Installation of Hadoop 3.1.3 in ubuntu 18.04/19.04/19.10

Step 1: Installation of openJDK-8

$ Sudo apt install openjdk-8-jdk openjdk-8-jre

$ java -version

$ sudo apt install vim openssh-server openssh-client

Step 2: Adding the Jdk path to the path variable

Open ~/.bashrc and add

$ sudo vim ~/.bashrc

#go to the last line and add the following

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

export PATH=$PATH:$JAVA_HOME

##save and exit

Inform the OS about the modification

$ source ~/.bashrc

Type

$ echo $JAVA_HOME

$ echo $PATH

Step 3: Add a dedicated user for the HADOOP

$ sudo adduser hadoop

$ sudo usermod -aG sudo hadoop

(Just in case)

$sudo visudo

User privilege specification

root ALL=(ALL:ALL) ALL

hadoop ALL=(ALL:ALL) ALL

(to get out , Ctlr+x , Y, enter)

Step 4: Once the user is added, login to the user "Hadoop" to generate the ssh key for passwordless login (hadoop@machinename)

$ sudo su - hadoop

$ ssh-keygen -t rsa

$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys

$ chmod 0600 ~/.ssh/authorized_keys

Check the login to localhost using ssh is valid

$ ssh localhost

IMPORTANT

Once the connection is made, logout from ssh

$ exit

Step 5: Download the latest binary from Hadoop site

" hadoop-3.1.3.tar.gz "

$ tar -xvzf **hadoop-3.1.3.tar.gz**

$ mv **hadoop-3.1.3 /usr/local/hadoop**

Step 6: Setup the path variables for hadoop

$ sudo vim /etc/profile.d/hadoop_java.sh

Add the following lines to it

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

export HADOOP_HOME=/usr/local/hadoop

export HADOOP_HDFS_HOME=$HADOOP_HOME

export HADOOP_MAPRED_HOME=$HADOOP_HOME

export YARN_HOME=$HADOOP_HOME

export HADOOP_COMMON_HOME=$HADOOP_HOME

export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native

export PATH=$PATH:$JAVA_HOME/bin:$HADOOP_HOME/bin:$HADOOP_HOME/sbin

exportHADOOP_OPTS="$HADOOP_OPTS -Djava.library.path=$HADOOP_HOME/lib/native"

Save and exit. Then source the file

$ source /etc/profile.d/hadoop_java.sh

Confirm your hadoop and hdfs version

$ hadoop version

$ hdfs version

Step 7: Configuring Hadoop

Navigate to /usr/local/hadoop/etc/hadoop and type ls

$ cd /usr/local/hadoop/etc/hadoop

$ hadoop@machine: /usr/local/hadoop/etc/hadoop: ls

Give the permission for the hadoop folder to hadoop user

$ sudo chown -R hadoop:hadoop /usr/local/hadoop

Step 7a: Specify JAVA_HOME in hadoop-env.sh (/usr/local/hadoop/etc/hadoop)

$ vim hadoop-env.sh

Add the following line in java implementation

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64 (54 line)

Save and exit

Step 7b: Modify core-site.xml to setup web portal for hadoop

Add the following lines to it

```
<configuration>
        <property>
        <name>fs.default.name</name>
        <value>hdfs://localhost:9000</value>
        <description>The default file system URI</description>
  </property>
  <property>
```

```
        <name>hadoop.tmp.dir</name>
        <value>/usr/local/hadoop/htemp</value>
  </property>
</configuration>
```

Step 7c: Modify hdfs-site.xml to setup namenode and datanode path and replication factor

Create a folder for namenode and datanode usage

$ ls

Give the permission for the hdfs and htemp folder to hadoop user

$ sudo chown -R hadoop:hadoop /usr/local/hadoop/hdfs

sudo chown -R hadoop:hadoop /usr/local/hadoop/htemp

Modify hdfs-site.xml and add the following lines inside

```
<configuration>
  <property>
        <name>dfs.replication</name>
        <value>1</value>
  </property>

  <property>
        <name>dfs.name.dir</name>
        <value>file:/usr/local/hadoop/hdfs/namenode</value>
  </property>

  <property>
        <name>dfs.data.dir</name>
        <value>file:/usr/local/hadoop/hdfs/datanode</value>
  </property>
</configuration>
```

Step 7d: Configure the mapreduce framework by editing the mapred-site.xml

Modify the mapred-site.xml and add the following lines

```
 <configuration>
        <property>
        <name>mapreduce.framework.name</name>
```

<value>yarn</value>
</property>
<property>
<name>mapreduce.application.classpath</name>
<value>$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*:$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/lib/*</value>
</property>
</configuration>

Step 7e: Configure the YARN resource manager by editing the yarn-site.xml

<configuration>
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
<property>
<name>yarn.nodemanager.env-whitelist</name>
<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
</property>
</configuration>

Step 8: Format the namenode using the command

$ hdfs namenode -format

Test HDFS configuration (/usr/local/hadoop/sbin/)

$ ./start-dfs.sh
$ ./start-yarn.sh
$./start-all.sh

Check the availability of all the nodes by typing

$ jps

12293 Jps
9877 NameNode
10085 DataNode
10953 NodeManager

10590 ResourceManager
10335 SecondaryNameNode

Step 9: Access the Web portal for hadoop management by typing in the following IP address in the browser

http://localhost:9870

Step 10: Check the hadoop cluster overview at

http://localhost:8088

Execute $HADOOP_HOME/sbin - ./stop-all.sh