Name: Muhammad Abeer

Roll no: 19P-0061
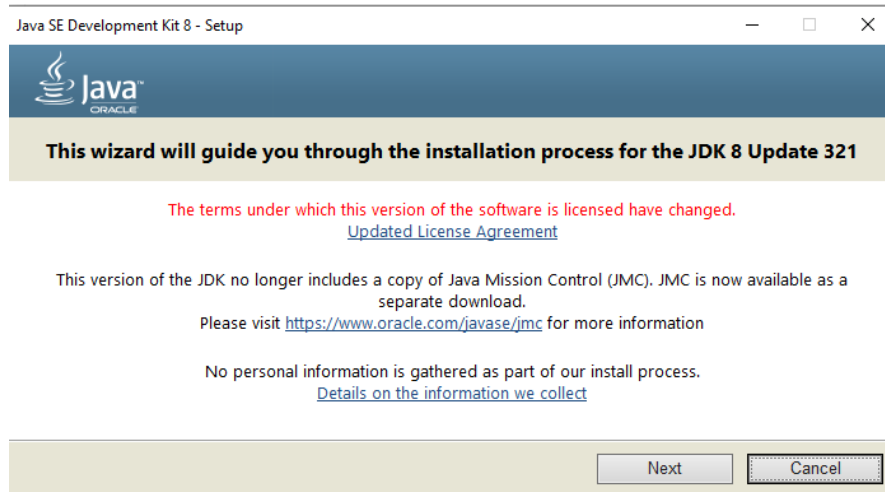
Section: 6-B

Deploying HADOOP on Windows

Step 1

Download and set up Java.

Go to: https://www.oracle.com/java/technologies/downloads/#java8

Make a directory in C naming Java and install it there. Then for installation of it's jdk files, keep the default directory i.e, **program files/java/jdk-17.0.1**

After successful completion of installation, go to **program files/java** and cut that **jdk-17.0.1** and paste it in **C://java.**

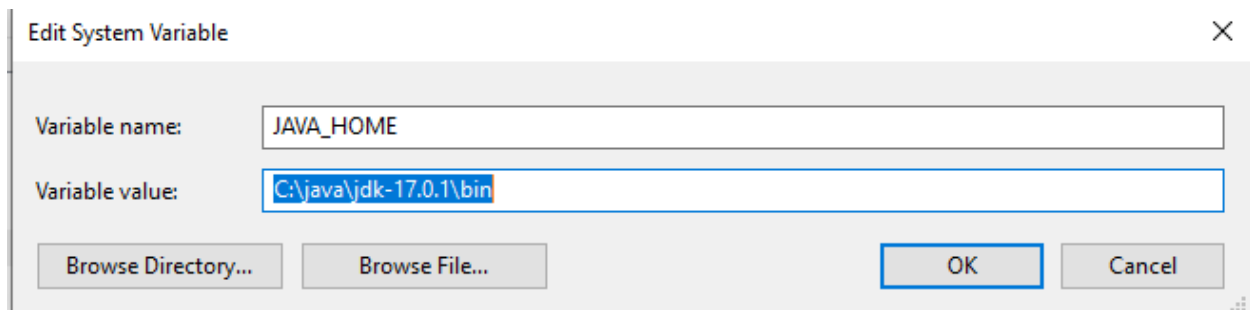Go to system environment variables by typing env in start.
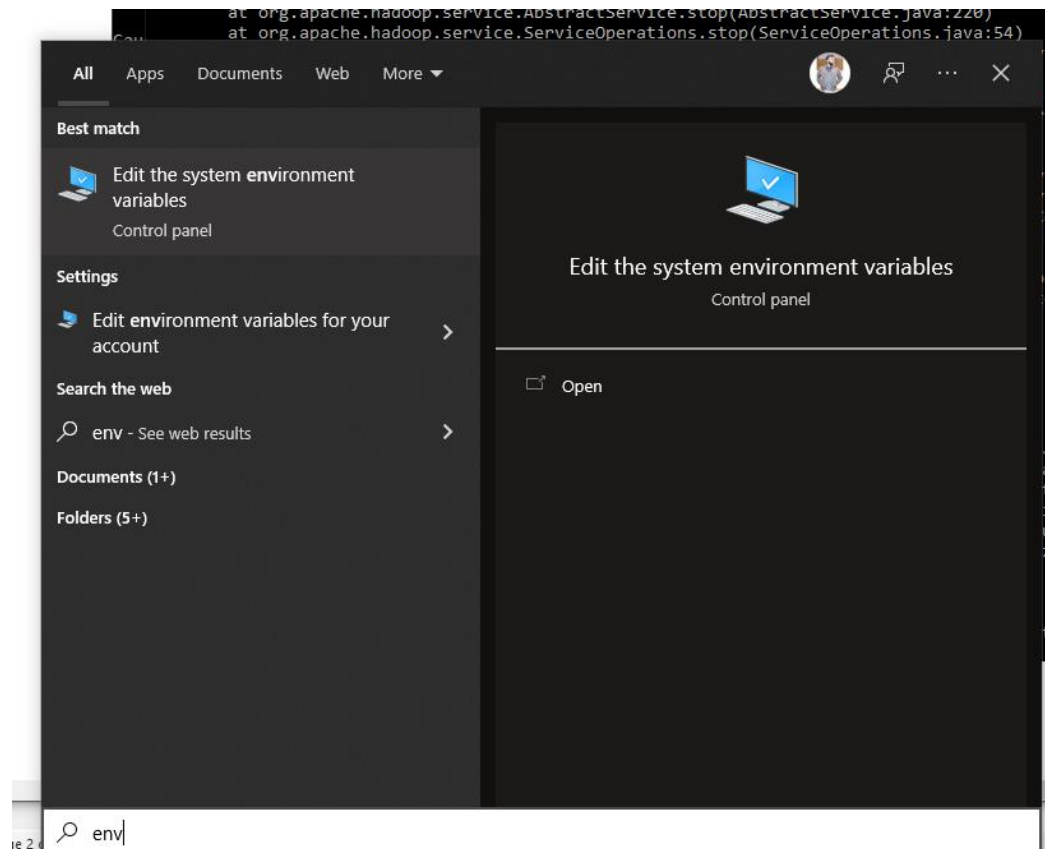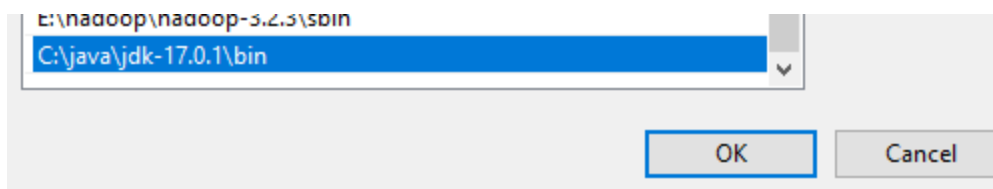
It will look like this
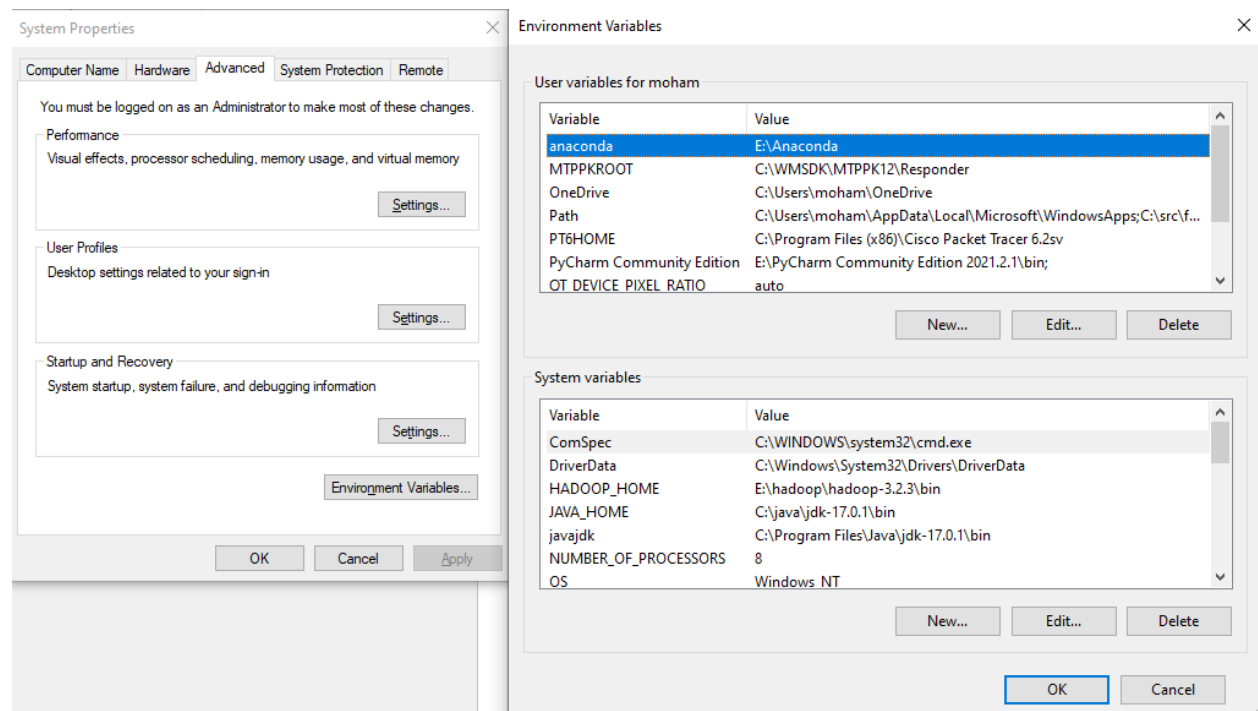
In system Variables, click on new and type new variable JAVA_HOME and set it's value to

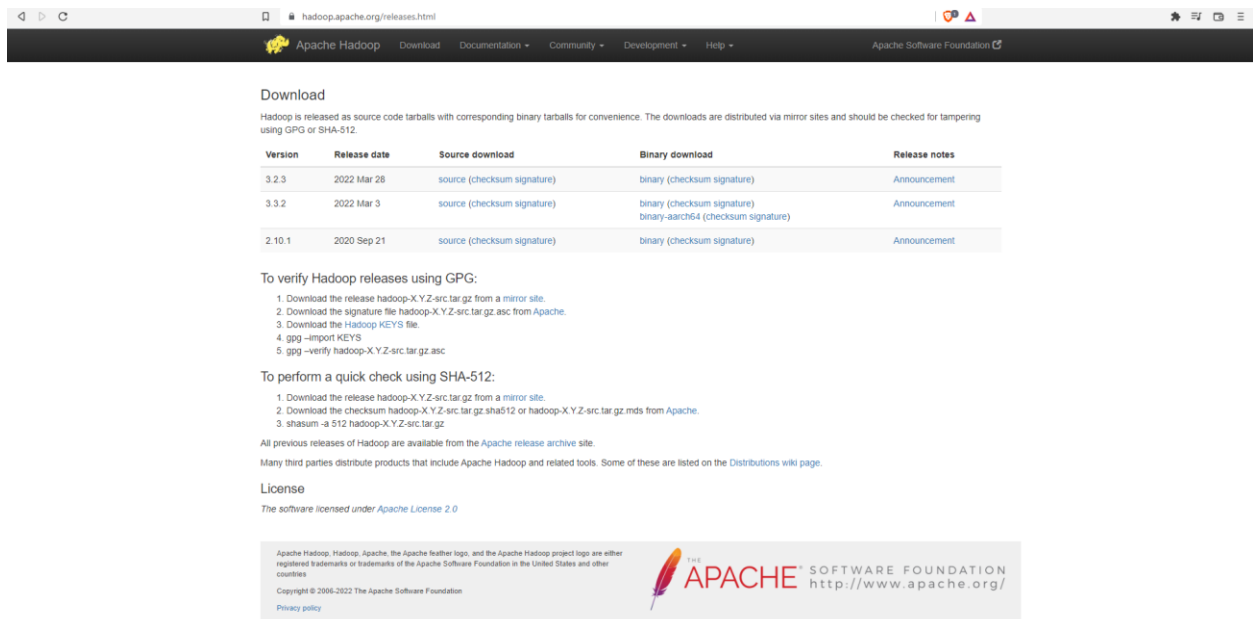**C:\Program Files\Java\jdk-17.0.1\bin**

Edit System Variable			✕
Variable name:	JAVA_HOME		
Variable value:	C:\java\jdk-17.0.1\bin		
Browse Directory...	Browse File...	OK	Cancel

Also add new value for path variable

C:\Program Files\Java\jdk-17.0.1\bin

E:\hadoop\hadoop-3.2.3\sbin
C:\java\jdk-17.0.1\bin

OK   Cancel

Step 2

Download and Set up HADOOP

Go to https://hadoop.apache.org/releases.html

Select your version and click on Binary to download it.

After downloading, Install it in your desired directory,

After installing go to Hadoop directory and create a new folder named **data.**

Open data folder and create 2 new folders named **namenode & datanode.**
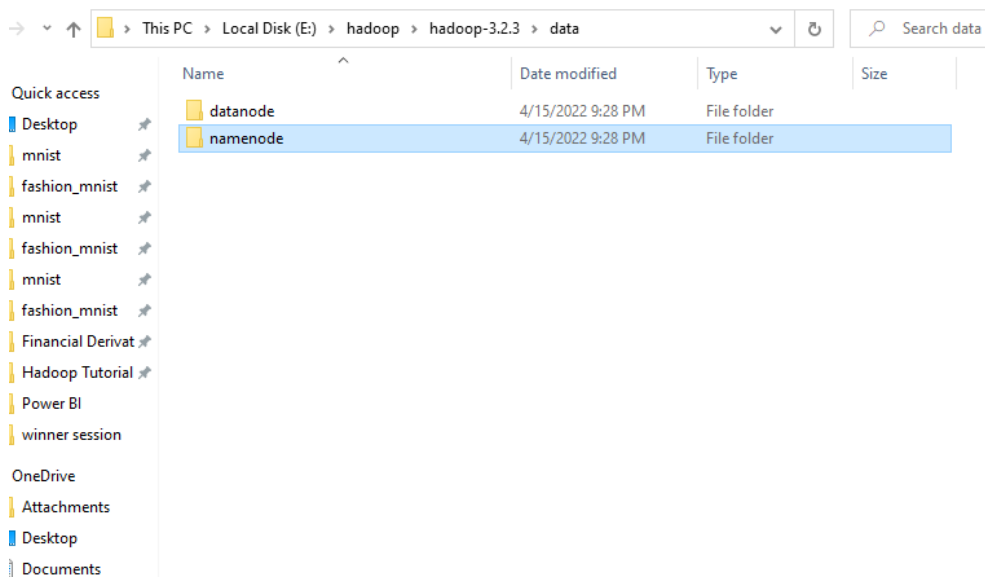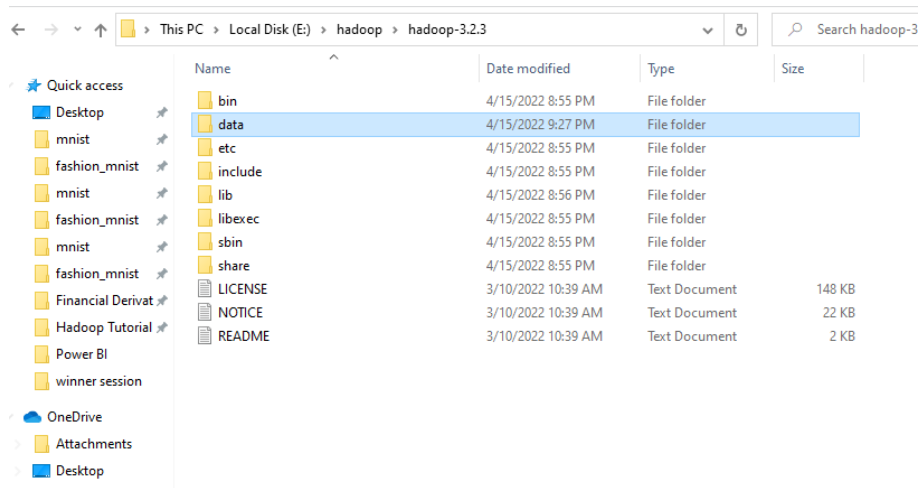
Go to **hadoop\hadoop-3.2.3\etc\hadoop** and open following files in VSCODE Editor:

1. **Core-site.xml**
2. **Mapred-site.xml**
3. **Yarn-site.xml**
4. **Hdfs-site.xml**
5. **Hadoop-env.cmd**
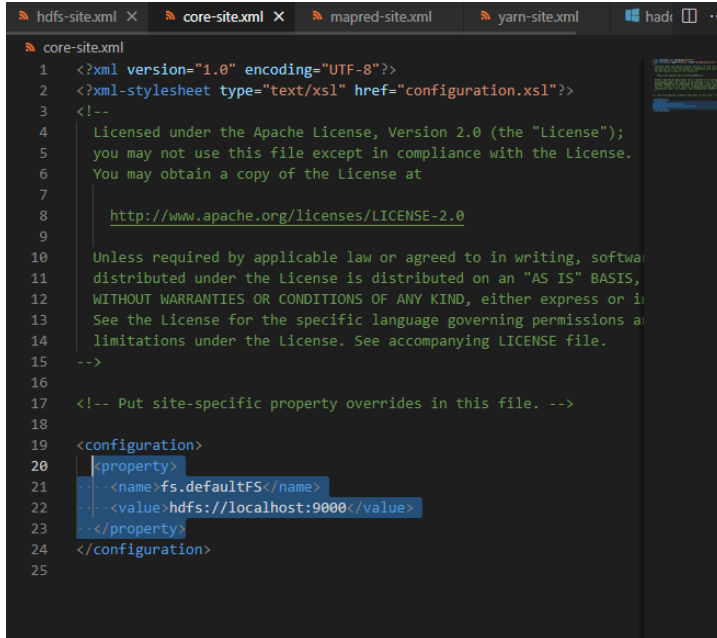
In core-site.xml file commit following changes under <configuration>

<property>

   <name>fs.defaultFS</name>

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

</property>

In Mapred-site.xml commit following changes under <configuration>

<property>

  <name>mapreduce.framework.name</name>

  <value>yarn</value>

</property>

In yarn-site.xml commit following changes under <configuration>

<property>

   <name>yarn.nodemanager.aux-services</name>
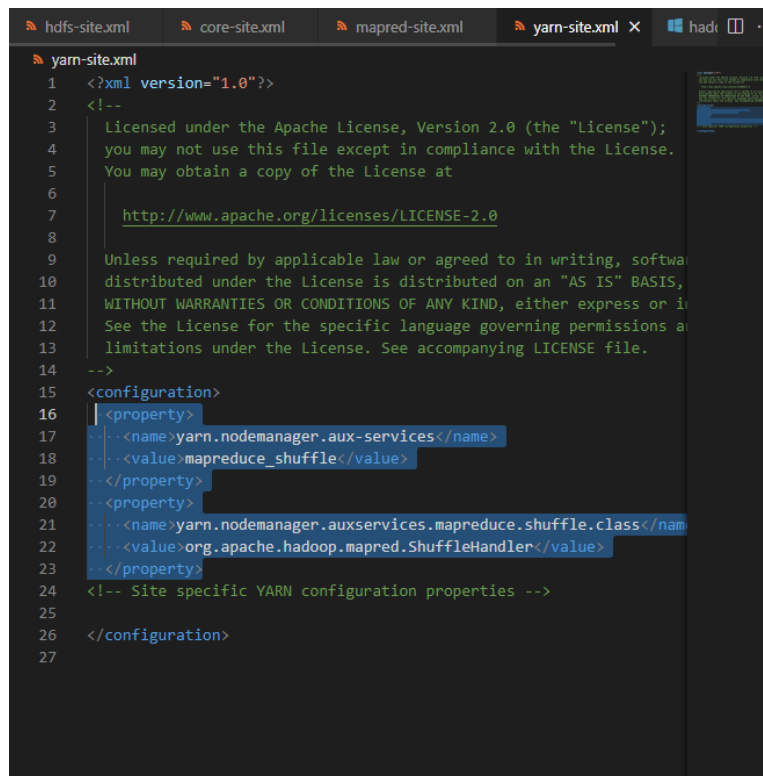
   <value>mapreduce_shuffle</value>

 </property>

 <property>

  <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>

  <value>org.apache.hadoop.mapred.ShuffleHandler</value>

 </property>

```xml
<?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, softwa
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or i
  See the License for the specific language governing permissions a
  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</nam
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
<!-- Site specific YARN configuration properties -->


</configuration>
```

In hdfs-site.xml commit following changes under <configuration>

<property>

 <name>dfs.replication</name>

 <value>1</value>

</property>

<property>

 <name>dfs.namenode.name.dir</name>

 <value>E:\hadoop\hadoop-3.2.3\data\namenode</value>

</property>

<property>

 <name>dfs.datanode.data.dir</name>

 <value>E:\hadoop\hadoop-3.2.3\data\datanode</value>

</property>

In Hadoop-env.cmd set JAVA_HOME=C:\java\jdk-17.0.1

Now go again to System environment variables and add the following values to PATH variable

E:\hadoop\hadoop-3.2.3\bin

E:\hadoop\hadoop-3.2.3\sbin

Now go to https://drive.google.com/file/d/1AMqV4F5ybPF4ab4CeK8B3AsjdGtQCdvy/view to download a bin file of Hadoop which you will replace with the old Hadoop bin file. The reason is that the old bin contains some missing files by default. After that open cmd and enter **hdfs namenode -format** to check the successful implementation of path values.

Now to Run Hadoop, open **hadoop\hadoop-3.2.3\sbin** folder in cmd and type

start-all.cmd

Hadoop running will look like this