Date : 15/10/22 Name : Siva Prakash G

EX:9 Roll.No:19UITE006

Reg.No:920419205041

**Installation of Single Node Hadoop Cluster**

**PROCEDURAL STEPS**

**Step 1:** Download the following Packages

1. hadoop-3.2.2.tar.gz

**Link:** <https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.2.2/hadoop-3.2.2.tar.gz>

1. 7zip to unzip the tar.gz file

**Link:** <https://www.7-zip.org/download.html>

1. Java 8 (JDK-8U333 & JRE-8U333)

**Link:** <https://www.oracle.com/java/technologies/javase/javase8u211-later-archive-downloads.html>

1. Hadoop dll Files

**Link:** <https://github.com/cdarlint/winutils/archive/refs/heads/master.zip>

1. hadoop-hdfs-3.2.2.jar

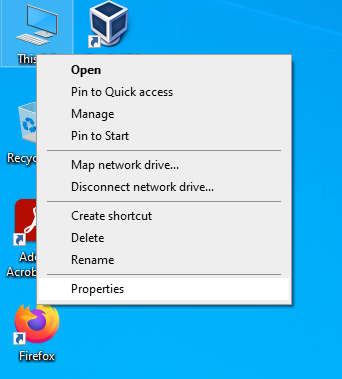
**Link:** <https://jar-download.com/artifacts/org.apache.hadoop/hadoop-hdfs/3.2.2/source-code>

**Step 2:** Install / Extract the following Packages:

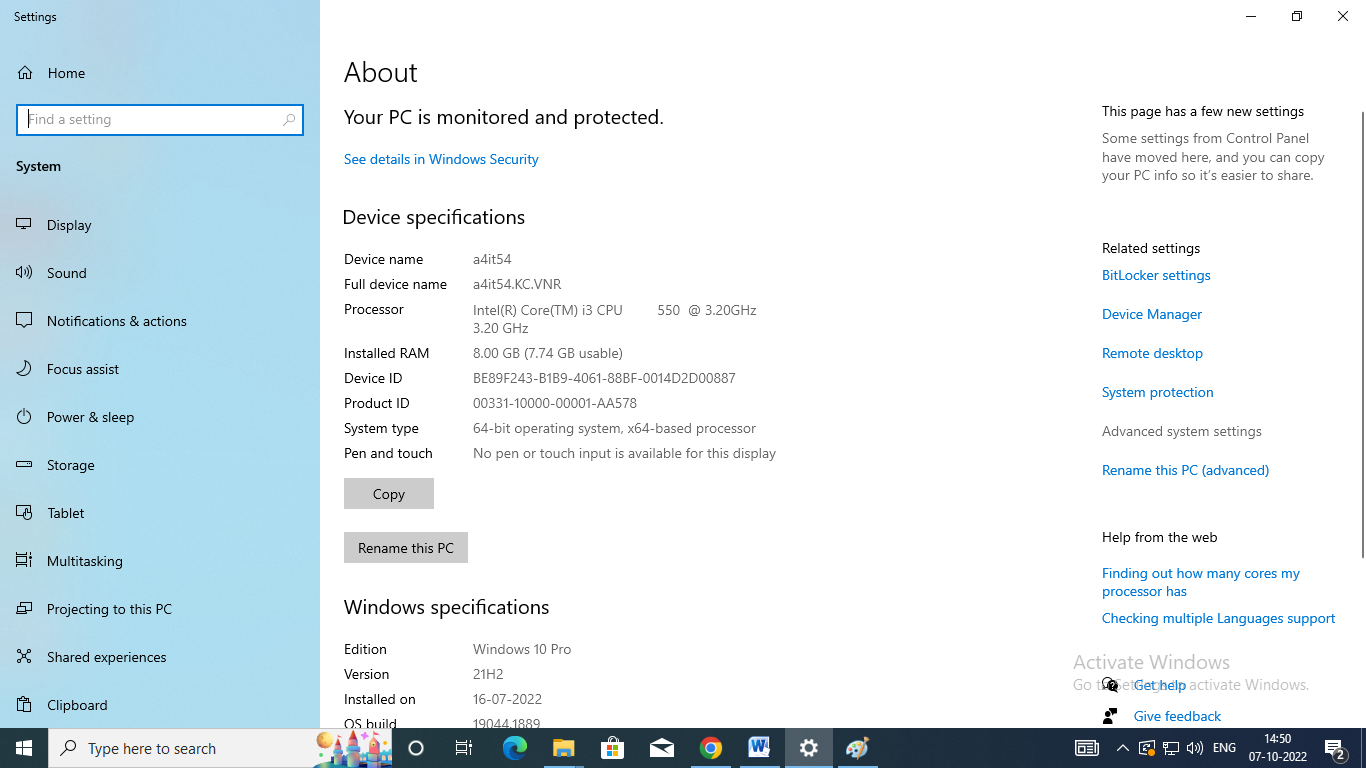
1. Install Java 8 in the location **c:\JAVA**. Also create two folders namely **“jdk1.8.0\_333”** and **“jre1.8.0\_333”**.
2. Create a folder **“hadoop-env”** in **D:\**
3. Paste the downloaded “hadoop-3.2.2.tar.gz” into “d:\ hadoop-env”
4. Unzip the “hadoop-3.2.2.tar.gz” using 7zip (do the unzip function two times)
5. Copy the downloaded Hadoop dll files to the location “D:\hadoop-env\hadoop-3.2.2\bin”
6. Copy the downloaded hadoop-hdfs-3.2.2.jar to “D:\hadoop-env\hadoop-3.2.2\share\hadoop\hdfs”
7. Edit the JAVA\_HOME = C:\JAVA\jdk1.8.0\_333 in the path “**D:\hadoop-env\hadoop-3.2.2\etc\hadoop\hadoop-env.cmd**

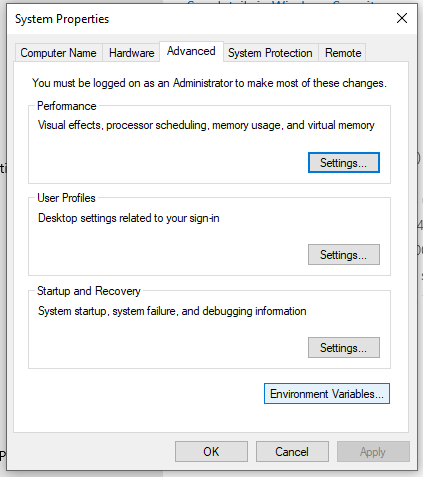
**Step 3:** Set the path for Java and Hadoop in system environment variables

To edit environment variables, go to Control Panel 🡪 System and Security 🡪 System (or) right-click “This PC” 🡪 Properties (My Computer icon) and click on the “Advanced system settings” link.

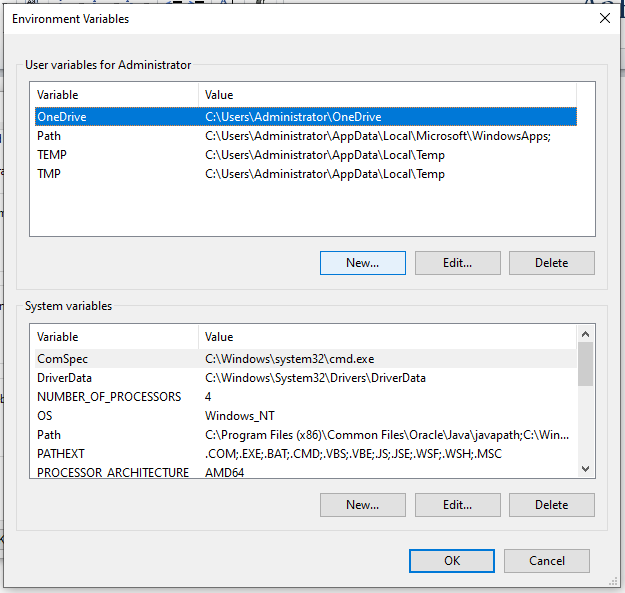


When the “Advanced system settings” dialog appears, go to the “Advanced” tab and click on the “Environment variables” button located on the bottom of the dialog.



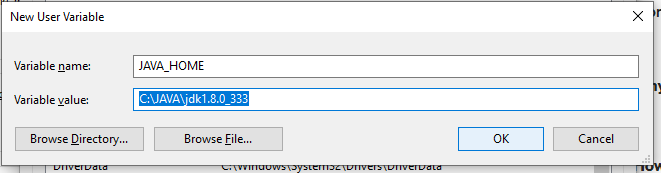


In the “Environment Variables” dialog, press the “New” button to add a new variable.

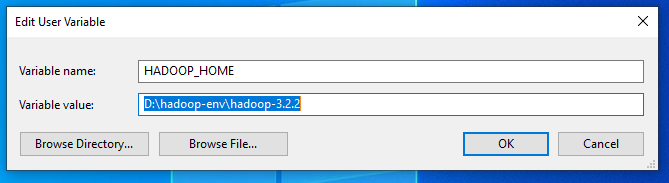


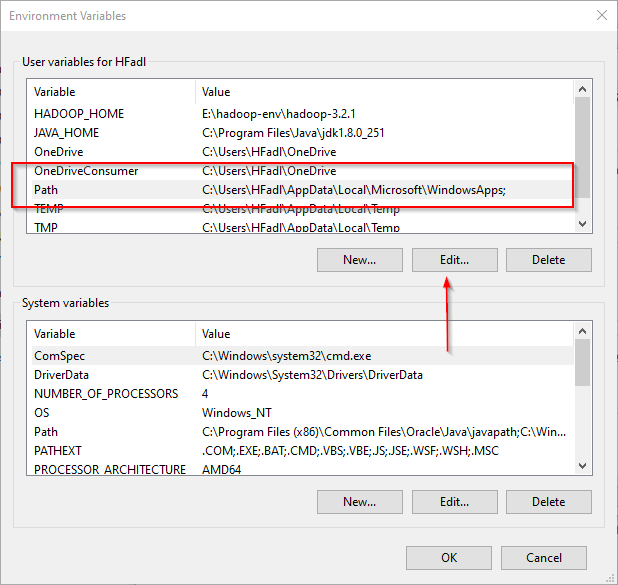
Now, Edit the PATH variable to add the Java and Hadoop binaries paths as shown in the following screenshots.

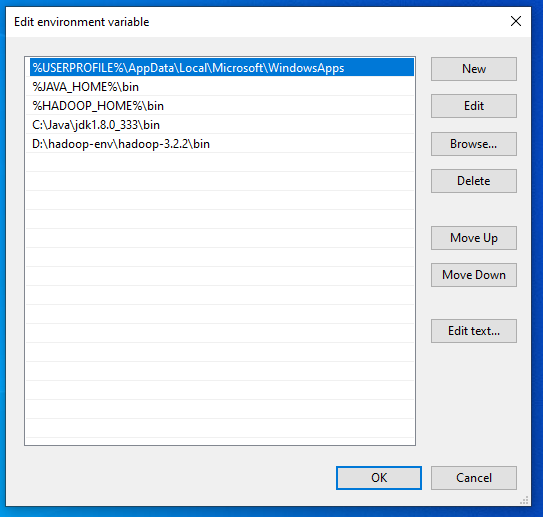
**JAVA\_HOME: JDK installation folder path**



**HADOOP\_HOME: Hadoop installation folder path**

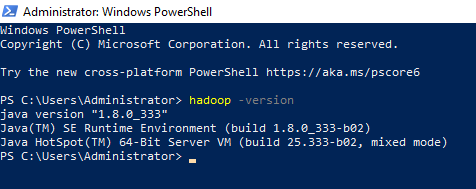






**Step 4:** Open Command Prompt as Administrator and run the following command:

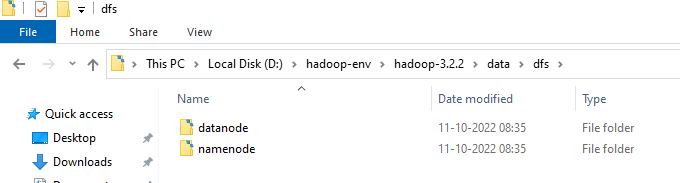
**hadoop –version**



**Step 4:** Create the following folders

D:\hadoop-env\hadoop-3.2.2\**data\dfs\namenode**

D:\hadoop-env\hadoop-3.2.2\**data\dfs\datanode**



**Step 5:** Configuring Hadoop cluster

There are **four files** to configure Hadoop cluster:

Location of the File: **“D:\hadoop-env\hadoop-3.2.2\etc\hadoop”**

**File Name: “hdfs-site.xml”**

Add the following properties within the <configuration></configuration> element:

*<property>*

*<name>dfs.replication</name>*

*<value>1</value>*

*</property>*

*<property>*

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

*<value>****file:///D:/hadoop-env/hadoop-3.2.2/data/dfs/namenode****</value>*

*</property>*

*<property>*

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

*<value>****file:///D:/hadoop-env/hadoop-3.2.2/data/dfs/datanode****</value>*

*</property>*

**File Name: “core-site.xml”**

Add the following properties within the <configuration></configuration> element:

*<property>*

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

*<value>****hdfs://localhost:9820****</value>*

*</property>*

**File Name: “mapred-site.xml”**

Add the following properties within the <configuration></configuration> element:

*<property>*

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

*<value>yarn</value>*

*<description>MapReduce framework name</description>*

*</property>*

**File Name: “yarn-site.xml”**

Add the following properties within the <configuration></configuration> element:

*<property>*

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

*<value>mapreduce\_shuffle</value>*

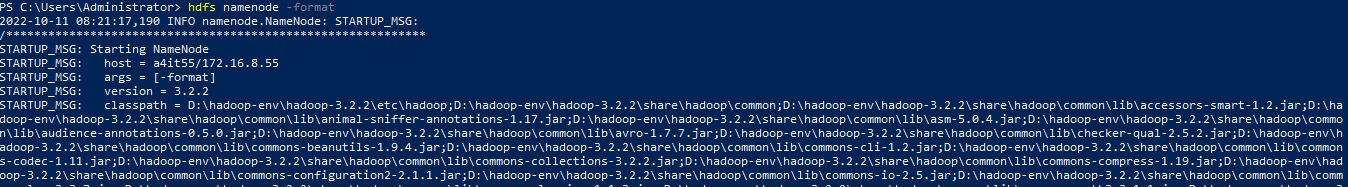
*<description>Yarn Node Manager Aux Service</description>*

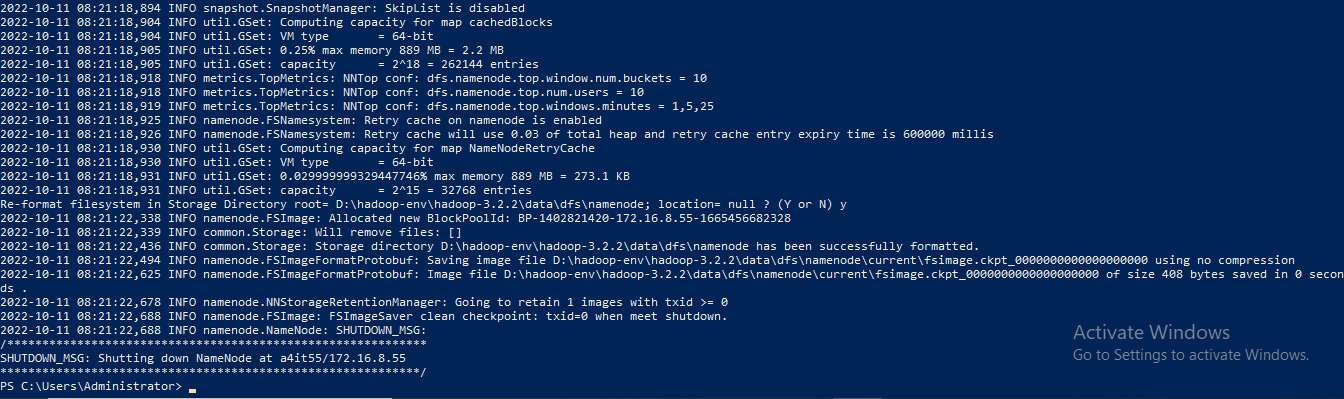
*</property>*

**Step 6:** Format the namenode

After finishing the configuration, format the name using the following command

**hdfs namenode –format**

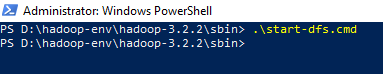




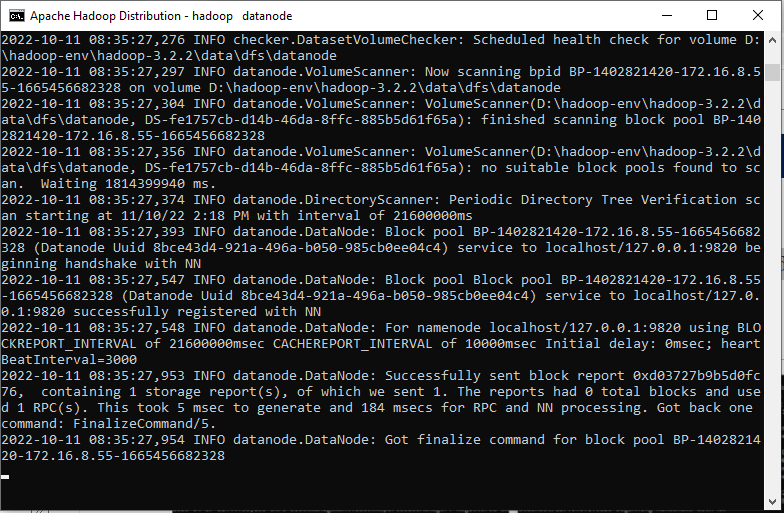
**Step 7:** Start the Hadoop nodes

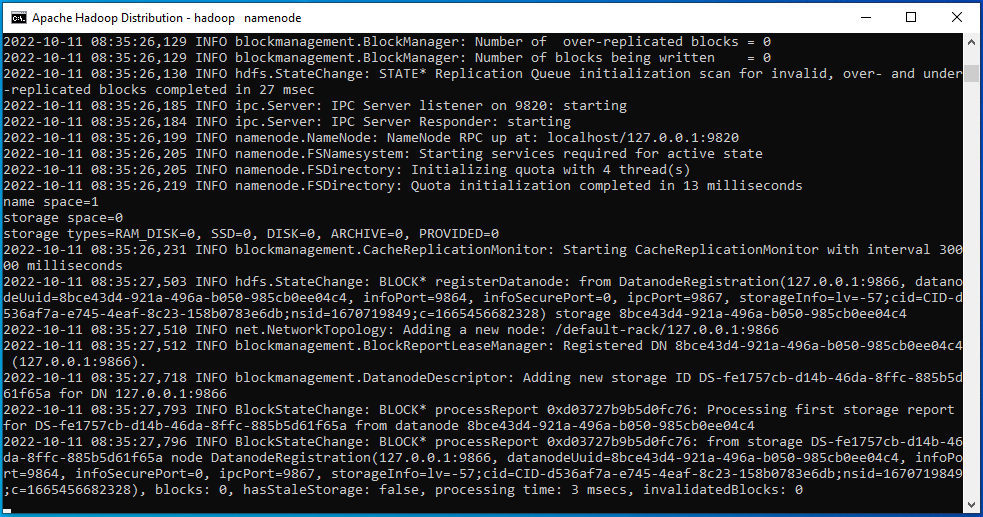
Now navigate to the location “D:\hadoop-env\hadoop-3.2.2\sbin” in powershell and then run the following command to start the Hadoop nodes:

**.\start-dfs.cmd**



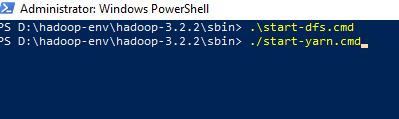
Two command prompt windows will open (one for the **namenode** and one for the **datanode**) as follows:



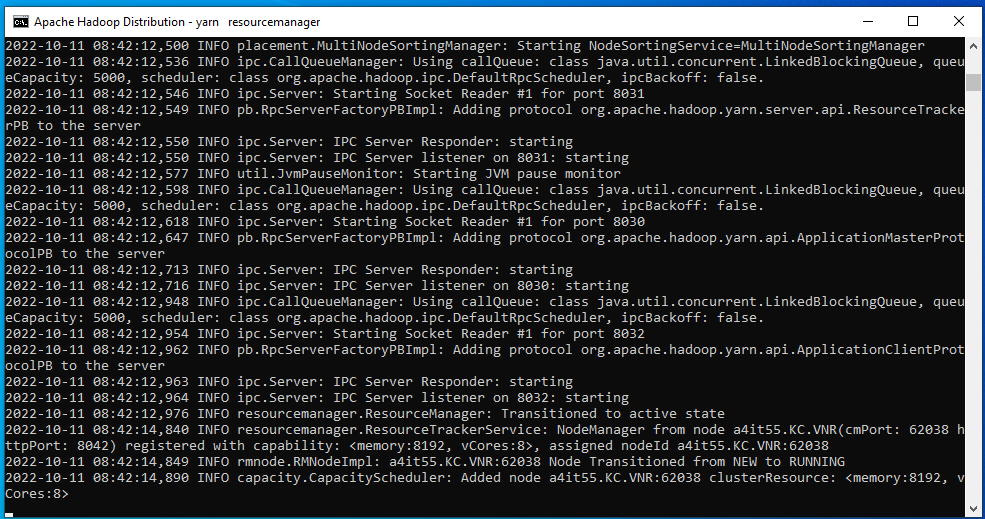


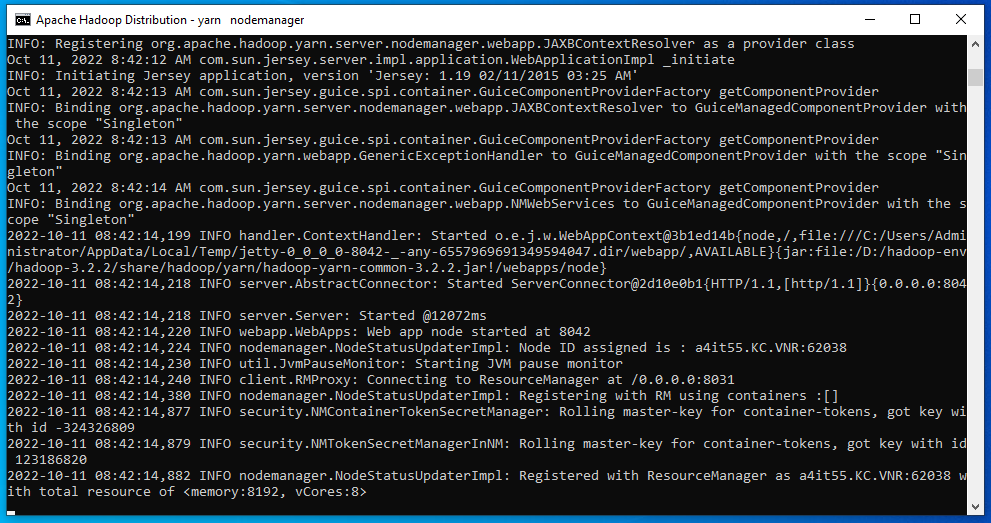
Next, Start the Hadoop Yarn service using the following command:

**./start-yarn.cmd**



Two command prompt windows will open (one for the **resource manager** and one for the **node manager**) as follows:





To make sure that all services started successfully, Run the following command:

**jps**

It should display the following services:

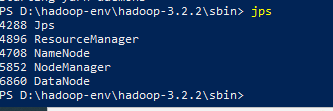
14560 DataNode

4960 ResourceManager

5936 NameNode

768 NodeManager

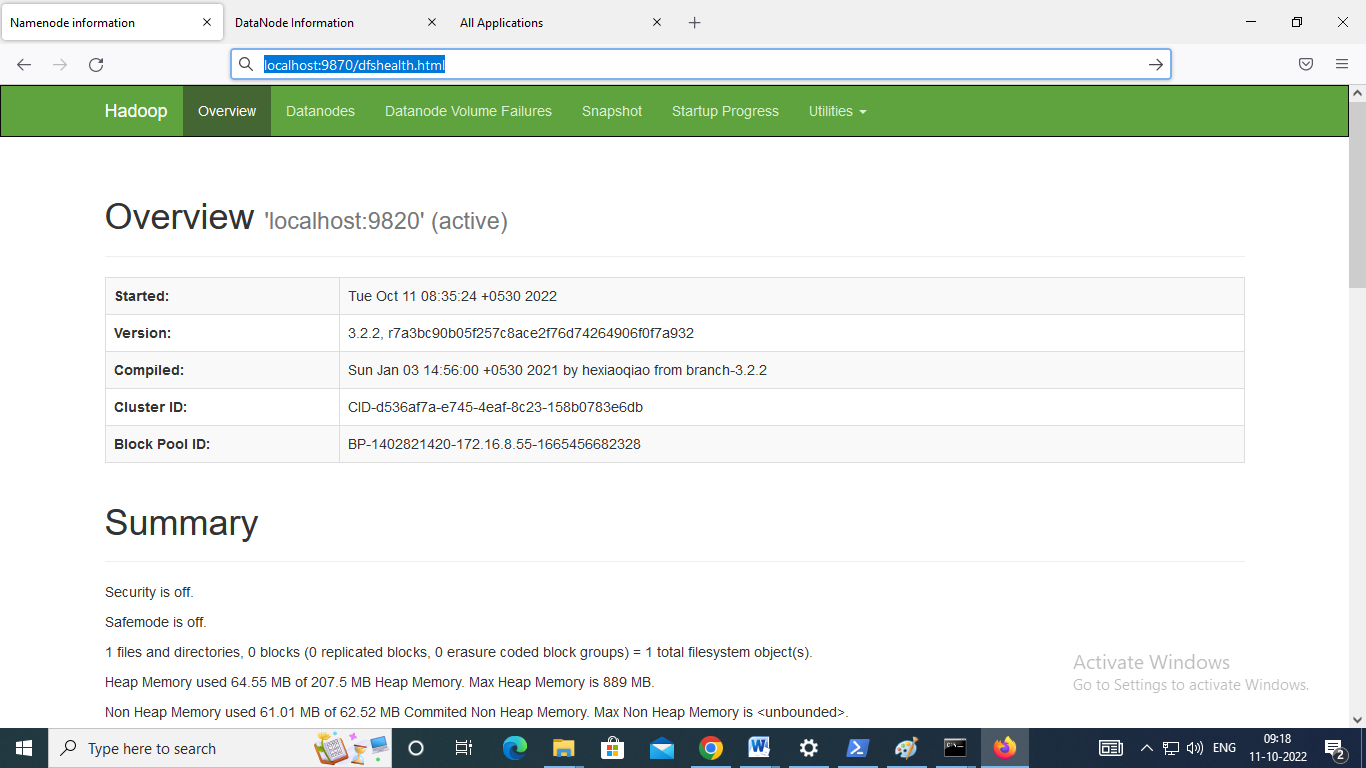
14636 Jps



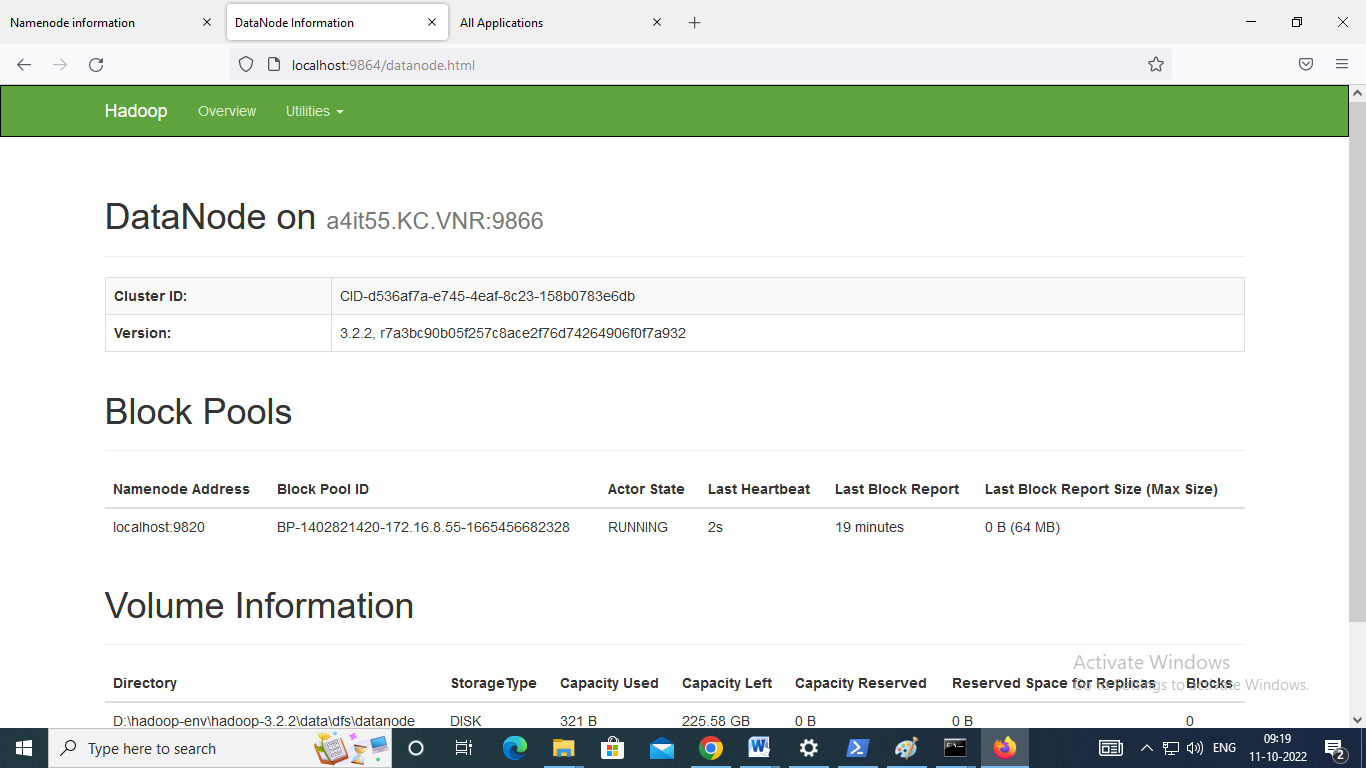
**Output**

Hadoop Web UI

<http://localhost:9870/dfshealth.html>



<http://localhost:9864/datanode.html>



<http://localhost:8088/cluster>

