

Practical 1

Aim: -Installing & Setting Environment Variables for Working with Apache Hadoop

Theory: Apache Hadoop is an open-source software framework used for distributed storage and processing of large data sets using the MapReduce programming model. It is designed to scale up from a single server to thousands of machines, each offering local computation and storage. Hadoop is part of the Apache Software Foundation and has become a key technology in the world of Big Data.

Key Components of Hadoop:

1)Hadoop Distributed File System (HDFS):

HDFS is a distributed file system that stores data across multiple machines. It is designed to handle large files by breaking them into smaller blocks and distributing them across different nodes in a cluster.

Features include fault tolerance, high throughput, and the ability to handle large volumes of data.

2)MapReduce:

MapReduce is the programming model that allows for processing large data sets with a distributed algorithm on a cluster. It divides the data processing into two phases:

Map phase: Processes the input data and converts it into a set of key-value pairs.

Reduce phase: Takes the output from the map phase and combines it to produce the final result.

3)YARN (Yet Another Resource Negotiator):

YARN is the resource management layer of Hadoop that schedules and manages resources across the cluster. It allows multiple data processing engines (e.g., MapReduce, Spark) to run simultaneously.

4)Hadoop Common:

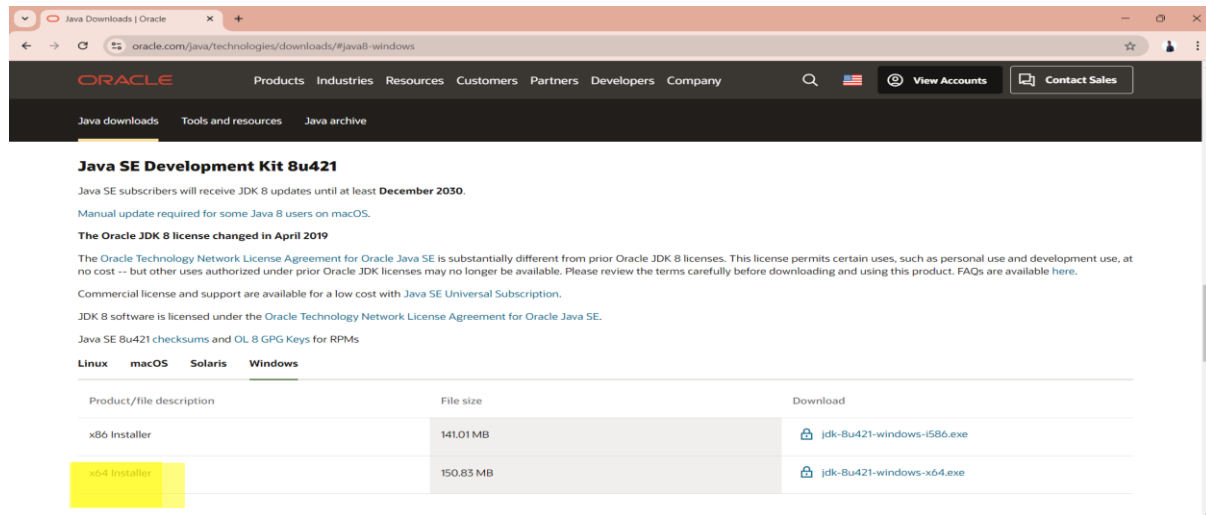
These are the common utilities and libraries that support the other Hadoop modules. It includes essential Java libraries and scripts required to start Hadoop

Steps:

JAVA INSTALLATION & SETTING ENVIRONMENT

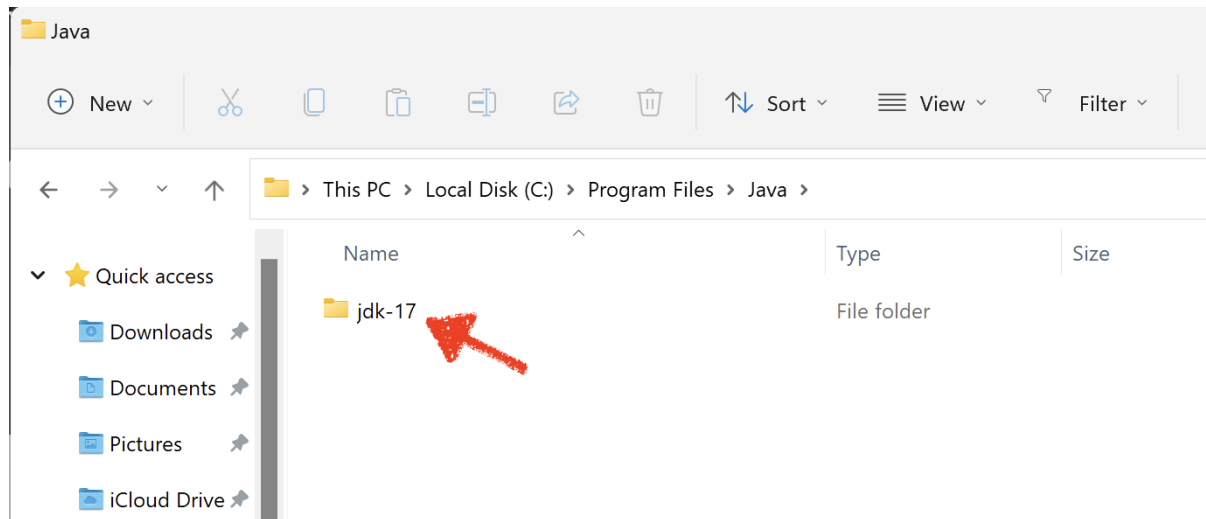
1)Download Java

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

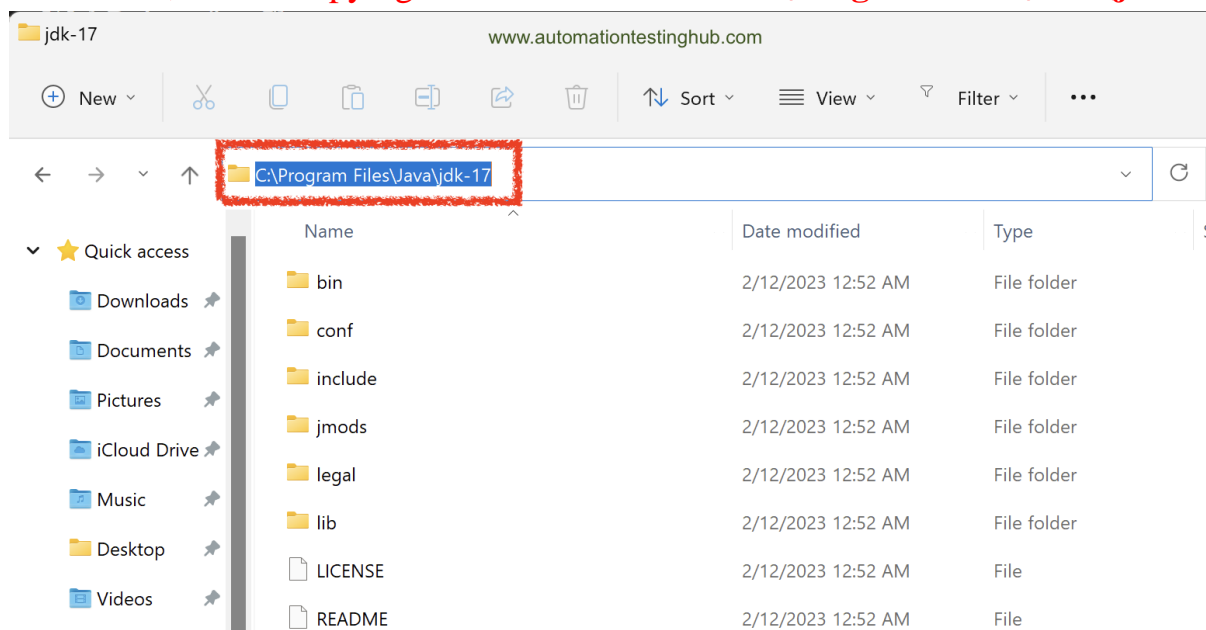


2)Setting System Environment Variable

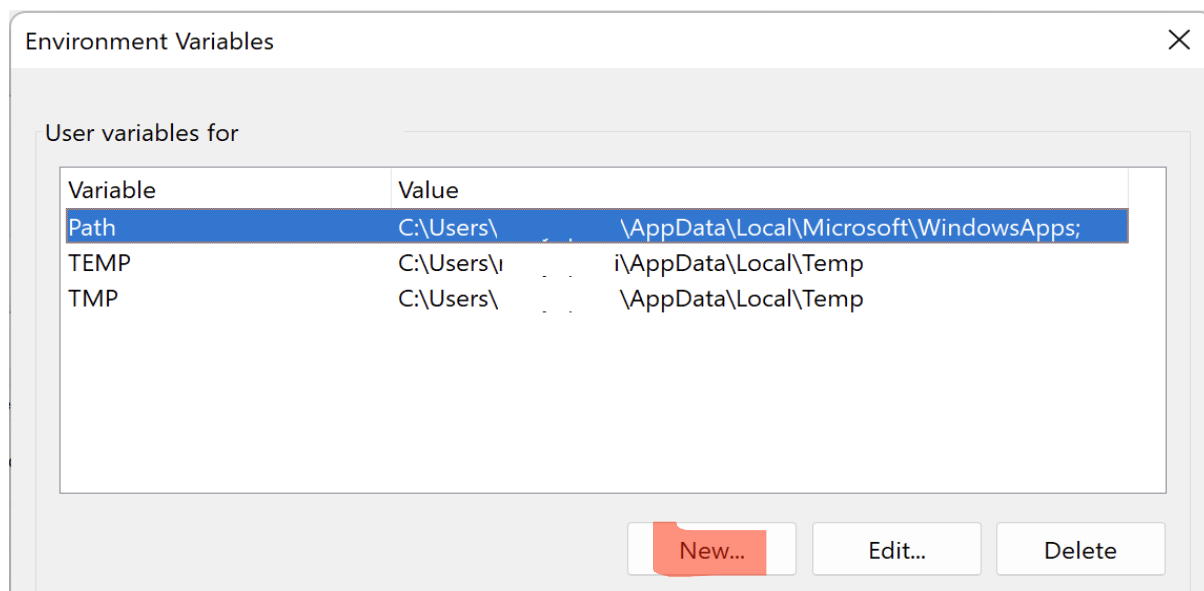
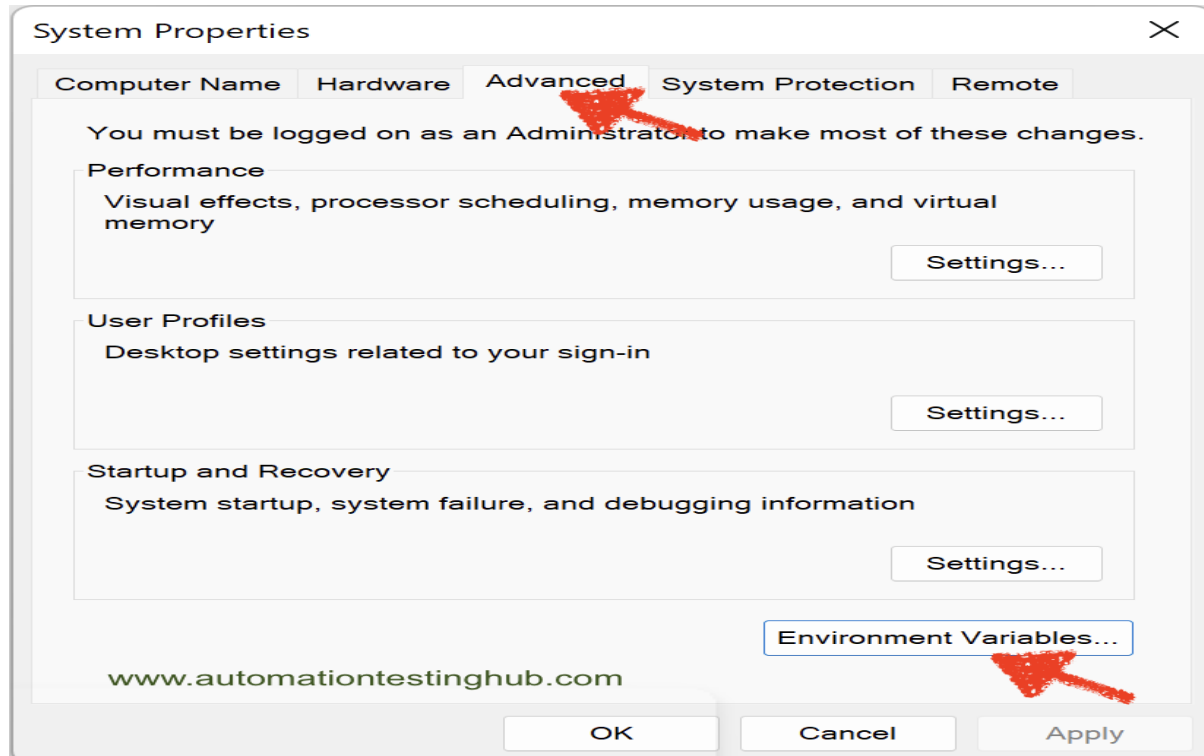
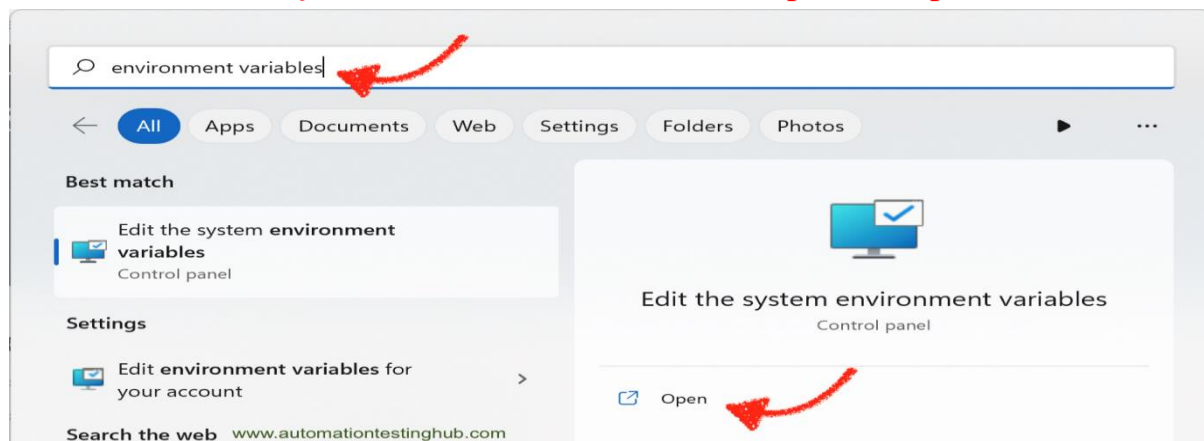
After installation now

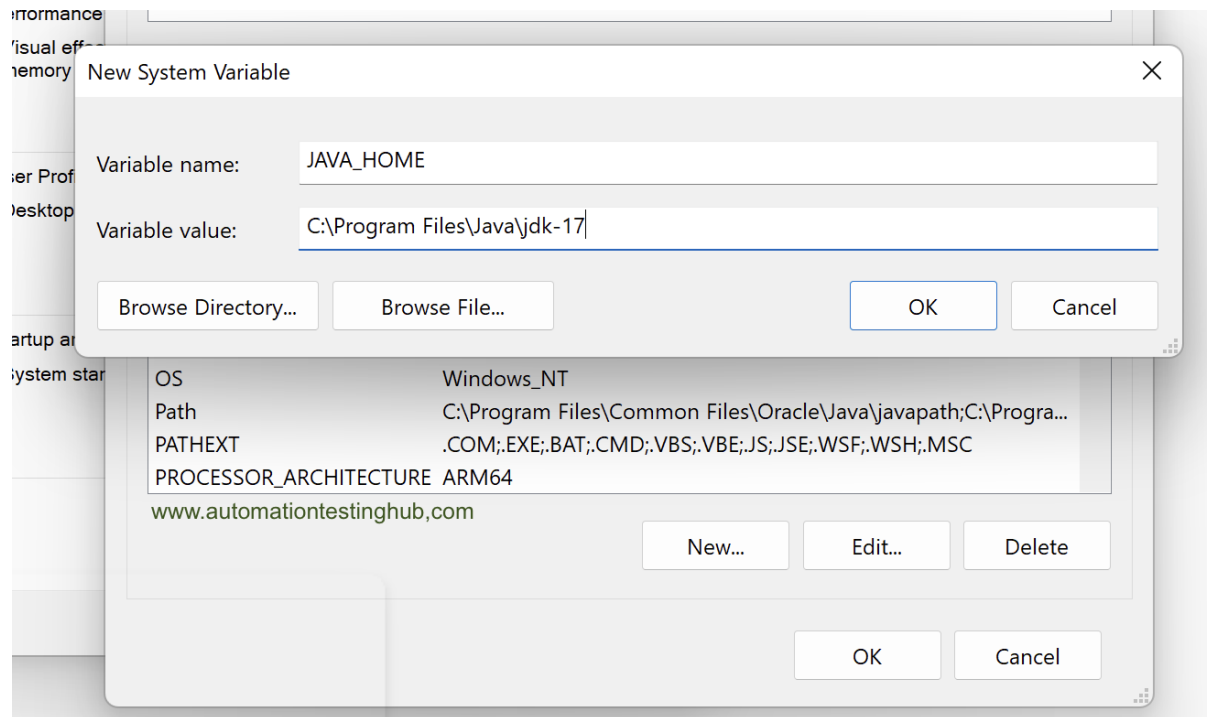


Now open the JDK folder. If you have multiple JDK folders, then open the one which represents the latest version, and copy the folder location. As per the below screenshot, we are copying this folder location – **C:\Program Files\Java\jdk-17**

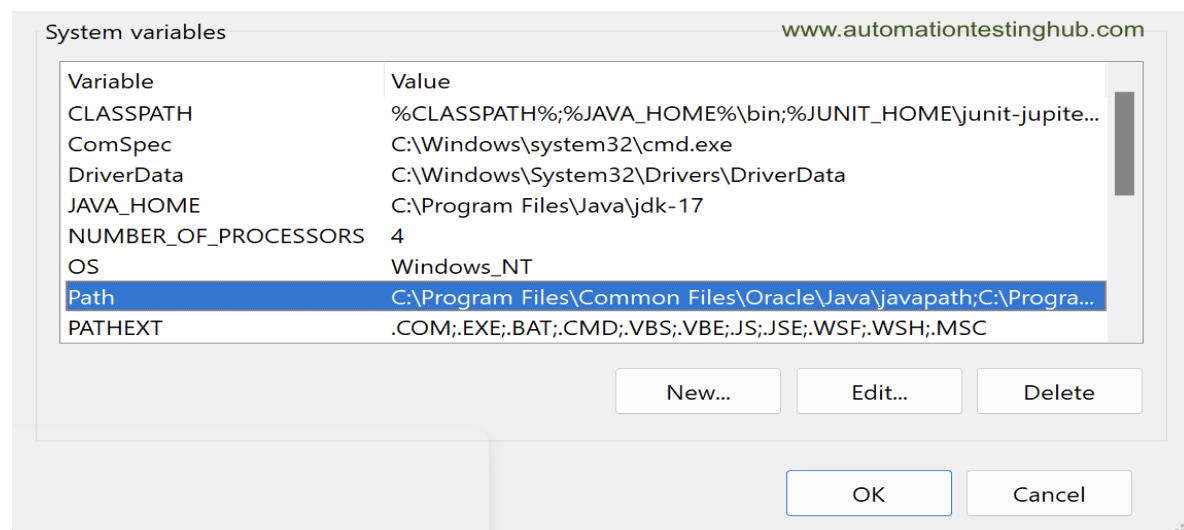
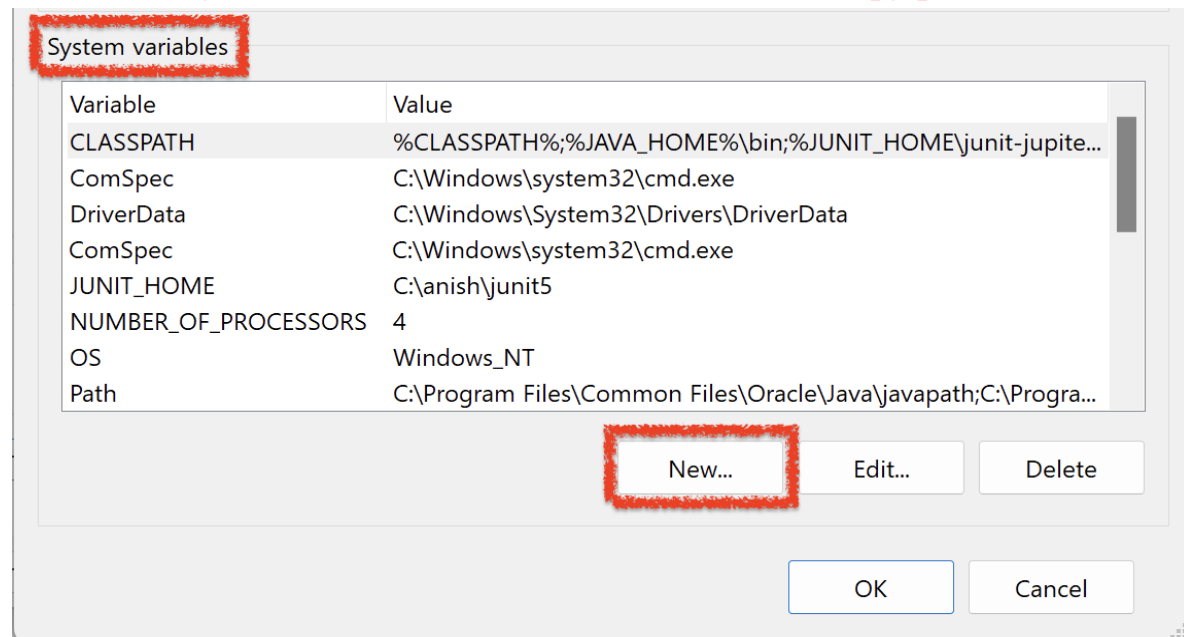


You have to now open environment variables window in your machine. To do this, click on **Search icon** and then type **environment variables** in the search bar. Click on **Edit the system environment variables** option to open it





Then Now System Variable-> Path -> Edit -> New -> Copy paste same link

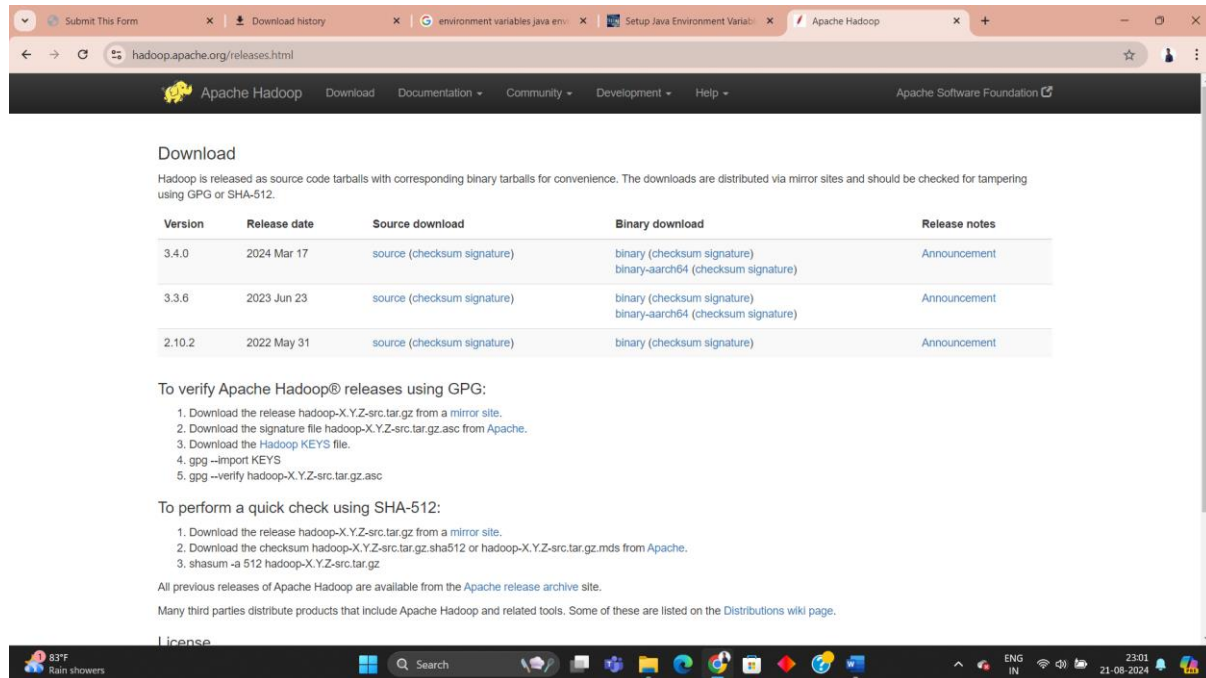


HADOOP INSTALLATION & SETTING ENVIRONMENT

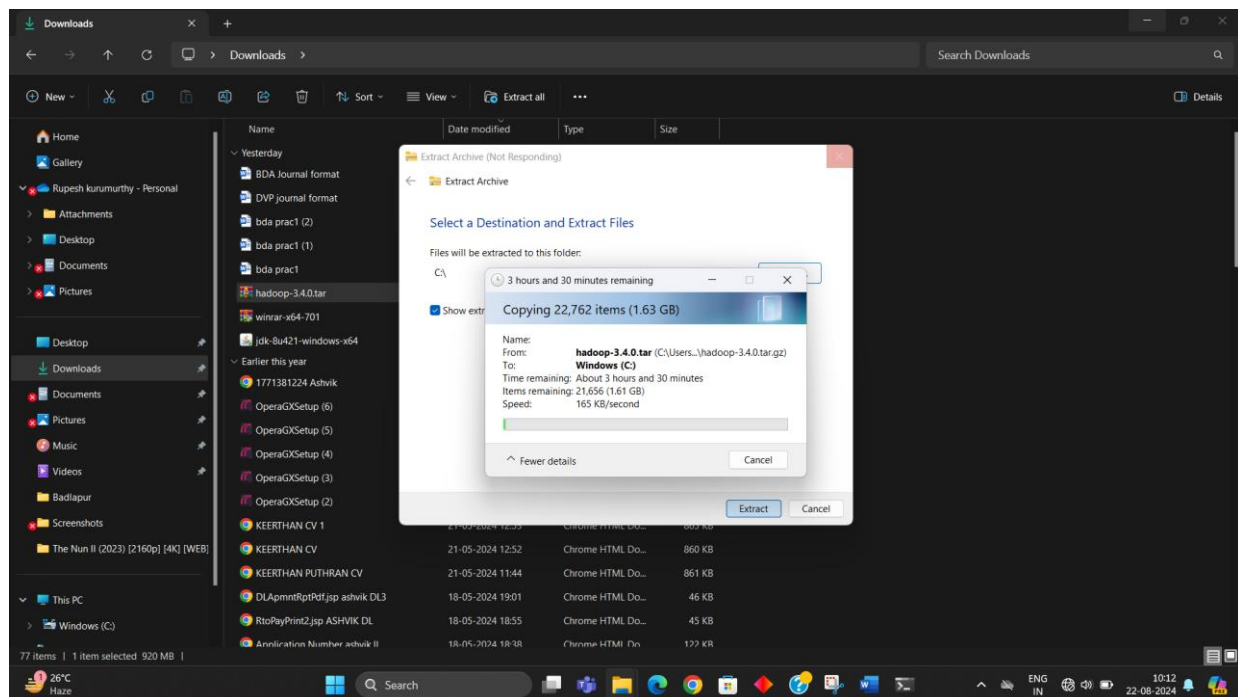
1)Download Hadoop

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

Click binary download



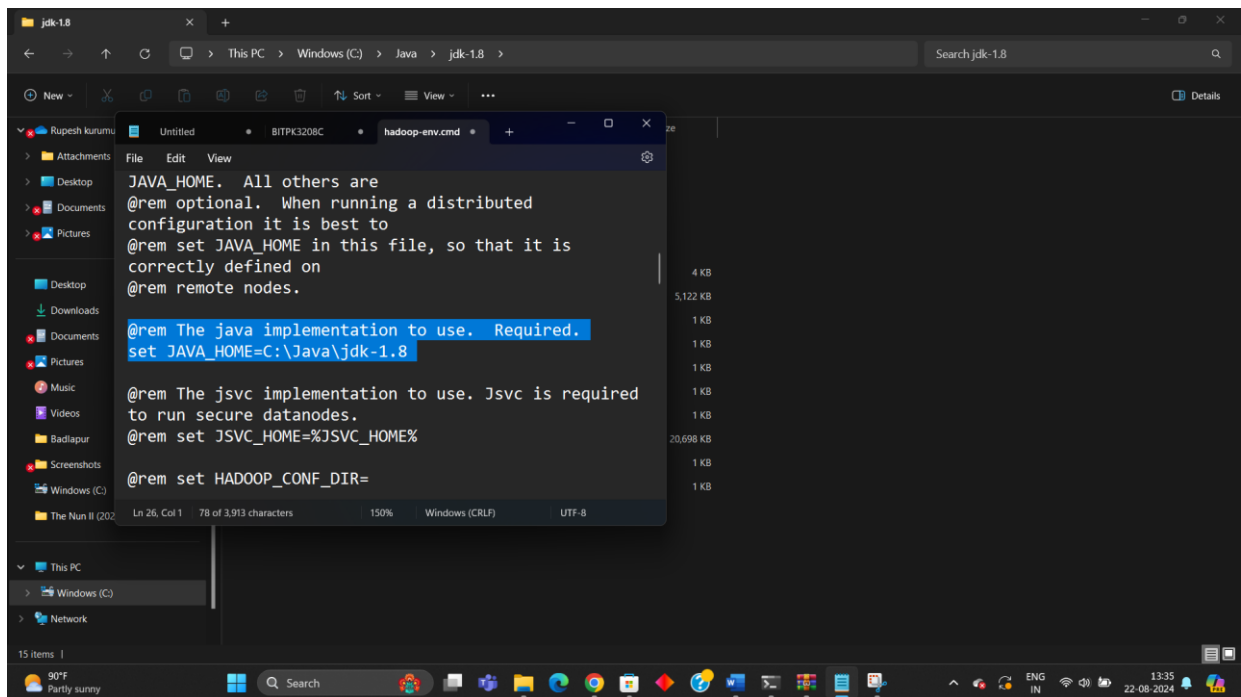
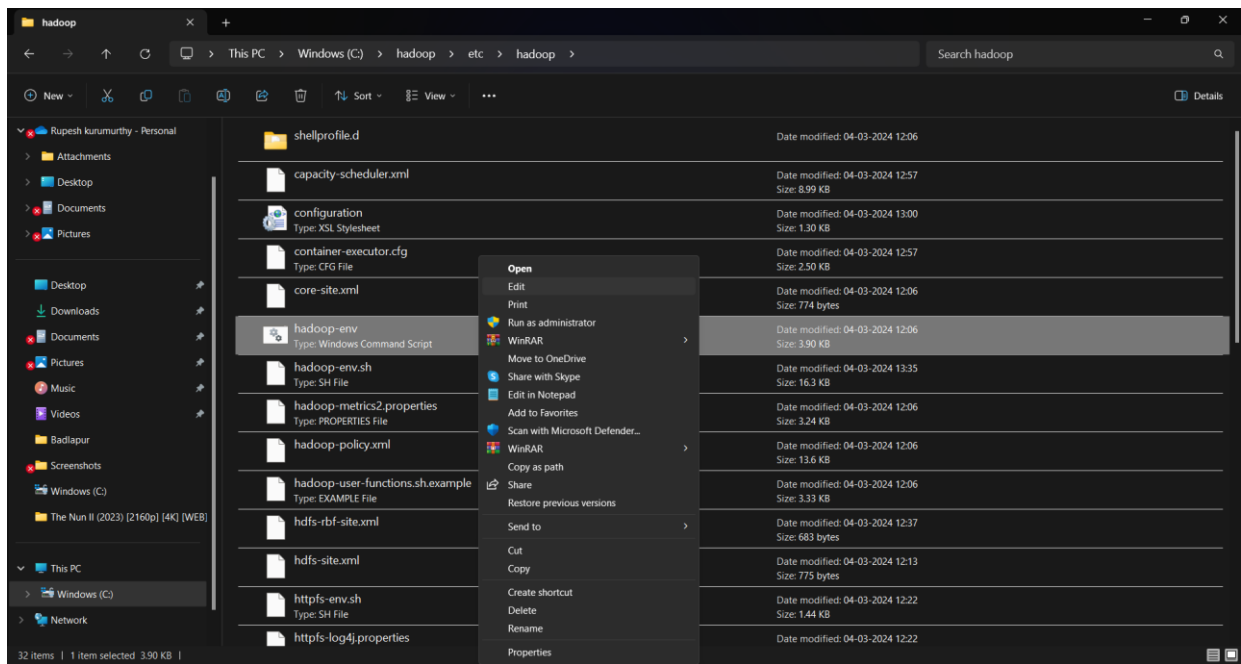
Extract it to C drive



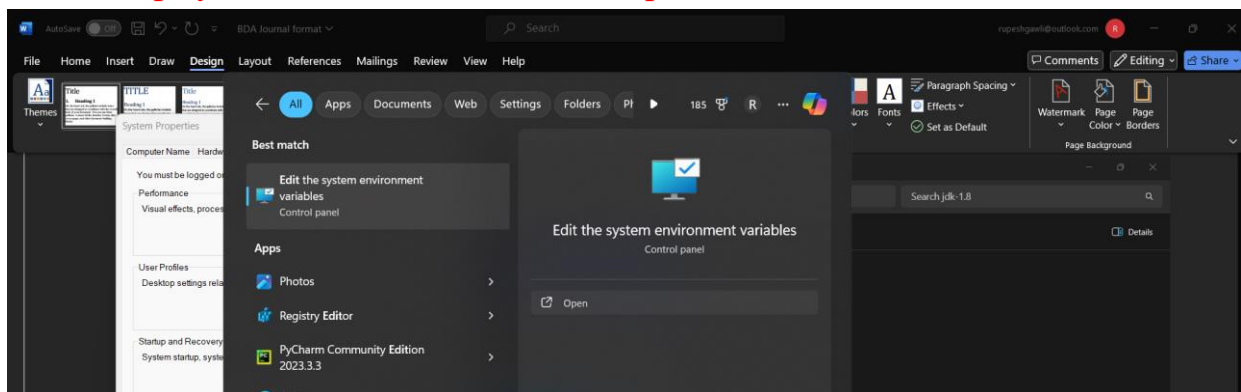
2)Setting System Environment Variable

After Installation

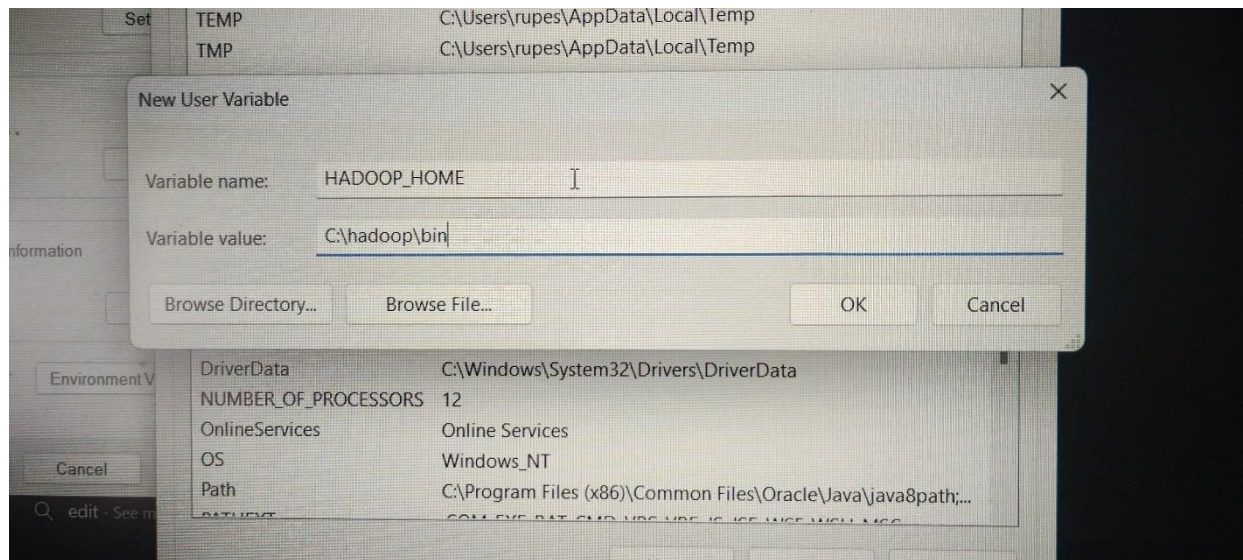
Gotowindows(c):->Hadoop->etc->Hadoop->Hadoop-env



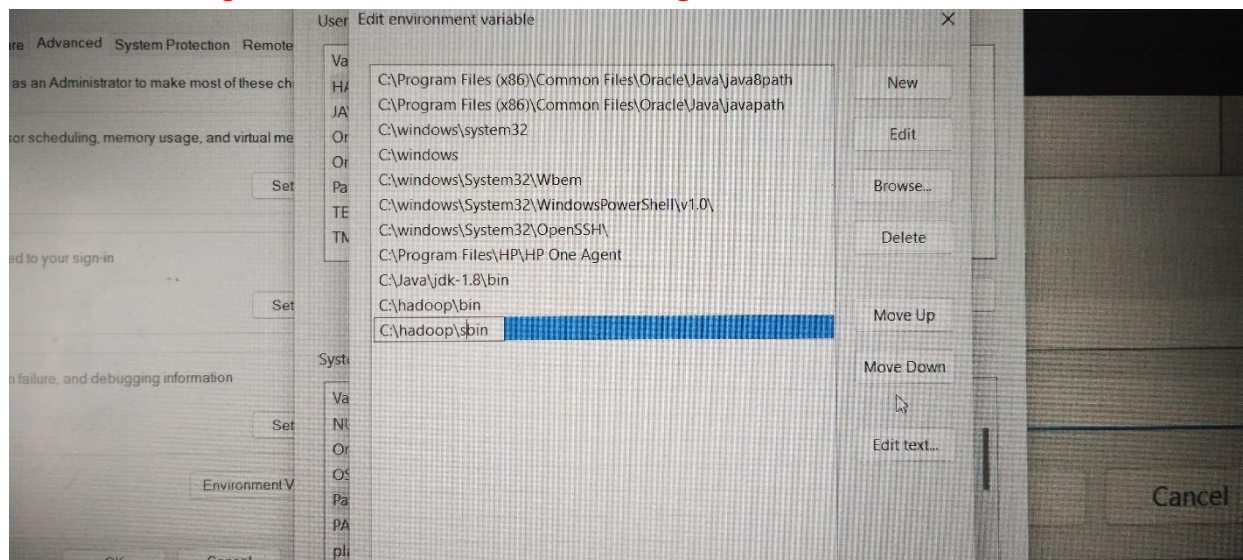
Then setup system environment for Hadoop



First Edit New User Variable As like we done for JAVA_HOME



Now System variable. Here pastes the path but in these we have to paste two link 1st the normal path and the second one adding “s” before Bin



After Installation Now Check In command prompt

```

Command Prompt
fs          run a generic filesystem user client
version     print the version
jar <jar>    run a jar file
            note: please use "yarn jar" to launch
            YARN applications, not this command.
checknative [-a|-h] check native hadoop and compression libraries availability
conftest    validate configuration XML files
distch path:owner:group:permission distributed metadata changer
distcp <srcurl> <desturl> copy file or directories recursively
archive -archiveName NAME -p <parent path> <src>* <dest> create a hadoop archive
classpath   prints the class path needed to get the
            Hadoop jar and the required libraries
credential  prints the java.library.path
jnipath     show auth_to_local principal conversion
kerbname    diagnose kerberos problems
kdiag       manage keys via the KeyProvider
key         get/set the log level for each daemon
daemonlog   run the class named CLASSNAME
or
CLASSNAME

Most commands print help when invoked w/o parameters.

C:\Users\Admin>hadoop -version
java version "1.8.0_421"
Java(TM) SE Runtime Environment (build 1.8.0_421-b09)
Java HotSpot(TM) 64-Bit Server VM (build 25.421-b09, mixed mode)

C:\Users\Admin>hadoop version
Hadoop 3.4.0
Source code repository git@github.com:apache/hadoop.git -r bd8b77f398f626bb7791783192ee7a5dffaec760
Compiled by root on 2024-03-04T06:35Z
Compiled on platform linux-x86_64
Compiled with protoc 3.21.12
From source with checksum f7fe694a3613358b38812ae9c31114e
This command was run using /C:/hadoop/share/hadoop/common/hadoop-common-3.4.0.jar

C:\Users\Admin>

```

3)Edit Specific File as given below

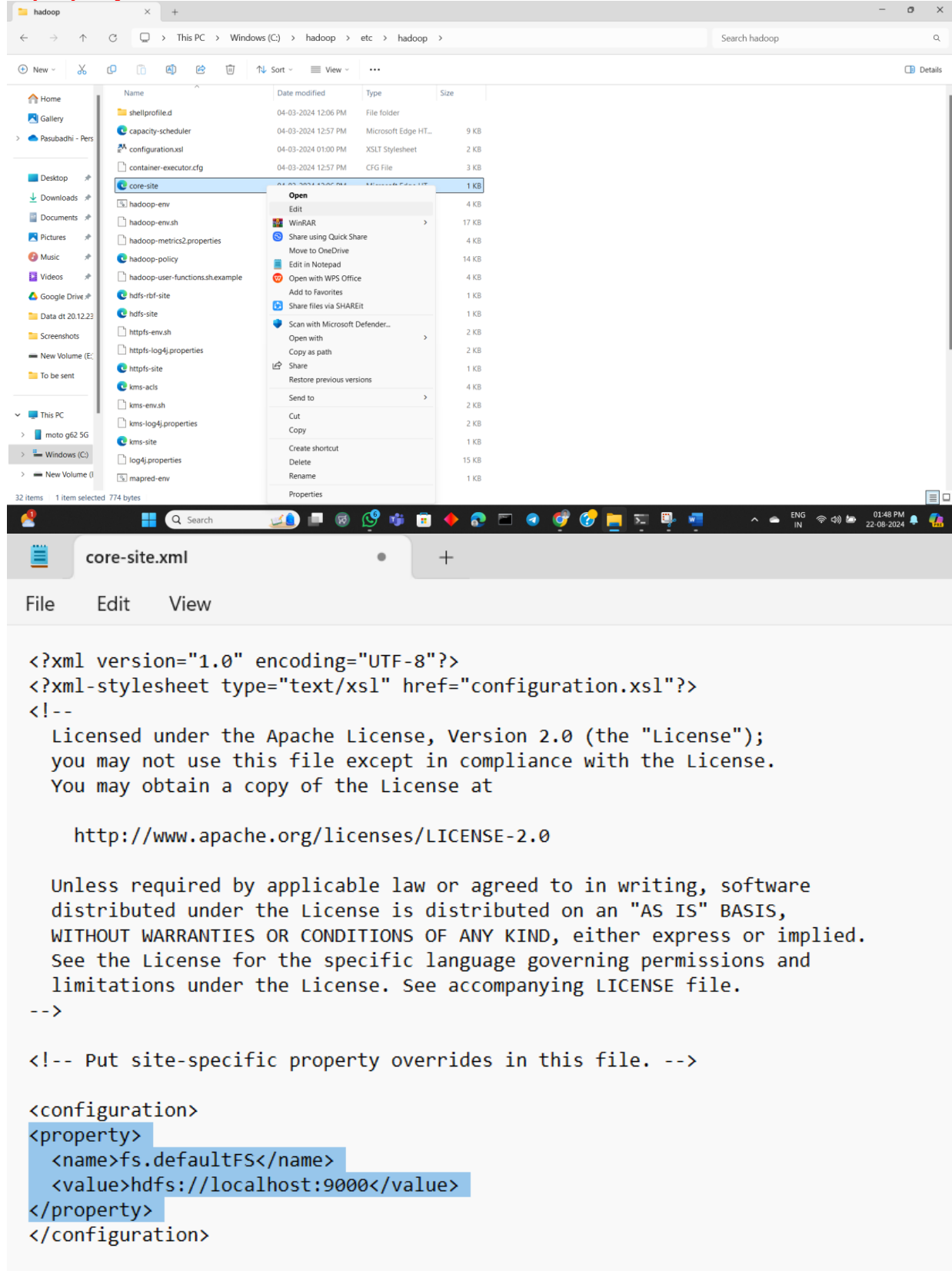
A)# For core-site.xml

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

</property>



```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

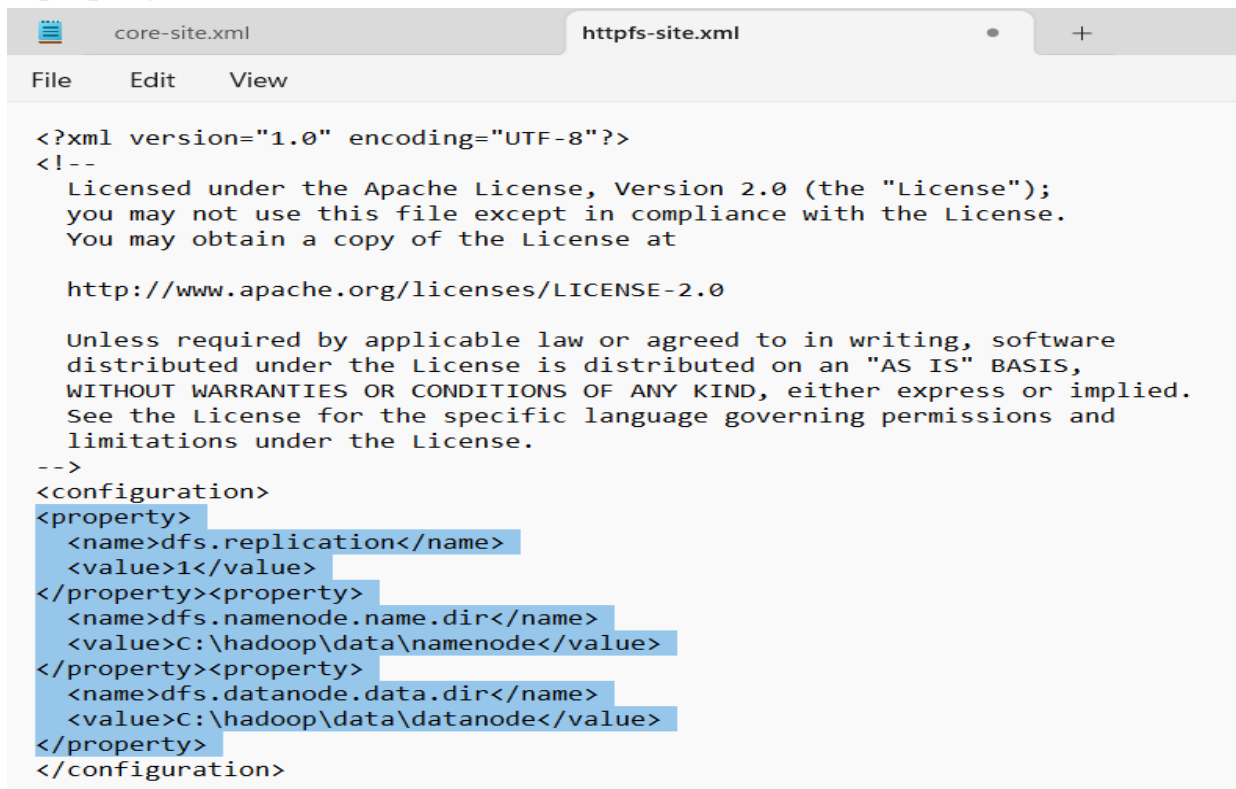
Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

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


B) # For hdfs-site.xml or https-site.xml

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



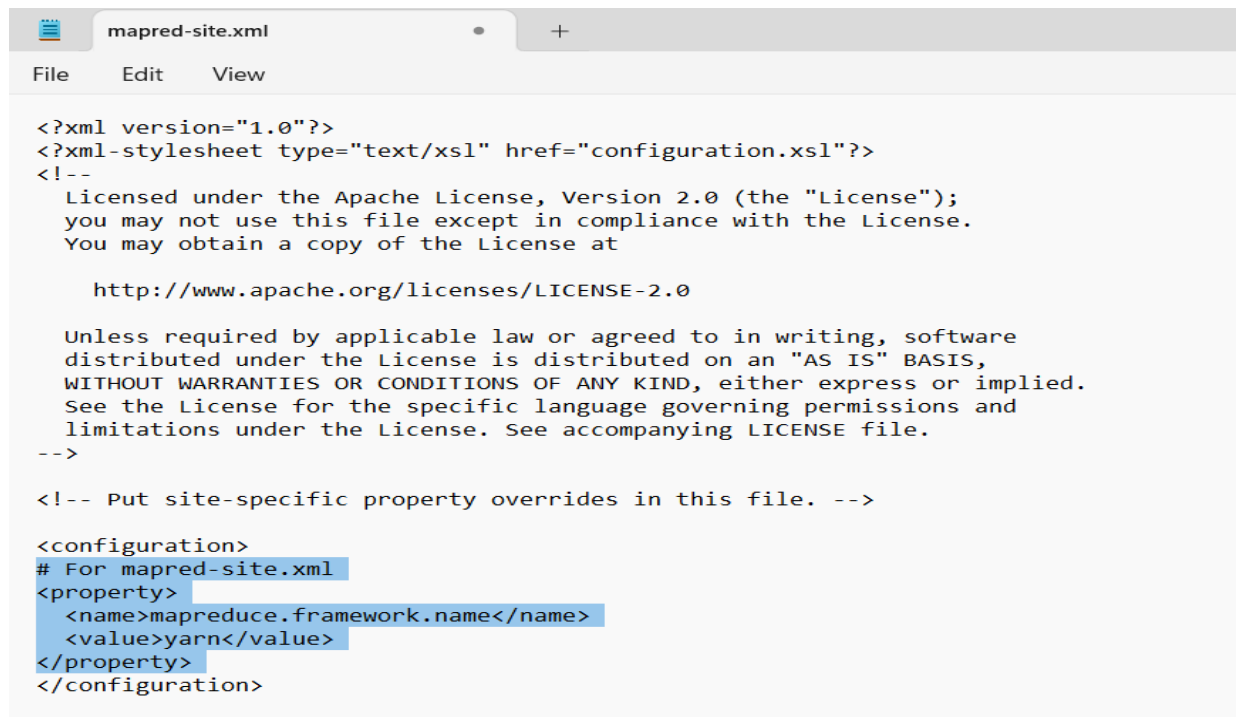
```
<?xml version="1.0" encoding="UTF-8"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License.
-->
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property><property>
    <name>dfs.namenode.name.dir</name>
    <value>C:\hadoop\data\namenode</value>
  </property><property>
    <name>dfs.datanode.data.dir</name>
    <value>C:\hadoop\data\datanode</value>
  </property>
</configuration>
```

C)# For mapred-site.xml

```
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
```



```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

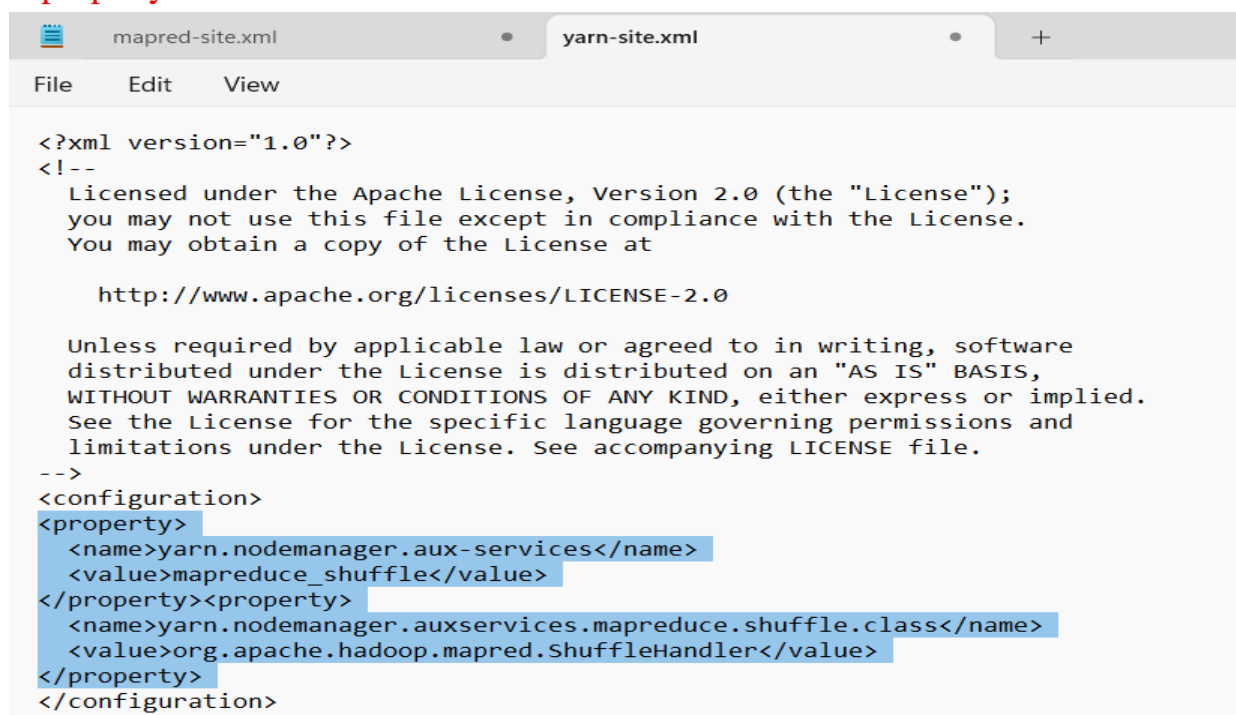
  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
# For mapred-site.xml
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
</configuration>
```

D)# For yarn-site.xml

```
<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>
```



```
<?xml version="1.0"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->
<configuration>
<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>
```

4)Delete Hadoop Old Bin & Deploy New Bin


Link:

https://drive.google.com/file/d/1nCN_jK7EJF2DmPUUxgOggvJ6k6tkYz/view

5)Download

a) MSVCR120 dll

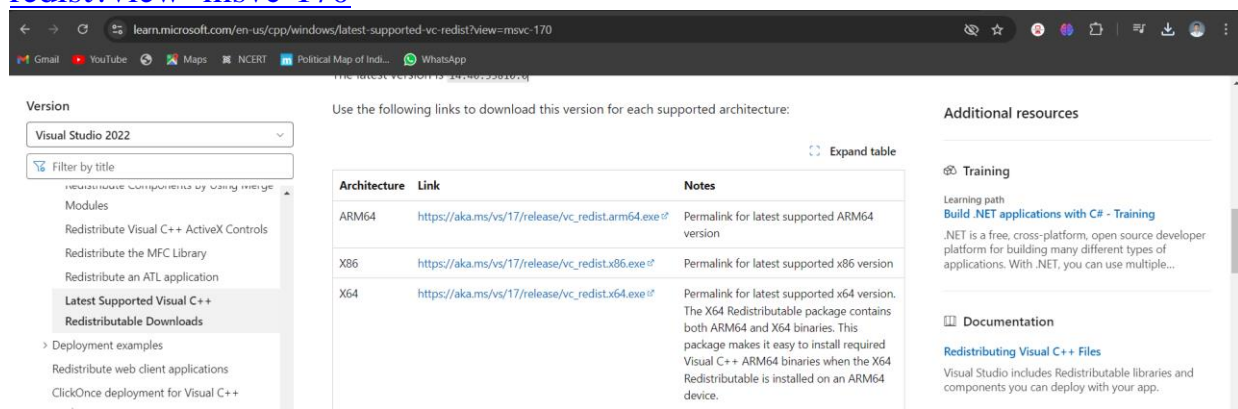
LINK: <https://www.dll-files.com/msvcr120.dll.html>

Version	12.0.40664.0	MD5:	b70474fe249402e251a94753b742788c
Architecture	64	SHA-1:	f53b3c21adf75dc84977067869253e207f1b9795
File size	0.92 MB		
Language	U.S. English		
Company	Microsoft Corporation	 Download	
Description	Microsoft® C Runtime Library	Zip file size:	0.44 MB

After Download Copy paste this file to windows->System32

b)MSVC-170

Link: <https://learn.microsoft.com/en-us/cpp/windows/latest-supported-vc-redist?view=msvc-170>



Version: Visual Studio 2022

Use the following links to download this version for each supported architecture:

Architecture	Link	Notes
ARM64	https://aka.ms/vs/17/release/vc_redist.arm64.exe	Permalink for latest supported ARM64 version
X86	https://aka.ms/vs/17/release/vc_redist.x86.exe	Permalink for latest supported x86 version
X64	https://aka.ms/vs/17/release/vc_redist.x64.exe	Permalink for latest supported x64 version. The X64 Redistributable package contains both ARM64 and X64 binaries. This package makes it easy to install required Visual C++ ARM64 binaries when the X64 Redistributable is installed on an ARM64 device.

Additional resources:

- Training: Build .NET applications with C# - Training
- Documentation: Redistributing Visual C++ Files

5)Commands

Type the Following Command in CMD

- I. hdfs namenode -format
- II. cd\
- III. cd hadoop
- IV. cd sbin
- V. start-dfs.cmd
- VI. jps
- VII. start-yarn.cmd
- VIII. jps

6)Brows in Google Chrome

- I. localhost:9870
- II. localhost:8088

7)To Stop

Type in cmd: Stop-all.cmd