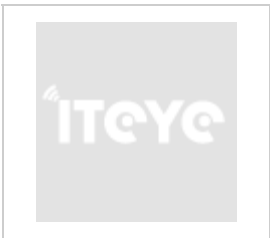


samwalt

永久域名 <http://samwalt.iteye.com>



samwalt

浏览: 7474 次

性别:

来自: 上海

我现在离线

[详细资料](#) [留言簿](#)

搜索本博客

最近访客 [>>更多访客](#)



[vlinux](#)



[birdflyzhi](#)



[level100](#)



[a123159521](#)

博客分类

- 全部博客 (30)
- 笔试试试 (1)
- ruby (13)
- 算法 (2)
- java (3)
- hadoop (5)
- 数据库 (2)

我的留言簿 [>>更多留言](#)

- 忙，有段时间没有上来了。说一下你的问题吧
-- by [samwalt](#)
- 您好，我遇见了一个跟你在<http://eclipsecpforum.group>...

2011-02-28

◀ [VIM的几个常用插件](#) | [安装rails: no such file to load zlib 错...](#) ▶

Linux集群上安装hadoop

博客分类: [hadoop](#)

[Hadoop](#) [Linux](#) [Java](#) [XML](#) [防火墙](#)

两台机器组成一个小实验集群，一台master，一台slave1。

OS: Red Hat Enterprise Linux 5

Java: jdk1.6

hadoop: 0.20.2

master机器的配置情况:

编辑/etc/hosts文件，加上master和slave1的ip地址

编辑hadoop-0.20.2/conf/masters文件，加上"master";

编辑hadoop-0.20.2/conf/slaves文件，加上"slave1";

编辑hadoop-0.20.2/conf/hadoop-env.sh文件，
加上JAVA_HOME路径:

Shell代码

```
1. export JAVA_HOME=/usr/java6/jdk1.6.0_20
```

编辑hadoop-0.20.2/conf/core-site.xml文件，

Xml代码

```
1. <property>
2.     <name>fs.default.name</name>
3.     <value>hdfs://master:9000</value>
```

-- by [hanzhen143437](#)

其他分类

- [我的收藏](#) (16)
- [我的代码](#) (0)
- [我的论坛主题帖](#) (8)
- [我的所有论坛帖](#) (144)
- [我的精华良好帖](#) (0)

最近加入群组

- [lucene爱好者](#)

存档

- [2011-10](#) (2)
- [2011-09](#) (1)
- [2011-08](#) (3)
- [更多存档...](#)



```
4.         </property>
5.         <property>
6.             <name>hadoop.tmp.dir</name>
7.             <value>/hadoop</value>
8.         </property>
```

/root/tmp 是分布式文件系统中的目录结构，在本地的文件系统中也会自动创建该目录结构。

编辑hadoop-0.20.2/conf/hdfs-site.xml文件，

Xml代码

```
1. <property>
2.     <name>dfs.replication</name>
3.     <value>2</value>
4. </property>
5.
6. <property>
7.     <name>dfs.name.dir</name>
8.     <value>/root/hdfs/name</value>
9. </property>
10.
11. <property>
12.     <name>dfs.data.dir</name>
13.     <value>/root/hdfs/data</value>
14. </property>
```

编辑hadoop-0.20.2/conf/mapred-site.xml文件，

Xml代码

```
1. <property>
2.     <name>mapred.job.tracker</name>
3.     <value>master:9001</value>
4. </property>
5. <property>
6.     <name>mapred.job.reuse.jvm.num.tasks</name>
7.     <value>10</value>
8. </property>
```


在master上执行start-all.sh脚本，如果slave1机器上的java进程不能启动。
解决方法：

关闭防火墙，

Java代码

```
1. service iptables stop
```

重新格式化namenode


Java代码 

```
1. | hadoop namenode -format
```

机器需要用ssh来启动和停止各个datanode上的进程，而且是不输入密码的方式，故需要配置SSH使用无密码公钥认证的方式。

在namenode上需要生成一个密钥对，即一个私钥，一个公钥。将公钥拷贝到datanode上，这样，比如当namenode向datanode发起ssh连接的时候，datanode上就会生成一个随机数并用namenode的公钥对这个随机数进行加密，并发送给namenode；namenode收到这个加密的数以后用私钥进行解密，并将解密后的数发送回datanode，datanode确认解密的数无误后就允许namenode进行连接了。这就完成了一次公钥认证过程。


首先在namenode上生成密钥对：

Java代码 

```
1. | ssh-keygen -t rsa
```

这个命令将为namenode上的登录用户生成其密钥对，询问其保存路径时直接回车采用默认路径，当提示要为生成的密钥输入passphrase的时候，直接回车，也就是将其设定为空密码。生成的密钥对id_rsa，id_rsa.pub，默认存储在.ssh目录下。然后将id_rsa.pub的内容复制到每个机器(也包括本机)的.ssh/authorized_keys文件中，如果机器上已经有authorized_keys这个文件了，就在文件末尾加上id_rsa.pub中的内容，如果没有authorized_keys这个文件，直接cp或者scp就好了。


下面的操作假设各个机器上都没有authorized_keys文件。

Java代码 

```
1. | chmod 644 authorized_keys
```


必须保证authorized_keys只对其所有者有读写权限，其他人不允许有写的权限，否则SSH是不会工作的。

至此各个机器上的SSH配置已经完成，可以测试一下了，比如namenode向datanode发起ssh连接

Java代码 

```
1. | ssh slave1
```

别忘了测试本机

Java代码 

```
1. | ssh master
```



◀ [VIM的几个常用插件](#) | [安装rails: no such file to load zlib 错...](#) ▶

17:19 | [评论 / 浏览 \(0 / 132\)](#) | 分类:[编程语言](#) | [相关推荐](#) [▶ MORE](#)

评论

发表评论



[您还没有登录,请您登录后再发表评论](#)