

# Install Hadoop 2.9.1 on Windows 10

Here, I am going to install Apache Hadoop 2.9.1 on Windows 10 platform.



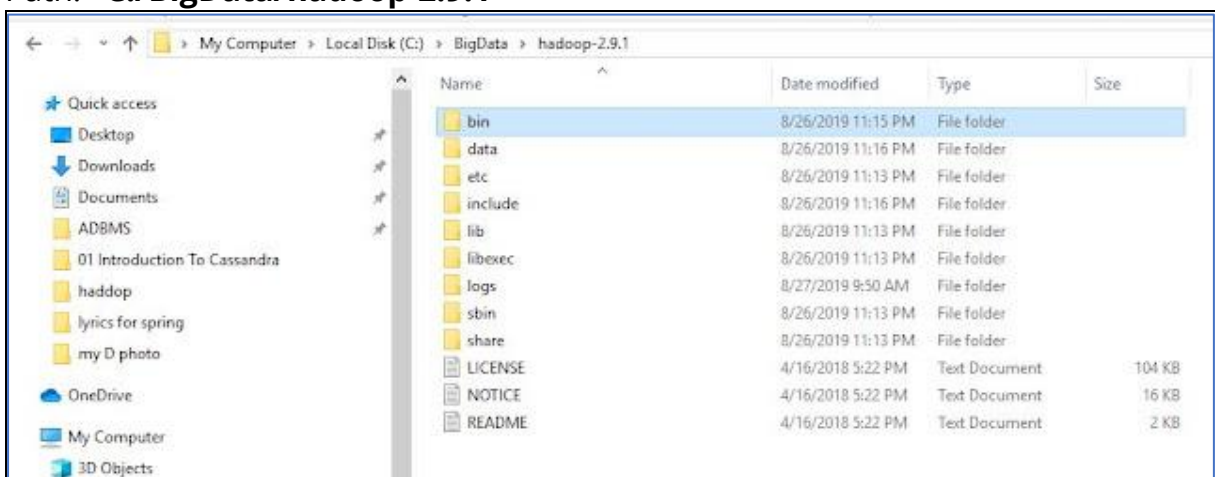
First download the **Hadoop 2.9.1** from the below link.

<https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-2.9.1/hadoop-2.9.1.tar.gz>



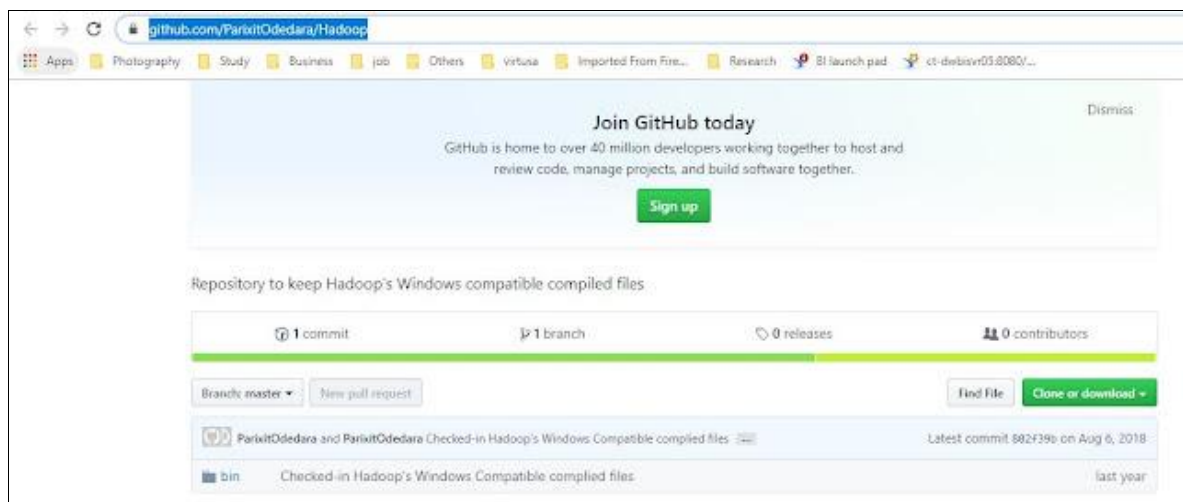
Create a folder path as below and copy the downloaded msi into this folder.

Path:- '**C:/BigData/hadoop-2.9.1**'



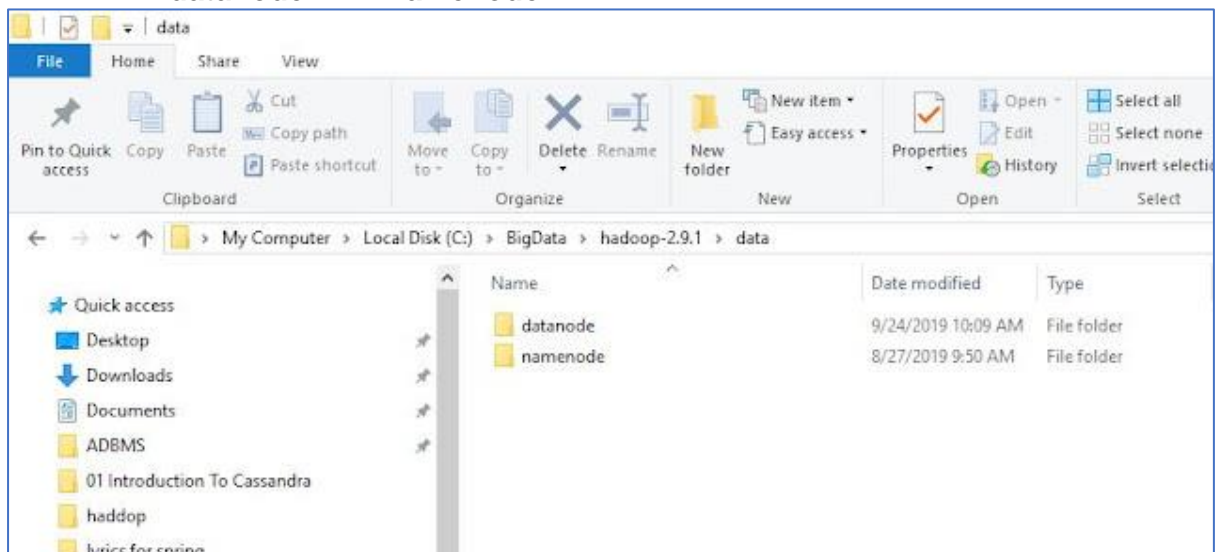
Then download the windows compatible binaries from the git hub repo.

Link:- <https://github.com/ParixitOdedara/Hadoop>



**Extract the zip and copy** all the files present under bin folder to C:\BigData\hadoop-2.9.1\bin. Replace the existing files as well.

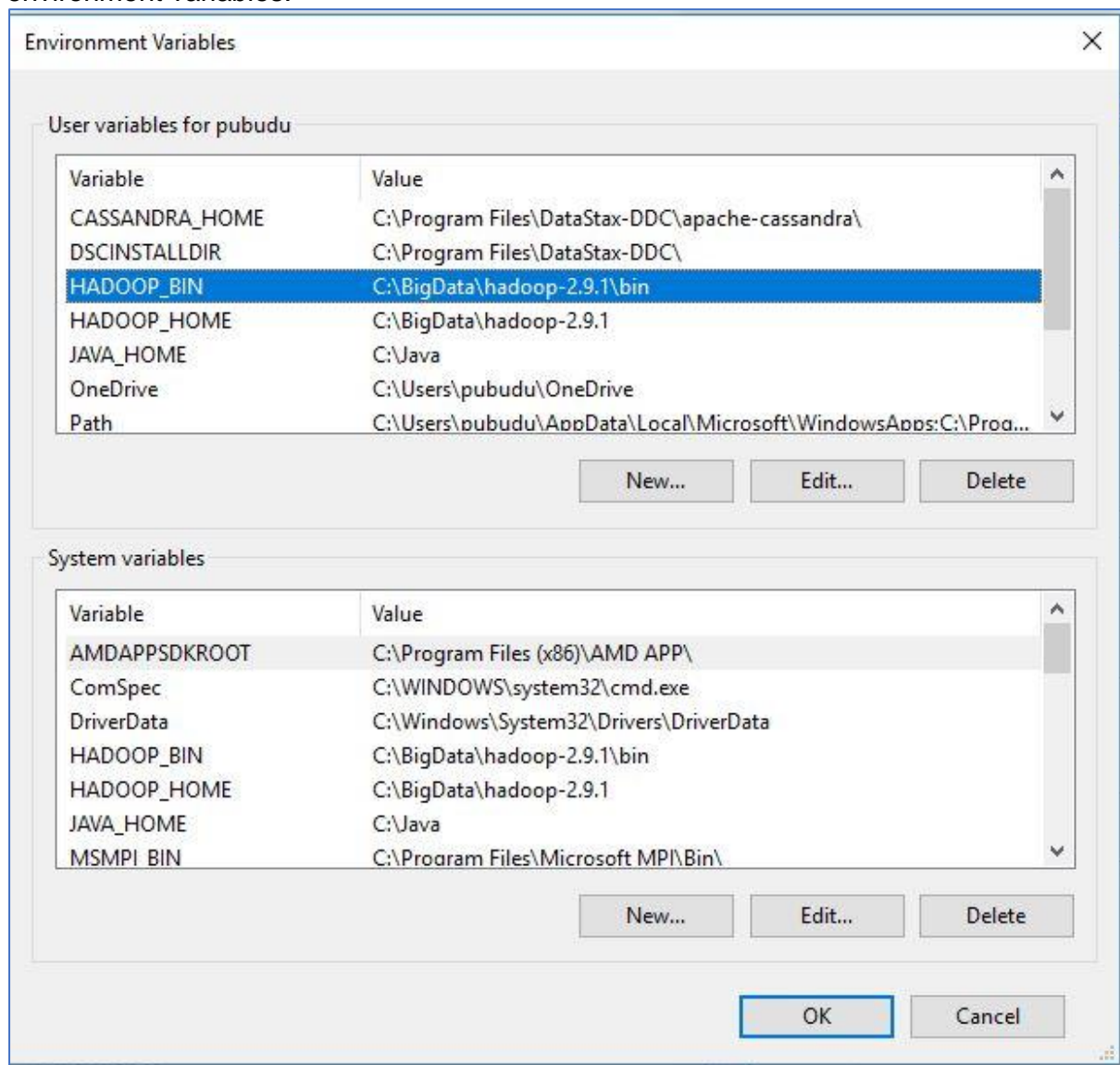
Go to **C:/BigData/ hadoop-2.9.1** and create a folder '**data**'. Inside the 'data' folder create two folders '**datanode**' and '**namenode**'.



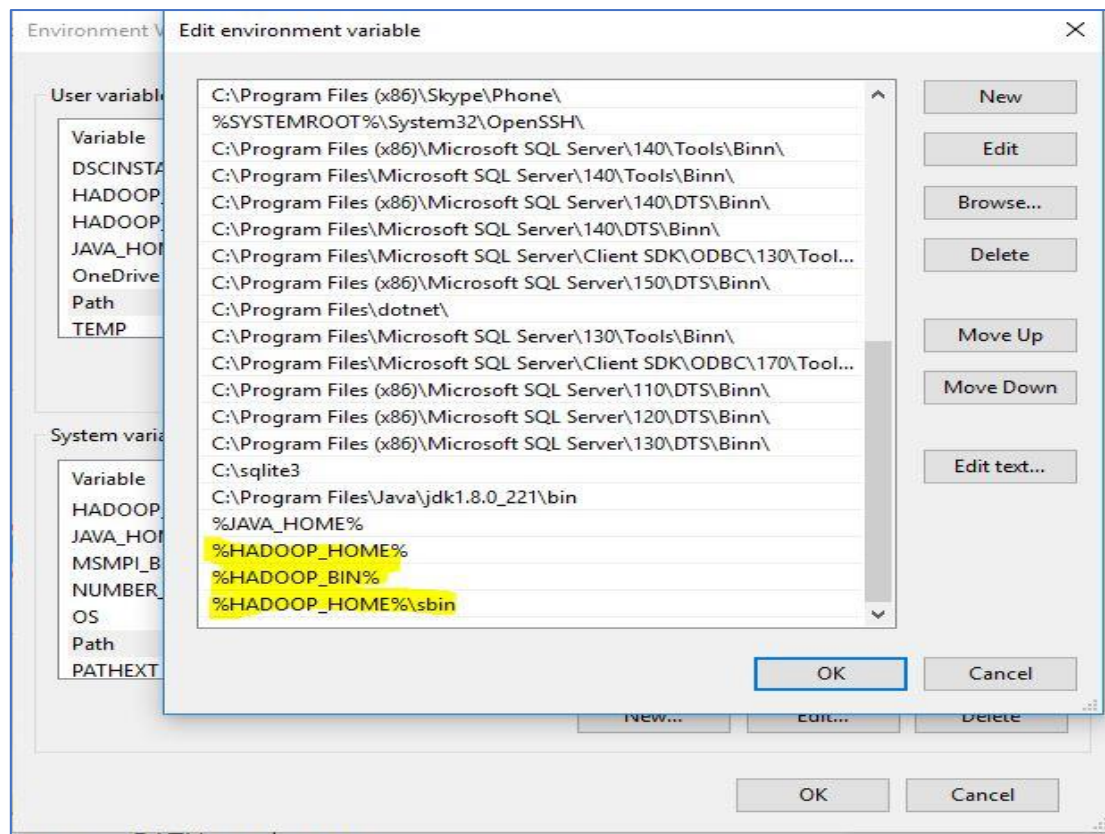
Then Set Hadoop Environment Variables

***HADOOP\_HOME="C:\BigData\hadoop-2.9.1"***  
***HADOOP\_BIN="C:\BigData\hadoop-2.9.1\bin"***  
***JAVA\_HOME=<JDK installation location>"***

To set these variables, go to My Computer or This PC. Right click --> Properties --> Advanced System settings --> Environment variables. Click New to create a new environment variables.



Then edit PATH Environment Variable



To validate the above setting, **open new cmd** and check the output.

**echo %HADOOP\_HOME%**

**echo %HADOOP\_BIN%**

**echo %PATH%**

```

Microsoft Windows [Version 10.0.17134.1006]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\pubudu>echo %HADOOP_HOME%
C:\BigData\hadoop-2.9.1

C:\Users\pubudu>echo %HADOOP_BIN%
C:\BigData\hadoop-2.9.1\bin

C:\Users\pubudu>echo %PATH%
C:\Program Files (x86)\Common Files\Oracle\Java\javapath;F:\Oracle\product\12.2.0\dbhome_1\bin;C:\Program Files\Microsoft
MPI\Bin\;C:\Program Files (x86)\AMD APP\bin\x86_64;C:\Program Files (x86)\AMD APP\bin\x86;C:\Program Files (x86)\Intel
\ICLS Client\;C:\Program Files\Intel\ICLS Client\;C:\WINDOWS\system32;C:\WINDOWS;C:\WINDOWS\System32\Wbem;C:\WINDOWS\Sys
tem32\WindowsPowerShell\v1.0\;C:\Program Files\Intel\Intel(R) Management Engine Components\DAL;C:\Program Files\Intel\In
tel(R) Management Engine Components\IPT;C:\Program Files (x86)\Intel\Intel(R) Management Engine Components\DAL;C:\Progra
m Files (x86)\Intel\Intel(R) Management Engine Components\IPT;C:\Program Files\WIDCOMM\Bluetooth Software\;C:\Program Fi
les\WIDCOMM\Bluetooth Software\;C:\Program Files (x86)\ATI Technologies\ATI.ACE\Core-Static;C:\Program Files (x8
6)\Skype\Phone\;C:\WINDOWS\System32\OpenSSH\;C:\Program Files (x86)\Microsoft SQL Server\140\Tools\Binn\;C:\Program File
s\Microsoft SQL Server\140\Tools\Binn\;C:\Program Files (x86)\Microsoft SQL Server\140\Tools\Binn\;C:\Program Files\Microso
ft SQL Server\140\Tools\Binn\;C:\Program Files\Microsoft SQL Server\140\Tools\Binn\;C:\Program Files (x86
)\Microsoft SQL Server\150\Tools\Binn\;C:\Program Files\dotnet\;C:\Program Files\Microsoft SQL Server\130\Tools\Binn\;C:\P
rogram Files\Microsoft SQL Server\Client SDK\ODBC\170\Tools\Binn\;C:\Program Files (x86)\Microsoft SQL Server\110\Tools\B
inn\;C:\Program Files (x86)\Microsoft SQL Server\120\Tools\Binn\;C:\Program Files (x86)\Microsoft SQL Server\130\Tools\Binn\
;C:\sqlite3;C:\Program Files\Java\jdk1.8.0_221\bin;C:\Java;C:\BigData\hadoop-2.9.1;C:\BigData\hadoop-2.9.1\bin;C:\BigData
\hadoop-2.9.1\sbin;C:\Users\pubudu\AppData\Local\Microsoft\WindowsApps;C:\Program Files\Java\jdk1.8.0_221\bin;C:\Program
Files\Java\jdk1.7.0_25\bin;

C:\Users\pubudu>

```

To configure the Hadoop on windows we have to edit below mention files in the extracted location.



1. hadoop-env.cmd
2. core-site.xml
3. hdfs-site.xml
4. mapred-site.xml
5. yarn-site.xml

### Edit hadoop-env.cmd

File location:-C:\BigData\hadoop-2.9.1\etc\hadoop\hadoop-env.cmd

Need to add:-

```

set HADOOP_PREFIX=%HADOOP_HOME%
set HADOOP_CONF_DIR=%HADOOP_PREFIX%\etc\hadoop
set YARN_CONF_DIR=%HADOOP_CONF_DIR%
set PATH=%PATH%;%HADOOP_PREFIX%\bin

88 @rem      potential for a symlink attack.
89 set HADOOP_PID_DIR=%HADOOP_PID_DIR%
90 set HADOOP_SECURE_DN_PID_DIR=%HADOOP_PID_DIR%
91
92 @rem A string representing this instance of hadoop. %USERNAME% by default.
93 set HADOOP_IDENT_STRING=%USERNAME%
94 set HADOOP_PREFIX=%HADOOP_HOME%
95 set HADOOP_CONF_DIR=%HADOOP_PREFIX%\etc\hadoop
96 set YARN_CONF_DIR=%HADOOP_CONF_DIR%
97 set PATH=%PATH%;%HADOOP_PREFIX%\bin

```

### Edit core-site.xml

File Location:- C:\BigData\hadoop-2.9.1\etc\hadoop\core-site.xml

Need to add:-content within <configuration> </configuration> tags.

```

<configuration>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://0.0.0.0:19000</value>
  </property>
</configuration>

14      limitations under the License. See accompanying LICENSE file.
15      -->
16
17      <!-- Put site-specific property overrides in this file. -->
18
19  <configuration>
20    <property>
21      <name>fs.default.name</name>
22      <value>hdfs://0.0.0.0:19000</value>
23    </property>
24  </configuration>
25

```

### **Edit hdfs-site.xml**

File Location:- C:\BigData\hadoop-2.9.1\etc\hadoop\hdfs-site.xml.

Need to add;- below content within <configuration> </configuration> tags.

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>C:\BigData\hadoop-2.9.1\data\namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>C:\BigData\hadoop-2.9.1\data\datanode</value>
  </property>
</configuration>
```

### **Edit mapred-site.xml**

File location:- Open C:\BigData\hadoop-2.9.1\etc\hadoop\mapred-site.xml

Need to add;- below content within <configuration> </configuration> tags. If you don't see mapred-site.xml then open mapred-site.xml.template file and rename it to mapred-site.xml

```
<configuration>
  <property>
    <name>mapreduce.job.user.name</name>
    <value>%USERNAME%</value>
  </property>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
  <property>
    <name>yarn.apps.stagingDir</name>
    <value>/user/%USERNAME%/staging</value>
  </property>
  <property>
    <name>mapreduce.jobtracker.address</name>
```

```
<value>local</value>
</property>
</configuration>
```

## Editing yarn-site.xml

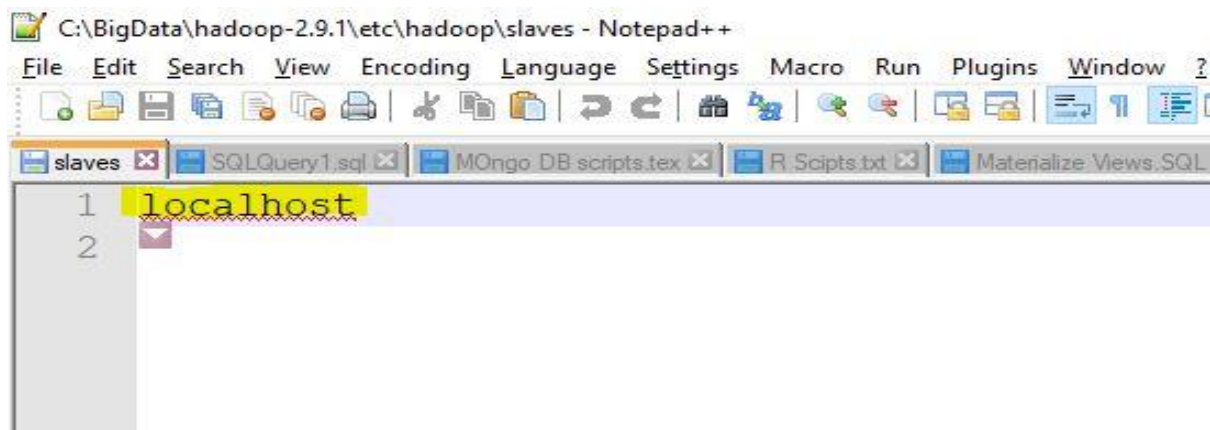
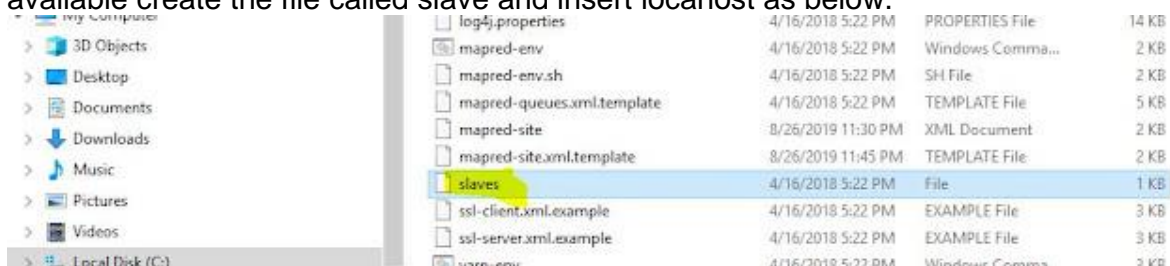
Right click on the file, select edit and paste the following content within <configuration> </configuration> tags.

**Note:- Below part already has the configuration tag, we need to copy only the part inside 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>
<!-- Site specific YARN configuration properties --></configuration>
```

## Additional Configuration:-

Check if C:\BigData\hadoop-2.9.1\etc\hadoop\slaves file is present, if that file not available create the file called slave and insert localhost as below.



## Node formatting

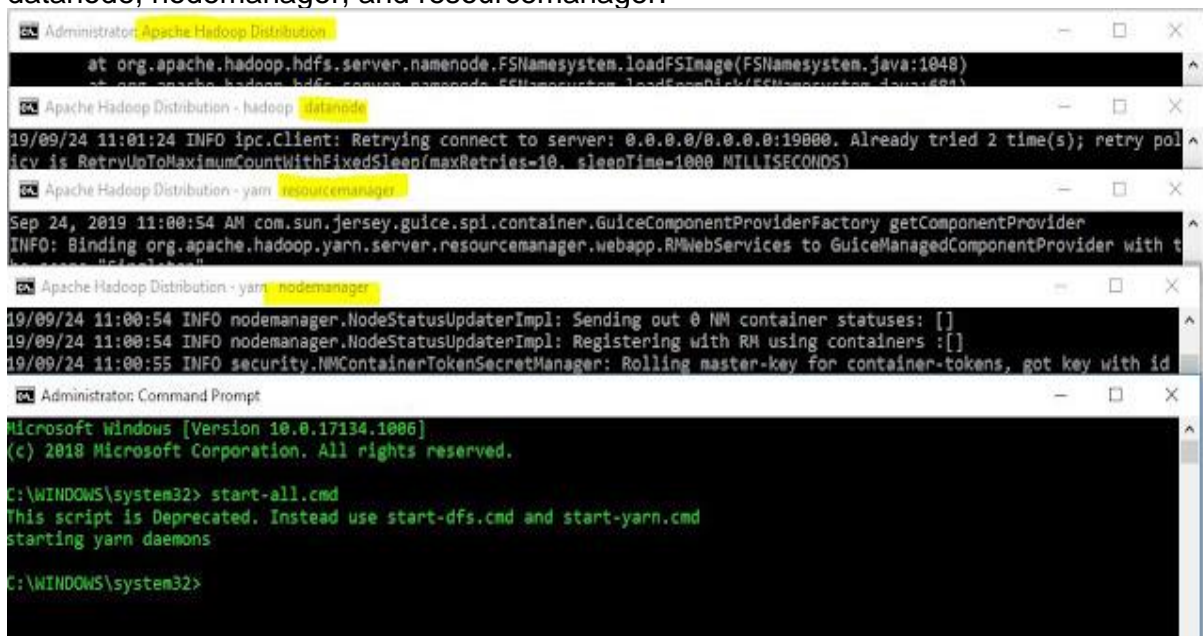
To format the node, **open the cmd** and execute the below command.

***hadoop namenode -format***

```
19/09/24 10:56:21 INFO util.ExitUtil: Exiting with status 1: java.io.IOException: Cannot remove current directory: C:\B1
gData\hadoop-2.9.1\data\namenode\current
19/09/24 10:56:21 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at DESKTOP-3JB06L8/192.168.56.1
*****/
C:\Users\pubudu>
```

To enable the Hadoop open the **CMD as Administrator** and type below command  
**start-all.cmd**

It will open 4 new windows cmd terminals for 4 daemon processes, namely namenode, datanode, nodemanager, and resourcemanager.



The screenshot shows four overlapping Windows command prompt windows. The top window is titled 'Administrator: Apache Hadoop Distribution' and shows a Java exception: 'java.io.IOException: Cannot remove current directory: C:\B1gData\hadoop-2.9.1\data\namenode\current'. The second window is titled 'Apache Hadoop Distribution - hadoop - datanode' and shows a log message: '19/09/24 11:01:24 INFO ipc.Client: Retrying connect to server: 0.0.0.0/0.0.0.0:19000. Already tried 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)'. The third window is titled 'Apache Hadoop Distribution - yarn - resourcemanager' and shows a log message: 'Sep 24, 2019 11:00:54 AM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider INFO: Binding org.apache.hadoop.yarn.server.resourcemanager.webapp.RMWebServices to GuiceManagedComponentProvider with t...'. The bottom window is titled 'Administrator: Command Prompt' and shows the command 'start-all.cmd' being executed, with a message: 'This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd starting yarn daemons'. The prompt is 'C:\WINDOWS\system32>'.

Then you have successfully install the hadoop 2.9.1 on windows platform.

Now you can access all the Hadoop components via web urls.



To access Resource Manager go to <http://localhost:8088> from your web browser.

To access Node Manager go to <http://localhost:8042> from your web browser.

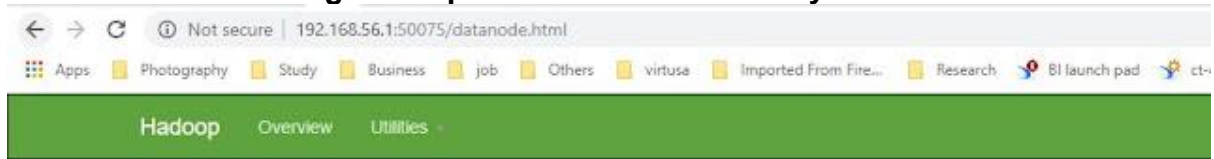
To access Name Node go to <http://localhost:50070> from your web browser.

## Overview '0.0.0.0:19000' (active)

Started:	Tue Sep 24 12:50:12 +0530 2019
Version:	2.9.1, re30710aea4e6e55e69372929106cf119af06fd0e
Compiled:	Mon Apr 16 15:03:00 +0530 2018 by root from branch-2.9.1
Cluster ID:	CID-dff52b8b-b137-4888-bdb8-982a44236747
Block Pool ID:	BP-1503339017-192.168.56.1-1569309564854

## Summary

To access Data Node go to <http://localhost:50075> from your web browser.



Open a new Windows Command Prompt and run below commands.

```
C:\hadoop-2.9.1\hadoop-2.9.1>cd bin
```

**ls:** This command is used to list all the files. Use `lsr` for recursive approach. It is useful when we want a hierarchy of a folder.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /
```

Found 2 items

```
drwxr-xr-x - LENOVO supergroup    0 2023-02-27 20:06 /sampledir
```

```
drwxr-xr-x - LENOVO supergroup    0 2023-02-27 19:57 /test
```

**mkdir:** To create a directory. In Hadoop `dfs` there is no home directory by default.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -mkdir /user
```

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -mkdir /user/Lenovo
```

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /
```

Found 3 items

```
drwxr-xr-x - LENOVO supergroup    0 2023-02-27 20:06 /sampledir
```

```
drwxr-xr-x - LENOVO supergroup    0 2023-02-27 19:57 /test
drwxr-xr-x - LENOVO supergroup    0 2023-02-27 20:23 /user
```

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -lsr /user

lsr: DEPRECATED: Please use 'ls -R' instead.

drwxr-xr-x - LENOVO supergroup    0 2023-02-27 20:23 /user/Lenovo
```

**touchz:** It creates an empty file.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -touchz /user/myfile.txt

C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls -R /user

drwxr-xr-x - LENOVO supergroup    0 2023-02-27 20:23 /user/Lenovo
-rw-r--r--  1 LENOVO supergroup    0 2023-02-27 20:26 /user/myfile.txt
```

**copyFromLocal (or) put:** To copy files/folders from local file system to hdfs store. This is the most important command. Local filesystem means the files present on the OS.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -put C:/hadoop-2.9.1/hadoop-2.9.1/Sample.txt /user

C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls -R /user

drwxr-xr-x - LENOVO supergroup    0 2023-02-27 20:23 /user/Lenovo
-rw-r--r--  1 LENOVO supergroup   12 2023-02-27 20:30 /user/Sample.txt
-rw-r--r--  1 LENOVO supergroup    0 2023-02-27 20:26 /user/myfile.txt
```

**cat:** To print file contents.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -cat /user/Sample.txt

Hello Hadoop
```

**copyToLocal (or) get:** To copy files/folders from hdfs store to local file system.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -get /user/Sample.txt ../HadoopExamples
```

**cp:** This command is used to copy files within hdfs.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -mkdir /user_copied
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -cp /user /user_copied
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /user_copied

Found 1 items

drwxr-xr-x - LENOVO supergroup      0 2023-02-27 20:48 /user_copied/user
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /user

Found 3 items

drwxr-xr-x - LENOVO supergroup      0 2023-02-27 20:23 /user/Lenovo
-rw-r--r--  1 LENOVO supergroup    12 2023-02-27 20:30 /user/Sample.txt
-rw-r--r--  1 LENOVO supergroup      0 2023-02-27 20:26 /user/myfile.txt
```

**mv:** This command is used to move files within hdfs.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -mv /user/myfile.txt /user_copied
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /user

Found 2 items

drwxr-xr-x - LENOVO supergroup      0 2023-02-27 20:23 /user/Lenovo
-rw-r--r--  1 LENOVO supergroup    12 2023-02-27 20:30 /user/Sample.txt
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /user_copied

Found 2 items

-rw-r--r--  1 LENOVO supergroup      0 2023-02-27 20:26 /user_copied/myfile.txt
drwxr-xr-x - LENOVO supergroup      0 2023-02-27 20:48 /user_copied/user
```

**rmr:** This command deletes a file from HDFS *recursively*. It is very useful command when you want to delete a *non-empty directory*.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -rmr /user_copied

rmr: DEPRECATED: Please use '-rm -r' instead.

Deleted /user_copied
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /user_copied

ls: `/user_copied': No such file or directory
```

**du:** It will give the size of each file in directory.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -du /user  
0 /user/Lenovo  
12 /user/Sample.txt
```

**dus:** This command will give the total size of directory/file.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -dus /user  
dus: DEPRECATED: Please use 'du -s' instead.  
12 /user
```

**stat:** It will give the last modified time of directory or path. In short it will give stats of the directory or file.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -stat /user  
2023-02-27 15:20:27
```

**setrep:** This command is used to change the replication factor of a file/directory in HDFS. By default it is 3 for anything which is stored in HDFS (as set in *hdfs core-site.xml*).

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /user  
Found 2 items  
drwxr-xr-x - LENOVO supergroup    0 2023-02-27 20:23 /user/Lenovo  
-rw-r--r--  6 LENOVO supergroup   12 2023-02-27 20:30 /user/Sample.txt
```