

Running Hadoop On Ubuntu Linux (Single-Node Cluster)

Expt No: 2

Feb 20, 2019

Author: Subalakshmi Shanthosi S (186001008)

Aim

To configure and install Hadoop in Ubuntu 16.04 LTS OS flavour.

Software's Used

- Ubuntu 16.04 LTS
- Hadoop 1.0.3

Description

Installation of Oracle VirtualBox with guest Operating System as Ubuntu 16.04. Installation of necessary packages namely - openssh-server, openssh-client, java, hadoop in the created virtualbox instance.

Procedure

1. Launch Ubuntu 16.04 LTS.
2. Login to the OS with sudo permission and install the following packages using apt-get command.
 - openssh-server
 - openssh-client
 - java jdk 8
 - javac compiler
 - hadoop 1.0.3
3. Create a new user with sudo permission (hduser:hadoop).
4. Log into the hduser and do the following:
 - Copy the hadoop executable to /usr/local directory.
 - Install and configure appropriate environment variables and parameters in the following configuration files:
 - conf/hadoop-env.sh : Configure JAVA_HOME and HADOOP_HOME with appropriate values.
 - conf/core-site.xml : Configure hadoop default temp directory and default file system.
 - conf/mapred-site.xml : JobTracker name and port number.
 - conf/hdfs-site.xml : Default replication factor specification.
 - Format the namenode by specified dfs.name.dir by running command : /usr/local/hadoop/bin/hadoop namenode -format .
 - Starting the local hadoop single node cluster by running command: /usr/local/hadoop/bin/start-all.sh .
 - To check the current running Hadoop Processes by running command : jps .
 - Stopping the local hadoop single node cluster by running command : /usr/local/hadoop/bin/stop-all.sh .

Output

```
admin@ssn-c6:~$ sudo apt-get install openssh-server
[sudo] password for admin:
Reading package lists... Done
Building dependency tree
Reading state information... Done
openssh-server is already the newest version (1:7.2p2-4ubuntu2.7).
The following packages were automatically installed and are no longer required:
  libdbusmenu-gtk4 libllvm3.8 libqnt-glib1
Use 'sudo apt autoremove' to remove them.
0 upgraded, 0 newly installed, 0 to remove and 4 not upgraded.
admin@ssn-c6:~$ sudo apt-get install openssh-client
Reading package lists... Done
Building dependency tree
Reading state information... Done
openssh-client is already the newest version (1:7.2p2-4ubuntu2.7).
openssh-client set to manually installed.
The following packages were automatically installed and are no longer required:
  libdbusmenu-gtk4 libllvm3.8 libqnt-glib1
Use 'sudo apt autoremove' to remove them.
0 upgraded, 0 newly installed, 0 to remove and 4 not upgraded.
admin@ssn-c6:~$ sudo apt-get install vi
admin@ssn-c6:~$ sudo apt-get install virtu
```

Figure 1: Install openssh-server,openssh-client in Ubuntu OS.

```
hduser@client-VirtualBox: ~/jdk1.8.0_171/bin
hduser@client-VirtualBox:~/jdk1.8.0_171/bin$ export JAVA_HOME=/home/hduser/jdk1.
3.0_171/bin
hduser@client-VirtualBox:~/jdk1.8.0_171/bin$ export PATH=$PATH:$JAVA_HOME
hduser@client-VirtualBox:~/jdk1.8.0_171/bin$ echo $PATH
/usr/local/bin:/usr/bin:/bin:/usr/local/games:/usr/games:/snap/bin:/home/hduser/
jdk1.8.0_171/bin:/home/hduser/jdk1.8.0_171/bin
hduser@client-VirtualBox:~/jdk1.8.0_171/bin$
```

Figure 2: Setting Java Home environment variable to the specified download path of JDK-1.7.

```
client@client-VirtualBox:~$ sudo addgroup hadoop
Adding group 'hadoop' (GID 1001) ...
Done.
client@client-VirtualBox:~$ sudo adduser --ingroup hadoop hduser
Adding user 'hduser' ...
Adding new user 'hduser' (1001) with group 'hadoop' ...
Creating home directory '/home/hduser' ...
Copying files from '/etc/skel' ...
Enter new UNIX password:
Retype new UNIX password:
passwd: password updated successfully
Changing the user information for hduser
Enter the new value, or press ENTER for the default
  Full Name []:
    Room Number []:
    Work Phone []:
    Home Phone []:
    other []:
Is the information correct? [Y/n]
client@client-VirtualBox:~$
client@client-VirtualBox:~$ su - hduser
```

Figure 3: Adding a dedicated hadoop system user.

```
hduser@client-VirtualBox: ~  
hduser@client-VirtualBox:~$ ssh-keygen -t rsa -P ""  
Generating public/private rsa key pair.  
Enter file in which to save the key (/home/hduser/.ssh/id_rsa):  
Created directory '/home/hduser/.ssh'.  
Your identification has been saved in /home/hduser/.ssh/id_rsa.  
Your public key has been saved in /home/hduser/.ssh/id_rsa.pub.  
The key fingerprint is:  
SHA256:vnkUKXUrfJx9B/DpAEj2uW29mZTH0np5C0m09uGt00o hduser@client-VirtualBox  
The key's randomart image is:  
+---[RSA 2048]-----+  
| .000+... |  
| ..=00.... |  
| 000.00. |  
| .+0.B 0 |  
| S..0+ @ |  
| .. = B |  
| .. . E.. |  
| O. .+*+ |  
| O. ...*+ |  
+---[SHA256]-----+  
hduser@client-VirtualBox:~$
```

Figure 4: Configuring SSH in newly created user.

```
hduser@client-VirtualBox: ~  
hduser@client-VirtualBox:~$ cat /proc/sys/net/ipv6/conf/all/disable_ipv6  
1  
hduser@client-VirtualBox:~$
```

Figure 5: Disabling IPv6 in the newly created user account.

```
# Log Martian Packets  
#net.ipv4.conf.all.log_martians = 1  
  
# /etc/sysctl.conf  
# disable ipv6  
net.ipv6.conf.all.disable_ipv6 = 1  
net.ipv6.conf.default.disable_ipv6 = 1  
net.ipv6.conf.lo.disable_ipv6 = 1
```

Figure 6: Disabling IPv6 in the newly created user account.

```

hduser@client-VirtualBox: /usr/local
hduser@client-VirtualBox: /usr/local$ sudo wget -e use_proxy=yes https_proxy=proxy.ssn.net:8080 https://archive.apache.org/dist/hadoop/core/hadoop-1.0.3/hadoop-1.0.3.tar.gz
--2019-04-10 11:13:29-- http://https-proxy-proxy.ssn.net:8080/
Resolving proxy.ssn.net (proxy.ssn.net)... 192.168.2.5
Connecting to proxy.ssn.net (proxy.ssn.net)|192.168.2.5|:8080... connected.
Proxy request sent, awaiting response... 403 Forbidden
2019-04-10 11:13:30 ERROR 403: Forbidden.

--2019-04-10 11:13:30-- https://archive.apache.org/dist/hadoop/core/hadoop-1.0.3/hadoop-1.0.3.tar.gz
Connecting to proxy.ssn.net (proxy.ssn.net)|192.168.2.5|:8080... connected.
Proxy request sent, awaiting response... 200 OK
Length: 62428860 (60M) [application/x-gzip]
Saving to: 'hadoop-1.0.3.tar.gz.1'

hadoop-1.0.3.tar.gz 28%[====>] 16.95M 942KB/s eta 57s

```

Figure 7: Installation of Hadoop 1.0.3 in new user login.

```

<configuration>
<!-- conf/core-site.xml -->
<property>
  <name>hadoop.tmp.dir</name>
  <value>/app/hadoop/tmp</value>
  <description>A base for other temporary directories.</description>
</property>
<property>
  <name>fs.default.name</name>
  <value>hdfs://localhost:54310</value>
  <description>The name of the default file system. A URI whose scheme and authority determine the FileSystem implementation. The uri's scheme determines the config property (fs.SCHEME.impl) naming the FileSystem implementation class. The uri's authority is used to determine the host, port, etc. for a filesystem.</description>
</property>
|
</configuration>

```

Figure 8: Configuring hadoop core-site.xml .

```

<configuration>
<!-- conf/mapred-site.xml -->
<property>
  <name>mapred.job.tracker</name>
  <value>localhost:54311</value>
  <description>The host and port that the MapReduce job tracker runs at. If "local", then jobs are run in-process as a single map and reduce task.
  </description>
</property>
</configuration>

```

Figure 9: Configuring Hadoop MapReduce.

```

<configuration>
<!-- conf/hdfs-site.xml -->
<property>
  <name>dfs.replication</name>
  <value>1</value>
  <description>Default block replication.
    The actual number of replications can be specified when the file
    is created.
    The default is used if replication is not specified in create time.
  </description>
</property>
</configuration>

```

Figure 10: Configuring Hadoop HDFS Site.

```

hduser@client-VirtualBox: /usr/local/hadoop/bin$ ./start-all.sh
Warning: $HADOOP_HOME is deprecated.

starting namenode, logging to /usr/local/hadoop/libexec/../logs/hadoop-hduser-na
menode-client-VirtualBox.out
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:2/CujIsBcbST/wahKi0ViZl0Vs2GPzw5AZzzkfbNBfE.
Are you sure you want to continue connecting (yes/no)? yes
localhost: Warning: Permanently added 'localhost' (ECDSA) to the list of known h
osts.
localhost: starting datanode, logging to /usr/local/hadoop/libexec/../logs/hadoo
p-hduser-datanode-client-VirtualBox.out
localhost: starting secondarynamenode, logging to /usr/local/hadoop/libexec/../l
ogs/hadoop-hduser-secondarynamenode-client-VirtualBox.out
starting jobtracker, logging to /usr/local/hadoop/libexec/../logs/hadoop-hduser-
jobtracker-client-VirtualBox.out
localhost: starting tasktracker, logging to /usr/local/hadoop/libexec/../logs/ha
doo-hduser-tasktracker-client-VirtualBox.out
hduser@client-VirtualBox: /usr/local/hadoop/bin$

```

Figure 11: Starting hadoop NameNode,Datanode,JobTracker and TaskTracker.

```

hduser@client-VirtualBox: /usr/local/hadoop/bin$ jps
2706 Jps
2582 TaskTracker
2360 SecondaryNameNode
2443 JobTracker
2062 NameNode
2223 DataNode

```

Figure 12: Showing hadoop NameNode,Datanode,JobTracker and TaskTracker Processes.

```
hduser@client-VirtualBox: /usr/local/hadoop/bin$ ./stop-all.sh
Warning: $HADOOP_HOME is deprecated.

stopping jobtracker
localhost: stopping tasktracker
stopping namenode
localhost: stopping datanode
localhost: stopping secondarynamenode
hduser@client-VirtualBox: /usr/local/hadoop/bin$
```

Figure 13: Stopping hadoop NameNode,DataNode,JobTracker and TaskTracker.

Result

Thus the hadoop single node cluster is successfully created in Ubuntu 16.04 OS version and required packages are installed.