EXP NO:01 **DATE:**

EXPLORATION OF INSTALLTION OF HADOOP

AIM : To install and setup Hadoop environment in windows

STEPS :

STEP 1: To install Hadoop the primary task is to setup and install java environment

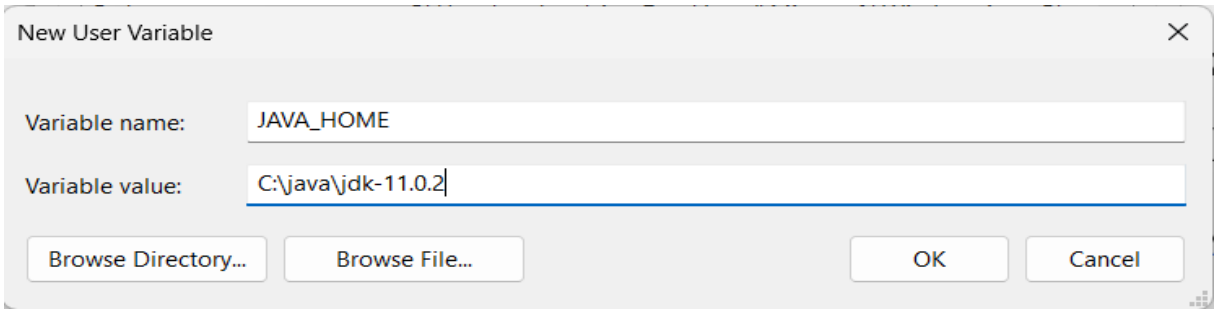
STEP 2: The java version that needed to be installed depends on the Hadoop's version. Here we are installing the latest version of Hadoop which is 3.3.6 which supports java version varying from 8-11(runtime only).

STEP 3: use the following link to install java
https://www.oracle.com/java/technologies/downloads/#java8-windows

Linux macOS Solaris Windows

Product/file description	File size	Download
x86 Installer	136.83 MB	jdk-8u381-windows-i586.exe
x64 Installer	145.55 MB	jdk-8u381-windows-x64.exe

STEP 4: After installing java setup the java environment in environmental variables directing the bin folder inside the java folder (**C:\java\jdk-11.0.2\bin**) copy the path till bin folder and paste it in the environmental variable define the new path and add the bin folder location as **JAVA_HOME="C:\java\jdk11.0.2\bin"** and apply the changes



STEP 5: Now after setting up the java environment check the setup has been successfully set by using **java -version** command in your command prompt and it should display the version of java you have installed.

```
C:\Windows\System32>java -version
java version "11.0.20" 2023-07-18 LTS
Java(TM) SE Runtime Environment 18.9 (build 11.0.20+9-LTS-256)
Java HotSpot(TM) 64-Bit Server VM 18.9 (build 11.0.20+9-LTS-256, mixed mode)
```

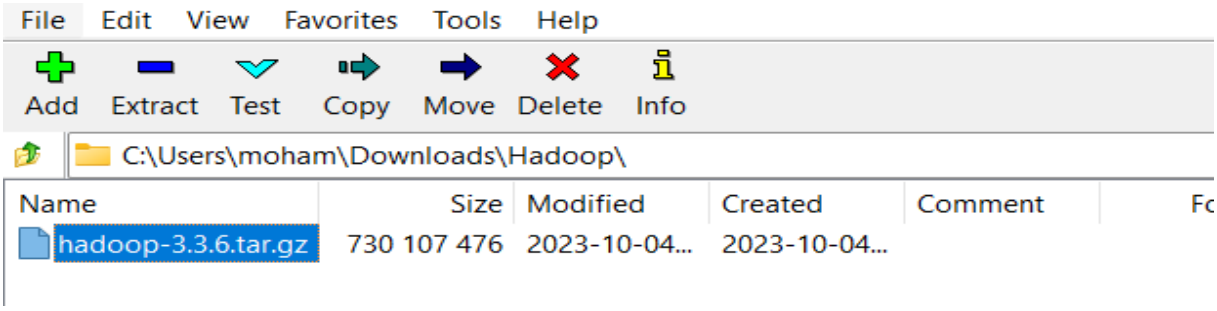
STEP 7: Hadoop is Unix distribution-based file with tar.gz extension we have to extract the file using the 7-zip manager which supports multiple formats follow this link to install 7-zip https://7-zip.org/

STEP 8: Now install the Notepad++ text editor which is further used to modify or edit the configuration file within Hadoop as per our requirement

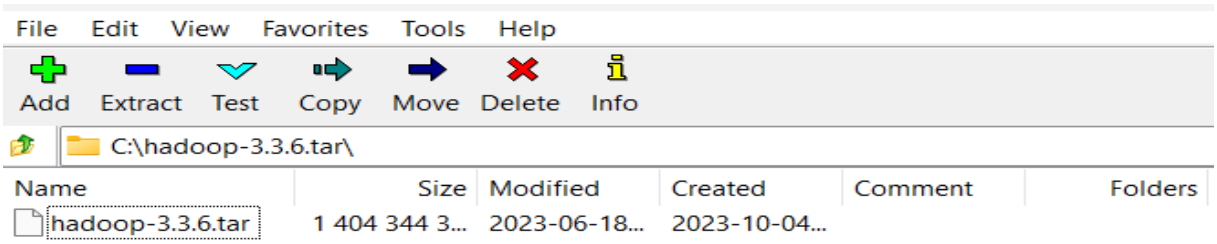
STEP 9: After installing and setting up all the required application install Hadoop from the official Apache Hadoop website https://hadoop.apache.org/releases.html download the binary download which can run directly without any need for compilation.

Version	Release date	Source download	Binary download	Release notes
3.3.6	2023 Jun 23	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
3.2.4	2022 Jul 22	source (checksum signature)	binary (checksum signature)	Announcement
2.10.2	2022 May 31	source (checksum signature)	binary (checksum signature)	Announcement

STEP 10: Run 7-zip manager as administrator and navigate to the path where Hadoop is located for extract the compiled binary download of Hadoop



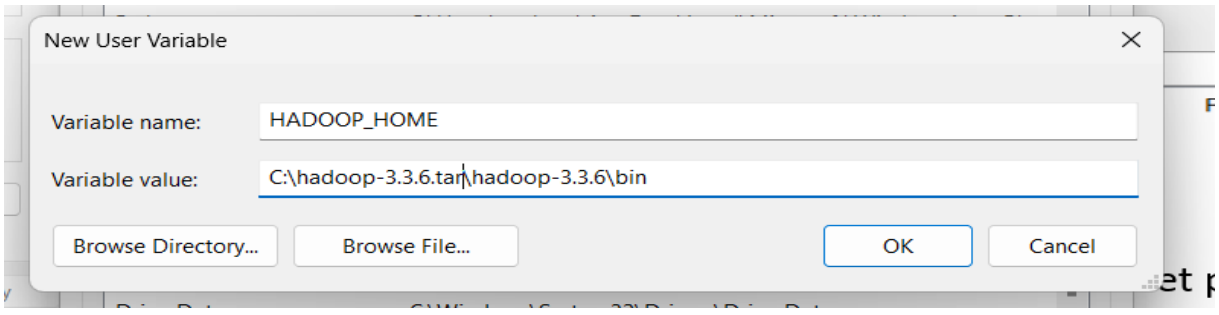
STEP 11: After doing the extraction process there is another compressed file with in the extracted file extract that as well.



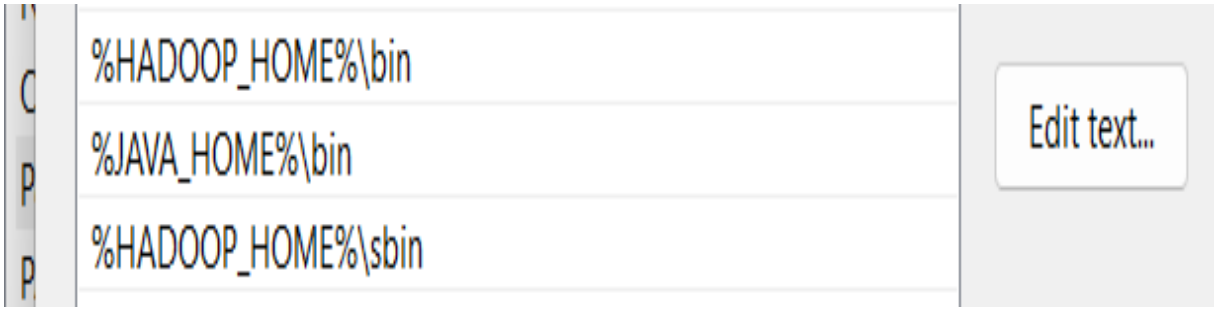
STEP 12: From the extracted folder replace the bin file with the reliable windows supported configured file here is the drive link to download the bin file

https://drive.google.com/file/d/1kVhX9snOZ3oLUxDjh3AVI8fcRnEWAAE4/view

STEP 13: Setup the Hadoop environment in environment variable and set path location as **HADOOP_HOME= "C:\hadoop-3.3.6.tar\hadoop-3.3.6\bin"**



STEP 14: Add the Hadoop bin and sbin path location by editing the path. And add the bin, sbin location there



STEP 15: Now open etc folder inside the Hadoop folder and locate the file Hadoop-env.cmd and set the java home location

```
@rem remote nodes.

@rem The java implementation to use.  Required.
set JAVA_HOME=%JAVA_HOME%
set JAVA_HOME=C:\java\java8

@rem The jsvc implementation to use. Jsvc is required to run secure datanodes.
@rem set JSVC_HOME=%JSVC_HOME%

@rem set HADOOP_CONF_DIR=
```

STEP 16: Edit the following configuration XML files core-site.xml, hdfs-site.xml, mapred-site.xml, yarn-site.xml are used to configure the behaviour of your Hadoop Cluster and save them.

STEP 17: Starting from core-site.xml edit it using notepad++ and the following the program to configure.

PROGRAM (CORE-SITE.XML)

<configuration>

<property>

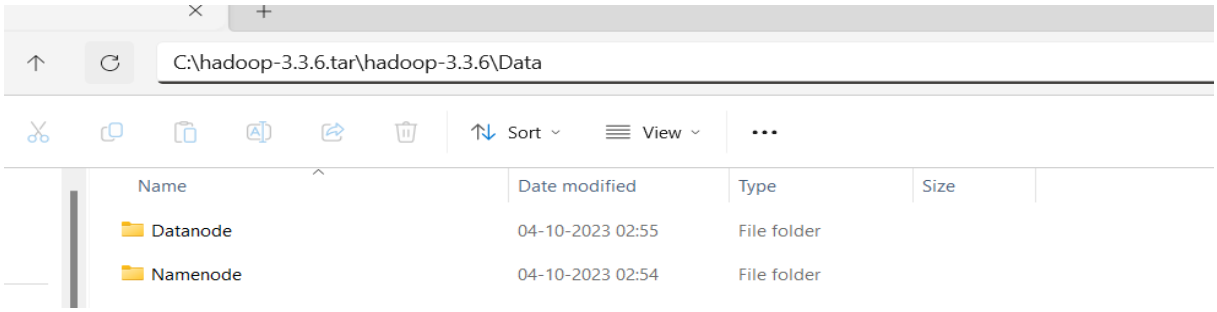
<name>fs.default.name</name>

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

</property>

</configuration>

STEP 18: To edit the hdfs-site.xml create Date folder and then within the Data folder create Namenode, Datenode which are used to manage and cluster flow date and log files.



STEP 19: Now open the hdfs-site.xml in notepad++ and add the following program

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

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

<value> C:/hadoop-3.3.6.tar/hadoop-3.3.6/Data/Datanode </value>

</property>

<property>

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

<value>> C:/hadoop-3.3.6.tar/hadoop-3.3.6/Data/Datanode </value>

</property>

</configuration>

STEP 20: Edit mapred-site.xml file using notepad++ add this following program

<configuration>

<property>

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

<value>yarn</value>

</property>

</configuration>

STEP 21: Edit the yarn-site.xml with following program and save it.

<configuration>

<property>

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

<value>mapreduce_shuffle</value>

</property>

<property>

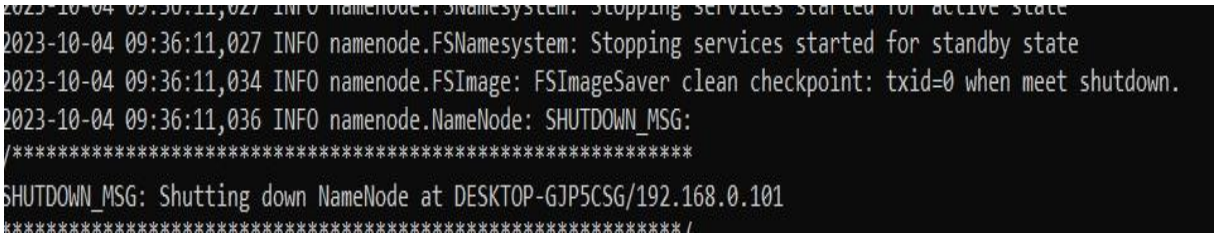
<name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>

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

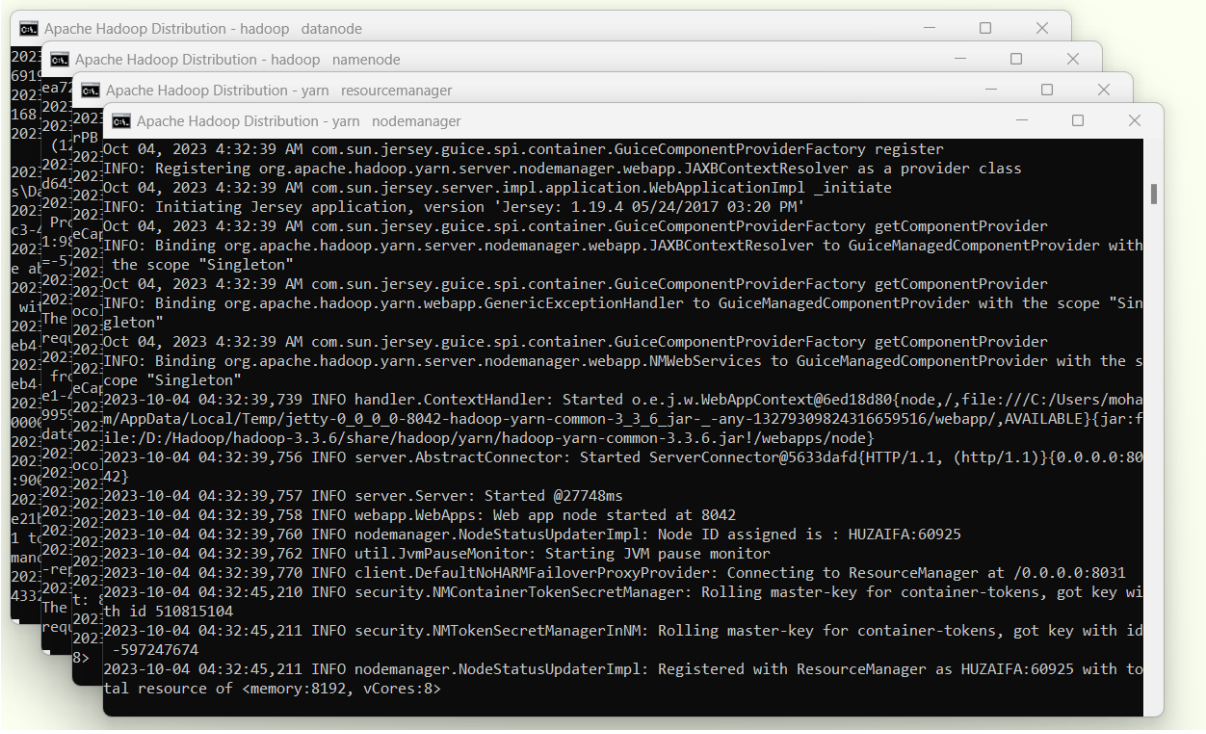
</property>

</configuration>

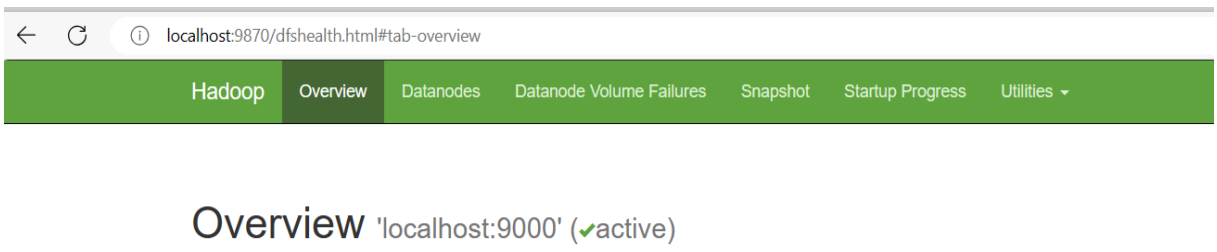
STEP 22: Now save them and open command prompt as administrator and run the following command **hdfs namenode -format** to format the contents of namenode



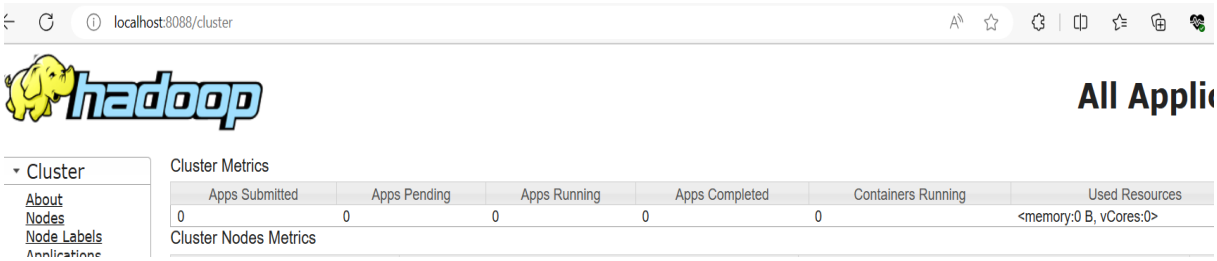
STEP 23: Open command prompt as administrator and run the following command **start-all** to check the whole configuration files are configured correctly and Hadoop environment has been set successfully.



STEP 24: Now we can access the Namenode and Datanode as web user interface (web-UI) by using the following localhost address **localhost:9870**



STEP 25: To access the Datanode use the following localhost address **localhost:8088**



RESULT: Thus, the program to install and setup Hadoop has been installed and verified successfully.