Assignment No- 1

Aim : Installing Hadoop 3.2.4 Single node cluster on Windows 10

Theory :

Introduction: Hadoop Ecosystem is a platform or a suite which provides various services to solve the big data problems. It includes Apache projects and various commercial tools and solutions. There are four major elements of Hadoop i.e. HDFS, MapReduce, YARN, and Hadoop Common. Most of the tools or solutions are used to supplement or support these major elements. All these tools work collectively to provide services such as absorption, analysis, storage and maintenance of data etc.

There are two ways to install Hadoop,

- Single node
- Multi node

1. **Single Node:**

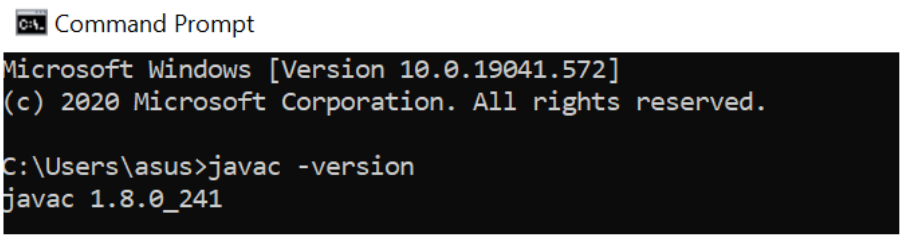
Single node cluster means only one DataNode running and setting up all the NameNode, DataNode, ResourceManager and NodeManager on a single machine.

2. **Multi Node:**

Multi node cluster, there are more than one DataNode running and each DataNode is running on different machines. The multi node cluster is practically used in organizations for analyzing Big Data. In real time when we deal with petabytes of data, it needs to be distributed across hundreds of machines to be processed.

Installing Single node Hadoop cluster

1. **Prerequisites to install Hadoop on windows**
    1. Install Java
        - Hadoop 3 requires a Java 8 installation
        - Java 8 runtime environment (JRE)
        - Java 8 development Kit (JDK)

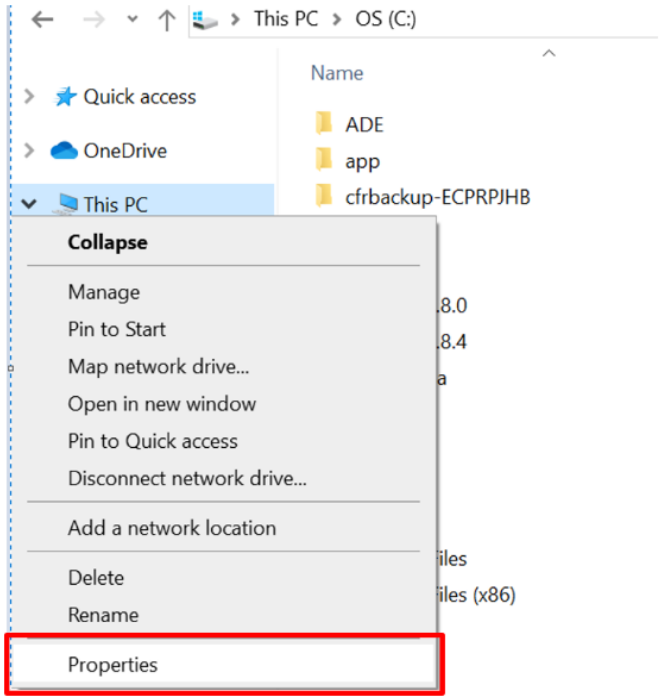


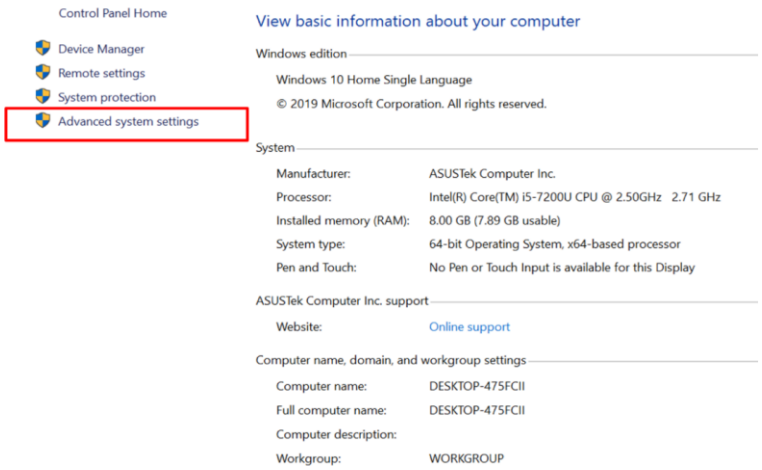
2. **Download Hadoop**

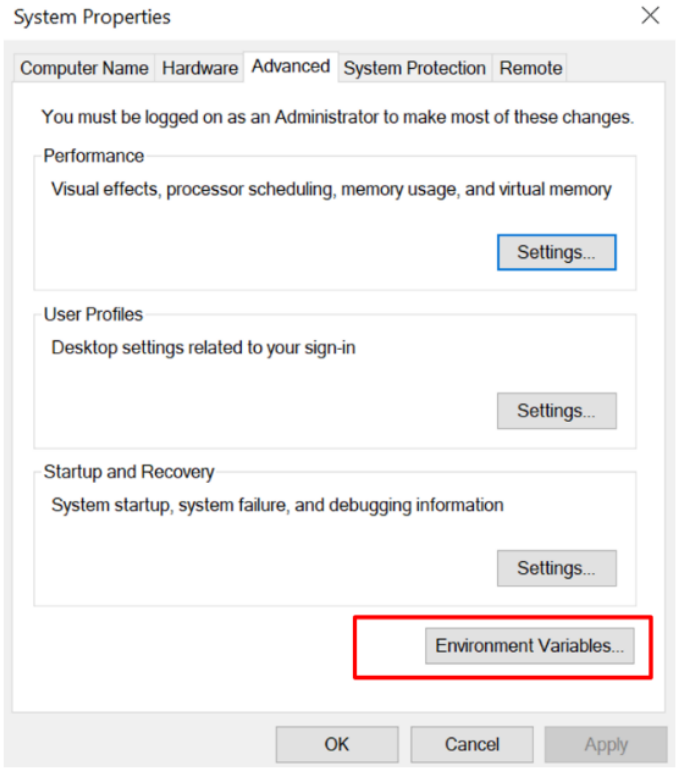
Download Hadoop binaries and unzip it.

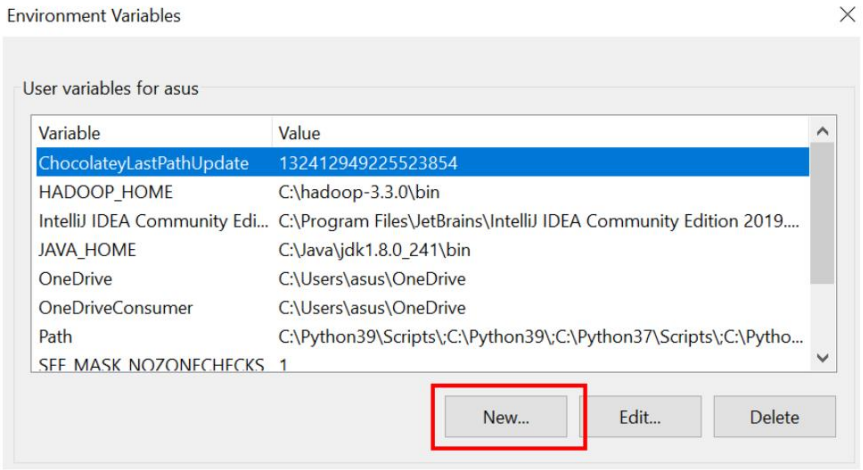
3. Setting up environment variables

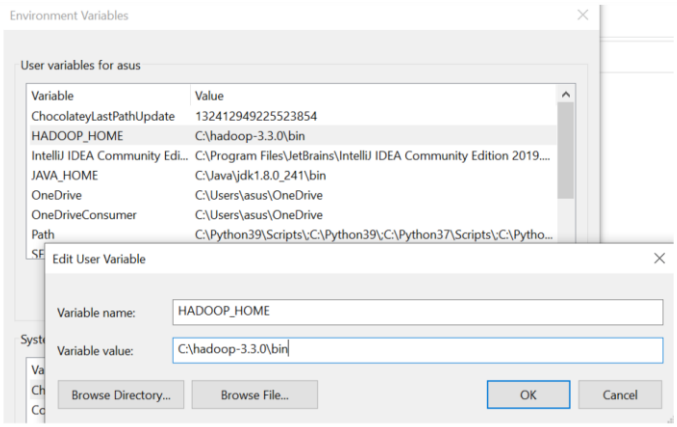
After installing Hadoop and its prerequisites, we should configure the environment variables to define Hadoop and Java default paths

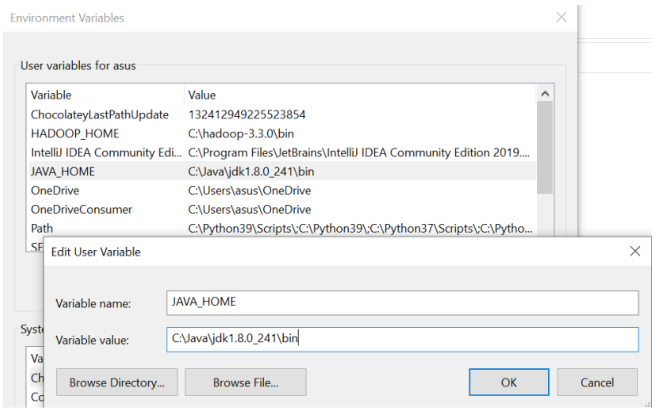


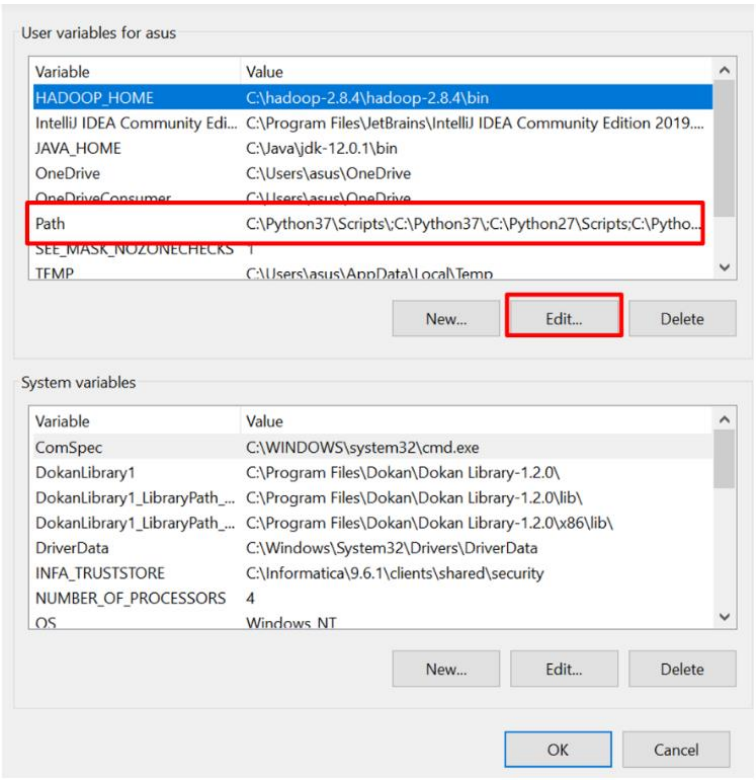


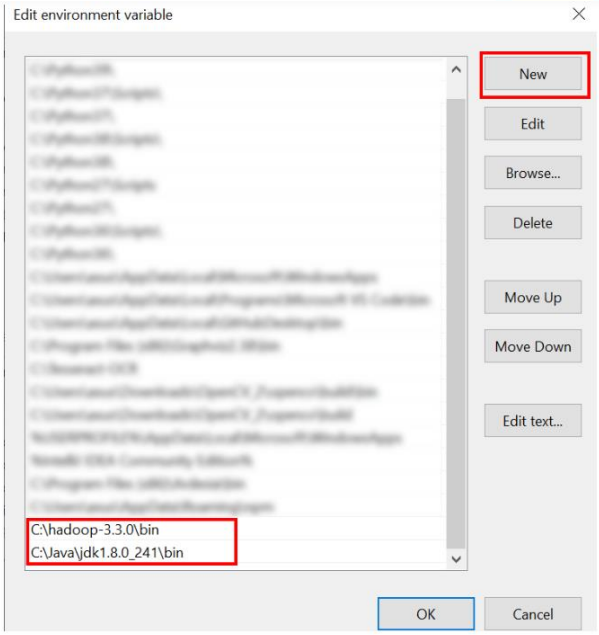

System Properties
Computer Name
Hardware
Advanced
System Protection
Remote
You must be logged on as an Administrator to make most of these changes.
Performance
Visual effects, processor scheduling, memory usage, and virtual memory
Settings...
User Profiles
Desktop settings related to your sign-in
Settings...
Startup and Recovery
System startup, system failure, and debugging information
Settings...
Environment Variables...
OK
Cancel
Apply


Environment Variables
User variables for asus
Variable
Value
ChocolateyLastPathUpdate
132412949225523854
HADOOP_HOME
C:\hadoop-3.3.0\bin
IntelliJ IDEA Community Edi...
C:\Program Files\JetBrains\IntelliJ IDEA Community Edition 2019....
JAVA_HOME
C:\Java\jdk1.8.0_241\bin
OneDrive
C:\Users\asus\OneDrive
OneDriveConsumer
C:\Users\asus\OneDrive
Path
C:\Python39\Scripts\;C:\Python39\;C:\Python37\Scripts\;C:\Pytho...
SEE_MASK_NOZONECHECKS
1
New...
Edit...
Delete


Environment Variables
User variables for asus
Variable
Value
ChocolateyLastPathUpdate
132412949225523854
HADOOP_HOME
C:\hadoop-3.3.0\bin
IntelliJ IDEA Community Edi...
C:\Program Files\JetBrains\IntelliJ IDEA Community Edition 2019....
JAVA_HOME
C:\Java\jdk1.8.0_241\bin
OneDrive
C:\Users\asus\OneDrive
OneDriveConsumer
C:\Users\asus\OneDrive
Path
C:\Python39\Scripts\;C:\Python39\;C:\Python37\Scripts\;C:\Pytho...
Edit User Variable
Variable name:
HADOOP_HOME
Variable value:
C:\hadoop-3.3.0\bin
Browse Directory...
Browse File...
OK
Cancel


Environment Variables
User variables for asus
Variable
Value
ChocolateyLastPathUpdate
132412949225523854
HADOOP_HOME
C:\hadoop-3.3.0\bin
IntelliJ IDEA Community Edi...
C:\Program Files\JetBrains\IntelliJ IDEA Community Edition 2019....
JAVA_HOME
C:\Java\jdk1.8.0_241\bin
OneDrive
C:\Users\asus\OneDrive
OneDriveConsumer
C:\Users\asus\OneDrive
Path
C:\Python39\Scripts\;C:\Python39\;C:\Python37\Scripts\;C:\Pytho...
Edit User Variable
Variable name:
JAVA_HOME
Variable value:
C:\Java\jdk1.8.0_241\bin
Browse Directory...
Browse File...
OK
Cancel


User variables for asus
Variable
Value
HADOOP_HOME
C:\hadoop-2.8.4\hadoop-2.8.4\bin
IntelliJ IDEA Community Edi...
C:\Program Files\JetBrains\IntelliJ IDEA Community Edition 2019....
JAVA_HOME
C:\Java\jdk-12.0.1\bin
OneDrive
C:\Users\asus\OneDrive
Path
C:\Python37\Scripts\;C:\Python37\;C:\Python27\Scripts;C:\Pytho...
TEMP
C:\Users\asus\AppData\Local\Temp
New...
Edit...
Delete
System variables
Variable
Value
ComSpec
C:\WINDOWS\system32\cmd.exe
DokanLibrary1
C:\Program Files\Dokan\Dokan Library-1.2.0\
DokanLibrary1_LibraryPath_...
C:\Program Files\Dokan\Dokan Library-1.2.0\lib\
DokanLibrary1_LibraryPath_...
C:\Program Files\Dokan\Dokan Library-1.2.0\x86\lib\
DriverData
C:\Windows\System32\Drivers\DriverData
INFA_TRUSTSTORE
C:\Informatica\9.6.1\clients\shared\security
NUMBER_OF_PROCESSORS
4
OS
Windows_NT
New...
Edit...
Delete
OK
Cancel


Edit environment variable
New
Edit
Browse...
Delete
Move Up
Move Down
Edit text...
C:\hadoop-3.3.0\bin
C:\Java\jdk1.8.0_241\bin
OK
Cancel

4. Configuring Hadoop cluster

For Hadoop Configuration we need to modify Six files that are listed below-

1. Core-site.xml

2. Mapred-site.xml

3. Hdfs-site.xml

4. Yarn-site.xml

5. Hadoop-env.cmd

6. Create two folders datanode and namenode

1. **Edit file C:/Hadoop-3.3.0/etc/hadoop/core-site.xml,**

paste the xml code in folder and save

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

3. **Rename "mapred-site.xml.template" to "mapred-site.xml" and edit this file C:/Hadoop-3.3.0/etc/hadoop/mapred-site.xml, paste xml code and save this file.**

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
```

```
    </configuration>
```

Edit file C:\Hadoop-3.3.0/etc/hadoop/hdfs-site.xml,

paste xml code and save this file.

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>/hadoop-3.3.0/data/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>/hadoop-3.3.0/data/datanode</value>
  </property>
</configuration>
```

Edit file C:/Hadoop-3.3.0/etc/hadoop/yarn-site.xml,

paste xml code and save this file.

```
<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>
```

Hadoop-env.cmd configuration

Edit file C:/Hadoop-3.3.0/etc/hadoop/hadoop-env.cmd

by closing the command line

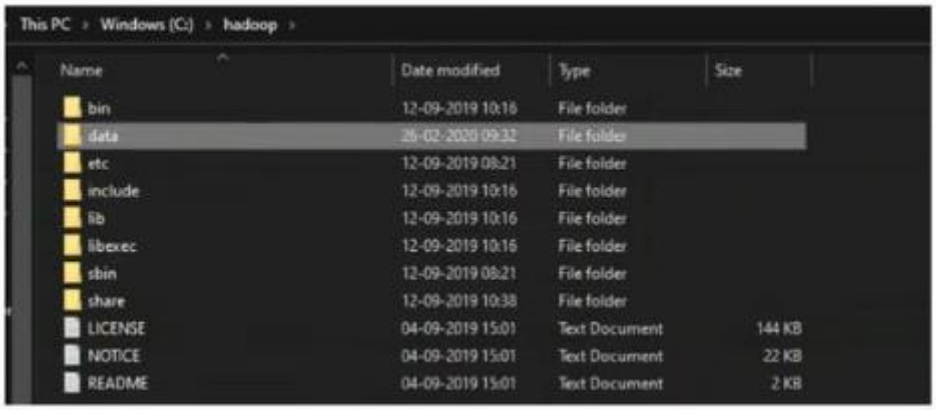
"JAVA_HOME=%JAVA_HOME%" instead of set "JAVA_HOME=C:\Java"

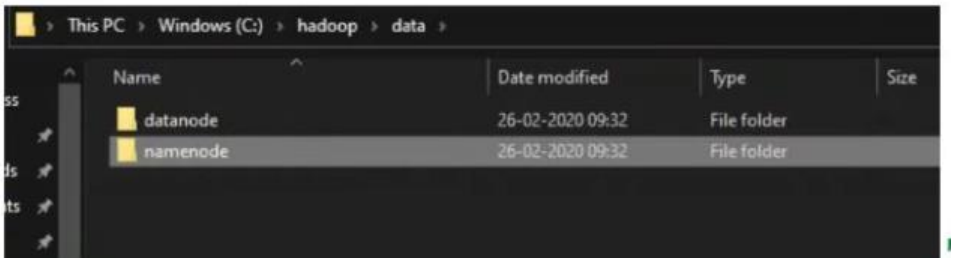
Create datanode and namenode folders

1. Create folder "data" under "C:\Hadoop-2.8.0"

2. Create folder "datanode" under "C:\Hadoop-2.8.0\data"

3. Create folder "namenode" under "C:\Hadoop-2.8.0\data"





4. **Formatting Name node**

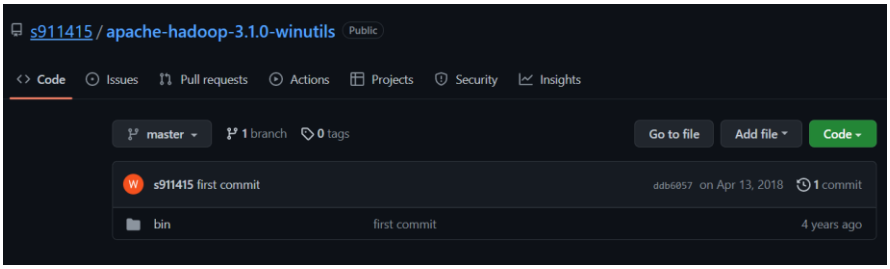
Download https://github.com/s911415/apache-hadoop-3.1.0-winutils

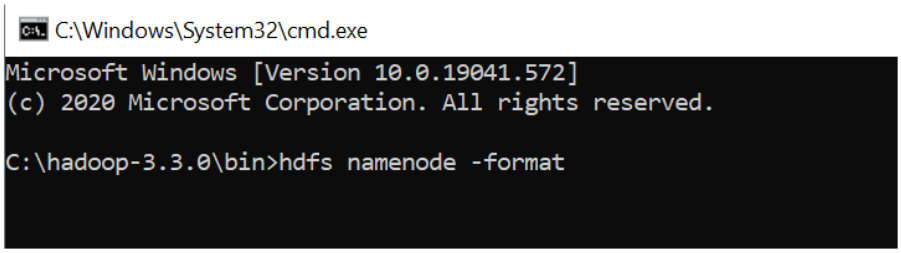
– Copy folder bin and replace existing bin folder in

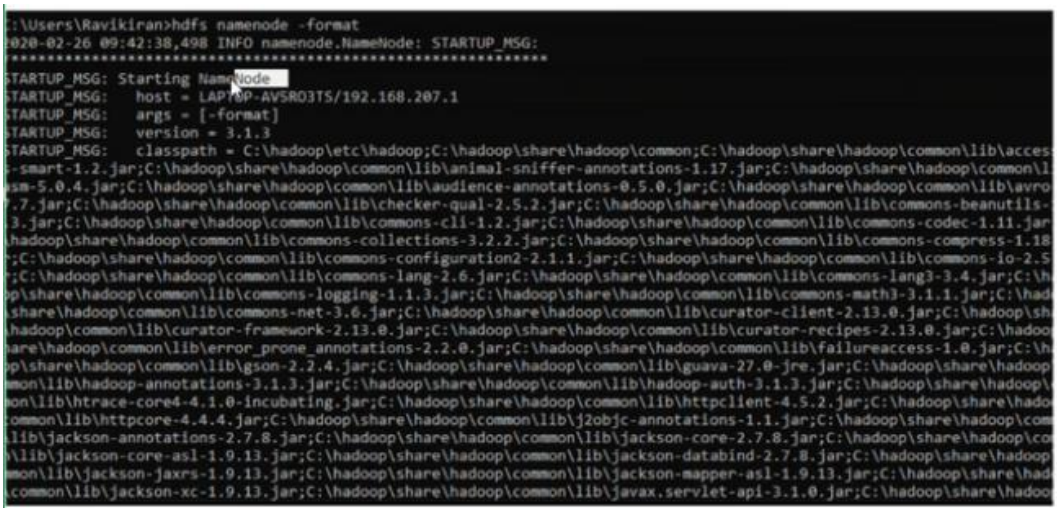
C:\Hadoop-3.3.0\bin

– Format the NameNode

– Open cmd and type command “hdfs namenode –format”



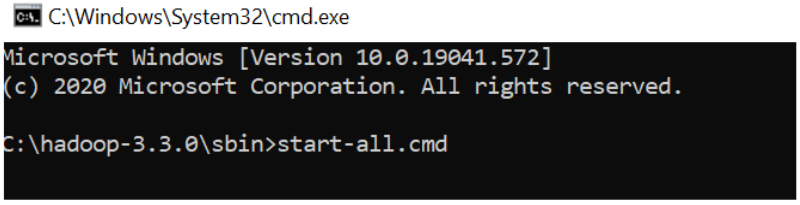




This is showing successful installation of hadoop

5. **Testing the setup**

Open command window (cmd) and typing command “start-all.cmd”



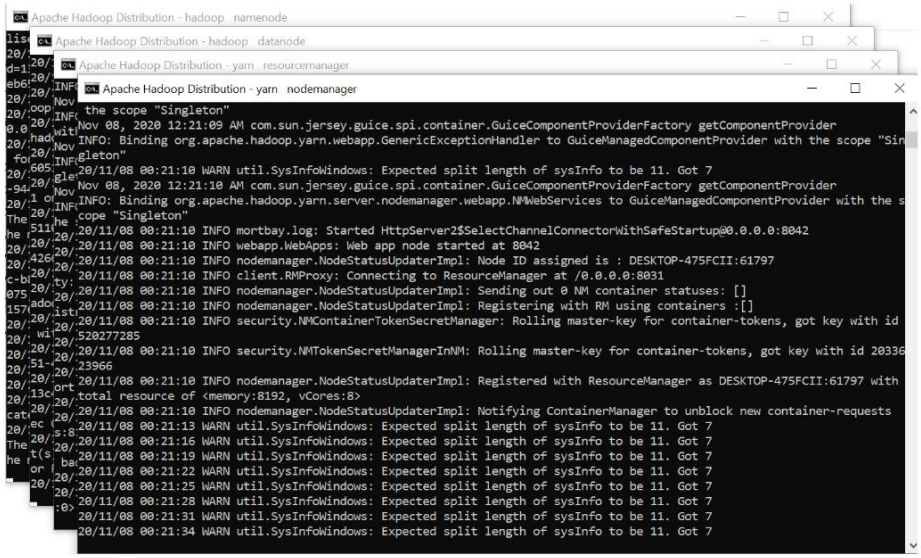
Make sure these apps are running

– Hadoop Namenode

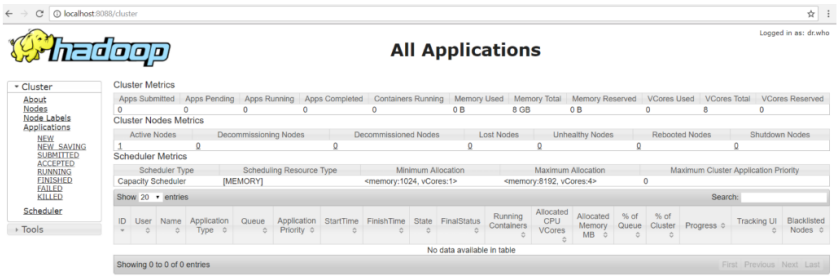
– Hadoop datanode

– YARN Resource Manager

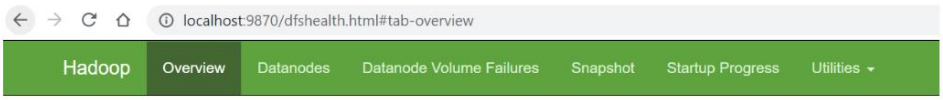
– YARN Node Manager



Open: http://localhost:8088



Open: http://localhost:9870



Conclusion : Thus I have successfully install hadoop in Windows 10 OS.