An easy to go guide for installing the Hadoop in Windows 10



Image taken from Google images

**1. Prerequisites**

1. Hardware Requirement  
   \* RAM — Min. 8GB, if you have SSD in your system then 4GB RAM would also work.  
   \* CPU — Min. Quad core, with at least 1.80GHz
2. JRE 1.8 — Offline installer for JRE
3. Java Development Kit — 1.8
4. A Software for Un-Zipping like 7Zip or Win Rar  
   \* I will be using a 64-bit windows for the process, please check and download the version supported by your system x86 or x64 for all the software.
5. Download Hadoop zip  
   \* I am using Hadoop-2.9.2, you can use any other STABLE version for hadoop.

A screenshot of a computer

Description automatically generated

Fig. 1:- Download Hadoop 2.9.2

Once we have Downloaded all the above software, we can proceed with next steps in installing the Hadoop.

**2. Unzip and Install Hadoop**

After Downloading the Hadoop, we need to Unzip the hadoop-2.9.2.tar.gz file.

A screenshot of a computer

Description automatically generated

Fig. 2:- Extracting Hadoop Step-1

Once extracted, we would get a new file hadoop-2.9.2.tar.  
Now, once again we need to extract this tar file.

A screenshot of a computer

Description automatically generated

Fig. 3:- Extracting Hadoop Step-2

* Now we can organize our Hadoop installation, we can create a folder and move the final extracted file in it. For Eg. :-

A screenshot of a computer

Description automatically generated

Fig. 4:- Hadoop Directory

* Please note while creating folders, DO NOT ADD SPACES IN BETWEEN THE FOLDER NAME.(it can cause issues later)
* I have placed my Hadoop in D: drive you can use C: or any other drive also.

**3. Setting Up Environment Variables**

Another important step in setting up a work environment is to set your Systems environment variable.

To edit environment variables, go to Control Panel > System > click on the “Advanced system settings” link  
Alternatively, We can Right click on This PC icon and click on Properties and click on the “Advanced system settings” link  
Or, easiest way is to search for Environment Variable in search bar and there you GO…😉

A screenshot of a computer

Description automatically generated

Fig. 5:- Path for Environment Variable

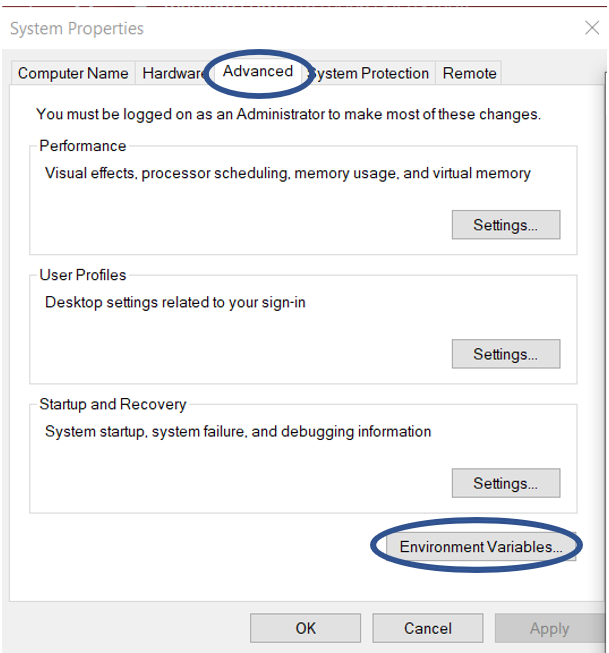


Fig. 6:- Advanced System Settings Screen

**3.1 Setting JAVA\_HOME**

* Open environment Variable and click on “New” in “User Variable”

A screenshot of a computer

Description automatically generated

Fig. 7:- Adding Environment Variable

* On clicking “New”, we get below screen.

A screenshot of a computer

Description automatically generated

Fig. 8:- Adding JAVA\_HOME

* Now as shown, add JAVA\_HOME in variable name and path of Java(jdk) in Variable Value.
* Click OK and we are half done with setting JAVA\_HOME.

**3.2 Setting HADOOP\_HOME**

* Open environment Variable and click on “New” in “User Variable”

A screenshot of a computer

Description automatically generated

Fig. 9:- Adding Environment Variable

* On clicking “New”, we get below screen.

A screenshot of a computer

Description automatically generated

Fig. 10:- Adding HADOOP\_HOME

* Now as shown, add HADOOP\_HOME in variable name and path of Hadoop folder in Variable Value.
* Click OK and we are half done with setting HADOOP\_HOME.

*Note:- If you want the path to be set for all users you need to select “New” from System Variables.*

**3.3 Setting Path Variable**

* Last step in setting Environment variable is setting Path in System Variable.

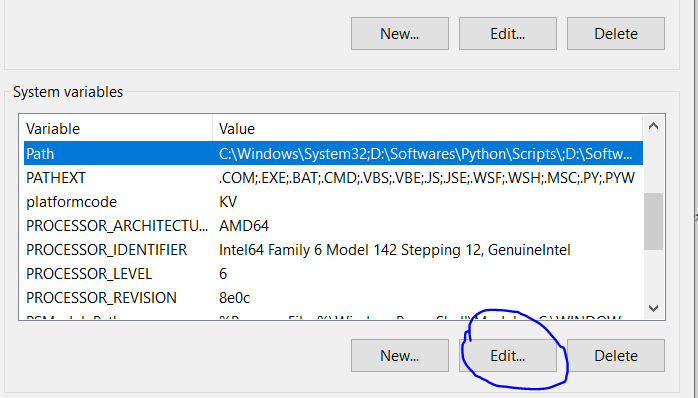


Fig. 11:- Setting Path Variable

* Select Path variable in the system variables and click on “Edit”.

A screenshot of a computer

Description automatically generated

Fig. 12:- Adding Path

* Now we need to add these paths to Path Variable one by one:-  
  \* %JAVA\_HOME%\bin  
  \* %HADOOP\_HOME%\bin  
  \* %HADOOP\_HOME%\sbin
* Click OK and OK. & we are done with Setting Environment Variables.

**3.4 Verify the Paths**

* Now we need to verify that what we have done is correct and reflecting.
* Open a **NEW** Command Window
* Run following commands

echo %JAVA\_HOME%   
echo %HADOOP\_HOME%  
echo %PATH%

**4. Editing Hadoop files**

Once we have configured the environment variables next step is to configure Hadoop. It has 3 parts:-

**4.1 Creating Folders**

We need to create a folder data in the hadoop directory, and 2 sub folders namenode and datanode

A screenshot of a computer

Description automatically generated

Fig. 13:- Creating Data Folder

* Create **DATA folder**in the Hadoop directory

A screenshot of a computer

Description automatically generated

Fig. 14:- Creating Sub-folders

* Once DATA folder is created, we need to create 2 new folders namely, **namenode and datanode**inside the data folder
* These folders are important because files on HDFS resides inside the datanode.

**4.2 Editing Configuration Files**

Now we need to edit the following config files in hadoop for configuring it :-  
(We can find these files in Hadoop -> etc -> hadoop)

\* core-site.xml  
\* hdfs-site.xml  
\* mapred-site.xml  
\* yarn-site.xml  
\* hadoop-env.cmd

**4.2.1 Editing core-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>fs.defaultFS</name>  
 <value>hdfs://localhost:9000</value>  
 </property>  
</configuration>

**4.2.2 Editing hdfs-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.  
Also replace PATH~1 and PATH~2 with the path of namenode and datanode folder that we created recently(step 4.1).***

<configuration>  
 <property>  
 <name>dfs.replication</name>  
 <value>1</value>  
 </property>  
 <property>  
 <name>dfs.namenode.name.dir</name>  
 <value>PATH~1\namenode</value>  
 <final>true</final>  
 </property>  
 <property>  
 <name>dfs.datanode.data.dir</name>  
 <value>PATH~2\datanode</value>  
 <final>true</final>  
 </property>  
</configuration>

**4.2.3 Editing mapred-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>mapreduce.framework.name</name>  
 <value>yarn</value>  
 </property>  
</configuration>

**4.2.4 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>

**4.2.5 Verifying hadoop-env.cmd**Right click on the file, select edit and check if the JAVA\_HOME is set correctly or not.  
We can replace the JAVA\_HOME variable in the file with your actual JAVA\_HOME that we configured in the System Variable.

set JAVA\_HOME=%JAVA\_HOME%  
 OR  
set JAVA\_HOME="C:\Program Files\Java\jdk1.8.0\_221"

**4.3 Replacing bin**

Last step in configuring the hadoop is to download and replace the bin folder.  
\* Go to this GitHub Repo and download the bin folder as a zip.  
\* Extract the zip and copy all the files present under bin folder to %HADOOP\_HOME%\bin

***Note:- If you are using different version of Hadoop then please search for its respective bin folder and download it.***

**5. Testing Setup**

Congratulation..!!!!!  
We are done with the setting up the Hadoop in our System.

Now we need to check if everything works smoothly…

**5.1 Formatting Namenode**

Before starting hadoop we need to format the namenode for this we need to start a NEW Command Prompt and run below command

hadoop namenode -format

A screen shot of a computer

Description automatically generated

Fig. 15:- Formatting Namenode

***Note:- This command formats all the data in namenode. So, its advisable to use only at the start and do not use it every time while starting hadoop cluster to avoid data loss.***

**5.2 Launching Hadoop**

Now we need to start a new Command Prompt remember to run it as administrator to avoid permission issues and execute below commands

***start-all.cmd***

A screenshot of a computer program

Description automatically generated

Fig. 16:- start-all.cmd

This will open 4 new cmd windows running 4 different Daemons of hadoop:-  
\* Namenode  
\* Datanode  
\* Resourcemanager  
\* Nodemanager

A screenshot of a computer

Description automatically generated

Fig. 17:- Hadoop Deamons

***Note:- We can verify if all the daemons are up and running using jps command in new cmd window.***

**6. Running Hadoop (Verifying Web UIs)**

**6.1 Namenode**  
Open localhost:50070 in a browser tab to verify namenode health.

A screenshot of a computer

Description automatically generated

Fig. 18:- Namenode Web UI

**6.2 Resourcemanger**Open localhost:8088 in a browser tab to check resourcemanager details.

A screenshot of a computer

Description automatically generated

Fig. 19:- Resourcemanager Web UI

**6.3 Datanode**Open localhost:50075 in a browser tab to checkout datanode.

A screenshot of a computer

Description automatically generated

Fig. 20:- Datanode Web UI