Hadoop Installation Process

we are using here ubuntu may be you are using Window machine.

I suggest either completely shift on ubuntu or used virtual box. I am assuming you are using window machine.

please follow the following step to setup Hadoop.

1) Download Virtual Box

https://www.virtualbox.org/wiki/Downloads

2) Download Ubuntu

https://ubuntu.com/download/desktop#newsletter-signup

3) Install ubuntu on virtual box

4) by default you will get Mozilla Firefox

5) Download JDK

https://www.oracle.com/in/java/technologies/downloads/

6) Download Hadoop

https://downloads.apache.org/hadoop/common/hadoop-3.3.6/

7) setup env variable (we will see in the class how to do that)

Hadoop File Configurations

```
sudo apt-get install vim (Install USER Friendly Editer)

nano .bashrc (Set the java Path in your Home Path)

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export HADOOP_HOME=/home/nooman/hadoop-3.3.6
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
```

```
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export HADOOP_YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$JAVA_HOME/bin:$HADOOP_HOME/bin:$PATH
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/nati
ve"
echo "welcome"




 source .bashrc (Execute the bashrc file)
 echo JAVA_HOME (Check the java path)

================================================================
================================================================
===============


2. Modify Hadoop Configuration Files
NAMENODE ----> core-site.xml
RESOURCE MANGER ----> mapperd-site.xml
SECONDARYNAMENODE ---->
DATANODE ----> slaves
NODEMANGER ----> slaves & yarn-site.xml


nano etc/hadoop/core-site.xml

<property>
<name>fs.default.name</name>
<value>hdfs://localhost:50000</value>
</property>

nano etc/hadoop/yarn-site.xml
```

```
<property>
<name>yarn.nodemanager.aux-services</name> <value>mapreduce_s
huffle</value>
</property>
<property>
<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</
name> <value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
<property>
<description>The hostname of the RM.</description>
<name>yarn.resourcemanager.hostname</name>
<value>localhost</value>
</property>
<property>
<description>The address of the applications manager interfac
e in the RM.</description>
<name>yarn.resourcemanager.address</name>
<value>localhost:8032</value>
</property>

nano etc/hadoop/hdfs-site.xml

<property>
<name>dfs.namenode.name.dir</name>
<value>/home/username/hadoop2-dir/namenode-dir</value>
</property>
<property>
<name>dfs.datanode.data.dir</name>
<value>/home/username/hadoop2-dir/datanode-dir</value>
</property>

nano etc/hadoop/mapred-site.xml

<property>
<name>mapreduce.framework.name</name>
```

```
<value>yarn</value>
</property>

nano etc/hadoop/hadoop-env.sh
export JAVA_HOME=/home/username/jdk1.8.0_45

nano etc/hadoop/mapred-env.sh
export JAVA_HOME=/home/username/jdk1.8.0_ 45

nano etc/hadoop/yarn-env.sh
export JAVA_HOME=/home/username/jdk1.8.0_45

nano etc/hadoop/slaves
localhost

Install the ssh key
(Generates, Manages and Converts Authentication keys)
 sudo apt-get install openssh-server
 ssh-keygen -t rsa
(Setup passwordless ssh to localhost and to slaves )
 cd .ssh
 ls
 cat id_rsa.pub >> authorized_keys (copy the .pub)
(Copy the id_rsa.pub from NameNode to authorized_keys in all
machines)
 ssh localhost
(Asking No Password )

============================================================
============================================================
===============



3. Format NameNode
 cd hadoop-2.9.1
 bin/hadoop namenode -format (Your Hadoop File System Ready)
```

```
================================================================
================================================================
===============

4. Start All Hadoop Related Services
 sbin/start-all.sh
(Starting Daemon’s For DFS & YARN)
NameNode
DataNode
SecondaryNameNode
ResourceManager
NodeManager


(check the Browser Web GUI )
NameNode - http://localhost:50070/
Resource Manager - http://localhost:8088/

================================================================
================================================================
===============

5.Stop All Hadoop and Yarn Related Services
 sbin/stop-all.sh
```