-- Install single node hadoop cluster (Hadoop\_ver2.7.0)

Pre-requisites:

1. Ubuntu OS image desktop 18.04, 20.04 or 22.04
2. VMware Workstation or VMware player /virtualbox
3. (virtualization) from the CPU should be enabled.

// Install jdk

sudo apt-get install default-jdk

sudo apt install openjdk-11-jdk

// check java version

Java -version

// check the installed java path

Update-alternatives --config java

//ssh install

sudo apt-get install ssh

// append the generated public keys from id\_rsa.pub to authorized\_keys

ssh-keygen -t dsa

// Copy the keygen into the authorized\_keys folder

cat ~/.ssh/id\_dsa.pub >> ~/.ssh/authorized\_keys

// Download Apache Hadoop

Hadoop version 2.7.0

// check localhost

Ssh localhost

// Install the configs

Sudo vim ~/.bashrc

Go to the end of the file of bashrc and apply the following config

Export JAVA\_HOME=”<your\_java\_path>”

Export HADOOP\_HOME=”/usr/local/Hadoop

Export PATH = $PATH:$HADOOP\_HOME/bin

Export PATH = $PATH:$HADOOP\_HOME/sbin

Export HADOOP\_MAPRED\_HOME = $HADOOP\_HOME

Export HADOOP\_COMMON\_HOME=$HADOOP\_HOME

Export HADOOP\_HDFS\_HOME=$HADOOP\_HOME

Export YARN\_HOME=$HADOOP\_HOME

Export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_HOME/lib/native

Export HADOOP\_OPTS=”-Djava.library.path=$HADOOP\_HOME/lib”

//Save and exit

Source ~/.bashrc

Perform the following tasks

// Move and copy the mapred-site.xml.template to mapred-site.xml

Cd $HADOOP\_HOME/etc/Hadoop

Sudo cp mapred-site.xml.template mapred-site.xml

<configuration>

<property>

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

<value>yarn</value>

</property>

</configuration>

// open yarn-site.xml

<configuration>

<property>

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

<value>mapreduce\_shuffle</value>

</property>

//update core-site.xml

<property>

<name>fs.default.name</name>

<value>hdfs://127.0.0.1:9000</value>

</property>

</configuration>

//update hdfs-site.xml

<configuration>

<property>

<name>dfs.replication</name>

<value>1 </value>

</property>

<property>

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

<value>file://home/ani/hadoop/hdfs/namenode</value>

</property>

<property>

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

<value>file://home/ani/hadoop/hdfs/namenode/datanode</value>

</property>

</configuration>

// Open the Hadoop env shell & update java\_home

Export JAVA\_HOME=<your\_java\_path>

// create the local directory for namenode and datanode

Mkdir -p /home/ani/Hadoop/hdfs/namenode

Mkdir -p /home/ani/Hadoop/hdfs/namenode/datanode

// Give permission

Sudo chown -R ani:ani /usr/local/Hadoop

// format the namenode

Hdfs namenode -format

// start the Hadoop services

Start-dfs.sh

Start-yarn.sh

//Check out the status

jps

// access hadoop UI from browser

Localhost:50070

http://localhost:8088 - YARN resource manager

// Make HDFS directories required to execute Mapreduce jobs

Hdfs dfs -mkdir /user/Hadoop/input

Hdfs dfs -put data.txt /user/Hadoop/input

// Apply wordcount sample

Cd #HADOOP\_HOME

Hadoop jar share/Hadoop/mapreduce/Hadoop-mapreduce-examples-2.7.0.jar wordcount /user/Hadoop/input /user/Hadoop/output

Hdfs dfs -ls /user/Hadoop/output

//check the contents

Hdfs dfs -cat /user/Hadoop/output/part-r-00000

Create jar file from java

Jar cf wc.jar input-files