# HADOOP

# Ip address

ifconfig

sudo apt update

# Java

sudo apt install openjdk-8-jdk

ls /usr/lib/anjaliasunil/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 Desktop/course/softwares

cd Desktop/course/softwares

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

tar -xzvf hadoop-3.3.4.tar.gz

# Stand alone mode

ls

nano ~/.bashrc

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

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

source ~/.bashrc

hadoop version

nano ~/.bashrc

export HADOOP\_HOME=$HOME/Desktop/course/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/anjaliasunil/hadoopinfra/hdfs/namenode </value>

</property>

<property>

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

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

</property>

</configuration>

hadoop classpath

nano yarn-site.xml

<configuration>

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<property>

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

<value>/home/anjaliasunil/Desktop/course/softwares/hadoop-3.3.4/etc/hadoop:/home/anjaliasunil/Desktop/course/softwares/hadoop-3.3.4/share/hadoop/common/lib/\*:/home/anjaliasunil/Desktop/course/softwares/hadoop-3.3.4/share/hadoop/common/\*:/home/anjaliasunil/Desktop/course/softwares/hadoop-3.3.4/share/hadoop/hdfs:/home/anjaliasunil/Desktop/course/softwares/hadoop-3.3.4/share/hadoop/hdfs/lib/\*:/home/anjaliasunil/Desktop/course/softwares/hadoop-3.3.4/share/hadoop/hdfs/\*:/home/anjaliasunil/Desktop/course/softwares/hadoop-3.3.4/share/hadoop/mapreduce/\*:/home/anjaliasunil/Desktop/course/softwares/hadoop-3.3.4/share/hadoop/yarn:/home/anjaliasunil/Desktop/course/softwares/hadoop-3.3.4/share/hadoop/yarn/lib/\*:/home/anjaliasunil/Desktop/course/softwares/hadoop-3.3.4/share/hadoop/yarn/\*</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/course/softwares/hadoop-3.3.4</value>

</property>

<property>

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

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

</property>

<property>

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

<value>HADOOP\_MAPRED\_HOME=$HOME/Desktop/course/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

# HIVE

cd Desktop/course/softwares

wget https://apache.osuosl.org/hive/hive-3.1.2/apache-hive-3.1.2-bin.tar.gz

tar -xzf apache-hive-3.1.2-bin.tar.gz

nano ~/.bashrc

export HIVE\_HOME=$HOME/Desktop/course/softwares/hive

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

export CLASSPATH=$CLASSPATH:$HADOOP\_HOME/lib/\*:$HIVE\_HOME/lib/\*

source ~/.bashrc

cd $HIVE\_HOME/conf

nano hive-env.sh

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

# Metastore - Apache Derby configuration

cd Desktop/course/softwares

wget http://archive.apache.org/dist/db/derby/db-derby-10.4.2.0/db-derby-10.4.2.0-bin.tar.gz

tar zxvf db-derby-10.4.2.0-bin.tar.gz

mv db-derby-10.4.2.0-bin derby

nano ~/.bashrc

export DERBY\_HOME=$HOME/Desktop/course/softwares/derby

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

export CLASSPATH=$CLASSPATH:$DERBY\_HOME/lib/derby.jar:$DERBY\_HOME/lib/derbytools.jar

source ~/.bashrc

mkdir $DERBY\_HOME/data

# Configure Derby for Hive

cd $HIVE\_HOME/conf

nano hive-site.xml

<configuration>

<property>

<name>javax.jdo.option.ConnectionURL</name>

<value>jdbc:derby:;databaseName=metastore\_db;create=true</value>

<description>JDBC connect string for a JDBC metastore </description>

</property>

</configuration>

nano jpox.properties

javax.jdo.PersistenceManagerFactoryClass =

org.jpox.PersistenceManagerFactoryImpl

org.jpox.autoCreateSchema = false

org.jpox.validateTables = false

org.jpox.validateColumns = false

org.jpox.validateConstraints = false

org.jpox.storeManagerType = rdbms

org.jpox.autoCreateSchema = true

org.jpox.autoStartMechanismMode = checked

org.jpox.transactionIsolation = read\_committed

javax.jdo.option.DetachAllOnCommit = true

javax.jdo.option.NontransactionalRead = true

javax.jdo.option.ConnectionDriverName = org.apache.derby.jdbc.ClientDriver

javax.jdo.option.ConnectionURL = jdbc:derby://hadoop1:1527/metastore\_db;create = true

javax.jdo.option.ConnectionUserName = APP

javax.jdo.option.ConnectionPassword = mine

source ~/.bashrc

# Setup Hive in HDFS

$HADOOP\_HOME/bin/hadoop fs -mkdir /tmp

$HADOOP\_HOME/bin/hadoop fs -mkdir /user

$HADOOP\_HOME/bin/hadoop fs -mkdir /user/hive

$HADOOP\_HOME/bin/hadoop fs -mkdir /user/hive/warehouse

$HADOOP\_HOME/bin/hadoop fs -chmod g+w /tmp

$HADOOP\_HOME/bin/hadoop fs -chmod g+w /user/hive/warehouse

cd $HIVE\_HOME

bin/schematool -initSchema -dbType derby

bin/hive

#PIG

cd Desktop/course/softwares

wget https://archive.apache.org/dist/pig/pig-0.17.0/pig-0.17.0.tar.gz

tar -xzf pig-0.17.0.tar.gz

mv pig-0.17.0 pig

nano ~/.bashrc

export PIG\_HOME=$HOME/Desktop/course/softwares/pig

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

export CLASSPATH=$CLASSPATH:$PIG\_HOME/lib/\*

export PIG\_CLASSPATH=$PIG\_HOME/conf:$HADOOP\_HOME/etc/hadoop

export PIG\_CONF\_DIR=$PIG\_HOME/conf

export PIG\_CLASSPATH=$PIG\_CONF\_DIR:$PATH

source ~/.bashrc

start-dfs.sh

start-yarn.sh

hdfs dfs -put /home/anjaliasunil/hadoopdata/empdata.csv /user/data/empdata\_pig.csv