

# CentOS 下 Hadoop-0.20.2 集群配置文档

张鹏

## 一、集群网络环境介绍

集群包含三个节点：1 个 namenode，2 个 datanode，节点之间局域网连接，可以相互 ping 通。节点 IP 地址分布如下：

Namenode: zhangpeng@192.168.90.91

Datanode1: zhangruixue@192.168.90.94

Datanode2: guanxiansun@192.168.90.205

三台节点上均是 CentOS 系统，并且有一个相同的用户 hadoop(具有管理员权限)，在 /home/hadoop/ 目录下均有一个 hadoopinstall 目录，用于存放 hadoop-0.20.2 安装文件，目录结构是 /home/hadoop/hadoopinstall/hadoop-0.20.2。

在 /etc/hosts 上添加主机名和相应的 IP 地址：

```
192.168.90.129  zhangpeng
```

```
192.168.90.94  zhangruixue
```

```
192.168.90.205  guanxiansun
```

## 二、SSH 无密码验证配置

Hadoop 需要使用 SSH 协议，namenode 使用 SSH 无密码登录并启动 datanode 进程，同样原理，datanode 上也能使用 SSH 无密码登录到 namenode（这样也能在 datanode 上进行启动或关闭 hadoop）。

### 1、安装和启动 SSH 协议

所有机器上安装 SSH 协议并启动服务，在所有机器上执行以下命令：

```
$ sudo yum install ssh  安装 SSH 协议
```

```
$ sudo yum install rsync  (rsync 是一个远程数据同步工具，可通过 LAN/WAN 快速同步多台主机间的文件)
```

```
$ ssh  sudo /etc/init.d/ssh restart  启动服务
```

命令执行完毕，各台机器之间可以通过密码验证相互登陆。

### 2、配置 Namenode 无密码登录所有 Datanode

#### (1) 原理

Namenode 作为客户端，要实现无密码公钥认证，连接到服务端 datanode 上时，需要在 namenode 上生成一个密钥对，包括一个公钥和一个私钥，而后将公钥复制到 datanode 上。当 namenode 通过 ssh 连接 datanode 时，datanode 就会生成一个随机数并用 namenode 的公钥对随机数进行加密，并发送给 namenode。namenode 收到加密数之后再用私钥进行解密，并将解密数回传给 datanode，datanode 确认解密数无误之后就允许 namenode 进行连接了。这就是一个公钥认证过程，其间不需要用户手工输入密码。重要过

程是将客户端 namenode 公钥复制到 datanode 上。

(2) Namenod 机器上生成密码对

在 Namenode 节点上执行以下命令:

```
hadoop@zhangpeng:~$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hadoop/.ssh/id_rsa): 默认路径
Enter passphrase (empty for no passphrase): 回车, 空密码
Enter same passphrase again:
Your identification has been saved in /home/hadoop/.ssh/id_rsa.
Your public key has been saved in /home/hadoop/.ssh/id_rsa.pub.
```

这将在/home/hadoop/.ssh/目录下生成一个私钥 id\_rsa 和一个公钥 id\_rsa.pub。

在 namenode 节点上做如下配置

```
hadoop@zhangpeng:~/.ssh$ cp id_rsa.pub authorized_keys (namenode 的公钥)
```

使用 SSH 协议将 namenode 的公钥信息 authorized\_keys 复制到所有 DataNode 的.ssh 目录下(.ssh 下最初没有 authorized\_keys, 如果有, 则需要追加)。

```
hadoop@zhangpeng:~$ scp authorized_keys hadoop@zhangruixue:/home/hadoop/.ssh
hadoop@zhangpeng:~$ scp authorized_keys hadoop@guanxiansun:/home/hadoop/.ssh
```

这样配置过后, namenode 可以无密码登录所有 datanode, 可以通过命令 “ssh 节点 ip 地址” 来验证。

### 3、配置 Datanode 无密码登录 Namenode

和 namenode 无密码登录所有 Datanode 原理一样, 把 Datanode 的公钥复制到 Namenode 的.ssh 目录下。

```
hadoop@zhangruixue:~$ scp authorized_keys hadoop@zhangpeng:/home/hadoop/.ssh
hadoop@guanxiansun:~$ scp authorized_keys hadoop@zhangpeng:/home/hadoop/.ssh
```

这样也能在 Datanode 上关闭和启动 Hadoop 服务。

注意: 整个过程中只涉及到创建密钥, 复制公钥, 添加公钥内容, 没有更改配置文件, 实际上须在 Namenode 上编辑配置文件/etc/ssh/sshd\_config 开启其公钥验证:

```
RSAAuthentication yes
PubkeyAuthentication yes
```

至此, namenode 能无密码验证登陆每个 datanode。每个 datanode 也能无密码验证 SSH 登陆到 namenode。

## 三、JDK 安装和 Java 环境变量配置

### 1、CentOS 下安装 JDK 1.6

所有机器上安装 JDK 1.6, 在所有机器上执行以下命令:

```
$ sudo yum install jdk
```

JDK 默认安装在/usr/java/jdk1.6.0\_14 目录里。

## 2、Java 环境变量配置

在所有机器上执行如下命令：vim /etc/profile。并加入以下内容，配置环境变量

```
# set java environment
export JAVA_HOME="/usr/java/jdk1.6.0_14/"
export JAVA_HOME="/usr/java/jdk1.6.0_14/jre"
export CLASSPATH=.:$JAVA_HOME/lib:$JRE_HOME/lib:$CLASSPATH
export PATH=$JAVA_HOME/bin:$JRE_HOME/bin:$PATH
```

保存并退出，执行以下命令使配置生效：

```
[hadoop@zhangpeng ~]$ source /etc/profile
[hadoop@zhangruixue ~]$ source /etc/profile
[hadoop@guanxiansun ~]$ source /etc/profile
```

配置完毕，在命令行中使用命令“java -version”可以判断是否成功。

## 四、Hadoop 集群配置

### 1、下载 *hadoop-0.20.2.tar.gz*

在 **namenode** 上下载 *hadoop-0.20.2.tar.gz*，将其解压到/home/hadoop/hadoopinstall 目录下，使用如下命令：

```
[hadoop@zhangpeng hadoopinstall]$ tar -xvzf hadoop-0.20.2.tar
```

然后在 hadoopinstall 下创建 tmp 文件夹并把 Hadoop 的安装路径添加到/etc/profile 中修改/etc/profile 文件(配置 java 环境变量的文件)，将以下语句添加到末尾，并使其有效(source /etc/profile)：

```
# set hadoop path
export HADOOP_HOME=/home/hadoop/hadoopinstall/hadoop-0.20.1
export PATH=$HADOOP_HOME/bin:$PATH
```

### 2、编辑 *Hadoop* 的配置文件

(1) 配置 *hadoop-env.sh*

```
[hadoop@zhangpeng conf]$ vim hadoop-env.sh
```

末尾加入以下语句

```
# set java environment
export JAVA_HOME=/usr/java/jdk1.6.0_14/
```

hadoop 配置文件在 conf 目录下，之前的版本 Hadoop 的配置文件主要是 Hadoop-default.xml 和 Hadoop-site.xml。由于 hadoop 发展迅速，代码量急剧增加，代码开发分为了 core,hdfs 和 map/reduce 三部分，配置文件也被分成了三个 core-site.xml、hdfs-site.xml、mapred-site.xml。

(2) 配置 *core-site.xml* 文件

```
<configuration>
  <property>
```

```

        <name>hadoop.tmp.dir</name>
        <value>/home/hadoop/hadoopinstall/tmp</value>
        (注意, 请先在 hadoopinstall 目录下建立 tmp 文件夹)
        <description>A base for other temporary directories.</description>
    </property>
<!-- file system properties -->
    <property>
        <name>fs.default.name</name>
        <value>hdfs://192.168.90.91:9000</value>
    </property>
</configuration>

```

### (3) 配置 hdfs-site.xml 文件

```

<configuration>
    <property>
        <name>dfs.replication</name>
        <value>1</value>
        (replication 是数据副本数量, 默认为 3, datanode 少于 3 台就会报错)
    </property>
</configuration>

```

### (4) 配置 mapred-site.xml 文件

```

<configuration>
    <property>
        <name>mapred.job.tracker</name>
        <value>http://192.168.90.91:9001</value>
    </property>
</configuration>

```

### (5) 配置 masters 文件, 加入 namenode 的 ip 地址

```
192.168.90.91
```

### (6) 配置 slaves 文件, 加入所有 datanode 的 ip 地址

```
192.168.90.94
192.168.90.205
```

将 namenode 上配置好的 hadoop 所在文件夹 hadoopinstall 复制到 datanode 的 /home/hadoop/ 目录下 (实际上 masters, slaves 文件时不必要的, 复制了也没问题)。

在 datanode 上修改 /etc/profile 文件 (配置 java 环境变量的文件), 将以下语句添加到末尾, 并使其有效 (source /etc/profile):

```

# set hadoop path
export HADOOP_HOME=/home/zk1/hadoopinstall/hadoop-0.20.1
export PATH=$HADOOP_HOME/bin:$PATH

```

## 五、Hadoop 集群启动

在 Namenode 格式化 hadoop:

```
[hadoop@zhangpeng ~]$ hadoop namenode -format
```

启动 hadoop:

```
[hadoop@zhangpeng ~]$ start-all.sh
```

可以通过以下启动日志看出, 首先启动 namenode 接着启动 datanode1, datanode2, 然后启动 secondarynamenode。再启动 jobtracker, 然后启动 tasktracker1, 最后启动 tasktracker2。

启动 hadoop 成功后, 在 namenode 中的 tmp 文件夹中生成了 dfs 文件夹, 在 datanode 中的 tmp 文件夹中均生成了 dfs 文件夹和 mapred 文件夹。

在 namenode 上用 java 自带的小工具 jps 查看进程

```
[hadoop@zhangpeng ~]$ /usr/java/jdk1.6.0_14/bin/jps
18978 JobTracker
21242 Jps
18899 SecondaryNameNode
18731 NameNode
```

在每个 datanode 上查看进程

```
[hadoop@zhangruixue ~]$ /usr/java/jdk1.6.0_14/bin/jps
17706 TaskTracker
20499 Jps
17592 DataNode
[hadoop@guanxiansun ~]$ /usr/java/jdk1.6.0_14/bin/jps
28550 TaskTracker
28436 DataNode
30798 Jps
```

在 namenode 上查看集群状态:

```
[hadoop@zhangpeng ~]$ hadoop dfsadmin -report
Configured Capacity: 123909840896 (115.4 GB)
Present Capacity: 65765638144 (61.25 GB)
DFS Remaining: 65765257216 (61.25 GB)
DFS Used: 380928 (372 KB)
DFS Used%: 0%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0
```

---

```
Datanodes available: 2 (2 total, 0 dead)
```

Name: 192.168.90.205:50010  
Decommission Status : Normal  
Configured Capacity: 61954920448 (57.7 GB)  
DFS Used: 258048 (252 KB)  
Non DFS Used: 47925145600 (44.63 GB)  
DFS Remaining: 14029516800(13.07 GB)  
DFS Used%: 0%  
DFS Remaining%: 22.64%  
Last contact: Thu Jul 22 15:11:28 CST 2010

Name: 192.168.90.94:50010  
Decommission Status : Normal  
Configured Capacity: 61954920448 (57.7 GB)  
DFS Used: 122880 (120 KB)  
Non DFS Used: 10219057152 (9.52 GB)  
DFS Remaining: 51735740416(48.18 GB)  
DFS Used%: 0%  
DFS Remaining%: 83.51%  
Last contact: Thu Jul 22 15:11:27 CST 2010

## 在 Windows 上安装 Hadoop 教程

— 见 2010.1.6 [www.hadoopor.com/hadoopor@foxmail.com](http://www.hadoopor.com/hadoopor@foxmail.com)

### 1. 安装 JDK

不建议只安装 JRE，而是建议直接安装 JDK，因为安装 JDK 时，可以同时安装 JRE。MapReduce 程序的编写和 Hadoop 的编译都依赖于 JDK，光 JRE 是不够的。

JRE 下载地址：[http://www.java.com/zh\\_CN/download/manual.jsp](http://www.java.com/zh_CN/download/manual.jsp)

JDK 下载地址：<http://java.sun.com/javase/downloads/index.jsp>，下载 Java SE 即可。

### 2. 安装 Cygwin

在安装 Cygwin 之前，得先下载 Cygwin 安装程序 setup.exe。

Cygwin 安装程序下载地址：<http://www.cygwin.com/setup.exe>，当然也可以从<http://www.cygwin.cn/setup.exe>下载 Cygwin 安装程序，不过如果在安装过程中，遇到如下图所示的错误，则只能从<http://www.cygwin.com/setup.exe>下载，本教程下载的是 Cygwin 1.7.1 版本。



Cygwin 安装程序 setup.exe 的存放目录可随意无要求。当 setup.exe 下载成功后，运行 setup.exe，弹出如下图所示的对话框：

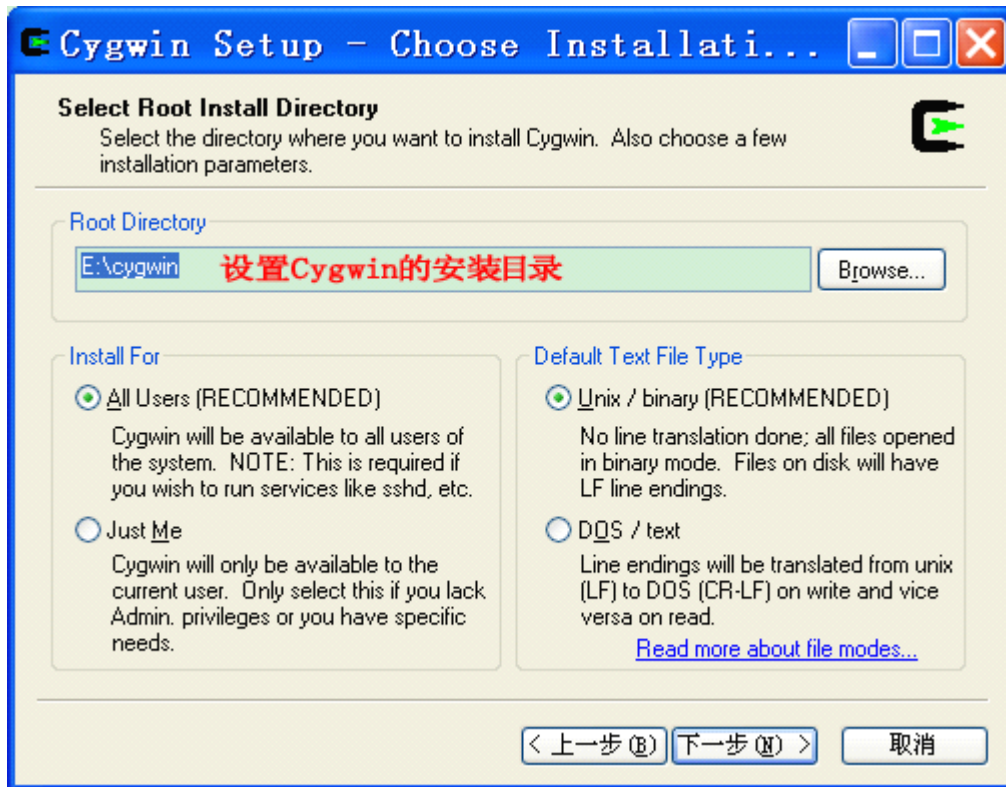


在上图所示的对话框中，直接点击“下一步”，进入如下图所示的对话框：

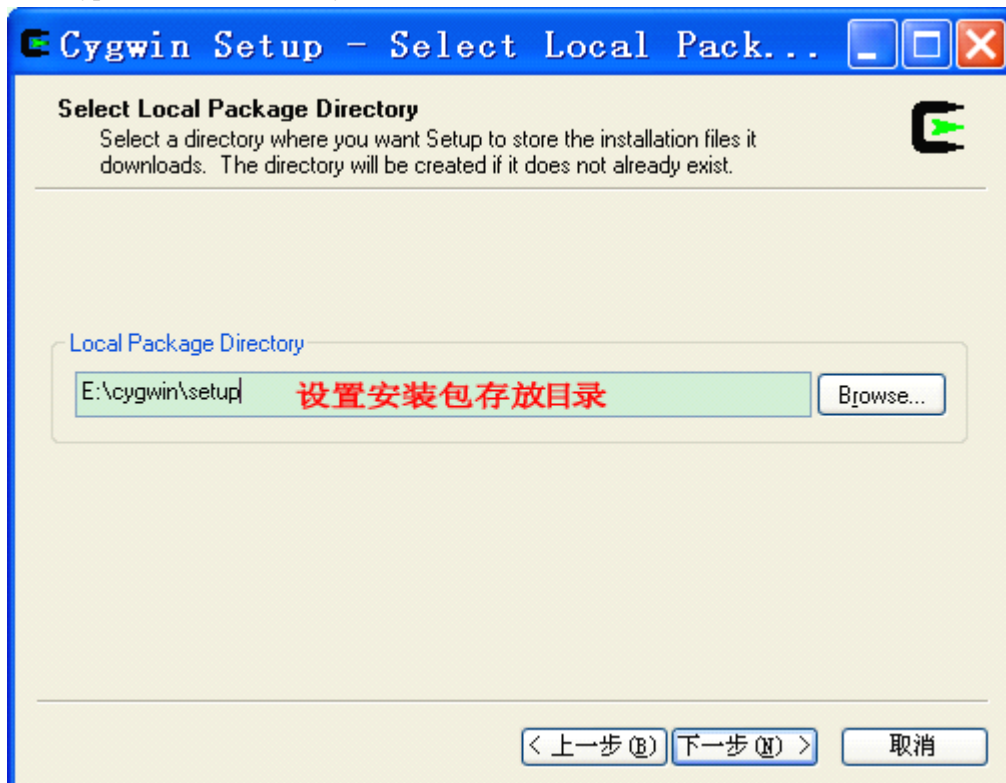


在上图所示的对话框中，选择“Install from Internet”，然后点击“下一步”，进入如下图所示对话框：

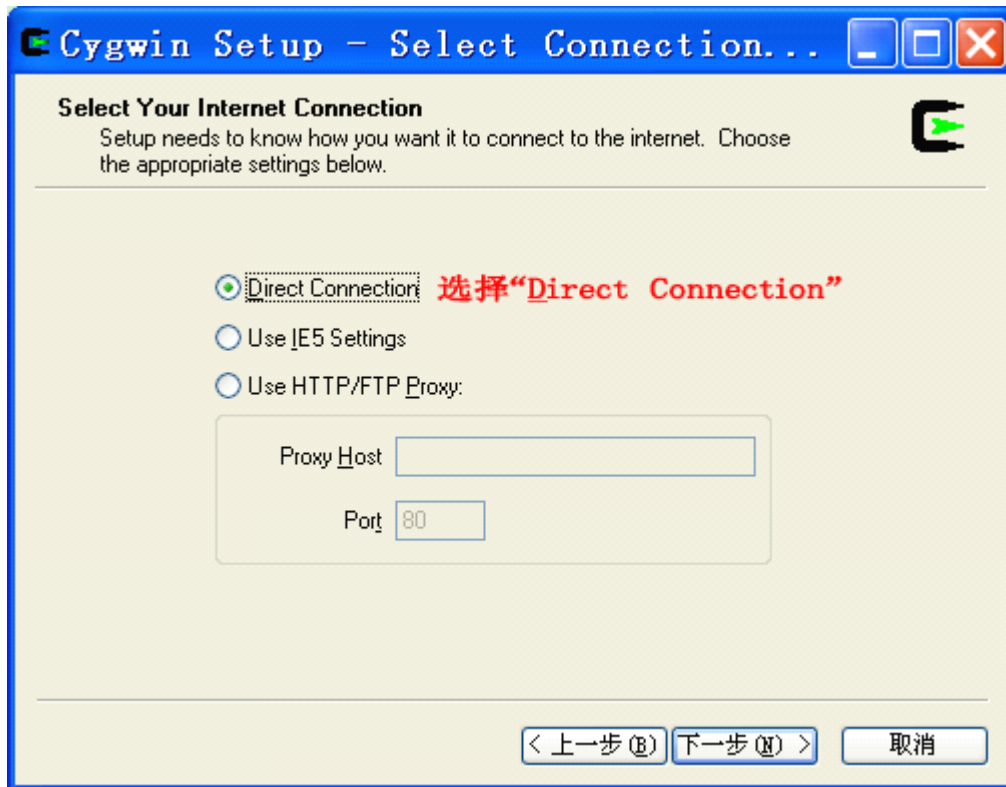




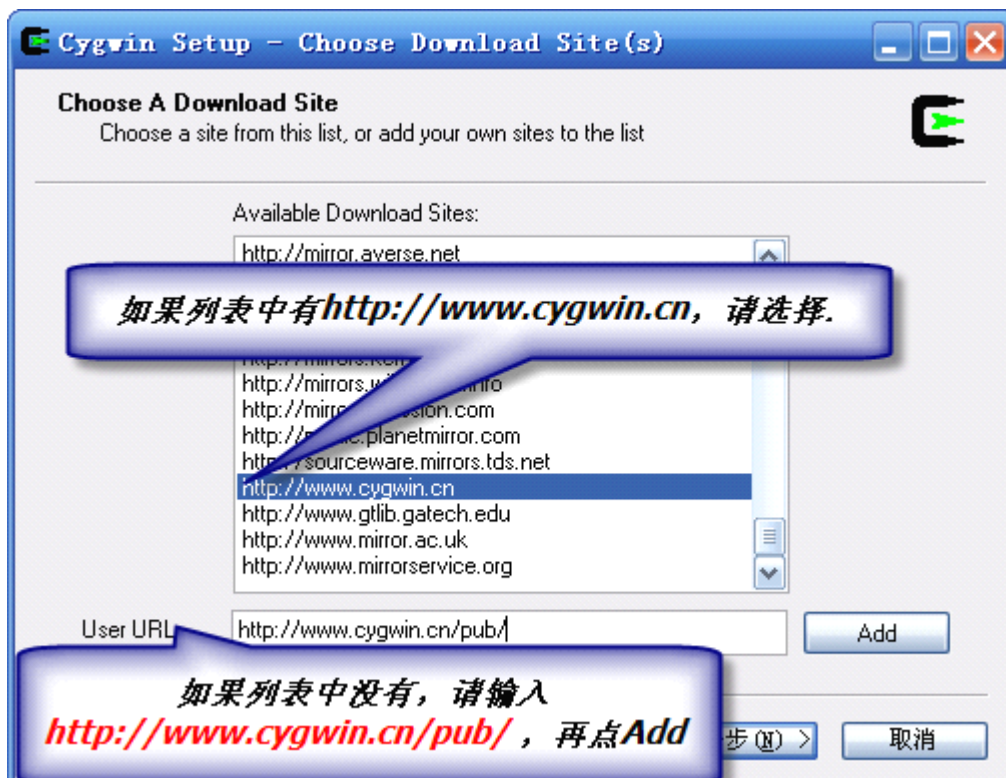
在上图所示的对话框中，设置 Cygwin 的安装目录，Install For 选择 “All Users”，Default Text File Type 选择 “Unix/binary”，然后点击 “下一步”，进入如下图所示对话框：



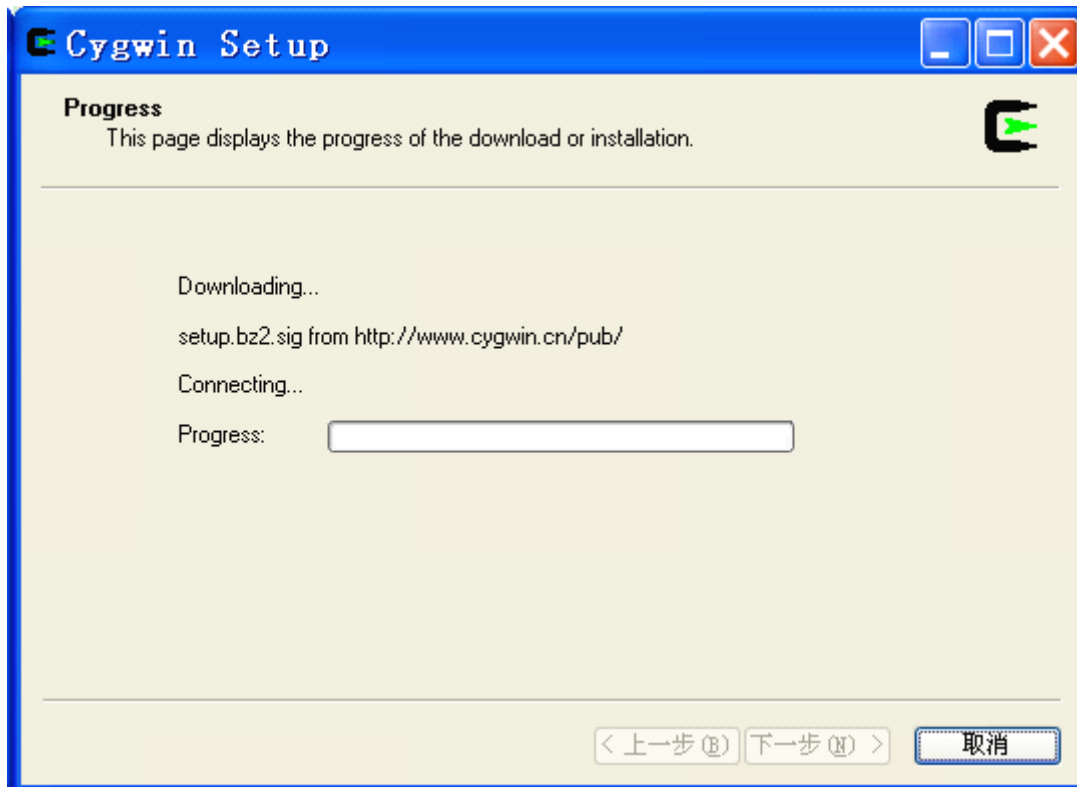
在上图所示的对话框中，设置 Cygwin 安装包存放目录，然后点击 “下一步”，进入如下图所示对话框：



在上图所示的对话框中，选择“Direct Connection”，然后点击“下一步”，进入如下图所示对话框：



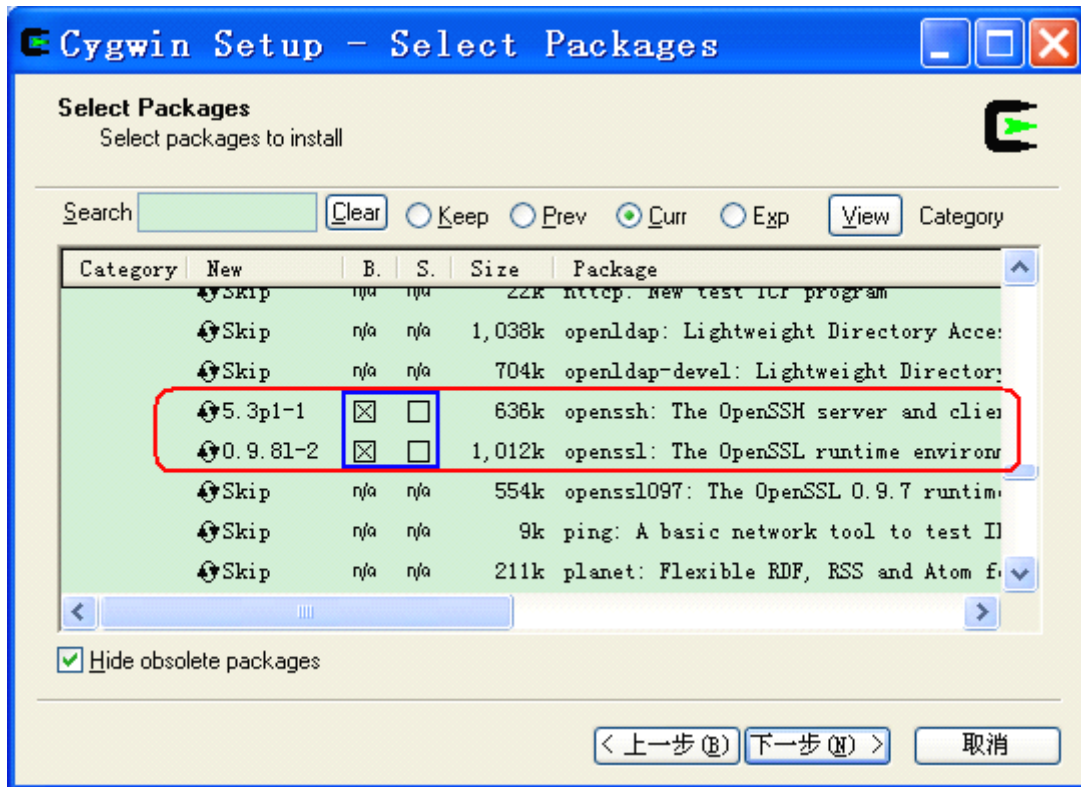
在上图所示的对话框中，点击“下一步”，将进入如下图所示的对话框：



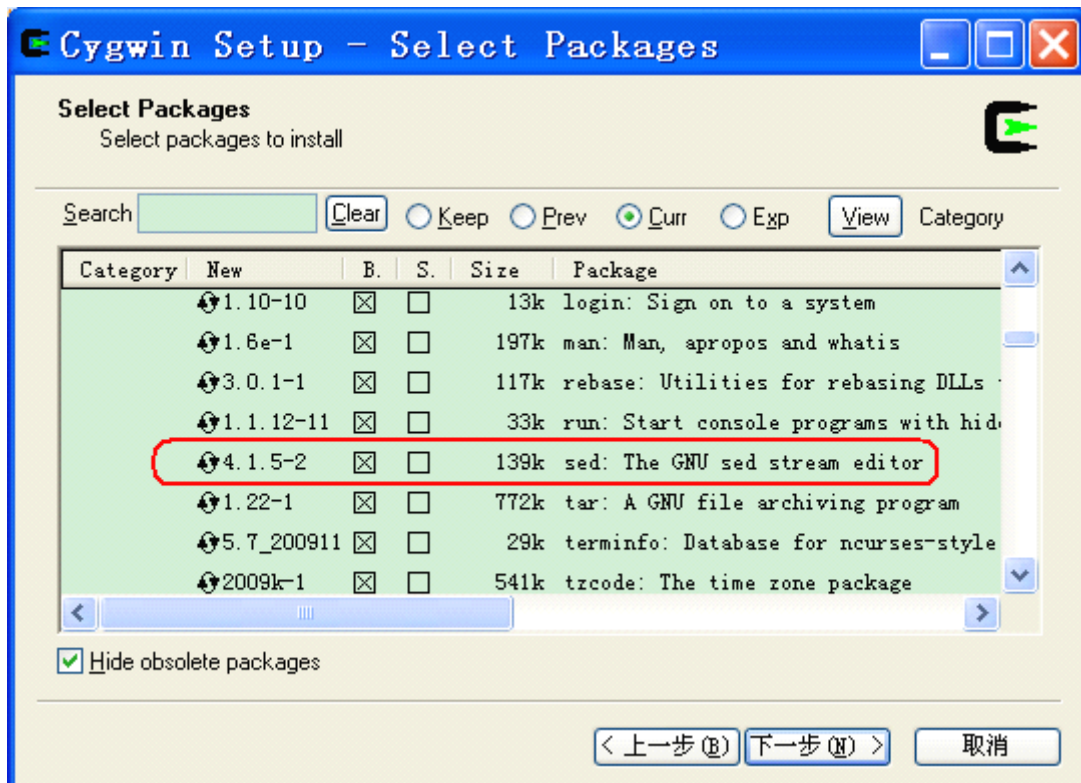
在上图所示的对话框过程中，可能会弹出如下图所示的“Setup Alert”对话框，直接点击“确定”即可。



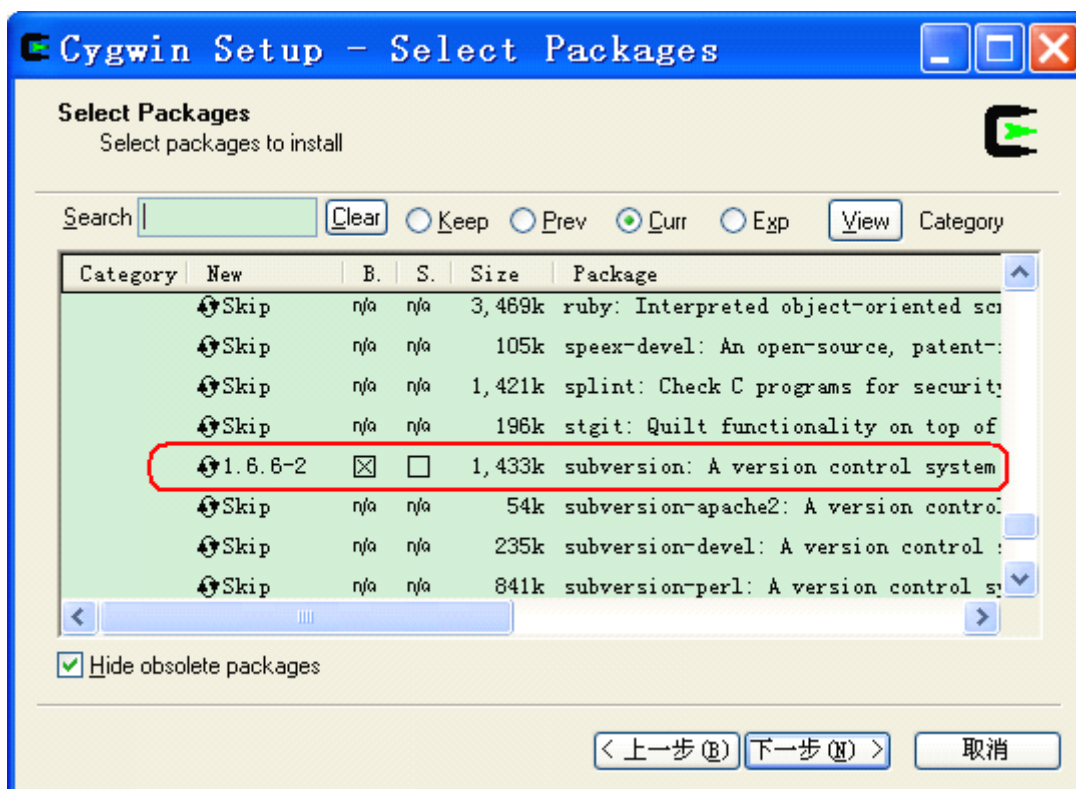
进入“Select Packages”对话框后，必须保证“Net Category”下的“OpenSSL”被安装，如下图所示：



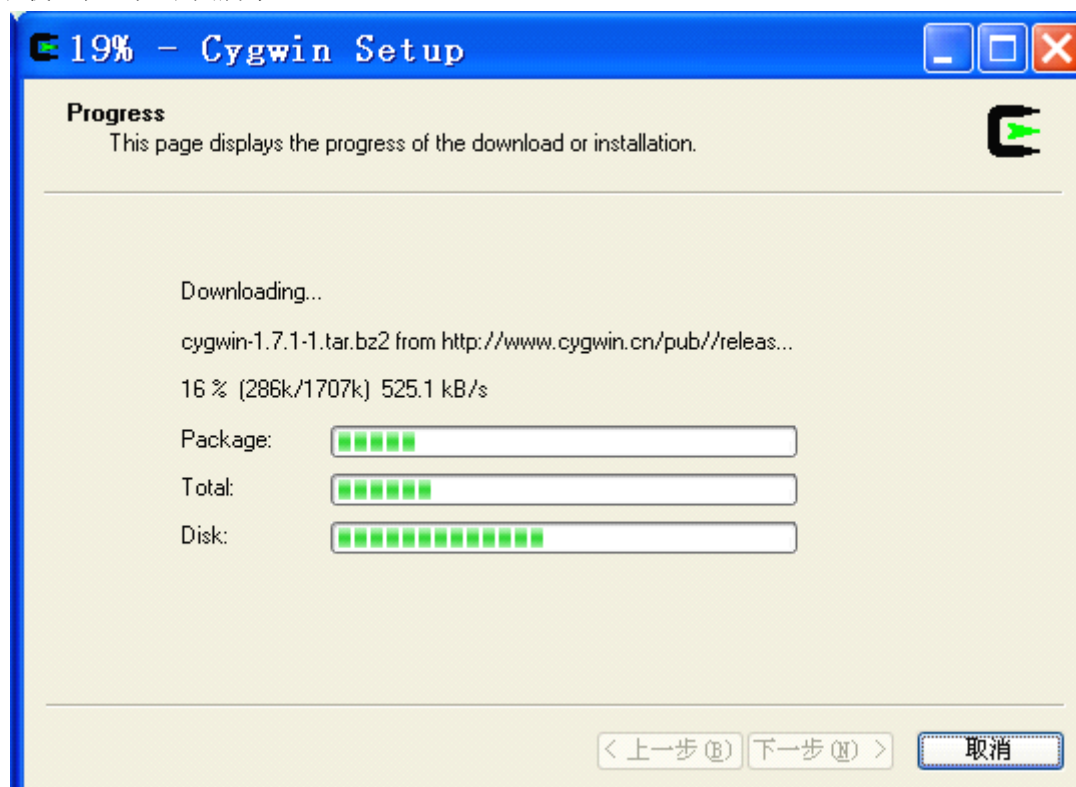
如果还打算在 eclipse 上编译 Hadoop，则还必须安装“Base Category”下的“sed”，如下图所示：



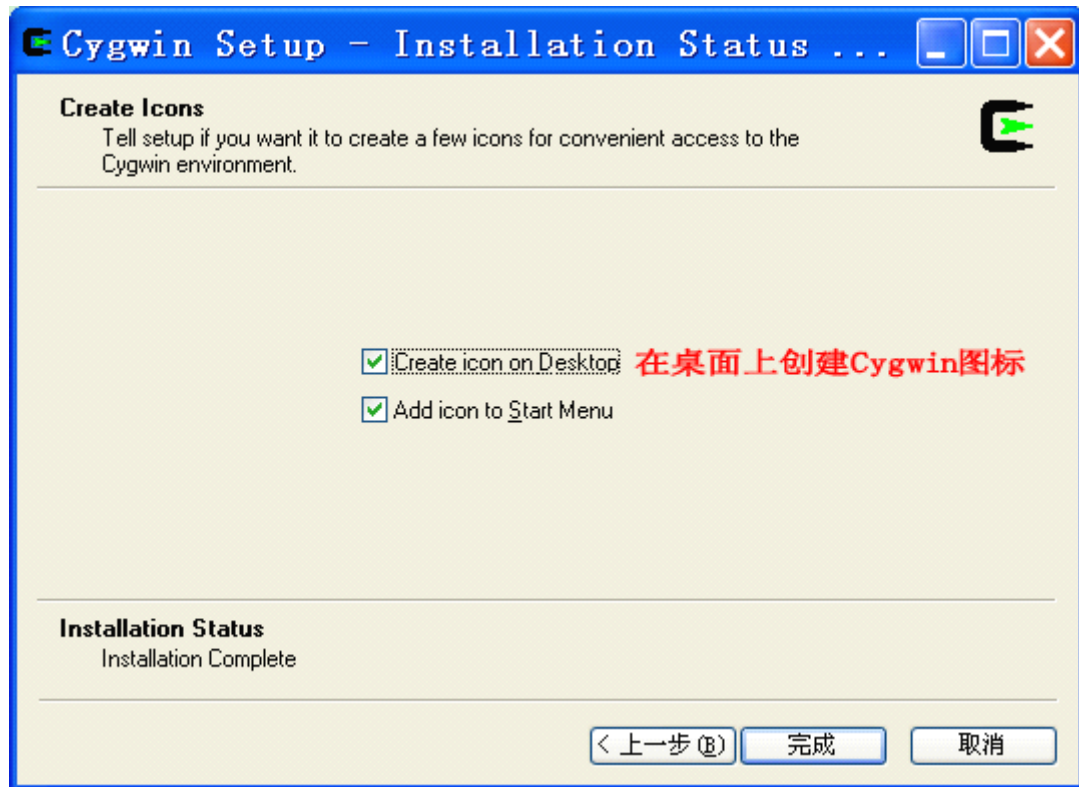
另外，还建议将“Editors Category”下的 vim 安装，以方便在 Cygwin 上直接修改配置文件；“Devel Category”下的 subversion 建议安装，如下图所示：



当完成上述操作后，点击“Select Packages”对话框中“下一步”，进入 Cygwin 安装包下载过程，如下图所示：



等待安装包下载完毕，当下载完后，会自动进入到如下图所示的对话框：

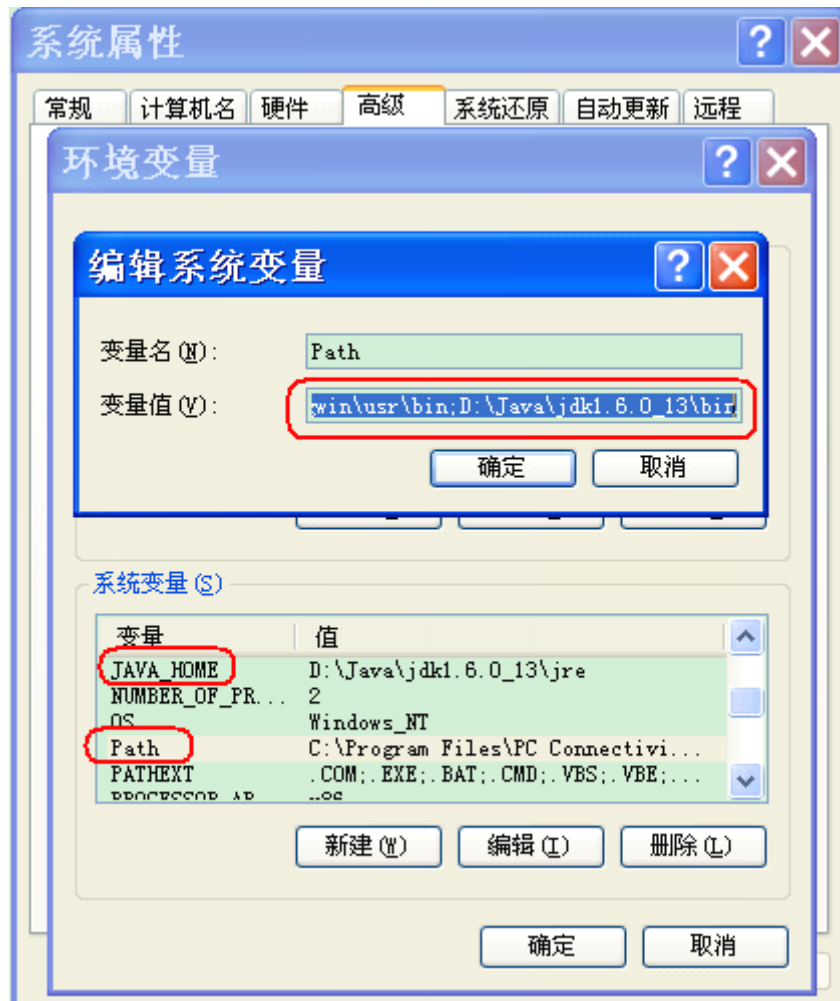


在上图所示的对话框中，选中“Create icon on Desktop”，以方便直接从桌面上启动 Cygwin，然后点击“完成”按钮。至此，Cygwin 已经安装完，安装目录下的内容如下图所示：



### 3. 配置环境变量

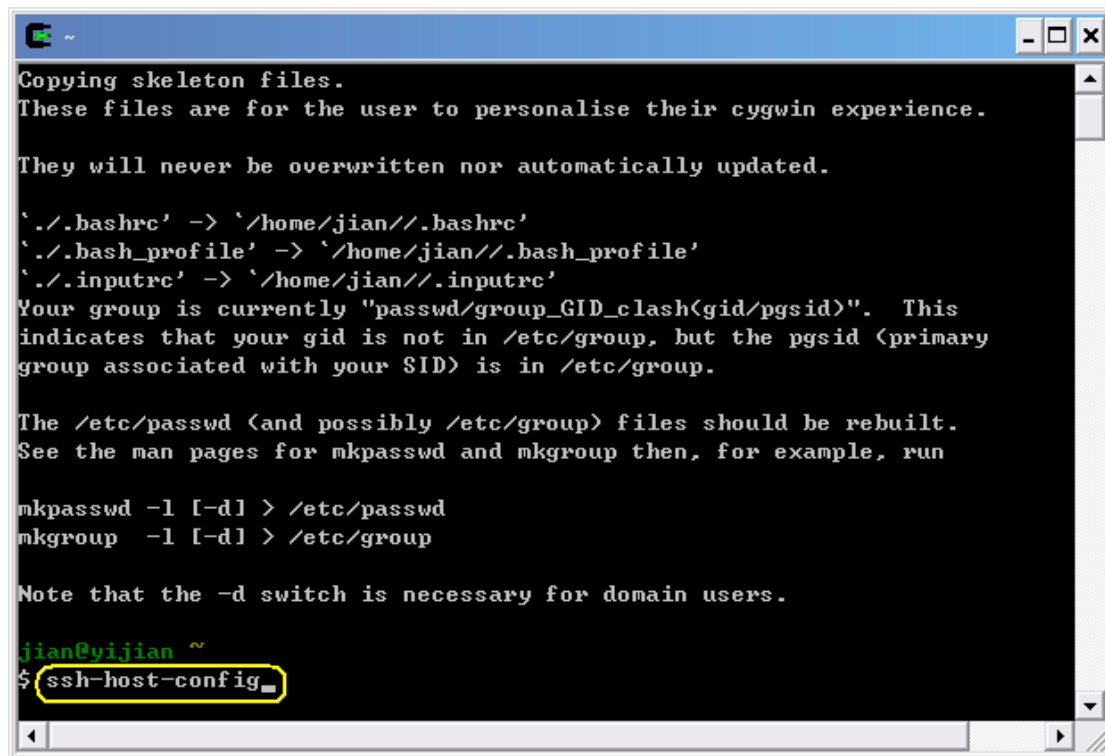
需要配置的环境变量包括 PATH 和 JAVA\_HOME: JAVA\_HOME 指向 JRE 安装目录; JDK 的 bin 目录, Cygwin 的 bin 目录, 以及 Cygwin 的 usr\bin 目录都必须添加到 PATH 环境变量中, 如下图所示:



## 4. 安装 sshd 服务

点击桌面上的 Cygwin 图标，启动 Cygwin，执行 **ssh-host-config** 命令，如下图所示：





```
Copying skeleton files.
These files are for the user to personalise their cygwin experience.

They will never be overwritten nor automatically updated.

'./.bashrc' -> '/home/jian/./.bashrc'
'./.bash_profile' -> '/home/jian/./.bash_profile'
'./.inputrc' -> '/home/jian/./.inputrc'
Your group is currently "passwd/group_GID_clash(gid/pgsid)". This
indicates that your gid is not in /etc/group, but the pgsid <primary
group associated with your SID> is in /etc/group.

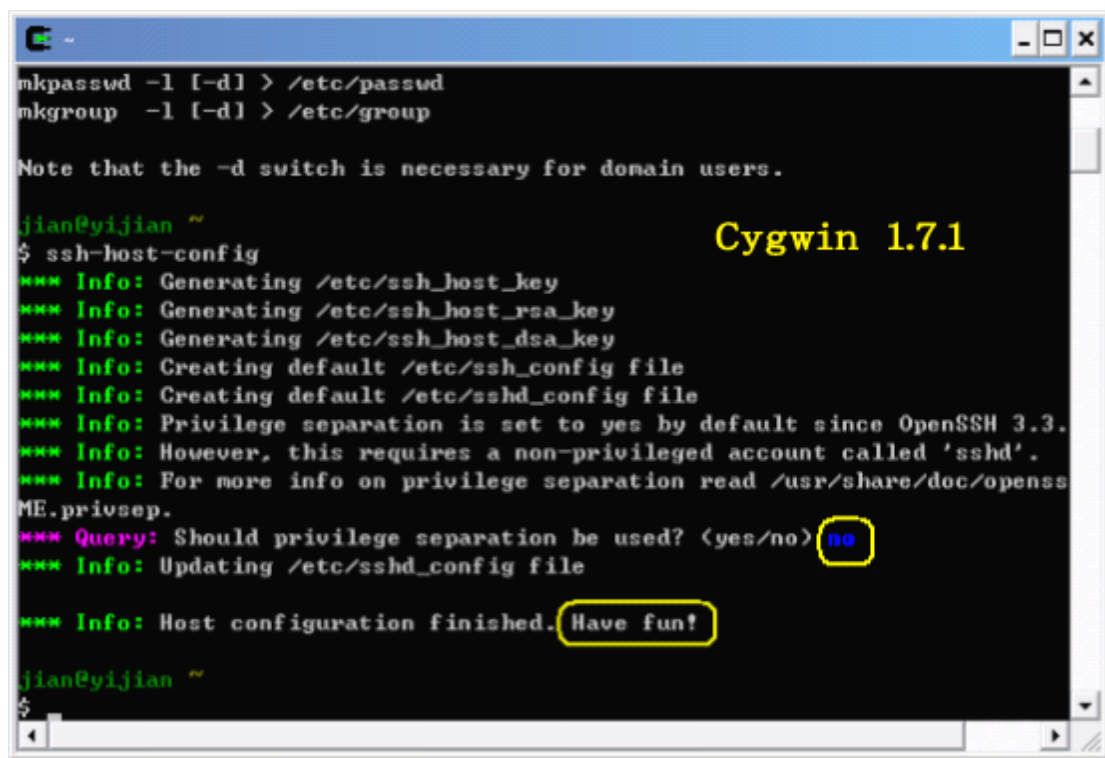
The /etc/passwd <and possibly /etc/group> files should be rebuilt.
See the man pages for mkpasswd and mkgroup then, for example, run

mkpasswd -l [-d] > /etc/passwd
mkgroup -l [-d] > /etc/group

Note that the -d switch is necessary for domain users.

jian@yijian ~
$ ssh-host-config
```

在执行 ssh-host-config 时，当要求输入 yes/no 时，选择输入 no，如下图所示：



```
mkpasswd -l [-d] > /etc/passwd
mkgroup -l [-d] > /etc/group

Note that the -d switch is necessary for domain users.

jian@yijian ~
$ ssh-host-config
*** Info: Generating /etc/ssh_host_key
*** Info: Generating /etc/ssh_host_rsa_key
*** Info: Generating /etc/ssh_host_dsa_key
*** Info: Creating default /etc/ssh_config file
*** Info: Creating default /etc/sshd_config file
*** Info: Privilege separation is set to yes by default since OpenSSH 3.3.
*** Info: However, this requires a non-privileged account called 'sshd'.
*** Info: For more info on privilege separation read /usr/share/doc/openssh
ME.privsep.
*** Query: Should privilege separation be used? <yes/no> no
*** Info: Updating /etc/sshd_config file

*** Info: Host configuration finished. Have fun!

jian@yijian ~
$
```

如果是 Cygwin 1.7 之前的版本，则 ssh-host-config 显示界面如下图所示：



```

$ ssh-host-config
Generating /etc/ssh_host_key
Generating /etc/ssh_host_rsa_key
Generating /etc/ssh_host_dsa_key
Generating /etc/ssh_config file
Privilege separation is set to yes by default since OpenSSH 3.3.
However, this requires a non-privileged account called 'sshd'.
For more info on privilege separation read /usr/share/doc/openssh/R
.

Should privilege separation be used? <yes/no> no
Generating /etc/sshd_config file
Added ssh to C:\WINDOWS\system32\drivers\etc\services

Warning: The following functions require administrator privileges!

Do you want to install sshd as service?
<Say "no" if it's already installed as service> <yes/no> yes

Which value should the environment variable CYGWIN have when
sshd starts? It's recommended to set at least "ntsec" to be
able to change user context without password.
Default is "ntsec". CYGWIN=ntsec

The service has been installed under LocalSystem account.
To start the service, call 'net start sshd' or 'cygrunsrv -S sshd'.

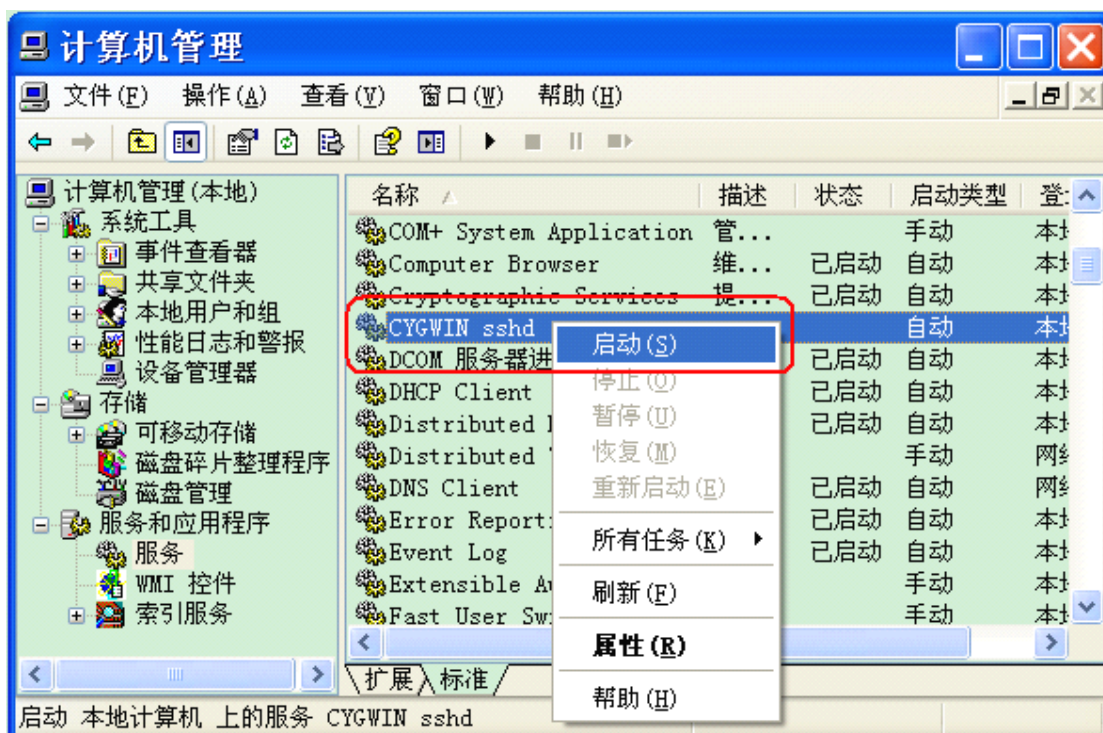
Host configuration finished. Have fun!

```

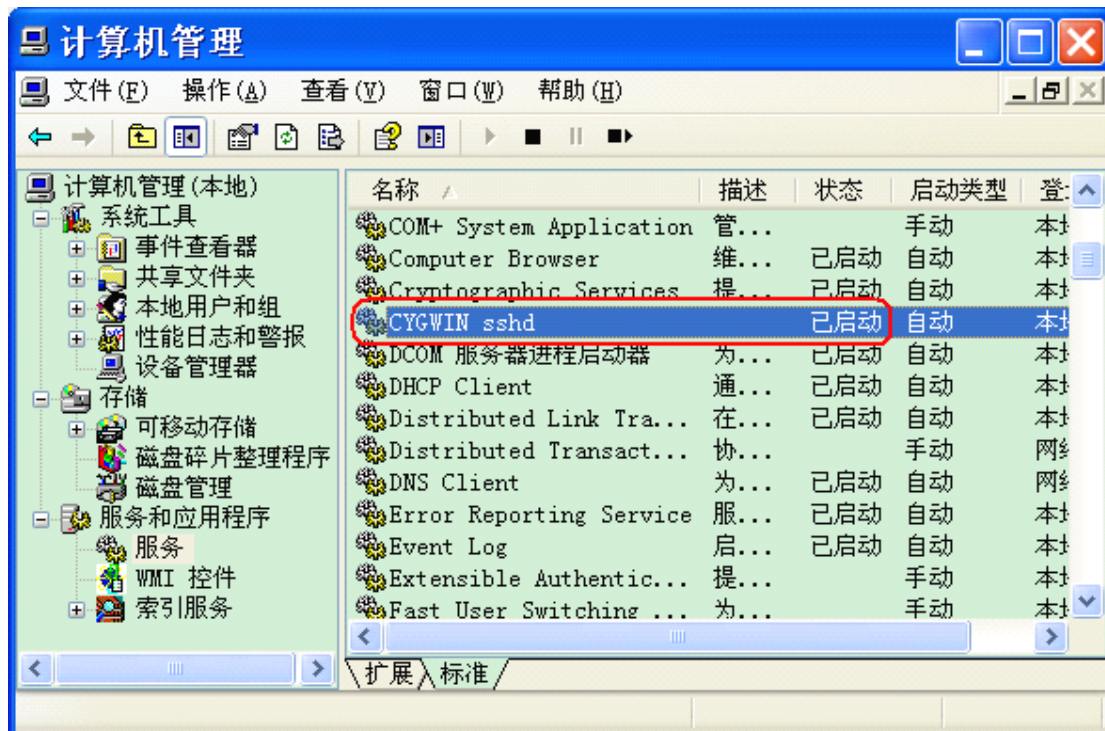
当看到“Have fun”时，一般表示 sshd 服务安装成功了，如上图所示。接下来，需要启动 sshd 服务。

## 5. 启动 sshd 服务

在桌面上的“我的电脑”图标上单击右键，点击“管理”菜单，进入 Windows 计算机管理，如下图所示：



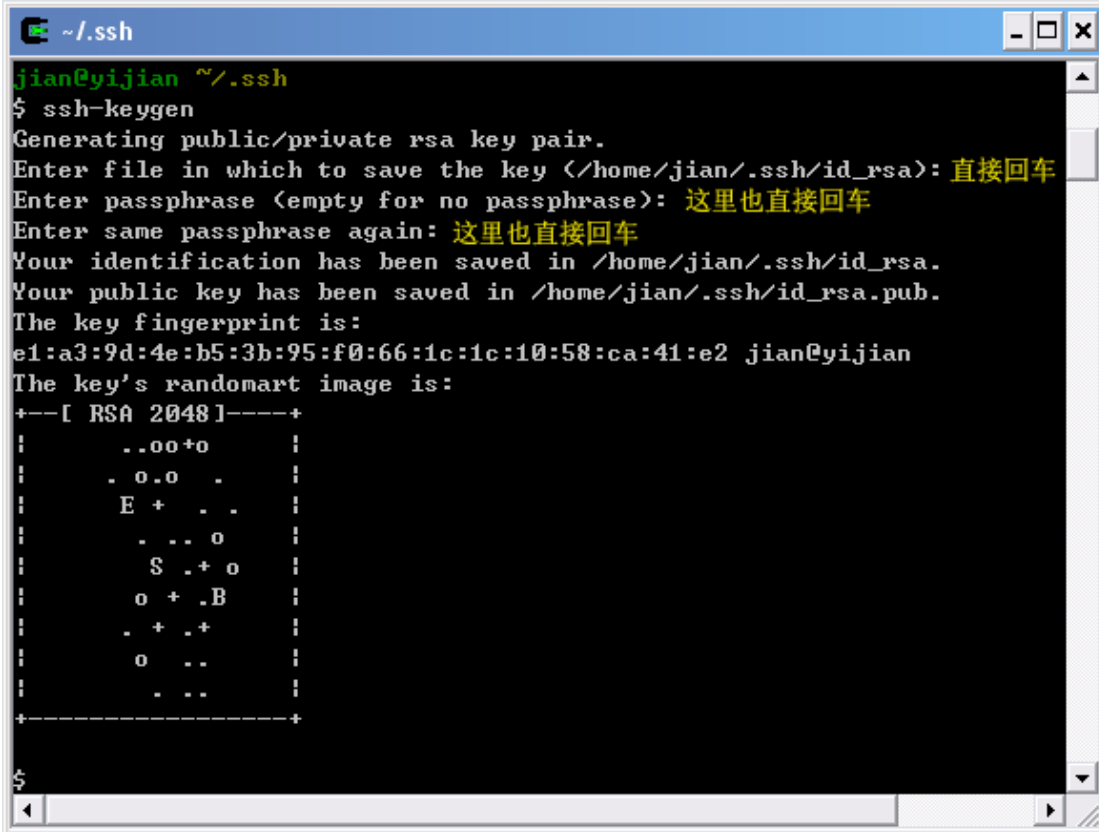
在上图所示的对话框中，选中“**CYGWIN sshd**”，弹出右键，并启动 CYGWIN sshd 服务，成功后，如下图所示：



当 CYGWIN sshd 的状态为“已启动”后，接下来就是配置 ssh 登录。

## 6. 配置 ssh 登录

执行 **ssh-keygen** 命令生成密钥文件，如下图所示：

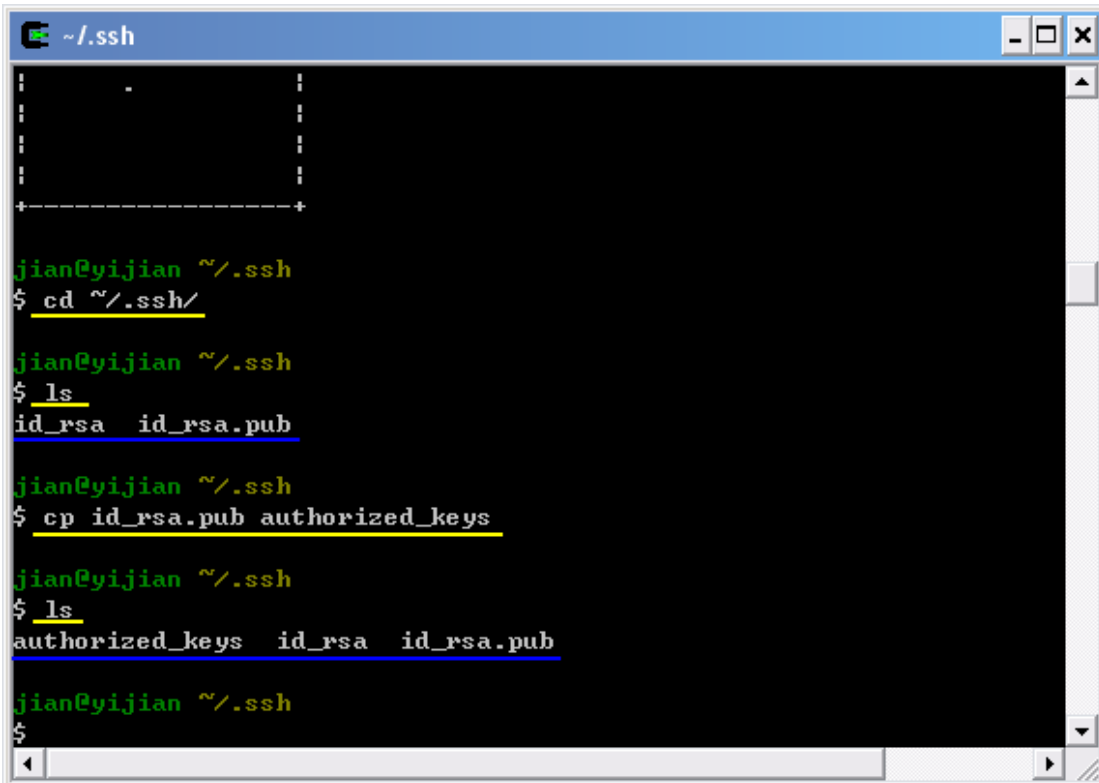


```

~/.ssh
jian@yijian ~/.ssh
$ ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/home/jian/.ssh/id_rsa): 直接回车
Enter passphrase (empty for no passphrase): 这里也直接回车
Enter same passphrase again: 这里也直接回车
Your identification has been saved in /home/jian/.ssh/id_rsa.
Your public key has been saved in /home/jian/.ssh/id_rsa.pub.
The key fingerprint is:
e1:a3:9d:4e:b5:3b:95:f0:66:1c:1c:10:58:ca:41:e2 jian@yijian
The key's randomart image is:
+--[ RSA 2048 ]-----+
|      ..oo+o      |
|     . o.o .      |
|    E + . .       |
|     . . . o      |
|    $ . + o       |
|    o + .B        |
|   . + .+         |
|    o ..          |
|     . ..         |
+-----+
$

```

在上图所示对话框中，需要输入时，直接按回车键即可，如果不出错，应当是需要三次按回车键。接下来生成 **authorized\_keys** 文件，按下图所示操作即可：



```

~/.ssh
jian@yijian ~/.ssh
$ cd ~/.ssh/

jian@yijian ~/.ssh
$ ls
id_rsa id_rsa.pub

jian@yijian ~/.ssh
$ cp id_rsa.pub authorized_keys

jian@yijian ~/.ssh
$ ls
authorized_keys id_rsa id_rsa.pub

jian@yijian ~/.ssh
$

```

正如下图所示，只需要两步操作，即可生成 **authorized\_keys** 文件：

```
cd ~/.ssh/
```

```
cp id_rsa.pub authorized_keys
```

完成上述操作后, 执行 **exit** 命令先退出 Cygwin 窗口, 如果不执行这一步操作, 下面的操作可能会遇到错误。接下来, 重新运行 Cygwin, 执行 **ssh localhost** 命令, 在第一次执行 ssh localhost 时, 会有如下图所示的提示, 输入 **yes**, 然后回车即可:

```

~/.ssh
Your group is currently "passwd/group_GID_clash(gid/pgsid)". This
indicates that your gid is not in /etc/group, but the pgsid (primary
group associated with your SID) is in /etc/group.

The /etc/passwd (and possibly /etc/group) files should be rebuilt.
See the man pages for mkpasswd and mkgroup then, for example, run

mkpasswd -l [-dl] > /etc/passwd
mkgroup -l [-dl] > /etc/group

Note that the -d switch is necessary for domain users.

jian@vijian ~
$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
RSA key fingerprint is d7:0f:a4:56:be:43:15:9c:f2:02:ac:24:62:5a:ac:
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'localhost' (RSA) to the list of known ho
Last login: Wed Jan  6 22:25:59 2010 from localhost
$

```

如果是 Windows 域用户, 这步操作可能会遇到问题, 错误信息如下:。

```

$ ssh localhost
Last login: Thu Jan  7 10:23:09 2010 from localhost
3 [main] -bash 5016 E:\cygwin\bin\bash.exe: *** fatal e
amically determine load address for 'WSAGetLastError' (handle
error 126
Connection to localhost closed.

```

这个错误暂无解决办法, 问题的解决情况, 可关注 Hadoop 技术论坛中的贴:  
<http://bbs.hadoopor.com/thread-348-1-1.html>(Cygwin 1.7.1 版本 ssh 问题)。否则, 如果成功, 执行 who 命令时, 可以看到如下图所示的信息:

```

$ who
jian      tty0      2010-01-06 22:51 (localhost)

```

至此, 配置 ssh 登录成功, 下面就可以开始安装 hadoop 了。

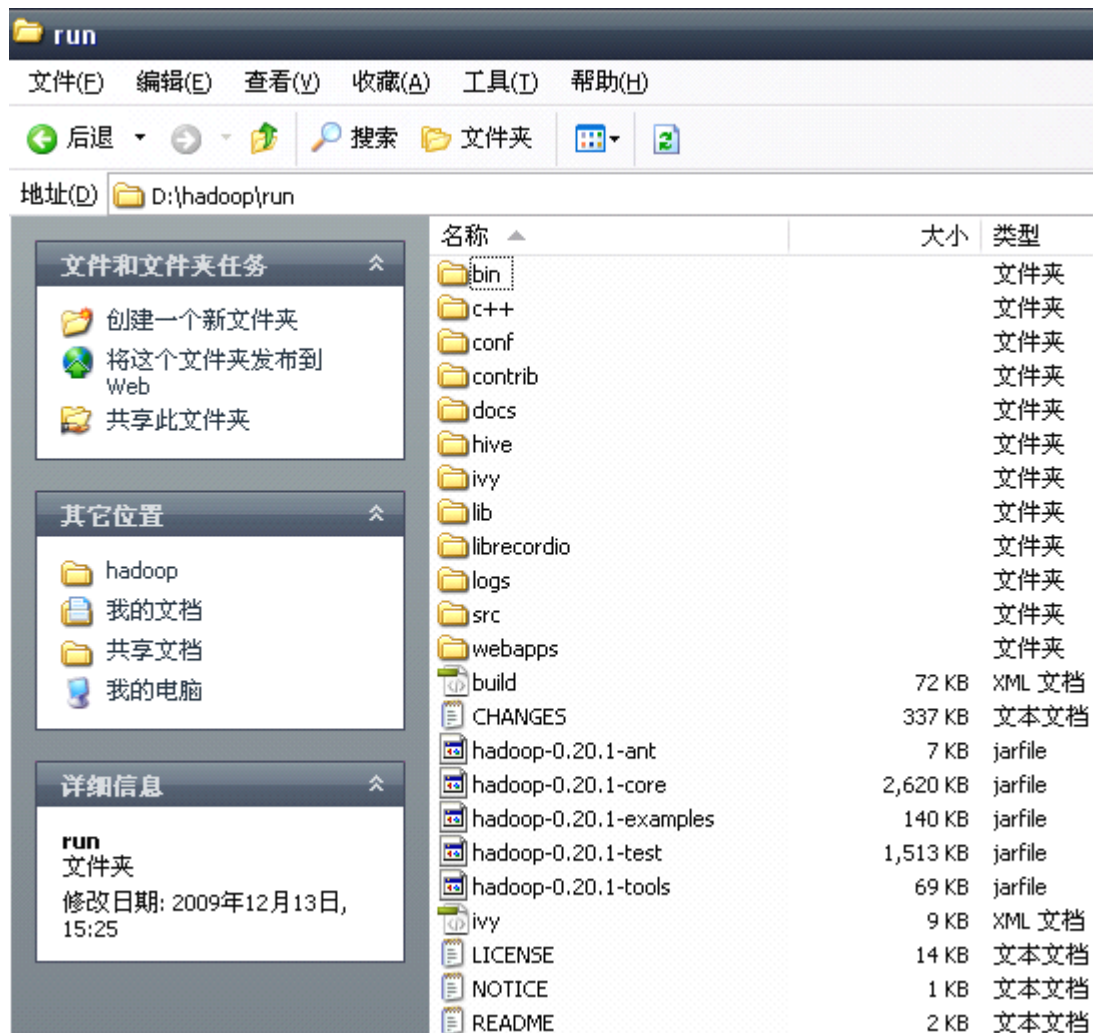
## 7. 下载 hadoop 安装包

hadoop 安装包下载地址:

<http://labs.xiaonei.com/apache-mirror/hadoop/core/hadoop-0.20.1/hadoop-0.20.1.tar.gz>

## 8. 安装 hadoop

将 hadoop 安装包 hadoop-0.20.1.tar.gz 解压到 **D:\hadoop\run** 目录(可以修改成其它目录)下, 如下图所示:



接下来, 需要修改 hadoop 的配置文件, 它们位于 conf 子目录下, 分别是 hadoop-env.sh、core-site.xml、hdfs-site.xml 和 mapred-site.xml 共四个文件。在 Cygwin 环境, masters 和 slaves 两个文件不需要修改。

- **修改 hadoop-env.sh**

只需要将 JAVA\_HOME 修改成 JDK 的安装目录即可, 请注意 JDK 必须是 1.6 或以上版本。

- **修改 core-site.xml**

为简化 core-site.xml 配置, 将 D:\hadoop\run\src\core 目录下的 core-default.xml 文件复制到 D:\hadoop\run\conf 目录下, 并将 core-default.xml 文件名改成 core-site.xml。修改 fs.default.name 的值, 如下所示:

```
<property>
  <name>fs.default.name</name>
  <value>hdfs://localhost:8888</value>
  <description>The name of the default file system
  scheme and authority determine the FileSystem im
  uri's scheme determines the config property (fs.
  the FileSystem implementation class. The uri's
  determine the host, port, etc. for a filesystem.
</property>
```

上图中的端口号 **8888**，可以改成其它未被占用的端口。

#### ● 修改 **hdfs-site.xml**

为简化 hdfs-site.xml 配置，将 D:\hadoop\run\src\hdfs 目录下的 hdfs-default.xml 文件复制到 D:\hadoop\run\conf 目录下，并将 hdfs-default.xml 文件名改成 hdfs-site.xml。不需要再做其它修改。

#### ● 修改 **mapred-site.xml**

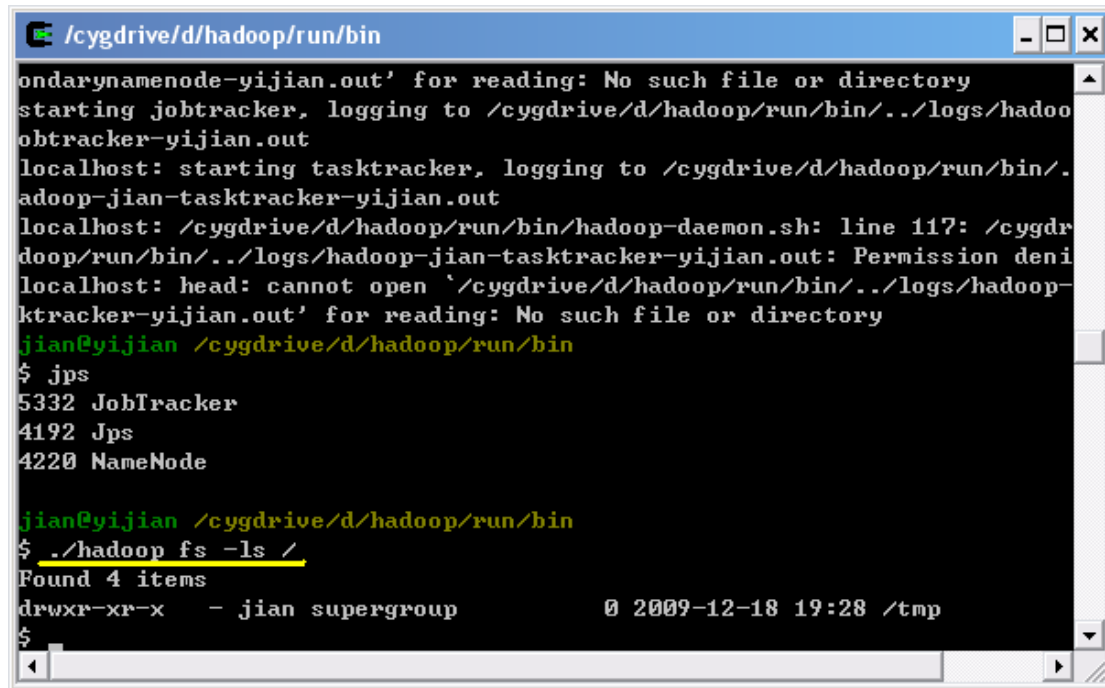
为简化 mapred-site.xml 配置，将 D:\hadoop\run\src\mapred 目录下的 mapred-default.xml 文件复制到 D:\hadoop\run\conf 目录下，并将 mapred-default.xml 文件名改成 mapred-site.xml。

```
<property>
  <name>mapred.job.tracker</name>
  <value>localhost:9999</value>
  <description>The host and port that the MapReduce job
  at. If "local", then jobs are run in-process as a si
  and reduce task.
</description>
</property>
```

上图中的端口号 **9999**，可以改成其它未被占用的端口。到这里，hadoop 宣告安装完毕，可以开始体验 hadoop 了！

## 9. 启动 hadoop

在 Cygwin 中，进入 hadoop 的 bin 目录，运行 ./start-all.sh 启动 hadoop，在启动成功之后，可以执行 ./hadoop fs -ls / 命令，查看 hadoop 的根目录，如下图所示：



```
/cygdrive/d/hadoop/run/bin
ondarynamenode-yijian.out' for reading: No such file or directory
starting jobtracker, logging to /cygdrive/d/hadoop/run/bin/../../logs/hadoo
obtracker-yijian.out
localhost: starting tasktracker, logging to /cygdrive/d/hadoop/run/bin/.
adoop-jian-tasktracker-yijian.out
localhost: /cygdrive/d/hadoop/run/bin/hadoop-daemon.sh: line 117: /cygdr
doop/run/bin/../../logs/hadoop-jian-tasktracker-yijian.out: Permission deni
localhost: head: cannot open '/cygdrive/d/hadoop/run/bin/../../logs/hadoop-
ktracker-yijian.out' for reading: No such file or directory
jian@yijian /cygdrive/d/hadoop/run/bin
$ jps
5332 JobTracker
4192 Jps
4220 NameNode

jian@yijian /cygdrive/d/hadoop/run/bin
$ ./hadoop fs -ls /
Found 4 items
drwxr-xr-x  - jian supergroup      0 2009-12-18 19:28 /tmp
$
```

如果运行 mapreduce，请参考其它文档，本教程的内容到此结束。