**Apache Hadoop 3.3.6 Installation on Ubuntu 22.04**

**Step 1 : Install Java Development Kit**

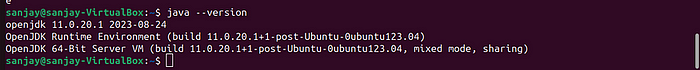
The default Ubuntu repositories contain Java 8 and Java 11 both. I am using Java 8 because hive only works on this version.Use the following command to install it.

sudo apt update && sudo apt install openjdk-8-jdk

**Step 2 : Verify the Java version :**

Once you have successfully installed it, check the current Java version:

java -version



**Step 3 : Install SSH :**

SSH (Secure Shell) installation is vital for Hadoop as it enables secure communication between nodes in the Hadoop cluster. This ensures data integrity, confidentiality, and allows for efficient distributed processing of data across the cluster.

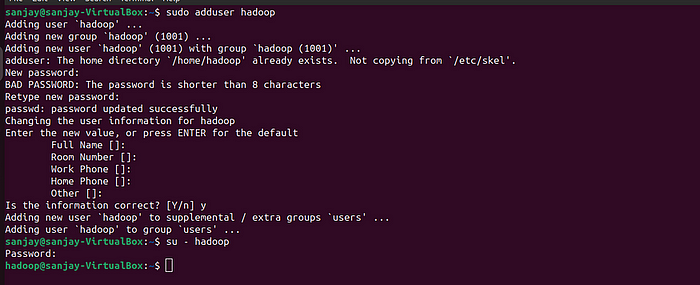
sudo apt install ssh

**Step 4 : Create the hadoop user :**

All the Hadoop components will run as the user that you create for Apache Hadoop, and the user will also be used for logging in to Hadoop’s web interface.

Run the command to create user and set password :

sudo adduser hadoop



**Step 5 : Switch user :**

Switch to the newly created hadoop user:

su - hadoop

**Step 6 : Configure SSH :**

Now configure password-less SSH access for the newly created hadoop user, so I didn’t enter key to save file and passpharse. Generate an SSH keypair first:

ssh-keygen -t rsa

A computer screen shot of a computer

Description automatically generated

**Step 7 : Set permissions :**

Copy the generated public key to the authorized key file and set the proper permissions:

cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys   
chmod 640 ~/.ssh/authorized\_keys

**Step 8 : SSH to the localhost**

ssh localhost

You will be asked to authenticate hosts by adding RSA keys to known hosts. Type yes and hit Enter to authenticate the localhost.

A computer screen shot of a computer program

Description automatically generated

**Step 9 : Switch user**

Again switch to hadoop

su - hadoop

**Step 10 : Install hadoop**

* Download hadoop 3.3.6

wget https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz

* Once you’ve downloaded the file, you can unzip it to a folder.

tar -xvzf hadoop-3.3.6.tar.gz

* Rename the extracted folder to remove version information. This is an optional step, but if you don’t want to rename, then adjust the remaining configuration paths.

mv hadoop-3.3.6 hadoop

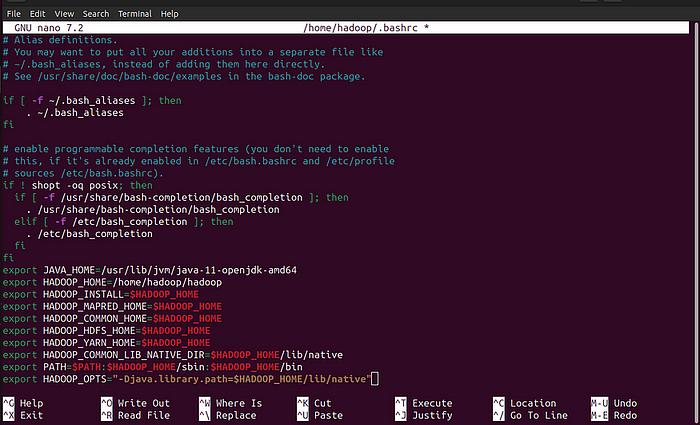


* Next, you will need to configure Hadoop and Java Environment Variables on your system. Open the ~/.bashrc file in your favorite text editor.Here I am using nano editior , to pasting the code we use ctrl+shift+v for saving the file ctrl+x and ctrl+y ,then hit enter:

nano ~/.bashrc

* Append the below lines to the file.

export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64  
export HADOOP\_HOME=/home/hadoop/hadoop  
export HADOOP\_INSTALL=$HADOOP\_HOME  
export HADOOP\_MAPRED\_HOME=$HADOOP\_HOME  
export HADOOP\_COMMON\_HOME=$HADOOP\_HOME  
export HADOOP\_HDFS\_HOME=$HADOOP\_HOME  
export HADOOP\_YARN\_HOME=$HADOOP\_HOME  
export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_HOME/lib/native  
export PATH=$PATH:$HADOOP\_HOME/sbin:$HADOOP\_HOME/bin  
export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_HOME/lib/native"



* Load the above configuration in the current environment.

source ~/.bashrc

* You also need to configure JAVA\_HOME in hadoop-env.sh file. Edit the Hadoop environment variable file in the text editor:

nano $HADOOP\_HOME/etc/hadoop/hadoop-env.sh

Search for the “export JAVA\_HOME” and configure it .

JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64

A screenshot of a computer

Description automatically generated

**Step 11 : Configuring Hadoop :**

* First, you will need to create the **namenode** and **datanode** directories inside the Hadoop user home directory. Run the following command to create both directories:

cd hadoop/

mkdir -p ~/hadoopdata/hdfs/{namenode,datanode}



* Next, edit the core-site.xml file and update with your system hostname:

nano $HADOOP\_HOME/etc/hadoop/core-site.xml

Change the following name as per your system hostname:

<configuration>  
 <property>  
 <name>fs.defaultFS</name>  
 <value>hdfs://localhost:9000</value>  
 </property>  
</configuration>

A screenshot of a computer

Description automatically generated

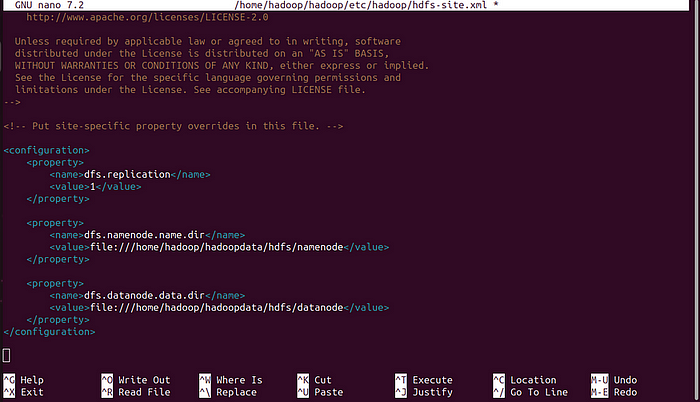
Save and close the file.

* Then, edit the hdfs-site.xml file:

nano $HADOOP\_HOME/etc/hadoop/hdfs-site.xml

* Change the NameNode and DataNode directory paths as shown below:

<configuration>  
 <property>  
 <name>dfs.replication</name>  
 <value>1</value>  
 </property>  
 <property>  
 <name>dfs.namenode.name.dir</name>  
 <value>file:///home/hadoop/hadoopdata/hdfs/namenode</value>  
 </property>  
 <property>  
 <name>dfs.datanode.data.dir</name>  
 <value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>  
 </property>  
 </configuration>

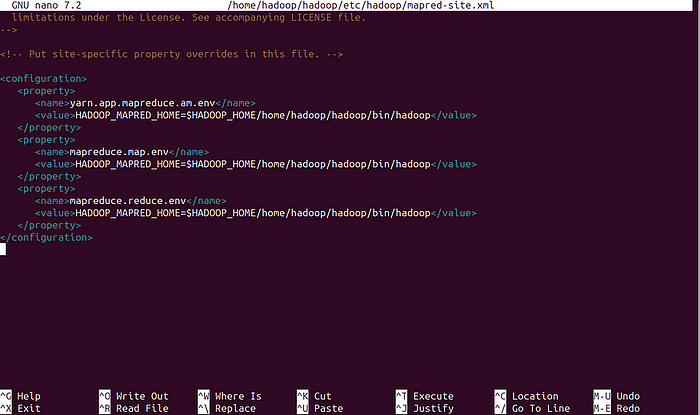


* Then, edit the mapred-site.xml file:

nano $HADOOP\_HOME/etc/hadoop/mapred-site.xml

* Make the following changes:

<configuration>  
 <property>  
 <name>yarn.app.mapreduce.am.env</name>  
 <value>HADOOP\_MAPRED\_HOME=$HADOOP\_HOME/home/hadoop/hadoop/bin/hadoop</value>  
 </property>  
 <property>  
 <name>mapreduce.map.env</name>  
 <value>HADOOP\_MAPRED\_HOME=$HADOOP\_HOME/home/hadoop/hadoop/bin/hadoop</value>  
 </property>  
 <property>  
 <name>mapreduce.reduce.env</name>  
 <value>HADOOP\_MAPRED\_HOME=$HADOOP\_HOME/home/hadoop/hadoop/bin/hadoop</value>  
 </property>  
</configuration>

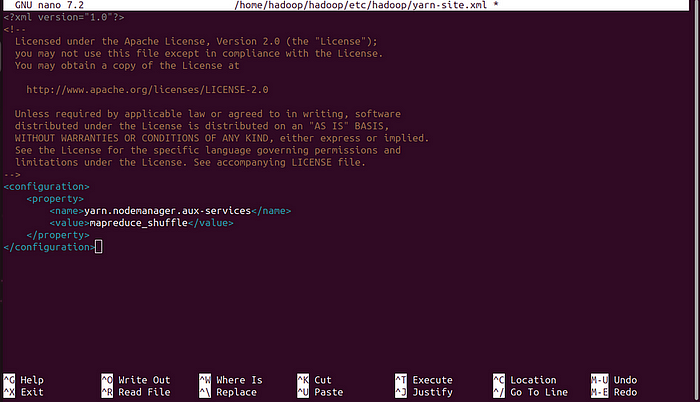


* Then, edit the yarn-site.xml file:

nano $HADOOP\_HOME/etc/hadoop/yarn-site.xml

* Make the following changes:

<configuration>  
 <property>  
 <name>yarn.nodemanager.aux-services</name>  
 <value>mapreduce\_shuffle</value>  
 </property>  
</configuration>



Save the file and close it .

**Step 12 : Start Hadoop cluster:**

* Before starting the Hadoop cluster. You will need to format the **Namenode** as a hadoop user.
* Run the following command to format the Hadoop **Namenode**:

hdfs namenode -format

* Once the **namenode** directory is successfully formatted with hdfs file system, you will see the message “Storage directory /home/hadoop/hadoopdata/hdfs/namenode has been successfully formatted”.

A screenshot of a computer

Description automatically generated

* Then start the Hadoop cluster with the following command.

start-all.sh

A screenshot of a computer program

Description automatically generated

* You can now check the status of all Hadoop services using the jps command:

jps

A black and purple background

Description automatically generated with medium confidence

**Step 13 : Access Hadoop Namenode and Resource Manager :**

* First we need to know our ip address,In Ubuntu we need to install net-tools to run ipconfig command, If you installing net-tools for the first time switch to default user :

sudo apt install net-tools

* Then run **ifconfig** command to know our ip address:

ifconfig

A computer screen shot of white text

Description automatically generated

Here my ip address is 192.168.1.6.

* To access the **Namenode**, open your web browser and visit the URL [http://your-server-ip:9870.](http://your-server-ip:9870./) You should see the following screen:

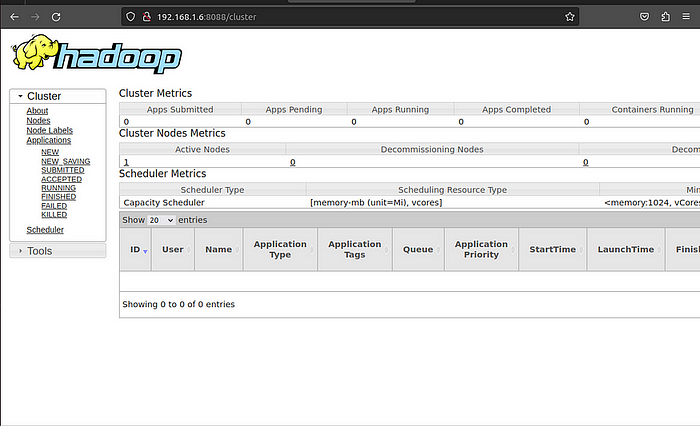
[**http://192.168.1.6:9870**](http://hadoop.tecadmin.net:9870/)

A screenshot of a computer

Description automatically generated

* To access Resource Manage, open your web browser and visit the URL [http://your-server-ip:8088.](http://your-server-ip:8088./) You should see the following screen:

[http://192.168.1.6:8088](http://hadoop.tecadmin.net:8088/)



**Step 13 :Verify the Hadoop Cluster :**

At this point, the Hadoop cluster is installed and configured. Next, we will create some directories in the HDFS filesystem to test the Hadoop.

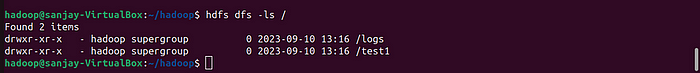
* Let’s create some directories in the HDFS filesystem using the following command:

hdfs dfs -mkdir /test1  
hdfs dfs -mkdir /logs

* Next, run the following command to list the above directory:

hdfs dfs -ls /

You should get the following output:



* Also, put some files to **hadoop** file system. For the example, putting log files from host machine to **hadoop** file system.

hdfs dfs -put /var/log/\* /logs/

You can also verify the above files and directory in the Hadoop web interface.

Go to the web interface, click on the Utilities => Browse the file system. You should see your directories which you have created earlier in the following screen:

A screenshot of a computer

Description automatically generated

**Step 14 : To stop hadoop services :**

To stop the Hadoop service, run the following command as a hadoop user:

stop-all.sh

A purple background with white text

Description automatically generated