Hadoop Practical

Created	@June 28, 2022 8:32 PM
Tags	
Property	

Download the Virtual Box from this link: www.virtualbox.org/wiki/Downloads

Run the downloaded file and click on next.

Go through the next steps and install the software.

~~If an error occurs, try this: Restart your computer and go to the BIOS setup. Go to UEFI Firmware > Advanced > CPU Setup > Intel Virtualization Technology > Then enable it. Save and exit.~~

In Oracle Virtual Box Manager, click on New.

Give the name of virtual machine as ubuntu-bigdata.

Click on Next.

Enter 3072 as the size of RAM and click on Next.
Click on Create.

Click on Next.

Click on Next.

Set the Hard Disk space to 20 GB.

Click on Create.

Download the Ubuntu ISO from this link: www.ubuntu.com/download/desktop

Select the newly created machine in the Oracle Virtual Box Manager.

Go to Settings > Under the Controller: IDE > Empty.

At the right hand side, click on the CD icon and click on Choose Virtual Optical Disk File.

Choose the downloaded Ubuntu ISO file and start the machine.

When the machine is loaded, click on Install ubuntu option.

Then select the language and click Continue.

Then o for Normal Installation and click on Continue.

Choose Erase disk and install Ubuntu (default), and click on Install now. A warning message will come, press Continue.

After that, select location and click on Continue.

Provide the information for the computer and click Continue.

Then wait for some time for the operating system to install.

After the installation is complete, a restart will be requested.

After restarting, you can find the desktop of the Ubuntu LTS 18.04.

To download JDK, please follow this link:
www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html

Accept the License Agreement and go for Linux x64 version (ending with tar.gz)

To download the JRE, please follow this link:
https://www.java.com/en/download/manual.jsp

Download the LinuxX64 version.

Open up the terminal by pressing Ctrl+Alt+T.

Create a .java folder at /usr/local/ using root permission. It will ask for the password for the Ubuntu system.

$sudo mkdir -p /usr/local/java

$cd ~/Downloads

$sudo cp -r jre*.tar.gz /usr/local/java

$sudo cp -r jdk*.tar.gz /usr/local/java

$cd /usr/local/java

$sudo tar xvzf jdk*.tar.gz

$sudo tar xvzf jre*.tar.gz

$sudo gedit /etc/profile

Then write these lines into the profile document.

export JAVA_HOME=/usr/local/java/jdk1.8.0_333

PATH=$PATH:$JAVA_HOME/bin

export JRE_HOME=/usr/local/java/jre1.8.0_333

PATH=$PATH:$JRE_HOME/bin

export PATH

Save and close the file.

Write the following lines on the terminal.

$sudo update-alternatives —install /usr/bin/java java /usr/local/java.jdk1.8.0_333/bin/java 2

$sudo update-alternatives —install /usr/bin/javac javac /usr/local/java//jdk1.8.0_333/bin/javac 2

$sudo update-alternatives —install /usr/bin/jar jar /usr/local/java//jdk1.8.0_333/bin/jar 2

$sudo update-alternatives —set java /usr/local/java/jdk1.8.0_333/bin/java

$sudo update-alternatives —set javac /usr/local/java/jdk1.8.0_333/bin/javac

$sudo update-alternatives —set jar /usr/local/java/jdk1.8.0_333/bin/jar

Note: If you are using the same Java version, the given lines will work fine. For different versions, you need to edit the version details in the lines.

To find the version of Java, open up a new terminal by pressing Ctrl+Alt+T and run the following commands.

$cd /usr/local/java/

$ls

Note the java version and close the terminal.

$. /etc/profile

$reboot

Open the terminal by pressing Ctrl+Alt+T and runs the following statements.

$java -version

$sudo apt-get update

$sudo apt-get upgrade —fix-missing

$sudo apt-get install ssh

$sudo apt-get install rsync

$ssh-keygen -t rsa

Press ‘Enter’ three times without entering anything.

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

$ssh localhost

Open the browser and go to this link: archive.apache.org/dist/hadoop/core/hadoop-2.4.1/

Click on the link with the extension .tar.gz

Create a new folder in the home directory named hadoop.

Extract the .tar.gz file in this folder.

In the terminal write the following command:

$sudo gedit /etc/profile

Add these lines to the file.

HADOOP_INSTALL=/home/bigdata/hadoop

PATH=$PATH:$HADOOP_INSTALL/bin

export PATH

Save and close the file.

$. /etc/profile

$exit

Open core-site.xml from the given path /home/bigdata/hadoop/etc/hadoop/core-site.xml

```
<configuration>
    <property>
        <name>fs.defaultFS</name>
        <value>hdfs://localhost:9000</value>
    </property>
    <property>
        <name>hadoop.tmp.dir</name>
        <value>/home/bigdata/hadoop/hadoop-2.4.1/temp</value>
```

```
    </property>
</configuration>
```

Save and close the file

Open hdfs-site.xml from the given path /home/bigdata/hadoop/etc/hadoop/hdfs-site.xml.

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

Open mapred-site.xml from the given path /home/bigdata/hadoop/etc/hadoop/mapred-site.xml

```
<configuration>
    <property>
        <name>mapred.job.tracker</name>
        <value>localhost:9001</value>
    </property>
</configuration>
```

Save and close the file.

Open hadoop-env.sh from the given path /home/bigdata/hadoop/etc/hadoop/hadoop-env.sh

```
export JAVA_HOME=/usr/local/java/jdk1.8.0_333
```

Save and close the file.

Open the terminal using Ctrl+Alt+T and write the following commands.

```
$sudo gedit ~/.bashrc
export JAVA_HOME=/usr/local/java/jdk1.8.0_333
export HADOOP_INSTALL=/home/bigdata/hadoop
```

export HADOOP_MAPRED_HOME=$HADOOP_INSTALL

export HADOOP_COMMON_HOME=$HADOOP_INSTALL

export HADOOP_HDFS_HOME=$HADOOP_INSTALL

export YARN_HOME=$HADOOP_INSTALL

export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_INSTALL/lib/native

export PATH=$PATH:$JAVA_HOME/bin

export PATH=$PATH:HADOOP_INSTALL/bin

export PATH=$PATH:$HADOOP_INSTALL/sbin

Save and close the file.

Open the terminal and run the following commands.

$. /etc/profile

$source ~/.bashrc

$hadoop namenode -format

$start-dfs.sh

$start-yarn.sh

$jps

$stop-dfs.sh

$stop-yarn.sh

Download the hadoopMyFiles folder.

Open a terminal and run the following commands:

$start-dfs.sh

$start-yarn.sh

$jps

$ls ~/hadoopMyFiles

$ls -l /hadoopMyFiles

$cat ~/hadoopMyFiles/Sample_File/txt

$cat ~/hadoopMyFiles/student_info.csv

$hdfs dfs -mkdir /hadoopMyFiles

$hadoop fs -mkdir /hadoopMyFiles1

$hdfs dfs -rm -r /hadoopMyFiles1

$hdfs dfs -ls /

$hadoop fs -ls /

$ls -l ~/hadoopMyFiles

$hdfs dfs -put ~/hadoopMyFiles/Sample_File.txt /hadoopMyFiles

$hadoop fs -ls /hadoopMyFiles

$hdfs dfs -copyFromLocal ~/hadoopMyFiles/student_info.csv /hadoopMyFiles

$hadoop fs -ls /hadoopMyFiles

$hdfs dfs -cp /hadoopMyFiles/Sample_File.txt /hadoopMyFiles/Sample_File1.txt

$ls -l ~/hadoopMyFiles

$hdfs dfs -mv /hadoopMyFiles/Sample_File1.txt /hadoopMyFiles/Sample_File2.txt

$hadoop fs -mkdir /hadoopMyFiles/Dir1

$hdfs dfs -mv /hadoopMyFiles/Sample_File2.txt /hadoopMyFiles/Dir1/File2.txt

$hadoop fs -cat /hadoopMyFiles/student_info.csv

$hdfs dfs -cat /hadoopMyFiles/student_info.csv

$hdfs dfs -ls /hadoopMyFiles

$ls -l ~/hadoopMyFiles

$hdfs dfs -get /hadoopMyFiles/Sample_File.txt ~/hadoopMyFiles/Sample1.txt

$hadoop fs -copyToLocal /hadoopMyFiles/student_info.csv
~/hadoopMyFiles/student1.csv

$ls -l ~/hadoopMyFiles

$hdfs dfs -rm -r /hadoopMyFiles

$hdfs dfs -count /

$echo $HOME

$echo $HADOOP_INSTALL

$cat $HOME/hadoopMyFiles/Sample_File.txt

$cat ~/hadoopMyFiles/Sample_File.txt

$cat /home/bigdata/hadoopMyFiles/Sample_File.txt

$hdfs dfs -mkdir /myinput

$hdfs dfs -put /home/bigdata/hadoopMyFiles/Sample_File.txt /myinput

$hadoop jar $HADOOP_INSTALL/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.4.1.jar wordcount /myinput /myoutput

$hdfs dfs -cat /myoutput/part-r-00000

$hdfs dfs -cat /myoutput/part*

$stop-dfs.sh

$stop-yarn.sh