1.Create two directories in /usr/local

hadoop@Hadoop-Dev:~$ sudo chmod 777 /usr/local

[sudo] password for hadoop:

hadoop@Hadoop-Dev:~$ cd /usr/local

hadoop@Hadoop-Dev:/usr/local$mkdir software

hadoop@Hadoop-Dev:/usr/local$mkdir hadoop-ecosystem

2.Update Ubuntu

hadoop@Hadoop-Dev:~$ sudo apt-get update

3.Install Java

hadoop@Hadoop-Dev:~$ sudo apt-get install openjdk-8-jdk

4.Install ssh

hadoop@Hadoop-Dev:~$ sudo apt-get install ssh

5. Download Hadoop3.2.1(hadoop-3.2.1.tar.gz) software from https://archive.apache.org/dist/hadoop/core/hadoop-3.2.1/ and move it to /usr/local.software directory.

6. Move the file hadoop-3.2.1.tar.gz to /usr/local/hadoop-ecosystem directory and extract it.

7. Goto /usr/local/hadoop-ecosystem/hadoop-3.2.1/etc/hadoop and copy paste the file mapred-site.xml.template and rename it as mapred-site.xml

8.Goto Terminal and edit bashrc file.

hadoop@Hadoop-Dev:/usr/local$ sudo gedit ~/.bashrc

9. Enter the below paths in .bashrc file

#JAVA path configuration

export JAVA\_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64

export PATH=$PATH:$JAVA\_HOME/bin

#HADOOP path configuration

export HADOOP\_HOME=/usr/local/hadoop-ecosystem/hadoop-3.2.1

#export HADOOP\_PREFIX=/usr/local/hadoop-ecosystem/hadoop-3.2.1

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\_CONF\_DIR=${HADOOP\_HOME}/etc/hadoop

# Native Path

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=${HADOOP\_HOME}/lib/native

export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_PREFIX/lib"

export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_COMMON\_LIB\_NATIVE\_DIR"

# Add Hadoop bin/ directory to PATH

export PATH=$PATH:$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin

10. Configuring the HDFS files.Goto /usr/local/hadoop-ecosystem/hadoop-2.9.2/etc/hadoop

They are:

core-site.xml

hdfs-site.xml

mapred-site.xml

hadoop-env.sh

yarn-site.xml

core-site.xml:

--------------

<configuration>

<property>

<name>hadoop.tmp.dir</name>

<value>/usr/local/hadoop-ecosystem/hadoop\_data/tmp</value>

<description> Parent directory for other temporary directories </description>

</property>

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

<description> The name of the default file system </description>

</property>

</configuration>

hdfs-site.xml:

---------------

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

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

<value>/usr/local/hadoop-ecosystem/hadoop\_data/hdfs/namenode</value>

</property>

<property>

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

<value>/usr/local/hadoop-ecosystem/hadoop\_data/hdfs/datanode</value>

</property>

</configuration>

mapred-site.xml:

----------------

<configuration>

<property>

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

<value>yarn</value>

</property>

</configuration>

yarn-site.xml:

-----------------

<configuration>

<!-- Site specific YARN configuration properties -->

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>

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

</property>

<property>

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

<value>localhost:8088</value>

</property>

</configuration>

hadoop-env.sh:

--------------

Just we need to update JAVA\_HOME(/usr/lib/jvm/java-1.8.0-openjdk-amd64)

11. To create a password less startup

$sudo apt-get install openssh-server openssh-client

$ssh-keygen -t rsa -P ""

Enter file in which to save the key (/home/hadoop/.ssh/id\_rsa): <just enter>

$cat /home/hadoop/.ssh/id\_rsa.pub >> /home/hadoop/.ssh/authorized\_keys

$sudo chmod 0650 /home/hadoop/.ssh/authorized\_keys

12.Format the disk to hdfs format.

$hdfs namenode -format

13. Start hdfs and yarn

$start-dfs.sh && start-yarn.sh

14. All the below mentioned daemons or processes should be up and running when you execute jps command with some process id’s.

$jps

13536 NameNode

14257 NodeManager

13715 DataNode

14101 ResourceManager

15097 Jps

13903 SecondaryNameNode