**Procedure for installing Hadoop**

OS used: Ubuntu 14.04 LTS 64 bit

After Ubuntu installation with default user girish

Configure IP-4 address:

192.168.31.36 / 255.255.255.0

Check whether Computer is connected to Internet

**Step 1)**Installing Java Oracle jdk1.8

Press Ctrl+Alt t (Go to terminal window)

$sudo apt-get update

$sudo apt-get install python-software-properties

$sudo apt-add-repository ppa:webupd8team/java

$sudo apt-get update

$sudo apt-get install oracle-java8-installer

$ java -version

Java will be installed at /usr/lib/jvm/java-8-oracle <Installation Directory of JAVAJDK>

JAVA\_HOME=/usr/lib/jvm/java-8-oracle

**Step 2)**Add a hadoop system user using following commands:

Create group hadoop

girish@ubuntu:~$ sudo addgroup hadoop (girish is default user created while installing linux)

Create user hduser

girish@ubuntu:~$ sudo adduser --ingroup hadoop hduser

Full Name: Team

There is a possibility of below mentioned error in this setup and installation process.

**"hduser is not in the sudoers file. This incident will be reported."**

Login as a root user

girish@ubuntu:~$sudo adduser hduser sudo (for adding user hduser to group sudo)

Relogin as hduser

girish@ubuntu:~$su hduser

hduser@ubuntu:~$

**Step 3)** Configure SSH

In order to manage nodes in a cluster, Hadoop require SSH access

hduser@ubuntu:~$

create a new key

hduser@ubuntu:~$ssh-keygen -t rsa -P ""

Enable SSH access to local machine using this key.

hduser@ubuntu:~$cat $HOME/.ssh/id\_rsa.pub >> $HOME/.ssh/authorized\_keys

Now test SSH setup by connecting to locahost as 'hduser' user.

hduser@ubuntu:~$ssh localhost

Please note, if you see below error in response to 'ssh localhost', then there is a possibility that SSH is not available on this system-

**To resolve this -**

Purge SSH using,

sudo apt-get purge openssh-server

It is good practice to purge before start of installation

hduser@ubuntu:~$sudo apt-get purge openssh-server

Install SSH using command-

sudo apt-get install openssh-server openssh-client

hduser@ubuntu:~$sudo apt-get install openssh-server openssh-client

**Step 4)**Next step is to download Hadoop

Visit https://archive.apache.org/dist/hadoop/common/hadoop-2.7.2/

Download hadoop-2.7.2.tar.gz (Stable version) **Select the tar.gz file ( not the file with src)**

Once download is complete, navigate to the directory containing the tar file (By default it is downloaded in Download folder)

hduser@ubuntu:~$ sudo tar xzf hadoop-2.7.2.tar.gz

**Now, rename hadoop-2.7.2 as hadoop and move it to /usr/local**

hduser@ubuntu:~$sudo mv hadoop-2.7.2 /usr/local/hadoop

hduser@ubuntu:~$sudo chown -R hduser:hadoop hadoop

**Step 5)** Modify **~/.bashrc** file

Login as hduser

Add following lines to end of file **~/.bashrc**

#Set HADOOP\_HOME

export HADOOP\_HOME=/usr/local/hadoop <Installation Directory of Hadoop>

#Set JAVA\_HOME

export JAVA\_HOME= /usr/lib/jvm/java-8-oracle <Installation Directory of JAVAJDK>

# Add bin/ and sbin directory of Hadoop to PATH

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

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_HOME/lib/native  
export HADOOP\_OPTS=”-Djava.library.path=$HADOOP\_HOME/lib”

hduser@ubuntu:~$gedit ~/.bashrc **(This is example for reference)**

#Set HADOOP\_HOME

export HADOOP\_HOME=/usr/local/hadoop

#Set JAVA\_HOME

export JAVA\_HOME=/usr/lib/jvm/java-8-oracle

# Add bin/ and sbin directory of Hadoop to PATH

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

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_HOME/lib/native  
export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_HOME/lib"

Now, source this environment configuration using below command

**. ~/.bashrc**

hduser@ubuntu:~$. ~/.bashrc

**Step 6)** Configurations related to HDFS

Set **JAVA\_HOME** inside file **$HADOOP\_HOME/etc/hadoop/hadoop-env.sh**

hduser@ubuntu:~ cd $HADOOP\_HOME/etc/hadoop

hduser@ubuntu:/usr/local/hadoop/etc/hadoop$ gedit hadoop-env.sh

# The java implementation to use.

export JAVA\_HOME=/usr/lib/jvm/java-8-oracle

There are two parameters in **$HADOOP\_HOME/etc/hadoop/core-site.xml**which need to be set-

**1.** **'hadoop.tmp.dir' -**Used to specify directory which will be used by Hadoop to store its data

files.

**2. 'fs.default.name' -**This specifies the default file system.

To set these parameters, open core-site.xml

**sudo gedit $HADOOP\_HOME/etc/hadoop/core-site.xml**

Copy below line in between tags <configuration></configuration>

<property>

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

<value>/app/hadoop/tmp</value>

<description>Parent directory for other temporary directories.</description>

</property>

<property>

<name>fs.defaultFS </name>

<value>hdfs://localhost:54310</value>

<description>The name of the default file system. </description>

</property>

Navigate to the directory **$HADOOP\_HOME/etc/hadoop**

Now, create the directory mentioned in core-site.xml

sudo mkdir -p <Path of Directory used in above setting>

hduser@ubuntu: /usr/local/hadoop/etc/hadoop$ sudo mkdir -p /app/hadoop/tmp

Grant permissions to the directory

hduser@ubuntu:/usr/local/hadoop/etc/hadoop$ sudo chown -R hduser:hadoop /app/hadoop/tmp

hduser@ubuntu:/usr/local/hadoop/etc/hadoop$ sudo chmod 750 /app/hadoop/tmp

**Step 7)** Map Reduce Configuration

Before you begin with these configurations, let's set HADOOP\_HOME path

sudo gedit /etc/profile.d/hadoop.sh

and Enter

export HADOOP\_HOME=/usr/local/hadoop

next make script executable

hduser@ubuntu:~$sudo chmod +x /etc/profile.d/hadoop.sh

Exit the Terminal and restart again

Type echo $HADOOP\_HOME. To verify the path

hduser@ubuntu:~$ echo $HADOOP\_HOME

/usr/local/hadoop

hduser@ubuntu:~$

Now copy files

sudo cp $HADOOP\_HOME/etc/hadoop/mapred-site.xml.template $HADOOP\_HOME/etc/hadoop/mapred-site.xml

Open the mapred-site.xml file

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

Add below lines of setting in between tags <configuration> and </configuration>

<property>

<name>mapreduce.jobtracker.address</name>

<value>localhost:54311</value>

<description>MapReduce job tracker runs at this host and port</description>

</property>

Open **$HADOOP\_HOME/etc/hadoop/hdfs-site.xml** as below,

**sudo gedit $HADOOP\_HOME/etc/hadoop/hdfs-site.xml**

Add below lines of setting between tags <configuration> and </configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

<description>Default block replication</description>

</property>

<property>

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

<value>/home/hduser/hdfs</value>

</property>

Create directory specified in above setting-

sudo mkdir -p <Path of Directory used in above setting>

sudo mkdir -p /home/hduser/hdfs

sudo chown -R hduser:hadoop <Path of Directory created in above step>

sudo chown -R hduser:hadoop /home/hduser/hdfs

sudo chmod 750 <Path of Directory created in above step>

sudo chmod 750 /home/hduser/hdfs

**Step 8)**Before we start Hadoop for the first time, format HDFS using below command

The first step to starting up your Hadoop installation is formatting the Hadoop filesystem which is implemented on top of the local filesystem of your cluster. You need to do this the first time you set up a Hadoop cluster. Do not format a running Hadoop filesystem as you will lose all the data currently in the cluster (in HDFS). To format the filesystem (which simply initializes the directory specified by the dfs.datanode.data.dir variable), run the following command:

**$HADOOP\_HOME/bin/hdfs namenode -format**

hduser@ubuntu:~$ $HADOOP\_HOME/bin/hdfs namenode -format

**Step 9)**Start Hadoop single node cluster using below command

hduser@ubuntu:~$ $HADOOP\_HOME/sbin/start-dfs.sh

hduser@ubuntu:~$ $HADOOP\_HOME/sbin/start-yarn.sh

Using **'jps'** tool/command, verify whether all the Hadoop related processes are running or not.

If Hadoop has started successfully then output of jps should show NameNode, NodeManager, ResourceManager, SecondaryNameNode, DataNode.

**Step 10)**Stopping Hadoop

hduser@ubuntu:~$ $HADOOP\_HOME/sbin/stop-dfs.sh

hduser@ubuntu:~$$HADOOP\_HOME/sbin/stop-yarn.sh

Or from any other connected computer open browser and type

[http://192.168.31.36:50070/dfshealth.html](http://192.168.31.36:50070/dfshealth.html#tab-overview) or

[http://192.168.31.36:50070/dfshealth.jsp](http://192.168.31.36:50070/dfshealth.html#tab-overview)