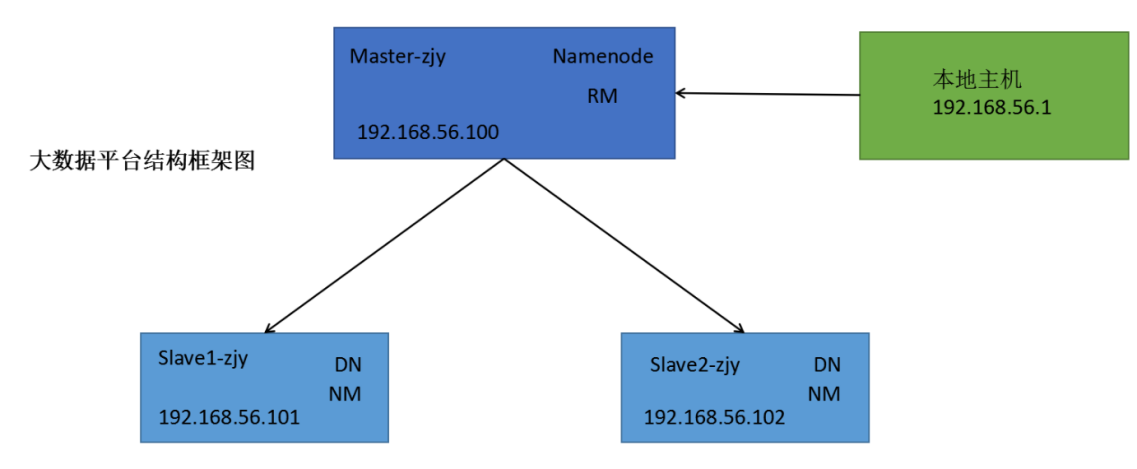
**Hadoop 2.x（Hadoop Yarn）安装部署实验**

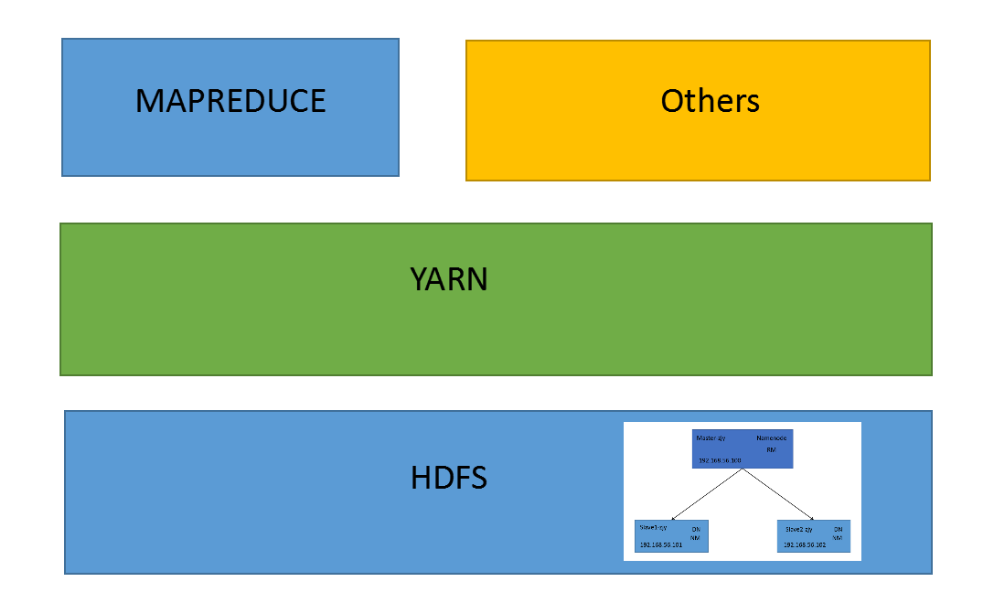
# 大数据HDFS平台的搭建

搭建大数据平台，我在虚拟机上设置三个节点，其中一个主节点Master-zjy，和两个分节点Slave-zjy和Slave2-zjy，其ip地址分别设置为192.168.56.100、192.168.56.101、192.168.56.102，配置完hadoop可以看到Master-zjy上的Namenode节点和Slave-zjy和Slave2-zjy上的Datanode，配置Yarn结束后可以看到Master-zjy上的ResourceManage节点和Slave-zjy和Slave2-zjy上的nodemanage。



YARN是一个资源管理、任务调度的框架，主要包含三大模块：ResourceManager（RM）、NodeManager（NM）、ApplicationMaster（AM）。其中，ResourceManager负责所有资源的监控、分配和管理；ApplicationMaster负责每一个具体应用程序的调度和协调；NodeManager负责每一个节点的维护。对于所有的applications，RM拥有绝对的控制权和对资源的分配权。而每个AM则会和RM协商资源，同时和NodeManager通信来执行和监控task。

RM相当于学院，NM相当于辅导员，AM相当于班长，学院负责所有资源的监控、分配和管理，辅导员负责节点的维护管理，班长负责班里应用程序的调度和协调。下图为Hadoop2.0的结构图。本次hadoop平台的搭建我就将按照这个结构图进行。



## 1.1、软件准备

Virtualbox、CentOS、Hadoop、JDK、Xshell、Xftp

## **1.2、安装虚拟机**

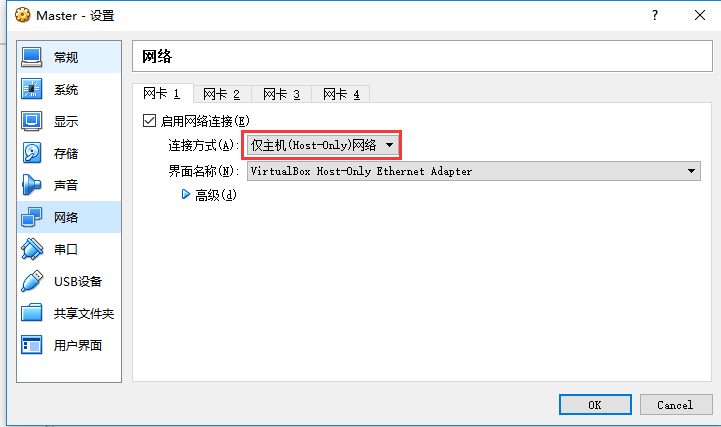
1. 、保留一个干净的虚拟机，随时可以提供复制。
2. 、在安装CentOs时，注意将其设置为基础服务器。

**1.3、查看主机的IP**



## **1.3、在虚拟机中设置网络**

（1）、在虚拟机中选用host-only模式。



（2）、设置虚拟机的网络

vi /etc/sysconfig/network

# Created by anaconda

NETWORKING=yes

GATEWAY=192.168.56.1



1. 、在虚拟机中配置主机ip

vi /etc/sysconfig/network-scripts/ifcfg-enp0s3

TYPE=Ethernet

IPADDR=192.168.56.100

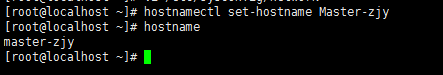
NETMASK=255.255.255.0



1. 、重启网络设置并将虚拟机主机名更改

service network restart

hostnamectl set-hostname master

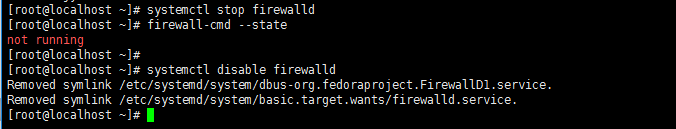


1. 、关闭windows和linux的防火墙。

systemctl stop firewalld

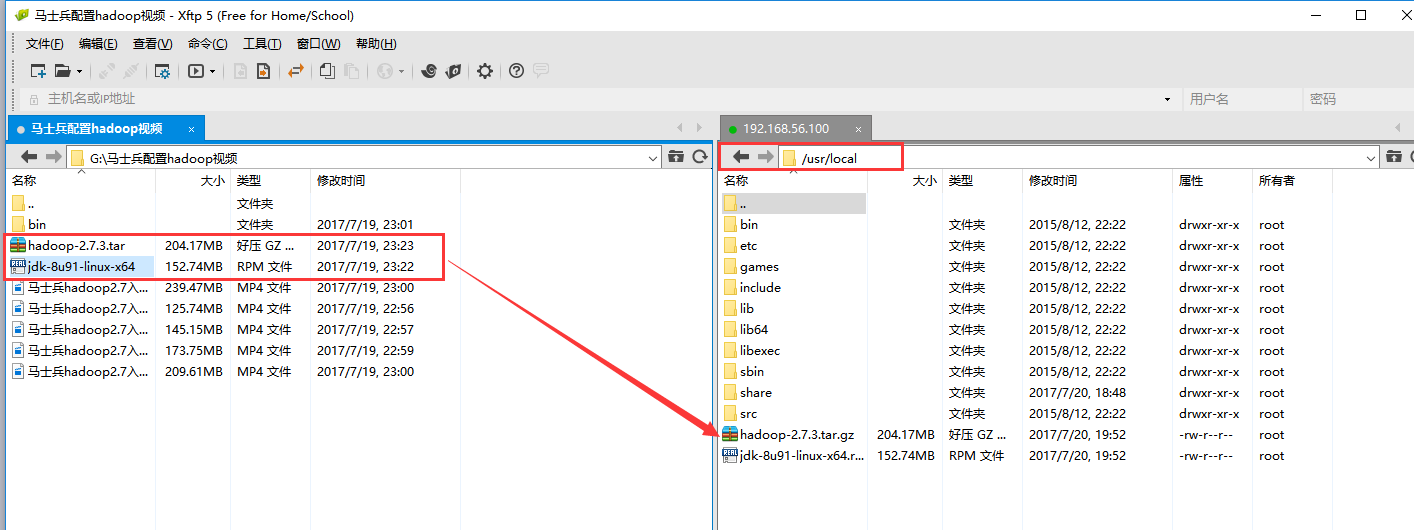
firewall-cmd --state

systemctl disable firewalld



## **1.4、Xshell和Xftp**

用Xshell和Xftp将下载好的jdk hadoop复制到master的usr/local/目录下。

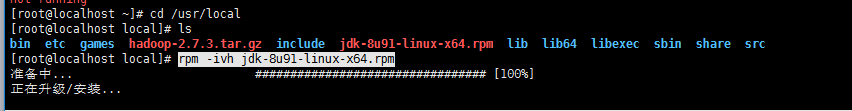


## **1.5、配置hadoop和java**

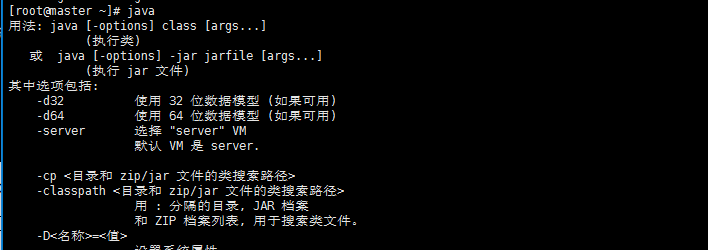
（1）、在/etc/local目录下解压jdk

cd /usr/local

rpm -ivh jdk-8u91-linux-x64.rpm



java当出现java的帮助文档时表示jdk安装完毕。



（2）、在/etc/local目录下解压hadoop

cd /usr/local

tar -xvf hadoop-2.7.3.tar.gz

更改mv hadoop-2.7.3 hadoop

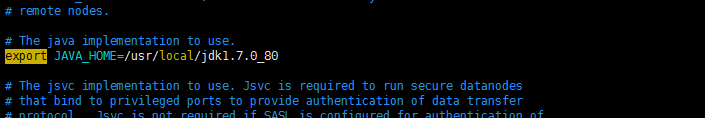




进入hadoop/etc配置hadoop-env.sh的java运行环境。

cd /usr/local/hadoop/etc/hadoop

vi hadoop-env.sh



（3）、将hadoop执行命令的环境加到path环境中vim /etc/profile

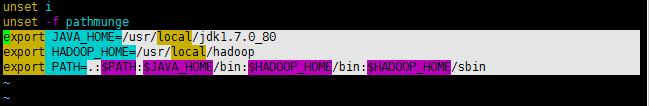
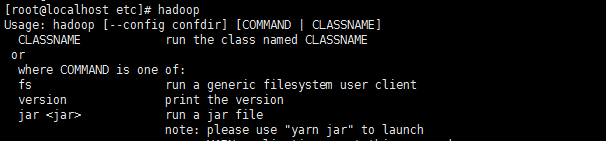
export JAVA\_HOME=/usr/local/jdk1.7.0\_80

export HADOOP\_HOME=/usr/local/hadoop

export PATH=.:$PATH:$JAVA\_HOME/bin:$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin

重新启动source /etc/profile

然后输入hadoop看到帮助文档表示hadoop配置成功。



（4）、配置hosts地址

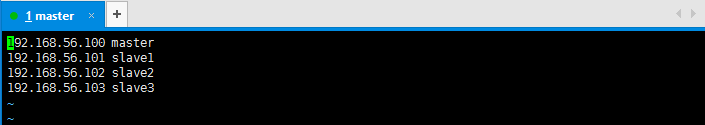
vim /etc/hosts

192.168.56.100 master

192.168.56.101 slave1

192.168.56.102 slave2

192.168.56.103 slave3



## **1.6、集群安装及配置文件修改**

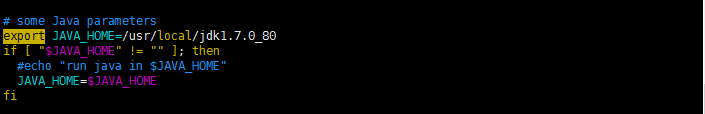
以下配置文件均在hadoop目录下进行的

cd /usr/local/hadoop/etc/hadoop

（1）、配置yarn-env.sh

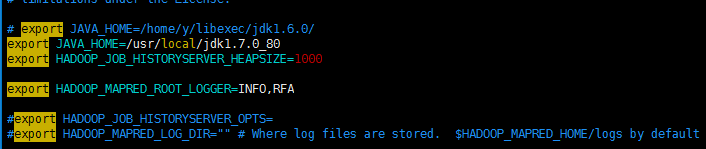
vim yarn-env.sh 指定JAVA\_HOME

export JAVA\_HOME=/usr/local/jdk1.7.0\_80



(2) 、vim指令打开mapred-env.sh指定JAVA\_HOME

export JAVA\_HOME=/usr/local/jdk1.7.0\_80



(3)、配置core-site

打开命令：vim core-site.xml

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://Master-zjy:9000</value>

<description>The name of the default file system</description>

</property>

<property>

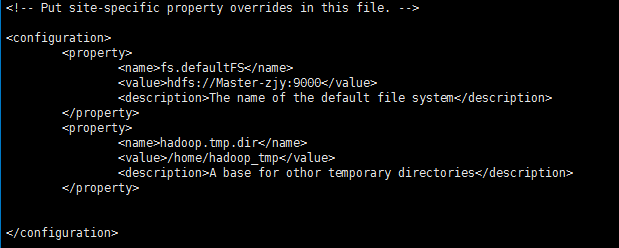
<name>hadoop.tmp.dir</name>

<value>/home/hadoop\_tmp</value>

<description>A base for othor temporary directories</description>

</property>

</configuration>



1. 、配置hdfs地址和端口

【dfs.namenode.secondary.http-address指定secondarynamenode在slave1】

vim 打开hdfs-site.xml vim hdfs-site.xml

<configuration>

<property>

<name>dfs.replication</name>

<value>3</value>

</property>

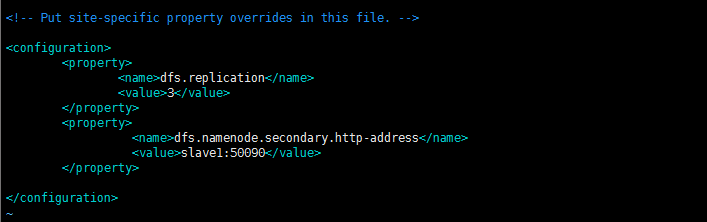
<property>

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

<value>slave1:50090</value>

</property>

</configuration>



(5)、配置vim mapred-site.xml

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



Vim打开mapred-site.xml： vim mapred-site.xml

<configuration>

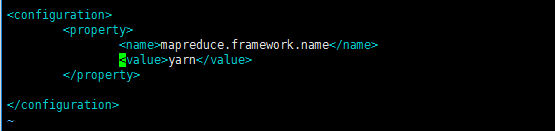
<property>

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

<value>yarn</value>

</property>

</configuration>



(6)、配置yarn-site.xml

Vim打开yarn-site.xml vim yarn-site.xml

<configuration>

<!-- Site specific YARN configuration properties -->

<property>

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

<value>Master</value>

</property>

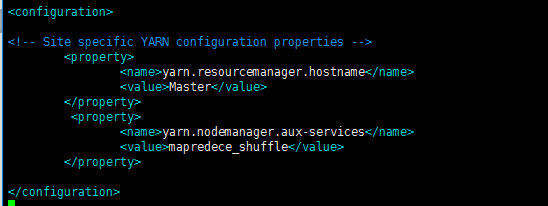
<property>

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

<value>mapreduce\_shuffle</value>

</property>

</configuration>



## **1.7、复制节点**



(1)、分别修改slave1网络地址为192.168.56.101

分别修改slave2网络地址为192.168.56.102

分别修改slave3网络地址为192.168.56.103

分别修改主机名：

hostnamectl set-hostname master

hostnamectl set-hostname slave1

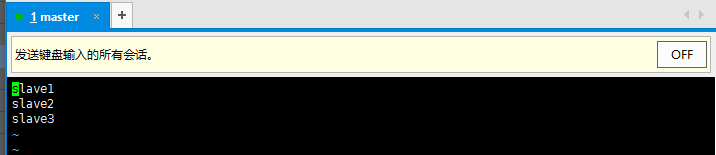
hostnamectl set-hostname slave2

hostnamectl set-hostname slave3

（2）、告诉Master-zjy下有3个节点

cd /usr/local/hadoop/etc/hadoop

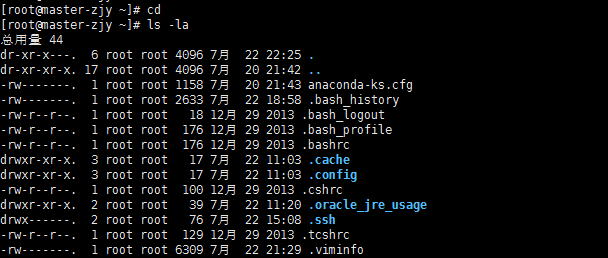
vim slaves



## **1.8、配置ssh免密码登录**

（1）、执行命令cd退回根目录：cd

（2）、运行ls -la (隐藏文件夹）

（3）、在根目录下运行ssh-keygen -t rsa连续四下回车生成密钥对私钥id\_rus和公钥id\_rsa.pub,默认存储在/root/.ssh中

（4）、将公钥分别发给slava1和slave2、slave3和自己master

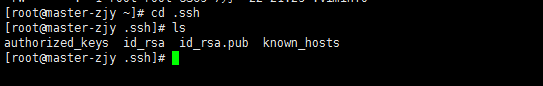
ssh-copy-id slave1

ssh-copy-id slave2

ssh-copy-id slave3

ssh-copy-id master

（5）、查看私钥和公钥



（6）、3台虚拟机分别执行命令：ssh 主机名或ip，如果都能登录则配置成功。

ssh master

ssh slave1

ssh slave2

**1.9、启动hadoop集群**

（1）、启动hadoop集群

格式化文件系统，在主节点上执行命令：hadoop namenode –format

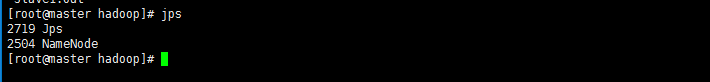
启动集群：start-dfs.sh

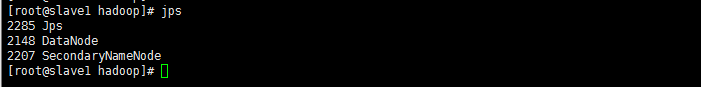
每个节点执行：jps

在主节点启动了NameNode，

slave1节点启动SecondaryNameNode和DataNode

其他分节点启动DataNode



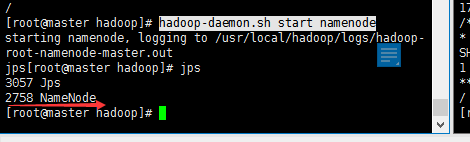


关闭集群：stop-dfs.sh

（2）、单独启动某个节点

hadoop-daemon.sh start namenode

hadoop-daemon.sh start datanode



# 二、Hadoop 2.x集群安装和启动配置

1. 、配置vim mapred-site.xml【步骤1只在master上配置】

cd /usr/local/hadoop/etc/hadoop

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



Vim打开mapred-site.xml：

vim mapred-site.xml

<configuration>

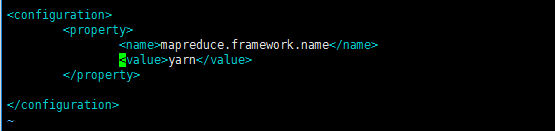
<property>

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

<value>yarn</value>

</property>

</configuration>



(2)、配置yarn-site.xml【步骤2在master和slave都要配置】

Vim打开yarn-site.xml

vim yarn-site.xml

<configuration>

<!-- Site specific YARN configuration properties -->

<property>

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

<value>master</value>

</property>

<property>

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

<value>mapreduce\_shuffle</value>

</property>

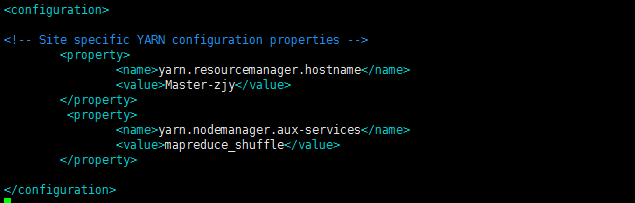
<property>

<name>yarn.nodemanager.aux-services.mapreduce\_shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

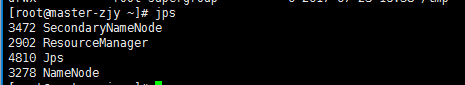
</configuration>



（3）、启动yarn集群

start-yarn.sh

主节点启动了ResourceManager，分节点上启动了NodeManager则启动成功

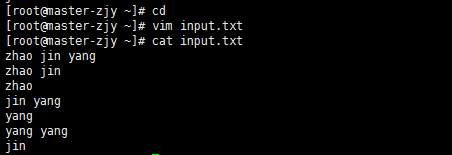


1. 、关闭yarn集群

stop-yarn.sh

# **三、MapReduce样例代码测试系统**

(1)、在根目录下生成input.txt文件，并编辑。

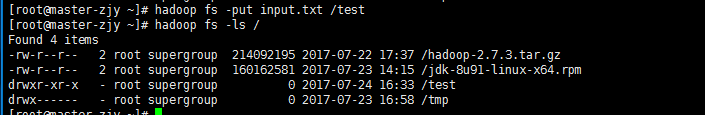


(2)、将其上传到hdfs下test目录下

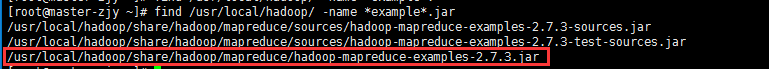
hadoop fs -mkdir /test //hdfs下创建test

hadoop fs -ls / //查看hdfs下根目录

fs -put input.txt /test //上传hdfs文件

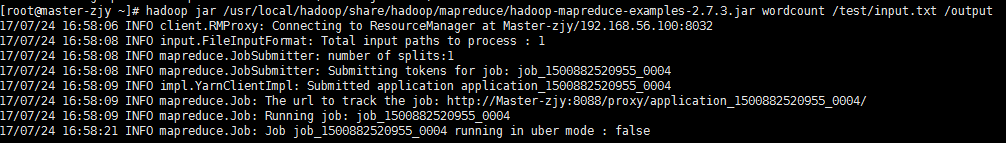


（3）、find /usr/local/hadoop -name \*example\*.jar 查找示例文件



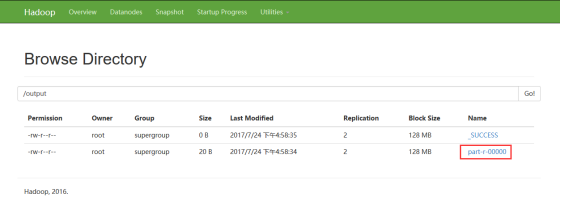
运行wordcount程序

hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.3.jar wordcount /test/input.txt /output

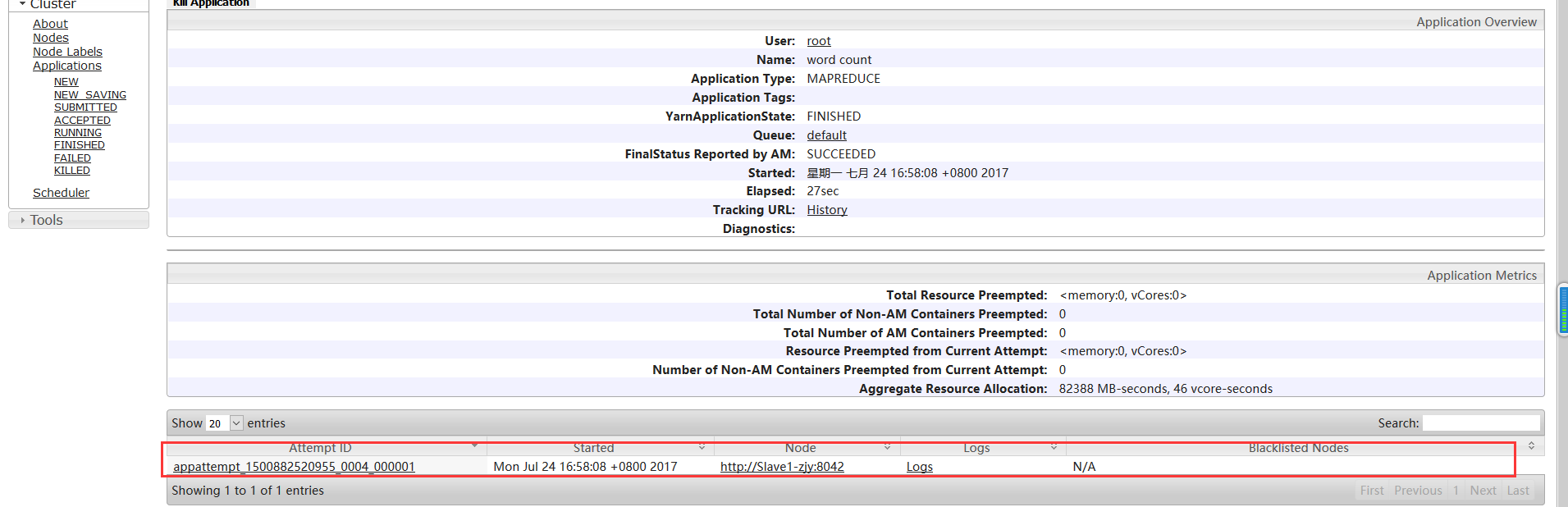


1. 、观察运行结果

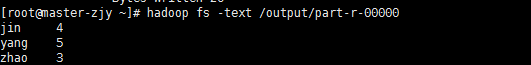
Hadoop hdfs端口：192.168.56.100:50070



Yarn 端口：192.168.56.100:8088



hadoop fs -text /output/part-r-00000



# 四、心得体会

本次作业的主要任务就是完成Hadoop 2.x（Hadoop Yarn）安装部署实验；Hadoop 2.x（Hadoop Yarn）集群系统版本安装和启动配置；使用Hadoop MapReduce V2样例代码快速测试系统。本次作业在进行hadoop平台搭建是遇到问题最多的一次，在搭建hadoop1.0时候，namenode和datanode节点总是不能同时启动，后来发现原因是配置文件总是自己写错，后来修改后就好了。

# 五、问题总结

namenode和datanode节点总是不能同时启动。

解决办法：

1. 、检查配置文档是否出错。
2. 、删除Slave1和slave2节点上目录/usr/local/hadoop/dfs/中的data文件。如果删除了Master上的name文件，需要对Master节点进行格式化再重新启动。

如果直接格式化还是不行，则需要先删除namenode历史的文件，再格式化重启。

**直接删除namenode的配置文件目录,在重新format**

**1,rm -rf ../hdfs/na\***

**2,./[hadoop](http://lib.csdn.net/base/hadoop" \o "Hadoop知识库" \t "http://blog.csdn.net/zzq900503/article/details/_blank) namenode -format**