在 Windows 上安装 Hadoop 教程

一见 2010.1.6 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

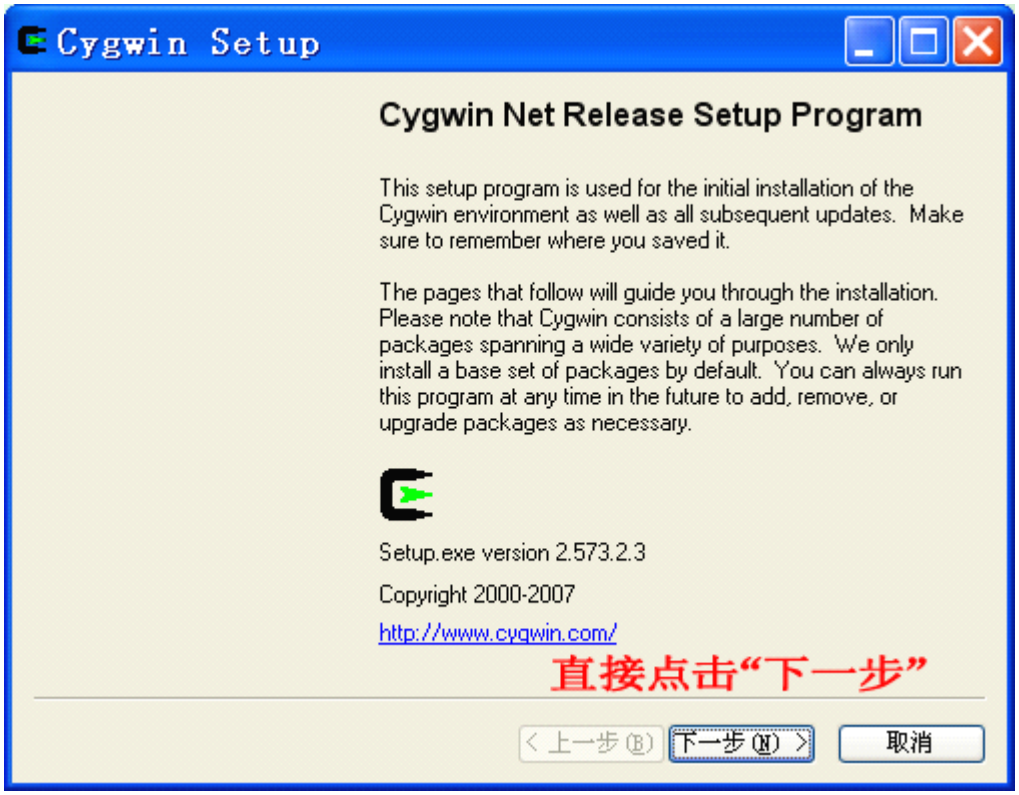
JDK 下载地址：http://java.sun.com/javase/downloads/index.jsp，下载 Java SE 即可。

2. 安装 Cygwin

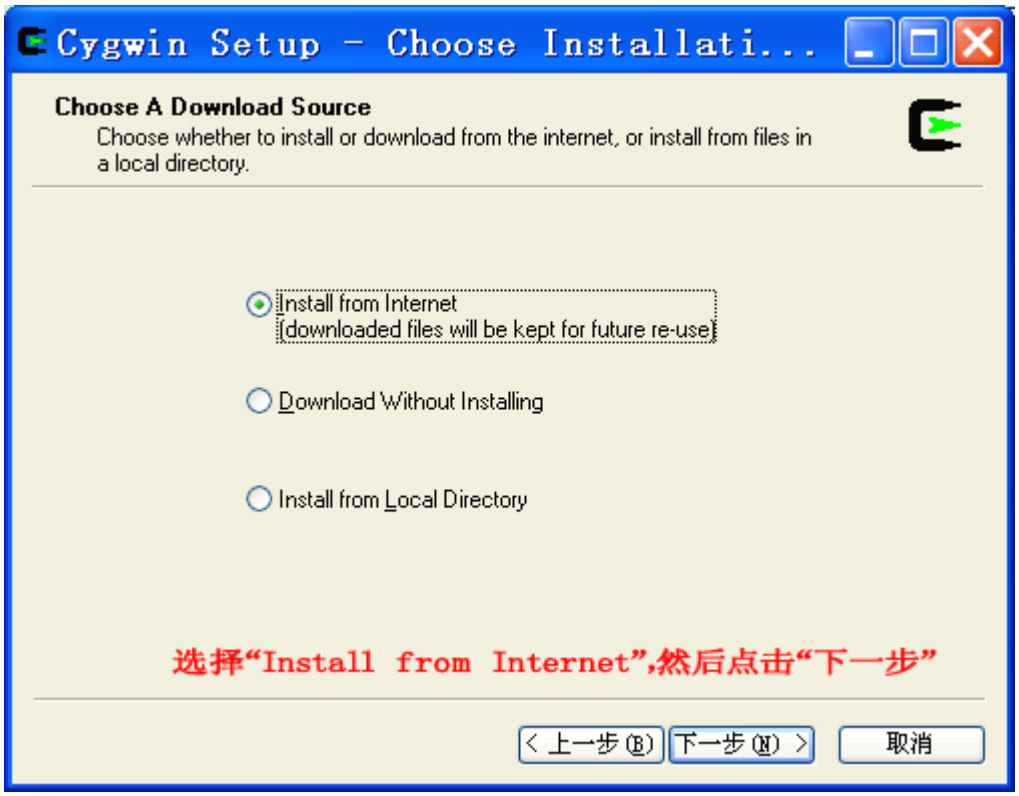
在安装 Cygwin 之前，得先下载 Cygwin 安装程序 setup.exe。

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。

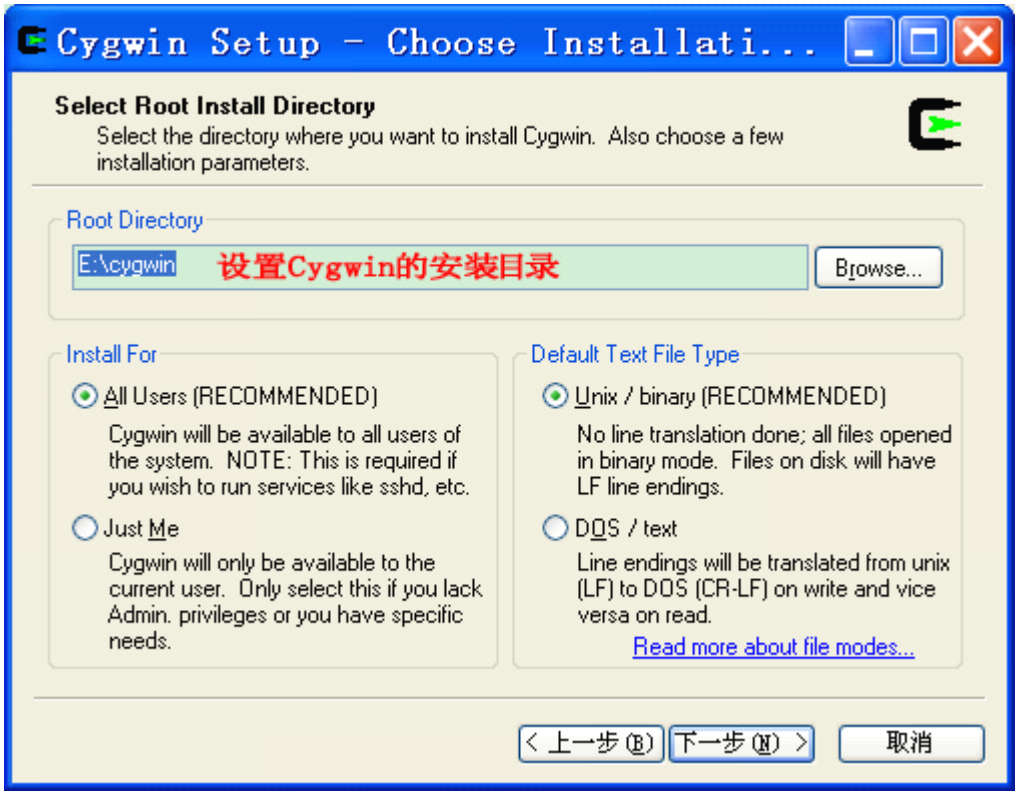
Cygwin 安装程序 setup.exe 的存放目录可随意无要求。当 setup.exe 下载成功后，运行 setup.exe，弹出如下图所示的对话框：



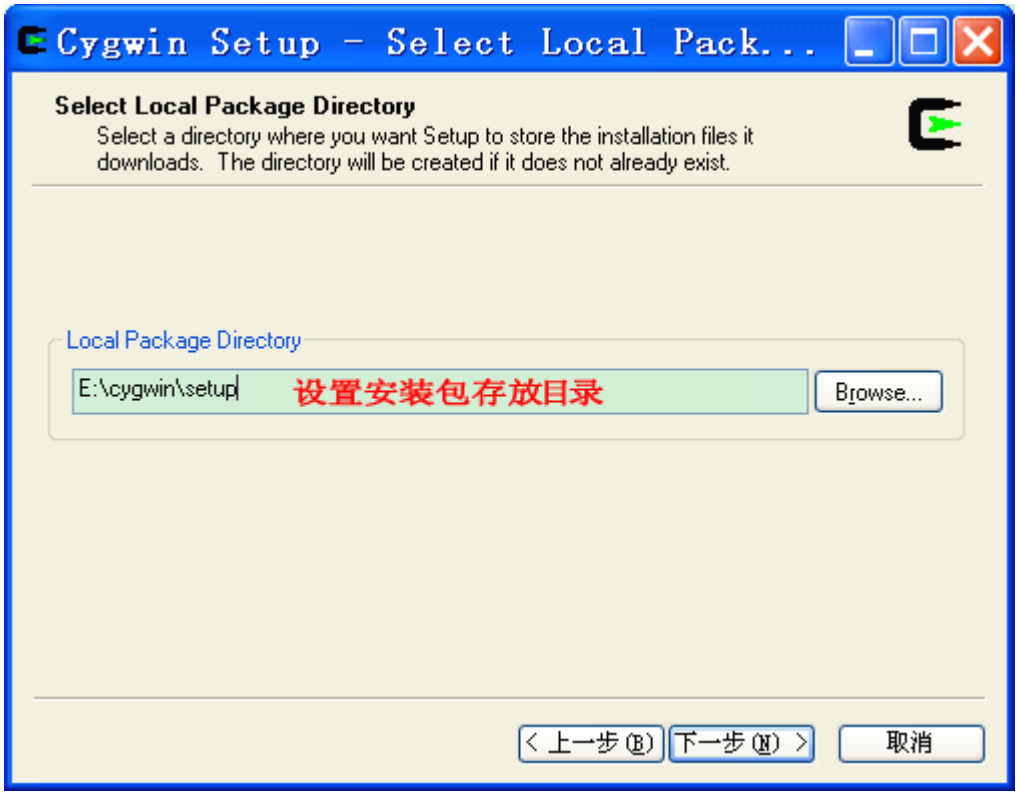
在上图所示的对话框中，直接点击“下一步”，进入如下图所示的对话框：



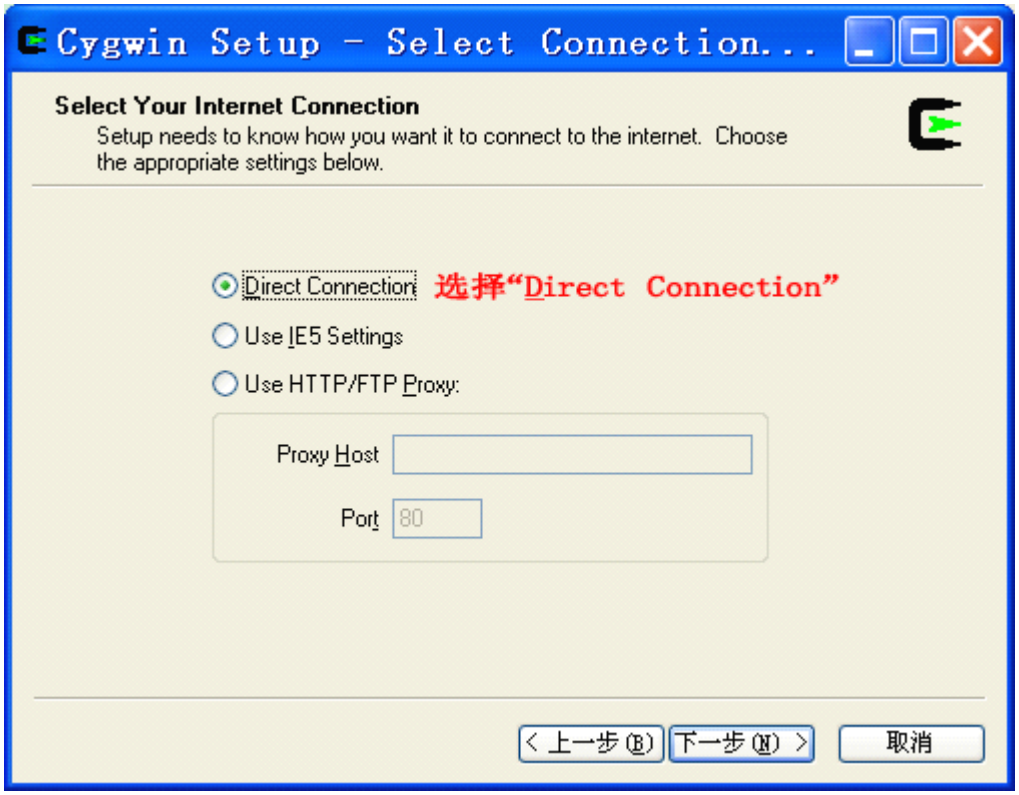
在上图所示的对话框中，选择“Install from Internet”，然后点击“下一步”，进入如下图所示对话框：



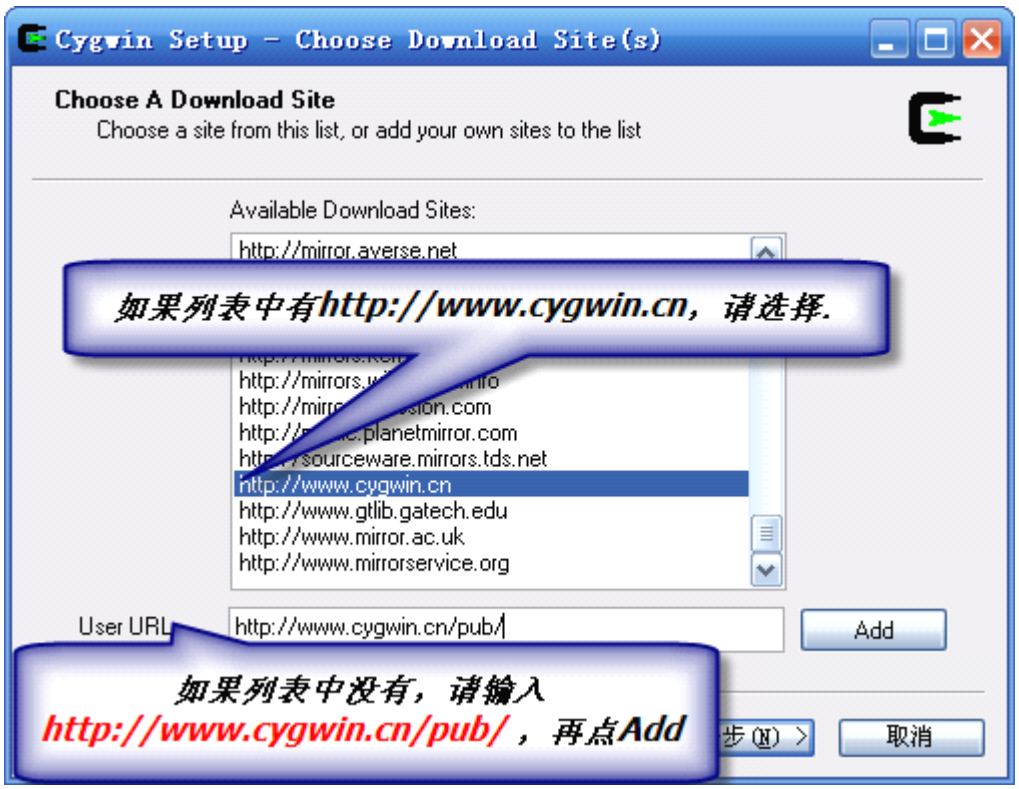
在上图所示的对话框中，设置 Cygwin 的安装目录，Install For 选择“All Users”，Default Text File Type 选择“Unix/binary”，然后点击“下一步”，进入如下图所示对话框：



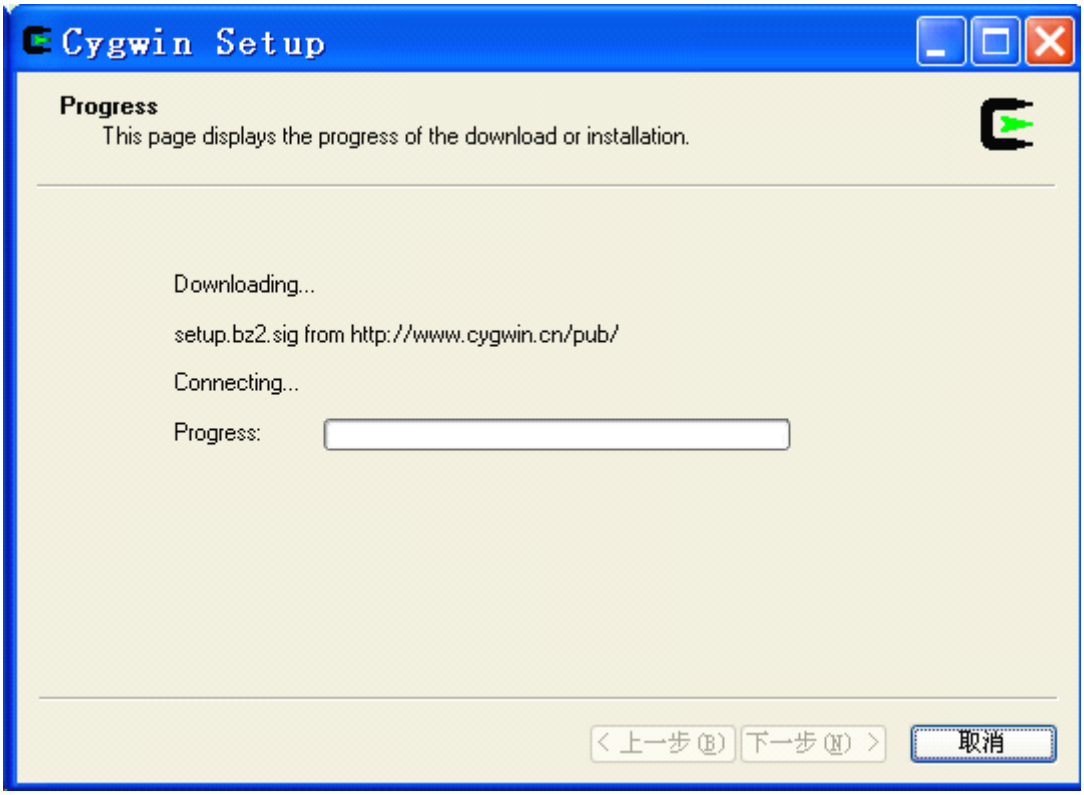
在上图所示的对话框中，设置 Cygwin 安装包存放目录，然后点击“下一步”，进入如下图所示对话框：



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：

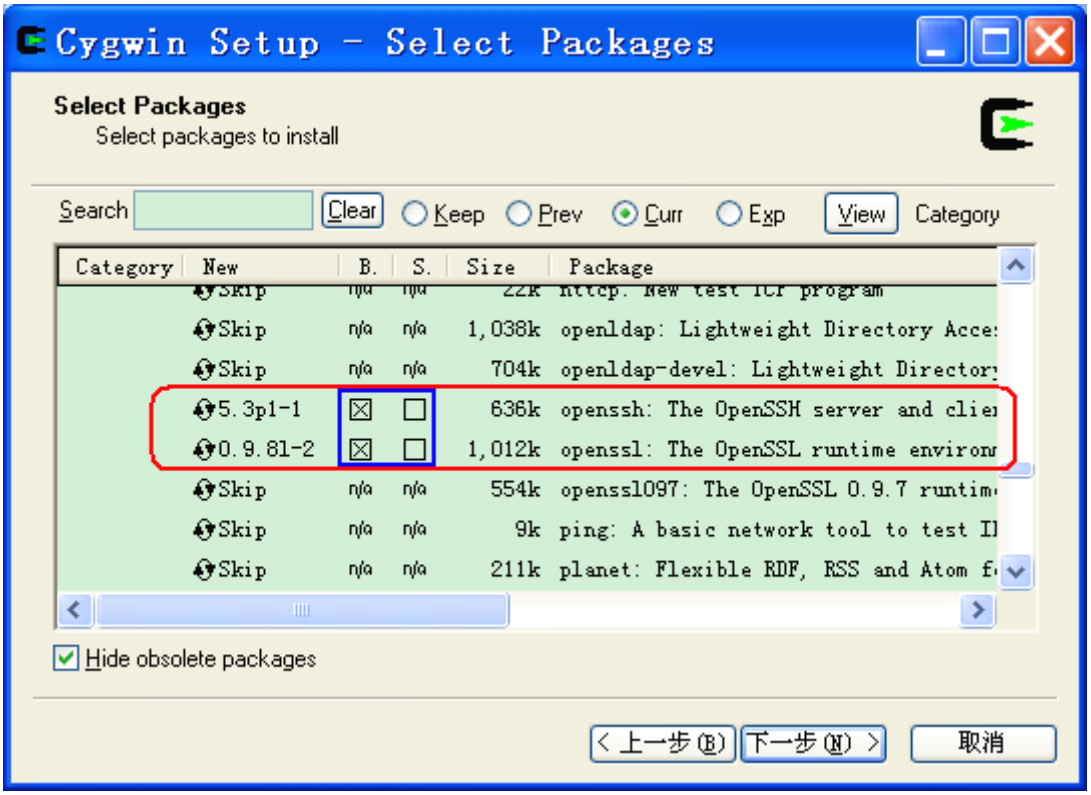


在上图所示的对话框中，点击“下一步”，将进入如下图所示的对话框：

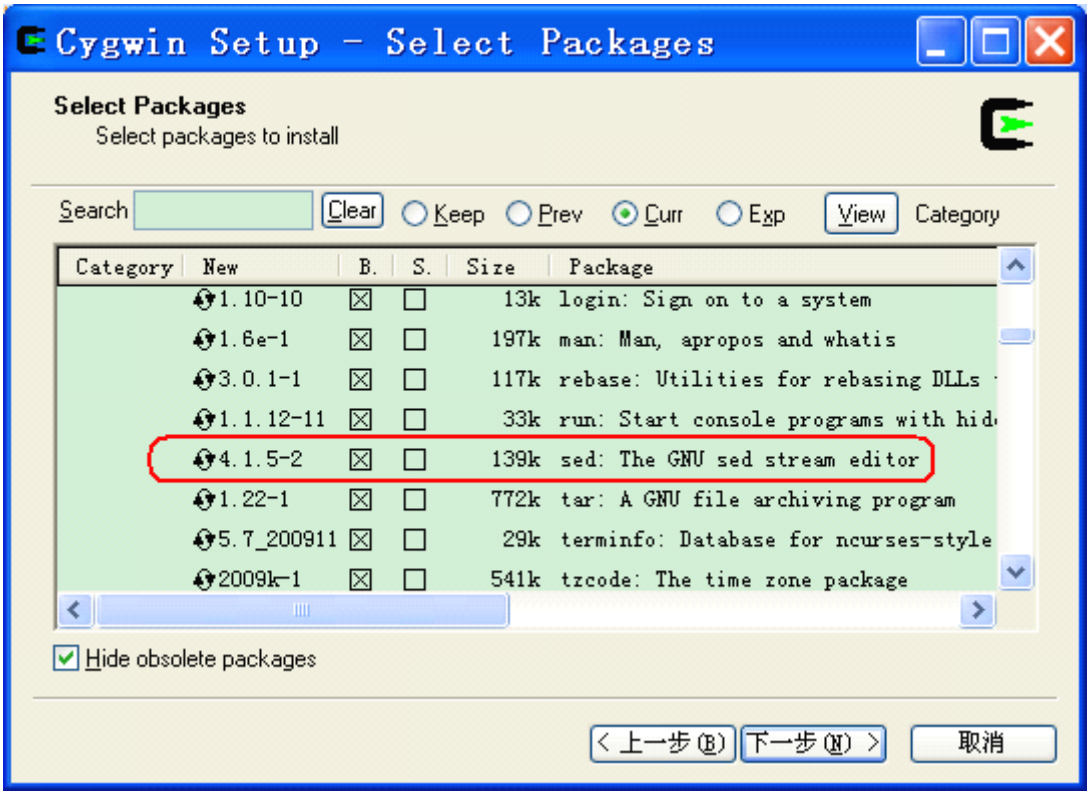


在上图所示的对话框过程中，可能会弹出如下图所示的“Setup Alert”对话框，直接点击“确定”即可。

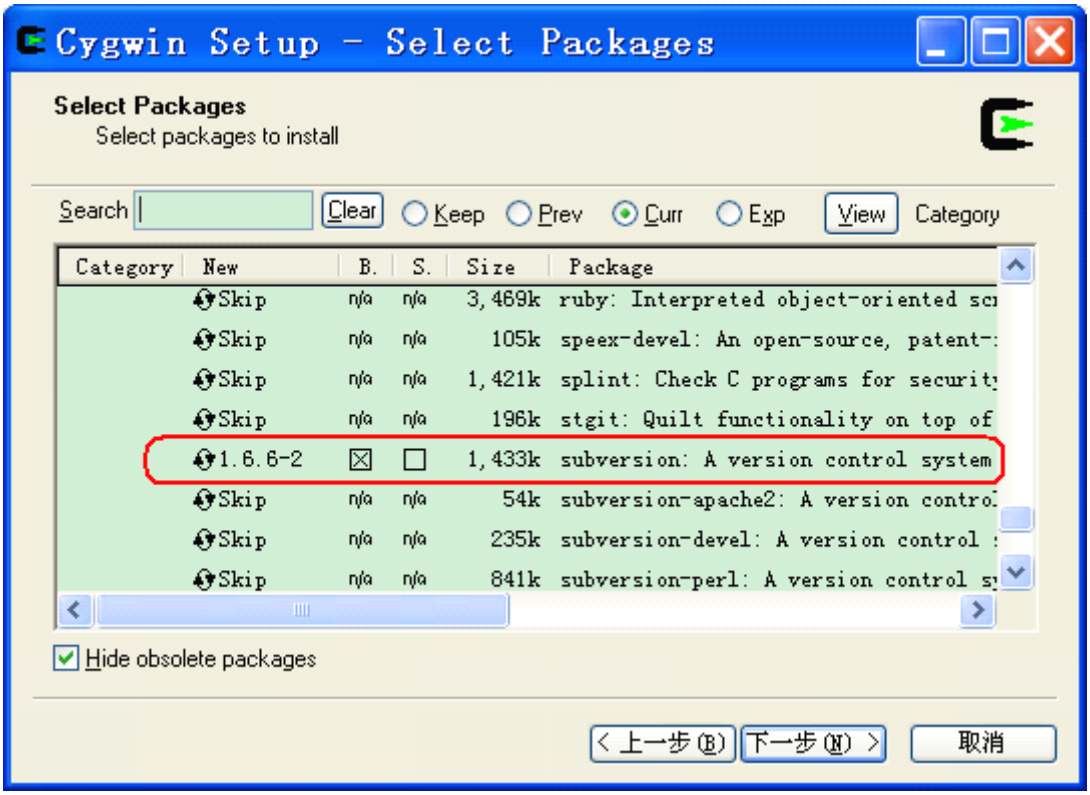
进入“Select Packages”对话框后，必须保证“Net Category”下的“OpenSSL”被安装，如下图所示：



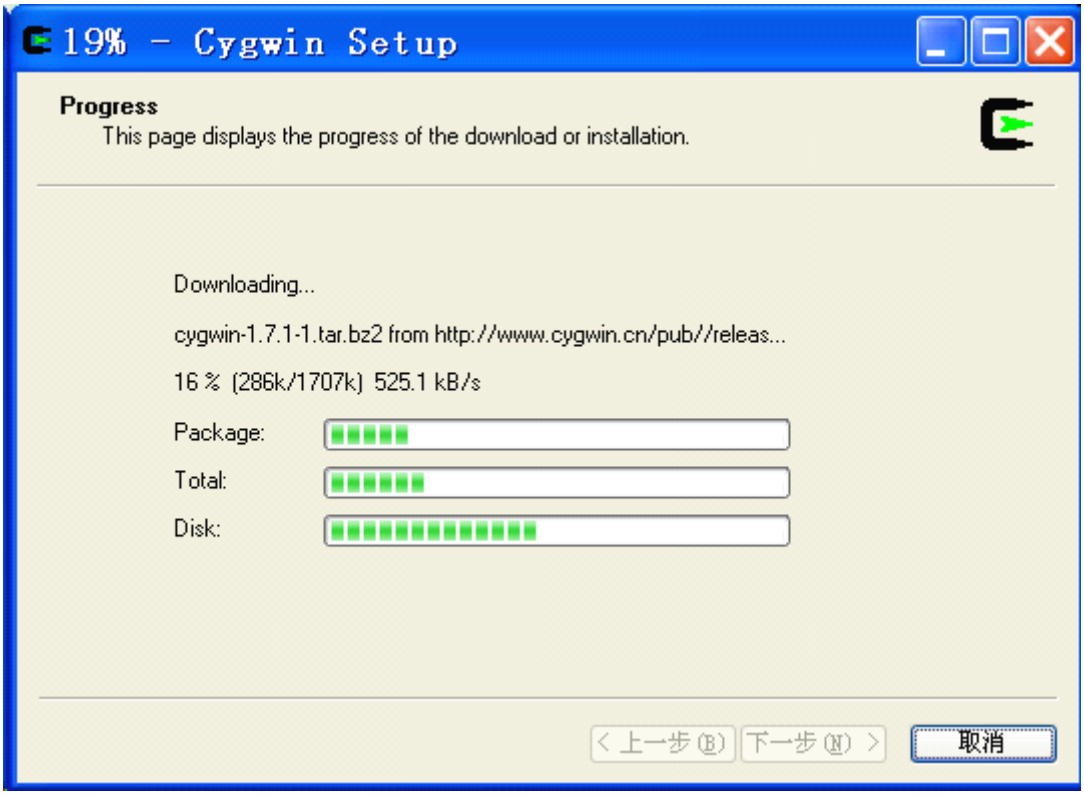
如果还打算在 eclipse 上编译 Hadoop，则还必须安装“Base Category”下的“sed”，如下图所示：



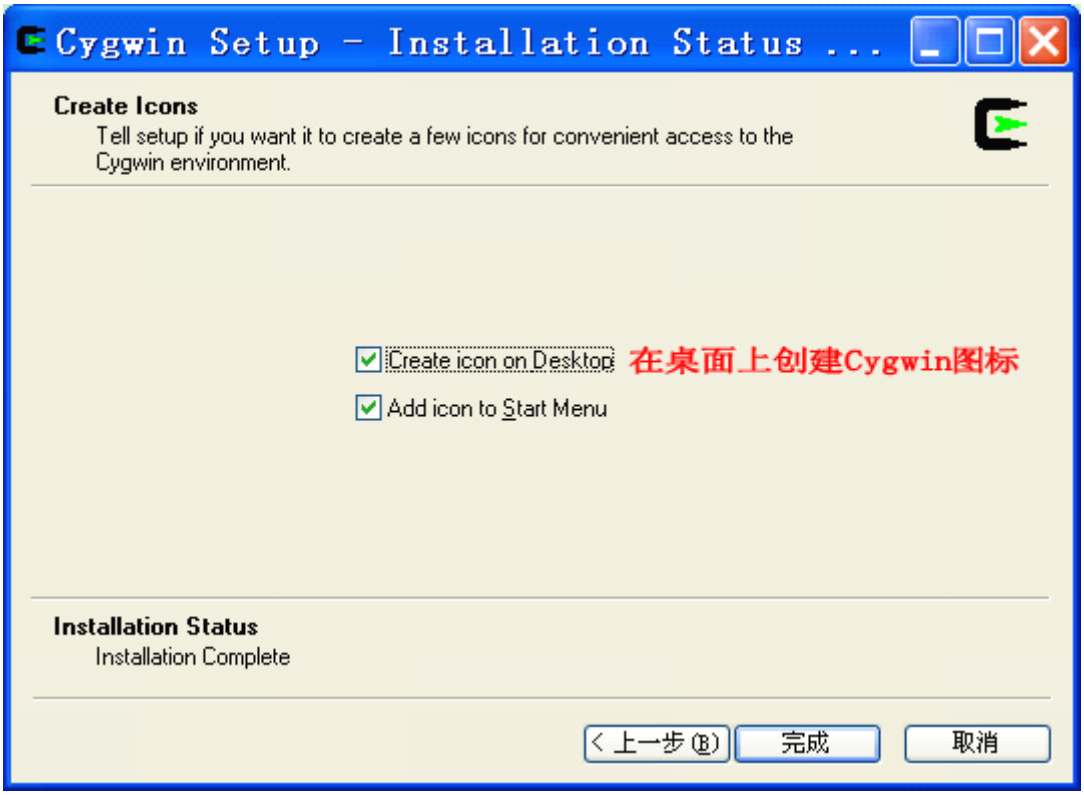
另外，还建议将“Editors Category”下的 vim 安装，以方便在 Cygwin 上直接修改配置文件；“Devel Category”下的 subversion 建议安装，如下图所示：



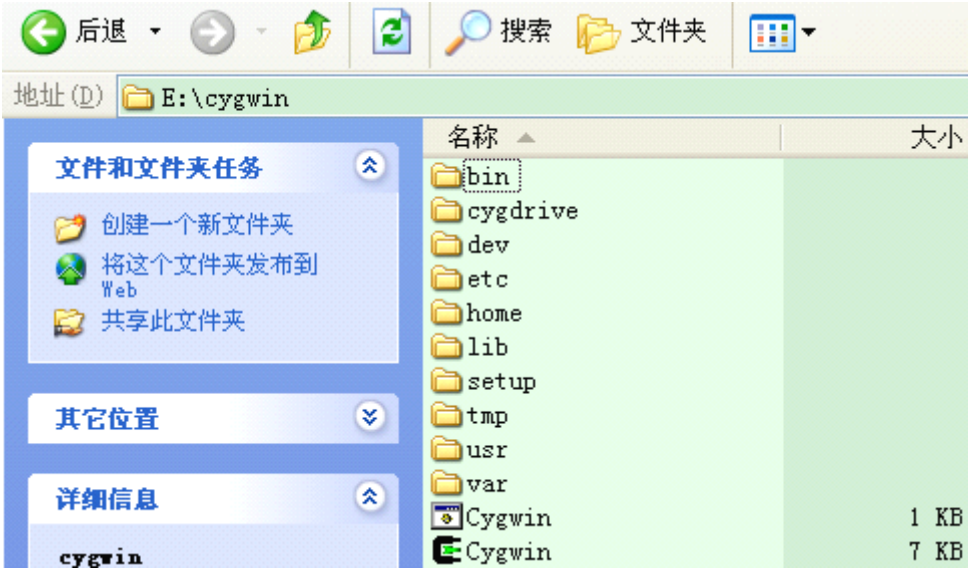
当完成上述操作后，点击“Select Packages”对话框中“下一步”，进入 Cygwin 安装包下载过程，如下图所示：



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：

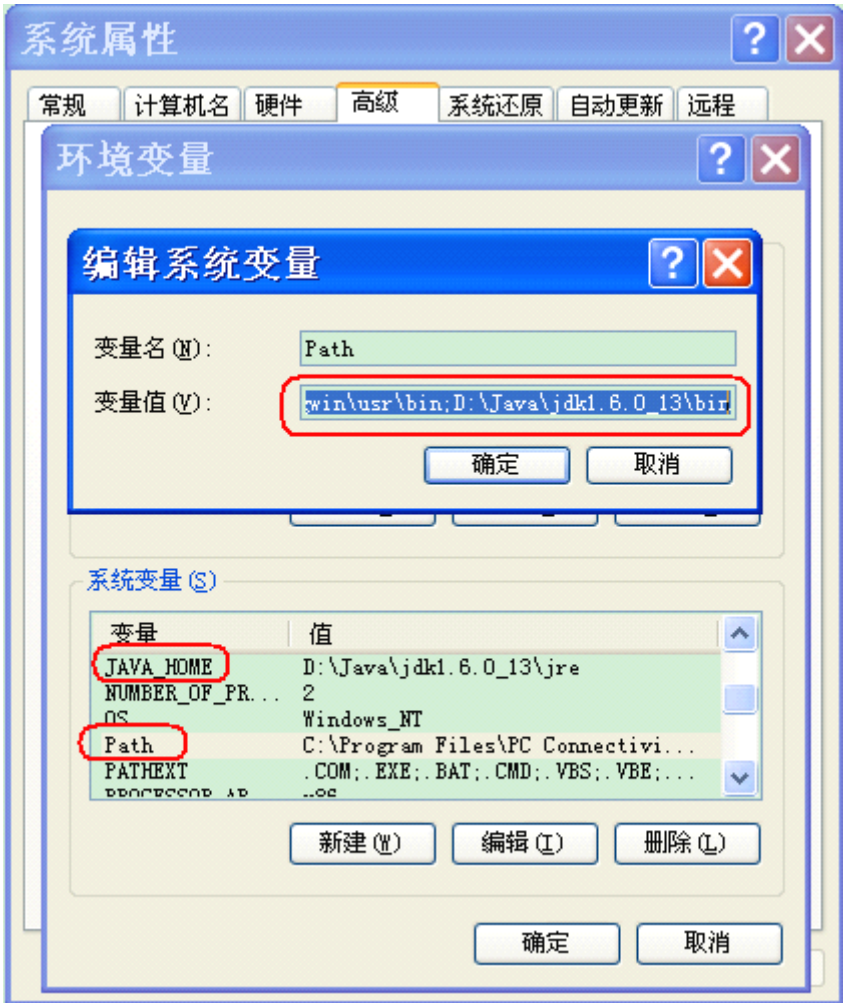


在上图所示的对话框中，选中“Create icon on Desktop”，以方便直接从桌面上启动 Cygwin，然后点击“完成”按钮。至此，Cgywin 已经安装完，安装目录下的内容如下图所示：



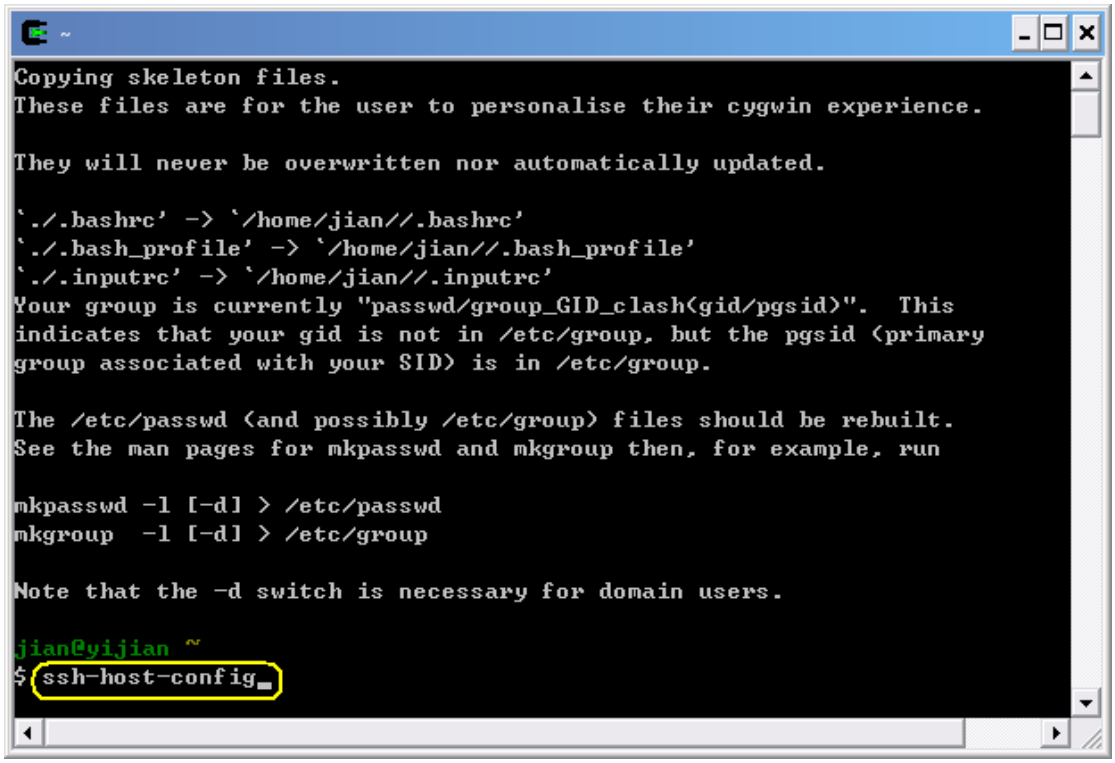
3. 配置环境变量

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：

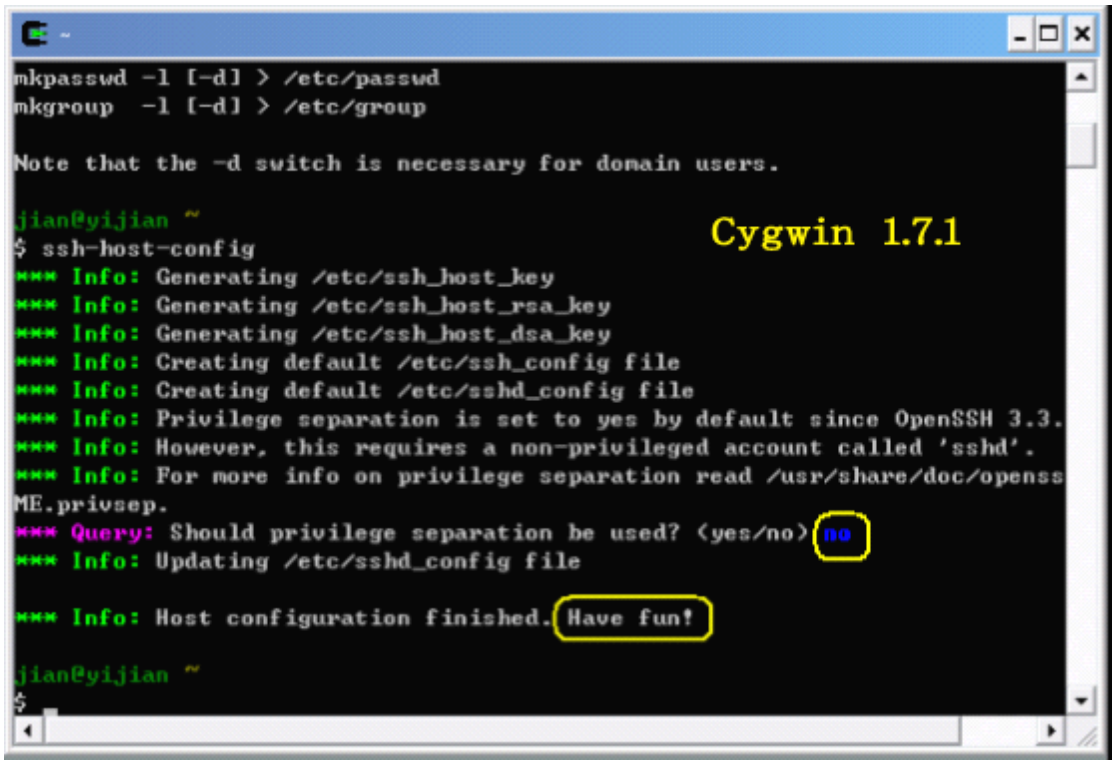


4. 安装 sshd 服务

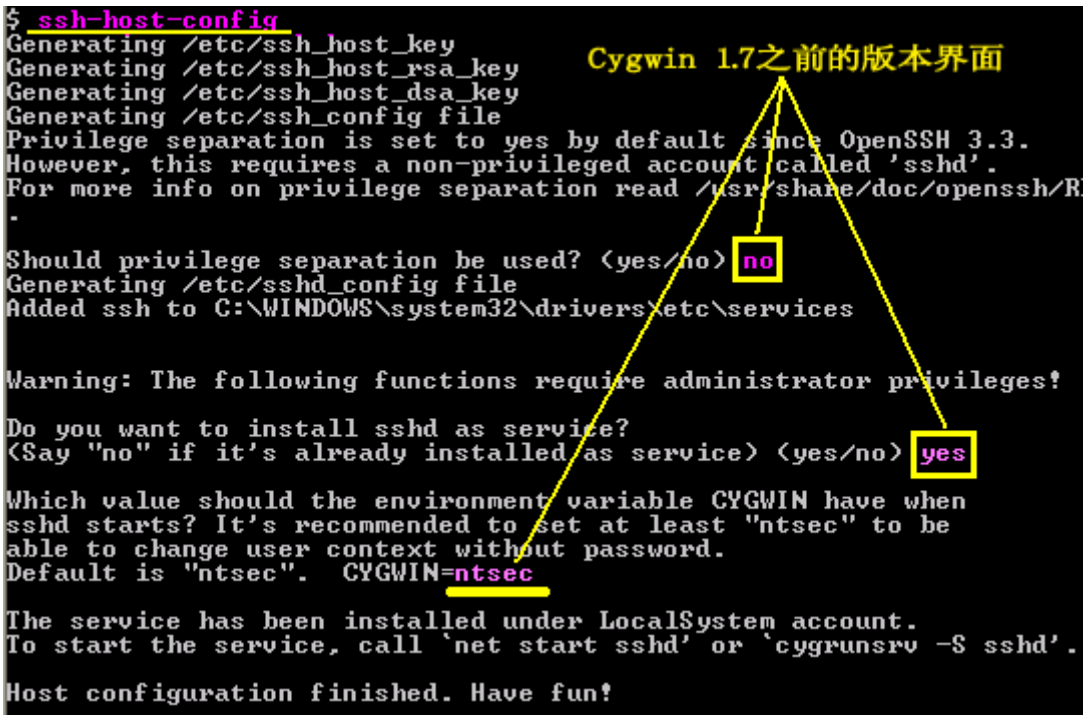
点击桌面上的 Cygwin 图标，启动 Cygwin，执行 **ssh-host-config** 命令，如下图所示：



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：



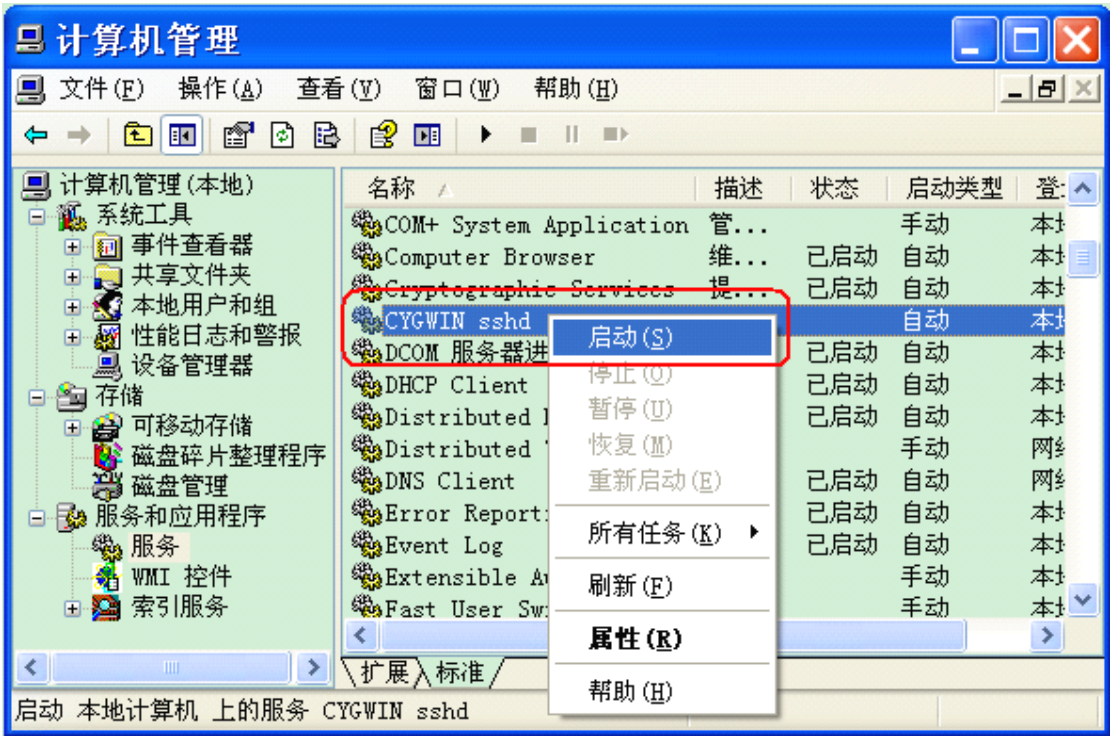
如果是 Cygwin 1.7 之前的版本，则 ssh-host-config 显示界面如下图所示：



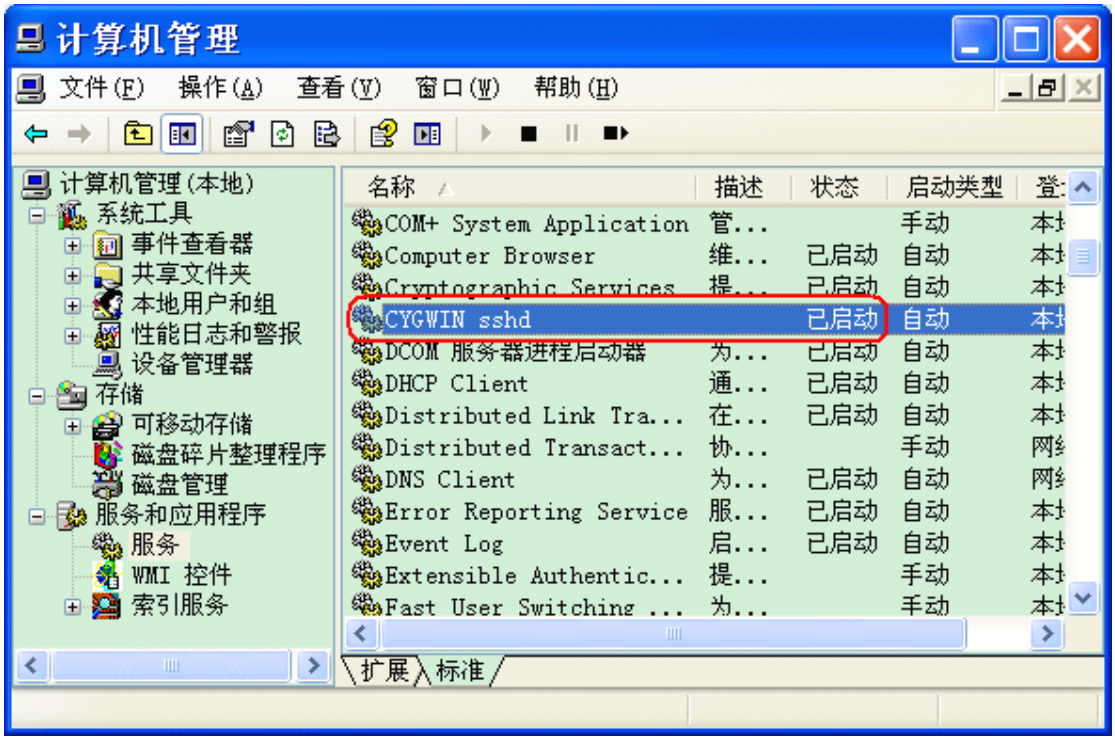
当看到“Have fun”时，一般表示 sshd 服务安装成功了，如上图所示。接下来，需要启动 sshd 服务。

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：



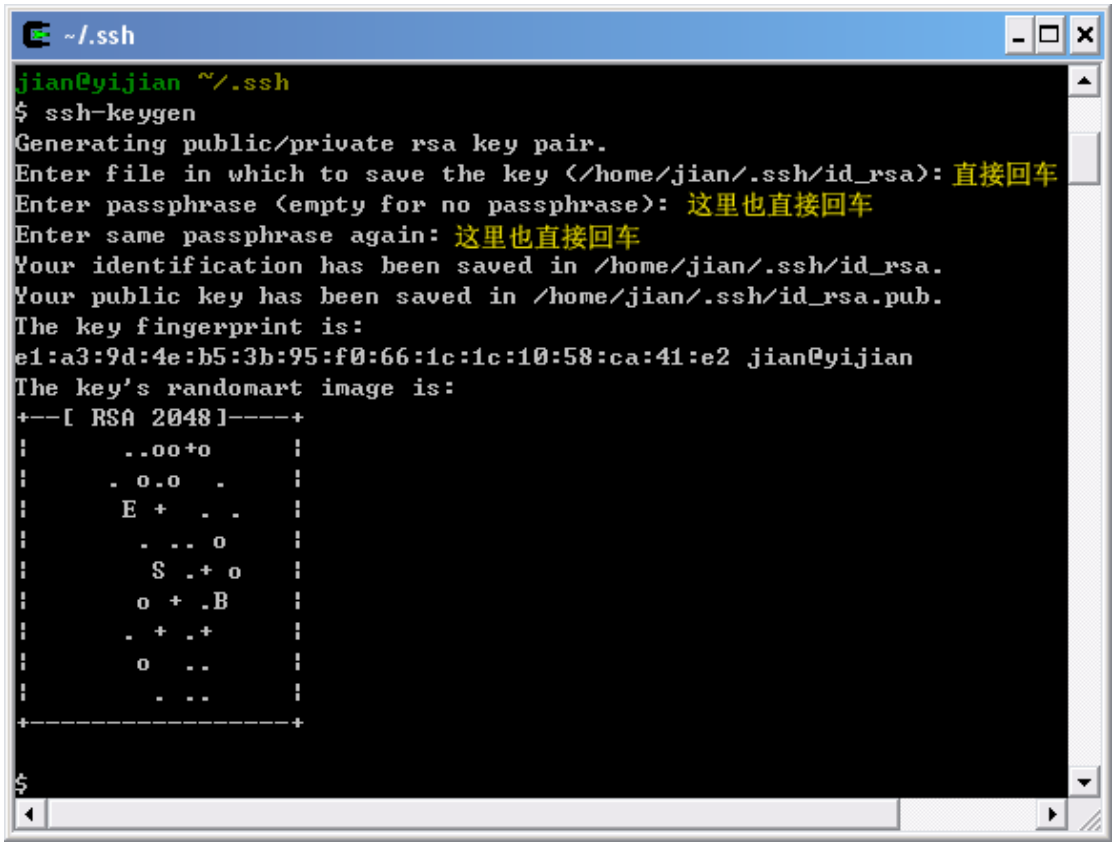
在上图所示的对话框中，选中“**CYGWIN sshd**”，弹出**右键**，并启动 CYGWIN sshd 服务，成功后，如下图所示：



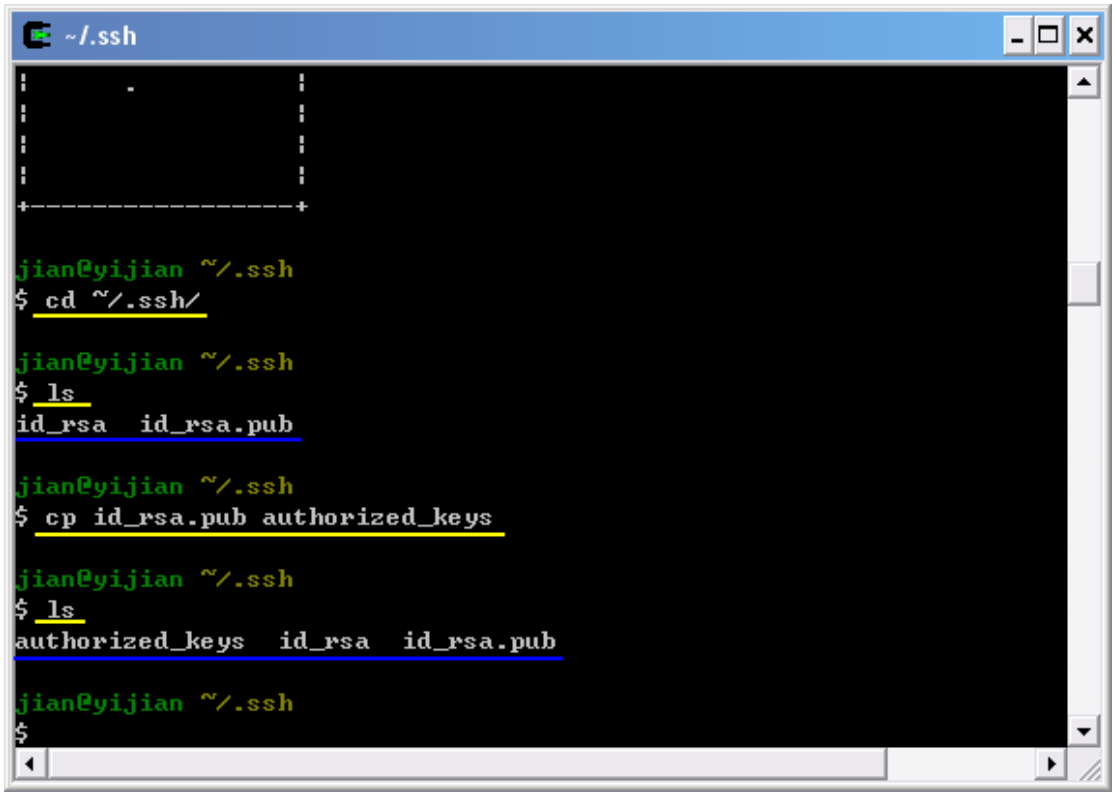
当 CYGWIN sshd 的状态为“**已启动**”后，接下来就是配置 ssh 登录。

6. 配置 ssh 登录

执行 **ssh-keygen** 命令生成密钥文件，如下图所示：



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：

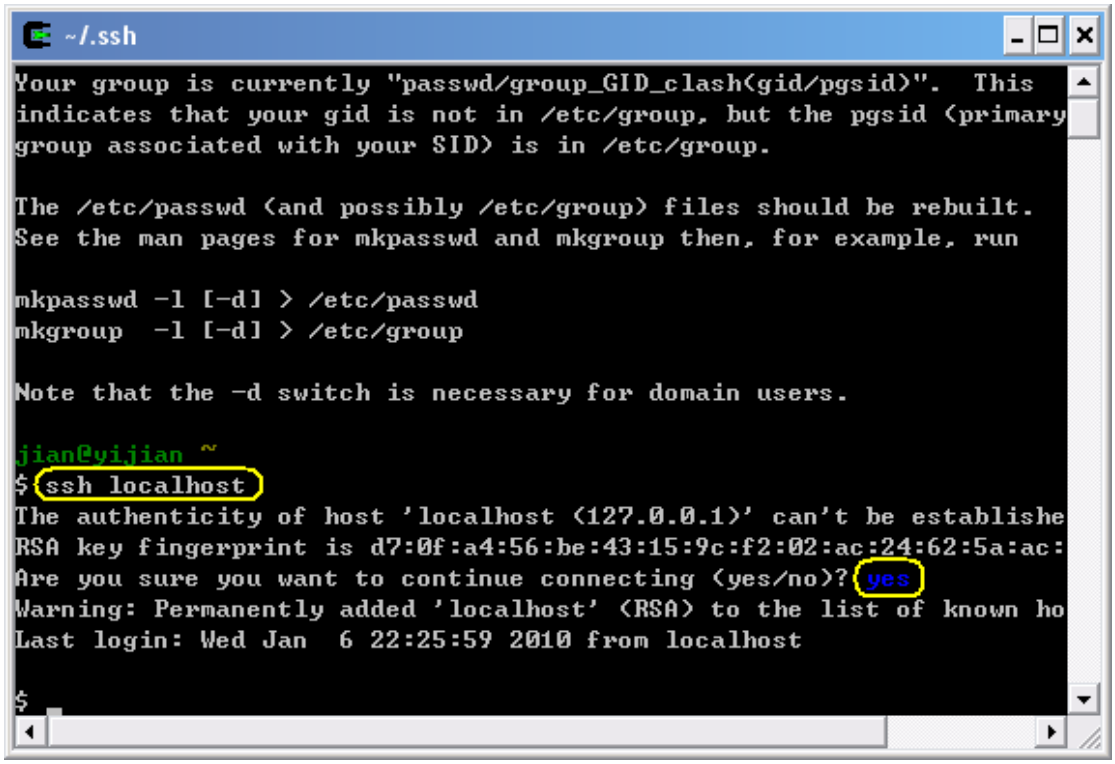


正如上图所示，只需要两步操作，即可生成 **authorized_keys** 文件：

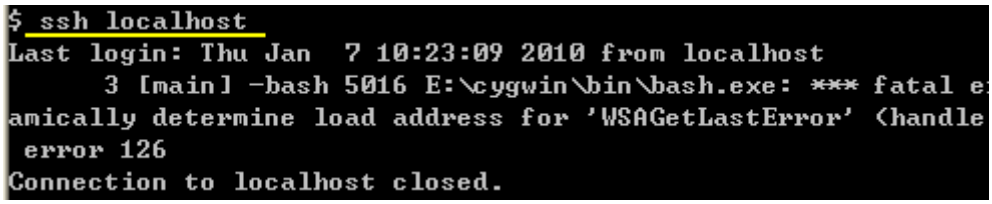
cd ~/..ssh/

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：



如果是 Windows 域用户，这步操作可能会遇到问题，错误信息如下：。



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：

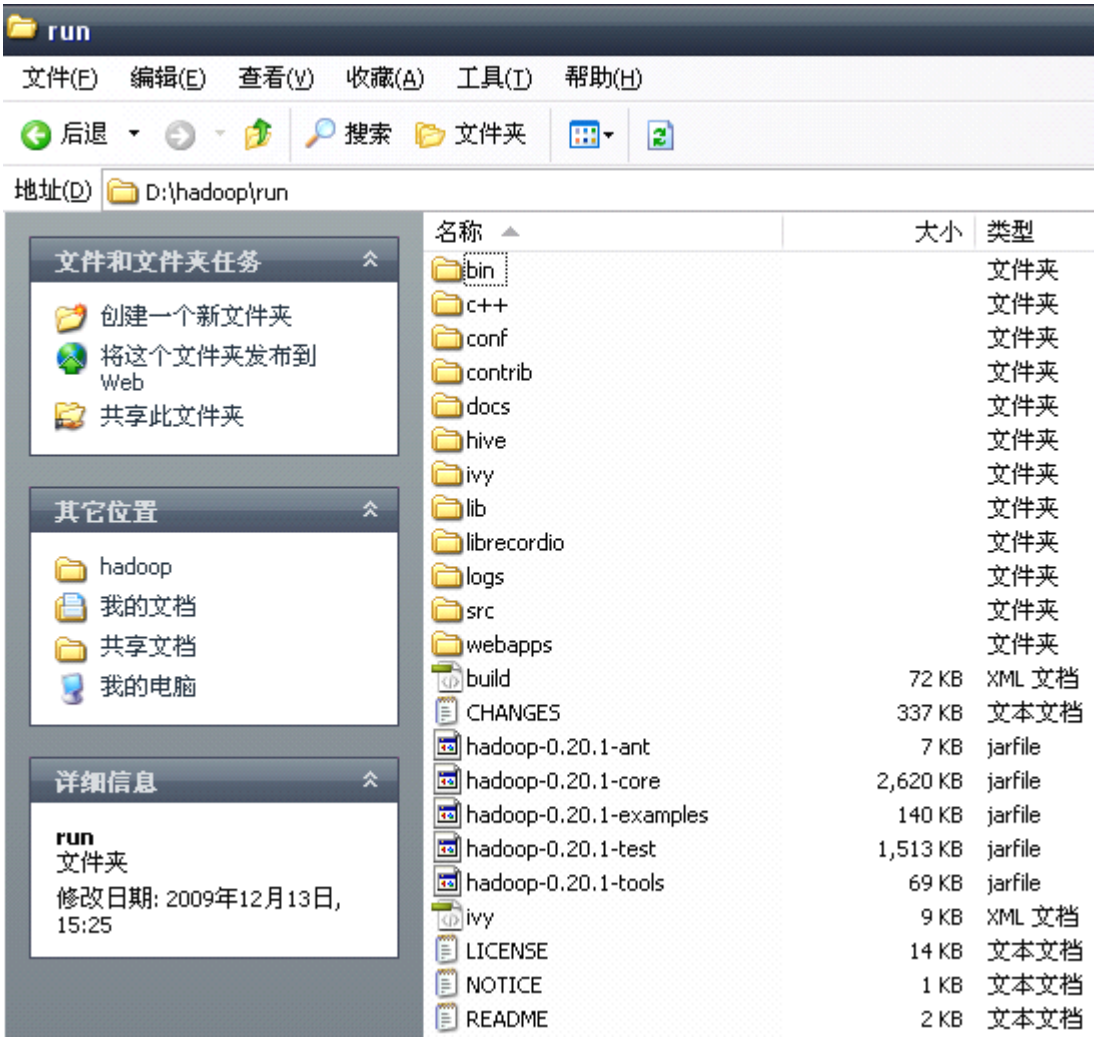
至此，配置 ssh 登录成功，下面就可以开始安装 hadoop 了。

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**

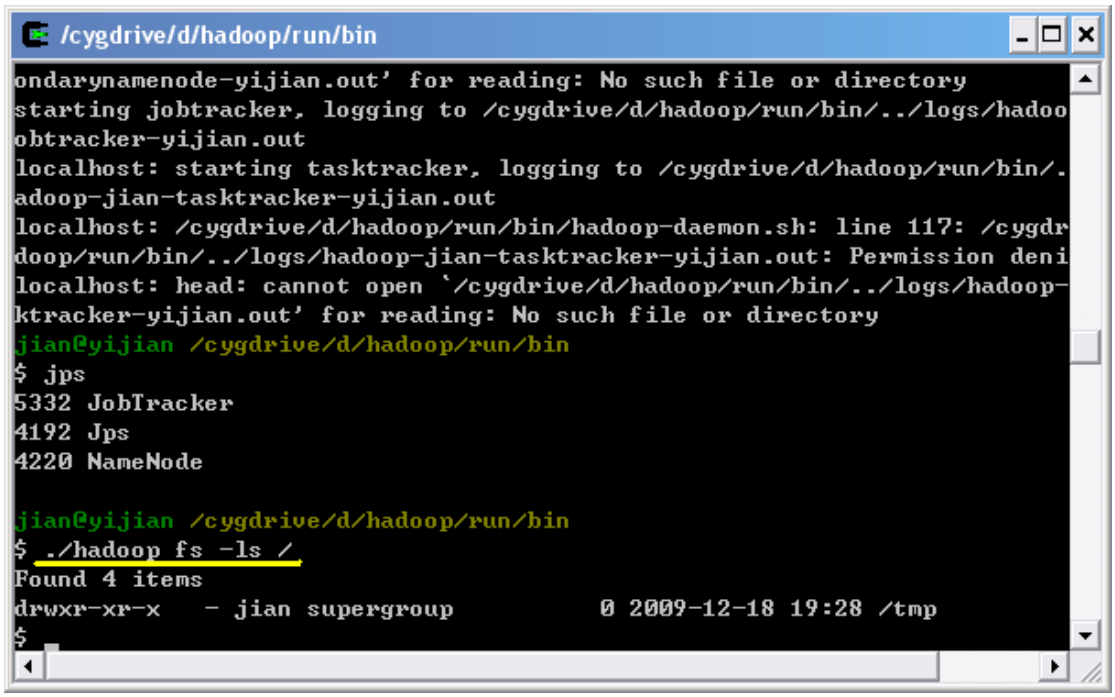
为简化 mapred-site.xml 配置，将 D:\hadoop\run\src\mapred 目录下的 mapred-default.xml 文件复制到 D:\hadoop\run\conf 目录下，并将 mapred-default.xml 文件名改成 mapred-site.xml。

```
<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：



如果运行 mapreduce，请参考其它文档，本教程的内容到此结束。