分布式系统实验报告

**实验一**

**Hadoop安装与配置**

姓名： 樊启元

学号： 7203610712

院系： 未来技术学院

**2023年 5月 3日**

**1.实验内容**

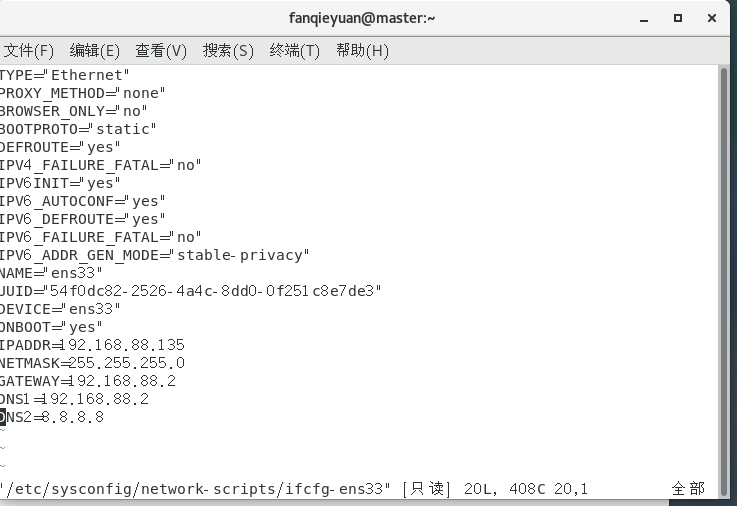
在自己计算机上搭建伪分布式Hadoop集群环境，环境搭建完毕后通过命令及Web访问方式查看集群的启动情况。

**实验具体步骤**（截图及说明）

1. 设置Linux系统静态IP地址

输入命令，按如下配置

vi /etc/sysconfig/network-scripts/ifcfg-ens33



1. 关闭防火墙



1. 修改主机名

vi /etc/hostname



然后重启

1. 配置主机名与IP映射

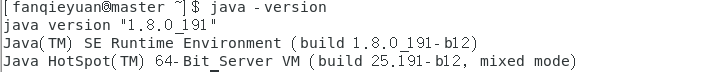
vi /etc/hosts



5.设置SSH免密登录



6.安装JDK



**Hadoop的安装与配置**

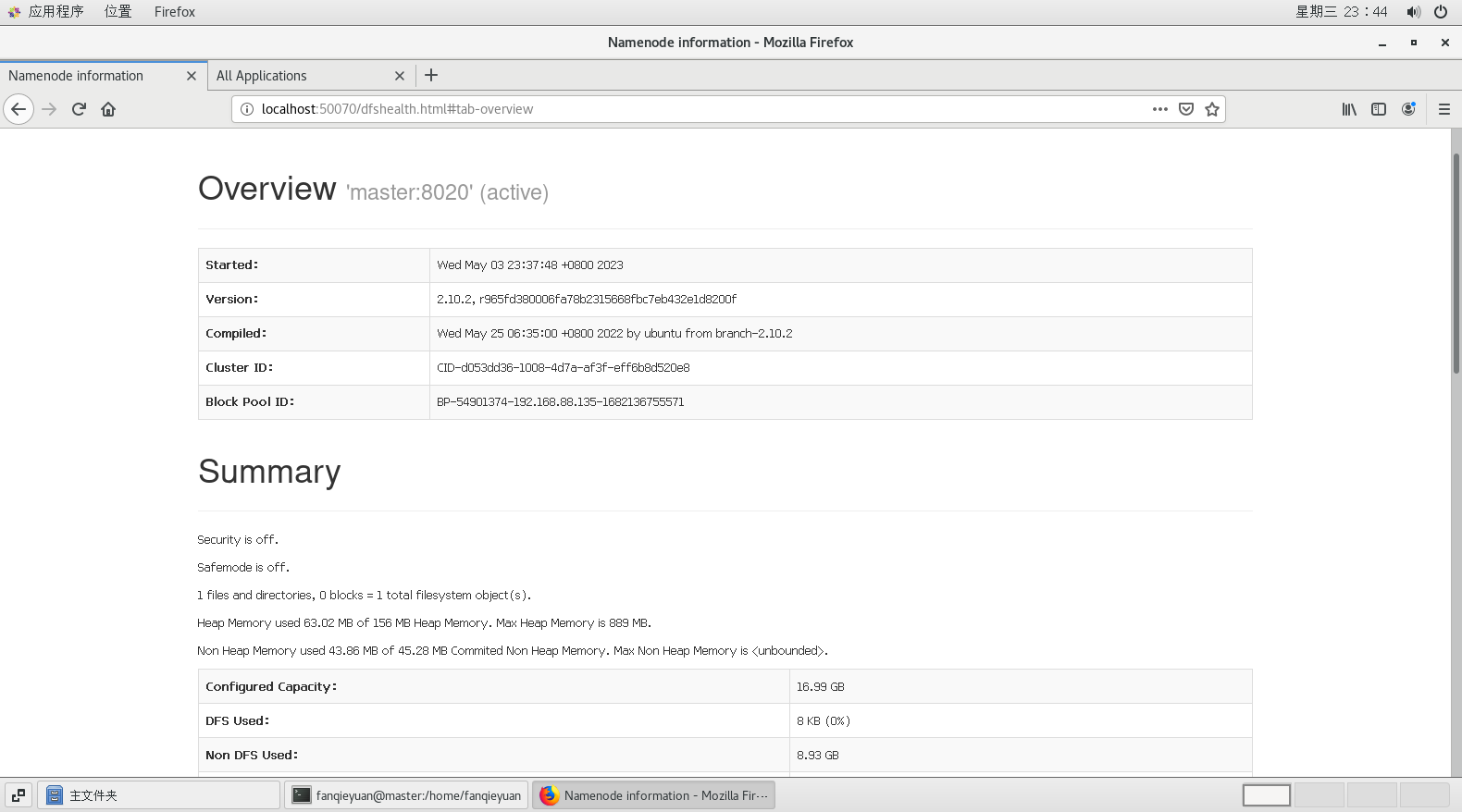
start-dfs.sh

start-yarn.sh

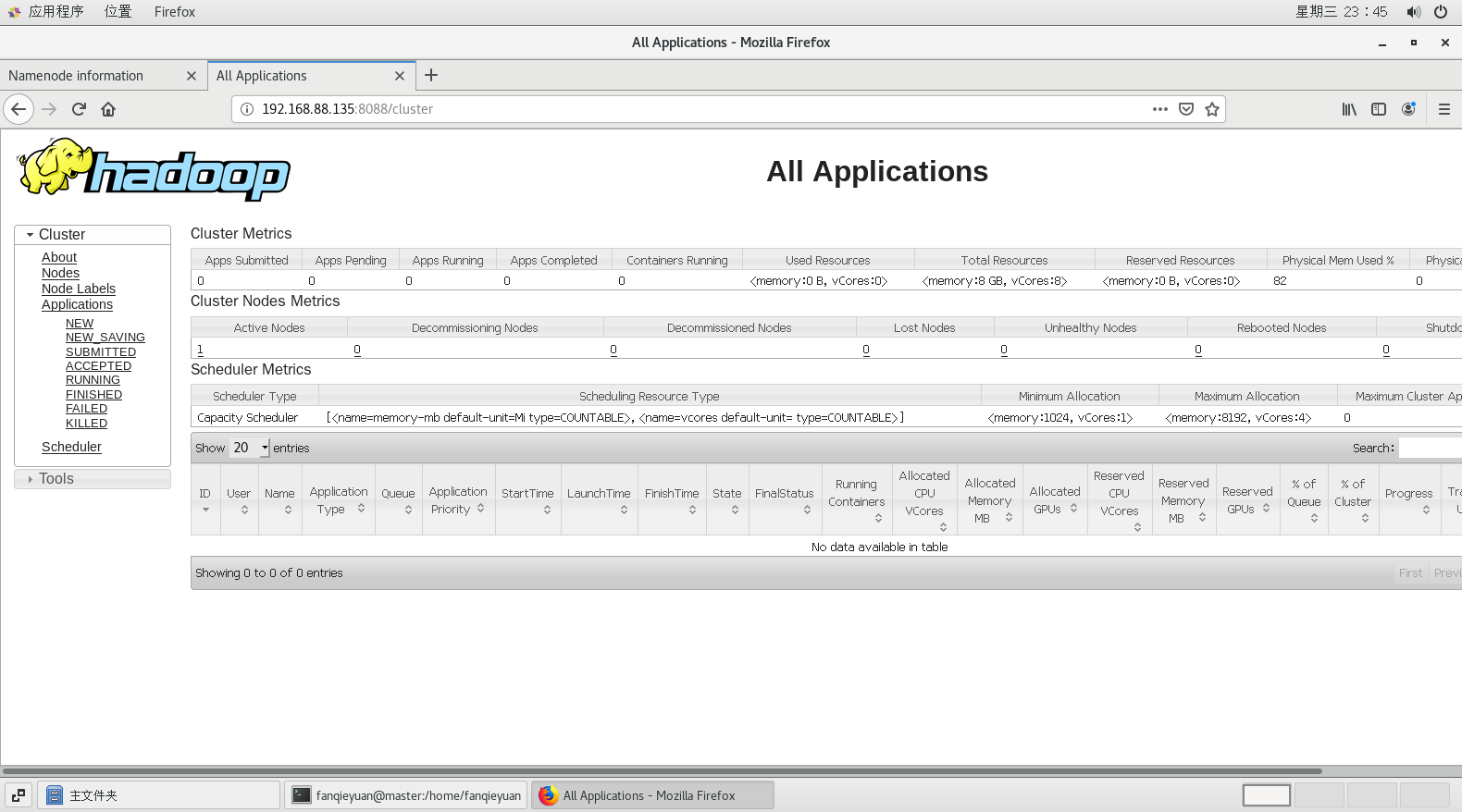
jps

查看网页：

http://localhost:50070/



<http://192.168.88.135:8088/>



**2.课后问题**

在伪分布式Hadoop集群环境中需要设置SSH免密登录，请简要说明免密登录设置的必要性和免密登录流程。

答：在伪分布式Hadoop集群环境中，SSH免密登录的设置是非常必要的。因为在Hadoop 集群中，每个节点都需要互相通信，并且需要不断地进行数据传输和处理。如果没有设置SSH免密登录，那么每次进行节点之间的通信或者数据传输时都需要输入密码，这样会增加很多不必要的操作和时间消耗。

SSH免密登录的设置流程如下：

1. 生成ssh公私钥对，使用ssh-keygen命令生成ssh公私钥对，江公钥复制到其他节点。
2. 在目标节点上添加用户及其公钥，可以使用ssh-copy-id命令将主节点上的公钥添加到其他节点的authorized\_keys文件中。
3. 测试SSH连接，确保从主节点可以通过ssh连接到其他节点，而无需输入密码。

通过以上步骤设置SSH免密登录后，就可以在Hadoop集群中进行节点之间的通信和数据传输，而无需输入密码，提高了工作效率。

**3. 个人建议及问题**

1. 配置过程中，静态ip配置出问题了。

将 ip 地址设置为与主机同一网段即可。

1. 中间配置core-site.xml 有中文符号，导致格式化报错。

将中文符号修改为英文即可。

1. 第一次配置完，网页中现实的可用资源全是0

检查发现是DataNode配置出了问题：



### 原因是多次格式化，导致clusterID不匹配引起的问题。

解决方法：

方案一: 保留现有数据

用NameNode节点的~/dfs/name/current/VERSION 中的namenode的clusterID替换所有datanode节点机器中~/dfs/data/current/VERSION中的clusterID。

重启启动hadoop:start-all.sh

这种方式不影响现有的数据，避免了重新的格式化。

方案二: 重新格式化

执行./stop-all.sh关闭集群

删除存放hdfs数据块的文件夹(hadoop/tmp/)，然后重建该文件夹

删除hadoop下的日志文件logs

执行hadoop namenode -format格式化hadoop

重启hadoop集群

原文链接：https://blog.csdn.net/m0\_67402564/article/details/126565378