

## 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:

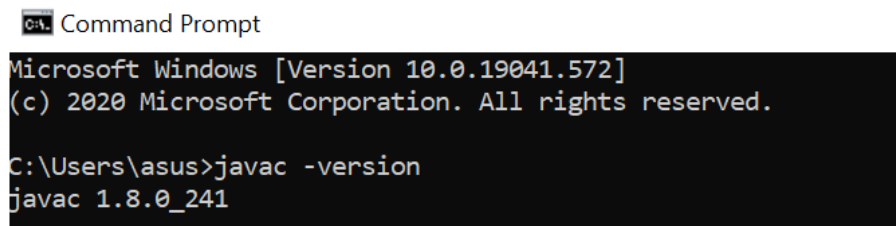
Multi node cluster, there are more than one DataNode running and each DataNode is running on different machines. The multi node cluster is practically used in organizations for analyzing Big Data. In real time when we deal with petabytes of data, it needs to be distributed across hundreds of machines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



```
C:\> Command Prompt
Microsoft Windows [Version 10.0.19041.572]
(c) 2020 Microsoft Corporation. All rights reserved.

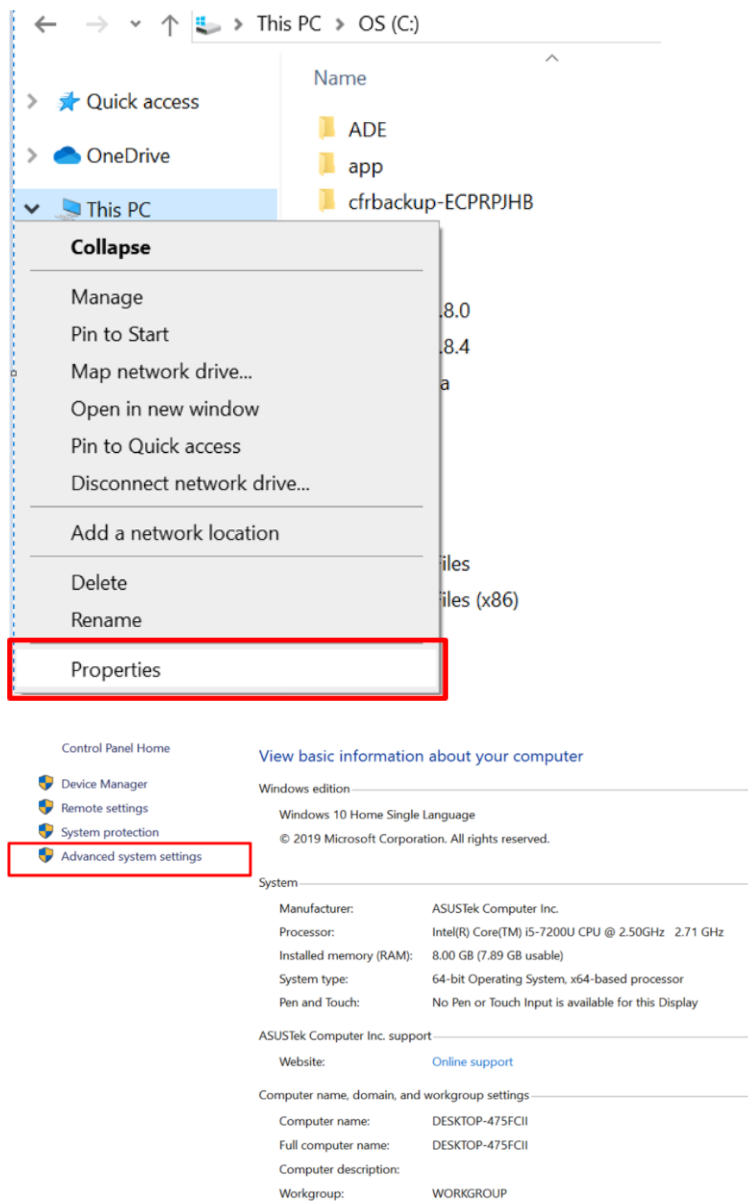
C:\Users\asus>javac -version
javac 1.8.0_241
```

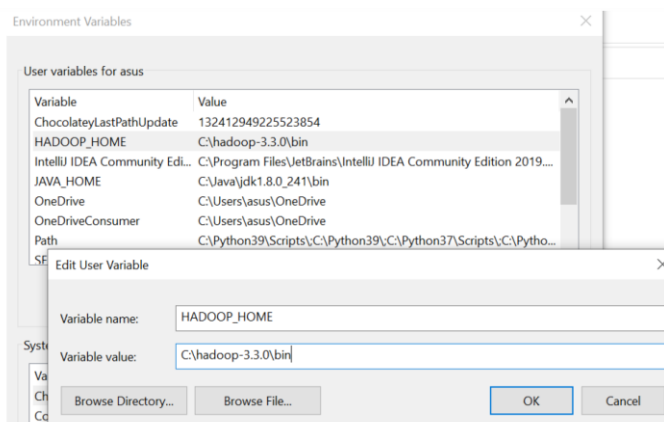
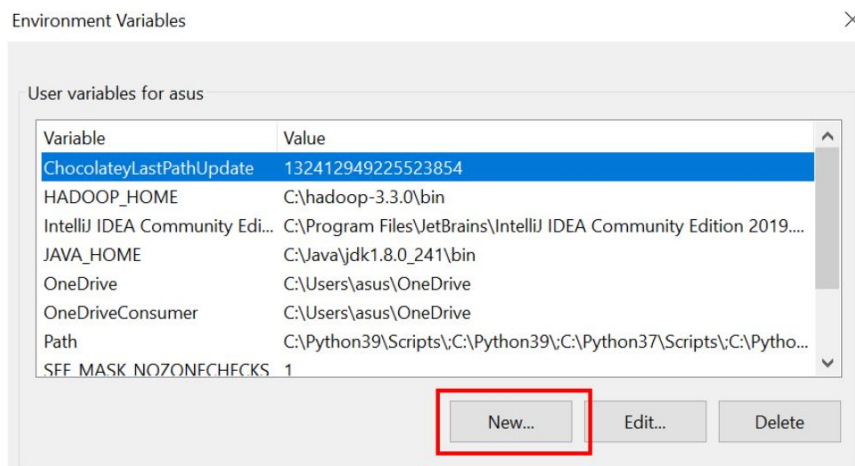
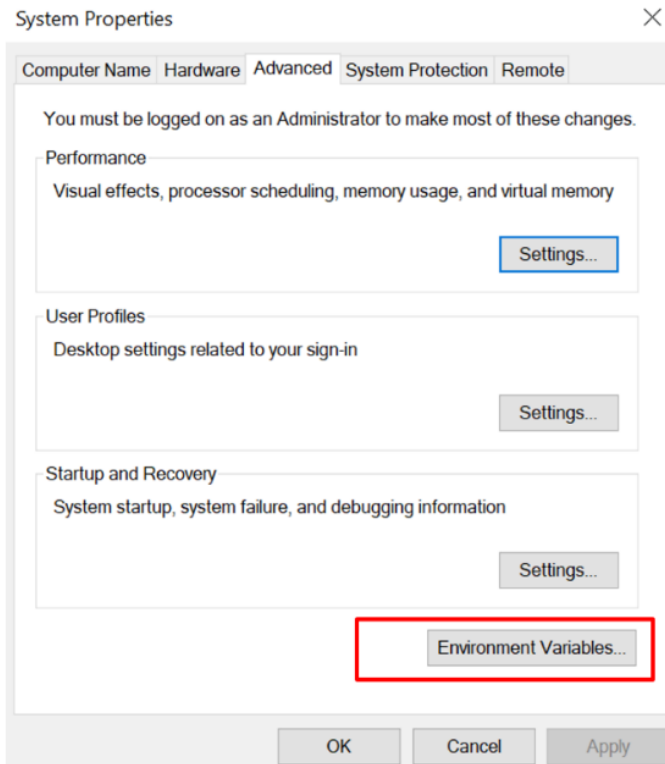
#### 2. Download Hadoop

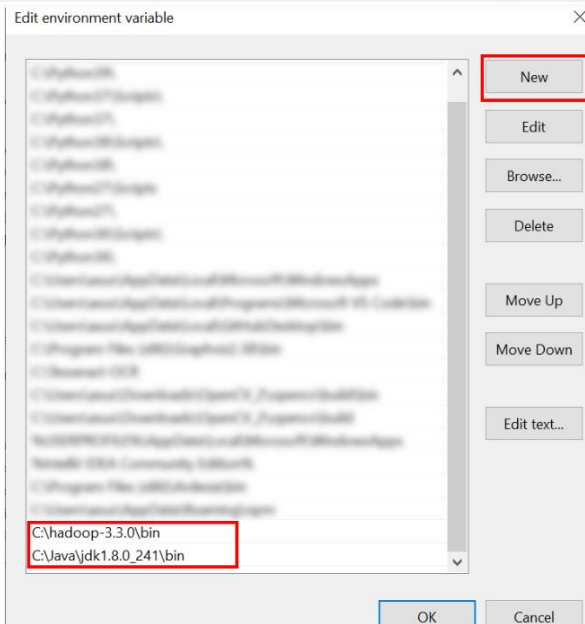
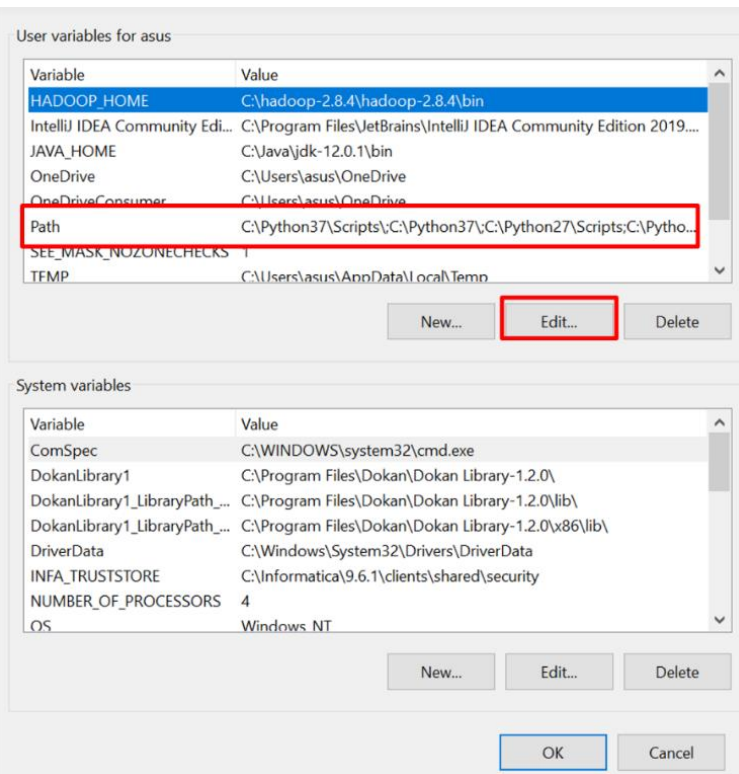
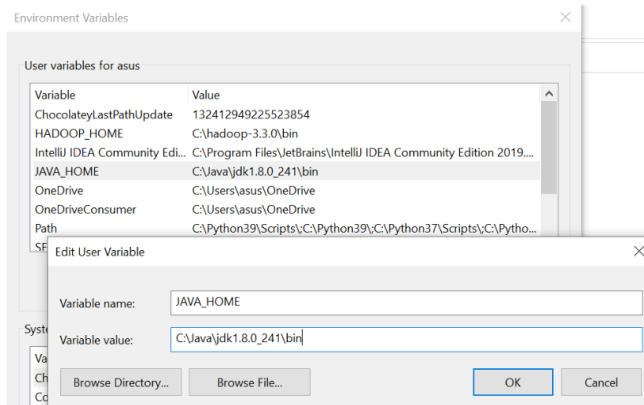
Download Hadoop binaries and unzip it.

### 3. Setting up environment variables

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







#### 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

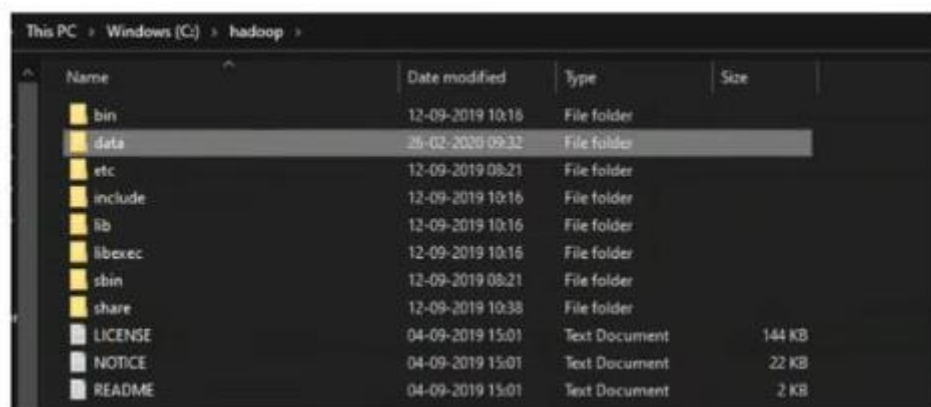
Edit file C:/Hadoop-3.3.0/etc/hadoop/hadoop-env.cmd

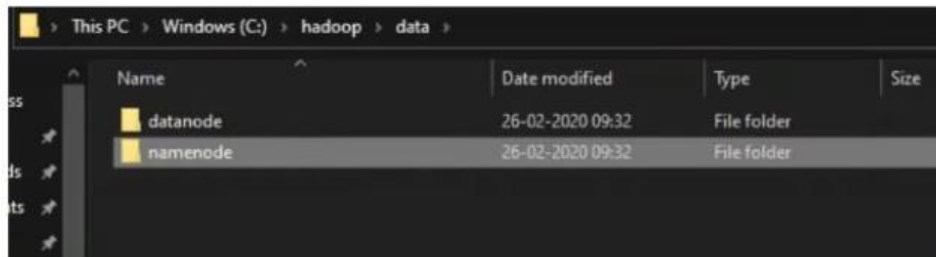
by closing the command line

**“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>

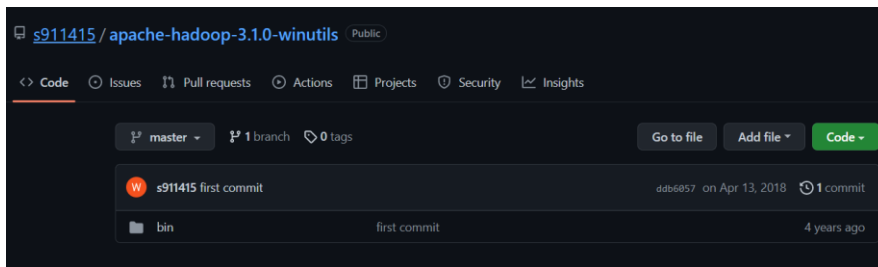
– Copy folder bin and replace existing bin folder in

C:\Hadoop-3.3.0\bin

– Format the NameNode

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





```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19041.572]
(c) 2020 Microsoft Corporation. All rights reserved.

C:\hadoop-3.3.0\bin>hdfs namenode -format
```

```
C:\Users\Ravikiran>hdfs namenode -format
2020-02-26 09:42:38,498 INFO namenode.NameNode: STARTUP_MSG:
=====
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = LAPTOP-AV5R03TS/192.168.207.1
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.1.3
STARTUP_MSG: classpath = C:\hadoop\etc\hadoop;C:\hadoop\share\hadoop\common;C:\hadoop\share\hadoop\common\lib\access-asm-5.0.4.jar;C:\hadoop\share\hadoop\common\lib\animal-sniffer-annotations-1.17.jar;C:\hadoop\share\hadoop\common\lib\asm-7.7.jar;C:\hadoop\share\hadoop\common\lib\checker-qual-2.5.2.jar;C:\hadoop\share\hadoop\common\lib\commons-beanutils-1.3.jar;C:\hadoop\share\hadoop\common\lib\commons-cli-1.2.jar;C:\hadoop\share\hadoop\common\lib\commons-codec-1.11.jar;C:\hadoop\share\hadoop\common\lib\commons-collections-3.2.2.jar;C:\hadoop\share\hadoop\common\lib\commons-compress-1.18.jar;C:\hadoop\share\hadoop\common\lib\commons-configuration2-2.1.1.jar;C:\hadoop\share\hadoop\common\lib\commons-io-2.5.jar;C:\hadoop\share\hadoop\common\lib\commons-lang-2.6.jar;C:\hadoop\share\hadoop\common\lib\commons-lang3-3.4.jar;C:\hadoop\share\hadoop\common\lib\commons-logging-1.1.3.jar;C:\hadoop\share\hadoop\common\lib\commons-math3-3.1.1.jar;C:\hadoop\share\hadoop\common\lib\commons-net-3.6.jar;C:\hadoop\share\hadoop\common\lib\curator-client-2.13.0.jar;C:\hadoop\share\hadoop\common\lib\curator-framework-2.13.0.jar;C:\hadoop\share\hadoop\common\lib\curator-recipes-2.13.0.jar;C:\hadoop\share\hadoop\common\lib\error_prone_annotations-2.2.0.jar;C:\hadoop\share\hadoop\common\lib\failureaccess-1.0.jar;C:\hadoop\share\hadoop\common\lib\gson-2.2.4.jar;C:\hadoop\share\hadoop\common\lib\guava-27.0-jre.jar;C:\hadoop\share\hadoop\common\lib\hadoop-annotations-3.1.3.jar;C:\hadoop\share\hadoop\common\lib\hadoop-auth-3.1.3.jar;C:\hadoop\share\hadoop\common\lib\htrace-core4-4.1.0-incubating.jar;C:\hadoop\share\hadoop\common\lib\httpclient-4.5.2.jar;C:\hadoop\share\hadoop\common\lib\httpcore-4.4.4.jar;C:\hadoop\share\hadoop\common\lib\j2objc-annotations-1.1.jar;C:\hadoop\share\hadoop\common\lib\jackson-annotations-2.7.8.jar;C:\hadoop\share\hadoop\common\lib\jackson-core-2.7.8.jar;C:\hadoop\share\hadoop\common\lib\jackson-core-asl-1.9.13.jar;C:\hadoop\share\hadoop\common\lib\jackson-databind-2.7.8.jar;C:\hadoop\share\hadoop\common\lib\jackson-jaxrs-1.9.13.jar;C:\hadoop\share\hadoop\common\lib\jackson-mapper-asl-1.9.13.jar;C:\hadoop\share\hadoop\common\lib\jackson-xc-1.9.13.jar;C:\hadoop\share\hadoop\common\lib\javax.servlet-api-3.1.0.jar;C:\hadoop\share\hadoop
```

**This is showing successful installation of hadoop**

## 5. Testing the setup

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

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19041.572]
(c) 2020 Microsoft Corporation. All rights reserved.

C:\hadoop-3.3.0\sbin>start-all.cmd
```


Make sure these apps are running

- Hadoop Namenode
- Hadoop datanode
- YARN Resource Manager

- YARN Node Manager

[illegible]

Open: <http://localhost:8088>



# All Applications

localhost:8080/custer

Logged in as: drwh

Cluster

About Nodes

Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	Vcores Used	Vcores Total	Vcores Reserved
0	0	0	0	0	0 B	8 GB	0 B	0	8	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
1	0	0	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
Queue Scheduler	[MEMORY]	<memory:1024, vcores:1>	<memory:8192, vcores:4>	0

Show 20 + entries

Search:

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU Vcores	Allocated Memory MB	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
No data available in table																	

Show 0 to 0 of 0 entries

First Previous Next Last

Open: <http://localhost:9870>

← → ↻ 🏠 ⓘ localhost:9870/dfshealth.html#tab-overview

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

## Overview 'localhost:9000' (✔active)

Started:	Sun Nov 08 16:53:46 +0530 2020
Version:	3.3.0, [REDACTED]9af
Compiled:	Tue Jul 07 00:14:00 +0530 2020 by brahma from branch-3.3.0
Cluster ID:	C [REDACTED]
Block Pool ID:	B [REDACTED]44

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