第一天:

1. 安装虚拟机
2. 安装centos环境
3. 修改ip地址为192.168.93.10
4. 到/etc/sysconfig/network文件里面添加主机名



1. 关闭防火墙: service iptables stop //临时关闭防火墙 chkconfig iptables off && setenforce 0//永久关闭防火墙及军用级防火墙
2. yum install -y lrzsz //安装上传、下载服务
3. yum -y install ntpdate
4. ntpdate -u ntp.api.bz
5. date "+%Y-%m-%d %H:%M:%S"
6. 添加用户名和密码adduser xulei passwd xulei 接着输入123456这个密码

第二天:

1.新建三个包分别如下

/server 软件安装目录

/soft 安装包

/datas/hadoopdata hadoop数据存放目录



2.把文件上传到/soft下面

3. 解压压缩文件tar -zxvf xxxx

4. 移动到指定目录下面 mv ./jdk1.8.0\_141 ~/server/jdk1.8

5. 修改文件 vi .bash\_profile

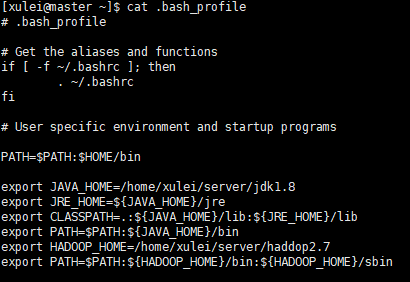
5. 到文件里面添加如下内容

export JAVA\_HOME=/home/xulei/server/jdk1.8

export JRE\_HOME=${JAVA\_HOME}/jre

export CLASSPATH=.:${JAVA\_HOME}/lib:${JRE\_HOME}/lib

PATH=$PATH:${JAVA\_HOME}/bin



6.使刚刚文件起效 source ~/.bash\_profile

7.修改mapred-env.sh ,在# export JAVA\_HOME=/home/y/libexec/jdk1.6.0/下面添加

export JAVA\_HOME=/home/xulei/server/jdk1.8

8. 修改hadoop-env.sh把export JAVA\_HOME=${JAVA\_HOME}改成自己的真实jdk

export JAVA\_HOME=/home/xulei/server/jdk1.8

9. 修改yarn-env.sh设置jdk路径：在# some Java parameters # export JAVA\_HOME=/home/y/libexec/jdk1.6.0/下面添加export JAVA\_HOME=/home/xulei/server/jdk1.8

10.修改文件

a.修改core-site.xml为如下内容

<configuration>

<!-- HDFS的默认路径，端口9000 -->

<property>

<name>fs.defaultFS</name>

<value>hdfs://master:9000</value>

</property>

<!-- hadoop运行时，存放数据的临时文件的路径 -->

<property>

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

<value>/home/xulei/datas/hadoopdata</value>

</property>

</configuration>

b.hdfs-site.xml：

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

</configuration>

c.yarn-site.xml：

<?xml version="1.0"?>

<configuration>

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<property>

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

<value>master:18040</value>

</property>

<property>

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

<value>master:18030</value>

</property>

<property>

<name>yarn.resourcemanager.resource-tracker.address</name>

<value>master:18025</value>

</property>

<property>

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

<value>master:18141</value>

</property>

<property>

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

<value>master:18088</value>

</property>

</configuration>

d. cp mapred-site.xml.template mapred-site.xml

mapred-site.xml.template

<?xml version="1.0"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<configuration>

<property>

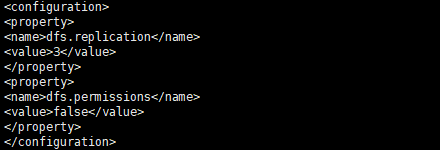
<name>mapreduce.framework.name</name>

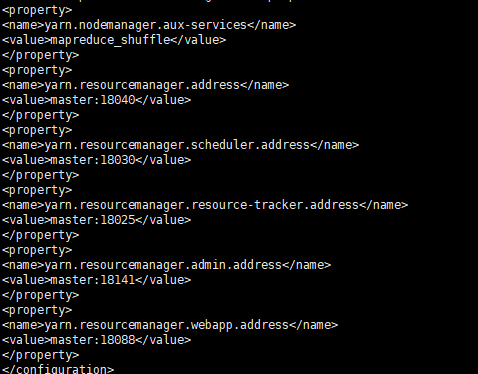
<value>yarn</value>

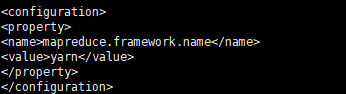
</property>

</configuration>









12. 第一次使用hadoop前 hadoop namenode –format

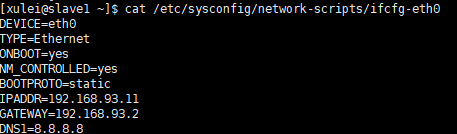
13. start-all.sh

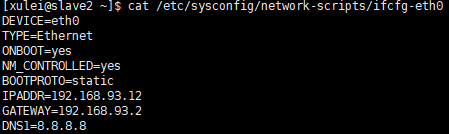
14.jps

第三天:

1.克隆两个从节点slave1: 192.168.93.11和slave2:192.168.93.12

2.到两个从节点的/etc/sysconfig/network-scripts/ifcfg-eth0文件里面修改ip和删除UUID/HWADDR这两行





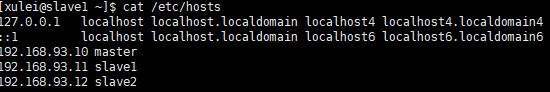
3. rm -rf /etc/udev/rules.d/70-persistent-net.rules

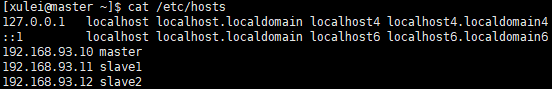
4. reboot

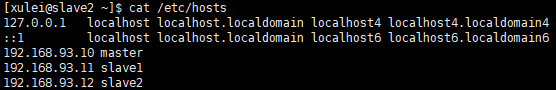
5.到两个从节点的/etc/sysconfig/network文件里面添加主机名





6. 到三个节点的/etc/hosts文件里面添加节点ip与主机名映射





7.步骤六可以用scp /etc/hosts slave1:/etc/ 从一个主机拷到另一个主机

8.在master节点生成公钥，分别拷贝到自己和其他两个从节点的~/.ssh/ authorized\_keys的文件里面通过下面命令(加密 SSH(DES、RSA(公钥，私钥)) master 参数一个公钥和私钥 把公钥派发给其他所有从节点下面免密用root用户)

a. ssh-keygen -t rsa (一系列回车)

b. cp ~/.ssh/id\_rsa.pub ~/.ssh/authorized\_keys

d. ssh-copy-id -i slave2

f. ssh-copy-id -i slave1

e. 远程登录到slave1 🡪 ssh slave1

9.在三个节点修改~/server/haddop2.7/etc/hadoop/slaves







10. 删除所有节点的日志和数据，输入下面命令

a. rm -rf ~/datas/hadoopdata/\*

b. rm -rf ~/server/hadoop2.7/logs/\*

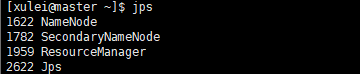
c. date "+%Y-%m-%d %H:%M:%S"

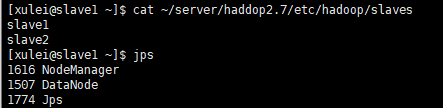
d. hadoop namenode –format

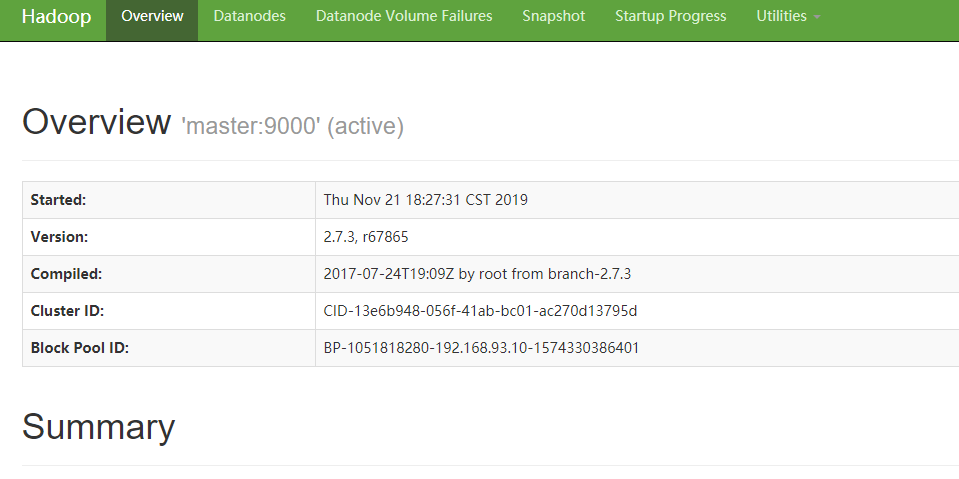
e. start-all.sh

f.jps

g.master的ip:50070







第四天：

一: hdfs 放到hdfs存储的文件，不适合修改，最多式追加

1.Hadoop fs -ls /

2.hadoop fs -mkdir /words

3.hadoop fs -cat /words/aa

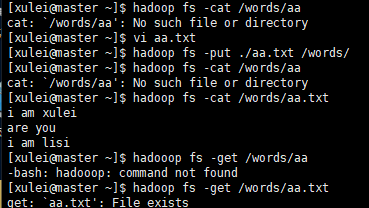
4.hadoop fs -rm /words

5.hadoop fs -cd /words

6.hadoop fs -put ./aa.txt /words/

7.hadoop fs -get /words/aa





二:Map Reduce 解释

slave1

i am zhanghai <i,1> <am,1> <zhanghai,1>

are you <are,1> <you,1>

i am lisi <i,1> <am,1> <lisi,1>

<i,(1,1)> <am,(1,1)><zhanghai,1> <are,1> <you,1><lisi,1>

slave2

you are name am <you,1> <are,1> <name,1> <am,1>

my name is wangwu <my,1> <name,1> <is,1> <wangwu,1>

<you,1> <are,1> <name,1,1> <am,1> <my,1> <is,1> <wangwu,1>

三:Window上面通过java通信

1.把 D:\作业\大数据\app\winutils-master\hadoop-2.7.1\bin 下面的 hadoop.dll 和 winutils.xml 拷贝到

D:\Program Files\hadoop-2.7.3\bin 和 C:\Windows\System32

2. 配置环境变量 1.HADOOP\_HOME 2. %HADOOP\_HOME%\bin\ 3.%HADOOP\_HOME%\sbin\

3.建工程

D:\Program Files\hadoop-2.7.3\share\hadoop\common 下面的三个jar包 D:\Program Files\hadoop-2.7.3\share\hadoop\common\lib重复D:\Program Files\hadoop-2.7.3\share\hadoop下面的hdfs mapreduce yarn

