Hdoop 的荐部署环境是 linux，但是我们想要在 windows 体验一下还是可以的，尤其是对我这样不是经常使用 Linux 系统的人。Windows 下安装配置 Hadoop 的教程还是挺多的，这里可以下载。

下面我说说安装配置过程中遇到的一些问题吧，给自己以后提供借鉴。

我的操作系统是 Windows 7，使用的 Cygwin 版本是 1.7.7-1，hadoop 版本是 0.20.2。

1. SSH 服务的安装

在安装 SSH 服务的过程中，如果按照网络教程上的步骤，在 Win7 下会有问题，安装完成后测试 ssh localhost，遇到 connection closed by ::1 的问题，通过网上搜索，搜索到一片 E 文的解决方法，里面详细讲解了从卸载已经安装的服务到成功安装的各个步骤，比较详细，我也上传到了这里，其中的 vista-cygwin.txt。有些不爽的是经过这些步骤，会给系统增加一个用户。当然 SSH 服务安装成功的就好啦。

2. SSH 连接 REMOTE HOST IDENTIFICATION HAS CHANGED 解决方法

说到 SSH 连接，我想到了以前使用 SSH 连接时候遇到的一个问题，远程一台机器被我重装了，虽然 IP 地址和用户名都没有变，但是用其他客户端连接就出现上述提示，这里需要将~/.ssh/know-hosts 文件中和该机器相关的 rsa 信息删除，然后再添加一下。删除命令：ssh-keygen –R XXX.XXX.XXX.XXX。当然也可以手动去删除。

3. Java SDK 路径配置的问题——hadoop-env.sh 配置

在 Hadoop 的配置文件，hadoop-env.sh 中，需要添加 Java SDK 的路径。要知道，Cygwin 中并没有安装 Java，那么只能使用 Windows 下安装的 JDK。看了网上的文章，只是提到说 export 后面的路径需要用双引号括起来，因为一般 JDK 都是安装在“Program Files”下的。但是我这么添加以后，就是不行，报错“JAVA_HOME is not set”还是“…Program cannot found”来着，不太记得了，反正就是启动 Hadoop 的时候报错。经过研究后知道，首先 Cygwin 下不能使用 windows 的路径直接访问，需要是 Linux 风格的目录，如果是访问 C 盘下安装的 JDK 的话，不能使用“C:/Program Files/Java/jdk1.6.0_22”，而需要使用这样“/cygdrive/c/Program Files/Java/jdk1.6.0_22”；其次是 hadoop 好像不能识别带空格的目录路径，那怎么办呢，想到了符号链接，通过在 Cygwin 下创建一个符号链接，指向我的 JDK 安装目录，使用这样的命令：ln “C:/Program Files/Java/jdk1.6.0_22” /usr/local/jdk1.6.0_22. 这样就在/usr/local/下创建了一个 Windows 下 JDK 目录的符号链接，如果使用：cd /usr/local/jdk1.6.0_22, 然后 ls 下，就会发现已经在 jdk 的安装目录下了。然后在 hadoop-env.sh 中 export JAVA_HOME=/usr/local/jdk1.6.0_22，这样就 OK 了。

4. hadoop-env.sh:line 2:$'/r':command not found 错误

如果将 hadoop-env.sh 用 Windows 下的编辑软件编辑以后，再使用的话可能出现上述错误，使用 dos2unix 工具修正下，在 conf/目录下执行“dos2unix hadoop-env.sh”就 OK 了。

5. Hadoop 启动的问题——process information unavailable

好不容易配置好了 hadoop，很兴奋，于是在赶紧启动 hadoop，”start-all.sh”，启动好像没有什么问题。使用 jps 列举下本机的 java 进程，发现怎么有几个进程显示“process information unavailable”，这是怎么回事？突然想起为什么要安装 SSH 呢，于是 SSH 连接到本机，然后再启动 hadoop，在 jps，OK 这下所有进程信息都列举出来了。具体是什么原因，我现在还不太清楚。

6. Hadoop 启动的问题——Server at localhost/127.0.0.1:8888 not available yet

经过上面的步骤，好像 Hadoop 运行没有问题了，TaskTracker、JobTrack、DataNode、NameNode 都启动了，好像没有问题了，但是执行“./hadoop fs –ls /”命令是，会出现“Retrying connect to server: localhost/127.0.0.1:8888. Already tried X time(s). … Server at localhost/127.0.0.1:8888 not available yet”，这又是为什么呢？查找后发现原来是 HDFS 在第一次用的时候需要初始化，安装完后就直接启动了，使用命令：./hadoop namenode –format, 进行文件系统的格式化，然后再“./start-all.sh”，再执行“./hadoop fs –ls /”，看到“Found 1 items
drwxr-xr-x - huangzhtao-pc/cyg_server supergroup 0 2010-11-27 17:26 /tmp”，总算成功了。

7. 浏览器访问的说明

浏览器下查看 Hadoop 系统情况的地址。

http://127.0.0.1:50070/ HDFS 情况

http://127.0.0.1:50060/ Task Tracker 情况

http://127.0.0.1:50030/ Job Tracker-Map/Reduce Administration

配置 hadoop 真是一个充满挑战的过程，遇到了很多问题，记录下来，供以后参考。