Hadoop2 伪分布环境搭建笔记

1. 获取包Hadoop2.4.1和JDK1.7.xx.tar.gz,并将包放置在linux中，修改节点的主机名

#vim /etc/sysconfig/network

设置HOSTNAME=h2m1 🡪 表示hadoop2的第一个master



将ip和主机名绑定

#vim /etc/hosts



1. 解压Hadoop2.4.1和JDK，并放置在/usr/local目录下

#tar –xzvf Hadoop2.4.1.tar.gz

#tar –xzvf jdk1.7.0\_75.tar.gz

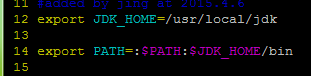
# mv jdk1.7.0\_75 /usr/local/jdk

# mv hadoop-2.4.1 /usr/local/hadoop241

1. 安装配置JDK

添加java的环境变量到系统中

# vim /etc/profile #编辑profile文件

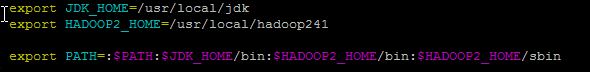


# source /etc/profile #使得环境变量修改生效

#java –version #验证java是否配置成功

1. 安装配置Hadoop2
   1. 将Hadoop的路径添加到环境变量中

# vim /etc/profile #编辑profile，



# source /etc/profile #使得修改生效

* 1. 修改Hadoop配置文件

# vim /usr/local/hadoop241/etc/hadoop/hadoop-env.sh #修改配置

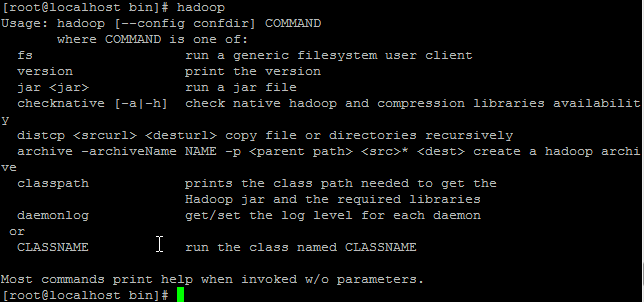
修改如图，指定安装的java主目录



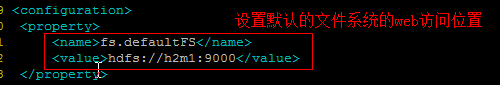
添加如图，指定HADOOP\_PREFIX



#hadoop #执行hadoop脚本，会提示usage信息



* 1. 伪分布式配置
     1. 修改/usr/local/hadoop241/ etc/hadoop/core-site.xml



* + 1. 修改/usr/local/hadoop241/ etc/hadoop/hdfs-site.xml



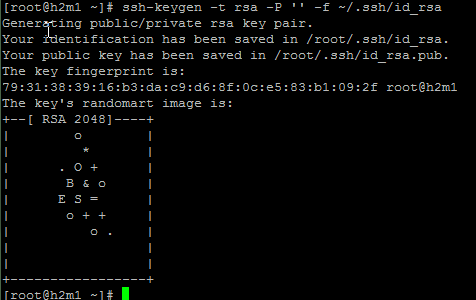
* 1. 配置ssh免密码登录

验证本机是否可以免密码登录

#ssh localhost

如果不能免密码登录，进行如下设置

# ssh-keygen -t rsa -P '' -f ~/.ssh/id\_rsa



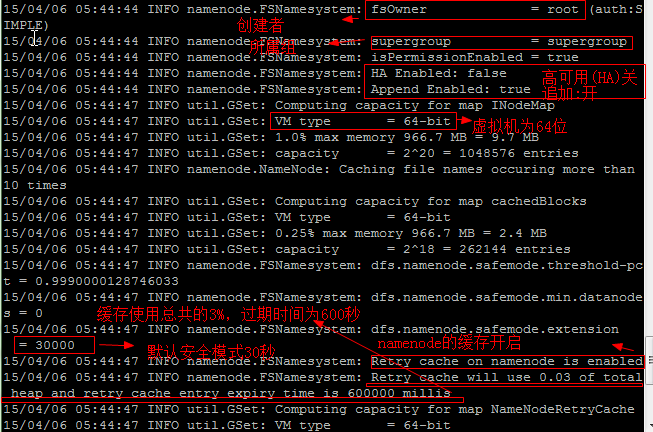
# cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys #生成免密码登录证书



已经可以免密码登录了

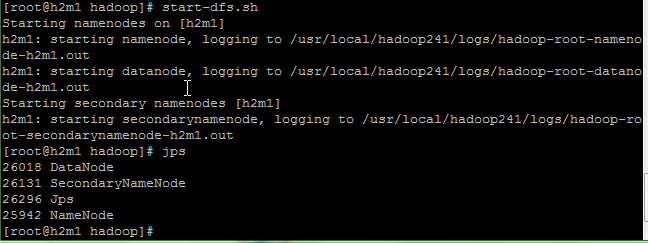
* 1. 在单节点上YARN平台的Job执行
     1. 格式化文件系统

# hdfs namenode -format



* + 1. 启动namenode和datanode

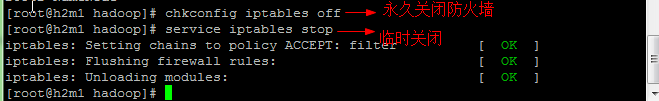
# start-dfs.sh

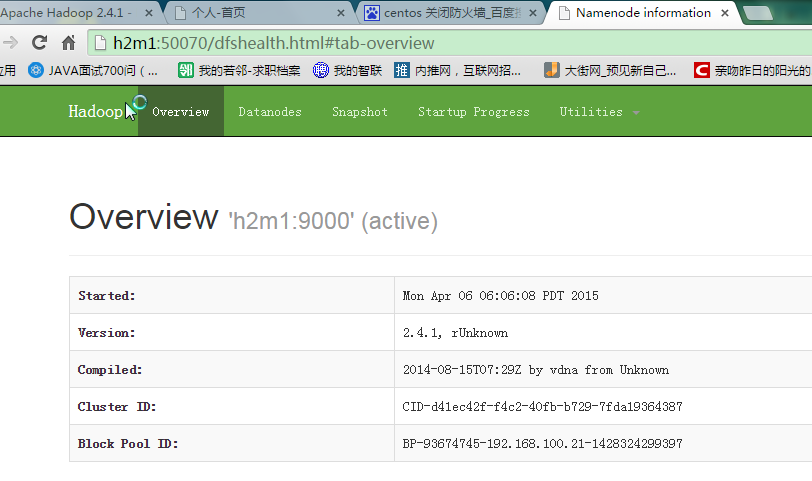


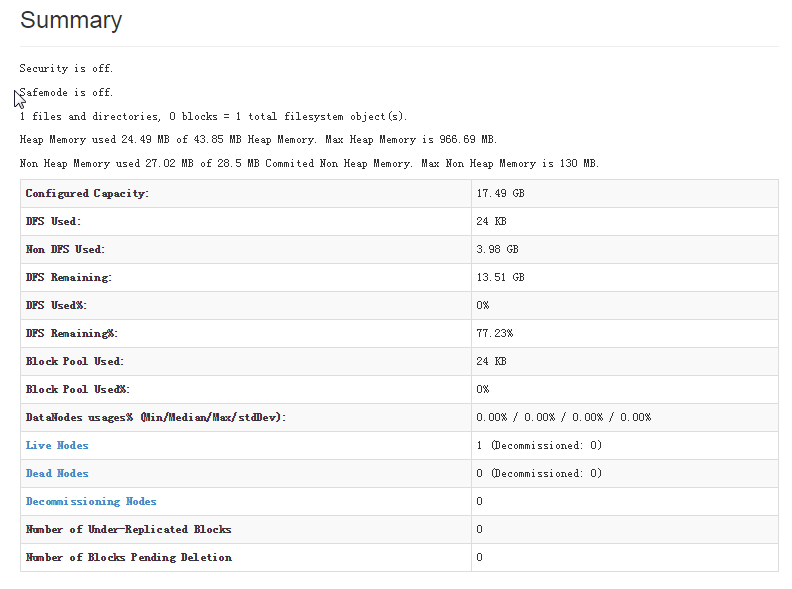
* + 1. 在浏览器上访问<http://h2m1:50070>

”h2m1”必须在浏览器所在的机器上与IP地址绑定

为了访问方便，关闭linux上的防火墙





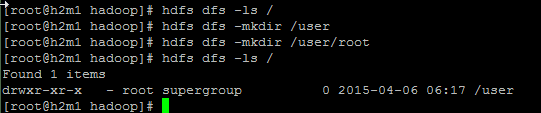




* + 1. 在hdfs上创建目录用来运行MapReduce任务

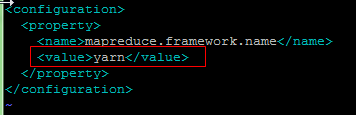
# hdfs dfs -mkdir /user

# hdfs dfs -mkdir /user/root

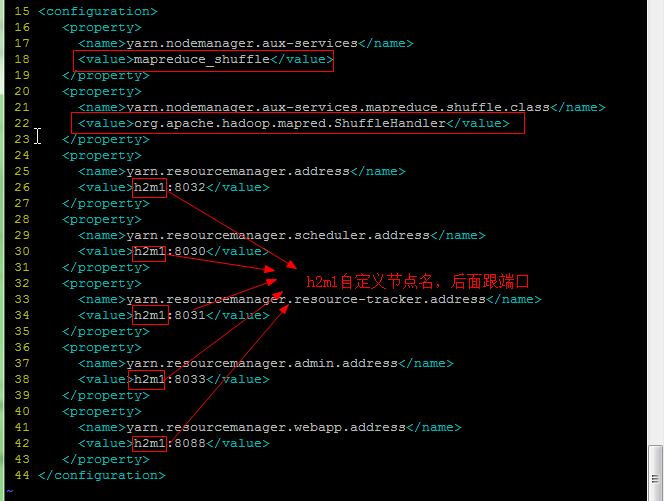


* + 1. 配置YARN

#vim /usr/local/hadoop241/etc/hadoop/mapred-site.xml



#vim /usr/local/hadoop241/etc/hadoop/yarn-site.xml

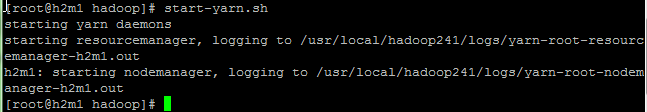


#vim /usr/local/hadoop241/etc/hadoop/yarn-env.sh

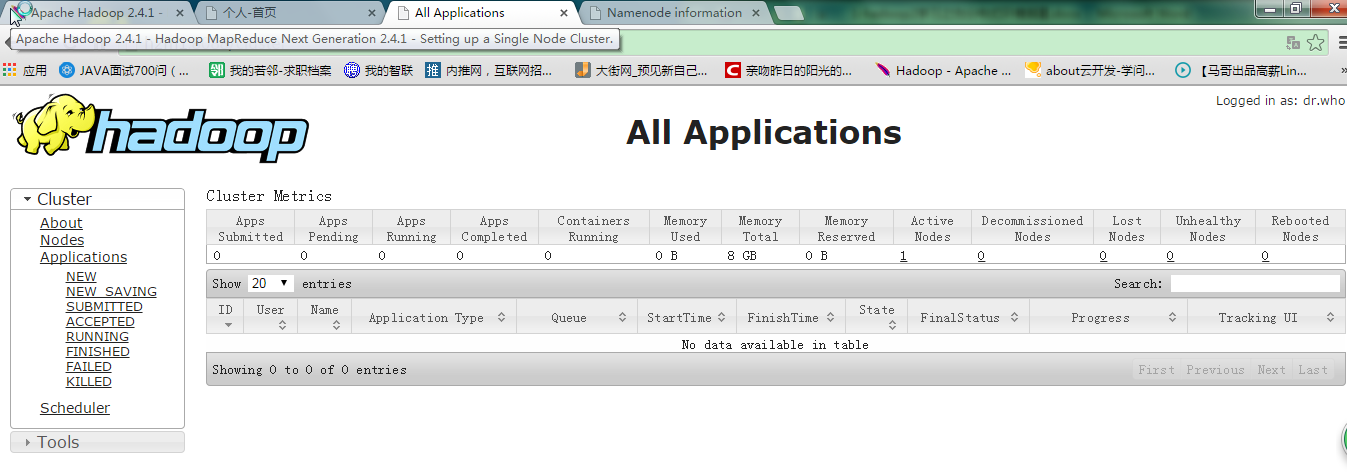


* + 1. 启动ResourceManager和NodeManager

# start-yarn.sh



* + 1. 在浏览器中查看ResourceManager <http://h2m1:8088>



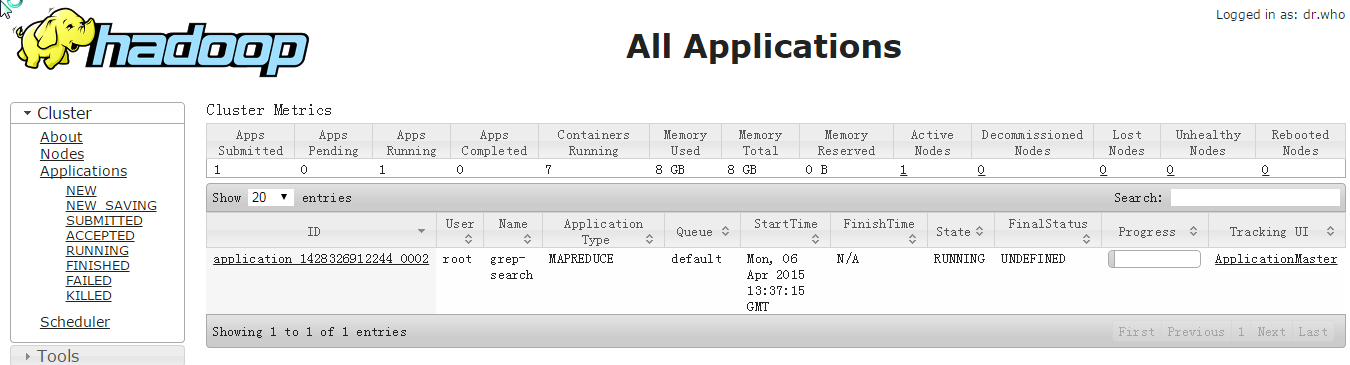
* + 1. 复制输入文件到DFS

# hdfs dfs -put etc/hadoop input

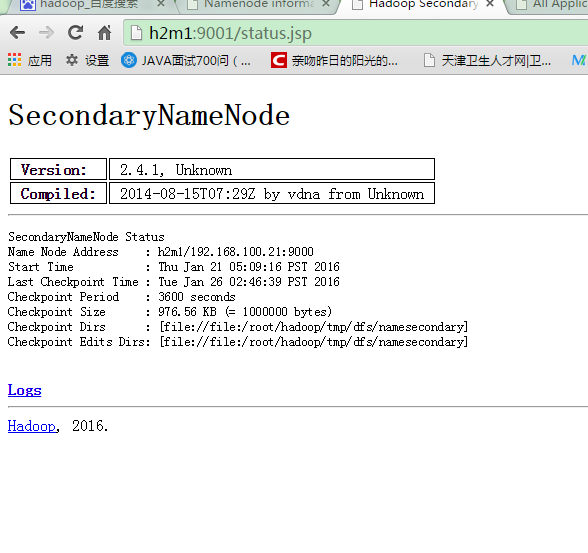
* + 1. 运行hadoop的实例程序

# hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.4.1.jar grep input output 'dfs[a-z.]+'

这是正在执行中，时间太长就给中断掉了，但终端上像是中断了，但实际上还在执行中，仍然占用着资源



* + 1. 浏览器中查看SecondaryNameNode <http://h2m1:9001>

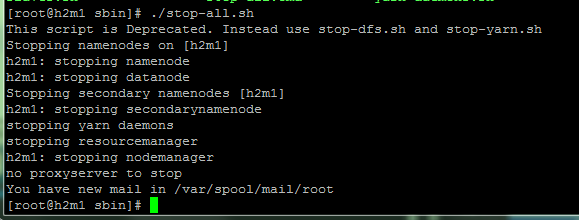


* + 1. 停止DFS服务和YARN服务

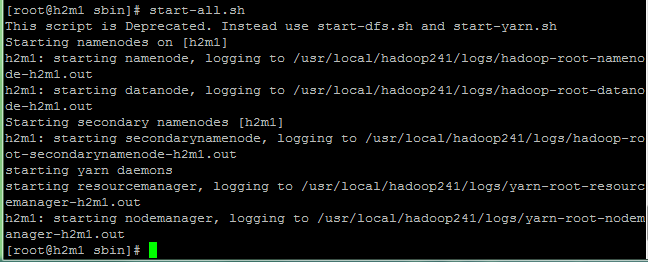
#stop-dfs.sh

#stop-yarn.sh

#stop-all.sh 执行结果

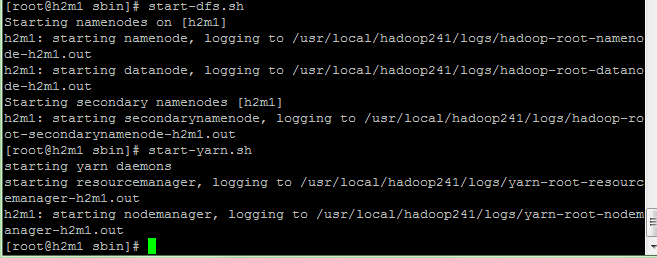


* + 1. 执行start-all.sh脚本



但该脚本已经被声明过时了，因此启动服务使用start-dfs.sh,start-yarn.sh

* + 1. start-dfs.sh和start-yarn.sh执行截图



* + 1. 注意点：

(1).修改了hdfs-site.xml或者core-site.xml文件后，需要重新格式化hdfs。否则50070端口没有显示内容

伪分布式模式的yarn安装就已经完成了