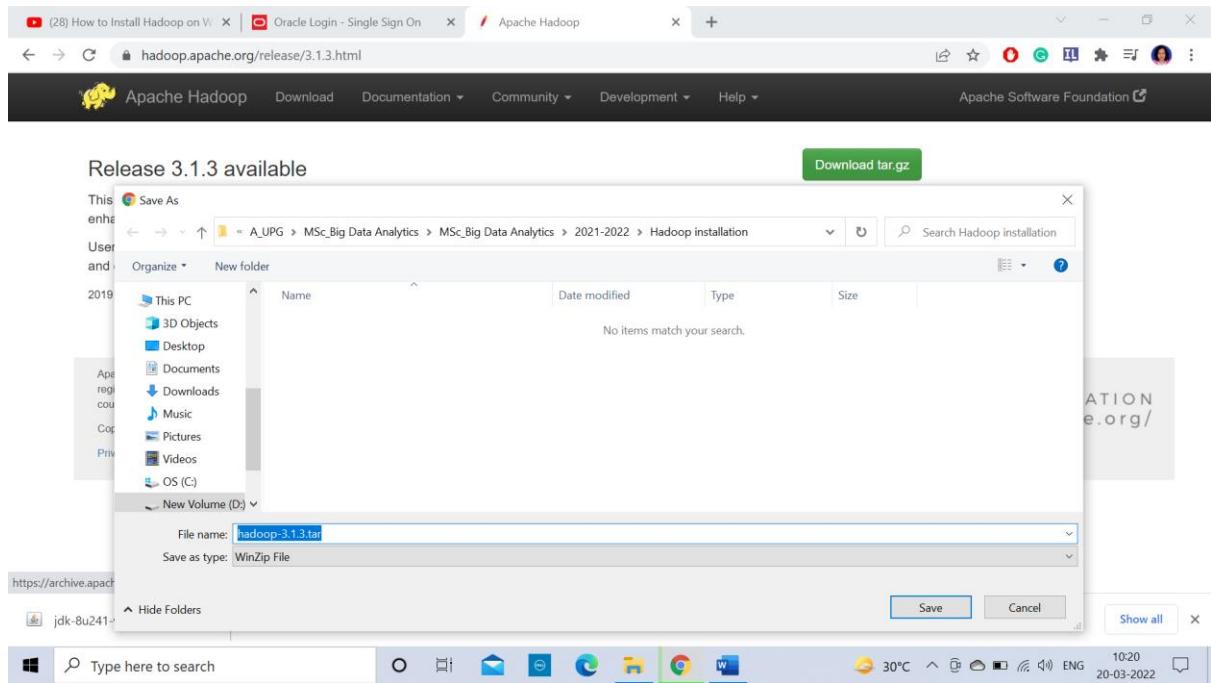
Download links and documentation:

- [Download tar.gz](#) (checksum signature)
- [Download src](#) (checksum signature)
- [Documentation](#)

Apache Software Foundation logo and links:

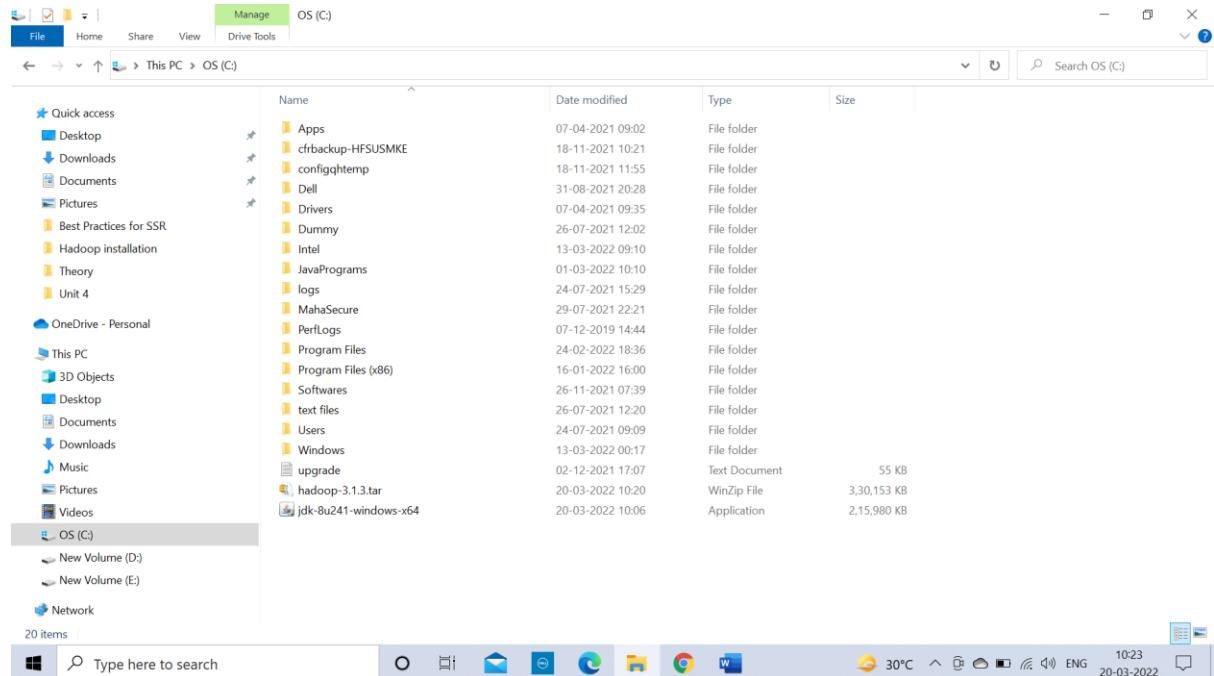
Apache Hadoop, Hadoop, Apache, the Apache feather logo, and the Apache Hadoop project logo are either registered trademarks or trademarks of the Apache Software Foundation in the United States and other countries.
Copyright © 2006-2022 The Apache Software Foundation
[Privacy policy](#)

The screenshot shows a Windows taskbar with the following pinned icons from left to right: File Explorer, Task View, Mail, Edge, File Explorer, Google Chrome, and Word.

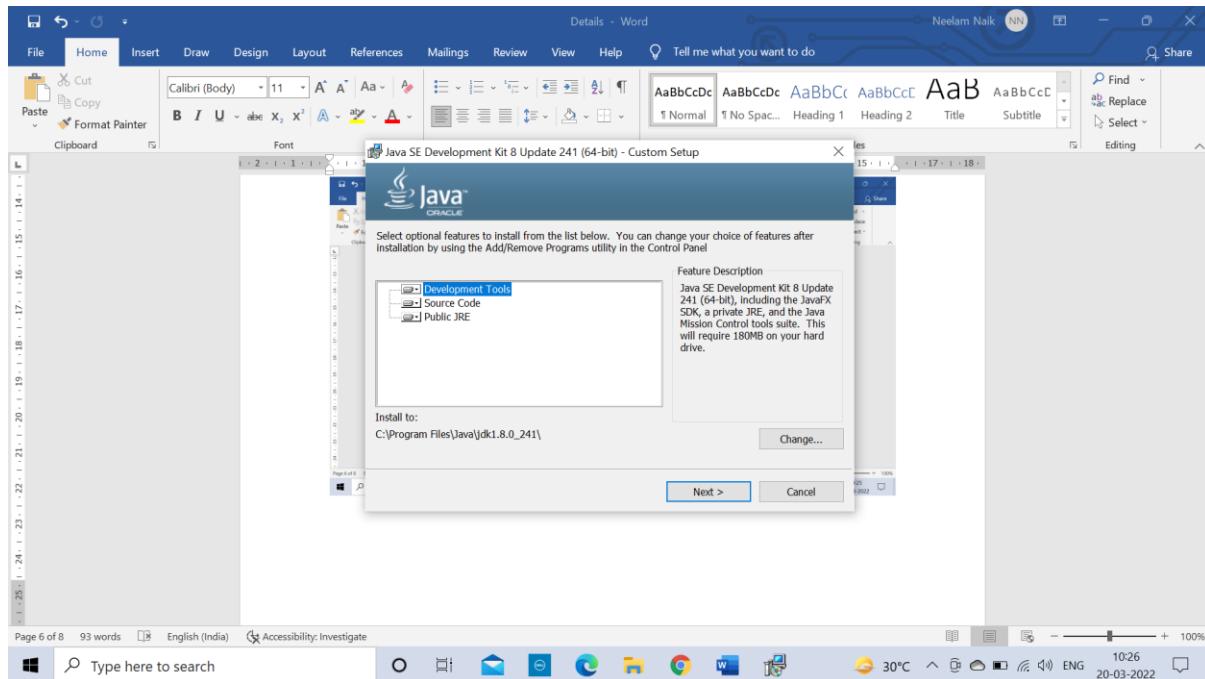
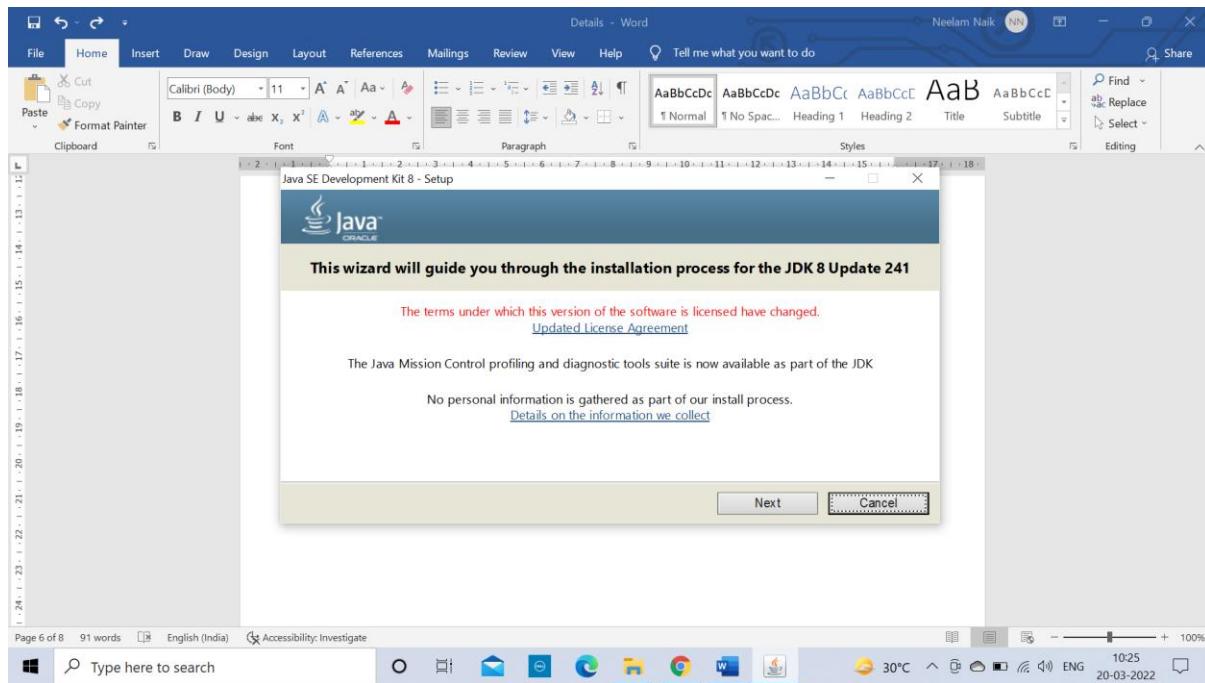


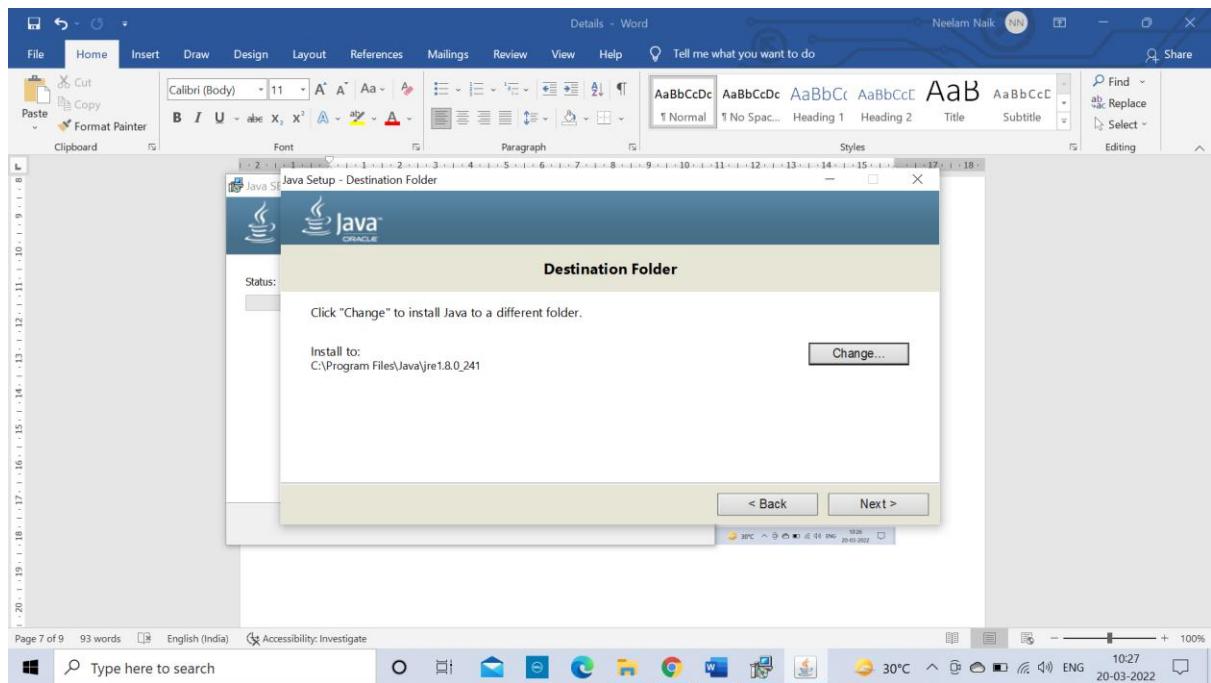
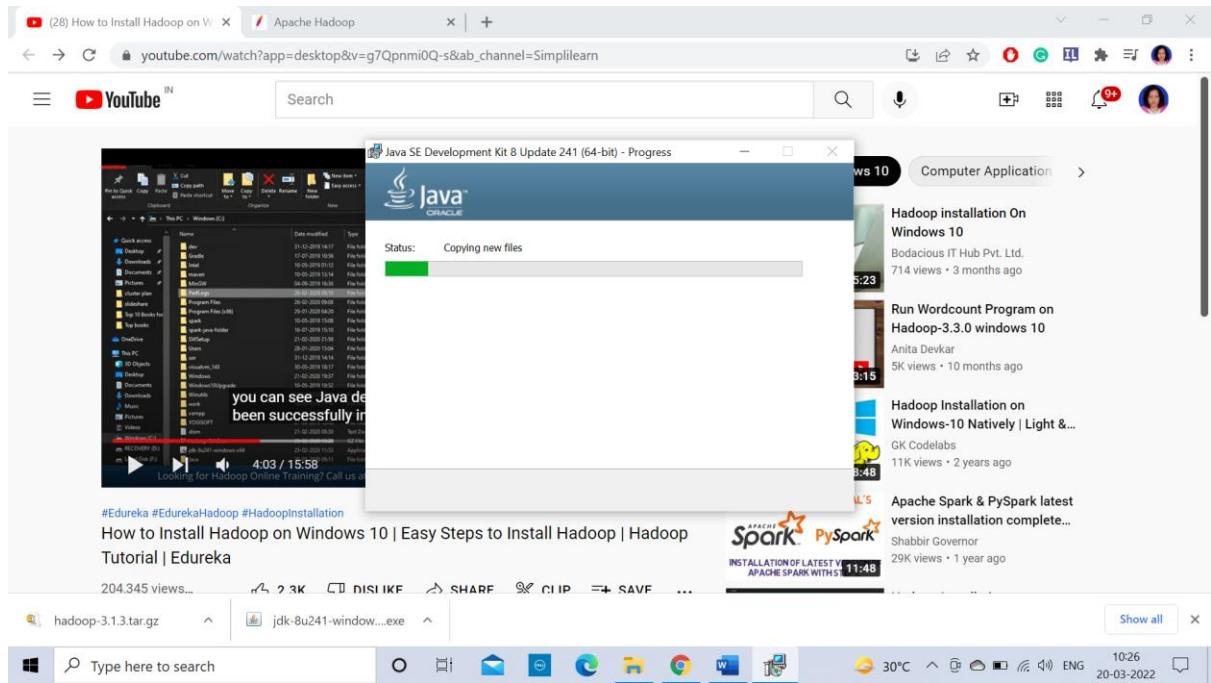
Hadoop downloaded

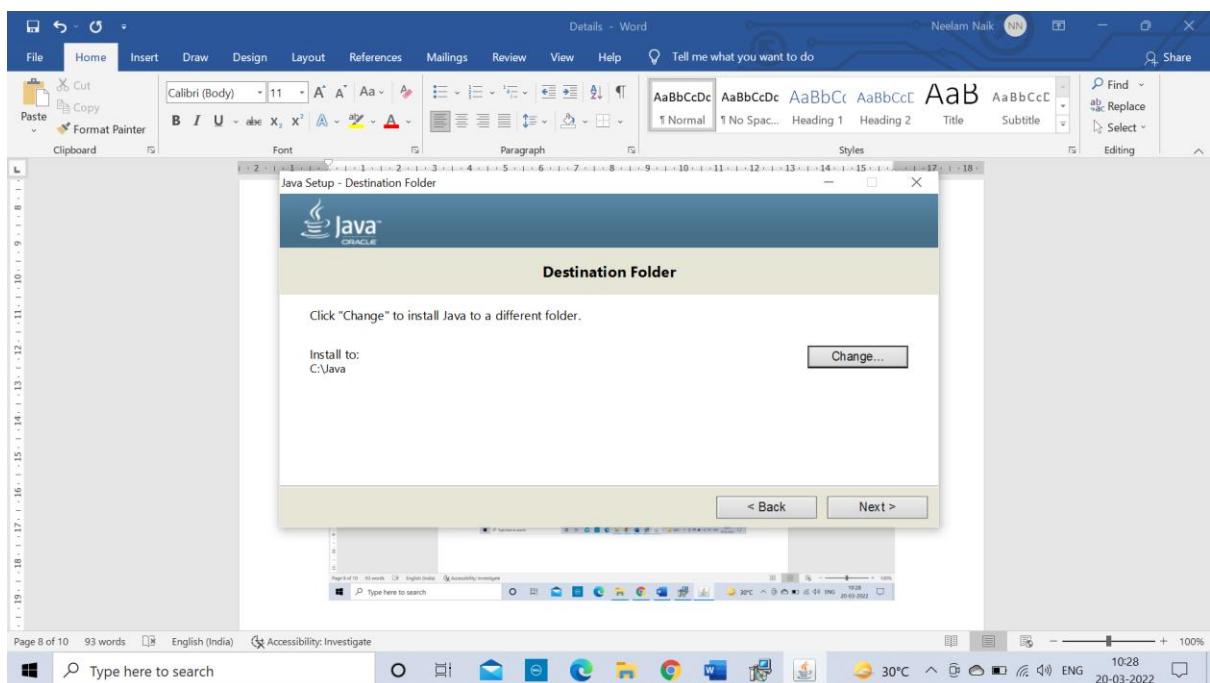
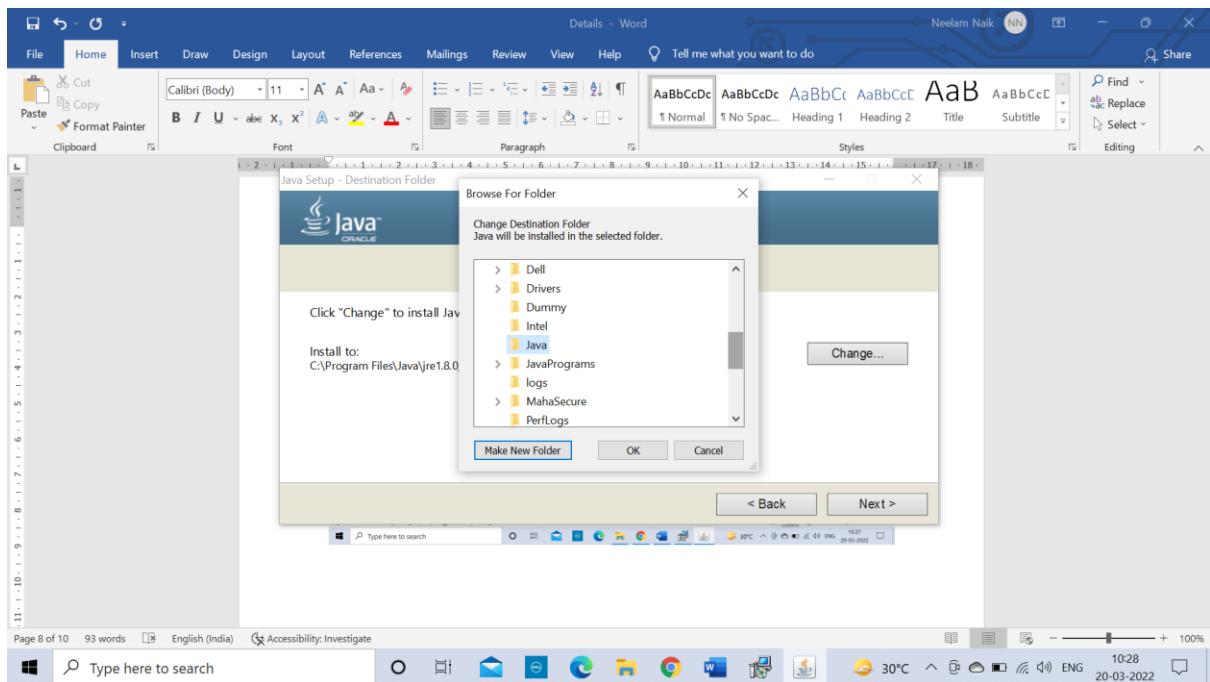
Move files to c drive

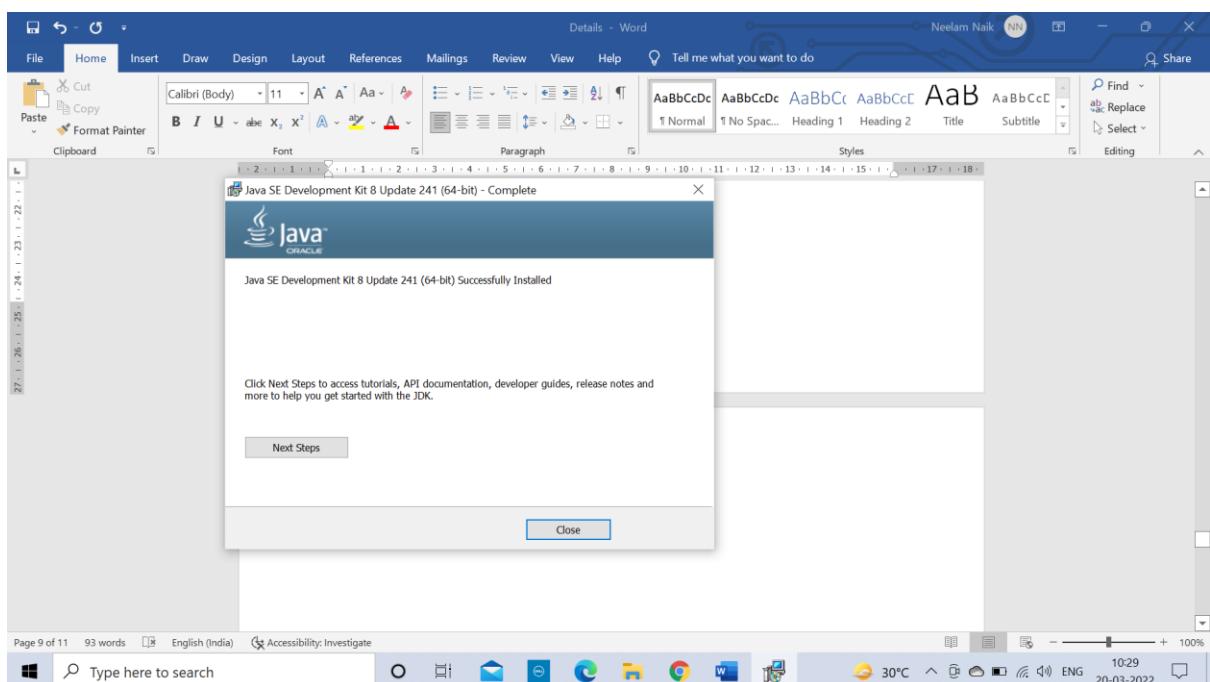
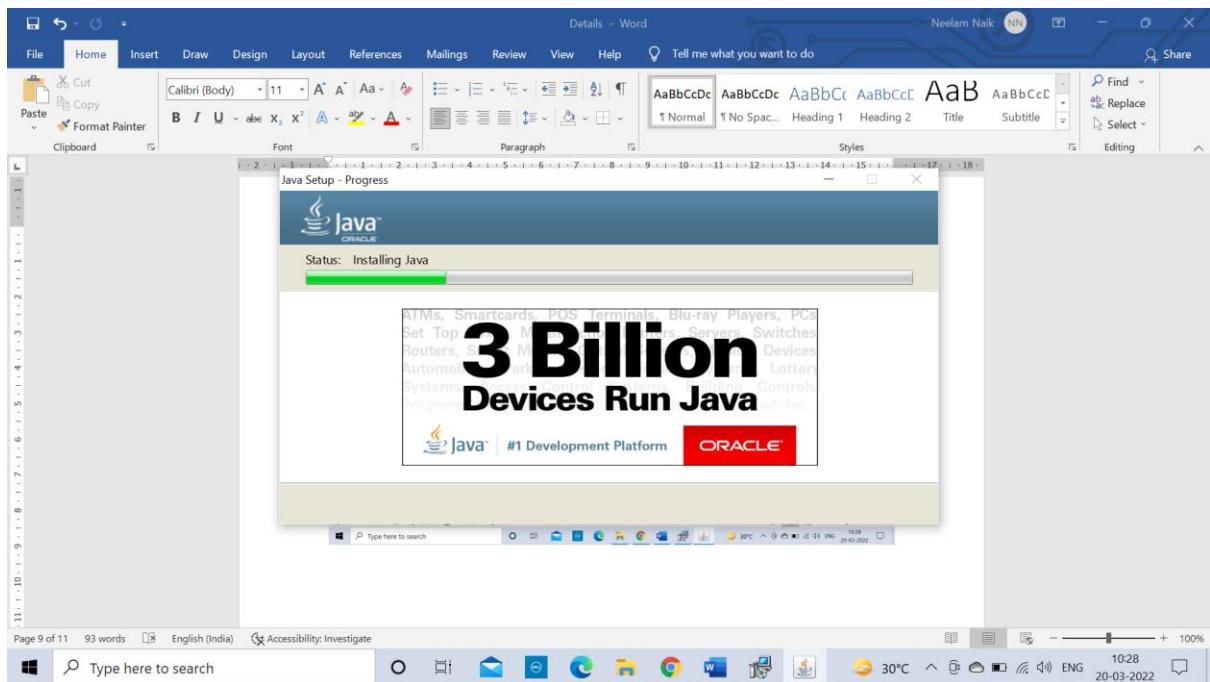


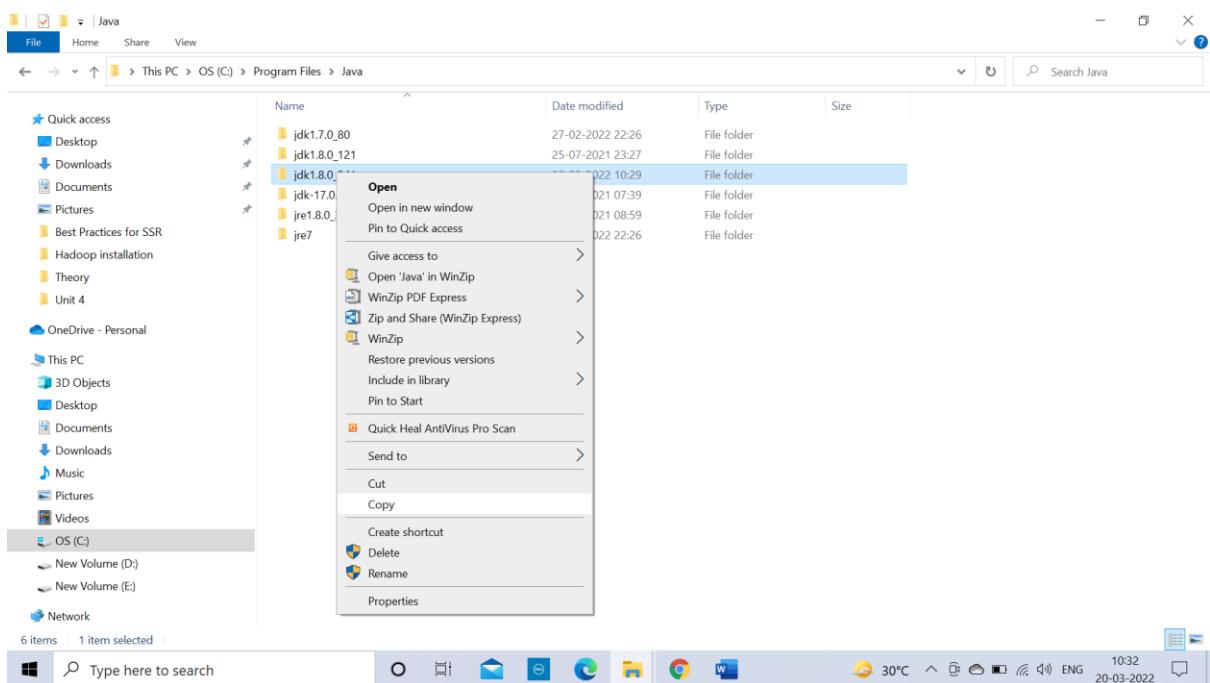
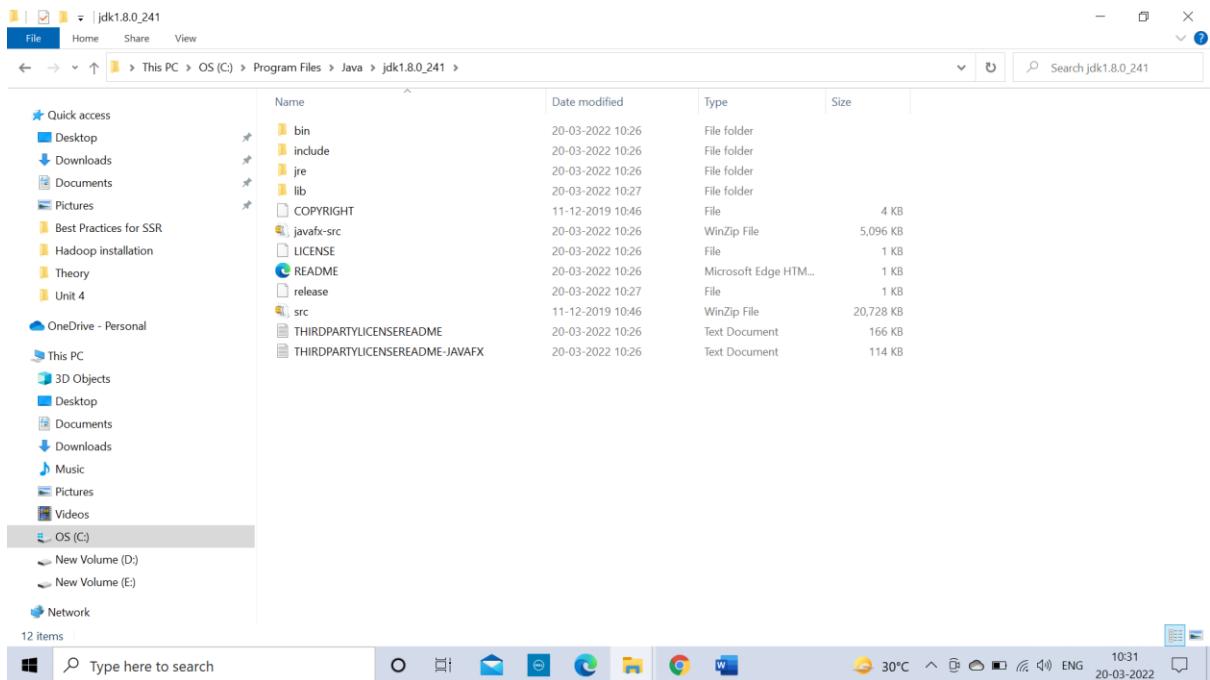
Step 3: Install Java



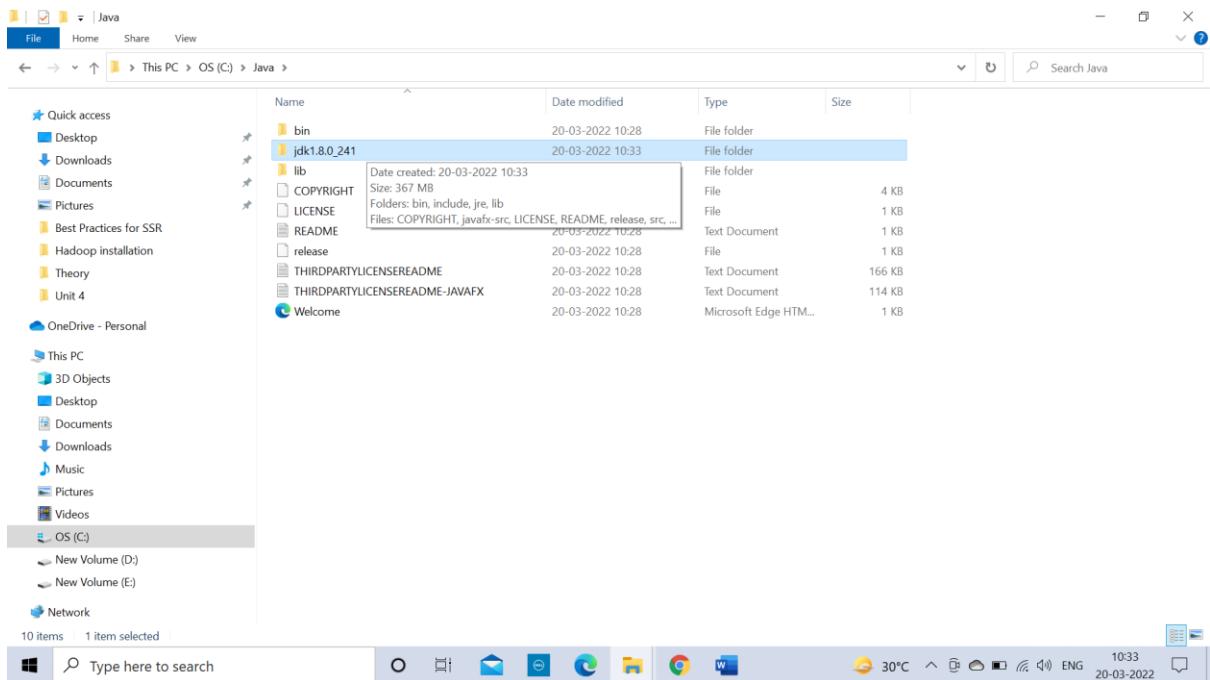








Paste it here

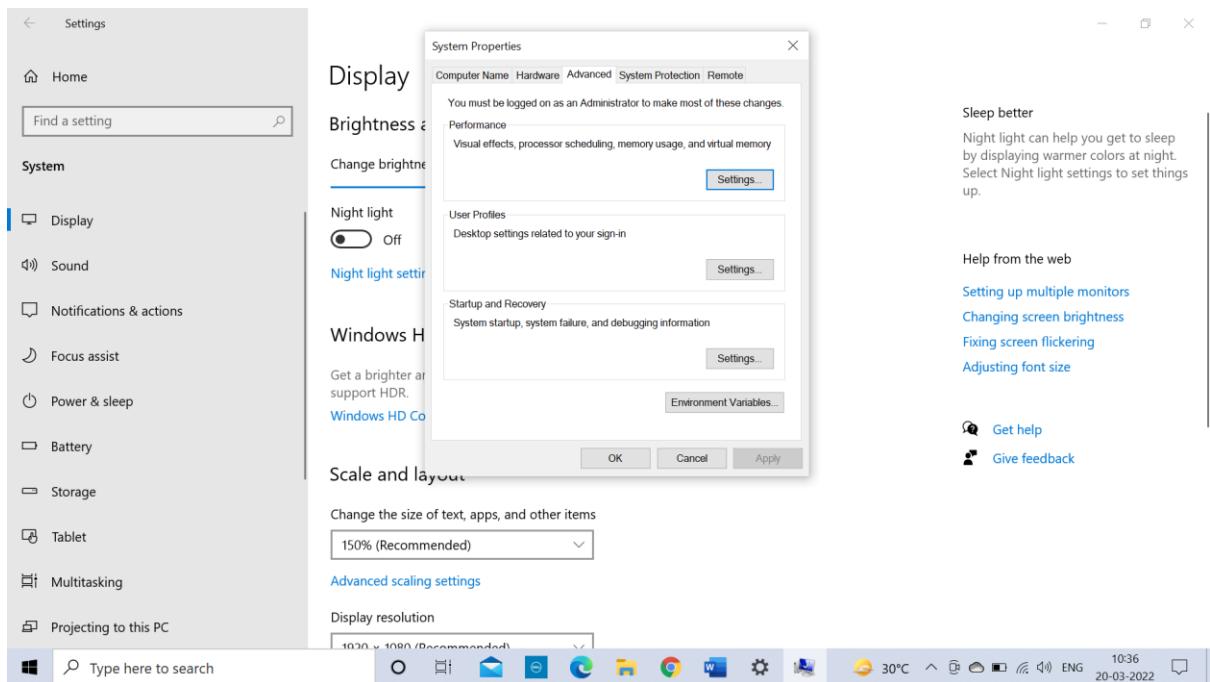


Now java will be available at:

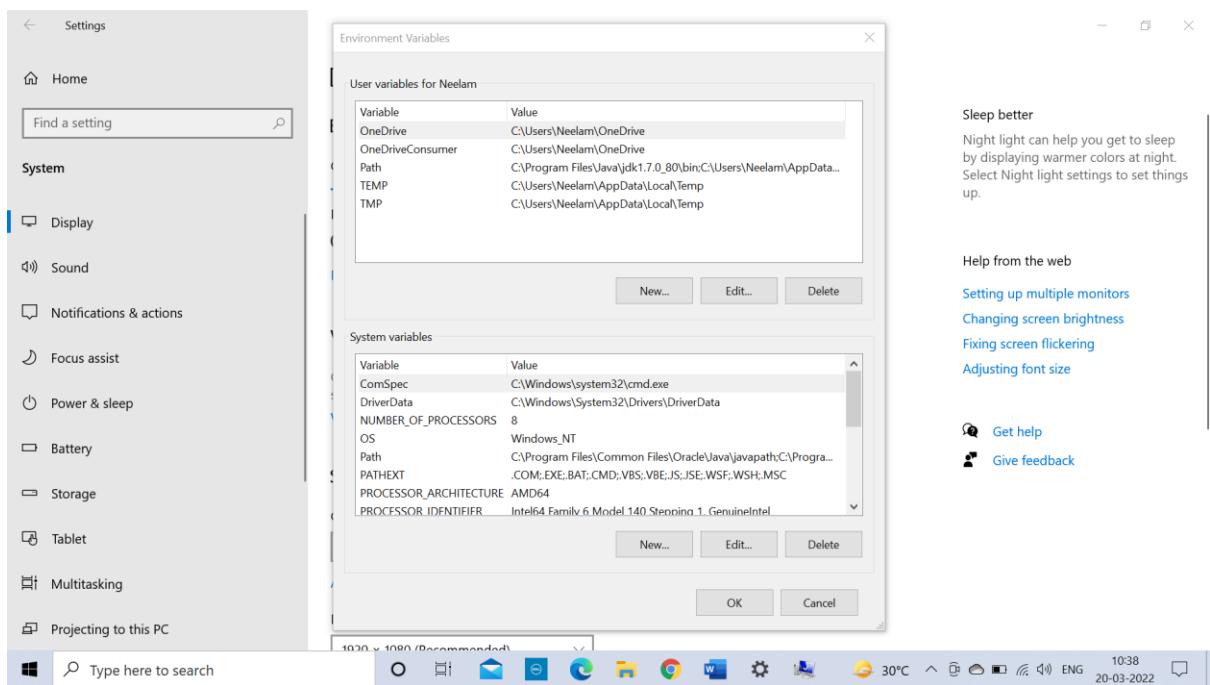
C:\Java

Step 4: Setting environment variables for java

Windows->settings->system->environment variables for system->edit the system environment variables

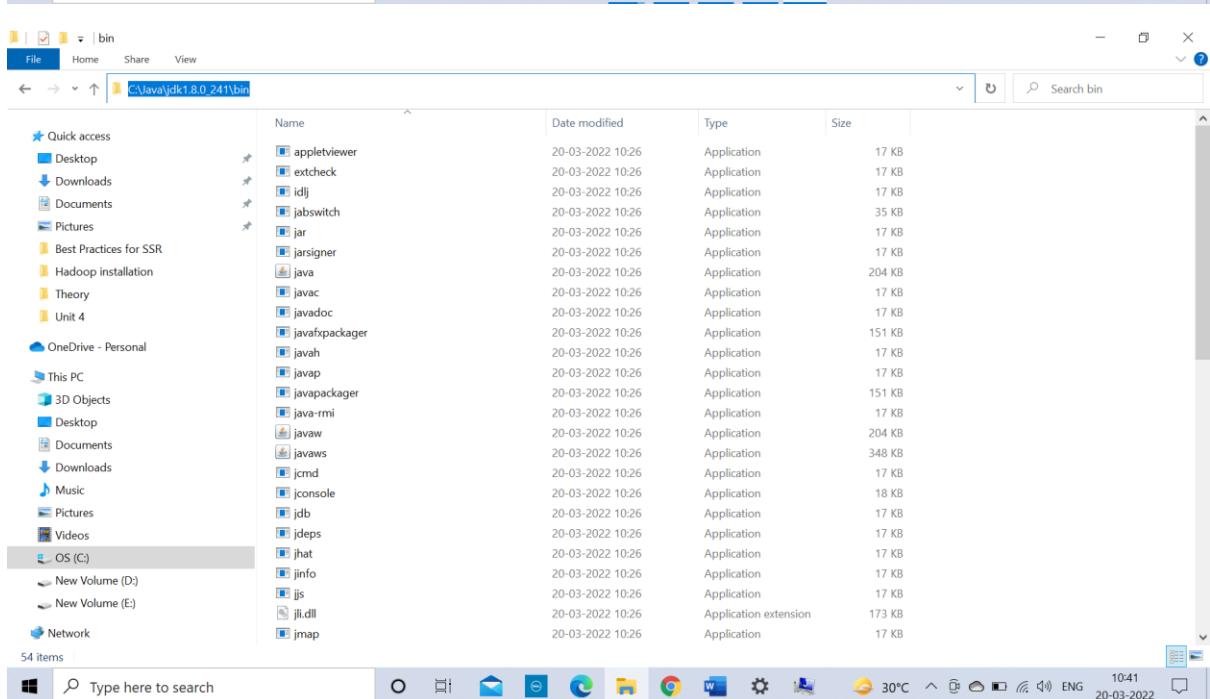
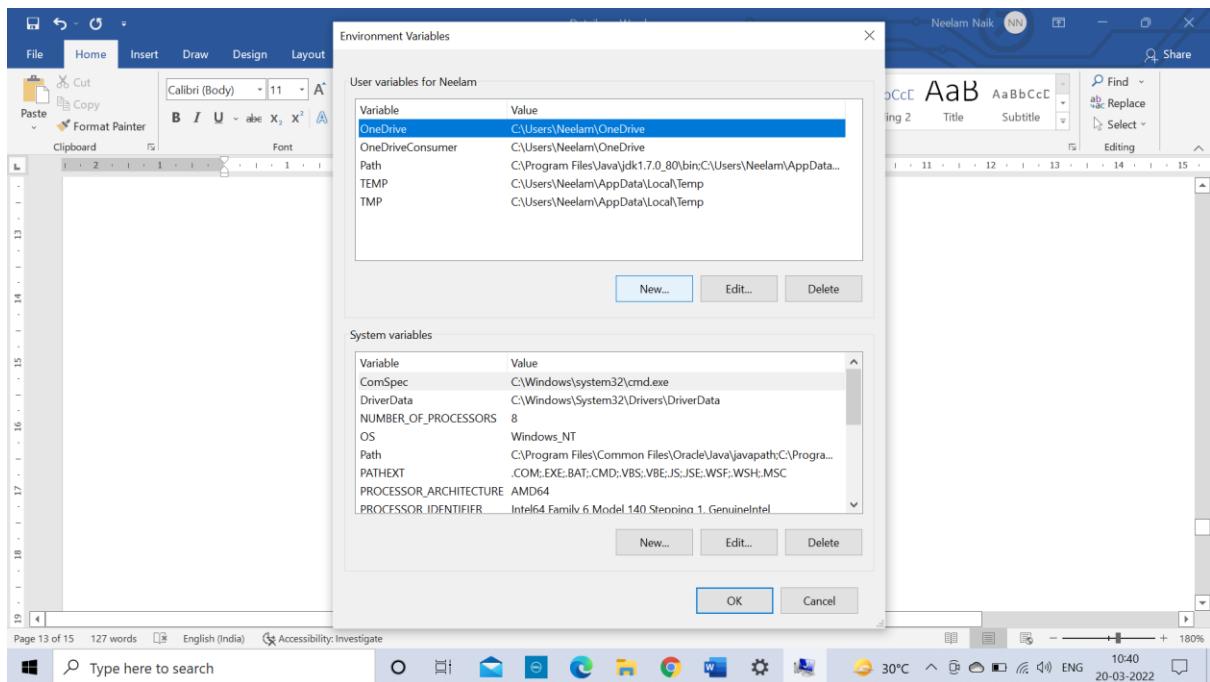


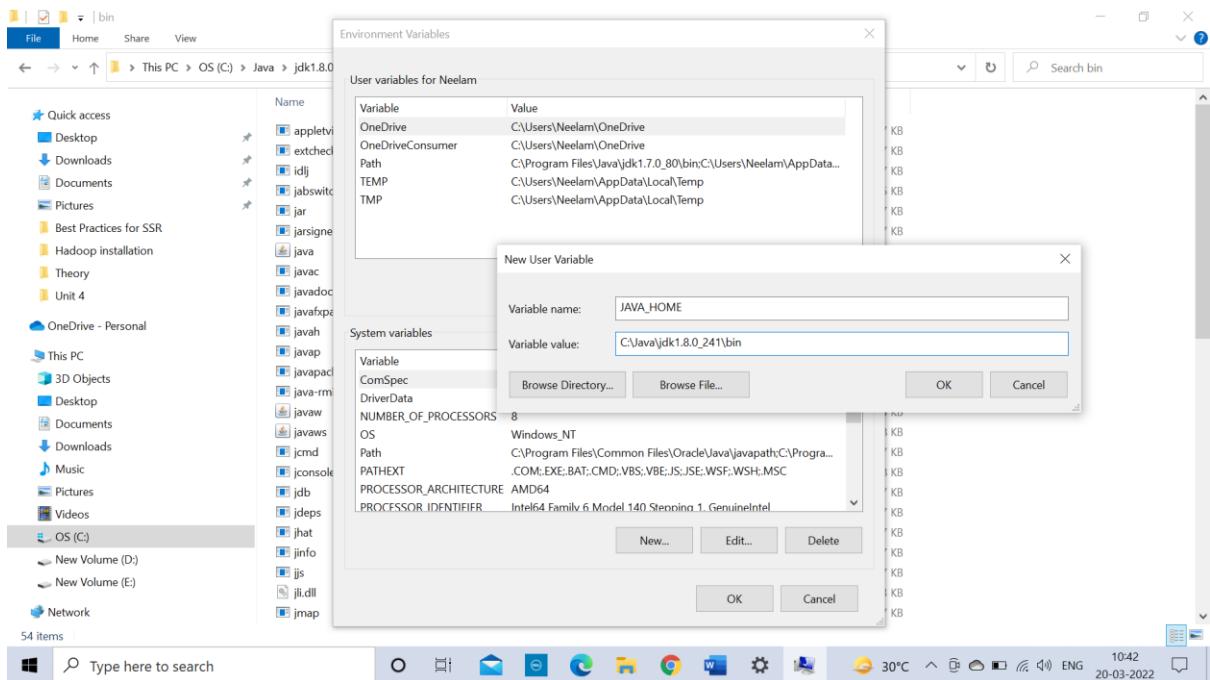
Click on environment variables



Set java home and path for java

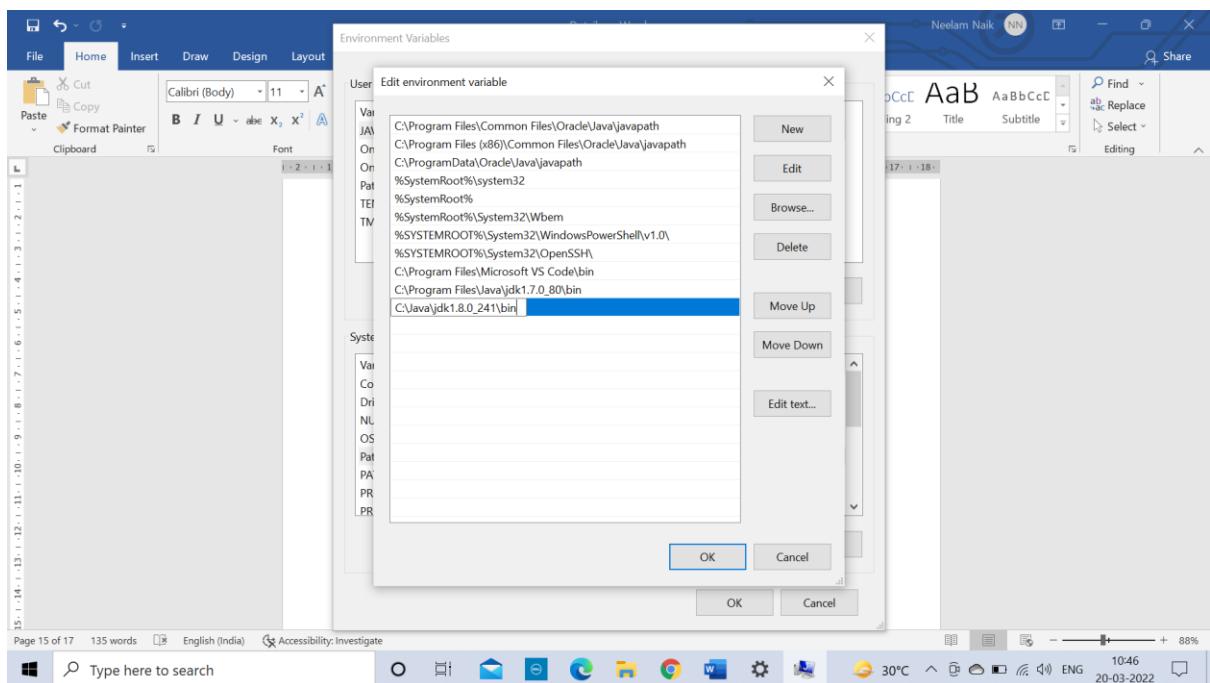
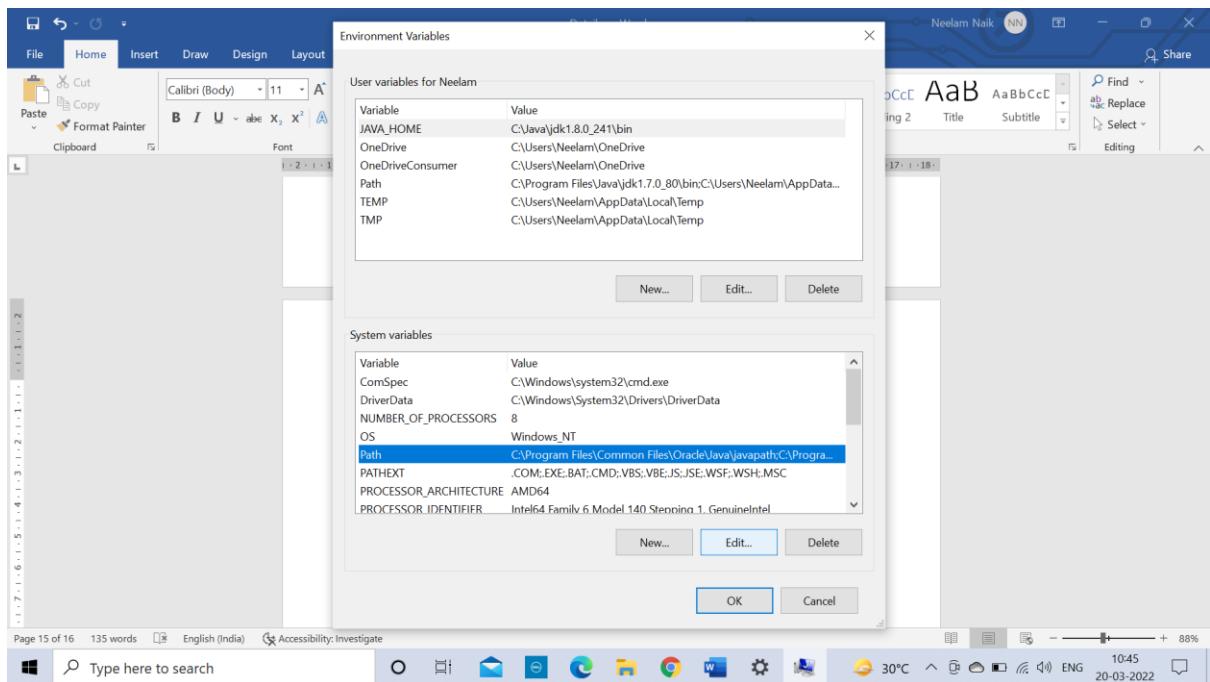
Set user variables





Set system variables

System variable->path->edit->new

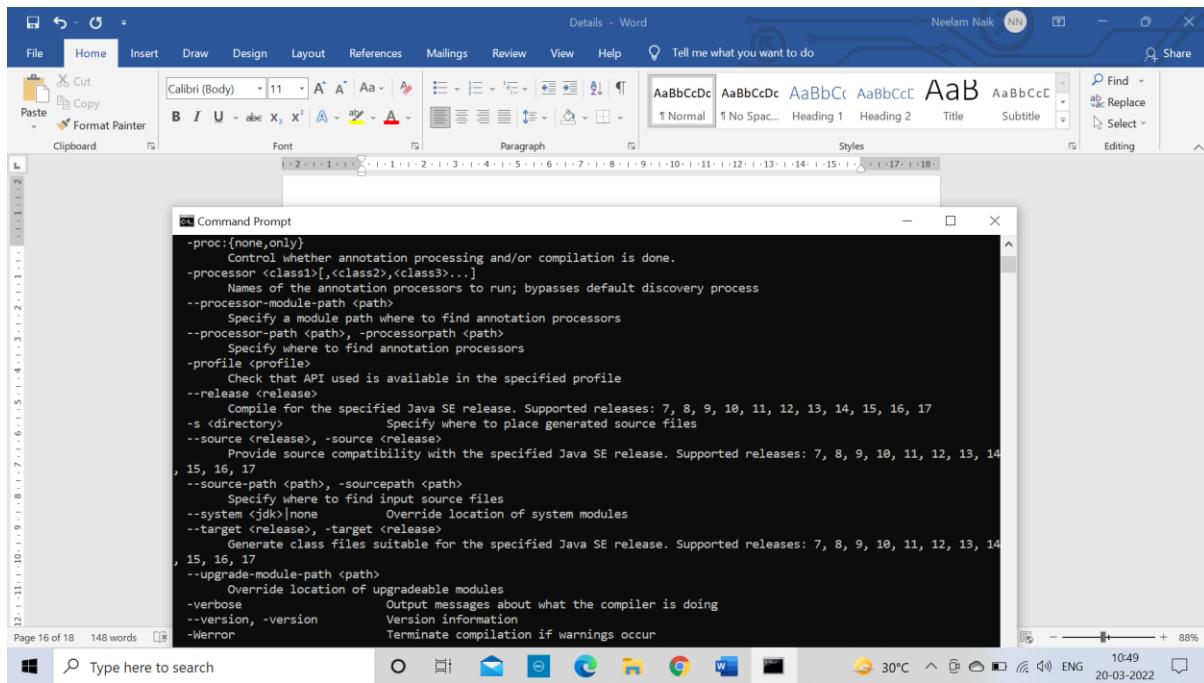
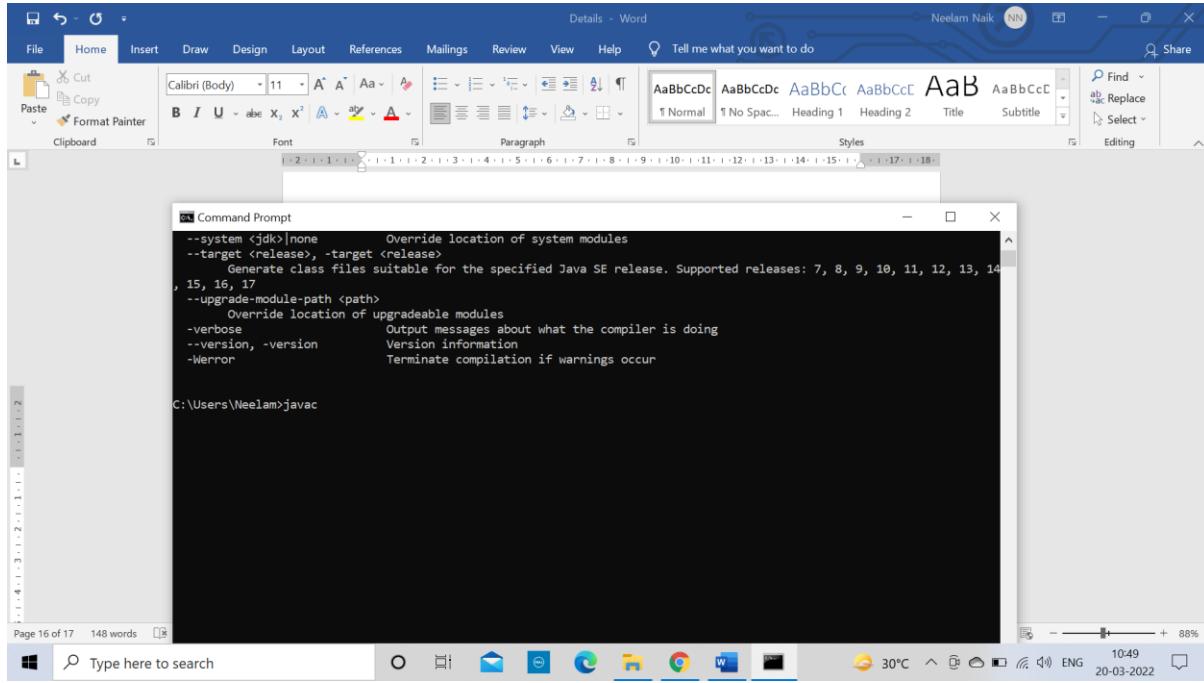


Click Ok-> OK->close

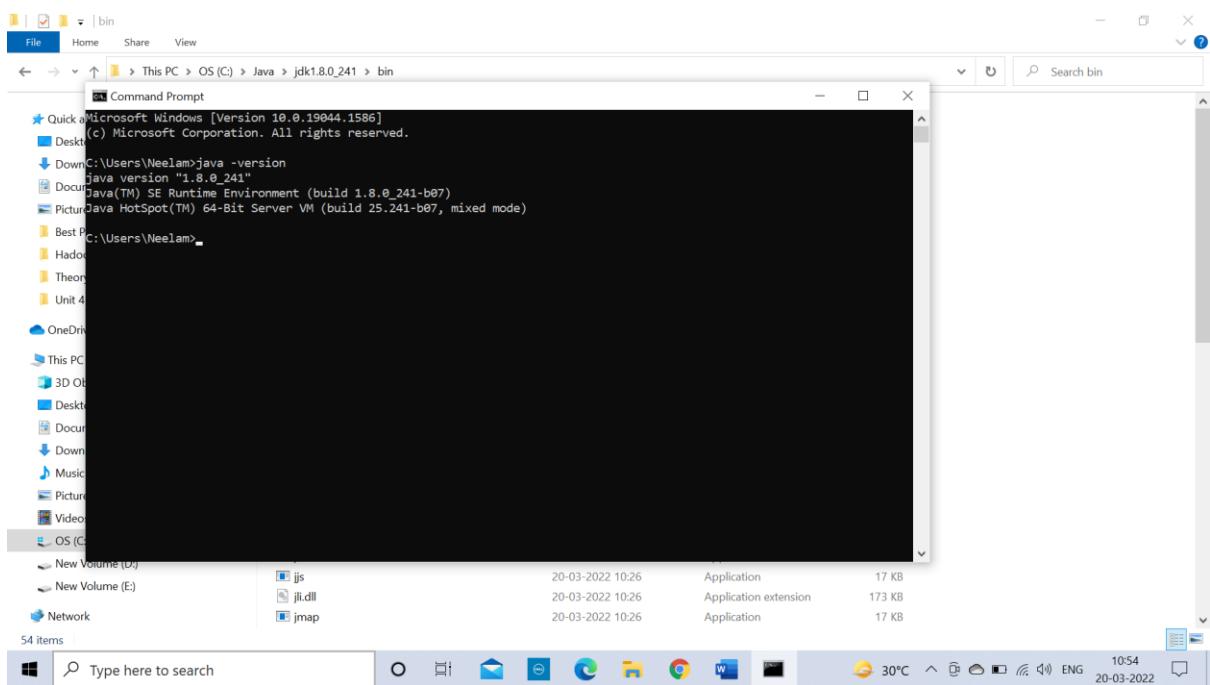
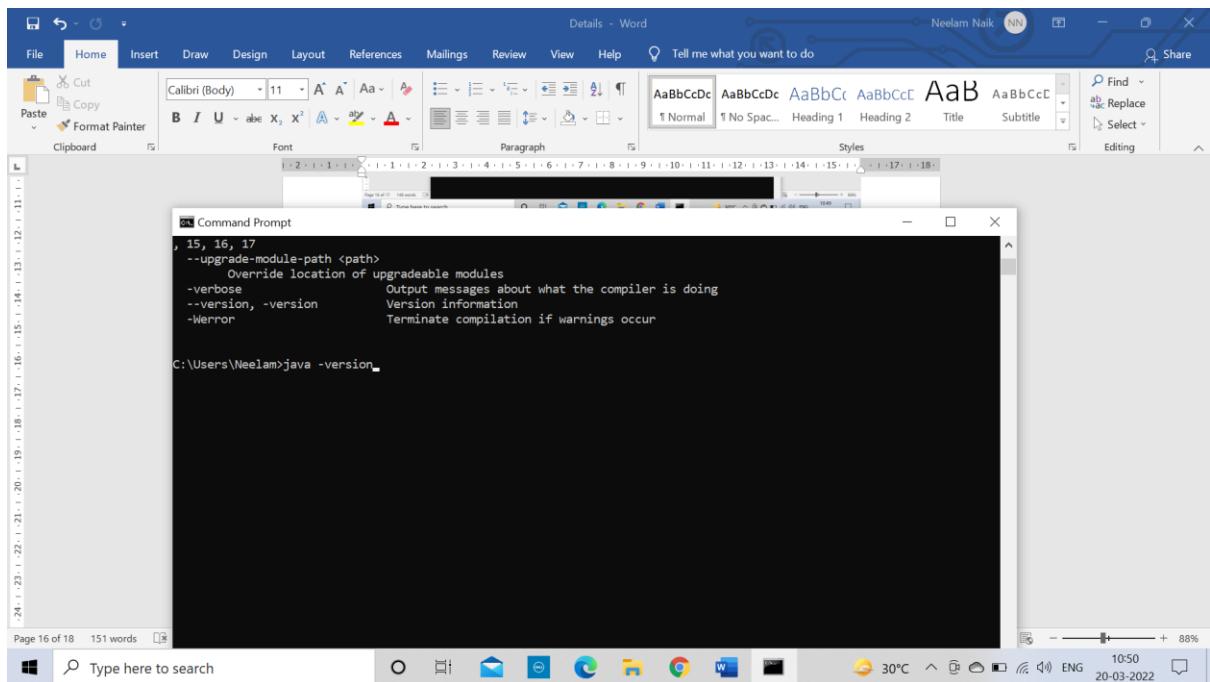
Java successfully installed

Step 5: Java version checking

Go to command prompt

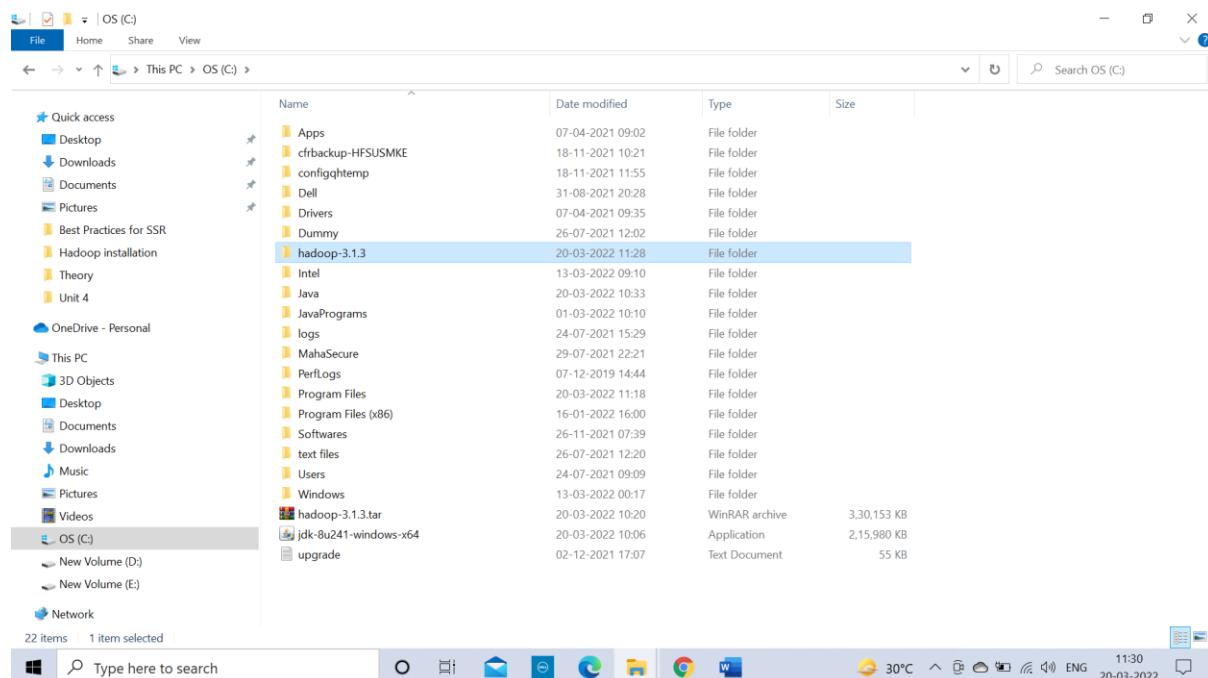
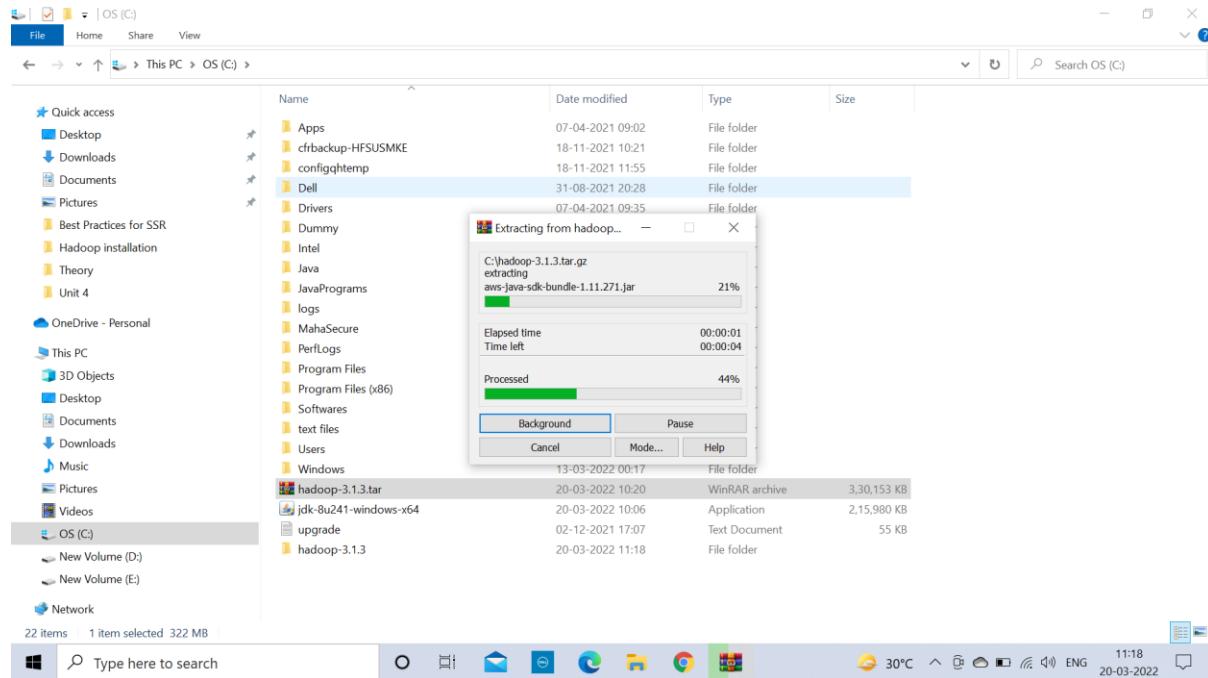


Check java version

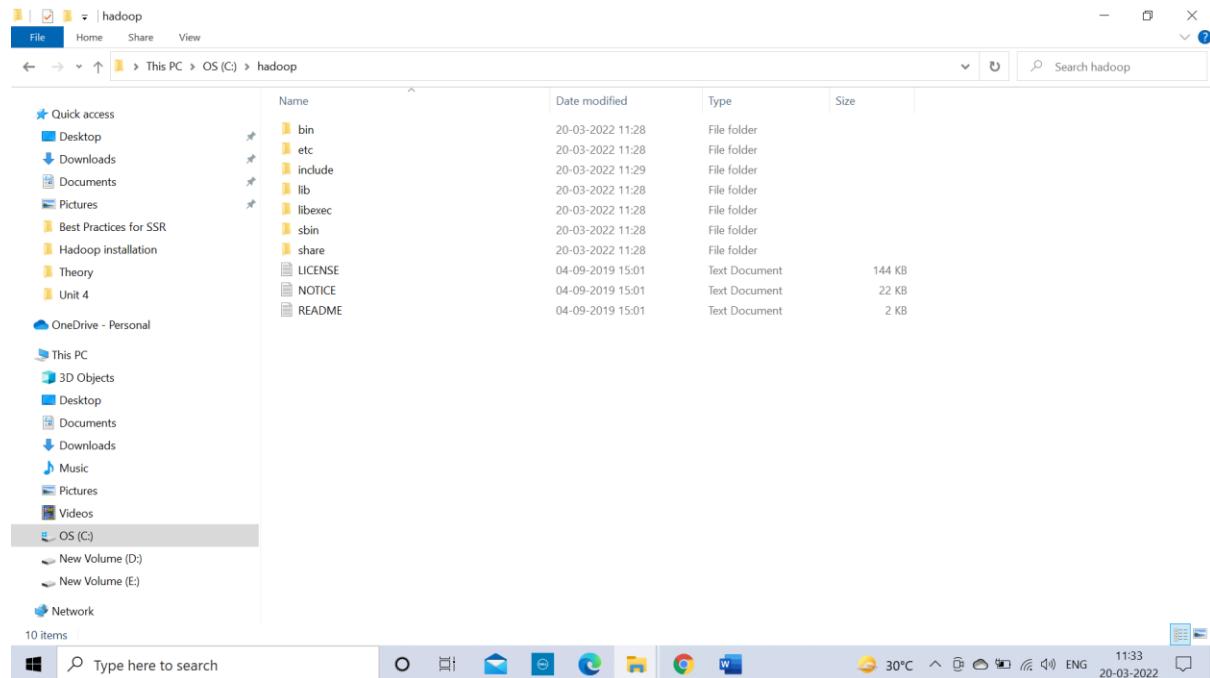


Step 6: Install Hadoop now to our local system

Unzip Hadoop setup file

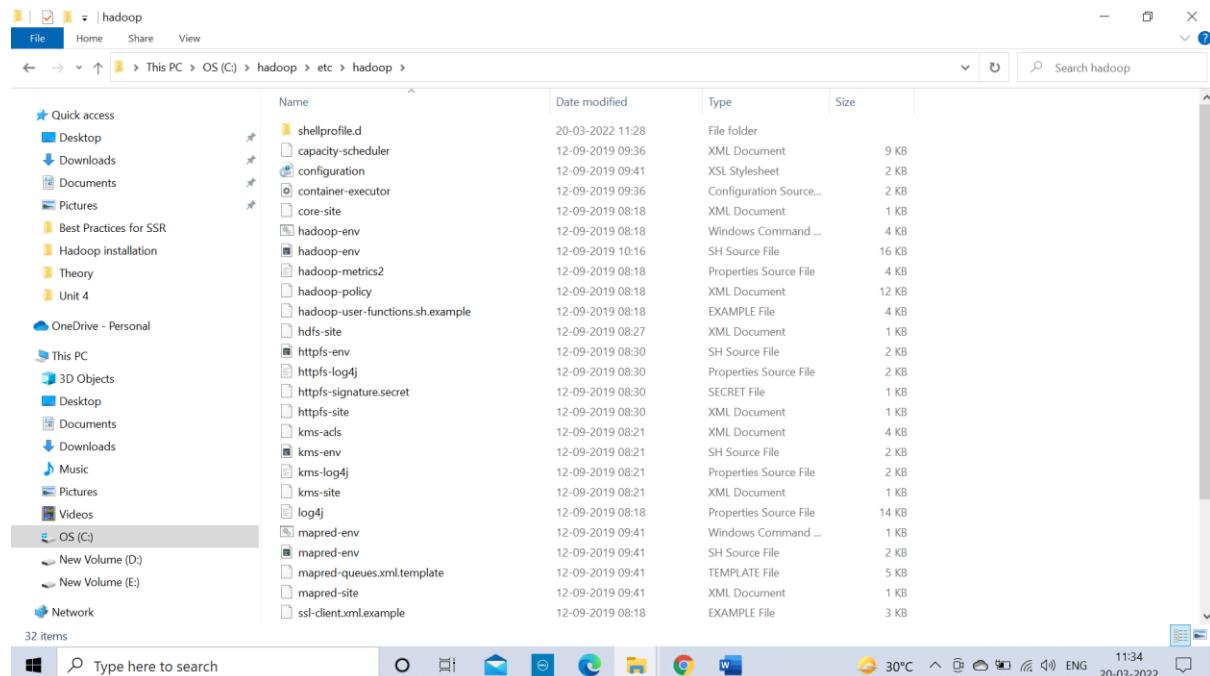


Rename Hadoop-3.1.3 folder as **hadoop**



Step 7: Configuration of Hadoop

Hadoop->etc-> Hadoop



Important files:

Core-site.xml

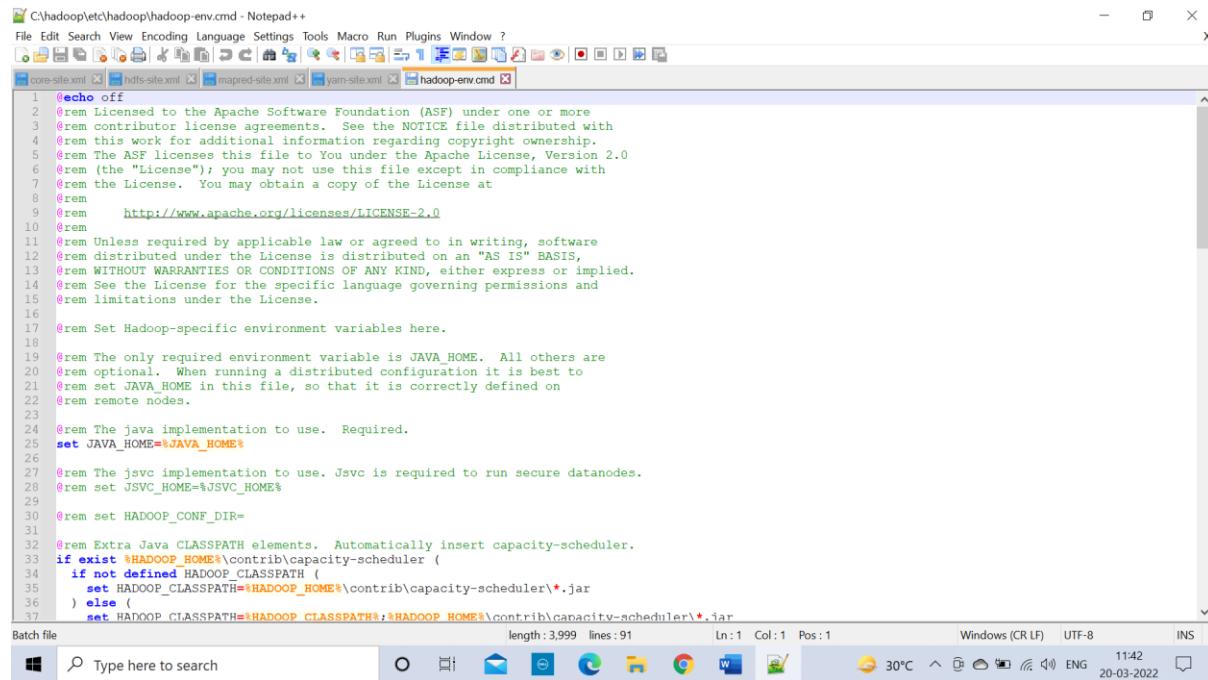
Hdfs-site.xml

Mapred-site.xml

Yarn-site.xml

Hadoop-env windows command prompt file

Open all these files in notepad++



```
core-site.xml  hdfs-site.xml  mapred-site.xml  yarn-site.xml  hadoop-env.cmd
1 @echo off
2 /* Licensed to the Apache Software Foundation (ASF) under one or more
3  * contributor license agreements. See the NOTICE file distributed with
4  * this work for additional information regarding copyright ownership.
5  * The ASF licenses this file to You under the Apache License, Version 2.0
6  * (the "License"); you may not use this file except in compliance with
7  * the License. You may obtain a copy of the License at
8  *
9  *      http://www.apache.org/licenses/LICENSE-2.0
10 /*
11 * Unless required by applicable law or agreed to in writing, software
12 * distributed under the License is distributed on an "AS IS" BASIS,
13 * WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
14 * See the License for the specific language governing permissions and
15 * limitations under the License.
16 /*
17 * Set Hadoop-specific environment variables here.
18 /*
19 * The only required environment variable is JAVA_HOME. All others are
20 * optional. When running a distributed configuration it is best to
21 * set JAVA_HOME in this file, so that it is correctly defined on
22 * remote nodes.
23 /*
24 * The java implementation to use. Required.
25 * set JAVA_HOME=%JAVA_HOME%
26 /*
27 * The jsvc implementation to use. Jsvc is required to run secure datanodes.
28 * set JSVC_HOME=%JSVC_HOME%
29 /*
30 * set HADOOP_CONF_DIR=
31 /*
32 * Extra Java CLASSPATH elements. Automatically insert capacity-scheduler.
33 * if exist "%HADOOP_HOME%\contrib\capacity-scheduler (
34 *   if not defined HADOOP_CLASSPATH (
35 *     set HADOOP_CLASSPATH=%HADOOP_HOME%\contrib\capacity-scheduler\*.jar
36 *   ) else (
37 *     set HADOOP_CLASSPATH=%HADOOP_CLASSPATH%;%HADOOP_HOME%\contrib\capacity-scheduler\*.jar
Batch file          length:3,999  lines:91           Ln:1  Col:1  Pos:1           Windows (CR LF)  UTF-8    INS
Type here to search          O  E-mail  E  G  M  W  S  30°C  ^  ⌂  ENG  11:42  20-03-2022
```

Core-site.xml

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

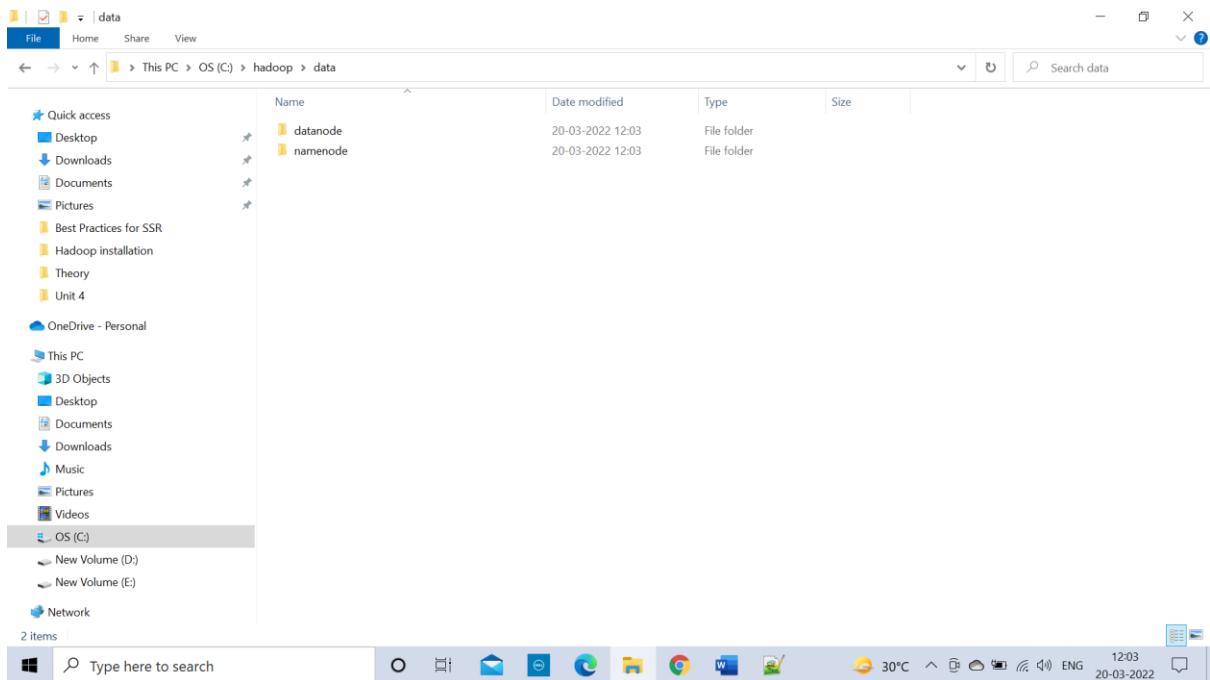
Mapred-site.xml

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

Yarn-site.xml

```
<configuration>
  <!-- Site specific YARN configuration properties -->
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
</configuration>
```

Create data directory and subdirectories in Hadoop

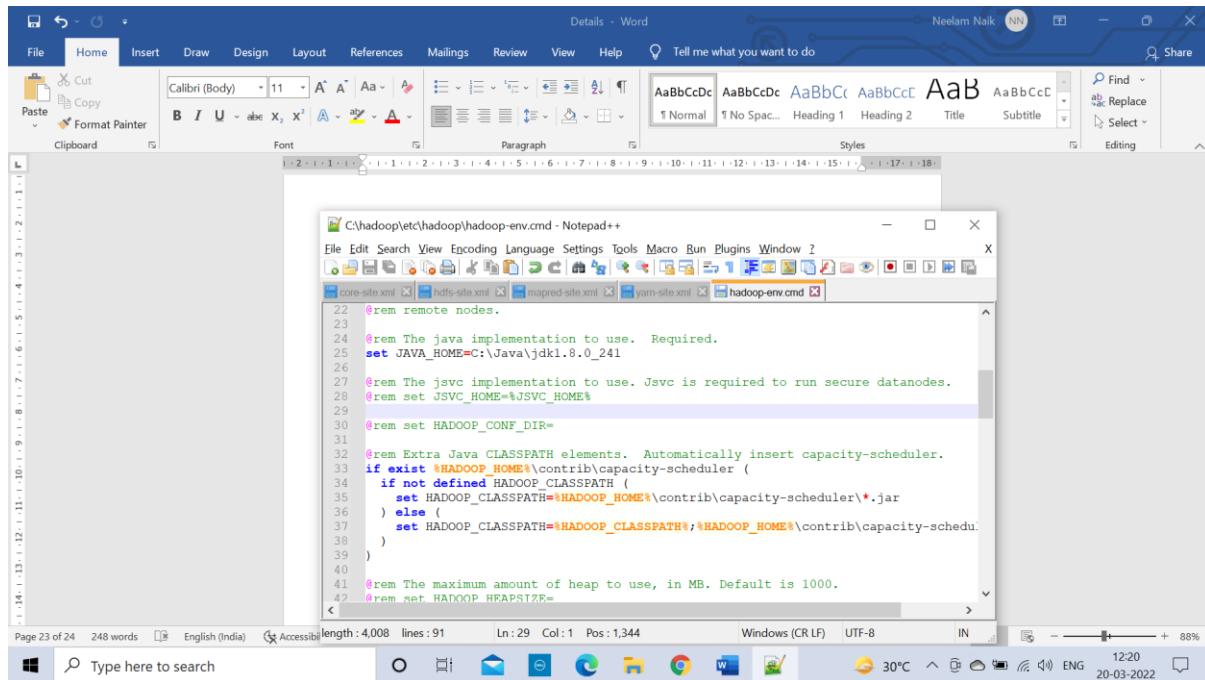


Hdfs-site.xml file

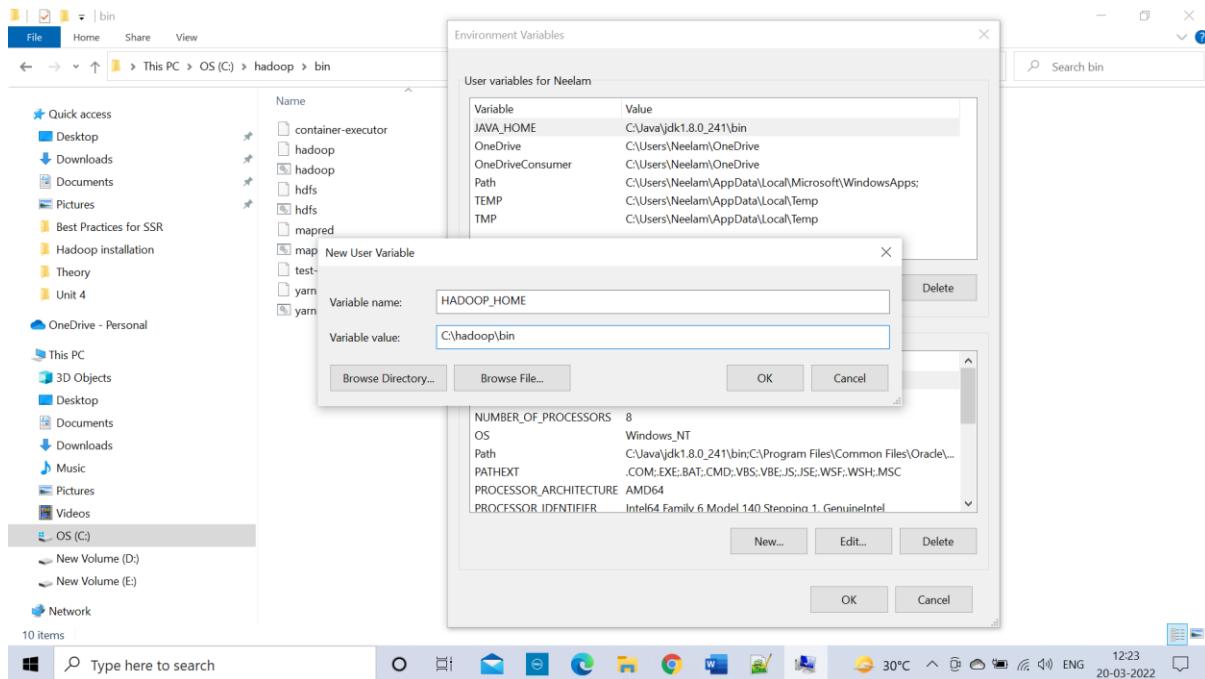
```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///C:/hadoop/data/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:///C:/hadoop/data/datanode</value>
  </property>
</configuration>
```

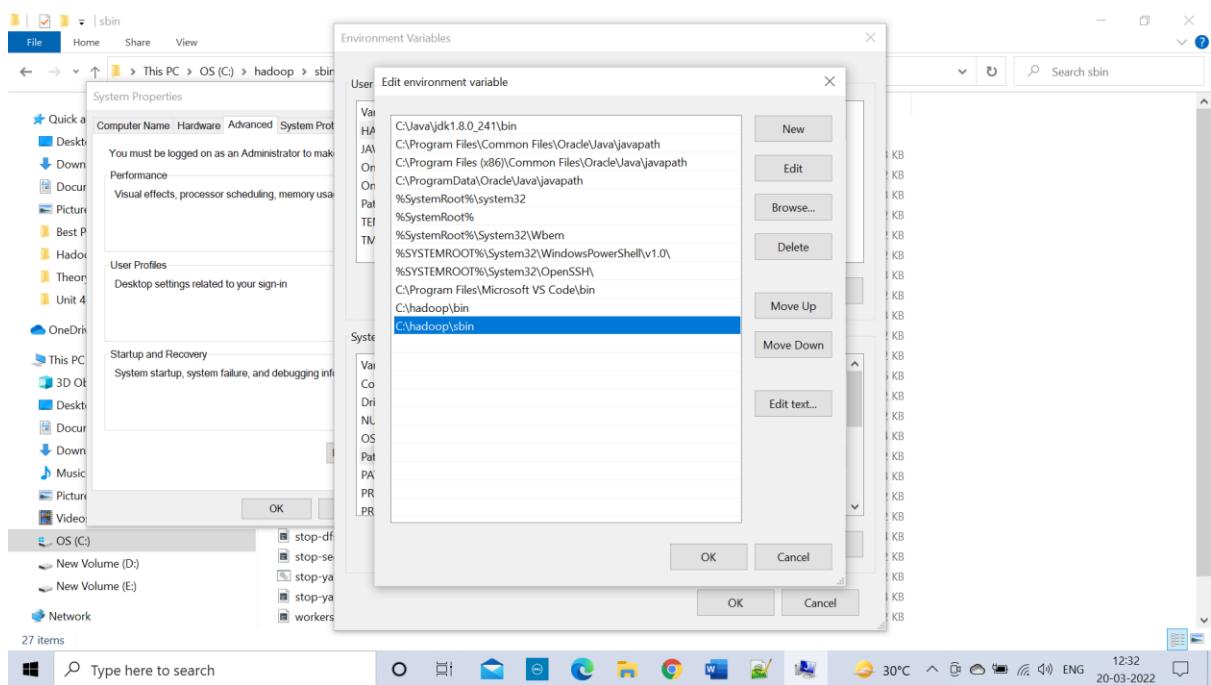
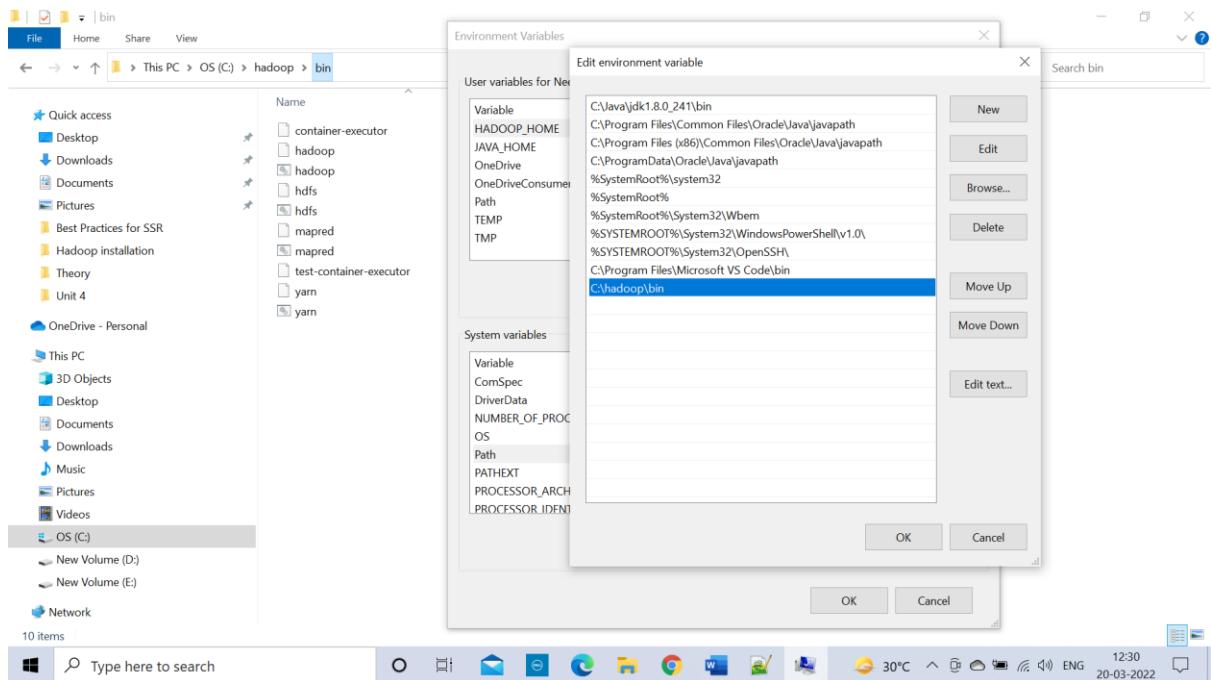
Hadoop-env.xml

```
set JAVA_HOME=C:\Java\jdk1.8.0_241
```

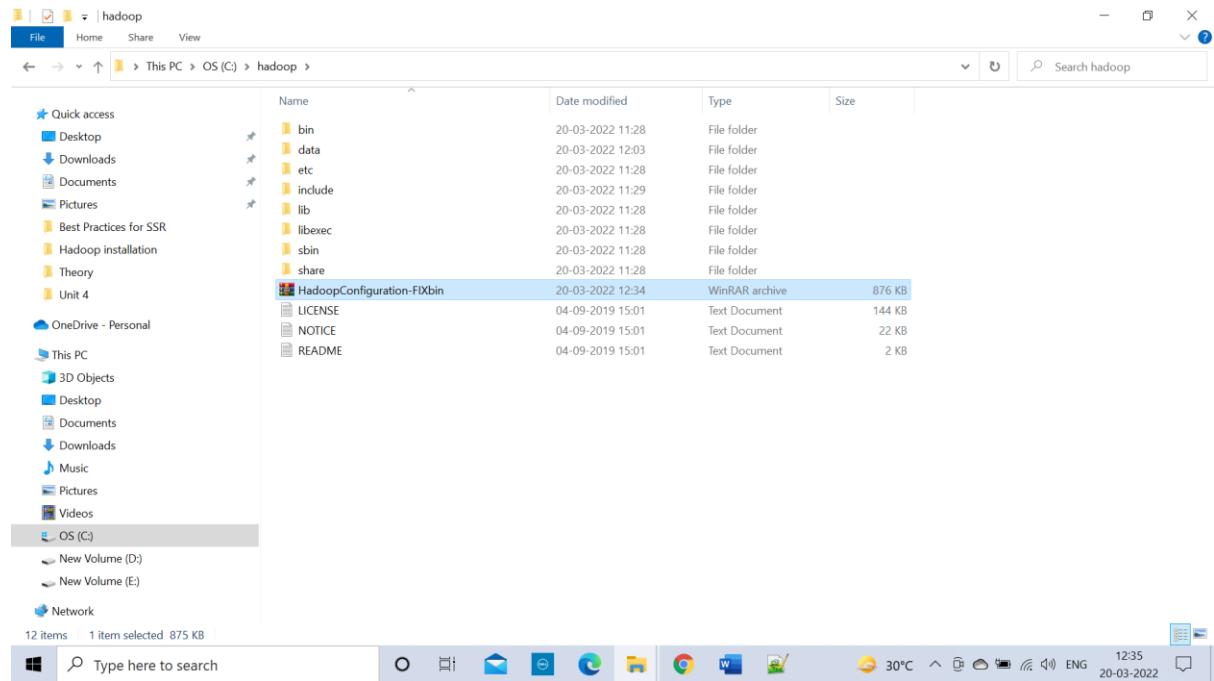


Set home and path for Hadoop

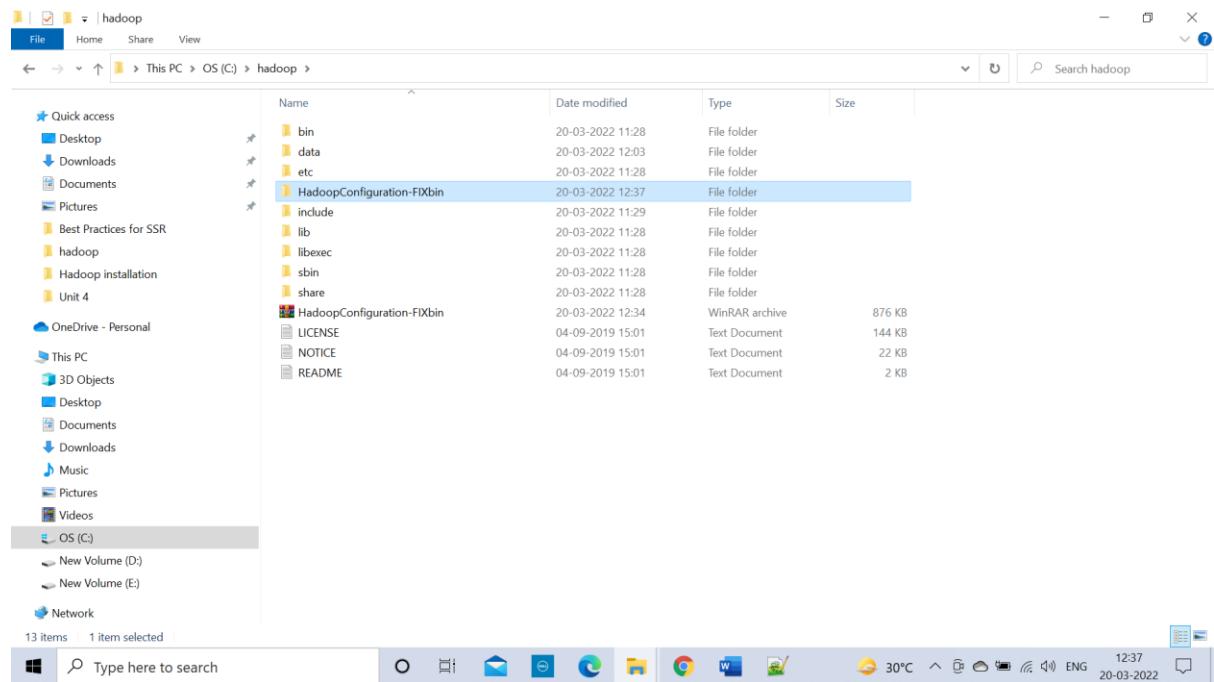




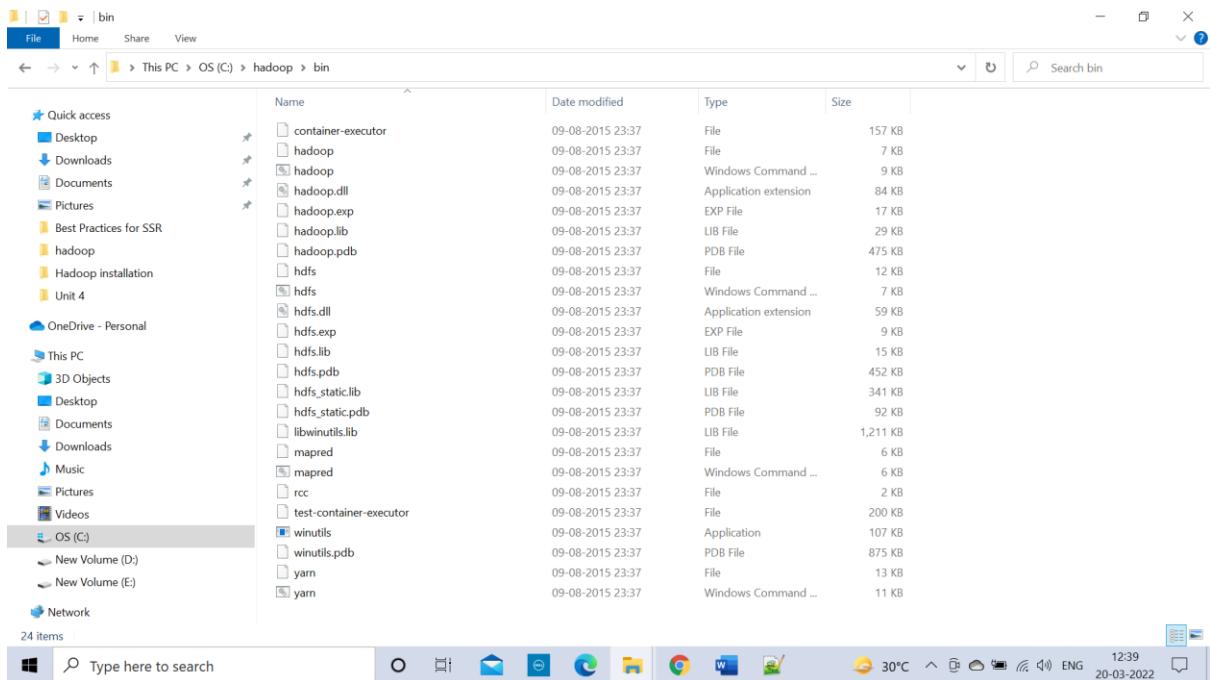
Other configuration files



Extract it



Copy bin folder from HadoopConfiguration-FIXbin folder and replace Hadoop\bin folder with this



Step 8: Verification of Hadoop installation

Go to command prompt.

Type:

hdfs namenode -format

set of files pop up on the terminal.

That means successful installation of Hadoop.

Details of Hadoop installation_for studnets - Word

File Home Insert Draw Design Layout References Mailings Review View Help Tell me what you want to do

Font Paragraph Styles Editing

Clipboard Command Prompt

```
Microsoft Windows [Version 10.0.19044.1586]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Neelam>hdfs namenode -format
```

Page 26 of 28 335 words English (India) Accessibility: Investigate

Type here to search

29°C ENG 23:27 24-03-2022

Details of Hadoop installation_for studnets - Word

File Home Insert Draw Design Layout References Mailings Review View Help Tell me what you want to do

Font Paragraph Styles Editing

Clipboard Command Prompt

```
2022-03-24 23:27:48,653 INFO util.GSet: 0.25% max memory 889 MB = 2.2 MB
2022-03-24 23:27:48,653 INFO util.GSet: capacity = 2^18 = 262144 entries
2022-03-24 23:27:48,653 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
2022-03-24 23:27:48,653 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2022-03-24 23:27:48,653 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2022-03-24 23:27:48,670 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2022-03-24 23:27:48,670 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
2022-03-24 23:27:48,670 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2022-03-24 23:27:48,670 INFO util.GSet: VM type = 64-bit
2022-03-24 23:27:48,670 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
2022-03-24 23:27:48,670 INFO util.GSet: capacity = 2^15 = 32768 entries
Re-format file system in Storage Directory root= C:\hadoop\data\namenode; location= null ? (Y or N) y
2022-03-24 23:27:53,813 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1639733029-192.168.1.28-1648144673813
2022-03-24 23:27:53,813 INFO common.Storage: Will remove files: [C:\hadoop\data\namenode\current\edits_inprogress_00000000000000000000, C:\hadoop\data\namenode\current\fsimage_00000000000000000000, C:\hadoop\data\namenode\current\fsimage_00000000000000000000, C:\hadoop\data\namenode\current\seen_txid, C:\hadoop\data\namenode\current\VERSION]
2022-03-24 23:27:53,844 INFO common.Storage: Storage directory C:\hadoop\data\namenode has been successfully formatted.
2022-03-24 23:27:53,881 INFO namenode.FSImageFormatProtobuf: Saving image file C:\hadoop\data\namenode\current\fsimage.ckpt_0000000000000000 using no compression
2022-03-24 23:27:53,981 INFO namenode.FSImageFormatProtobuf: Image file C:\hadoop\data\namenode\current\fsimage.ckpt_0000000000000000 of size 393 bytes saved in 0 seconds .
2022-03-24 23:27:53,992 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2022-03-24 23:27:53,992 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid = 0 when meet shutdown.
2022-03-24 23:27:53,992 INFO namenode.NameNode: SHUTDOWN_MSG:
*****
SHUTDOWN MSG: Shutting down NameNode at DESKTOP-M6JHUV/192.168.1.28
*****
```

C:\Users\Neelam>

Page 27 of 28 335 words English (India) Accessibility: Investigate

Type here to search

29°C ENG 23:27 24-03-2022

Namenode is successfully started.

Details of Hadoop installation_for studnets - Word

File Home Insert Draw Design Layout References Mailings Review View Help Tell me what you want to do

Normal No Spac... Heading 1 Heading 2 Title Subtitle Editing

Font Paragraph Styles

Clipboard Command Prompt

```
Microsoft Windows [Version 10.0.19044.1586]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Neelam>hdfs namenode -format
2022-03-24 23:27:47,372 INFO namenode.NameNode: STARTUP_MSG:
*****STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = DESKTOP-M6JHHUV/192.168.1.28
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.1.3
STARTUP_MSG: classpath = C:\hadoop\etc\hadoop;C:\hadoop\share\hadoop\common;C:\hadoop\share\hadoop\common\lib\accessor
e-smart-1.2.jar;C:\hadoop\share\hadoop\common\lib\animal-sniffer-annotations-1.17.jar;C:\hadoop\share\hadoop\common\lib\avro-1
.8.5.0.4.jar;C:\hadoop\share\hadoop\common\lib\audience-annotations-0.5.0.jar;C:\hadoop\share\hadoop\common\lib\avro-1
.7.7.jar;C:\hadoop\share\hadoop\common\lib\checker-qual-2.5.2.jar;C:\hadoop\share\hadoop\common\lib\commons-beanutils-1.9
.3.jar;C:\hadoop\share\hadoop\common\lib\commons-cli-1.2.jar;C:\hadoop\share\hadoop\common\lib\commons-codesrc-1.11.jar;C:
\hadoop\share\hadoop\common\lib\commons-collections-3.2.2.jar;C:\hadoop\share\hadoop\common\lib\commons-compress-1.18.ja
r;C:\hadoop\share\hadoop\common\lib\commons-configuration2-2.1.1.jar;C:\hadoop\share\hadoop\common\lib\commons-io-2.5.ja
r;C:\hadoop\share\hadoop\common\lib\commons-lang-2.6.jar;C:\hadoop\share\hadoop\common\lib\commons-lang3-3.4.jar;C:\hado
op\share\hadoop\common\lib\commons-logging-1.1.3.jar;C:\hadoop\share\hadoop\common\lib\commons-math3-3.1.1.jar;C:\hado
op\share\hadoop\common\lib\curator-framework-2.13.0.jar;C:\hadoop\share\hadoop\common\lib\curator-client-2.13.0.jar;C:\hado
op\share\hadoop\common\lib\error-prone-annotations-2.2.8.jar;C:\hadoop\share\hadoop\common\lib\failsafeaccess-1.0.jar;C:\hado
op\share\hadoop\common\lib\guava-27.0-jre.jar;C:\hadoop\share\hadoop\common\lib\hadoop-auth-3.1.3.jar;C:\hadoop\share\hadoop\co
mmon\lib\hadoop-annotations-3.1.3.jar;C:\hadoop\share\hadoop\common\lib\hadoop-auth-3.1.3.jar;C:\hadoop\share\hadoop\co
mmon\lib\hadoop-core-4.1.0-incubating.jar;C:\hadoop\share\hadoop\common\lib\httpclient-4.5.2.jar;C:\hadoop\share\hadoop\co
mmon\lib\hadoop-core-4.4.4.jar;C:\hadoop\share\hadoop\common\lib\jackson-core-2.7.8.jar;C:\hadoop\share\hadoop\common\lib\jackson
-commons-net-3.6.jar;C:\hadoop\share\hadoop\common\lib\jackson-core-asl-1.9.13.jar;C:\hadoop\share\hadoop\common\lib\jackson-databind-2.7.8.ja
r;C:\hadoop\share\hadoop\common\lib\jackson-jaxrs-1.9.13.jar;C:\hadoop\share\hadoop\common\lib\jackson-xc-1.9.13.jar;C:\hadoop\share\hadoop\co
mmon\lib\jaxws-rt-2.3.0.jar;C:\hadoop\share\hadoop\common\lib\javaj.servlet-api-3.1.0.jar;C:\hadoop\share\hadoop\co
```

Page 27 of 29 335 words English (India) Accessibility: Investigate

29°C 23:28 ENG 24-03-2022

Open new terminal and start all Hadoop daemons

Go to Hadoop location.

C:\hadoop\sbin

Details of Hadoop installation_for studnets - Word

File Home Insert Draw Design Layout References Mailings Review View Help Tell me what you want to do

Normal No Spac... Heading 1 Heading 2 Title Subtitle Editing

Font Paragraph Styles

Clipboard Command Prompt

```
Microsoft Windows [Version 10.0.19044.1586]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Neelam>cd..
C:\Users>cd..
C:\>cd hadoop
C:\hadoop>cd sbin
C:\hadoop\sbin>
```

Page 28 of 29 352 words English (India) Accessibility: Investigate

29°C 23:31 ENG 24-03-2022

Type the command

start-all.cmd

Details of Hadoop installation_for studnets - Word

File Home Insert Draw Design Layout References Mailings Review View Help Tell me what you want to do

Clipboard

c:\ Command Prompt

```
Microsoft Windows [Version 10.0.19044.1586]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Neelam>cd..

C:\Users>cd..

C:\>cd hadoop

C:\hadoop>cd sbin

C:\hadoop\sbin>start-all.cmd
```

Page 28 of 30 1 of 356 words English (India) Accessibility: Investigate

Type here to search

Details of Hadoop installation_for studnets - Word

File Home Insert Draw Design Layout References Mailings Review View Help Tell me what you want to do

Clipboard

Apache Hadoop Distribution - hadoop_namenode

```
2022-03-24 23:32:56,612 INFO nodemanager.NodeManager: SHUTDOWN_MSG:
*****SHUTDOWN MSG: Shutting down NodeManager at DESKTOP-M6JHUV/192.168.1.28
*****
```

Page 29 of 30 356 words English (India) Accessibility: Investigate

Type here to search

All the nodes will start successfully.

Also try the following command to start yarn daemons.

start-yarn.cmd

Microsoft Windows [Version 10.0.19044.1586]
(c) Microsoft Corporation. All rights reserved.

```
C:\Users\Neelam>cd..  
C:\Users>cd..  
C:\>cd hadoop  
C:\hadoop>cd sbin  
C:\hadoop\sbin>start-all.cmd  
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd  
starting yarn daemons  
C:\hadoop\sbin>start-yarn.cmd  
starting yarn daemons  
C:\hadoop\sbin>
```

A screenshot of a Microsoft Word document titled "Apache Hadoop Distribution". The document contains a large block of text representing Hadoop startup logs from March 24, 2022, at 23:35:27. The log details the startup of the NodeManager, listing various Java classpath entries and system properties. The Microsoft Word ribbon is visible at the top, and the status bar shows the date and time.