Check the version compatibility of cloudera components by clicking

https://www.cloudera.com/documentation/enterprise/release-notes/topics/cdh_vd_cdh_package_tarball_53.html

HDFS file system : localhost:50070

Reinstall the VMware tools

enabling the drag and drop and copy paste in VMWARE

1) sudo apt-get autoremove open-vm-tools

2) Install VMware Tools by following the usual method (Virtual Machine --> Reinstall VMWare Tools)

3) Reboot the VM

4) sudo apt-get install open-vm-tools-desktop

5) Reboot the VM, after the reboot copy/paste and drag/drop will work!

Recommended Platform:

- OS: Linux is supported as a development and production platform. You can use Ubuntu 14.04 or later (you can also use other Linux flavors like: CentOS, +-, etc.)
- Hadoop: Cloudera Distribution for Apache hadoop CDH5.x (you can use Apache hadoop 2.x)

Setup Platform

If you are using Windows/Mac OS you can create virtual machine and install Ubuntu using VMWare Player, alternatively you can create virtual machine and install Ubuntu using Oracle Virtual Box.

Prerequisites:

Install Java 7 (Recommended Oracle Java)

Install Python Software Properties

```
1 sudo apt-get install python-software-properties
```

Add Repository

```
1 sudo add-apt-repository ppa:webupd8team/java
```

Update the source list

```
1     sudo apt-get update
```

Install Java

```
1 sudo apt-get install openjdk-8-jre
```

Configure SSH

Install Open SSH Server-Client

```
1 sudo apt-get install openssh-server openssh-client
```

Generate Key Pairs

```
1   ssh-keygen -t rsa -P ""
```

Configure password-less SSH

```
cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
```

4.2.4. Check by SSH to localhost

```
    ssh localhost
```

Install Hadoop

Download Hadoop

https://www.cloudera.com/documentation/enterprise/release-notes/topics/cdh_vd_cdh_package_tarball_53.html

Untar Tar ball

```
tar -xvzf hadoop-2.5.0-cdh5.3.2.tar.gz
```

Note: All the required jars, scripts, configuration files, etc. are available in HADOOP_HOME directory (hadoop-2.5.0-cdh5.3.2)

Setup Configuration:

Edit .bashrc:

Edit .bashrc file located in user's home directory and add following parameters:

cd

sudo gedit ~/.bashrc

```
export HADOOP_PREFIX=" /usr/local/hadoop"

export PATH=$PATH:$HADOOP_PREFIX/bin

export PATH=$PATH:$HADOOP_PREFIX/sbin

export HADOOP_MAPRED_HOME=${HADOOP_PREFIX}

export HADOOP_COMMON_HOME=${HADOOP_PREFIX}

export HADOOP_HDFS_HOME=${HADOOP_PREFIX}

export YARN_HOME=${HADOOP_PREFIX}



export JAVA_HOME="/usr"

export PATH="$PATH:$JAVA_HOME/bin"
```

```
source ~/.bashrc
```

Note: After above step restart the terminal, so that all the environment variables will come into effect

Edit hadoop-env.sh:

Edit configuration file hadoop-env.sh (located in HADOOP_HOME/etc/hadoop) and set JAVA_HOME:

```
export JAVA_HOME=<path-to-the-root-of-your-Java-installation> (eg:
/usr/lib/jvm/java-7-oracle/)
```

```
export JAVA_HOME=/usr
```

Edit core-site.xml:

Edit configuration file core-site.xml (located in HADOOP_HOME/etc/hadoop) and add following entries:

```
<configuration>
    <property>
        <name>fs.defaultFS</name>
        <value>hdfs://localhost:9000</value>
    </property>
    <property>
        <name>hadoop.tmp.dir</name>
        <value>/usr/local/hadoop</value>
    </property>
</configuration>
```

Note: /home/hdadmin/hdata is a sample location; please specify a location where you have Read Write privileges

Edit hdfs-site.xml:

Edit configuration file hdfs-site.xml (located in HADOOP_HOME/etc/hadoop) and add following entries:

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

Edit mapred-site.xml:

Edit configuration file mapred-site.xml (located in HADOOP_HOME/etc/hadoop) and add following entries:

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

Edit yarn-site.xml:

Edit configuration file mapred-site.xml (located in HADOOP_HOME/etc/hadoop) and add following entries:

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

Start the Cluster:

Format the name node:

```
hdfs namenode -format
```

NOTE: This activity should be done once when you install hadoop, else It will delete all your data from HDFS

Start HDFS Services:

```
1    $sbin/start-dfs.sh
```

Start YARN Services:

```
1    $sbin/start-yarn.sh
```

Check whether services have been started

```
$jps
NameNode
DataNode
ResourceManager
NodeManager
```

Run Map-Reduce Jobs

Problem : datanode hasnt started :

cloudera@ubuntu:~$ jps

47840 NameNode

48433 NodeManager

48316 ResourceManager

48574 Jps

48116 SecondaryNameNode

cloudera@ubuntu:~$ cd /home/cloudera/hdata/dfs

cloudera@ubuntu:~/hdata/dfs$ ls

data name namesecondary

cloudera@ubuntu:~/hdata/dfs$ rm -Rf data/

cloudera@ubuntu:~/hdata/dfs$ ls

name namesecondary

cloudera@ubuntu:~/hdata/dfs$ rm -Rf name

cloudera@ubuntu:~/hdata/dfs$ rm -Rf namesecondary/

cloudera@ubuntu:~/hdata/dfs$ ls

cloudera@ubuntu:~$ jps

47840 NameNode

48433 NodeManager

48316 ResourceManager

48116 SecondaryNameNode

48839 Jps

cloudera@ubuntu:~$ stop-all.sh

7.1. Run word count example:

```
1 $ bin/hdfs dfs -mkdir /inputwords

2 $ bin/hdfs dfs -put <data-file> /inputwords

3 $ bin/yarn jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.5.0-
  cdh5.3.2.jar wordcount /inputwords /outputwords

4 $ bin/hdfs dfs -cat /outputwords/*
```

Play with HDFS Commands and perform various operations, Follow HDFS command Guide

8. Stop The Cluster

8.1. Stop HDFS Services:

```
1     $sbin/stop-dfs.sh
```

8.2. Stop YARN Services:

```
1     $sbin/stop-yarn.sh
```