# Spark集群搭建手册

## 操作系统设置

### 关闭防火墙

Chkconfig iptables off

### 设置selinux

修改/etc/selinux/config文件中的SELINUX=disabled

## **配置集群hosts**

在/etc/hosts文件中添加以下内容（IP地址和主机名以实际为准）

192.168.85.130 hadoop1

192.168.85.131 hadoop2

192.168.85.132 hadoop3

## **配置服务器间免密码登录**

1、分别在三台主机上执行ssh-keygen -t rsa

2、在hadoop1上先进入/root/.ssh ,然后执行cat id\_rsa.pub >> authorized\_keys

3、把其他主机上id\_rsa.pub文件内容追加到hadoop1主机上authorized\_keys里。

4、拷贝hadoop1主机上authorized\_keys文件到其他主机对应的目录里。

## **安装JDK**

rpm -ivh jdk-7u67-linux-x64.rpm

## **安装Scala**

解压Scala到相应的目录里，本次解压到/root目录里 tar -xvf scala-2.10.5.tgz。

## **安装Spark**

解压Spark到相应的目录里，本次解压到/root目录里，tar -xvf spark-1.6.0-bin-hadoop2.6.tgz。

## **配置环境变量**

1、在/etc/profile文件里添加以下内容：

export SPARK\_HOME=/root/spark-1.6.0-bin-hadoop2.6

export SCALA\_HOME=/root/scala-2.10.5

export PATH=$PATH:$SPARK\_HOME/bin:$SCALA\_HOME/bin:$MAVEN\_HOME/bin

1. 使/etc/profile文件生效

执行 source /etc/profile

## **配置Spark**

1. 配置spark-env.sh文件（文件在spark安装路径下的conf目录里），内容如下：

|  |
| --- |
| export SPARK\_HOME=/root/spark-1.6.0-bin-hadoop2.6  export HADOOP\_CONF\_DIR=/root/hadoop-2.6.0-cdh5.10.1/etc/hadoop  export JAVA\_HOME=/usr/java/default  export SCALA\_HOME=/root/scala-2.10.5  export SPARK\_MASTER\_IP=hadoop1  export SPARK\_WORKER\_MEMORY=1g  export SPARK\_WORKER\_CORES=1  export SPARK\_WORKER\_INSTANCES=1 |

1. 配置slaves文件如下(配置worker节点)：

hadoop1

hadoop2

hadoop3

1. 在三个节点上做同样的配置（也可以在hadoop1上配置完后，拷贝到其他节点上）。

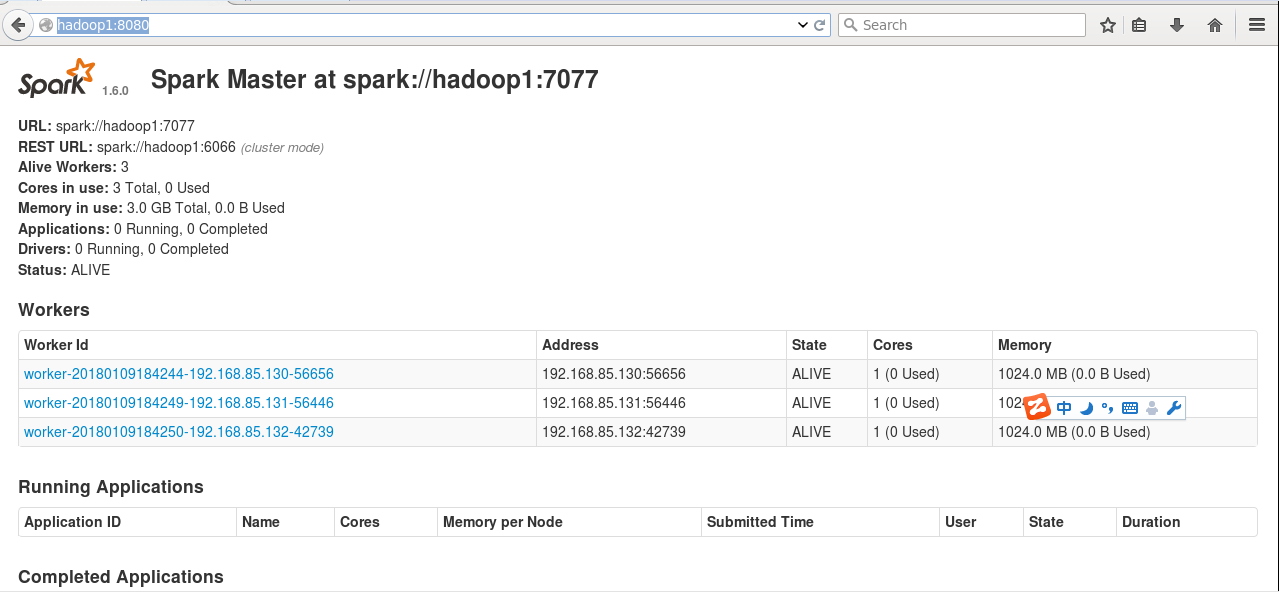
**九、**启动测试

1. 启动集群

在hadoop1上执行

/root/spark-1.6.0-bin-hadoop2.6/sbin/start-all.sh

1. 分别在三个节点上执行JPS命令，查看进程是否拉起
2. 在浏览器里输入http://hadoop1:8080/，如果能够正确打开，并能看到我们配置的三个worker，说明集群配置成功了。如下图所示：



1. 运行Spark自带示例进行测试

在hadoop1上执行run-example SparkPi。结果会显示很多信息，如果没有错误日子，说明成功。