EXNO-9 **DATE-15.10.22**

NAME-KARTHIKEYAN **REGNO-920419205018**

ROLLNO-19UITE012

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
2. 7zip to unzip the tar.gz file

   Link: https://www.7-zip.org/download.html
3. Java 8 (JDK-8U333 & JRE-8U333)

   Link: https://www.oracle.com/java/technologies/javase/javase8u211-later-archive-downloads.html
4. Hadoop dll Files

   Link: https://github.com/cdarlint/winutils/archive/refs/heads/master.zip
5. hadoop-hdfs-3.2.2.jar

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

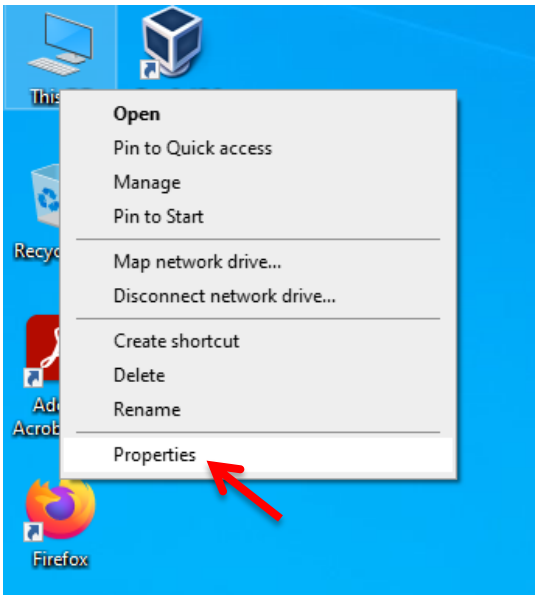
Step 2: Install / Extract the following Packages:

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"

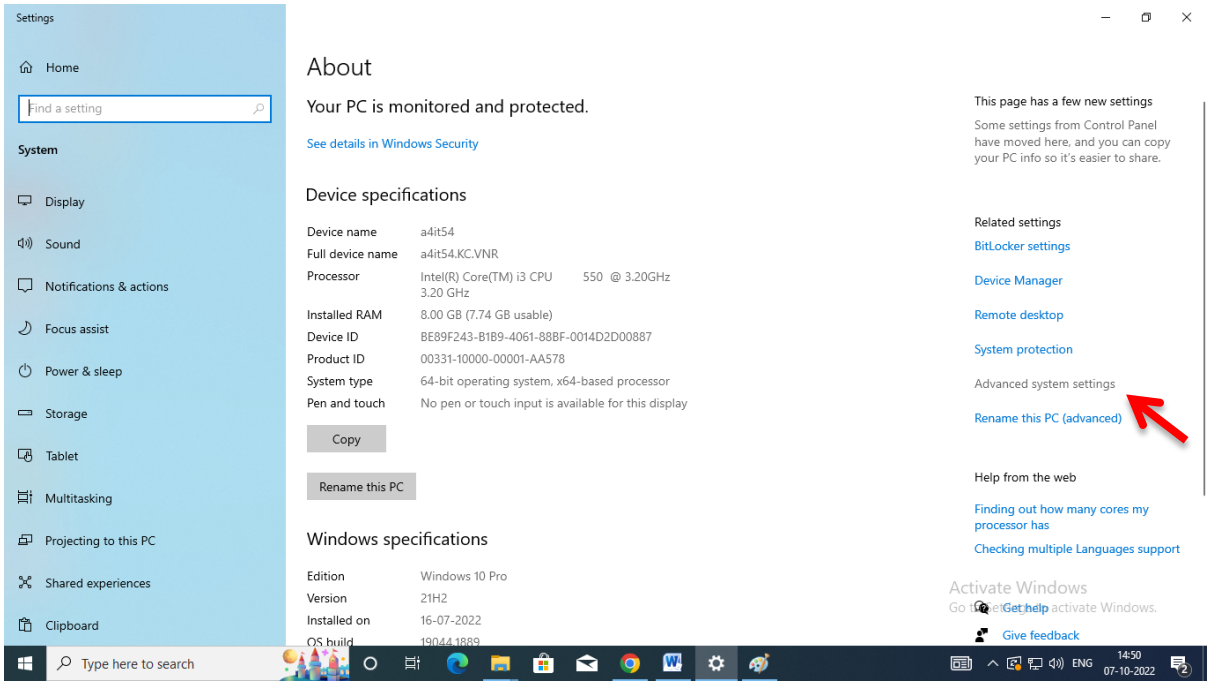
7. Edit the JAVA_HOME = C:\JAVA\jdk1.8.0_333 in the path "**D:\hadoop-env\hadoop-3.2.2\etc\hadoop\hadoop-env.cmd**

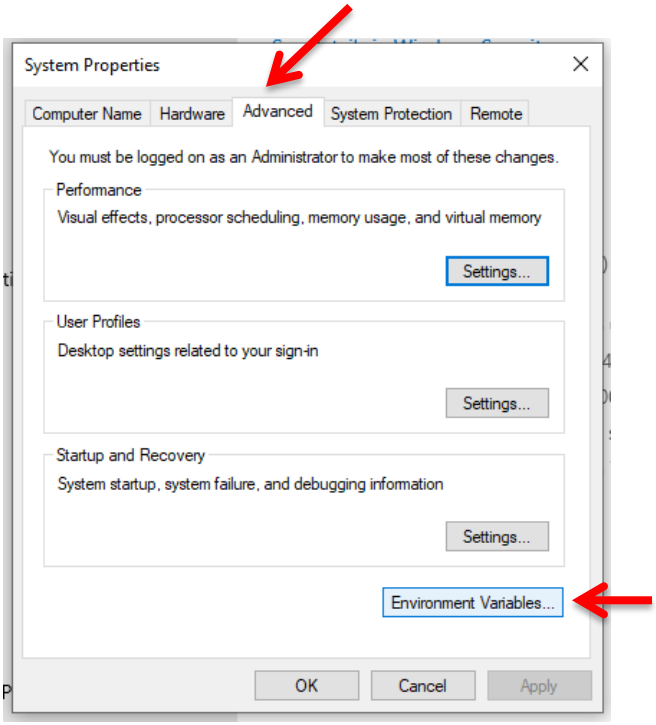
Step 3: Set the path for Java and Hadoop in system environment variables

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.

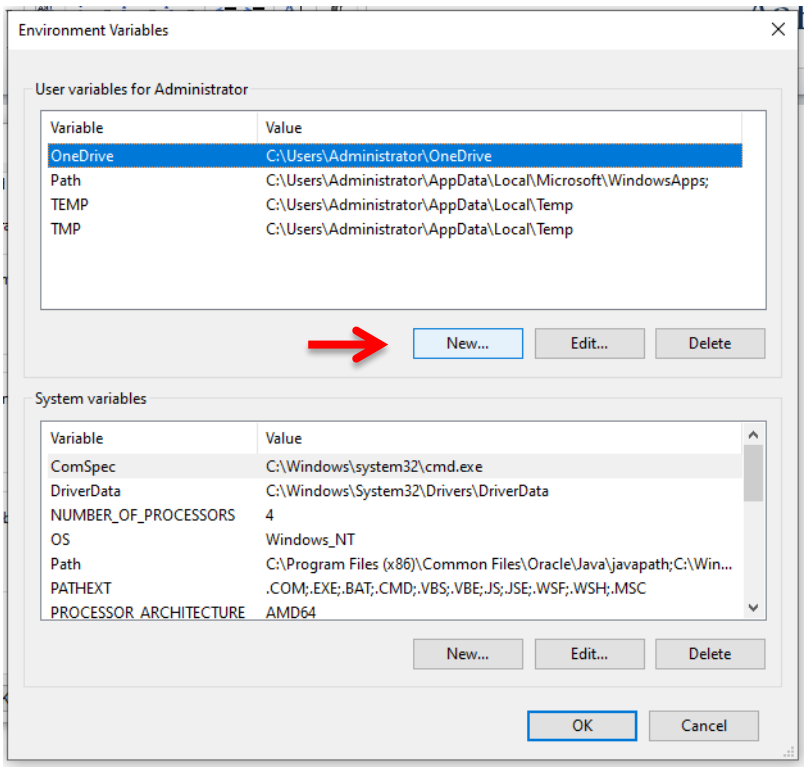


When the "Advanced system settings" dialog appears, go to the "Advanced" tab and click on the "Environment variables" button located on the bottom of the dialog.



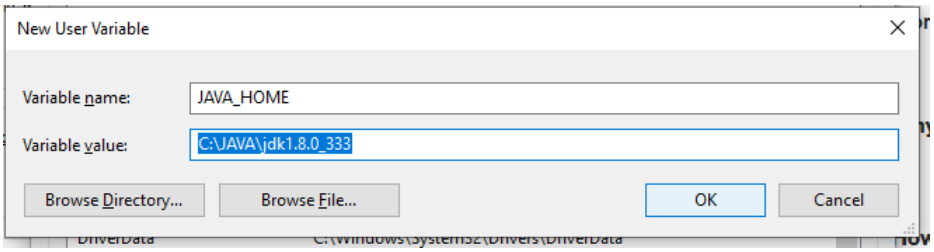


In the “Environment Variables” dialog, press the “New” button to add a new variable.

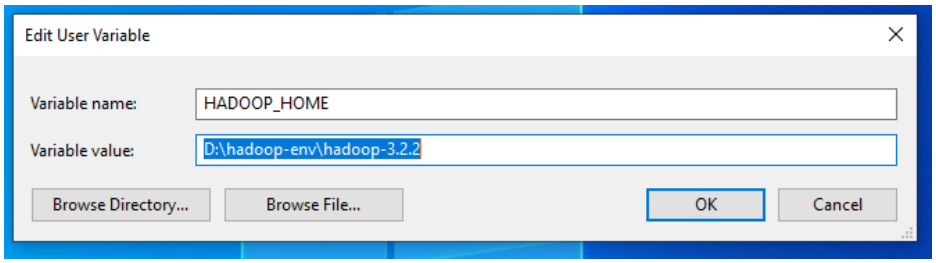


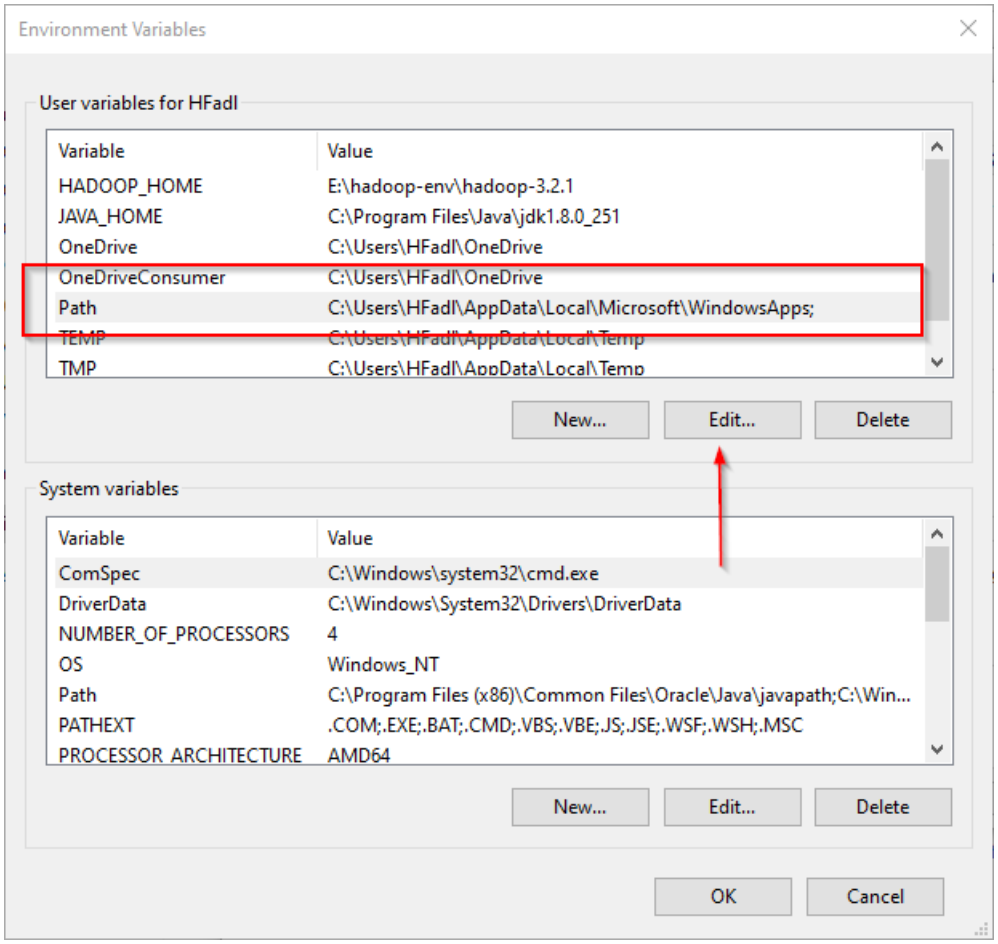
Now, Edit the PATH variable to add the Java and Hadoop binaries paths as shown in the following screenshots.

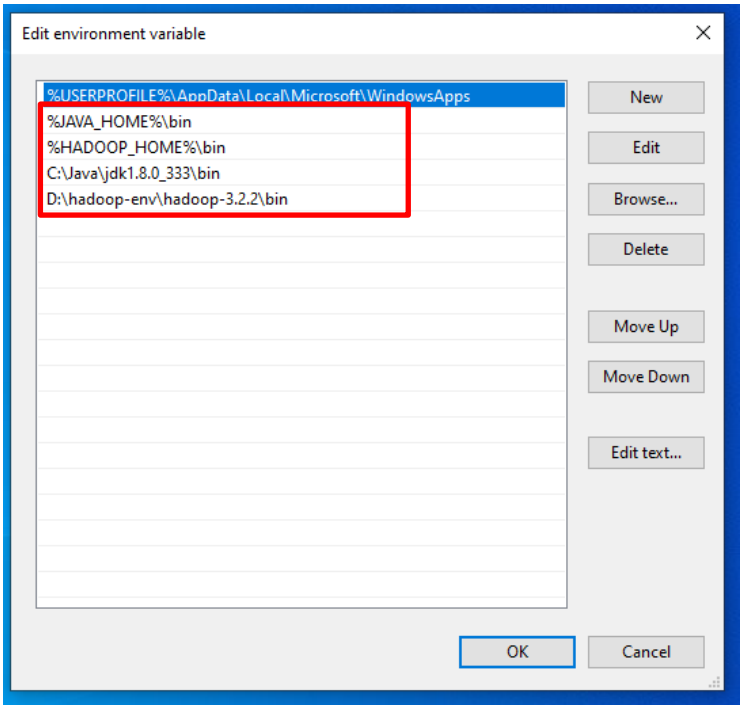
JAVA_HOME: JDK installation folder path



HADOOP_HOME: Hadoop installation folder path

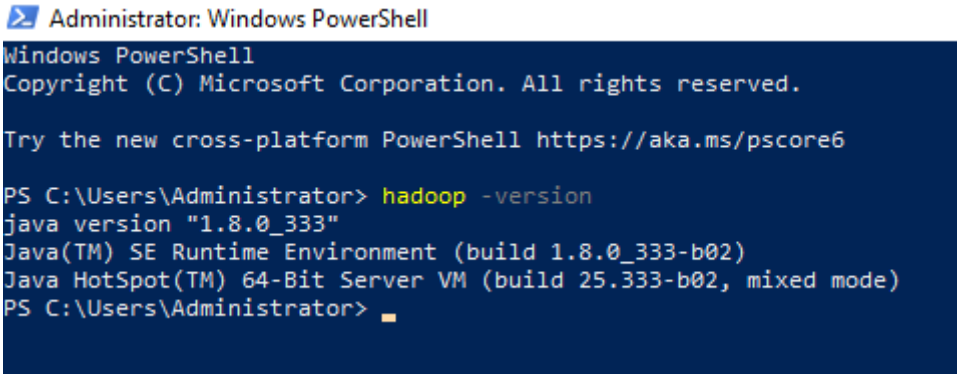






Step 4: Open Command Prompt as Administrator and run the following command:

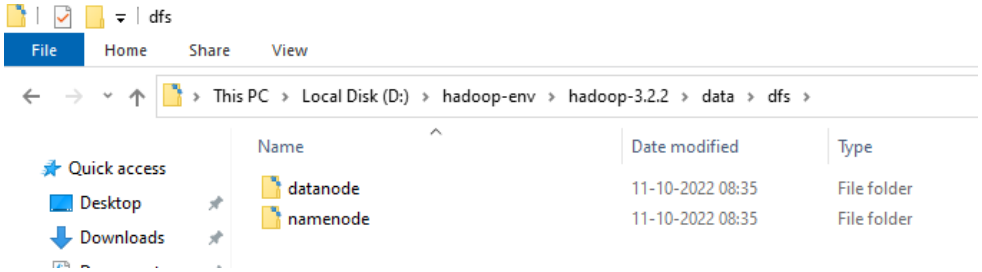
hadoop –version



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>

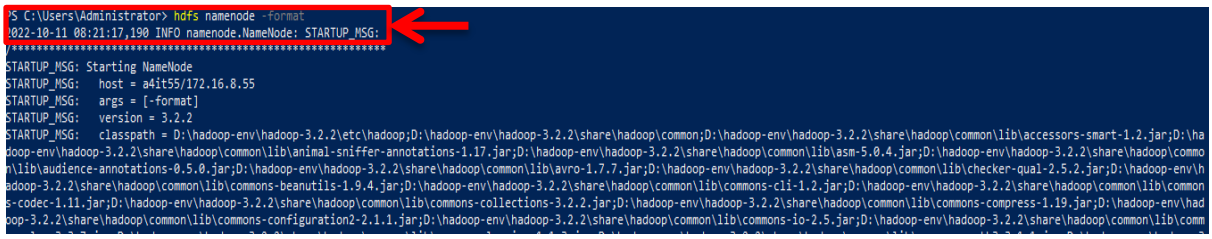
<description>Yarn Node Manager Aux Service</description>

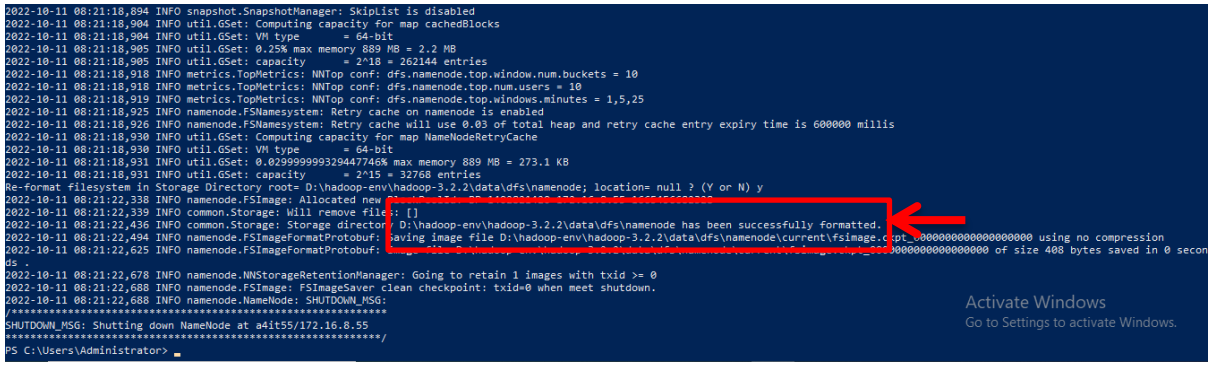
</property>

Step 6: Format the namenode

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

hdfs namenode –format

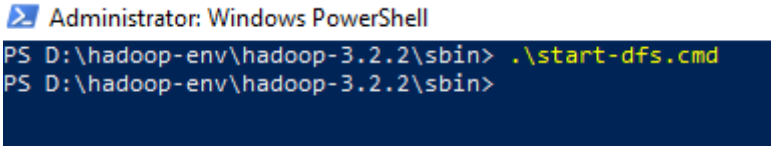




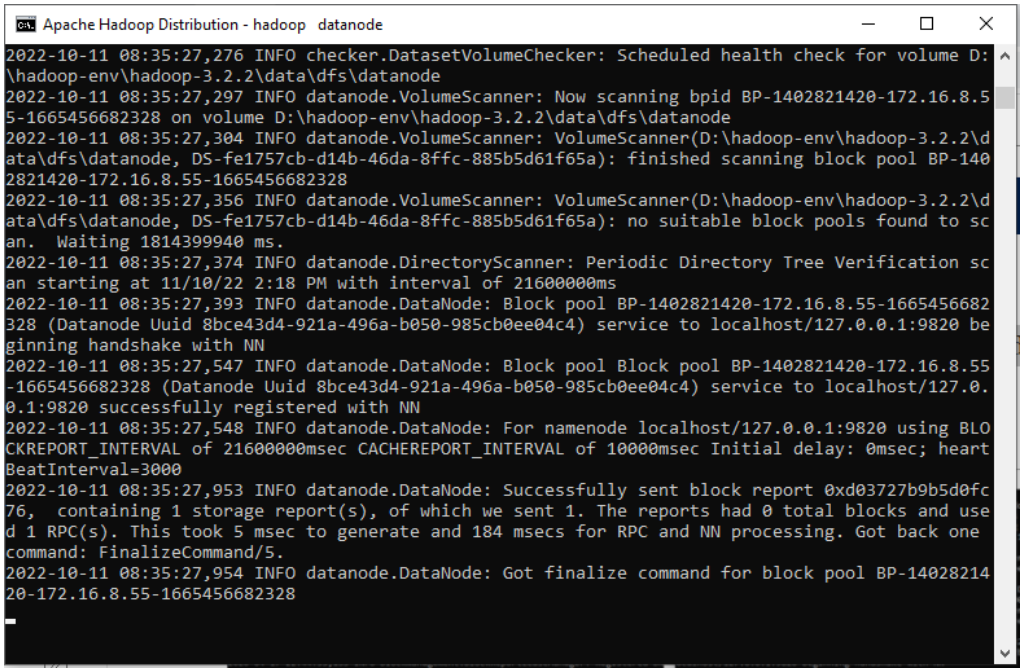
Step 7: Start the Hadoop nodes

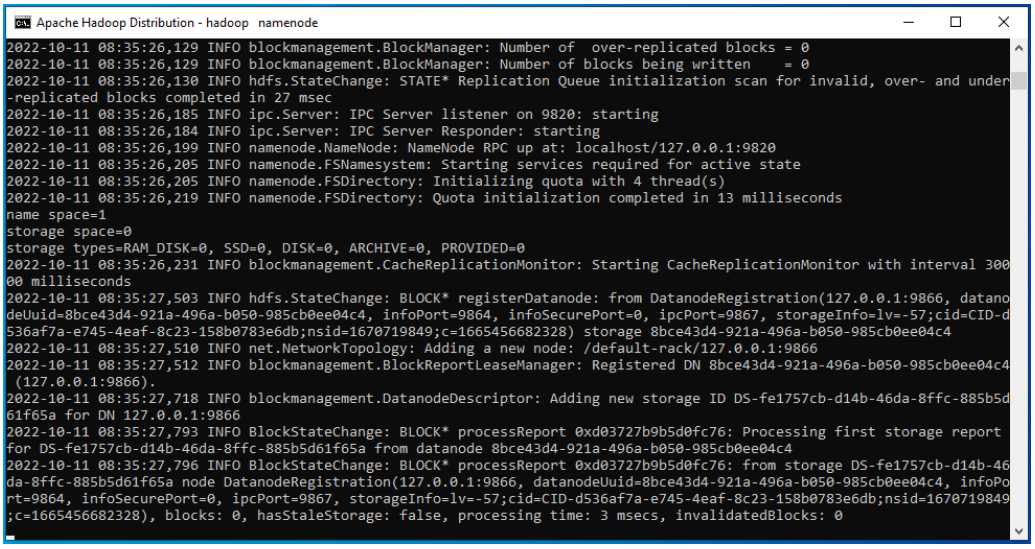
Now navigate to the location “D:\hadoop-env\hadoop-3.2.2\sbin” in powershell and then run the following command to start the Hadoop nodes:

.\start-dfs.cmd



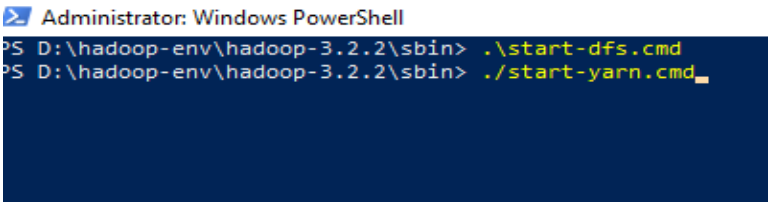
Two command prompt windows will open (one for the **namenode** and one for the **datanode**) as follows:



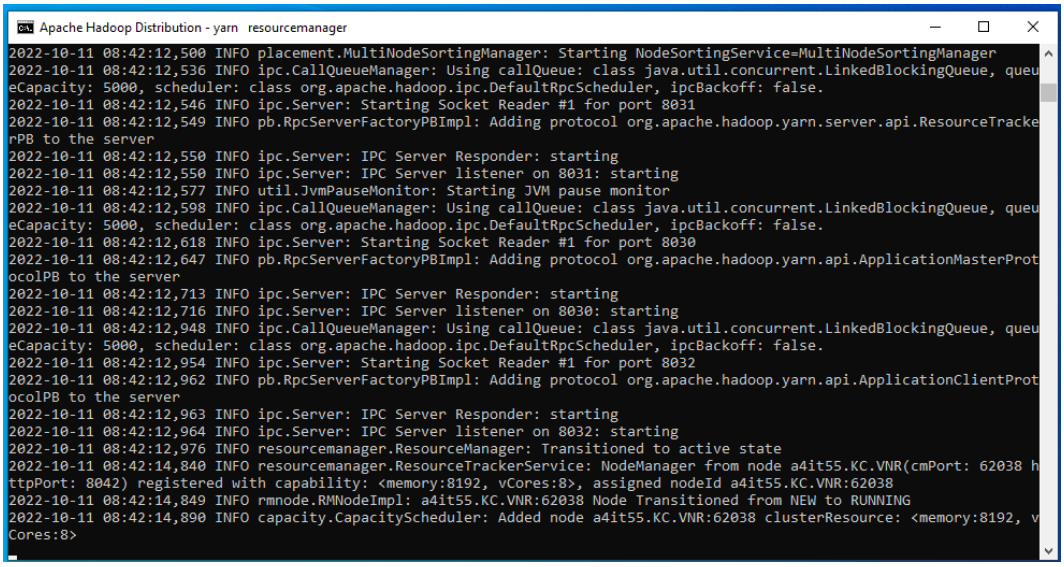


Next, Start the Hadoop Yarn service using the following command:

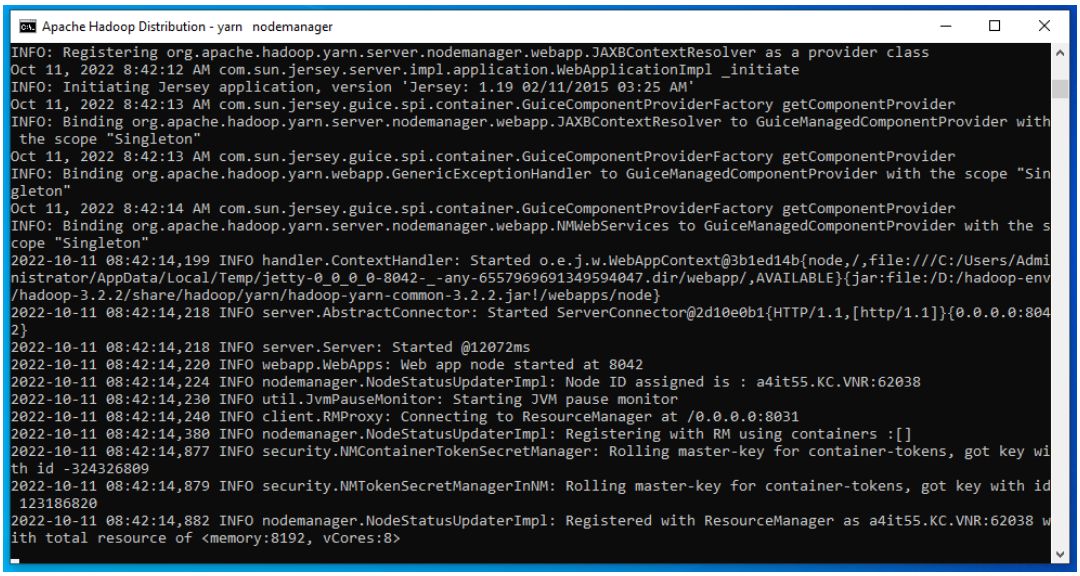
./start-yarn.cmd



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



```
Apache Hadoop Distribution - yarn  resourcemanager
2022-10-11 08:42:12,500 INFO placement.MultiNodeSortingManager: Starting NodeSortingService=MultiNodeSortingManager
2022-10-11 08:42:12,536 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queu
eCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2022-10-11 08:42:12,546 INFO ipc.Server: Starting Socket Reader #1 for port 8031
2022-10-11 08:42:12,549 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceTracke
rPB to the server
2022-10-11 08:42:12,550 INFO ipc.Server: IPC Server Responder: starting
2022-10-11 08:42:12,550 INFO ipc.Server: IPC Server listener on 8031: starting
2022-10-11 08:42:12,577 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2022-10-11 08:42:12,598 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queu
eCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2022-10-11 08:42:12,618 INFO ipc.Server: Starting Socket Reader #1 for port 8030
2022-10-11 08:42:12,647 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationMasterProt
ocolPB to the server
2022-10-11 08:42:12,713 INFO ipc.Server: IPC Server Responder: starting
2022-10-11 08:42:12,716 INFO ipc.Server: IPC Server listener on 8030: starting
2022-10-11 08:42:12,948 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queu
eCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2022-10-11 08:42:12,954 INFO ipc.Server: Starting Socket Reader #1 for port 8032
2022-10-11 08:42:12,962 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationClientProt
ocolPB to the server
2022-10-11 08:42:12,963 INFO ipc.Server: IPC Server Responder: starting
2022-10-11 08:42:12,964 INFO ipc.Server: IPC Server listener on 8032: starting
2022-10-11 08:42:12,976 INFO resourcemanager.ResourceManager: Transitioned to active state
2022-10-11 08:42:14,840 INFO resourcemanager.ResourceTrackerService: NodeManager from node a4it55.KC.VNR(cmPort: 62038 h
ttpPort: 8042) registered with capability: <memory:8192, vCores:8>, assigned nodeId a4it55.KC.VNR:62038
2022-10-11 08:42:14,849 INFO rmnode.RMNodeImpl: a4it55.KC.VNR:62038 Node Transitioned from NEW to RUNNING
2022-10-11 08:42:14,890 INFO capacity.CapacityScheduler: Added node a4it55.KC.VNR:62038 clusterResource: <memory:8192, v
Cores:8>
```



```
Apache Hadoop Distribution - yarn  nodemanager
INFO: Registering org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver as a provider class
Oct 11, 2022 8:42:12 AM com.sun.jersey.server.impl.application.WebApplicationImpl _initiate
INFO: Initiating Jersey application, version 'Jersey: 1.19 02/11/2015 03:25 AM'
Oct 11, 2022 8:42:13 AM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver to GuiceManagedComponentProvider with
 the scope "Singleton"
Oct 11, 2022 8:42:13 AM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponentProvider with the scope "Sin
gleton"
Oct 11, 2022 8:42:14 AM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices to GuiceManagedComponentProvider with the s
cope "Singleton"
2022-10-11 08:42:14,199 INFO handler.ContextHandler: Started o.e.j.w.WebAppContext@3b1ed14b{node,/,file:///C:/Users/Admi
nistrator/AppData/Local/Temp/jetty-0_0_0_0-8042-_-any-6557969691349594047.dir/webapp/,AVAILABLE}{jar:file:/D:/hadoop-env
/hadoop-3.2.2/share/hadoop/yarn/hadoop-yarn-common-3.2.2.jar!/webapps/node}
2022-10-11 08:42:14,218 INFO server.AbstractConnector: Started ServerConnector@2d10e0b1{HTTP/1.1,[http/1.1]}{0.0.0.0:804
2}
2022-10-11 08:42:14,218 INFO server.Server: Started @12072ms
2022-10-11 08:42:14,220 INFO webapp.WebApps: Web app node started at 8042
2022-10-11 08:42:14,224 INFO nodemanager.NodeStatusUpdaterImpl: Node ID assigned is : a4it55.KC.VNR:62038
2022-10-11 08:42:14,230 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2022-10-11 08:42:14,240 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8031
2022-10-11 08:42:14,380 INFO nodemanager.NodeStatusUpdaterImpl: Registering with RM using containers :[]
2022-10-11 08:42:14,877 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key wi
th id -324326809
2022-10-11 08:42:14,879 INFO security.NMTokenSecretManagerInNM: Rolling master-key for container-tokens, got key with id
 123186820
2022-10-11 08:42:14,882 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceManager as a4it55.KC.VNR:62038 w
ith total resource of <memory:8192, vCores:8>
```

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

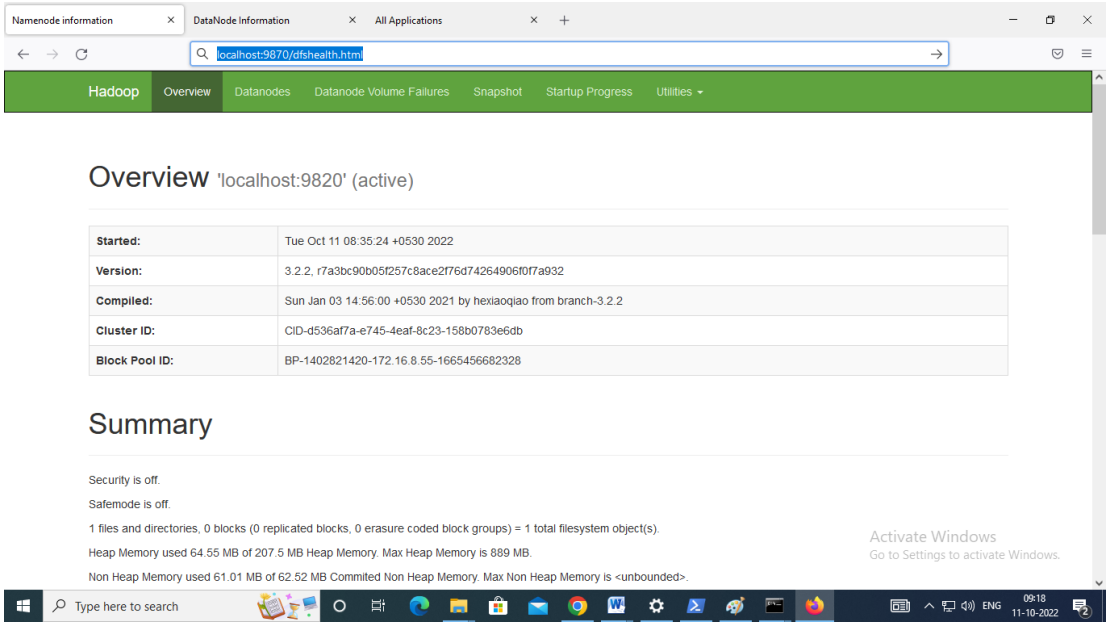
14636 Jps

```
PS D:\hadoop-env\hadoop-3.2.2\sbin> jps
4288 Jps
4896 ResourceManager
4708 NameNode
5852 NodeManager
6860 DataNode
PS D:\hadoop-env\hadoop-3.2.2\sbin>
```

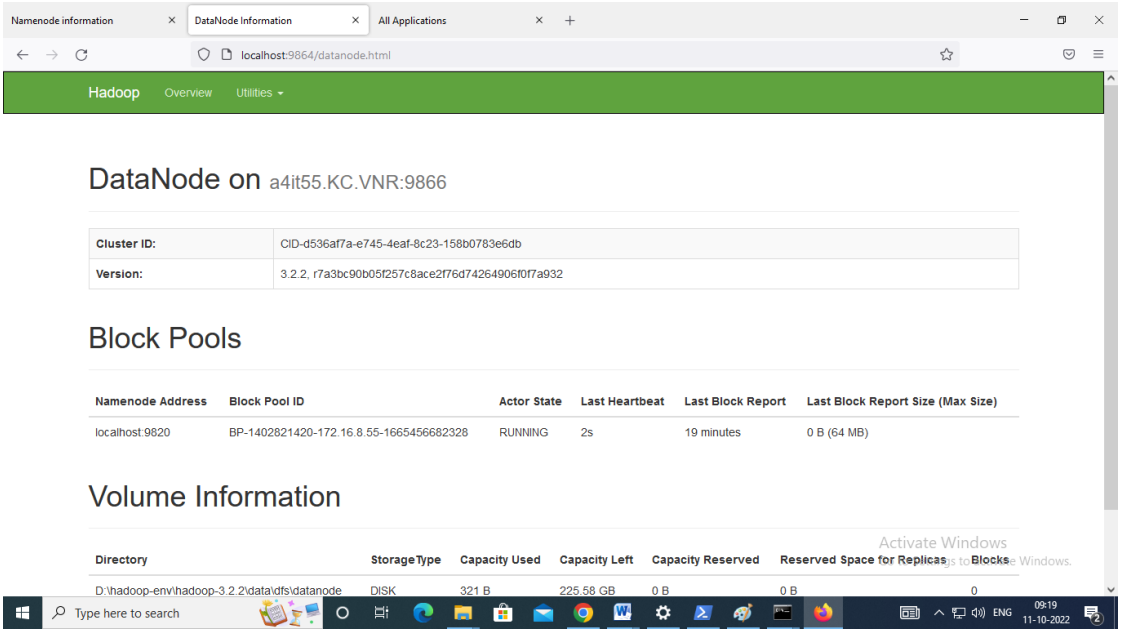
Output

Hadoop Web UI

http://localhost:9870/dfshealth.html



http://localhost:9864/datanode.html



http://localhost:8088/cluster