Ubuntu Installation Steps:

1. Install Oracle VM Virtual Box. Keep Default settings.
2. Install Ubuntu64ServerBasic (OVA-oracle virtual application).
3. For Step 2, open virtual box, File->Import appliance ->Next->Import. Choose the location where ova file is kept.

* Check for the memory. It should be exactly half or less than the system RAM.Otherwise,system will crash.

It will be now visible in left pane with icon->powered off.

To start it, click on start.

1. Select virtual machine, go to settings->
2. Go to general->check version is set to Ubuntu 64 bit. If this option is not available, go to Reboot (press F2).

* Set Virtualization and enable all options.

6. Go to Settings->System->check the slider.It should be on Green color.

7. Go to Network, set settings for adapters.Enable adapter 1(Host only n/w,Name:virtual host only)and adapter2(bridge adapter,Name:Broadcom n/w adapter(or whatever internet n/w you have connected).

Adapter 3 and Adapter 4 disable it.

Note:If adapter 1 not working,then host files of windows are missing.

We can also check this setting In file->preferences after start of the virtual machine.

8.Now,click on start.

If blank window is coming,then this is due to ubuntu32 selected .We need it to change to ubuntu64.

Settings->General->check version.

Reboot your system to solve this.Press F2->Virtualization->enable all options.

Again,restart the machine.ubuntu64 wiil,be visible now.Select it..

Note.Also check,another virtual machine like Hyper V should not be enable .To disable it,

Go to Control Panel->Programs and features->turn Windows features on/off->uncheck hyper v.

Restart the machine.

9.Type username:hduser

Pwd:hduser

Now,we can run the linus commands on it.

Run the command:ifconfig eth0->to check the IP of virtual machine.

192.168.56.123

**Putty:It’s a** remote shell to connect to server.

Double click on putty and give ip:192.168.56.123

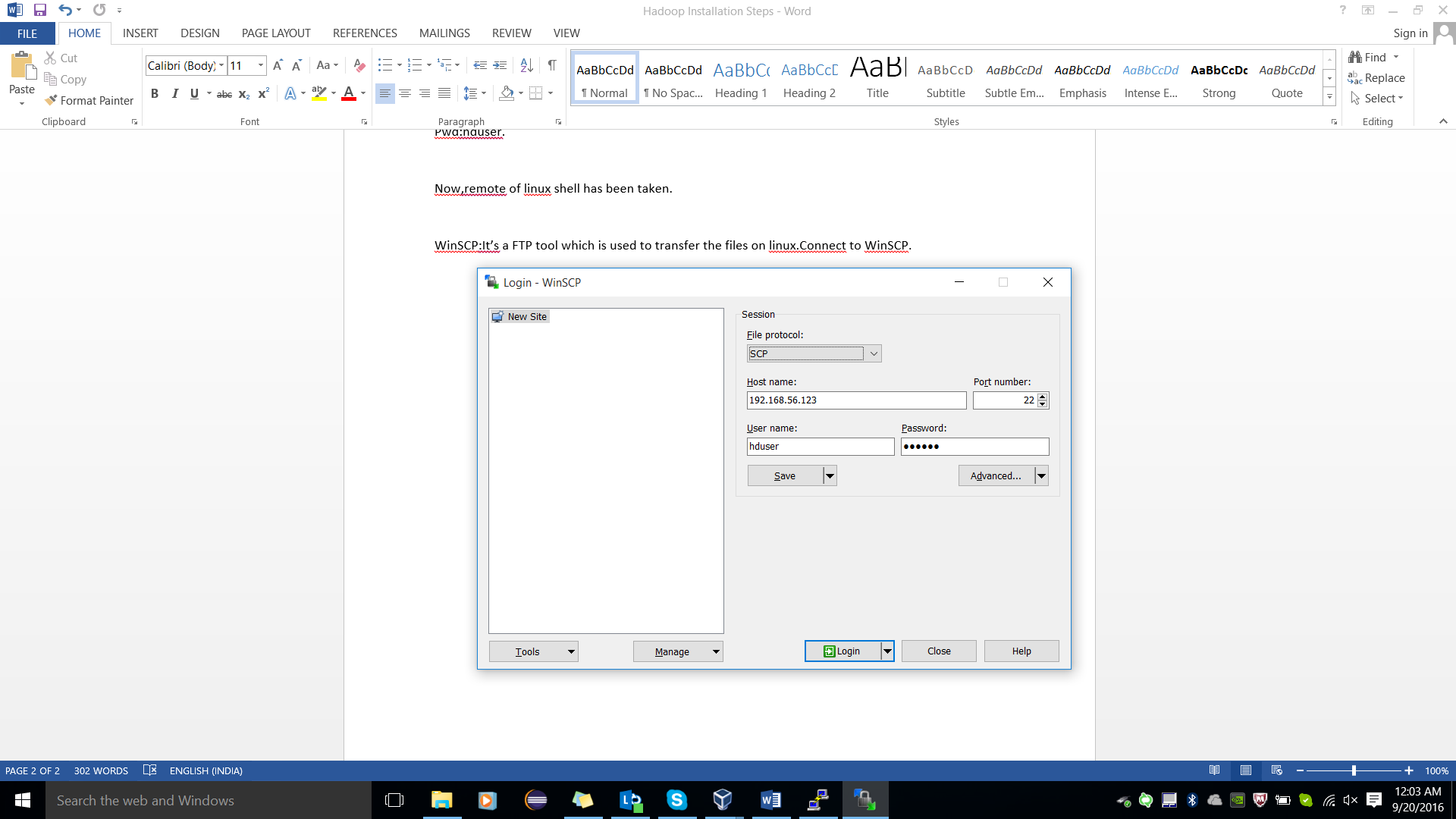
Connection type:SSH

Id:hduser

Pwd:hduser.

Now,remote of linux shell has been taken.

WinSCP:It’s a FTP tool which is used to transfer the files on linux.Connect to WinSCP.



Pwd:hduser

Now,install Hadoop into linux terminal,

We are using 2.7.1 version of Hadoop,

Copy the Hadoop-2.7.1.tar file into WinSCP(right hand side) by copy and paste.

The location will be,/home/hduser.

We can check the location,on putty by running the command:

->pwd

/home/hduser

Extract this tar file on linux by running the following command:

tar -xzvf hadoop-2.7.1.tar.gz

x-extract

z-dual archieve,more than one time file has been archieved.

v-virbose,Whatever it extracts,it will show every step.

f-forcefully,till the time it don’t extract final folder,keeps extracting.

Now, check it on WinSCP ,Hadoop folder has been extracted..

Now,specify in Linux, where the Hadoop is installed.For this,we need to set the Path.For this,we have environment file in Linux.

To set the environment variable,use bashrc file.

->sudo nano ~/.bashrc

sudo->admin rights.

nano->edit the file.

~->home directory

We have bash-profile file also which contains the data of per profile.

Bashrc is at the upper level and bash –profile comes under bashrc which contains the login information of individual users.

->sudo nano ~/.bashrc

When this file gets opened,set the following environment variables;

sudo nano ~/.bashrc

Note:Use Ctrl+x to come out of the editor.

Append the following variables in this file:

export HADOOP\_PREFIX="/home/hduser/hadoop-2.7.1/"

export PATH=$PATH:$HADOOP\_PREFIX/bin

export PATH=$PATH:$HADOOP\_PREFIX/sbin

export HADOOP\_COMMON\_HOME=${HADOOP\_PREFIX}

export HADOOP\_MAPRED\_HOME=${HADOOP\_PREFIX}y

export HADOOP\_HDFS\_HOME=${HADOOP\_PREFIX}

export YARN\_HOME=${HADOOP\_PREFIX}

After saving the file, run the source command to refresh the values of the environment variables.

source ~/.bashrc

Edit the /home/hduser/hadoop-2.7.1/etc/hadoop/core-site.xml file and add the Hadoop HDFS URI (NameNode and its port) property under configuration tag.

To edit type,nano filepath(e.g. nano /home/hduser/hadoop-2.7.1/etc/hadoop/core-site.xml)

Note:while doing the copy and paste of the command,no space should be there,otherwise the node may not start.

<property>

<name>fs.defaultFS</name>

<value>hdfs://192.168.56.123:8020</value>

<final>true</final>

</property>

Edit the /home/hduser/hadoop-2.7.1/etc/hadoop/hdfs-site.xml file and add the Hadoop HDFS properties:

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:///home/hduser/hadoop-2.7.1/hadoop\_data/dfs/namenode</value>

</property>

<property>

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

<value>file:///home/hduser/hadoop-2.7.1/hadoop\_data/dfs/datanode</value>

</property>

</configuration>

Rename the mapred-site.xml.template file through WinScp to mapred-site.xml

Edit the /home/hduser/hadoop-2.7.1/etc/hadoop/mapred-site.xml file and specify the MapReduce framework as YARN.

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

Edit the /home/hduser/hadoop-2.7.1/etc/hadoop/yarn-site.xml file and specify the YARN properties.

<configuration>

<property>

<name>yarn.resourcemanager.address</name>

<value>192.168.56.123:8032</value>

</property>

<property>

<name>yarn.resourcemanager.scheduler.address</name>

<value>192.168.56.123:8030</value>

</property>

<property>

<name>yarn.resourcemanager.resource-tracker.address</name>

<value>192.168.56.123:8031</value>

</property>

<property>

<name>yarn.resourcemanager.admin.address</name>

<value>192.168.56.123:8033</value>

</property>

<property>

<name>yarn.resourcemanager.webapp.address</name>

<value>192.168.56.123:8088</value>

</property>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.aux-services.mapreduce\_shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

<property>

<name>yarn.resourcemanager.scheduler.class</name>

<value>org.apache.hadoop.yarn.server.resourcemanager.scheduler.capacity.CapacityScheduler</value>

</property>

</configuration>

Last step before running the Hadoop services is to format the NameNode. Before executing the format command, make sure that the dfs.namenode.name.dir and dfs.datanode.data.dir directories specified in hdfs-site.xml file does not exists. The hdfs command exists in the /home/hduser/hadoop-2.7.1/bin directory.

hdfs namenode –format

It will generate the folder named hadoop\_data we can check it in winscp .The path where it stores is:

/home/hduser/hadoop-2.7.1…Refresh once only then it will be visible.

Now,Hadoop has been installed .Run below commands to start the services on Hadoop.

->JPS is the command to check which all services are running.

->jps

Service Command

Namenode hadoop-daemon.sh start namenode

Datanode hadoop-daemon.sh start datanode

Resourcemanager yarn-daemon.sh start resourcemanager

Nodemanager yarn-daemon.sh start nodemanager

Check with jps whether all the services has started or not.

After that,open chrome

<http://192.168.56.123:50070/>

HDFS will be visible .Go to Browse->utilities to see all files.

To work on cloudera,go to oracle VM,file->import alliance->import it.It has all the things installed under it.

No need to make use of putty.winscp still used to transfer the data.

No need to move jar files on Hadoop.Direct give the path of the jar file

Cloudera is the environment where all the things are installed under it.