实验题目

[数据存储：HDFS](http://mooc.study.163.com/course/2001353003?tid=2001512000)

实验要求

1、学员链接集群环境

2、通过给定的端口号安装HDFS(注意用给定的端口后，避免产生冲突)

3、为了合理分配资源，在yarn安装环节，请用教学环节中的yarn，不需要单独配置

4、安装完成后，运行hadoop fs  -ls 命令，验证是否安装成功

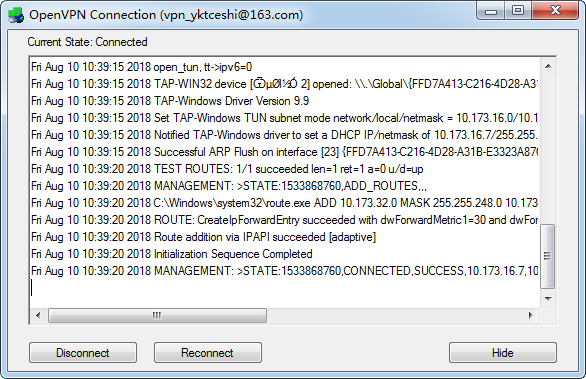
5、运行hadoop fs -mkdir 命令，创建专属文件夹

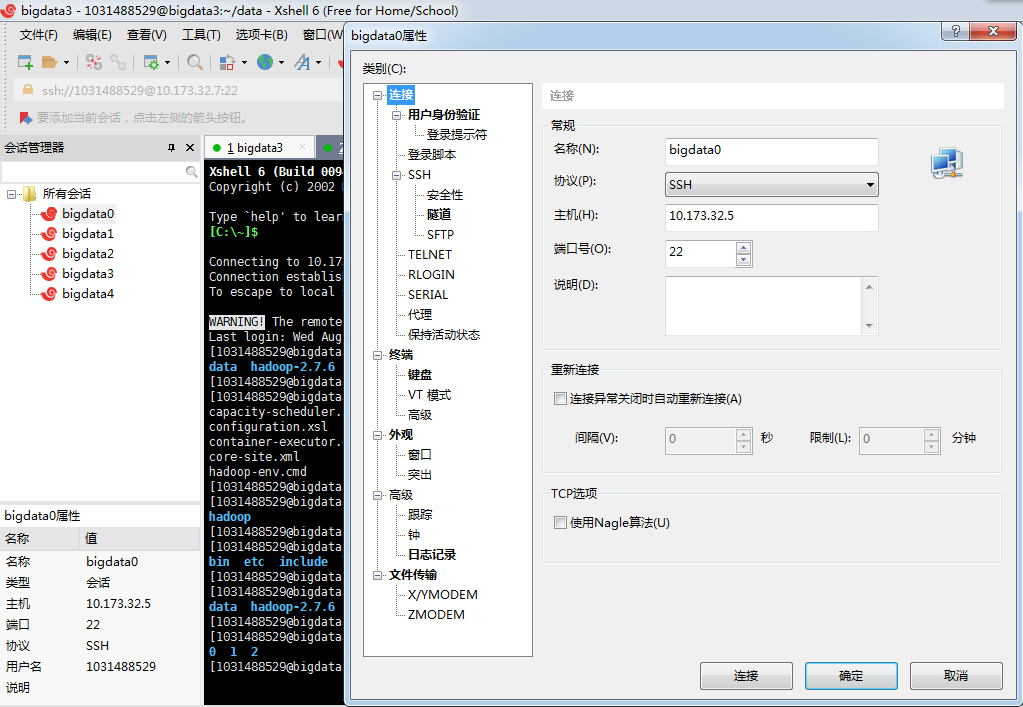
6、获取教学机tmp目录下的日志文件

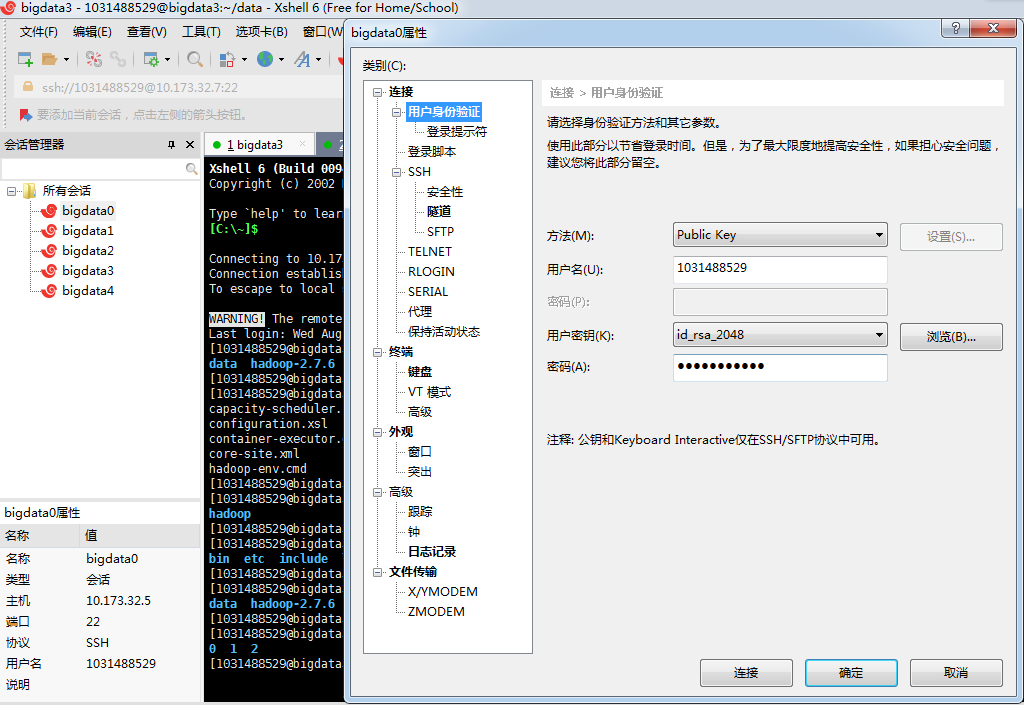
7、本地上传日志文件到HDFS的专属文件夹中

操作步骤

一、Xshell链接集群截图：



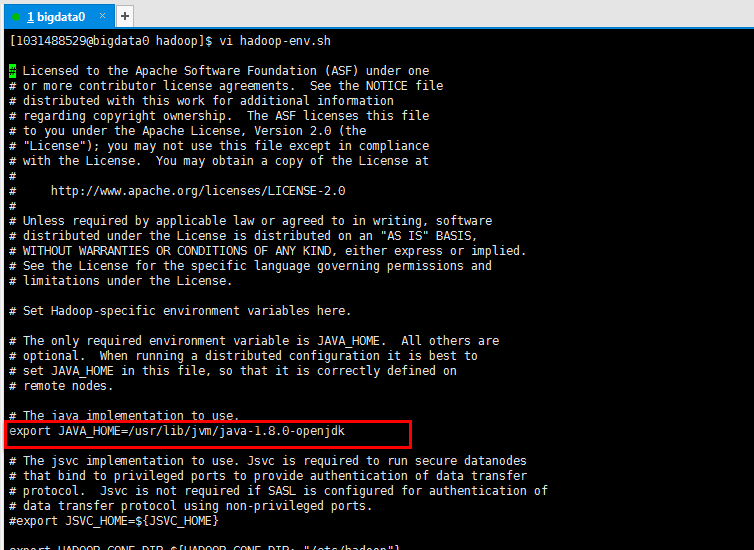


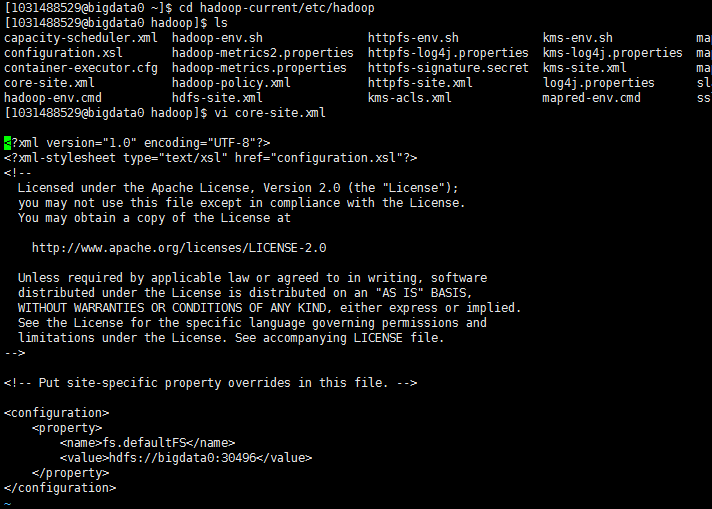


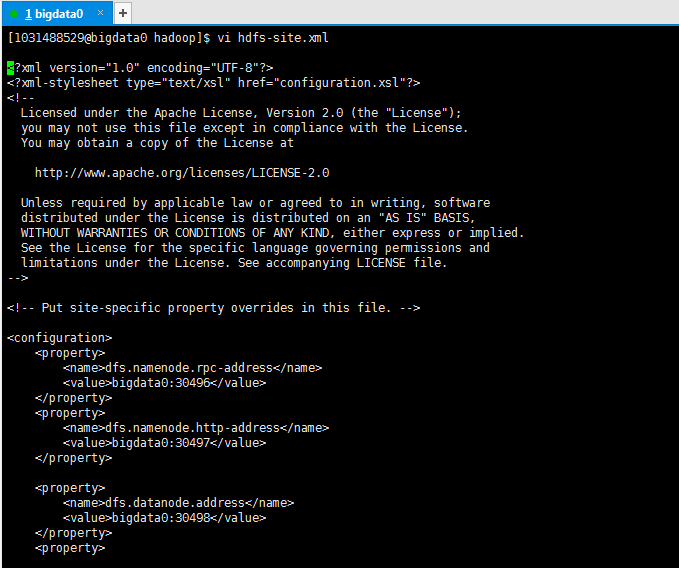
二、配置HDFS:

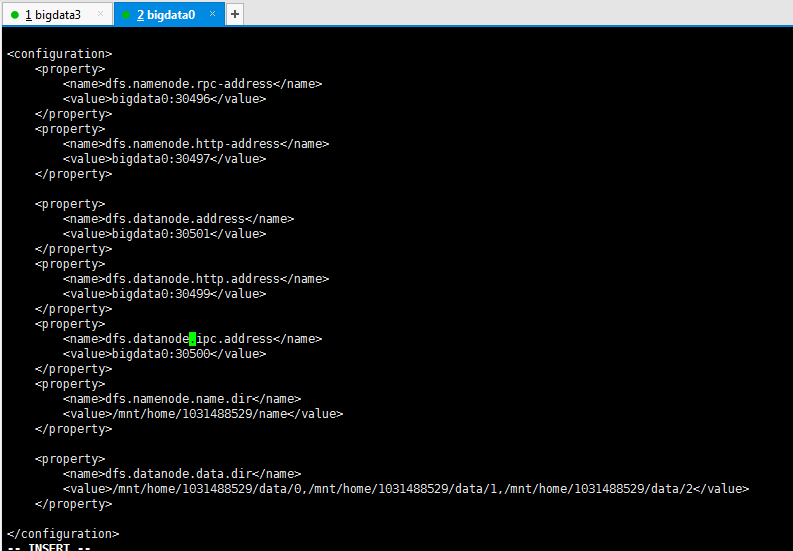
Bigdata0配置步骤：有些步骤当时没有截图，从笔记截图

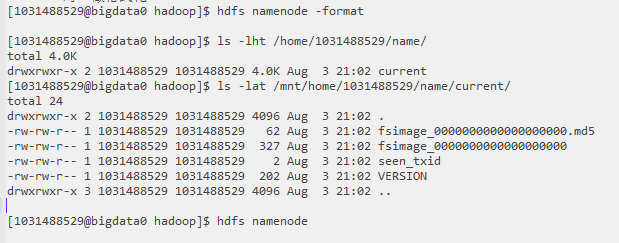


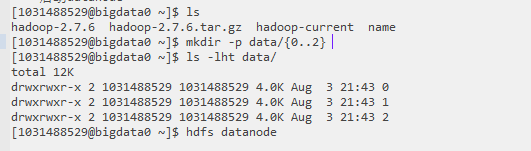




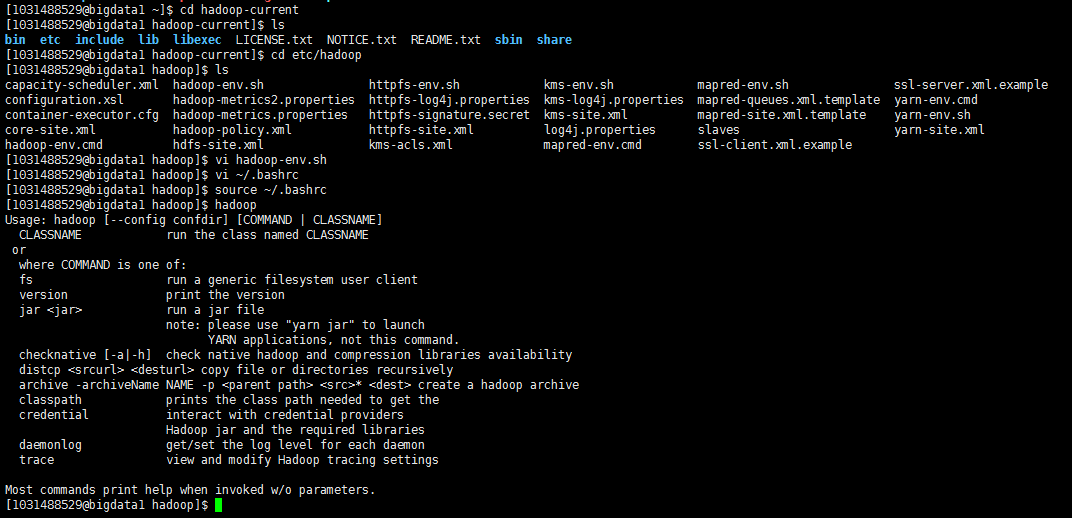


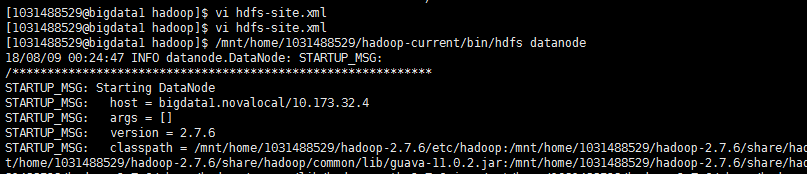


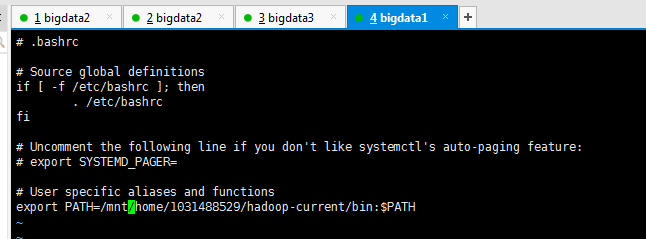


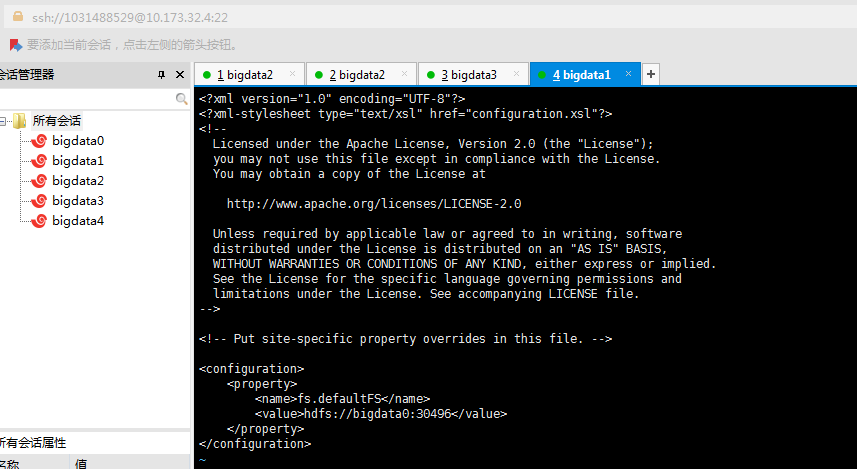


Bigdata1 配置步骤: copy和解压hadoop安装包的步骤当时没截图，参见bigdata2

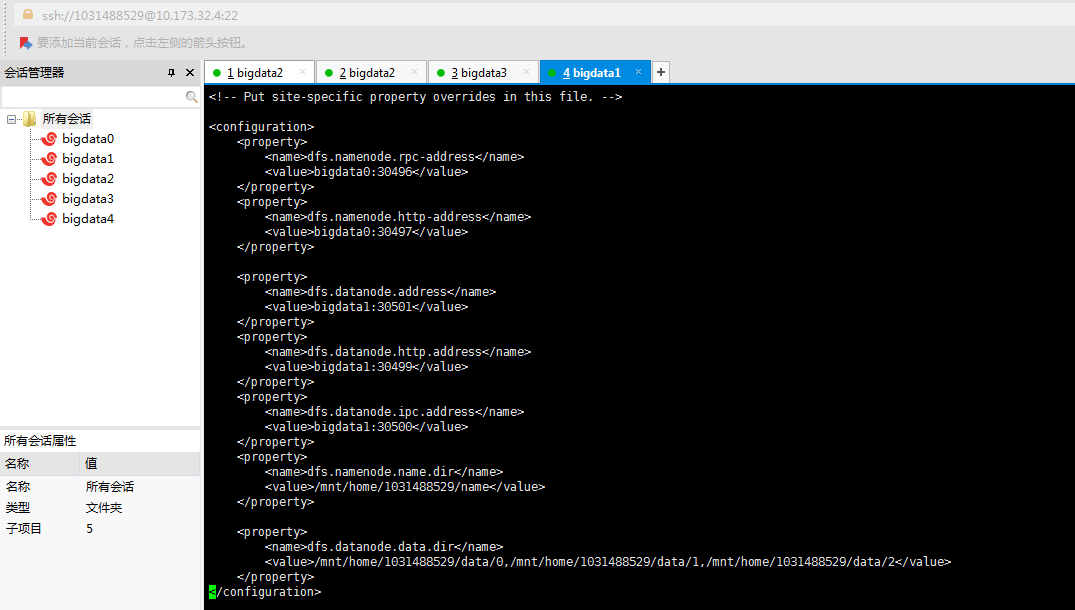




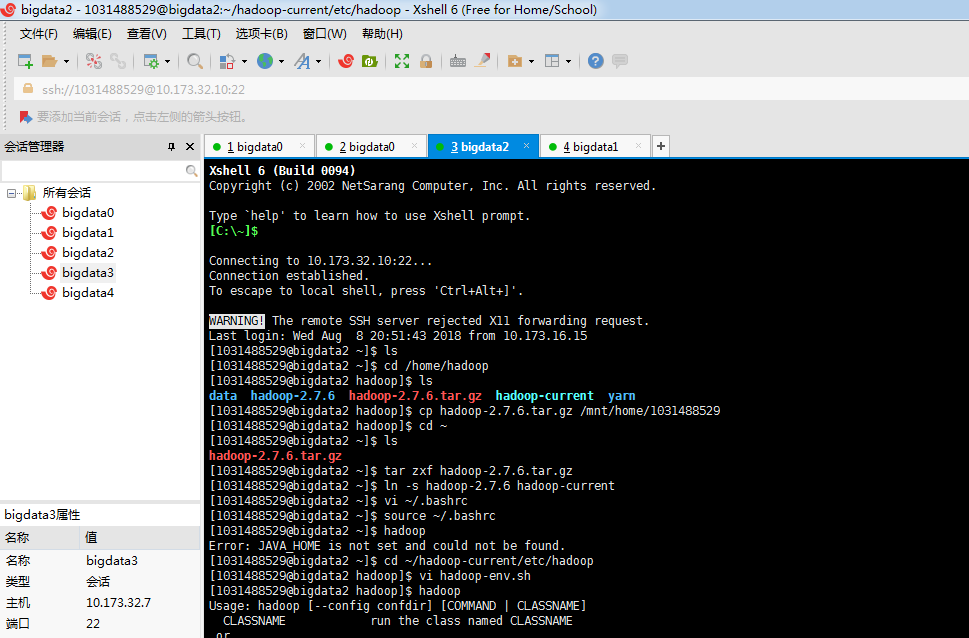


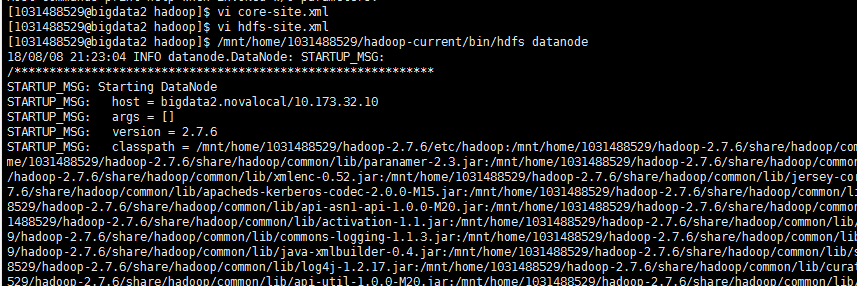


因为bigdata1:

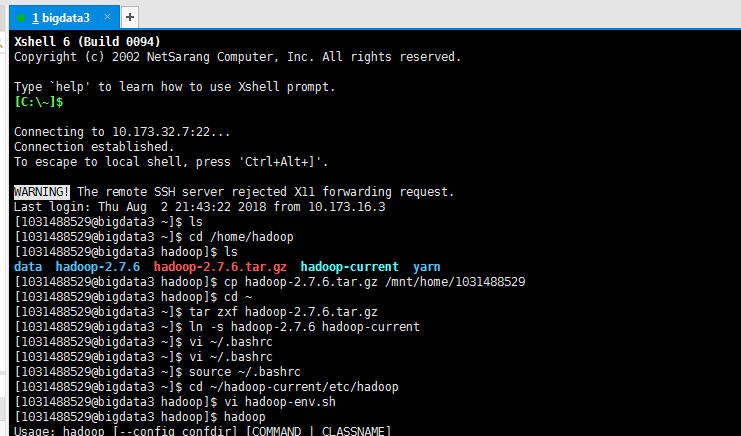


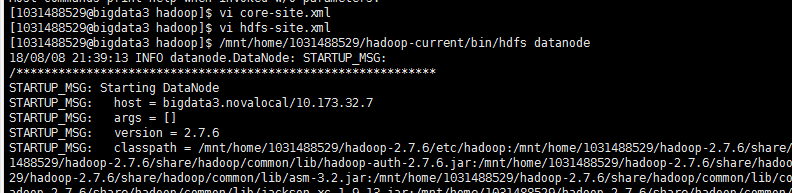
Bigdata2配置步骤：





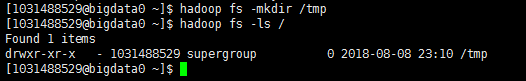
Bigdata3配置步骤：

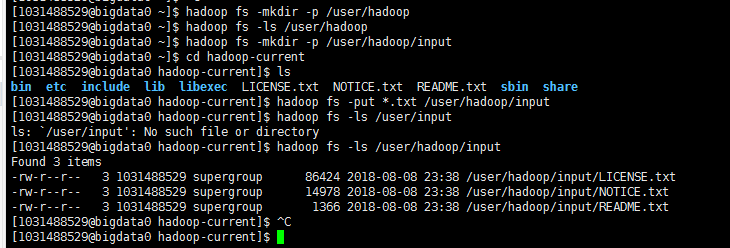


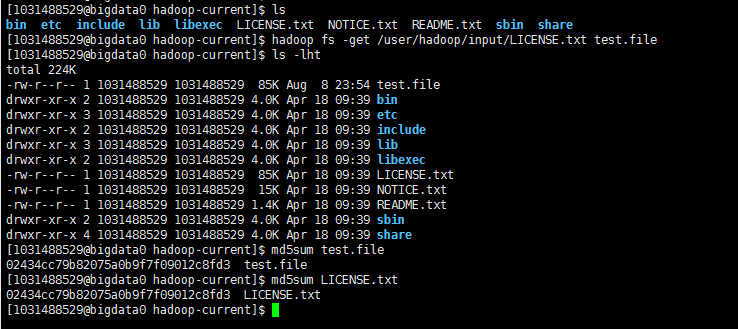


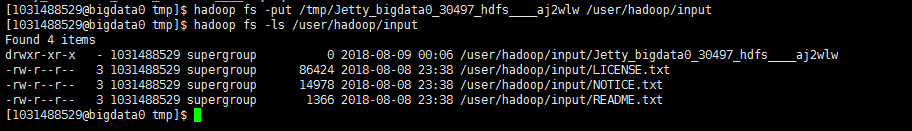
三．Hadoop创建专属文件夹，上传文件：



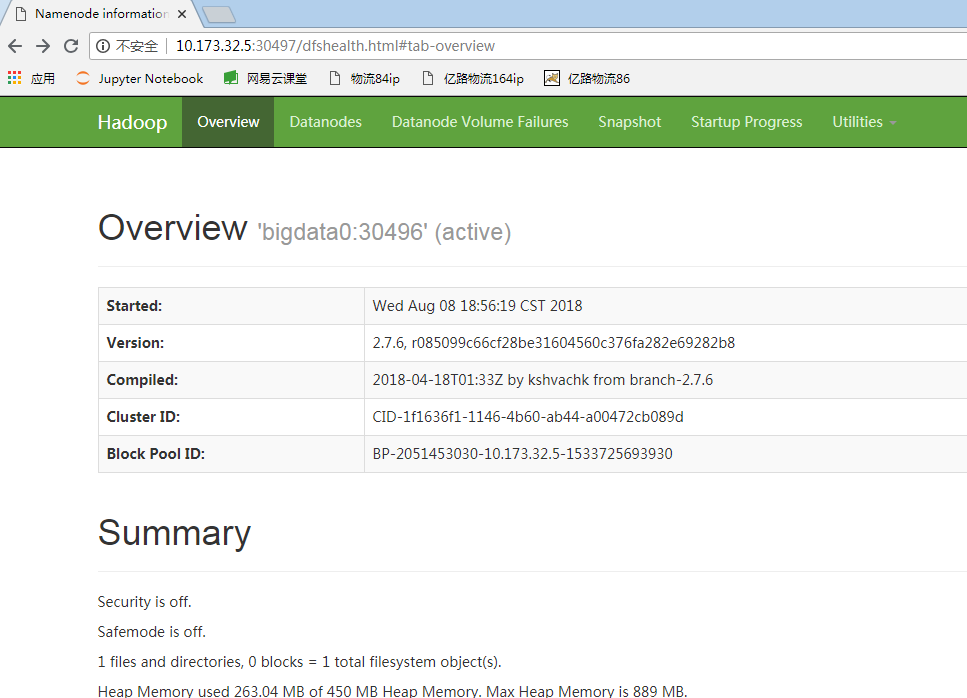


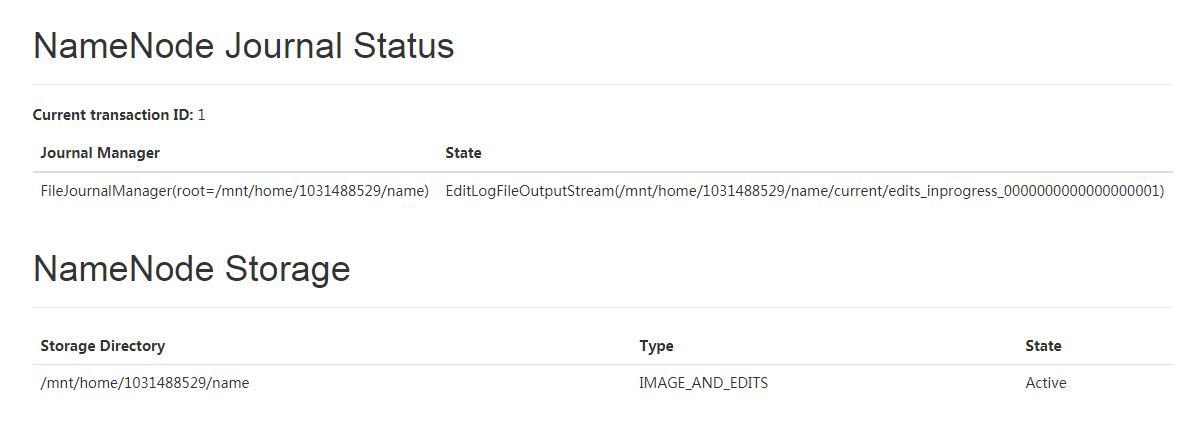


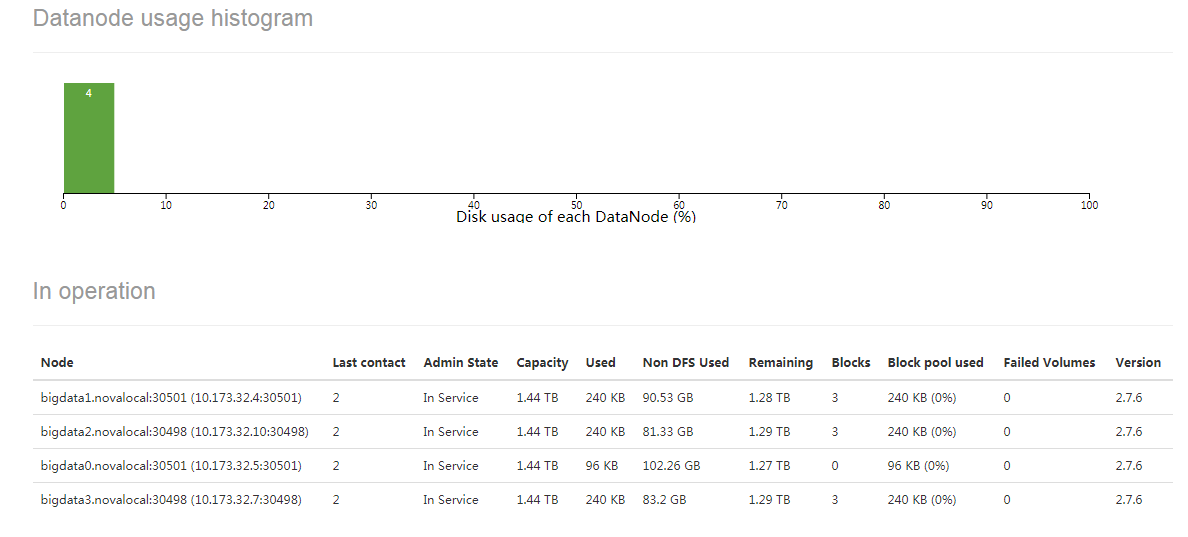




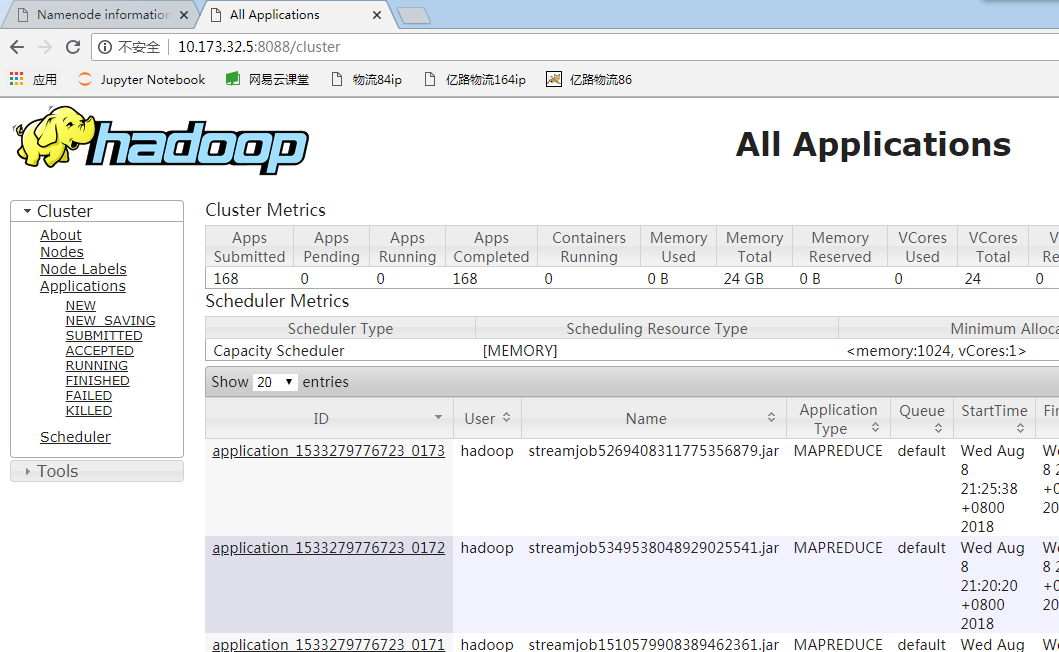
实验结果





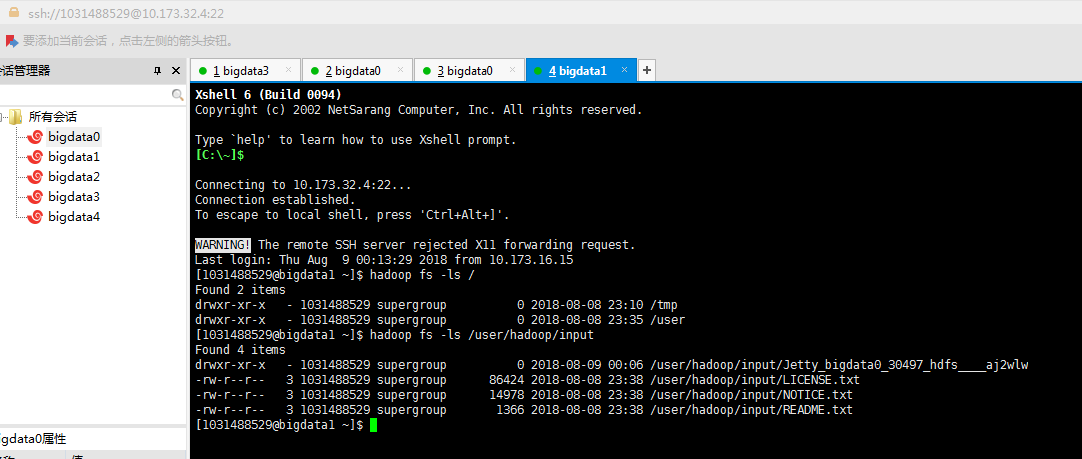


链接老师配好的yarn

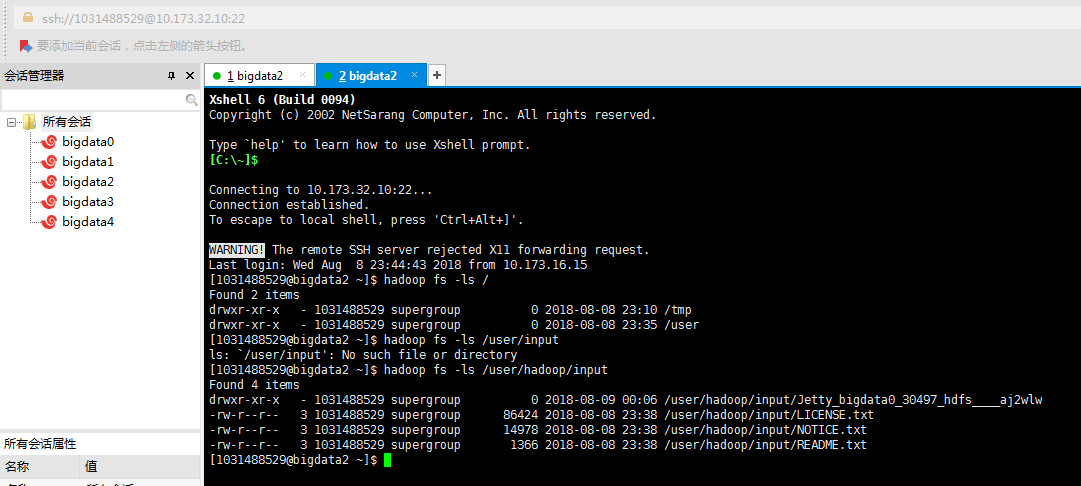


上传文件后各datanode结果：

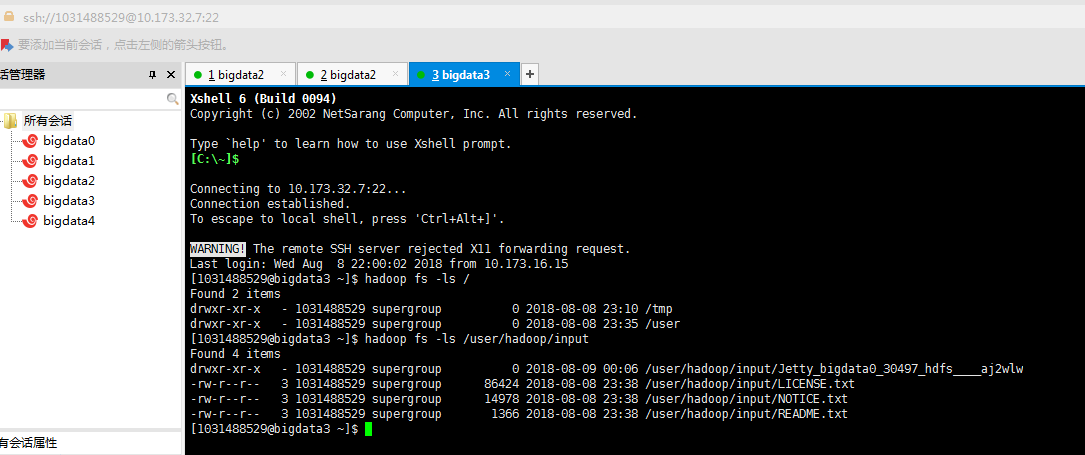
Bigdata1:



Bigdata2:



Bigdata3:



总结反思

1，自己创建的软连接、目录，可能有问题，可以尝试删掉重新创建；

2，开始没有理解namenode和datanode的结构，导致bigdata1的hdfs-site.xml 里面的namenode配错，从头把bigdata1重新配了一次，才理解；

3，启动datanode失败，开始时束手无策，后来学会看异常，针对异常去定位原因；

4，遇到问题要及时找老师和群里的同学请教；

5，一开始接触新的领域，要勤记笔记，把操作步骤保存起来，写下自己的理解，并作为后续操作的参考；

6，对hdfs的理解：

基本的hdfs架构包括一个namenode和几个datanode，比如bigdata0~3这4个节点上的datanode，都指向bigdata0上的namenode；

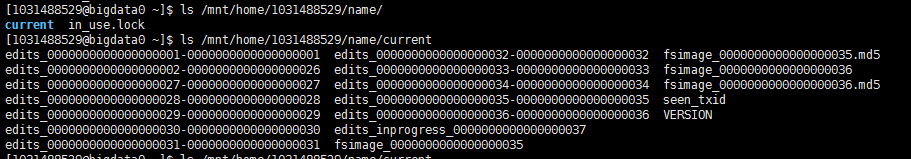
每个节点除了基本的环境配置外，**关键是配置core-site.xml和hdfs-site.xml这两个文件**，比如core-site.xml中在fs.defaultFS配置namenode地址；涉及namenode的配置全部指向bigdata0同一个端口；

每个物理数据节点，都使用自己的地址，不同的节点上可以使用相同的端口，互不干扰；

dfs.datanode.data.dir属性中配置的三个路径，结果是文件更新时，hdfs会自动把文件生成3个备份，放到这三个路径下；

配置完namenode需要格式化再启动，datanode配置完直接启动；可以保持窗口让namenode/datanode运行，也可以在后台用daemon进程永久运行，直至手动关闭；

7，通过实例了解namenode结构：启动namenode才会有in\_use.lock



8，通过实例了解datanode目录结构：启动namenode才会有in\_use.lock

