ssh免密码登录

简单操作，也就是将本机的公钥拷贝到想要免登陆的机器上。

ssh-copy-id -i worker1

ssh-copy-id -i worker2

参考：

<https://blog.csdn.net/fanrenxiang/article/details/69212647>

hadoop配置文件

1. hadoop\_env.sh

配置java 环境变量 export JAVA\_HOME=/opt/java

2. core-site.xml

3. hdfs-site.xml

启动hdfs

先要格式化文件系统

./bin/hdfs namenode –format

然后启动hdfs

./sbin/start-dfs.sh

设置Yarn服务

设置Yarn服务、启动相关组件、运行和操作Yarn集群。

1. 启动ResourceManager守护进程和NodeManager守护进程

$sbin/start-yarn.sh

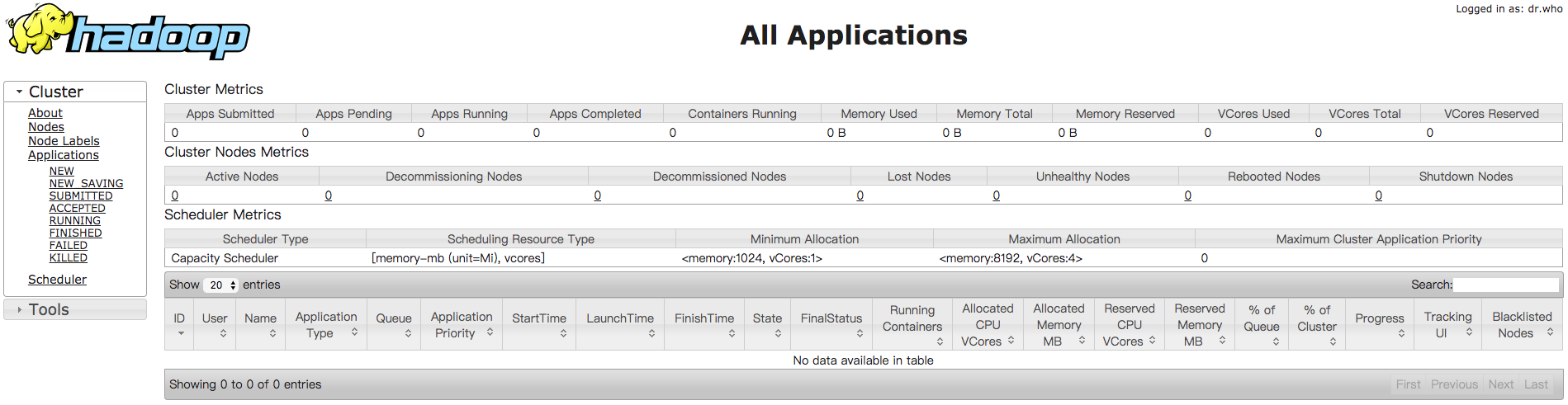
2、浏览ResourceManager的web界面，默认状态下地址为<http://localhost:8088>

3、运行MapReduce任务

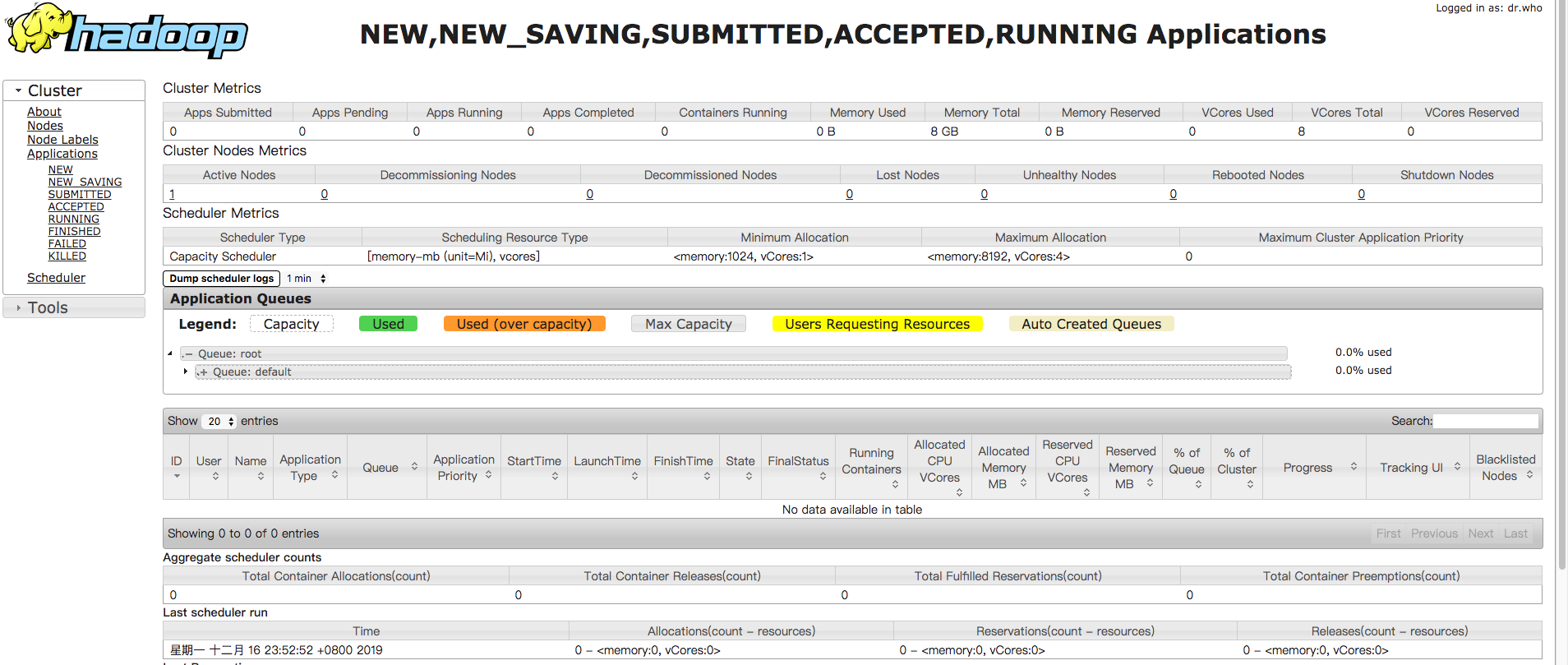
4、结束后，终止守护进程，运行：

$sbin/stop-yarn.sh

下图显示Yarn ResourceManager。



下图显示了集群中的资源队列，以及处于运行状态下的应用程序。除此之外，还可以在此处查看和监视运行中的各项任务。



hive

下载安装hive-3.1.2

<https://hive.apache.org/downloads.html>

hive-3.1.2版本与Hadoop3适配。

cd /usr/local/Cellar/hive/3.1.1/libexec/conf

cp hive-default.xml.template hive-site.xml #原始配置文件备份

Pip

下载安装Pig-0.17.0

<https://pig.apache.org/releases.html#Download>