1. Install Ubuntu
2. sudo apt-get update
3. 登录root
4. 安装SSH 并设置无密码登录

安装SSH

sudo apt-get install ssh openssh-server

启动SSH 服务

设置无密码登录

ssh-keygen -t dsa -P '' -f ~/.ssh/id\_dsa   
cat ~/.ssh/id\_dsa.pub >> ~/.ssh/authorized\_keys

测试无密码登录

ssh localhost

1. 关机防火墙

sudo ufw disable

1. 设置静态IP，重启网卡 service network restart (skip)

验证：主机ping客户机 客户机ping主机

1. 修改hostname

Vi /etc/sysconfig/network

绑定主机名 vi /etc/hosts

Ping hostname

1. 安装jdk

sudo apt-get install openjdk-7-jdk

cd /usr/lib/jvm

sudo ln -s java-7-openjdk-amd64 jdk

1. Add Hadoop Group and User

sudo addgroup hadoop

sudo adduser --ingroup hadoop hduser

sudo adduser hduser sudo

1. 安装hadoop

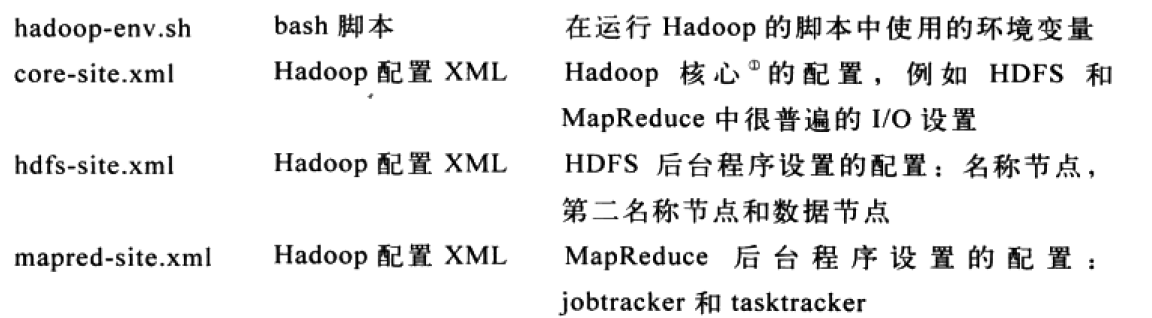
tar -zxvf hadoop-2.7.2.tar.gz

sudo chmod a+w /usr/local

mv hadoop-2.7.2 /usr/local/Hadoop

1. 设置伪分布模式

修改配置文件 HADOOP\_HOME/ etc/hadoop



sudo chmod a+w hadoop-env.sh

sudo chmod a+w core-site.xml

sudo chmod a+w hdfs-site.xml

z@ubuntu:~/hadoop/etc/hadoop$ cp mapred-site.xml.template mapred-site.xml

sudo chmod a+w mapred-site.xml

sudo chmod a+w      ~/.bashrc

sudo chmod a+w /etc/profile

sudo chmod a+w /etc/hosts

sudo chmod a+w /usr/local/hadoop\_store/hdfs

**add to ~/.bashrc /etc/profile**

export JAVA\_HOME=/usr/lib/jvm/jdk/

export HADOOP\_INSTALL=/usr/local/hadoop

export PATH=$PATH:$HADOOP\_INSTALL/bin

export PATH=$PATH:$HADOOP\_INSTALL/sbin

export HADOOP\_MAPRED\_HOME=$HADOOP\_INSTALL

export HADOOP\_COMMON\_HOME=$HADOOP\_INSTALL

export HADOOP\_HDFS\_HOME=$HADOOP\_INSTALL

export YARN\_HOME=$HADOOP\_INSTALL

**add to hadoop-env.sh**

export JAVA\_HOME=/usr/lib/jvm/jdk/

**add to core-site.xml**

Enter the following content in between the tag <configuration></configuration>

<property>

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

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

</property>

**Add to yarn-site.xml**

Enter the following content in between the tag <configuration></configuration>

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>

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

</property>

cp /usr/local/hadoop/etc/hadoop/mapred-site.xml.template /usr/local/hadoop/etc/hadoop/mapred-site.xml

**add to mapred-site.xml**

Enter the following content in between the tag <configuration></configuration>

<property>

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

<value>yarn</value>

</property>

~~<property>~~

~~<name>mapred.job.tracker</name>~~

~~<value>localhost:9001</value>~~

~~</property>~~

mkdir -p /usr/local/hadoop\_store/hdfs/namenode

mkdir -p /usr/local/hadoop\_store/hdfs/datanode

sudo chown hduser:hadoop -R /usr/local/hadoop\_store

sudo chown hduser:hadoop -R /usr/local

sudo chown hduser:hadoop -R /usr/local/hadoop\_store/hdfs

sudo chmod 777 -R /usr/local/hadoop\_store

sudo chmod 777 -R /usr/local

sudo chmod 777 -R /usr/local/hadoop\_store/hdfs

**add to hdfs-site.xml**

Enter the following content in between the tag <configuration></configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

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

<value>file:/usr/local/hadoop\_store/hdfs/namenode</value>

</property>

<property>

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

<value>file:/usr/local/hadoop\_store/hdfs/datanode</value>

</property>

1. Reboot ubuntu
2. Check if Hadoop configured correctly.

**hadoop version**

1. Format Namenode

cd /usr/local/hadoop/

hdfs namenode -format

1. Start Hadoop

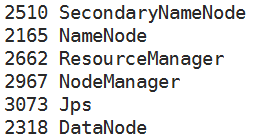
su hduser

/usr/local/hadoop/sbin/start-all.sh

/usr/local/hadoop/sbin/stop-all.sh

1. jps

should be six processes



1. 验证

管理界面：http://localhost:8088

NameNode界面：http://localhost:50070

HDFS NameNode界面：http://localhost:8042

<https://rstudio-pubs-static.s3.amazonaws.com/78508_abe89197267240dfb6f4facb361a20ed.html>

1. 文件（夹）操作

hadoop fs -mkdir /usr

hadoop fs -ls /

hadoop fs -put /root/install.log /usr

配置map reducer 开发环境

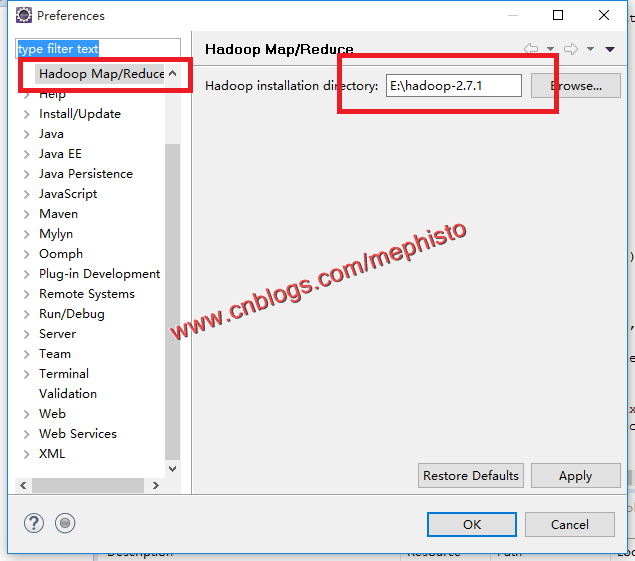
1. 安装eclipse on Ubuntu
2. Copy hadoop-eclipse-plugin-2.6.0.jar to eclipse\ plugins

<https://github.com/winghc/hadoop2x-eclipse-plugin>

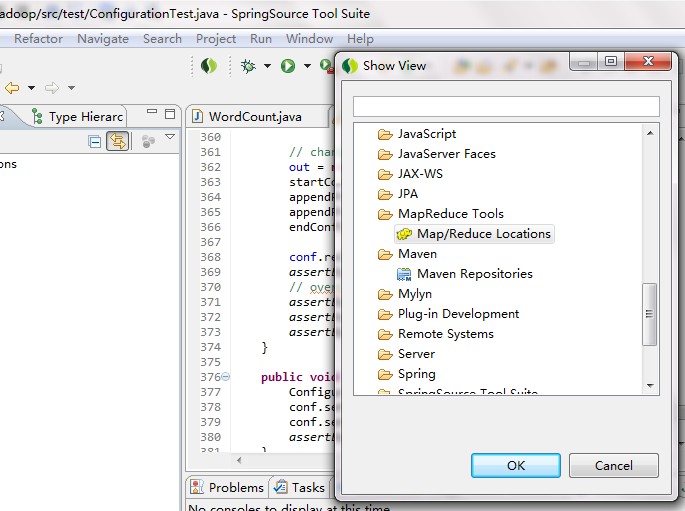
<http://www.cnblogs.com/shitouer/archive/2012/05/29/2522860.html>

进入Windows->Preferences

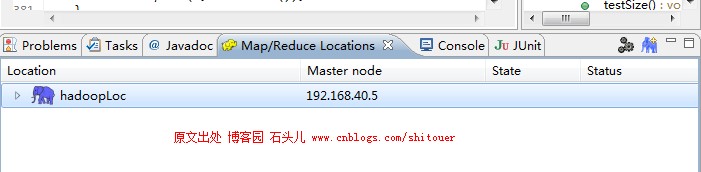
　　选择Hadoop的路径



配置相关参数，点击 Windows-->Show View -->Other --> Map/Reduce Location，打开Hadoop的MapReduce View

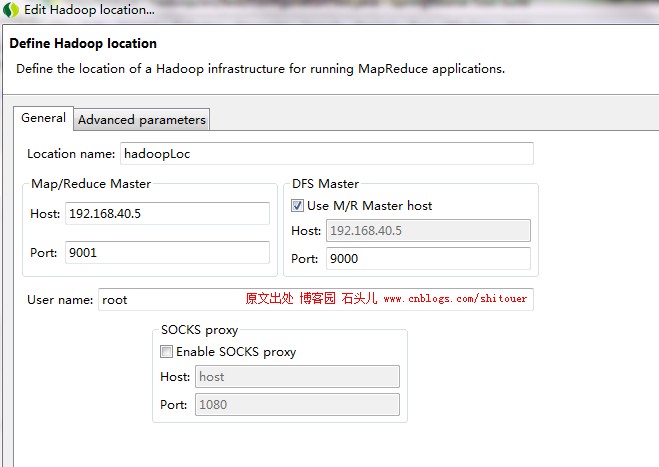


点击Ok之后，出现如下图



⑤ 上一步你不应该看到hadoopLoc, 应该什么都没有，右键点击空白处-->New Hadoop Location, 你会看到一个填写MapReduce Location参数的一个界面：

其中：



Location Name：

这个不用在意，就是对这个MapReduce的标示，只要能帮你记忆即可

Map/Reduce Master 部分相关定义：

Host：上一节搭建的集群中JobTracker所在的机器的IP地址

port：JobTracker的端口

两个参数就是 mapred-site.xml中mapred.job.tracker的ip和端口

DFS Master部分：

Host：就是上一节集群搭建中Namenode所在机器IP

Port：就是namenode的端口

这两个参数是在 core-site.xml里fs.default.name里面的ip和端口

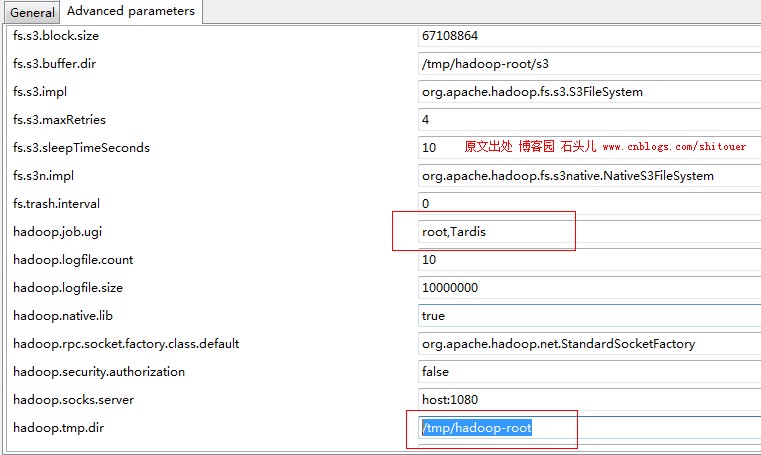
User Name：

就是搭建Hadoop集群是所用的用户名，我这里用的是root

⑥ 填写完以上信息以后，关闭Eclipse，然后重新启动。

这部分仅供参考

可看到大部分选项都已经预设好了，其中有一下几个选项，必须注意填写，其他的选项，要么默认，要么会根据我们填写的以下几个选项自动修改：



dfs.replication:

这个这里默认是3，但是要根据你的Datanode的个数来定，如果你仅有2个datanode，那么就写2，如果只有一个，就写1，大于或者等于3个，写3

hadoop.tmp.dir:

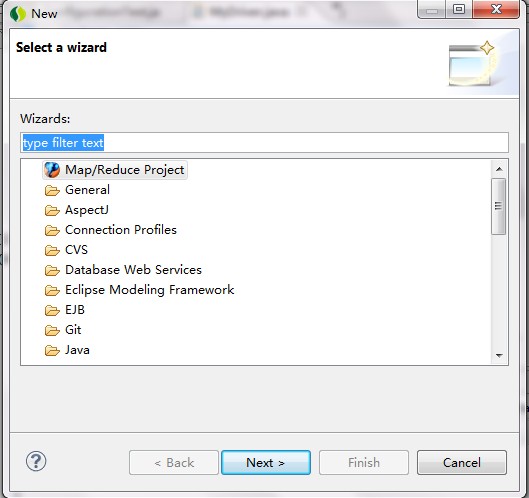
这个要跟上节建立的Hadoop集群配置中core-site.xml 中的hadoop.tmp.dir设置一样，上节填写的是/tmp/hadoop-root，这里也如是填写

hadoop.job.ugi:

填成root,Tardis，其中root是我们的用户名，后边的",Tardis"一定要这么写，不能改

这些都填写完了之后，点击保存，重新启动Eclipse，会发现，很多Advanced Parameters里的好多选项都根据hadoop.tmp.dir重新进行了设置。

建立Map/Reduce 工程



安装map/reducer 示例程序

在HDFS的/HadoopRoot/WordCount/in 中新建文件1, 内容如下

A

B

C

D

A

C

上传文件hadoop fs –put 1 /HadoopRoot/WordCount/in/1

执行hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar wordcount /HadoopRoot/WordCount/in /HadoopRoot/WordCount/out

查看执行结果hadoop fs -text /HadoopRoot/WordCount/out/out1/part-r-00000