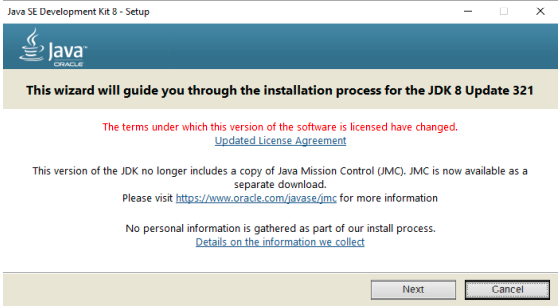
Rao Nauman

p19-0073

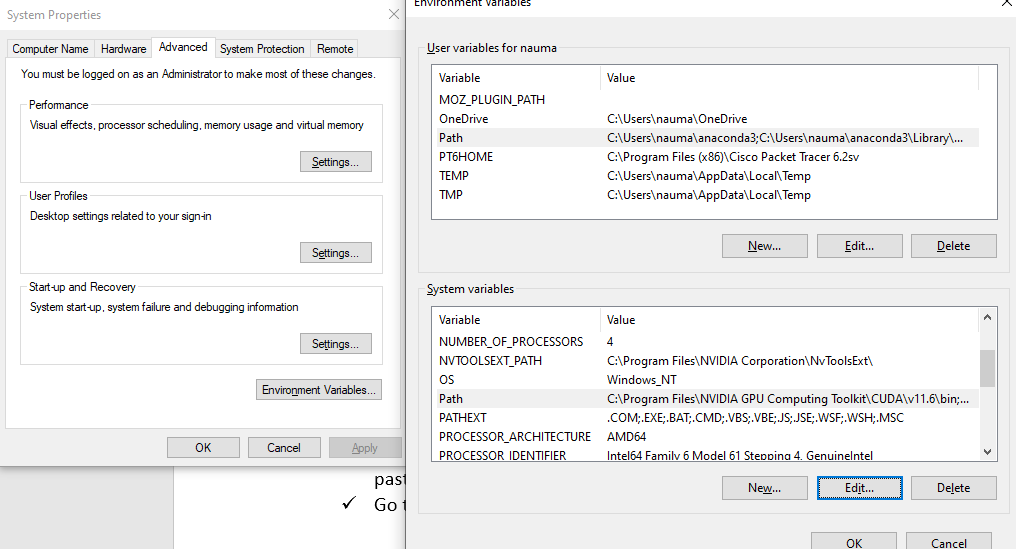
hadoop-deploying on windows

**Steps to follow:**

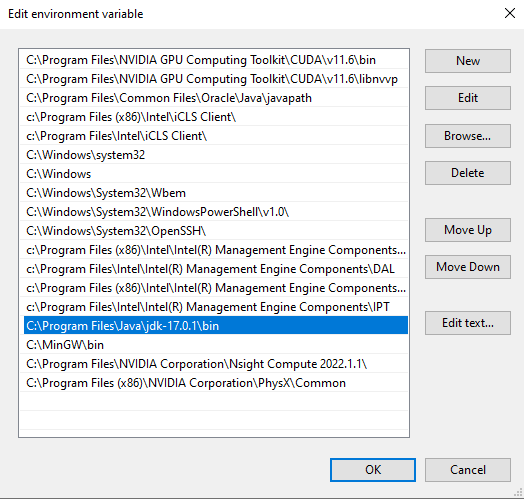
* Download and set up Java. Go to: <https://www.oracle.com/java/technologies/downloads/#java8>



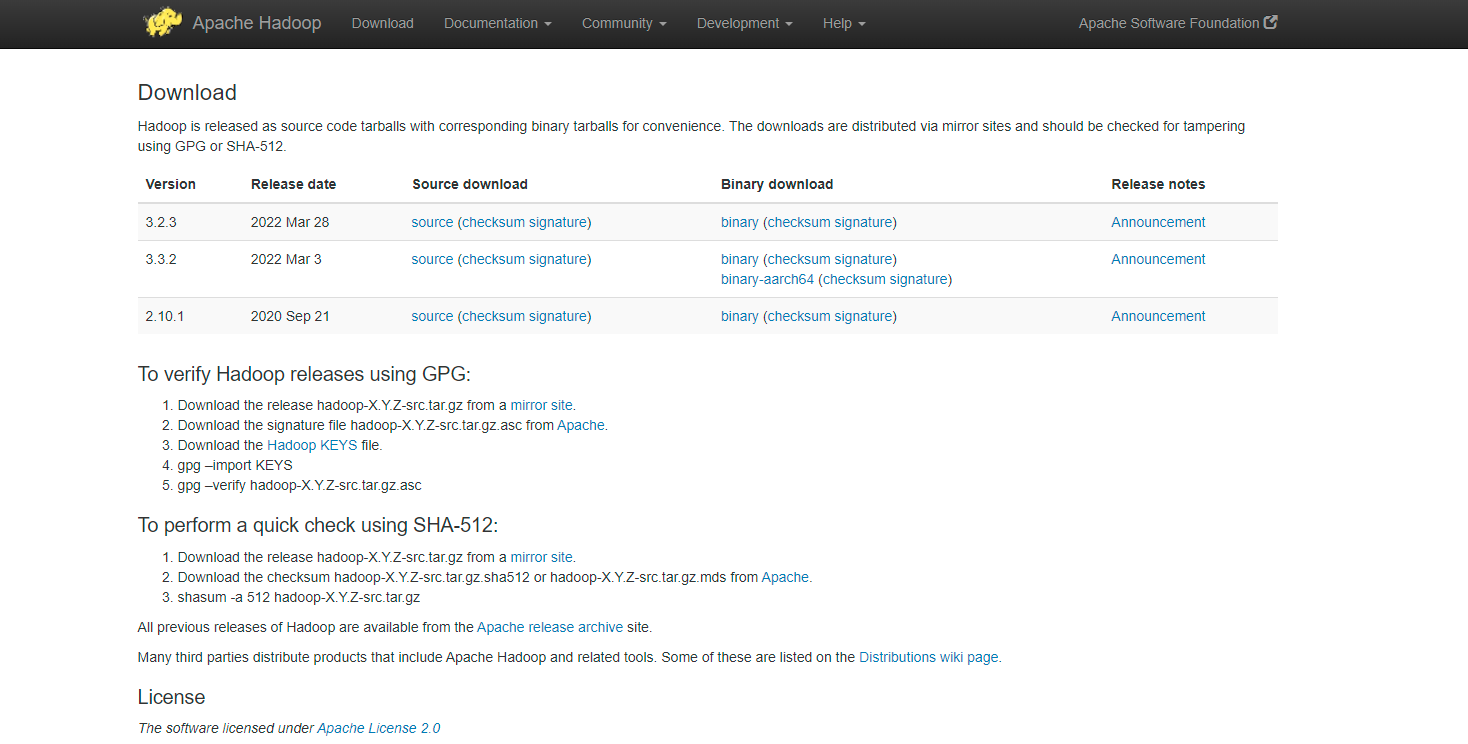
* Make a directory in C naming Java and install it there. Then for installation of it’s jdk files, keep the default directory i.e, **program files/java/jdk-17.0.1**
* After successful completion of installation, go to **program files/java** and cut that **jdk-17.0.1** and paste it in **C://java**.
* Go to system environment variables



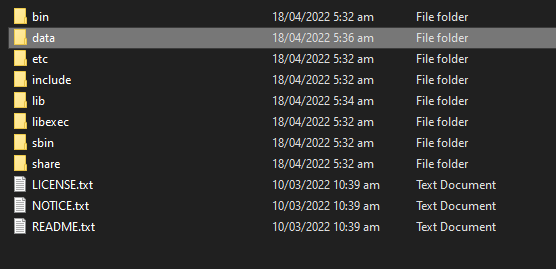
I have already set it up for something else:

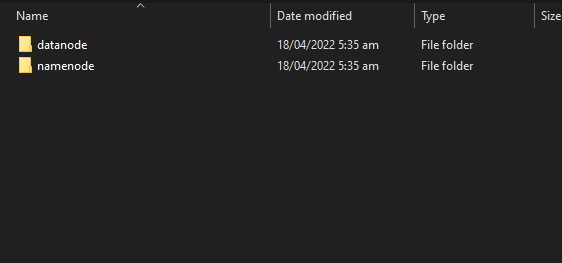


* **Download and Set up HADOOP**
* Go to <https://hadoop.apache.org/releases.html>



* 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**

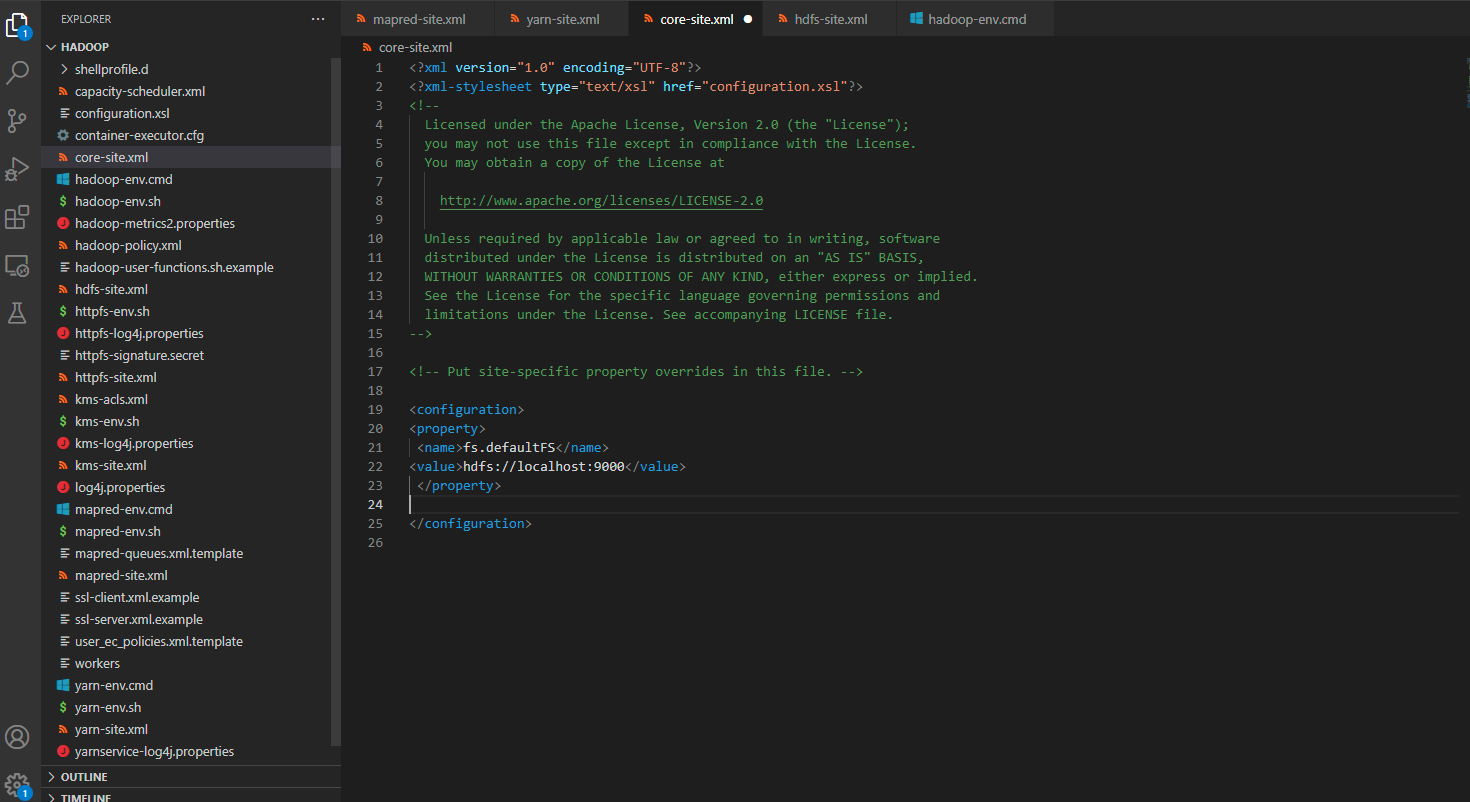
In core-site.xml file commit following changes under <configuration>

<property>

<name>fs.defaultFS</name>

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

</property>



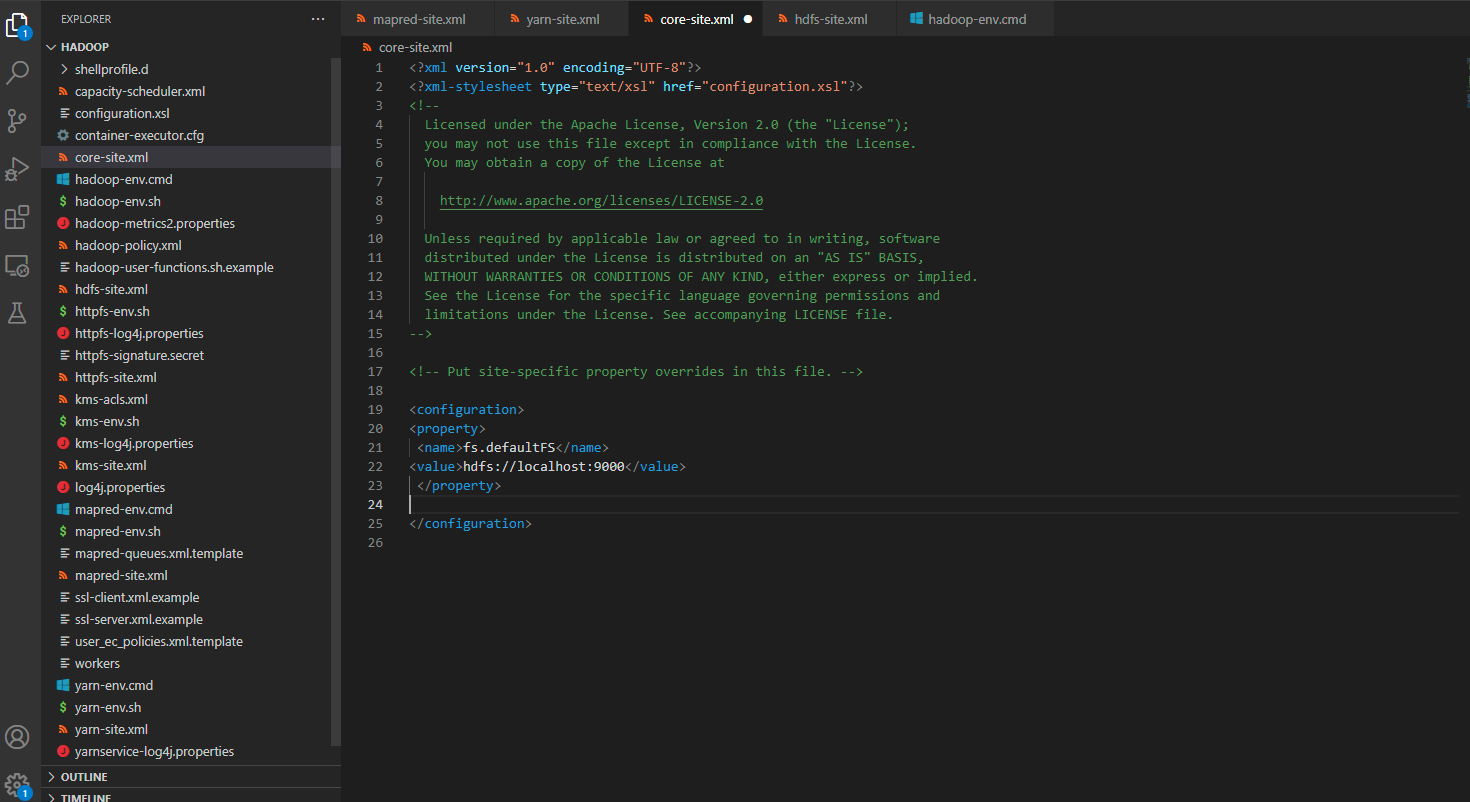
In Mapred-site.xml commit following changes under <configuration>

<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>

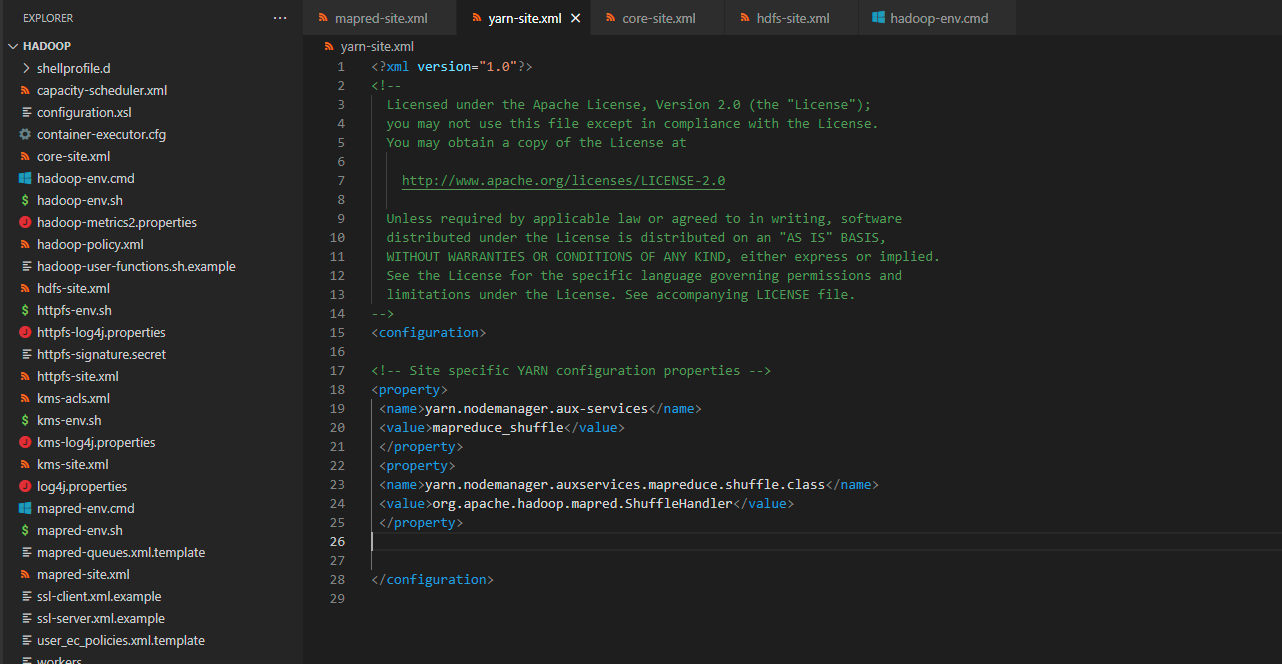
</property>

<property>

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

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

</property>



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>C:\Users\nauma\OneDrive\Documents\hadoop-3.2.3\data\namenode</value>

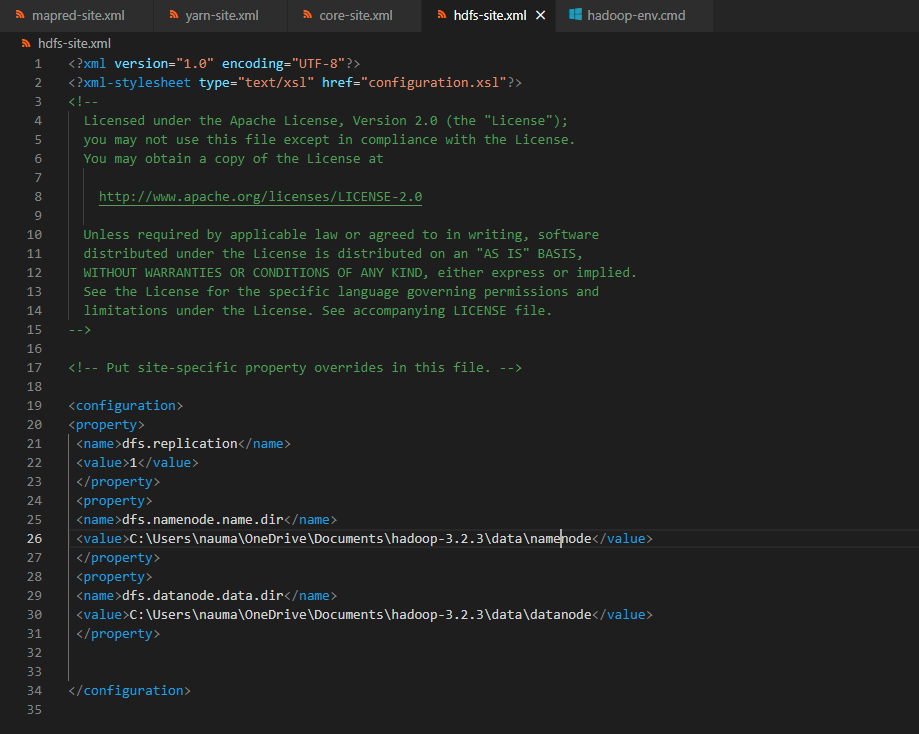
</property>

<property>

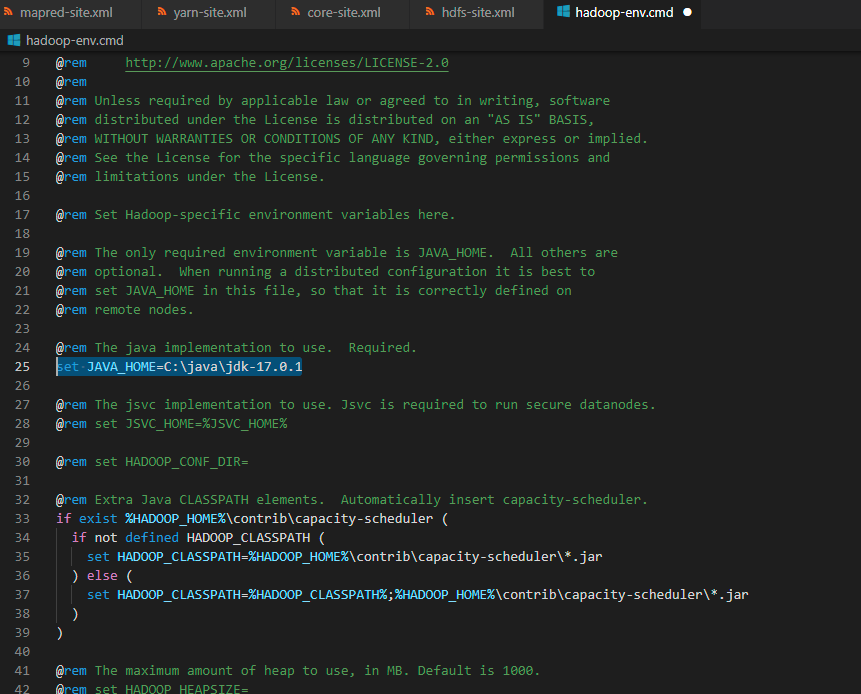
<name>dfs.datanode.data.dir</name>

<value>C:\Users\nauma\OneDrive\Documents\hadoop-3.2.3\data\datanode</value>

</property>

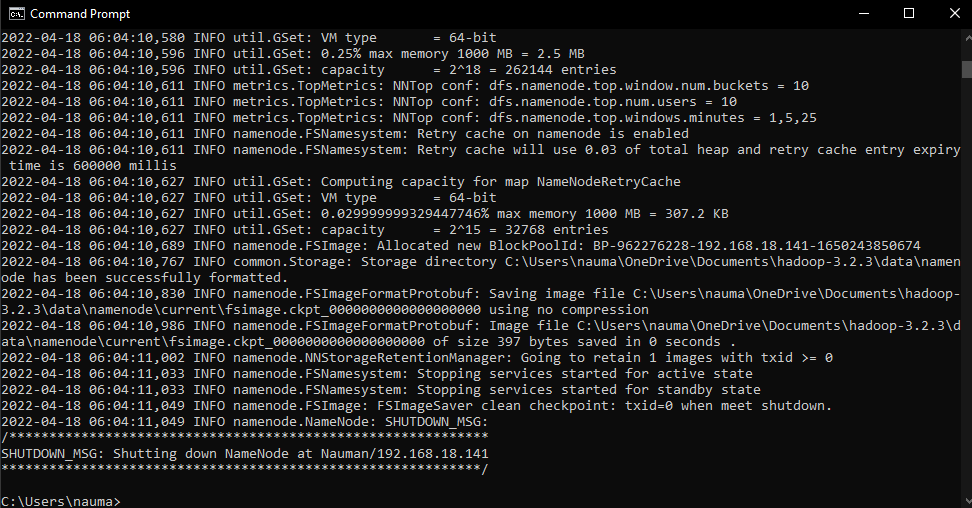


In Hadoop-env.cmd set JAVA\_HOME=C:\java\jdk-17.0.1



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

* **C:\Users\nauma\OneDrive\Documents\hadoop-3.2.3\bin**
* **C:\Users\nauma\OneDrive\Documents\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**

