**222BDA41**

**HADOOP INSTALLATION CODES**

Oracle Virtual Box - https://www.virtualbox.org/wiki/Downloads

Ubuntu Desktop LTS - https://ubuntu.com/download/desktop

Setup parameters -

# Ip address

ifconfig

sudo apt update

# Java

sudo apt install openjdk-8-jdk

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

nano ~/.bashrc

export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64

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

source .bashrc

echo $JAVA\_HOME

java -version #1.8

# Passwordless ssh

ssh localhost

sudo apt-get install openssh-server openssh-client

ssh localhost

ssh-keygen -t rsa -P ""

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

ssh localhost

exit

# Hadoop

mkdir -p course/softwares

cd course/softwares

wget https://dlcdn.apache.org/hadoop/common/hadoop-3.3.4/hadoop-3.3.4.tar.gz

mv hadoop-3.3.4.tar.gz course/softwares

tar -xzvf hadoop-3.3.4.tar.gz

'''

Local (or Standalone) mode: There are no daemons and everything runs on a single JVM.

Pseudo-Distributed mode: Each daemon(Namenode, Datanode etc) runs on its own JVM on a single host.

Distributed mode: Each Daemon run on its own JVM across a cluster of hosts

'''

# Stand alone mode

ls

nano ~/.bashrc

export HADOOP\_HOME=$HOME/Desktop/jes/softwares/hadoop-3.3.4

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

source .bashrc

hadoop version

# Word count problem in standalone mode

ls $HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.4.jar

mkdir wordcountex

# add text files inside that folder or cp $HADOOP\_HOME/\*.txt wordcountex

jar tf $HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.4.jar | grep wordcount -i

hadoop jar $HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.4.jar wordcount wordcountex texts\_output

cat texts\_output/part-r-00000 | sort -k 2 -nr | head n -5

# Pseudo Distributed mode

nano ~/.bashrc

export HADOOP\_HOME=$HOME/Desktop/jes/softwares/hadoop-3.3.4

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

export HADOOP\_HDFS\_HOME=$HADOOP\_HOME

export HADOOP\_MAPRED\_HOME=$HADOOP\_HOME

export HADOOP\_COMMON\_HOME=$HADOOP\_HOME

export HADOOP\_YARN\_HOME=$HADOOP\_HOME

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

export HADOOP\_INSTALL=$HADOOP\_HOME

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

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

export HADOOP\_SECURITY\_CONF\_DIR

source ~/.bashrc

hadoop version

cd $HADOOP\_HOME/etc/hadoop

nano hadoop-env.sh

export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64

nano core-site.xml

<configuration>

<property>

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

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

</property>

</configuration>

nano hdfs-site.xml

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

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

<value>file:///home/jesse/hadoopinfra/hdfs/namenode </value>

</property>

<property>

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

<value>file:///home/jesse/hadoopinfra/hdfs/datanode </value >

</property>

</configuration>

hadoop classpath

/home/jesse/Desktop/jes/softwares/hadoop-3.3.4/etc/hadoop:/home/jesse/Desktop/jes/softwares/hadoop-3.3.4/share/hadoop/common/lib/\*:/home/jesse/Desktop/jes/softwares/hadoop-3.3.4/share/hadoop/common/\*:/home/jesse/Desktop/jes/softwares/hadoop-3.3.4/share/hadoop/hdfs:/home/jesse/Desktop/jes/softwares/hadoop-3.3.4/share/hadoop/hdfs/lib/\*:/home/jesse/Desktop/jes/softwares/hadoop-3.3.4/share/hadoop/hdfs/\*:/home/jesse/Desktop/jes/softwares/hadoop-3.3.4/share/hadoop/mapreduce/\*:/home/jesse/Desktop/jes/softwares/hadoop-3.3.4/share/hadoop/yarn:/home/jesse/Desktop/jes/softwares/hadoop-3.3.4/share/hadoop/yarn/lib/\*:/home/jesse/Desktop/jes/softwares/hadoop-3.3.4/share/hadoop/yarn/\*

nano yarn-site.xml

<configuration>

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.application.classpath</name>

<value>output from hadoop classpath</value>

</property>

</configuration>

nano mapred-site.xml

<configuration>

<property>

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

<value>yarn</value>

</property>

<property>

<name>mapreduce.reduce.env</name>

<value>HADOOP\_MAPRED\_HOME=$HOME/Desktop/jes/softwares/hadoop-3.3.4</value>

</property>

<property>

<name>yarn.app.mapreduce.am.env</name>

<value>HADOOP\_MAPRED\_HOME=$HOME/Desktop/jes/softwares/hadoop-3.3.4</value>

</property>

<property>

<name>mapreduce.map.env</name>

<value>HADOOP\_MAPRED\_HOME=$HOME/Desktop/jes/softwares/hadoop-3.3.4</value>

</property>

</configuration>

cd ~

hdfs namenode -format

start-dfs.sh

start-yarn.sh

# Port to access Hadoop

http://localhost:9870/

http://localhost:8088/

stop-yarn.sh

stop-dfs.sh

# Word count problem in psuedo distributed mode

ls $HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.4.jar

mkdir wordcountex

# add text files inside that folder or cp $HADOOP\_HOME/\*.txt wordcountex

jar tf $HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.4.jar | grep wordcount -i

# add files to hdfs

hdfs dfs -mkdir -p /user/input/wcexample

hdfs dfs -put wordcountex/\*.txt /user/input/wcexample

# run wc from hdfs files and save into hdfs

hadoop jar $HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.4.jar wordcount /user/input/wcexample /user/output/ex2

hdfs dfs -cat /user/output/ex2/part-r-00000 | sort -k 2 -nr | head -n5