**Installing Hadoop 3.2.1 On Windows :**

**1.Prerequisites :**

First, we need to make sure that the following prerequisites are installed:

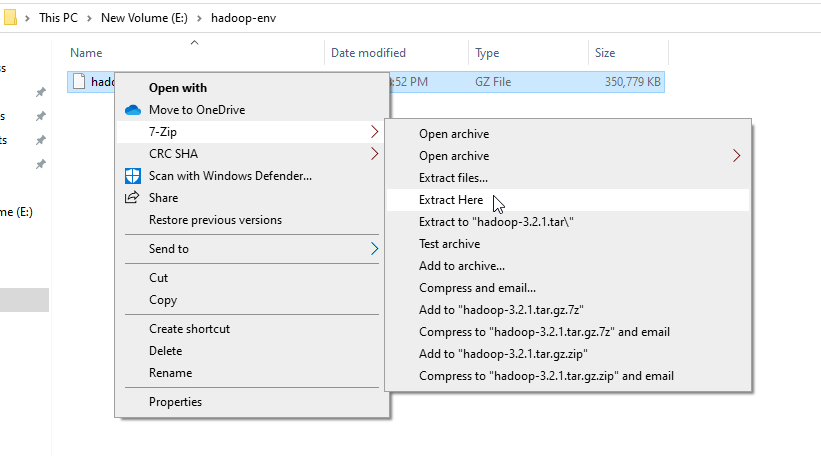
1. Java 8 runtime environment (JRE): [Hadoop 3 requires a Java 8 installation](https://cwiki.apache.org/confluence/display/HADOOP/Hadoop+Java+Versions). I prefer using the [offline installer](https://www.java.com/en/download/windows_offline.jsp).
2. [Java 8 development Kit (JDK)](https://www.oracle.com/java/technologies/javase-jdk8-downloads.html)
3. To unzip downloaded Hadoop binaries, we should install [7zip](https://www.7-zip.org/download.html).
4. I will create a folder “E:\hadoop-env” on my local machine to store downloaded files.

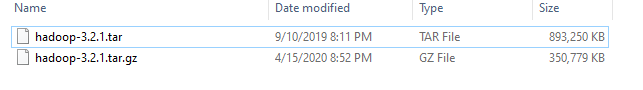
**2. Download Hadoop binaries :**

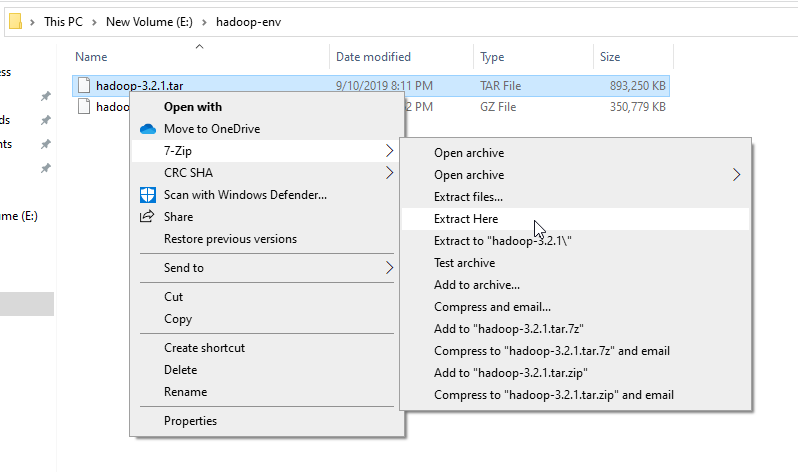
The first step is to download Hadoop binaries from the [official website](https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.2.1/hadoop-3.2.1.tar.gz). The binary package size is about 342 MB.



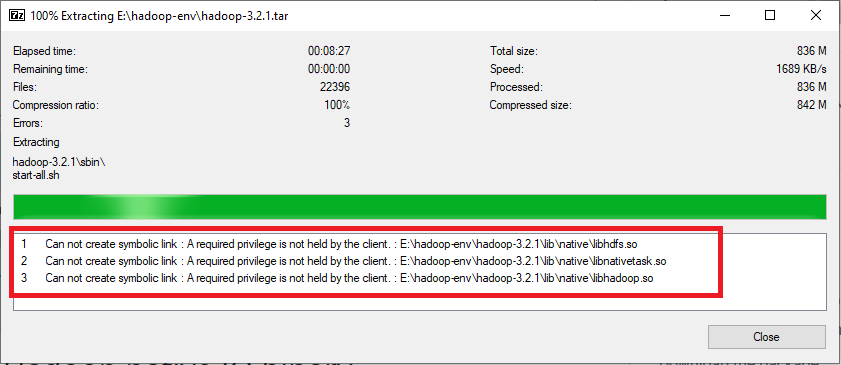
After finishing the file download, we should unpack the package using 7zip int two steps. First, we should extract the hadoop-3.2.1.tar.gz library, and then, we should unpack the extracted tar file:







The tar file extraction may take some minutes to finish. In the end, you may see some warnings about symbolic link creation. Just ignore these warnings since they are not related to windows.



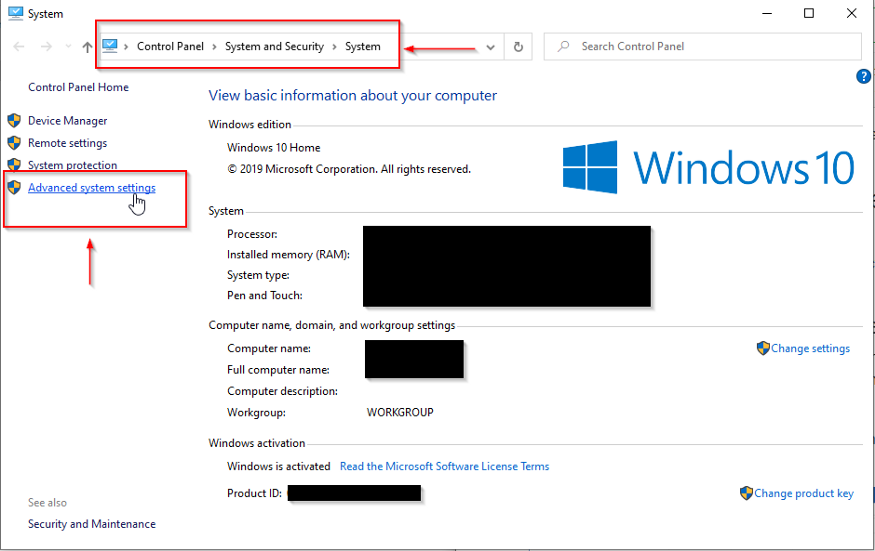
After unpacking the package, we should add the Hadoop native IO libraries, which can be found in the following GitHub repository: <https://github.com/cdarlint/winutils>.

Since we are installing Hadoop 3.2.1, we should download the files located in <https://github.com/cdarlint/winutils/tree/master/hadoop-3.2.1/bin> and copy them into the “hadoop-3.2.1\bin” directory.

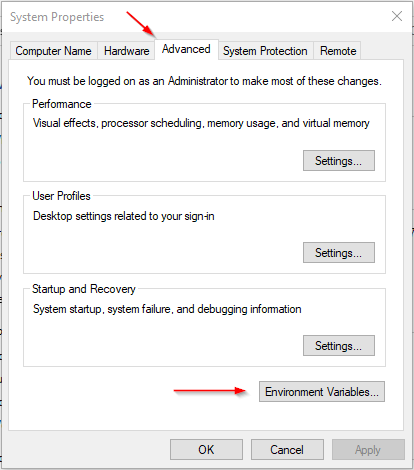
# 3. Setting up environment variables

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

To edit environment variables, go to Control Panel > System and Security > System (or right-click > properties on 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.



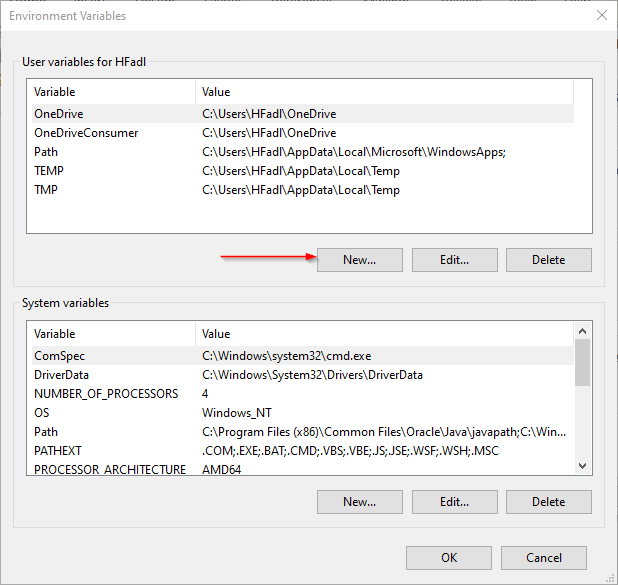
In the “Environment Variables” dialog, press the “New” button to add a new variable.

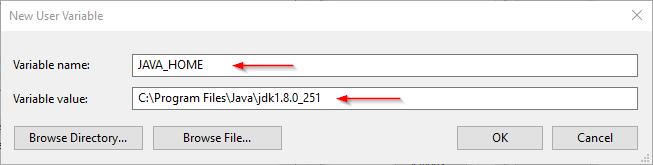
*Note: In this guide, we will add user variables since we are configuring Hadoop for a single user. If you are looking to configure Hadoop for multiple users, you can define System variables instead.*

There are two variables to define:

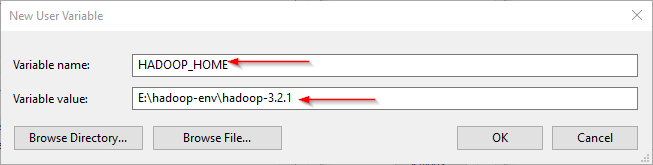
JAVA\_HOME: JDK installation folder path

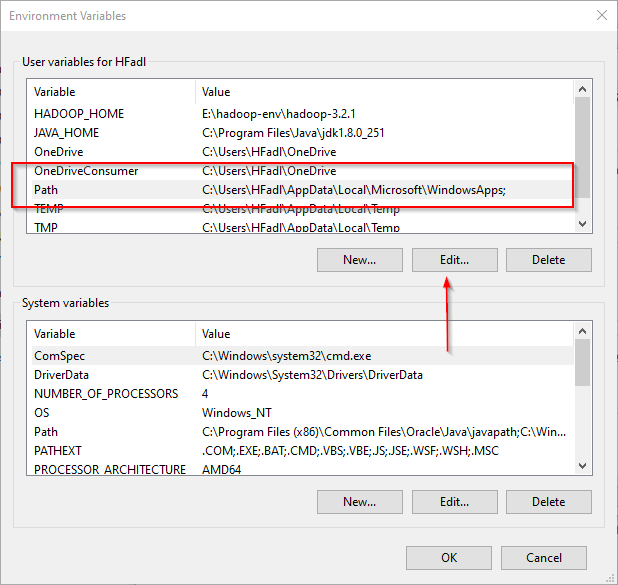
HADOOP\_HOME: Hadoop installation folder path

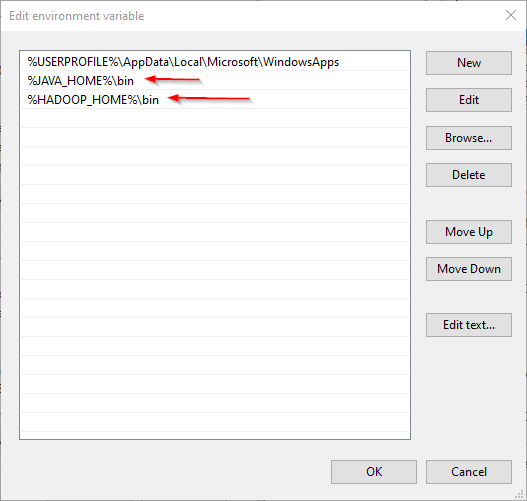




Now, we should edit the PATH variable to add the Java and Hadoop binaries paths as shown in the following screenshots.







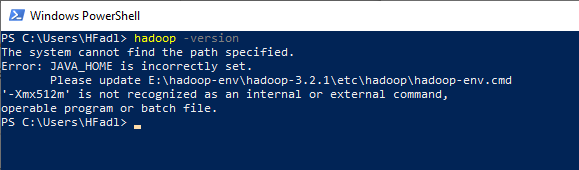
## **3.1. JAVA\_HOME is incorrectly set error**

Now, let’s open PowerShell and try to run the following command:

hadoop -version

In this example, since the JAVA\_HOME path contains spaces, I received the following error:

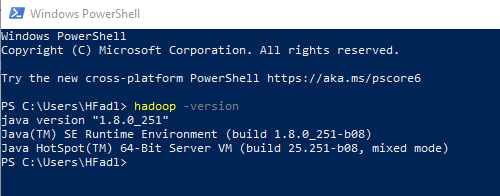
JAVA\_HOME is incorrectly set



To solve this issue, we should use the windows 8.3 path instead. As an example:

* Use “Progra~1” instead of “Program Files”
* Use “Progra~2” instead of “Program Files(x86)”

After replacing “Program Files” with “Progra~1”, we closed and reopened PowerShell and tried the same command. As shown in the screenshot below, it runs without errors.



# 4. Configuring Hadoop cluster

There are four files we should alter to configure Hadoop cluster:

1. %HADOOP\_HOME%\etc\hadoop\hdfs-site.xml
2. %HADOOP\_HOME%\etc\hadoop\core-site.xml
3. %HADOOP\_HOME%\etc\hadoop\mapred-site.xml
4. %HADOOP\_HOME%\etc\hadoop\yarn-site.xml

## ***4.1. HDFS site configuration***

As we know, Hadoop is built using a master-slave paradigm. Before altering the HDFS configuration file, we should create a directory to store all master node (name node) data and another one to store data (data node). In this example, we created the following directories:

* E:\hadoop-env\hadoop-3.2.1\data\dfs\namenode
* E:\hadoop-env\hadoop-3.2.1\data\dfs\datanode

Now, let’s open “hdfs-site.xml” file located in “%HADOOP\_HOME%\etc\hadoop” directory, and we should 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:///E:/hadoop-env/hadoop-3.2.1/data/dfs/namenode</value>

</property>

<property>

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

<value>file:///E:/hadoop-env/hadoop-3.2.1/data/dfs/datanode</value>

</property>

**4.2 Core site configuration**

<property>

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

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

</property>

**4.3 Map Reduce site configuration**

<property>

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

<value>yarn</value>

<description>MapReduce framework name</description>

</property>

**4.4 Yarn site configuration**

<property>

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

<value>mapreduce\_shuffle</value>

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

</property>

# 5. Formatting Name node

After finishing the configuration, let’s try to format the name node using the following command:

hdfs namenode -format

Due to a [bug in the Hadoop 3.2.1 release](https://issues.apache.org/jira/browse/HDFS-14890), you will receive the following error:

2020–04–17 22:04:01,503 ERROR namenode.NameNode: Failed to start namenode.

java.lang.UnsupportedOperationExceptionat

java.nio.file.Files.setPosixFilePermissions(Files.java:2044)

at org.apache.hadoop.hdfs.server.common.Storage$StorageDirectory.clearDirectory(Storage.java:452)

at org.apache.hadoop.hdfs.server.namenode.NNStorage.format(NNStorage.java:591)

at org.apache.hadoop.hdfs.server.namenode.NNStorage.format(NNStorage.java:613)

at org.apache.hadoop.hdfs.server.namenode.FSImage.format(FSImage.java:188)

at org.apache.hadoop.hdfs.server.namenode.NameNode.format(NameNode.java:1206)

at org.apache.hadoop.hdfs.server.namenode.NameNode.createNameNode(NameNode.java:1649)

at org.apache.hadoop.hdfs.server.namenode.NameNode.main(NameNode.java:1759)

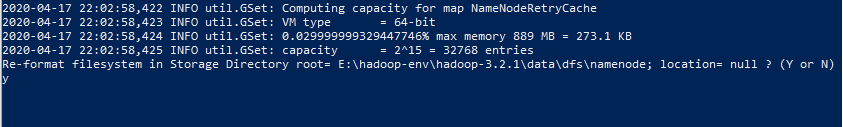
2020–04–17 22:04:01,511 INFO util.ExitUtil: Exiting with status 1: java.lang.UnsupportedOperationException

2020–04–17 22:04:01,518 INFO namenode.NameNode: SHUTDOWN\_MSG:

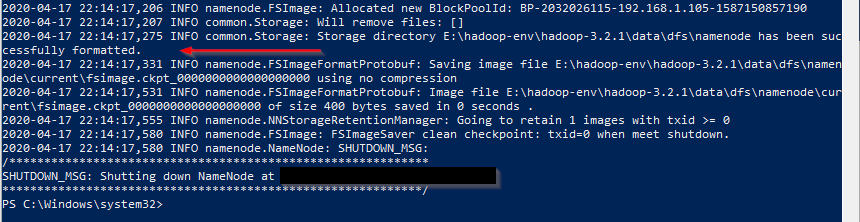
This issue will be solved within the next release. For now, you can fix it temporarily using the following steps ([reference](https://kontext.tech/column/hadoop/377/latest-hadoop-321-installation-on-windows-10-step-by-step-guide)):

1. Download hadoop-hdfs-3.2.1.jar file from the [following link](https://github.com/FahaoTang/big-data/blob/master/hadoop-hdfs-3.2.1.jar).
2. Rename the file name hadoop-hdfs-3.2.1.jar to hadoop-hdfs-3.2.1.bak in folder %HADOOP\_HOME%\share\hadoop\hdfs
3. Copy the downloaded hadoop-hdfs-3.2.1.jar to folder %HADOOP\_HOME%\share\hadoop\hdfs

Now, if we try to re-execute the format command (Run the command prompt or PowerShell as administrator), you need to approve file system format.



And the command is executed successfully:



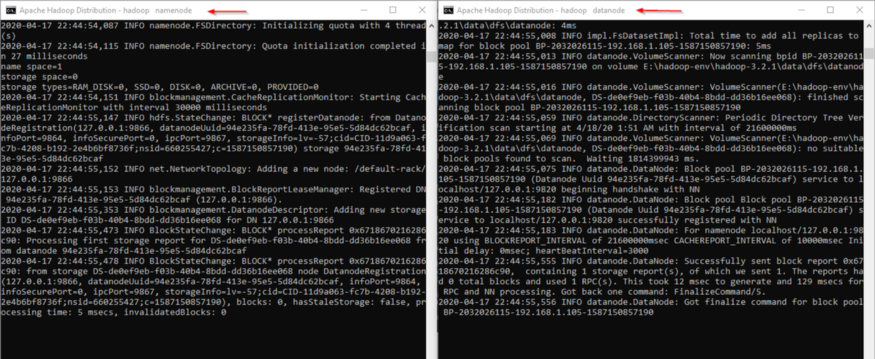
# 6. Starting Hadoop services

Now, we will open PowerShell, and navigate to “%HADOOP\_HOME%\sbin” directory. Then we will run the following command to start the Hadoop nodes:

.\start-dfs.cmd



Two command prompt windows will open (one for the name node and one for the data node) as follows:

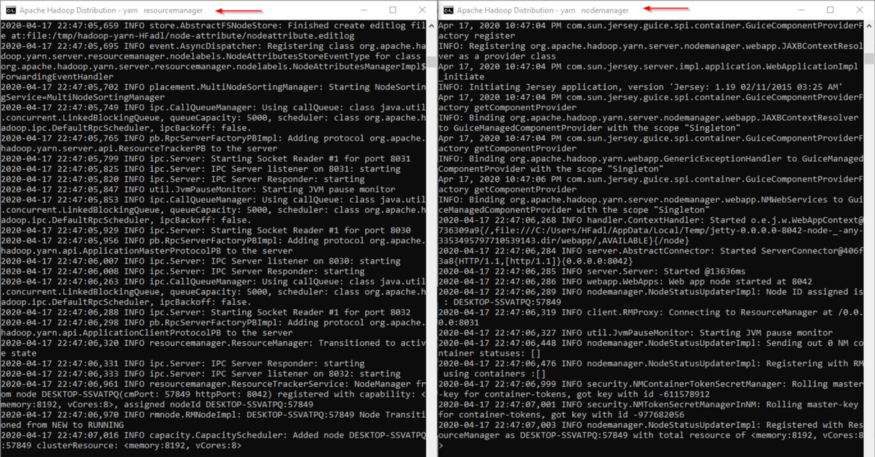


Next, we must start the Hadoop Yarn service using the following command:

./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, we can run the following command:

jps

It should display the following services:

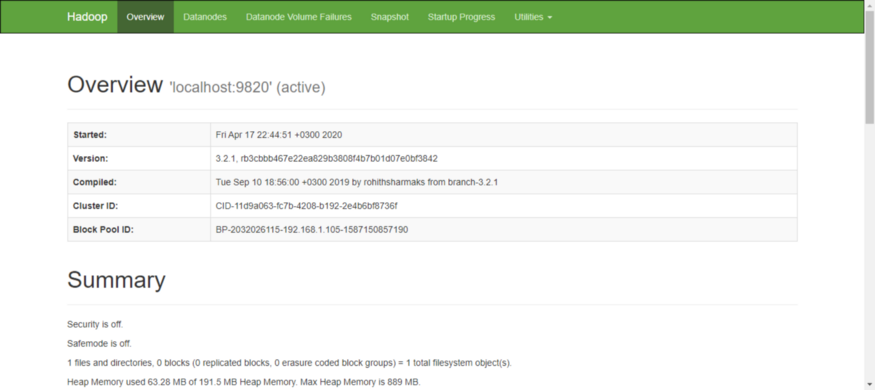
14560 DataNode  
4960 ResourceManager  
5936 NameNode  
768 NodeManager  
14636 Jps



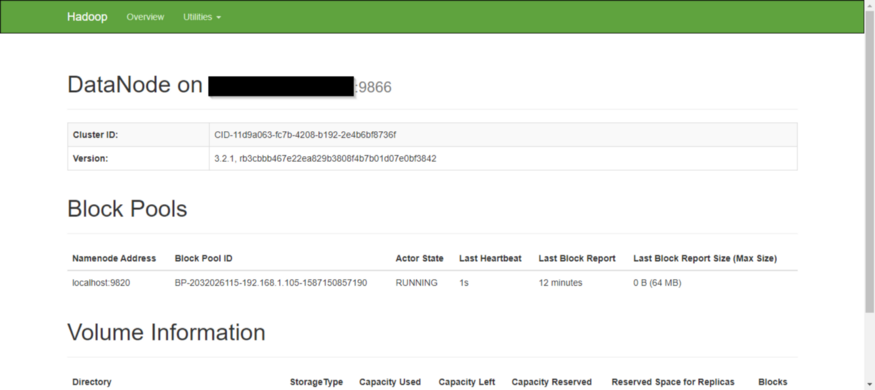
# 7. Hadoop Web UI

There are three web user interfaces to be used:

* Name node web page: <http://localhost:9870/dfshealth.html>



* Data node web page: <http://localhost:9864/datanode.html>



* Yarn web page: <http://localhost:8088/cluster>

