## Steps to Install Hadoop 3 on Ubuntu

### Prerequisites

First, download the Hadoop 3.1.2 from the below link:

[Hadoop 3.1.2](https://archive.apache.org/dist/hadoop/common/hadoop-3.1.2/hadoop-3.1.2.tar.gz)

Here are the steps for installing Hadoop 3 on ubuntu for your system:

**Step 1:** Install ssh on your system using the below command:

***sudo apt-get install ssh***

Type the password for the sudo user and then press **Enter**.

Type **‘Y’** and then press **Enter** to continue with the installation process.

**Step 2:** Install pdsh on your system using the below command:

***sudo apt-get install pdsh***

Type **‘Y’** and then press **Enter** to continue with the installation process.

**Step 3:** Open the .bashrc file in the nano editor using the following command:

***nano .bashrc***

Now set the **PDSH\_RCMD\_TYPE** environment variable to **ssh**

***export PDSH\_RCMD\_TYPE=ssh***

To save the changes you’ve made, press **Ctrl+O**. To exit the nano editor, press **Ctrl+X** and then press **‘Y’** to exit the editor.

**Step 4:** Now configure ssh. To do so, create a new key with the help of the following command (don’t copy-paste following command, rather type):

***ssh-keygen -t rsa -P ""***

Press **Enter** when asked the **file name**.

**Step 5:** Copy the content of the public key to authorized\_keys.

***cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys***

**Step 6:** Now examine the SSH setup by connecting to the localhost.

***ssh localhost***

Type **‘Y’** and then press **Enter** to continue with the connection.

**Step 7:** Update the source lists.

***sudo apt-get update***

**Step 8:** Now install Java 8 using the following command:

***sudo apt-get install openjdk-8-jdk***

Type **‘Y’** and then press **Enter** to finish with the installation process.

**Step 9:** To cross-check whether you have successfully installed Java on your machine or not, run the below command:

***java -version***

Please Download Hadoop from the link given in the first section and copy the setup in your home directory (/home/USER-NAME)

### Hadoop Installation on Ubuntu

**Step 10:** Now locate the Hadoop tar file in your system.

**Step 11:** Extract the **hadoop-3.1.2.tar.gz** file using the below command:

***tar xzf hadoop-3.1.2.tar.gz***

**Step 12:** Rename **hadoop-3.1.2.tar.gz** as **hadoop** for ease of use.

***mv hadoop-3.1.2.tar.gz hadoop***

**Step 14:** Now check the Java home path

***ls /usr/lib/jvm/java-8-openjdk-amd64/***

**Step 15:** Open the **hadoop-env.sh** file in the nano editor. This file is located in **~/hadoop/etc/hadoop** (Hadoop configuration directory).

***nano hadoop-env***

Now, Set JAVA\_HOME path:

***export JAVA\_HOME=<path-to-the-root-of-your-Java-installation> (eg: /usr/lib/jvm/java-8-openjdk-amd64/)***

To save the changes you’ve made, press **Ctrl+O**. To exit the nano editor, press **Ctrl+X** and then press **‘Y’** to exit the editor.

**Step 16:** Open the **core-site.xml** file in the nano editor. This file is also located in the **~/hadoop/etc/hadoop** (Hadoop configuration directory).

***nano core-site.xml***

Add the following configuration properties:

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

</property>

<property>

<name>hadoop.tmp.dir</name>

<value>/home/ibstraining/hdata</value>

</property>

</configuration>

**Step 17:** Open the **hdfs-site.xml** file in the nano editor. This file is also located in **~/hadoop/etc/hadoop** (Hadoop configuration directory):

***nano hdfs-site.xml***

Add the following entries in core-site.xml:

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

</configuration>

**Step 18:** Open the **mapred-site.xml** file in the nano editor. This file is also located in **~/hadoop/etc/hadoop** (Hadoop configuration directory).

***nano mapred-site.xml***

Add the following entries in core-site.html:

<configuration>

<property>

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

<value>yarn</value>

</property>

<property>

<name>yarn.app.mapreduce.am.env</name>

<value>HADOOP\_MAPRED\_HOME=/home/ibstraining/hadoop</value>

</property>

<property>

<name>mapreduce.map.env</name>

<value>HADOOP\_MAPRED\_HOME=/home/ibstraining/hadoop</value>

</property>

<property>

<name>mapreduce.reduce.env</name>

<value>HADOOP\_MAPRED\_HOME=/home/ibstraining/hadoop</value>

</property>

</configuration>

**Step 19:** Open the **yarn-site.xml** file in the nano editor. This file is also located in **~/hadoop/etc/hadoop** (Hadoop configuration directory).

nano yarn-site.xml

Add the following entries in the yarn-site.xml:

<configuration>

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

</configuration>

**Step 20:** Open the bashrc files in the nano editor using the following command:

***nano .bashrc***

Edit .bashrc file located in the user’s home directory and add the following parameters:

export HADOOP\_HOME="/home/ibstraining/hadoop"

export PATH=$PATH:$HADOOP\_HOME/bin

export PATH=$PATH:$HADOOP\_HOME/sbin

export HADOOP\_MAPRED\_HOME=${HADOOP\_HOME}

export HADOOP\_COMMON\_HOME=${HADOOP\_HOME}

export HADOOP\_HDFS\_HOME=${HADOOP\_HOME}

export YARN\_HOME=${HADOOP\_HOME}

To save the changes you’ve made, press **Ctrl+O**. To exit the nano editor, press **Ctrl+X** and then press **‘Y’** to exit the editor.

Now, source the bashrc file so that the changes will come into effect:

***source ~/.bashrc***

**Step 21:** Before starting Hadoop, we need to format HDFS, which can be done using the below command:

***hdfs namenode -format***

**Step 22:** Start the HDFS services:

***sbin/start-dfs.sh***

**Step 23:** Open the HDFS web console:

***localhost:9870***

**Step 24:** Now start the yarn services:

***sbin/start-yarn.sh***

$***jps***

NameNode

DataNode

ResourceManager

NodeManager

SecondaryNameNode

The **‘jps’** command is used to check whether all the Hadoop processes are running or not.

**Step 25:** Open the yarn web console:

***localhost:8088***