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 neccessary 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

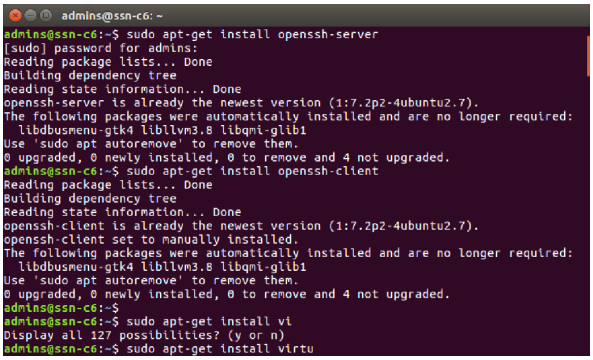

Figure 1: Install openssh-server,openssh-client in Ubuntu OS.

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

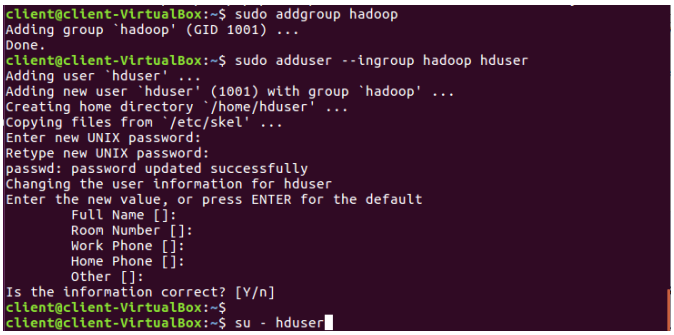

Figure 3: Adding a dedicated hadoop system user.

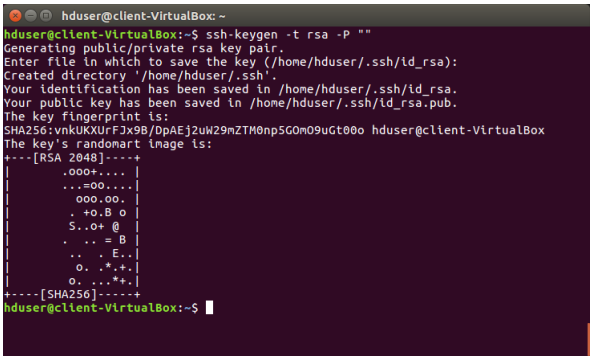

Figure 4: Configuring SSH in newly created user.

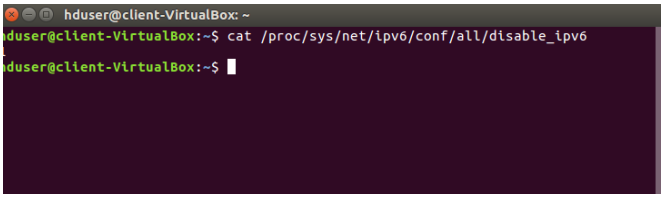

Figure 5: Disabling IPv6 in the newly created user account.

```
# 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.

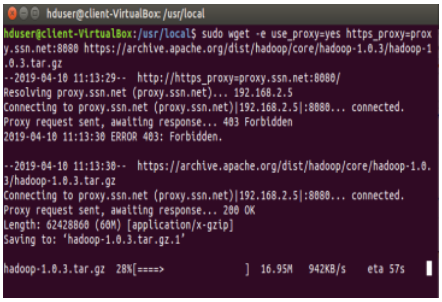

Figure 7: Installation of Hadoop 1.0.3 in new user login.

```
<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.

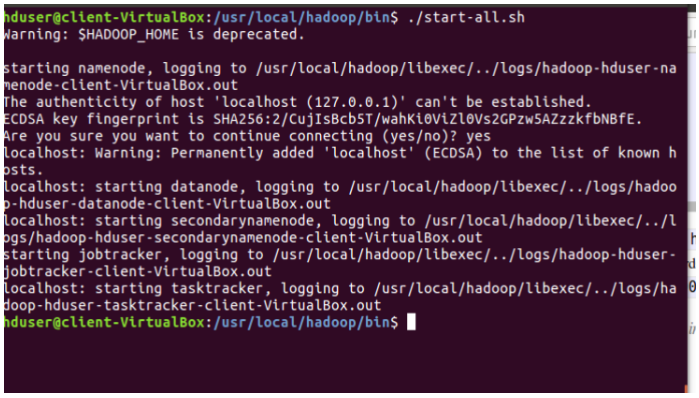

Figure 11: Starting hadoop NameNode,Datanode,JobTracker and TaskTracker.

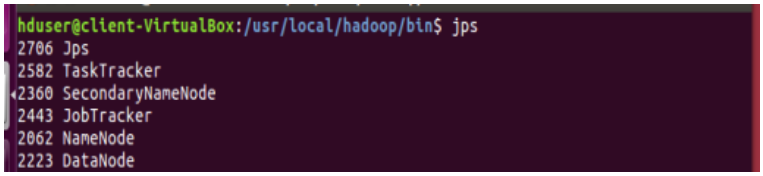

Figure 12: Showing hadoop NameNode,Datanode,JobTracker and TaskTracker Processes.

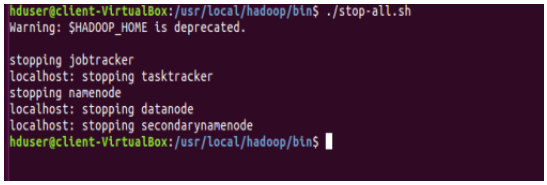

Figure 13: Stopping hadoop NameNode,DataNode,JobTracker and TaskTracker.

Result

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