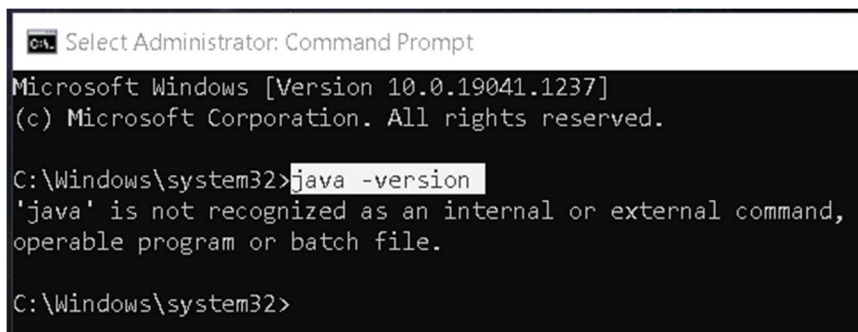


# Hadoop Installation:

Hadoop is written in JAVA. When you are using Hadoop, you need JDK installed on your machine. Prerequisites to install Hadoop on Windows 10 Operating System is Java and the version needed is 8.

## Part 1: Java

First check if your computer has Java installed on your Windows Operating System or not by using the command prompt (in administrator mode of course) and entering the command "java -version"



```
Microsoft Windows [Version 10.0.19041.1237]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>java -version
'java' is not recognized as an internal or external command,
operable program or batch file.

C:\Windows\system32>
```


If 'java' is not recognized as an internal or external command then you will need to get JDK from the Oracle website which is available for free.

You can choose any 1 from the link: <https://www.oracle.com/java/technologies/downloads/#java8-windows>

Or you can search for Oracle Java SE Development Kit 8u202 on Google and it will provide you the below link that I have used:

<https://www.oracle.com/in/java/technologies/javase/javase8-archive-downloads.html>

Scroll down on the webpage until you find the Windows 64-bit version JDK to download.

Windows x64	211.58 MB	 <a href="#">jdk-8u202-windows-x64.exe</a>
-------------	-----------	---

After you click on the JDK link it will ask you to agree to the terms and then ask you to sign in using your Oracle account.

Download Button with accept prompt:

×

You must accept the [Oracle Binary Code License Agreement for the Java SE Platform Products](#) to download this software.

☐ I reviewed and accept the Oracle Binary Code License Agreement for the Java SE Platform Products  
Required

Download jdk-8u202-windows-x64.exe 

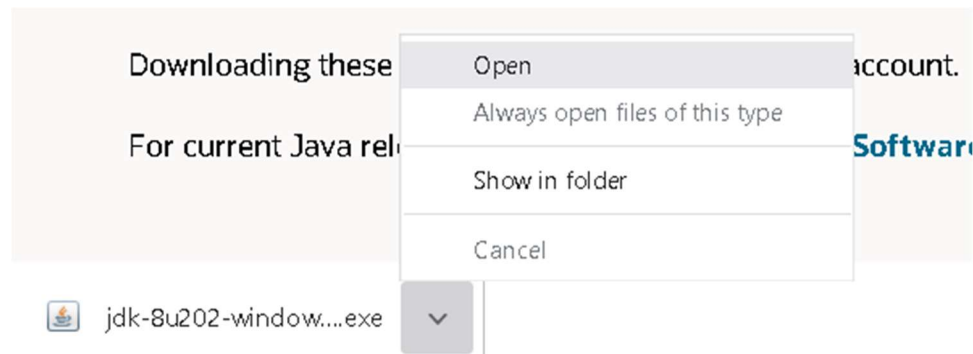
If you don't have an Oracle account simply choose the create account option.

## Don't have an Oracle Account?

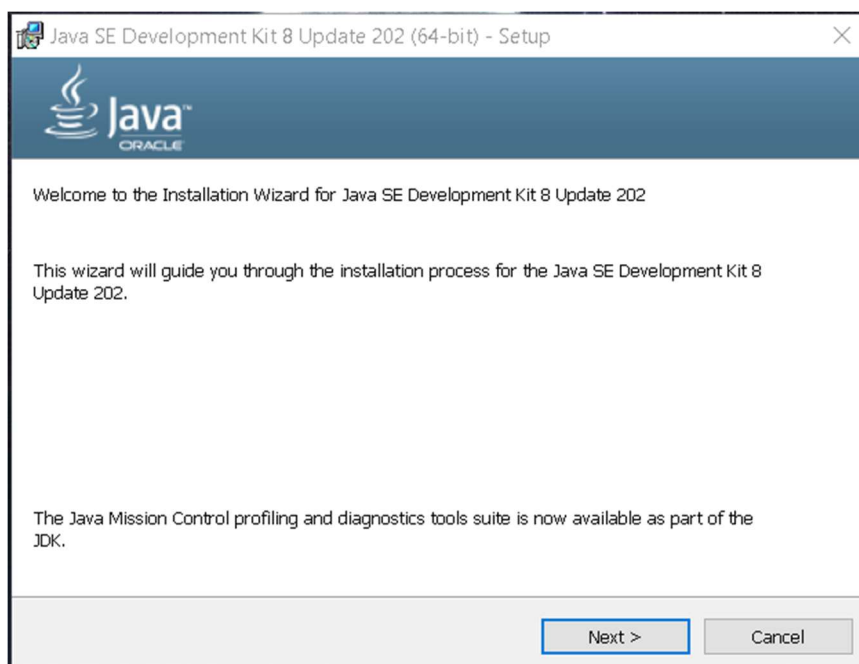
Create Account

© Oracle | [Terms of Use](#) | [Privacy Policy](#)

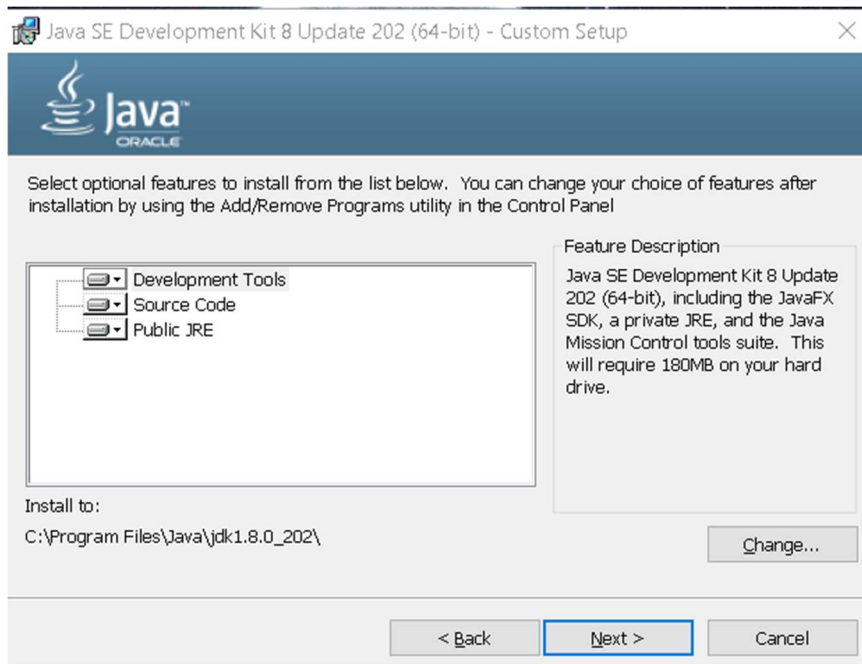
Once the download is complete you can simply open the file.



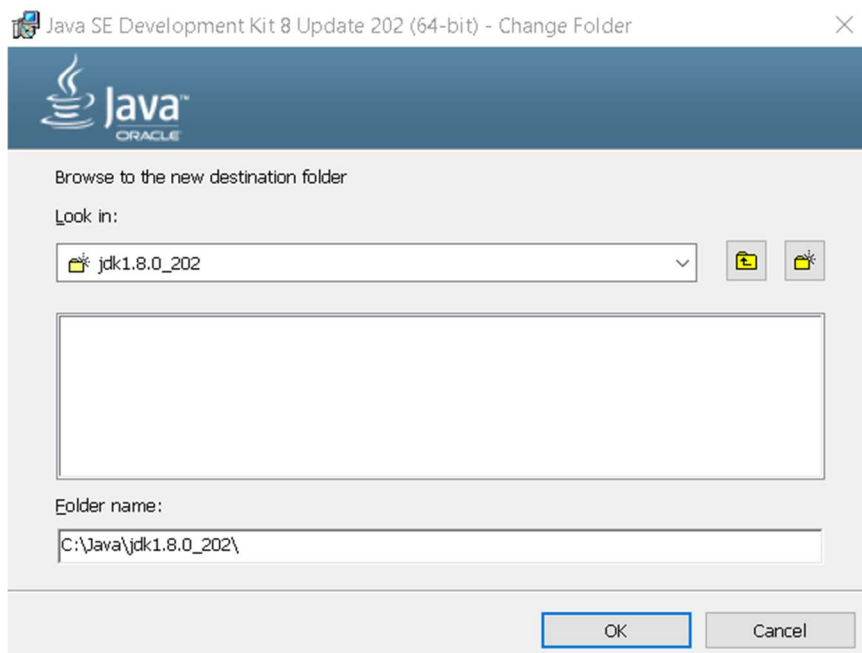
It takes few seconds for the setup window to appear after you have clicked on Open.



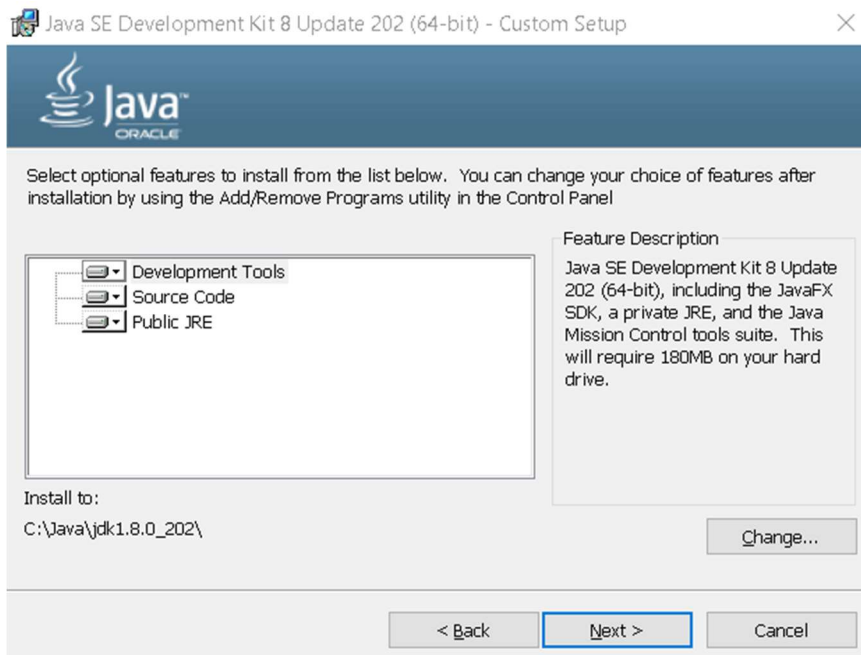
After clicking on the Next button, you will need to ensure that Java gets saved under the C drive directly and not under Program Files.



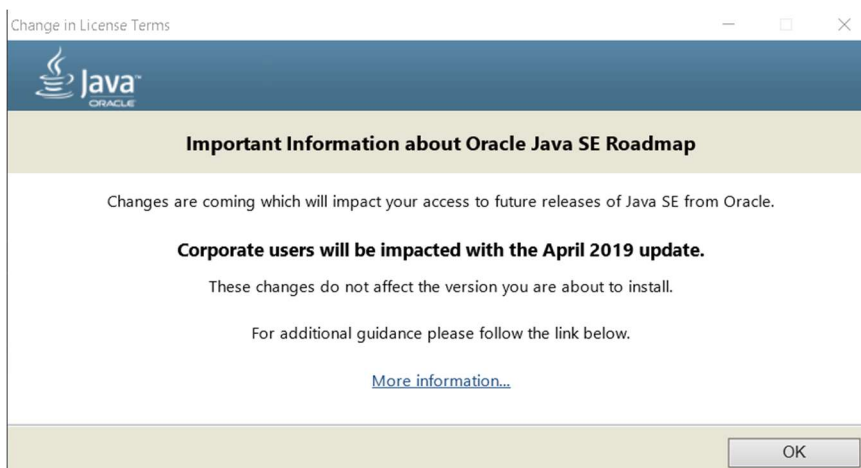
By default, it will take the Program Files option so we need to change the path to C drive before we click on the Next button.



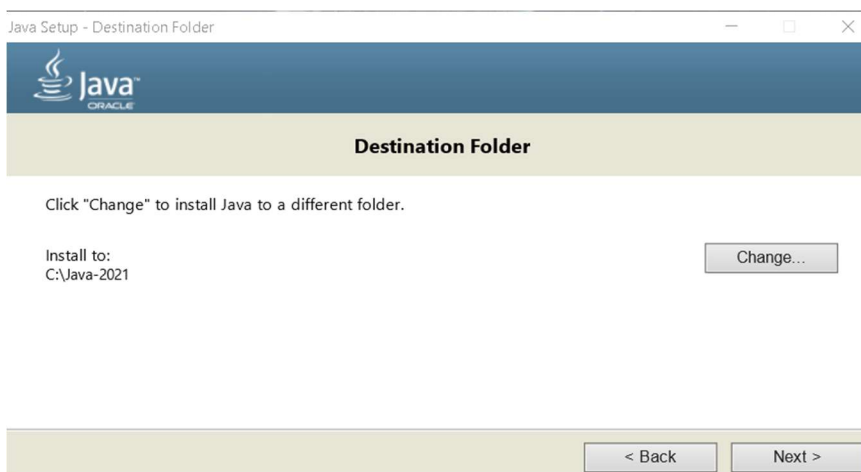
After editing the change folder click on the ok response. Post which you can confirm the new path and click on the Next button. We are making sure that the Java folder gets saved in the C drive directly and not inside Program Files to avoid any future errors.



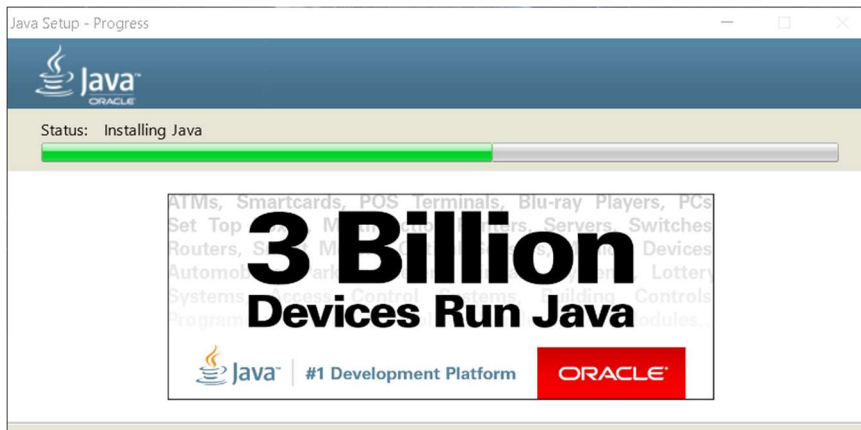
You can see a change in license terms pop up window which needs a simple OK button click.



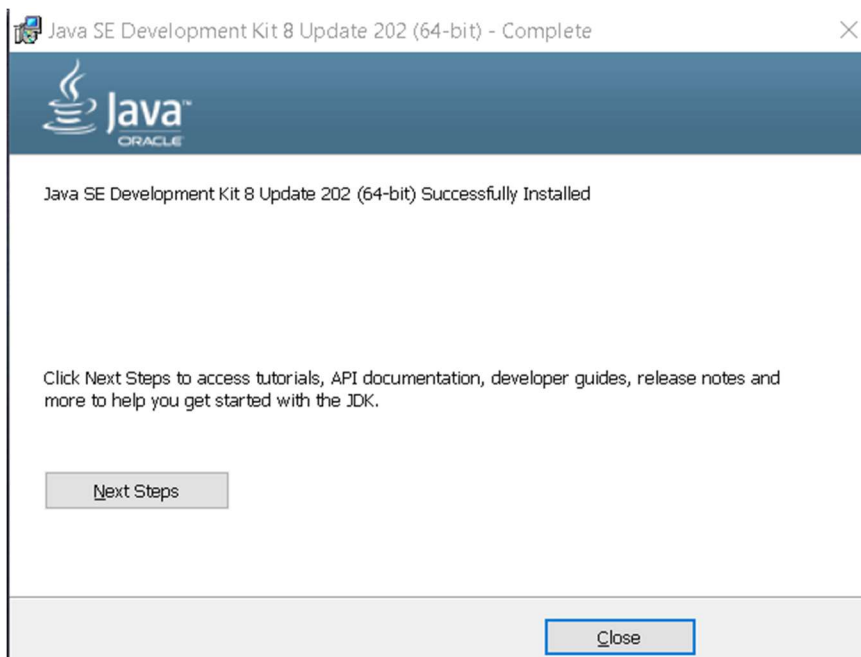
Choose your destination folder and click on the next button.



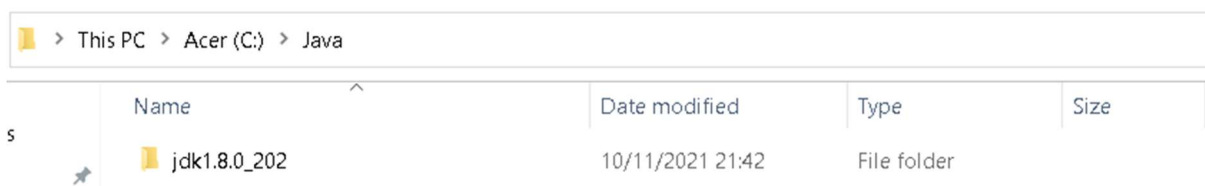
The setup gets initiated and shows the progress bar for it.



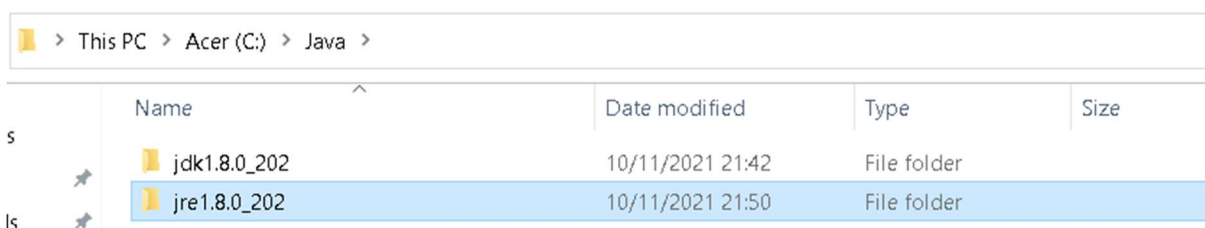
In the installation gets completed in seconds and provides us with the below prompt.



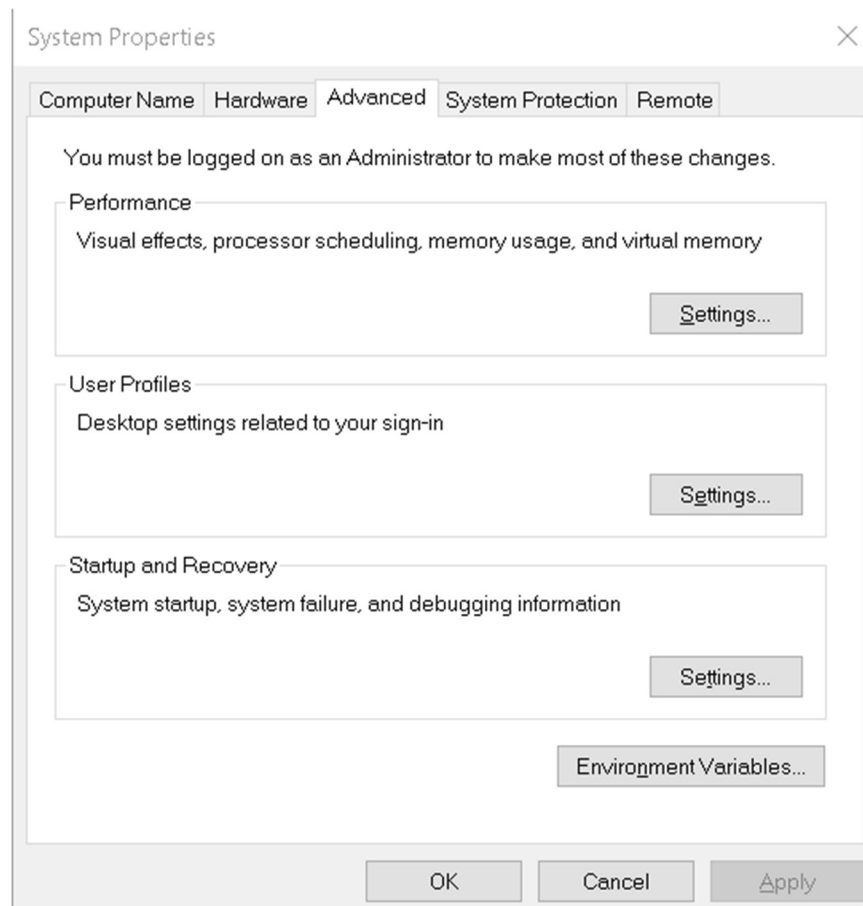
Then we need to go to our File Explorer and locate the Java folder in our C drive.



Here we will need to create a new folder named "jre1.8.0\_202" just below the "jdk1.8.0\_202" folder.

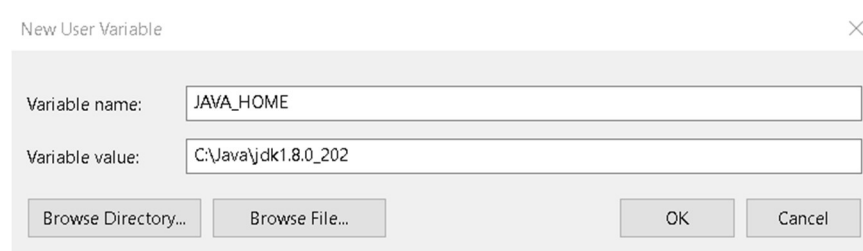


Then you will need to open System Properties by Right Clicking on This PC > Properties > System Protection > Advanced (tab) > Environment Variables

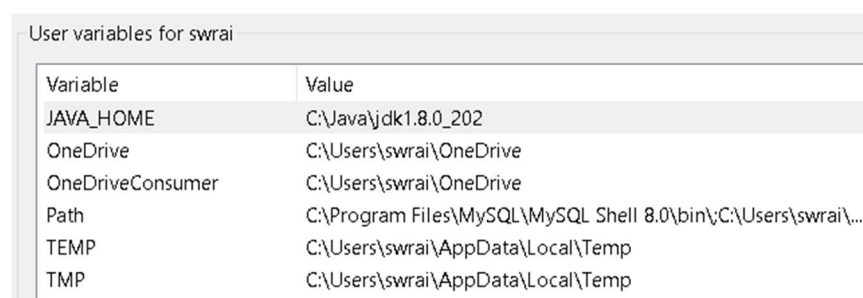


After clicking on the Environment Variables option, we will need create a new User Variable (Beware do not touch the System Variables)

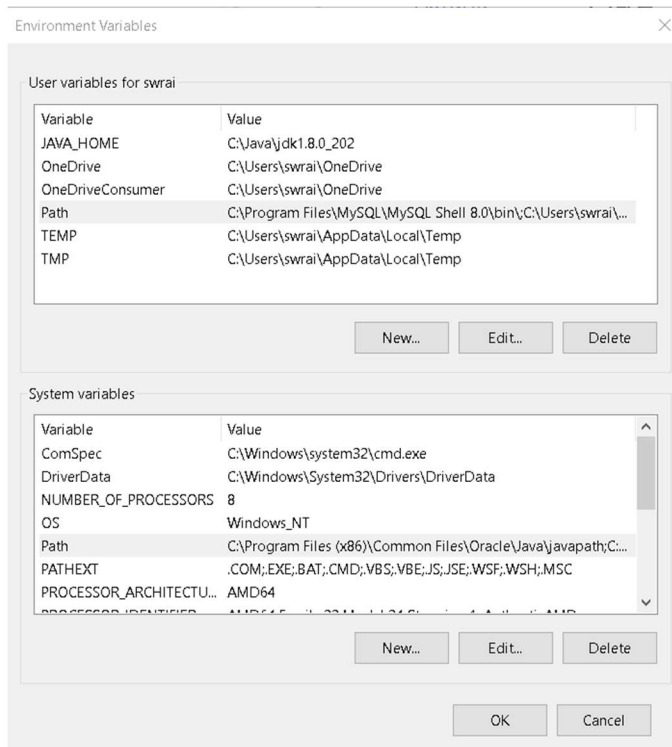
Type in the exact name and value as shown below.



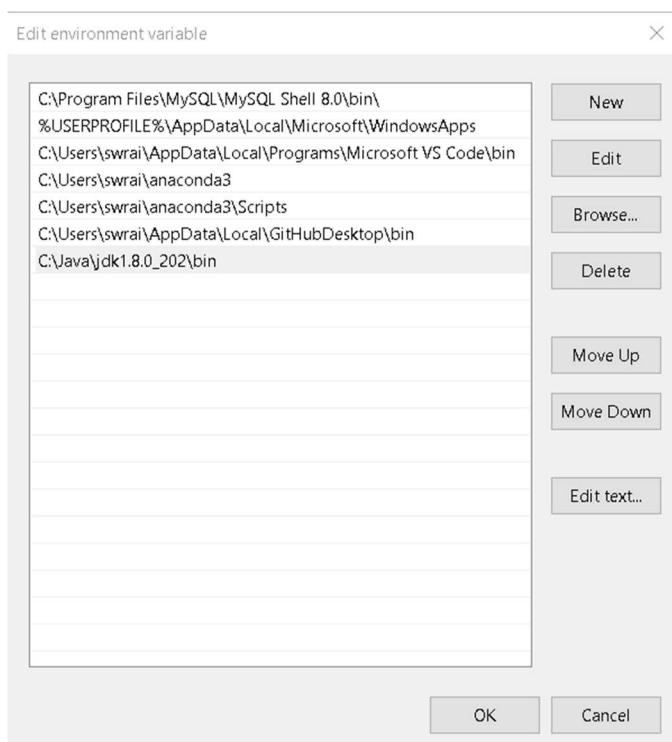
Once you click on "OK" you will get to see the newly created user variable in the list.



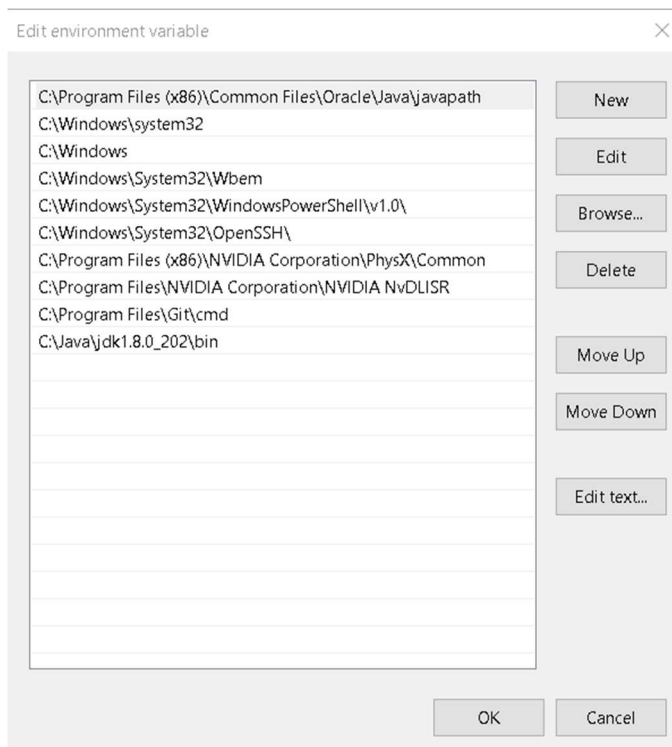
Now we will need to add Path “C:\Java\jdk1.8.0\_202\bin” for both user variable and system variable section. You can open the Path by simply double clicking on the name to add New.



User variable Path:



System variable Path:



Now simply click on OK buttons to close the dialogue boxes.

Finally, confirm whether the Java got installed correctly by following the first step and you should be able to see the version this time.

```
C:\> Select Administrator: Command Prompt
Microsoft Windows [Version 10.0.19041.1237]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32> java -version
java version "1.8.0_202"
Java(TM) SE Runtime Environment (build 1.8.0_202-b08)
Java HotSpot(TM) 64-Bit Server VM (build 25.202-b08, mixed mode)

C:\Windows\system32>
```

Here 1.8.0\_202 reflects the java version that we have successfully installed on the Windows 10 system.

## Part 2: Hadoop

Now to get Hadoop we will need to download the stable release of it that can be found here <https://hadoop.apache.org/releases.html>

The trick is to identify the most recent release and select the one prior to it since that will be the stable version.

Once you have decided on the version you will need to click on Binary download for it.



Version	Release date	Source download	Binary download	Release notes
3.3.1	2021 Jun 15	<a href="#">source (checksum signature)</a>	<a href="#">binary (checksum signature)</a> <a href="#">binary-aarch64 (checksum signature)</a>	<a href="#">Announcement</a>
3.2.2	2021 Jan 9	<a href="#">source (checksum signature)</a>	<a href="#">binary (checksum signature)</a>	<a href="#">Announcement</a>
2.10.1	2020 Sep 21	<a href="#">source (checksum signature)</a>	<a href="#">binary (checksum signature)</a>	<a href="#">Announcement</a>

It will then direct you to the mirror download link where you can get the tar.gz file.



COMMUNITY-LED DEVELOPMENT "THE APACHE WAY"

[Projects](#)
[People](#)
[Community](#)
[License](#)
[Sponsors](#)

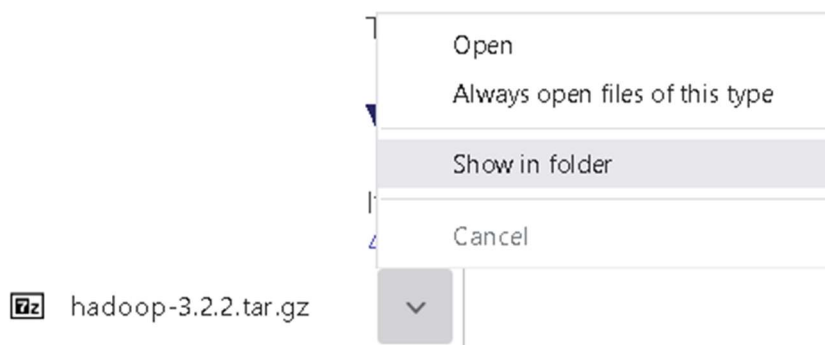


---

We suggest the following mirror site for your download:

<https://mirrors.estointernet.in/apache/hadoop/common/hadoop-3.2.2/hadoop-3.2.2.tar.gz>

Once the download is complete open the folder where the file has been saved.



This tar.gz file now needs to be extracted first. If you have WinZip you can do the extraction right away but for someone like me who did not have it available, I chose to use the 7zip for extracting files.

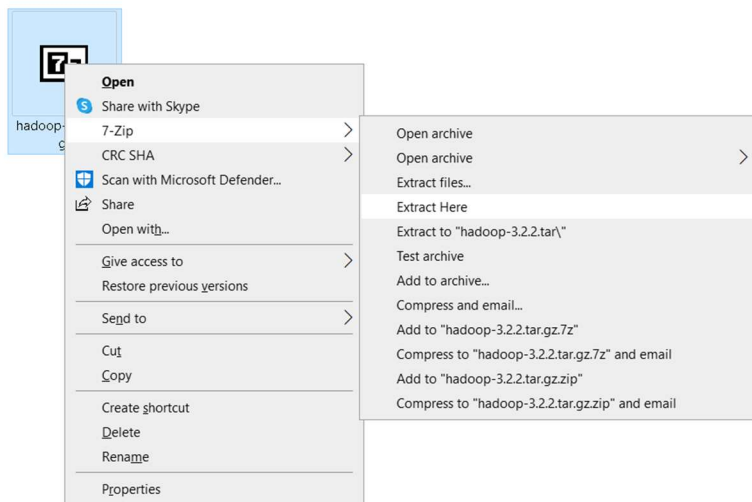
Look for 7-zip on the internet to land on their official webpage <https://www.7-zip.org/download.html>.

Choose the download link exe file and get it installed on your computer.

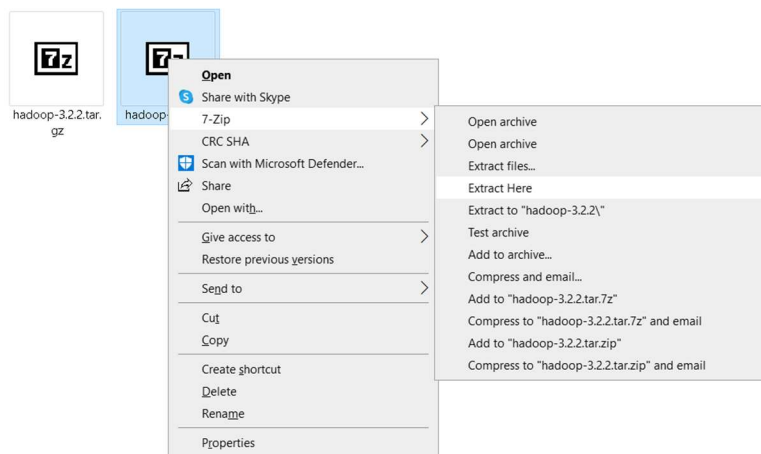
Download 7-Zip 19.00 (2019-02-21) for Windows:

Link	Type	Windows	Description
<a href="#">Download</a>	<a href="#">.exe</a>	<a href="#">64-bit x64</a>	<a href="#">7-Zip for 64-bit Windows x64 (Intel 64 or AMD64)</a>
<a href="#">Download</a>	<a href="#">.exe</a>	32-bit x86	7-Zip for 32-bit Windows
<a href="#">Download</a>	<a href="#">.7z</a>	x86 / x64	7-Zip Extra: standalone console version, 7z DLL, Plugin for Far Manager
<a href="#">Download</a>	<a href="#">.7z</a>	Any	7-Zip Source code
<a href="#">Download</a>	<a href="#">.7z</a>	Any / x86 / x64	LZMA SDK: (C, C++, C#, Java)
<a href="#">Download</a>	<a href="#">.msi</a>	64-bit x64	(alternative MSI installer) 7-Zip for 64-bit Windows x64 (Intel 64 or AMD64)
<a href="#">Download</a>	<a href="#">.msi</a>	32-bit x86	(alternative MSI installer) 7-Zip for 32-bit Windows

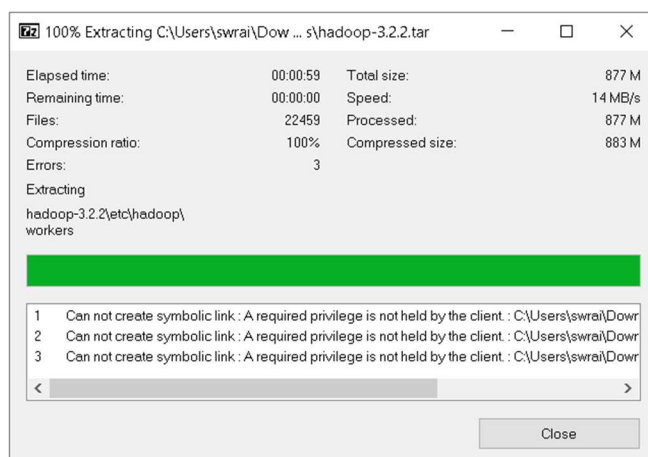
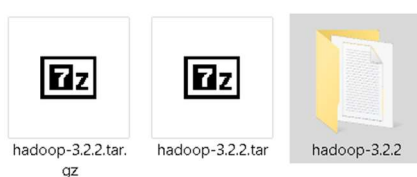
Now you will need to extract the Hadoop tar.gz file using the 7-zipFM option. This is basically for File Manager and provides you with prompts.



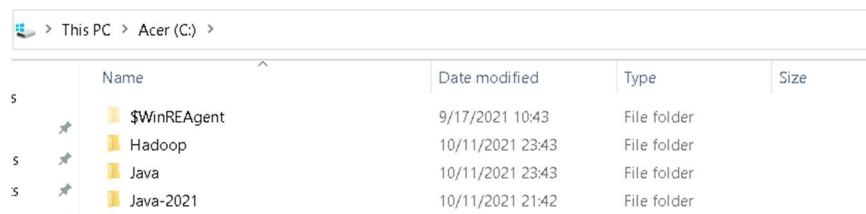
It will then give you a tar file that needs to be extracted again using the same process.



You will then get the Hadoop version named folder and a prompt of 100% extraction completion with 3 errors which can be ignored by clicking on the Close button.

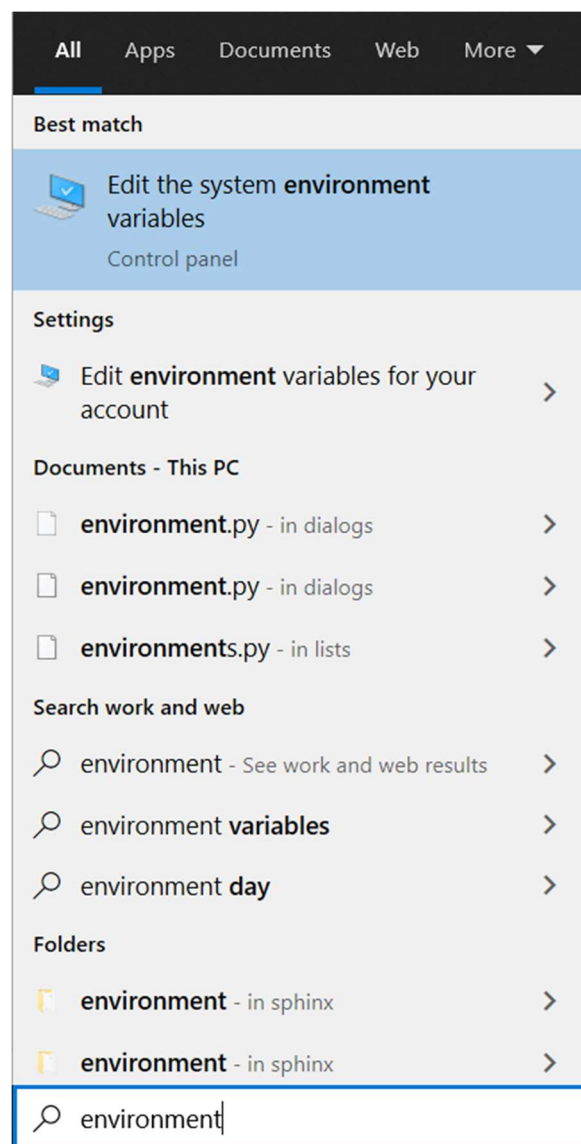


This Hadoop-3.2.2 folder now needs to be moved to your C drive exactly like your Java folder.

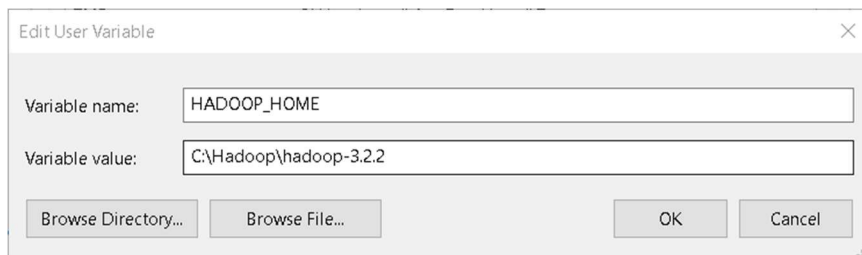


	Name	Date modified	Type	Size
5	\$WinREAgent	9/17/2021 10:43	File folder	
5	Hadoop	10/11/2021 23:43	File folder	
5	Java	10/11/2021 23:43	File folder	
5	Java-2021	10/11/2021 21:42	File folder	

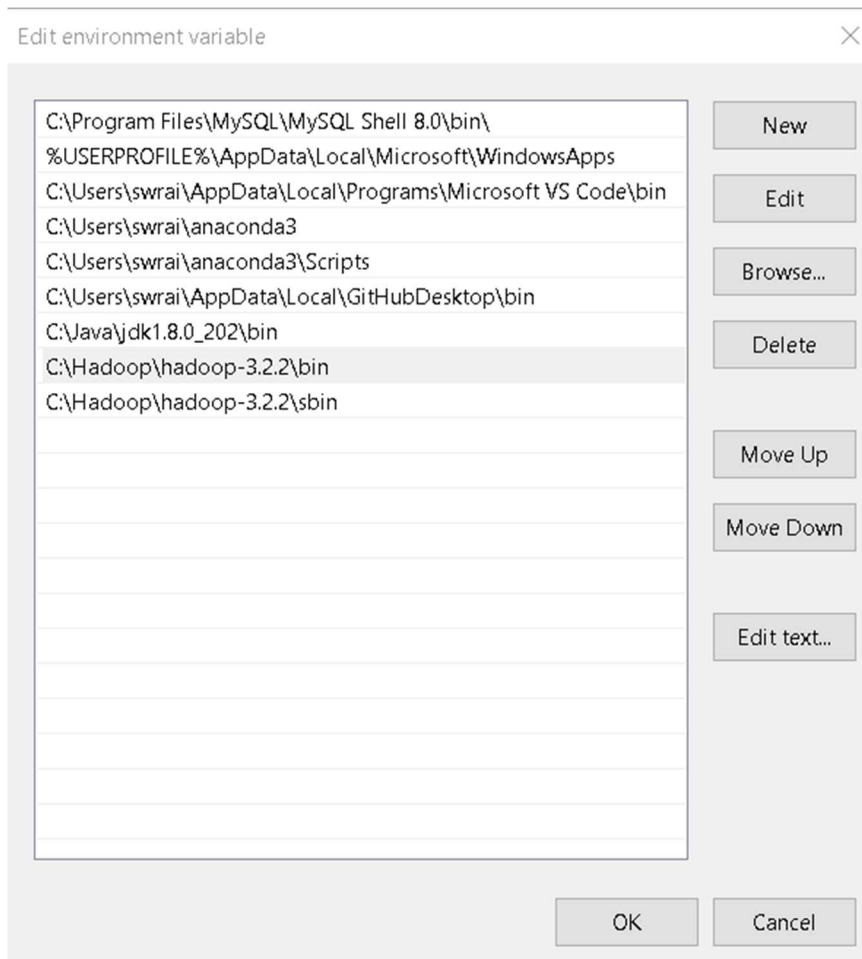
Again, like JAVA\_HOME we will need to create environmental variable for HADOOP\_HOME too. And easier way to get to it is by typing “environment” in your Windows 10 search box and the choosing the highlighted option “Edit the system environment variables”.



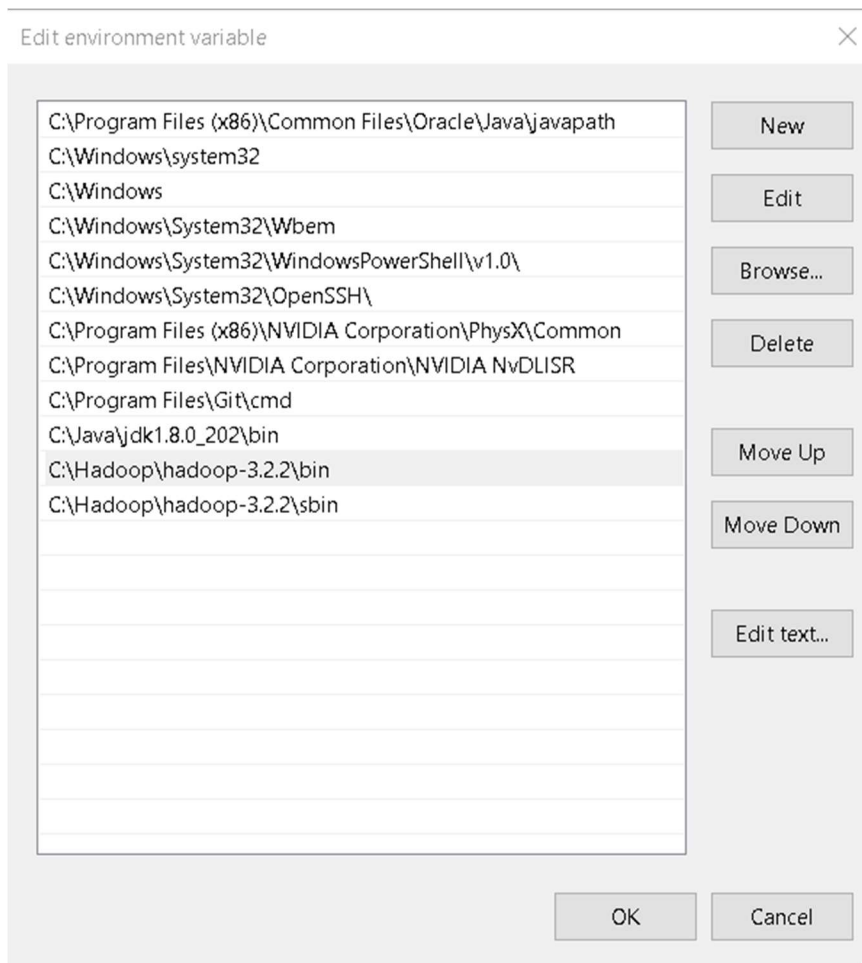
Create a new user variable name HADOOP\_HOME with the path as value.



Now we need to set the user variable path with bin and sbin locations.



Also need to set the system variable path with the bin and sbn locations.



Click on the “OK” buttons to close the System Properties dialogue box.

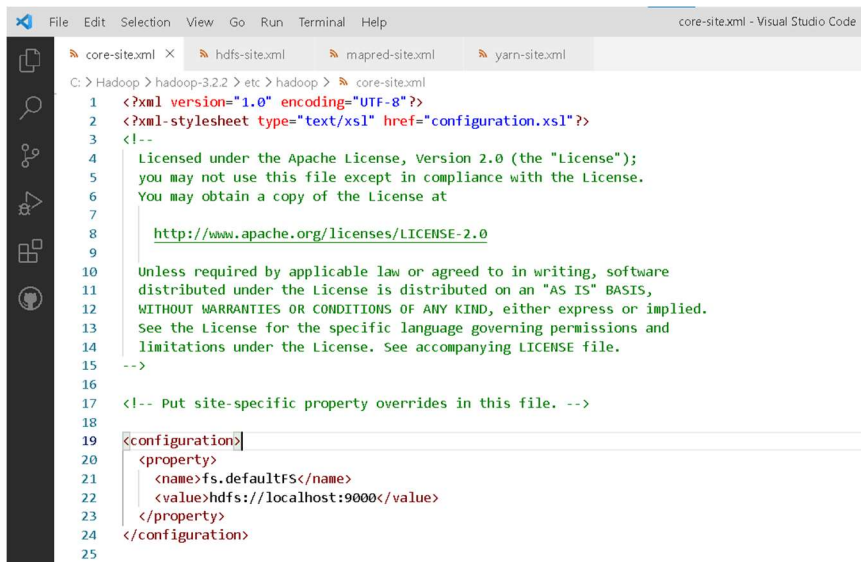
## Part 3: Configurations for Hadoop

It is time to edit some configuration files present in the Hadoop “etc” directory. For me the location is C:\Hadoop\hadoop-3.2.2\etc\hadoop

The 4 XML files that we need to edit are core-site, hdfs-site, mapred-site and yarn-site.

core-site.xml	10/11/2021 23:58	XML Document	1 KB
hadoop-policy.xml	1/3/2021 14:58	XML Document	12 KB
hdfs-site.xml	10/12/2021 00:14	XML Document	2 KB
httpfs-site.xml	1/3/2021 15:03	XML Document	1 KB
kms-acls.xml	1/3/2021 14:59	XML Document	4 KB
kms-site.xml	1/3/2021 14:59	XML Document	1 KB
mapred-site.xml	10/12/2021 00:00	XML Document	1 KB
yarn-site.xml	10/12/2021 00:06	XML Document	1 KB

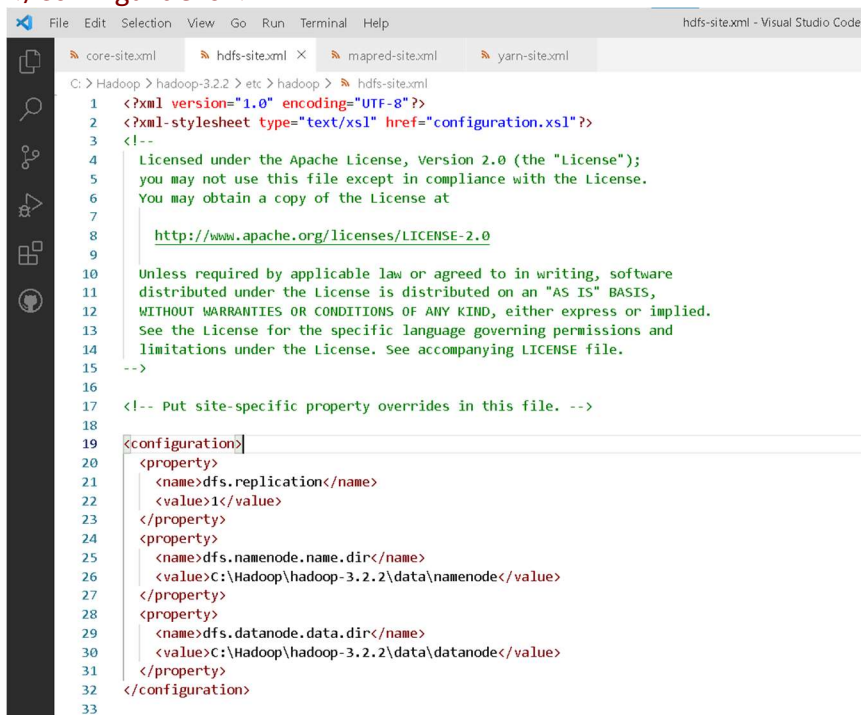
You can open these files in VS Code to edit and make changes. The XML files will have empty configuration location wherein we need to fill our own codes as shown below.



```
C:\> Hadoop > hadoop-3.2.2 > etc > hadoop > core-site.xml
1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
3 <!--
4 Licensed under the Apache License, Version 2.0 (the "License");
5 you may not use this file except in compliance with the License.
6 You may obtain a copy of the License at
7
8 http://www.apache.org/licenses/LICENSE-2.0
9
10 Unless required by applicable law or agreed to in writing, software
11 distributed under the License is distributed on an "AS IS" BASIS,
12 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13 See the License for the specific language governing permissions and
14 limitations under the License. See accompanying LICENSE file.
15 -->
16
17 <!-- Put site-specific property overrides in this file. -->
18
19 <configuration>
20   <property>
21     <name>fs.defaultFS</name>
22     <value>hdfs://localhost:9000</value>
23   </property>
24 </configuration>
25
```

Core-site XML code:

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



```
C:\> Hadoop > hadoop-3.2.2 > etc > hadoop > hdfs-site.xml
1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
3 <!--
4 Licensed under the Apache License, Version 2.0 (the "License");
5 you may not use this file except in compliance with the License.
6 You may obtain a copy of the License at
7
8 http://www.apache.org/licenses/LICENSE-2.0
9
10 Unless required by applicable law or agreed to in writing, software
11 distributed under the License is distributed on an "AS IS" BASIS,
12 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13 See the License for the specific language governing permissions and
14 limitations under the License. See accompanying LICENSE file.
15 -->
16
17 <!-- Put site-specific property overrides in this file. -->
18
19 <configuration>
20   <property>
21     <name>dfs.replication</name>
22     <value>1</value>
23   </property>
24   <property>
25     <name>dfs.namenode.name.dir</name>
26     <value>C:\Hadoop\hadoop-3.2.2\data\namenode</value>
27   </property>
28   <property>
29     <name>dfs.datanode.data.dir</name>
30     <value>C:\Hadoop\hadoop-3.2.2\data\datanode</value>
31   </property>
32 </configuration>
33
```

Hdfs-site XML code:

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
```

```

</property>
<property>
  <name>dfs.namenode.name.dir</name>
  <value>C:\Hadoop\hadoop-3.2.2\data\namenode</value>
</property>
<property>
  <name>dfs.datanode.data.dir</name>
  <value>C:\Hadoop\hadoop-3.2.2\data\datanode</value>
</property>
</configuration>

```

```

C:\Hadoop> hadoop-3.2.2> etc> hadoop> mapred-site.xml
1  <?xml version="1.0"?>
2  <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
3  <!--
4  Licensed under the Apache License, Version 2.0 (the "License");
5  you may not use this file except in compliance with the License.
6  You may obtain a copy of the License at
7
8  http://www.apache.org/licenses/LICENSE-2.0
9
10 Unless required by applicable law or agreed to in writing, software
11 distributed under the License is distributed on an "AS IS" BASIS,
12 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13 See the License for the specific language governing permissions and
14 limitations under the License. See accompanying LICENSE file.
15 -->
16
17 <!-- Put site-specific property overrides in this file. -->
18
19 <configuration>
20   <property>
21     <name>mapreduce.framework.name</name>
22     <value>yarn</value>
23   </property>
24 </configuration>
25

```

Mapred-site XML code:

```

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

```

```

C:\Hadoop> hadoop-3.2.2> etc> hadoop> yarn-site.xml
1  <?xml version="1.0"?>
2  <!--
3  Licensed under the Apache License, Version 2.0 (the "License");
4  you may not use this file except in compliance with the License.
5  You may obtain a copy of the License at
6
7  http://www.apache.org/licenses/LICENSE-2.0
8
9  Unless required by applicable law or agreed to in writing, software
10 distributed under the License is distributed on an "AS IS" BASIS,
11 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
12 See the License for the specific language governing permissions and
13 limitations under the License. See accompanying LICENSE file.
14 -->
15 <configuration>
16   <property>
17     <name>yarn.nodemanager.aux-services</name>
18     <value>mapreduce_shuffle</value>
19   </property>
20   <property>
21     <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
22     <value>org.apache.hadoop.mapred.ShuffleHandler</value>
23   </property>
24 </configuration>
25 <!-- Site specific YARN configuration properties -->

```



Yarn-site XML code:

```
<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>
```

Last file that needs to be configured is the Hadoop-env.cmd file. You can simply right click and choose the edit option to open the file in a notepad.

hadoop-env.cmd	10/12/2021 00:17	Windows Command Script	4 KB
mapred-env.cmd	1/3/2021 15:27	Windows Command Script	1 KB
yarn-env.cmd	1/3/2021 15:24	Windows Command Script	3 KB

Now look for the line that says JAVA\_HOME and remove everything between the % sign along with the % sign and replace it with the Java home location that was set in the environment variable as shown below.

```
@rem The java implementation to use. Required.
set JAVA_HOME=C:\Java\jdk1.8.0_202
```

Save all the XML and Notepad files with the configuration changes that were made.

Create a new folder named data under the bin file of your Hadoop directory for me the path is C:\Hadoop\hadoop-3.2.2

This PC > Acer (C:) > Hadoop > hadoop-3.2.2				
	Name	Date modified	Type	Size
	bin	10/12/2021 00:26	File folder	
	data	10/12/2021 00:08	File folder	
	etc	1/3/2021 14:59	File folder	
	include	1/3/2021 15:41	File folder	
	lib	1/3/2021 15:41	File folder	
	libexec	1/3/2021 15:41	File folder	
	logs	10/12/2021 00:31	File folder	
	sbin	1/3/2021 14:59	File folder	
	share	1/3/2021 16:16	File folder	
	LICENSE.txt	12/5/2020 20:39	Text Document	148 KB
	NOTICE.txt	12/5/2020 20:39	Text Document	22 KB
	README.txt	12/5/2020 20:39	Text Document	2 KB

Inside the data folder we will need to create 2 more folders called the datanode and namenode.



This PC > Acer (C:) > Hadoop > hadoop-3.2.2 > data				
	Name	Date modified	Type	Size
	datanode	10/12/2021 00:31	File folder	
	namenode	10/12/2021 00:31	File folder	

The last step will be to replace the original bin folder in Hadoop with proper configuration files that can be downloaded from <https://github.com/s911415/apache-hadoop-3.1.3-winutils> in zip format.

s911415 / apache-hadoop-3.1.3-winutils Public

<> Code Issues Pull requests Actions Projects Wiki Security Insights

master 1 branch 0 tags

Go to file Add file Code

s911415 first commit

bin first commit

Clone

HTTPS SSH GitHub CLI

<https://github.com/s911415/apache-hadoop-3.1.3-winutils>

Use Git or checkout with SVN using the web URL.

Open with GitHub Desktop

Download ZIP

Once the ZIP folder is extracted simply replace this extracted bin with the Hadoop bin folder.

This PC > Acer (C:) > Hadoop > hadoop-3.2.2				
	Name	Date modified	Type	Size
	bin	10/12/2021 00:26	File folder	
	data	10/12/2021 00:08	File folder	
	etc	1/3/2021 14:59	File folder	
	include	1/3/2021 15:41	File folder	
	lib	1/3/2021 15:41	File folder	
	libexec	1/3/2021 15:41	File folder	
	logs	10/12/2021 00:31	File folder	
	sbin	1/3/2021 14:59	File folder	
	share	1/3/2021 16:16	File folder	
	LICENSE.txt	12/5/2020 20:39	Text Document	148 KB
	NOTICE.txt	12/5/2020 20:39	Text Document	22 KB
	README.txt	12/5/2020 20:39	Text Document	2 KB

## Part 4: Checking Hadoop Installation

Here comes the part where we will cross check and confirm that all the above configurations were done correctly. To do so we need to open the command prompt in administrator mode.

### 1. Hadoop Version

```
C:\Windows\system32>hadoop version
Hadoop 3.2.2
Source code repository Unknown -r 7a3bc90b05f257c8ace2f76d74264906f0f7a932
Compiled by hexiaoqiao on 2021-01-03T09:26Z
Compiled with protoc 2.5.0
From source with checksum 5a8f564f46624254b27f6a33126ff4
This command was run using /C:/Hadoop/hadoop-3.2.2/share/hadoop/common/hadoop-common-3.2.2.jar
C:\Windows\system32>
```

## 2. Hdfs namenode -format

```
C:\Windows\system32>hdfs namenode -format
2021-10-12 14:20:26,127 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = MissWhite/192.168.1.212
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.2.2
STARTUP_MSG: classpath = C:\Hadoop\hadoop-3.2.2\etc\hadoop;C:\Hadoop\hadoop-3.2.2\share\hadoop\
-3.2.2\share\hadoop\common\lib\accessors-smart-1.2.jar;C:\Hadoop\hadoop-3.2.2\share\hadoop\common
otations-1.17.jar;C:\Hadoop\hadoop-3.2.2\share\hadoop\common\lib\asm-5.0.4.jar;C:\Hadoop\hadoop-3
n\lib\audience-annotations-0.5.0.jar;C:\Hadoop\hadoop-3.2.2\share\hadoop\common\lib\avro-1.7.7.ja
2\share\hadoop\common\lib\checker-qual-2.5.2.jar;C:\Hadoop\hadoop-3.2.2\share\hadoop\common\lib\c
```

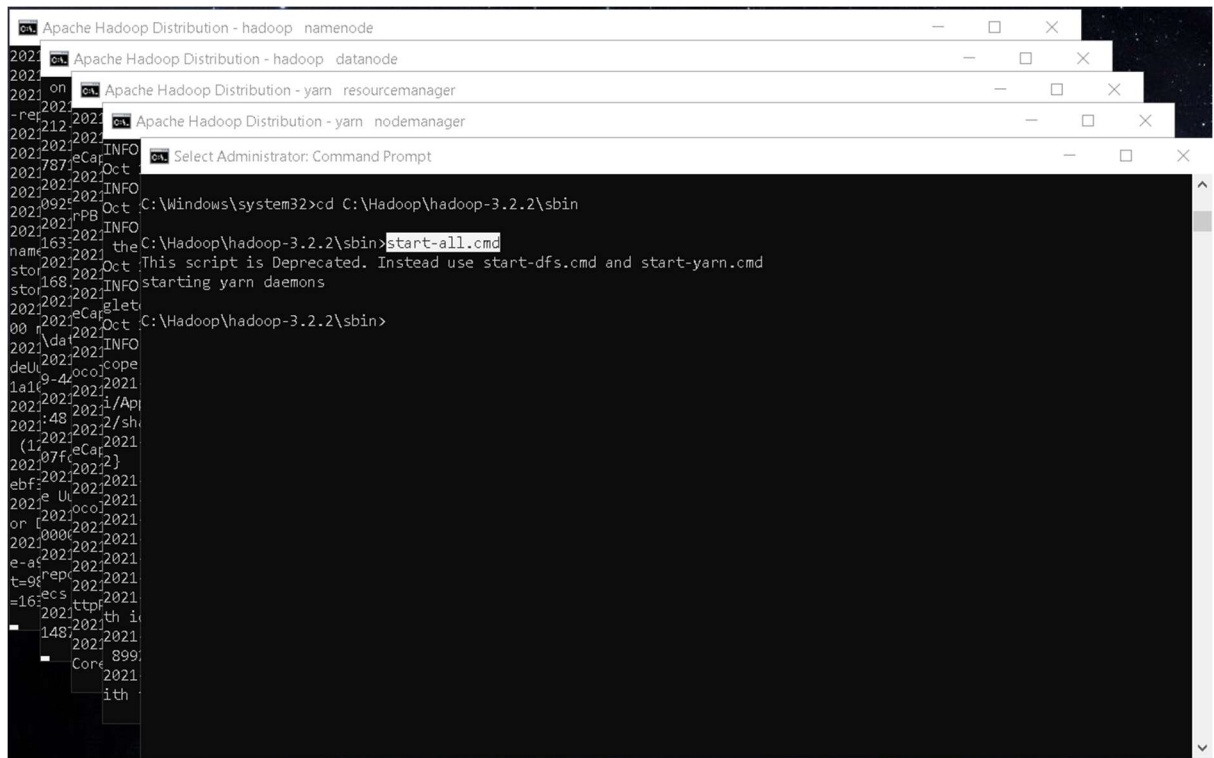
## 3. Now change the directory to sbin using cd option

```
C:\Windows\system32>cd C:\Hadoop\hadoop-3.2.2\sbin
C:\Hadoop\hadoop-3.2.2\sbin>
```

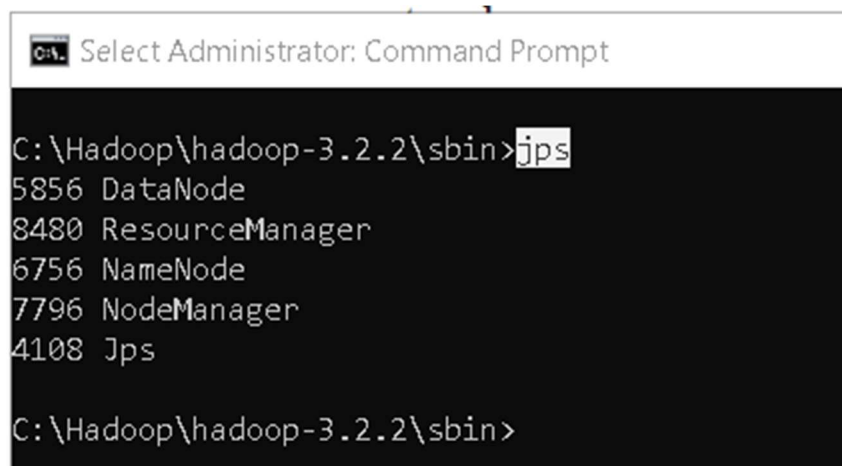
## 4. We will try to run all the Hadoop Daemons namely namenode, datanode, jobtracker and tasktracker. The command used is "start-all.cmd"

You will notice multiple CMD windows opening automatically and running for namenode, datanode, resourcemanager and nodemanager.

Allow them to keep running while you can bring your main Administrator: Command Prompt at the front.



5. JPS stands for Java Virtual Machine Process Status Tool or [JVM Process Status tool]



To access information about resource manager current jobs, successful and failed jobs visit this link in your browser - <http://localhost:8088/cluster>