Optional 1- Install Ubuntu in windows machine open below link. You can skip this if you already have Ubuntu installed.

https://theholmesoffice.com/installing-ubuntu-in-vmware-player-on-windows/

**Open the Ubuntu in VM and start-**

Optional 2 - Follow below link to make password less sudo

https://www.techiediaries.com/ubuntu/sudo-without-password-ubuntu-20-04/

**Note : Please run only those scripts from ubuntu console which are given in box**

Create software dir in /opt

| sudo mkdir -p /home/imran/software |
| --- |

Step 1:-

Download the hadoop binaries from given location-

| https://downloads.apache.org/hadoop/common/hadoop-3.3.3/hadoop-3.3.3.tar.gz |
| --- |

Step 2:- Install java

**Run below 2 commands on console to install default java -**

| sudo apt-get update  sudo apt-get install openjdk-8-jdk |
| --- |

Run below command to verify the version of the JDK:

| java -version |
| --- |

*Java will be installed at given location*

*/usr/lib/jvm/java-8-openjdk-amd64*

**Create a link for this, run below command-**

| ln -s /usr/lib/jvm/java-8-openjdk-amd64 /home/imran/jsdk |
| --- |

Step 4:- Copy the downloaded https://archive.apache.org/dist/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz file into a folder name /home/imran/software

| sudo cp ~/Downloads/hadoop-3.3.0.tar.gz /home/imran/software/ |
| --- |

Step4: Extract the hadoop-3.3.0.tar.gz file and rename it to **hadoop**.

| sudo tar -xvf /home/imran/software/hadoop-3.3.0.tar.gz  sudo ln -s hadoop-3.3.0 hadoop |
| --- |

Step5: Install password less ssh. Run below commands

| sudo apt-get install ssh   * + Enter Y   sudo apt-get install rsync  ssh-keygen |
| --- |

**(Note-**Above keygen command ask some input from imran, No need to provide any input ,just press enter whenever it ask for input**)**

| cp .ssh/id\_rsa.pub .ssh/authorized\_keys  ssh localhost   * + Enter yes |
| --- |

Step6**:** Set environment variables in .bashrc file, run below command in home directory

| nano .bashrc |
| --- |

A new console page will be open, press insert and some export commands into it.

| export HOME\_DIR=/home/imran  export JAVA\_HOME=$HOME\_DIR/jsdk  export HADOOP\_HOME=$HOME\_DIR/software/hadoop  export HADOOP\_CONF\_DIR=$HADOOP\_HOME/etc/hadoop  export PATH=$PATH:$JAVA\_HOME/bin:$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin:$HADOOP\_CONF\_DIR/ |
| --- |

**Set Up Configurations Files**

Go to the dir path -

| cd /home/imran/software/hadoop/etc/hadoop |
| --- |

**Open below files and update**

create a folder for yarn data

| sudo mkdir /home/imran/software/hadoop/yarndata |
| --- |

open /home/imran/hadoop/etc/hadoop/**core-site.xml either from GUI or nano**

| sudo nano /home/imran/hadoop/etc/hadoop/core-site.xml |
| --- |

**Copy below xml and replace/overwrite in core-site.xml file-**

| <?xml version="1.0" encoding="UTF-8"?>  <configuration>  <property>  <name>fs.defaultFS</name>  <value>hdfs://localhost:54310</value>  </property>  <property>  <name>hadoop.tmp.dir</name>  <value>/home/imran/software/hadoop/yarndata</value>  <description>A base for other temporary directories.</description>  </property>  </configuration> |
| --- |

To save the changes

| Control +S  Control +X |
| --- |

**Create 2 directories for datanodes**

| sudo mkdir /home/imran/software/hadoop/nndata  sudo mkdir /home/imran/software/hadoop/ddata |
| --- |

**Open** /home/imran/hadoop/etc/hadoop/**hdfs-site.xml**

| sudo nano /home/imran/hadoop/etc/hadoop/hdfs-site.xml |
| --- |

| <?xml version="1.0" encoding="UTF-8"?>  <configuration>  <property>  <name>dfs.replication</name>  <value>1</value>  </property>  <property>  <name>dfs.namenode.name.dir</name>  <value>/home/imran/software/hadoop/nndata</value>  <description>A base for other temporary directories.</description>  </property>  <property>  <name>dfs.datanode.name.dir</name>  <value>/home/imran/software/hadoop/dndata</value>  <description>A base for other temporary directories.</description>  </property>  </configuration>  </configuration> |
| --- |

To save the changes

| Control +S  Control +X |
| --- |

3. Create file /home/imran/hadoop/etc/hadoop/**mapred-site.xml** if not exist and overwrite it.

Create file

| sudo touch /home/imran/hadoop/etc/hadoop/**mapred-site.xml** |
| --- |

Open File

| sudo nano /home/imran/hadoop/etc/hadoop/**mapred-site.xml** |
| --- |

| <?xml version="1.0" encoding="UTF-8"?>  <configuration>  <property>  <name>mapreduce.framework.name</name>  <value>yarn</value>  </property>  <property>  <name>mapreduce.jobhistory.address</name>  <value>osboxes:10020</value>  </property>  <property>  <name>mapreduce.jobhistory.webapp.address</name>  <value>osboxes:19888</value>  </property>  <property>  <name>yarn.app.mapreduce.am.env</name>  <value>HADOOP\_MAPRED\_HOME=/home/imran/hadoop</value>  </property>  <property>  <name>mapreduce.map.env</name>  <value>HADOOP\_MAPRED\_HOME=/home/imran/hadoop</value>  </property>  <property>  <name>mapreduce.reduce.env</name>  <value>HADOOP\_MAPRED\_HOME=/home/imran/hadoop</value>  </property>  </configuration> |
| --- |

To save the changes

| Control +S  Control +X |
| --- |

4. Open and overwrite **yarn-site.xml**

| sudo nano /home/imran/hadoop/etc/hadoop/**yarn-site.xml** |
| --- |

| <?xml version="1.0" encoding="UTF-8"?>  <configuration>  <!-- Site specific YARN configuration properties -->  <property>  <name>yarn.nodemanager.aux-services</name>  <value>mapreduce\_shuffle</value>  </property>  <property>  <name>yarn.log-aggregation-enable</name>  <value>true</value>  </property>  <property>  <name>yarn.nodemanager.log-dirs</name>  <value>/tmp</value>  </property>  <property>  <name>yarn.log.dir</name>  <value>/tmp</value>  </property>  <property>  <name>yarn.log.server.url</name>  <value>http://osboxes:19888/jobhistory/logs</value>  </property>  <property>  <name>yarn.nodemanager.delete.debug-delay-sec</name>  <value>1200</value>  </property>  <property>  <name>yarn.nodemanager.resource.memory-mb</name>  <value>3072</value>  </property>  <property>  <name>yarn.scheduler.maximum-allocation-mb</name>  <value>2048</value>  </property>  </configuration> |
| --- |

To save the changes

| Control +S  Control +X |
| --- |

Now close the terminal and open it again. And run below commands

6.Open and add JAVA\_HOME in **hadoop-env.sh**

| sudo nano /home/imran/hadoop/etc/hadoop/hadoop-env.sh |
| --- |

| export JAVA\_HOME=/opt/software/jsdk |
| --- |

To save the changes

| Control +S  Control +X |
| --- |

7. Open and add JAVA\_HOME in **mapred-env.sh**

| sudo nano /home/imran/hadoop/etc/hadoop/mapred-env.sh |
| --- |

| export JAVA\_HOME=/opt/software/jsdk |
| --- |

To save the changes

| Control +S  Control +X |
| --- |

Run below commands

**hadoop namenode -format**

**start-dfs.sh**

**start-yarn.sh**

Copy below link in VM/ubuntu browser

For YARN UI

http://localhost:8088

For HDFS UI

http://localhost:9870/

After the completion run command

**jps**

You easily can see hadoop daemons running

We are done with the installation…… Good Luck