**步骤**

**一、编写xsync集群分发脚本：**

在/home/hadoop下创建bin目录，并在bin目录下创建脚本文件—vim xsync

#!/bin/bash

##1 获取输入参数个数，如果没有参数，直接退出

pcount=$#

if((pcount==0)); then

echo no args;

exit;

fi

#2 获取文件名称

p1=$1

fname=`basename $p1`

echo fname=$fname

#3 获取上级目录到绝对路径

pdir=`cd -P $(dirname $p1); pwd`

echo pdir=$pdir

#4 获取当前用户名称

user=`whoami`

#5 循环

for((host=162; host<164; host++)); do

echo -------------------hadoop$host --------------

rsync -rvl $pdir/$fname $user@hadoop$host:$pdir

done

之后修改脚本权限，使其具有执行权限：

chmod 777 xsync

调用脚本形式：

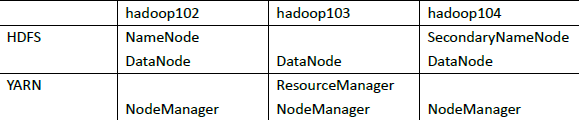
xsync /home/hadoop/bin

执行脚本文件：

xsync bin/

**二、集群文件配置**

集群部署规划：



修改配置文件：

**core-site.xml：**

<!--指定HDFS中NameNode的地址-->

<property>

<name>fs.defaultFS</name>

<value>hdfs://hadoop102:9000</value> — hadoop102根据实际情况修改

</property>

<!--指定Hadoop运行时产生文件的存储目录-->

<property>

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

<value>/opt/module/hadoop-2.7.2/data/tmp</value>

</property>

**hadoop-env.sh：**

export JAVA\_HOME=/opt/module/jdk1.8.0\_144

**hdfs-site.xml：**

<property>

<name>dfs.replication</name>

<value>3</value>

</property>

<!--指定Hadoop辅助名称节点主机配置-->

<property>

<name>dfs.namenode.secondary.http-address</name>

<value>hadoop164:50090</value>

</property>

**yarn-env.sh：**

export JAVA\_HOME=/opt/module/jdk1.8.0\_144

**yarn-site.xml：**

<!--Reducer获取数据的方式-->

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<!--指定YARN的ResourceManager的地址-->

<property>

<name>yarn.resourcemanager.hostname</name>

<value>hadoop103</value>

</property>

**mapred-env.sh：**

export JAVA\_HOME=/opt/module/jdk1.8.0\_144

**mapred-site.xml：**

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

vi mapred-site.xml

<!--指定MR运行在Yarn上-->

<property>

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

<value>yarn</value>

</property>

在集群上分发配置好的Hadoop配置文件：

xsync /hadoop —— 在Hadoop安装目录下的etc/目录下进行

**三、SSH免密登陆配置**

进入到 /home/hadoop/.ssh 目录下

生成公钥和私钥

ssh-keygen -t rsa

将公钥拷贝到免密登陆的目标机子上

ssh-copy-id hadoop162

除了拷贝到其他机子上，本机也要拷贝一份，在其他机子上需要重复一遍以上步骤

还要切换到root用户下在配置一遍

**四、群起集群及配置**

配置文件slaves：(在Hadoop安装目录下的/etc/hadoop文件夹下)

vim slaves

hadoop161

hadoop162

hadoop163

同步所有节点配置文件：

xsync slaves

启动集群：

第一次启动集群格式化namenode：

sbin/hdfs namenode -format

启动HDFS：

sbin/start-dfs.sh

启动YARN：

sbin/start-yarn.sh

NameNode和ResourceManger如果不是同一台机器，不能在NameNode上启动YARN，应该在ResouceManager所在的机器上启动YARN（本机配置的是用第三台启动）

整体停止HDFS：

stop-dfs.sh

整体停止YARN：

stop-yarn.sh

**五、集群基本测试**

上传文件到集群：

bin/hdfs dfs -mkdir -p /user/hadoop/input —— 创建HDFS目录

小文件：

bin/hdfs dfs -put wcinput/wc.input /user/hadoop/input —— 上传到HDFS上  
大文件：

bin/hadoop fs -put /opt/software/hadoop-2.7.2.tar.gz

下载文件：

bin/hadoop fs -get /user/atguigu/input/hadoop-2.7.2.tar.gz ./