**1.软件包准备**

Hadoop-0.20.203

jdk-1.7 或者1.6 (jdk-1.8可能造成容器类无法识别)

eclipse-jee-indigo-SR1-linux (32 位) (eclipse-luna版本可能也会出现问题)

**2.** **安装 Hadoop-Eclipse-Plugin**

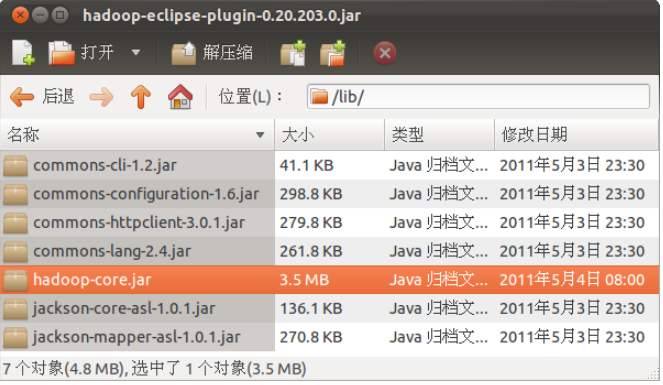
**将 Hadoop/contrib/eclipse-plugin 中的 hadoop-eclipse- plugin-0.20.203.0.jar复制到 Eclipse 安装目录的 plugin 文件夹中，运行 eclipse -clean 重启 Eclipse。**

**以下为引用可以自行尝试**

“但是在我实践尝试中，发现hadoop-0.20.203.0版本的该包如果直接复制到eclipse的插件目录中，在连接DFS时会出现错误，提示信息为： "error: failure to login",弹出的错误提示框内容为"An internal error occurred during: "Connecting to DFS hadoop". org/apache/commons/configuration/Configuration". 经过察看Eclipse的log，发现是缺少jar包导致的。进一步查找资料后，发现直接复制hadoop-eclipse-plugin-0.20.203.0.jar，该包中lib目录下缺少了jar包。

经过网上资料搜集，此处给出正确的安装方法：

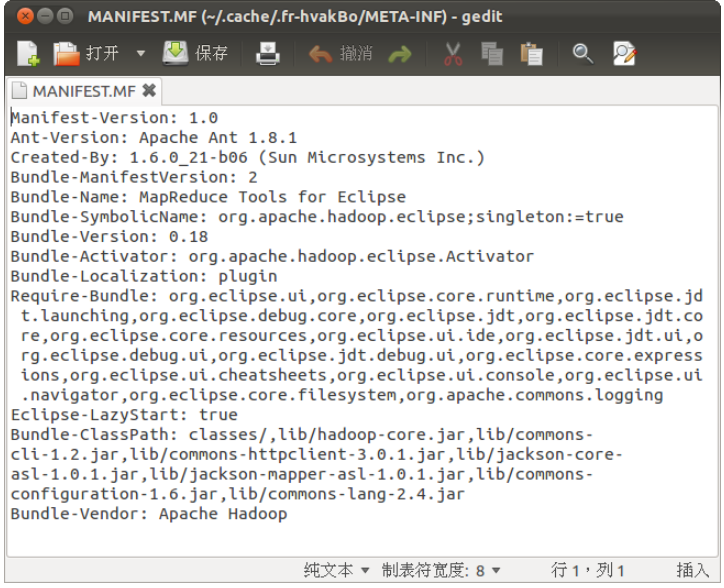
首先要对hadoop-eclipse-plugin-0.20.203.0.jar进行修改。用归档管理器打开该包，发现只有commons-cli-1.2.jar 和hadoop-core.jar两个包。将HADOOP\_HOME/lib目录下的 commons-configuration-1.6.jar , commons-httpclient-3.0.1.jar , commons-lang-2.4.jar , jackson-core-asl-1.0.1.jar 和 jackson-mapper-asl-1.0.1.jar 等5个包复制到hadoop-eclipse-plugin-0.20.203.0.jar的lib目录下，如下图：



然后，修改该包META-INF目录下的MANIFEST.MF，将classpath修改为一下内容：

Bundle-ClassPath: classes/,lib/hadoop-core.jar,lib/commons-cli-1.2.jar,lib/commons-httpclient-3.0.1.jar,lib/jackson-core-asl-1.0.1.jar,lib/jackson-mapper-asl-1.0.1.jar,lib/commons-configuration-1.6.jar,lib/commons-lang-2.4.jar

如下图：



这样就完成了对hadoop-eclipse-plugin-0.20.203.0.jar的修改。

最后，将hadoop-eclipse-plugin-0.20.203.0.jar复制到Eclipse的plugins目录下”

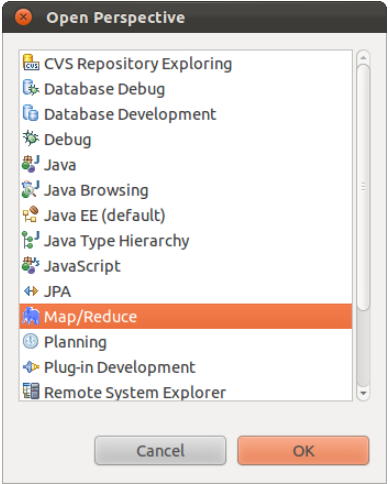
**以上为引用内容，经过尝试并没有有效编译生成可用jar包，于是在网上下载了一个编译通过的**hadoop-eclipse-plugin-0.20.203.0.jar**并复制到Eclipse的plugins目录下。**

**可以自行尝试生成jar包**

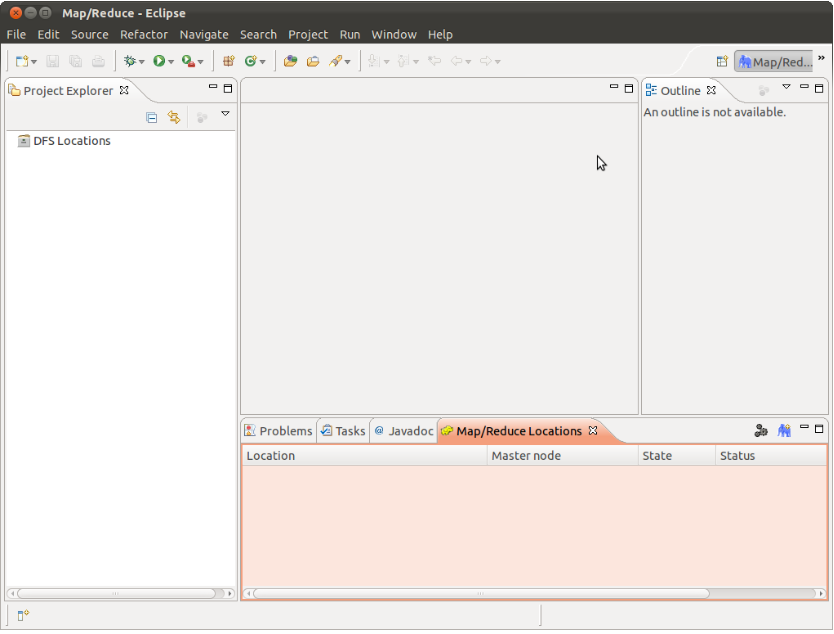
**3.** **在Eclipse中对该插件进行配置。**

**首先，打开Eclipse，设置好workspace。**

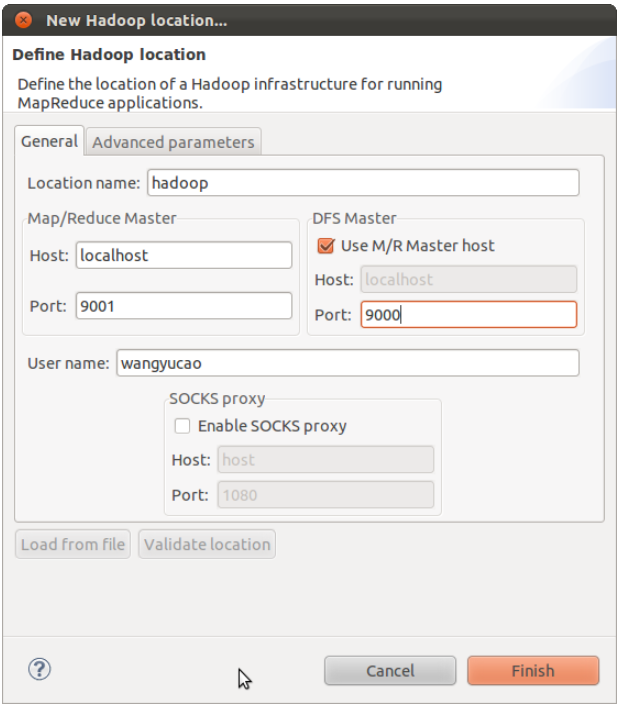
**选择window -> open perspective -> Other... ， 选择有大象图标的 Map/Reduce.**



**此时，就打开了Map/Reduce的开发环境。可以看到，右下角多了一个Map/Reduce Locations的框。如下图：**



**然后，设置Hadoop的环境参数。选择Map/Reduce Locations 标签，点击该标签最右边的大象图标，即那个齿轮状图标右侧的大象图标，打开参数设置页面，参数填写如下图：**



**Location Name ： 此处为参数设置名称，可以任意填写**

**Map/Reduce Master (此处为Hadoop集群的Map/Reduce地址，应该和mapred-site.xml中的mapred.job.tracker设置相同)**

**Host: localhost**

**port: 9001**

**DFS Master  (此处为Hadoop的master服务器地址，应该和core-site.xml中的 fs.default.name 设置相同)**

**Host: localhost**

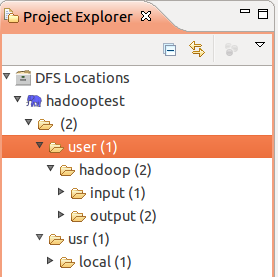
**Port: 9000**

**在advanced parameters中 参数**

**Hadoop.tmp.dir 也要填入core-site.xml的设置 （如果设置的话）**

**设置完成后，点击Finish就应用了该设置。**

**此时，在最左边的Project Explorer中就能看到DFS的目录，如下图所示。**

****

**4.** **建立文档并上传到HDFS 作为例子wordcount的input**

在Hadoop的dfs中新建文件夹：

hadoop dfs -mkdir /user/hadoop/input

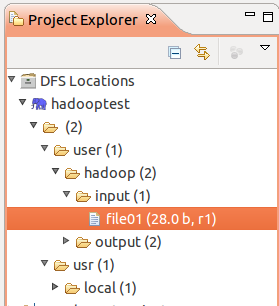
生成文件，和把文件放进dfs中

echo "Hello World Bye World" > file0 (自己建立个文本命名file0把Hello World Bye World 写进去，这里注意文本最后建在/usr文件夹下 否则可能存在上传时候的权限问题)

hadoop dfs -put file0 /user/hadoop/input

**可以登录**

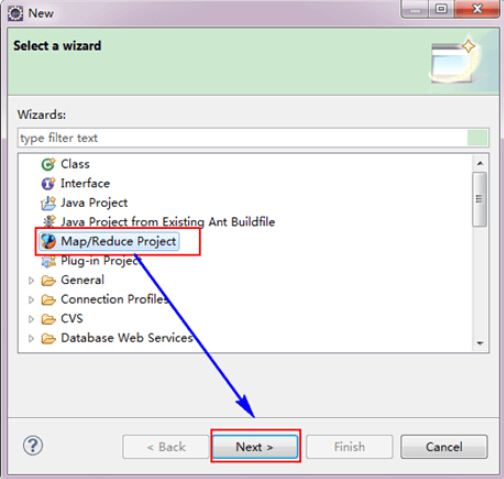
[**http://192.168.86.129:50070**](http://192.168.86.129:50070) **进行查看**

****

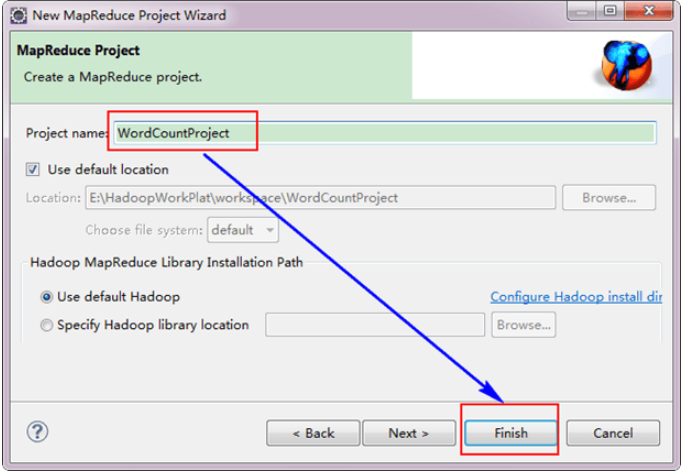
**注意此时是没有output文件夹的 output文件夹存放的是程序执行完的结果 是自动生成的（在程序中指定位置）**

**5. 在Eclipse中创建MapReduce项目**

点击 File 菜单，选择 Map/Reduce Project。

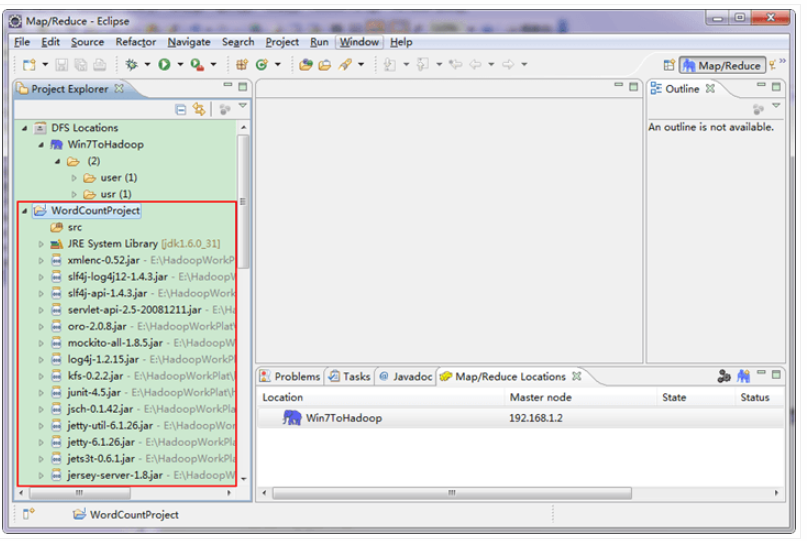


填写 MapReduce 工程的名字为 WordCountProject

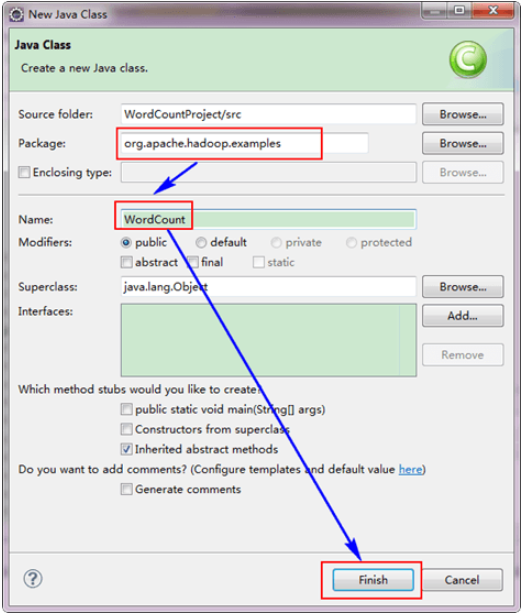


**这里注意 use default Hadoop Configure Hadoop install dir 这里填入的是整个hadoop的文件夹 因为里面会包含很多进行程序编译的jar包（前面好像有一个地方要填写一次，这里保持一致就行）**

**这样就已成功创建了MapReduce项目，左侧能看到刚才建立的项目。**



**接着右键点击刚创建的 WordCountProject，选择 New -> Class，并填写如下信息。**

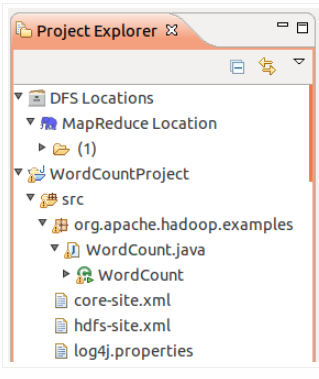


接着将 WordCount 的代码复制进去，默认在/usr/local/hadoop/src/examples/org/apache/hadoop/examples。

并将代码 main() 函数的 String[] otherArgs = new GenericOptionsParser(conf, args).getRemainingArgs(); 改为：

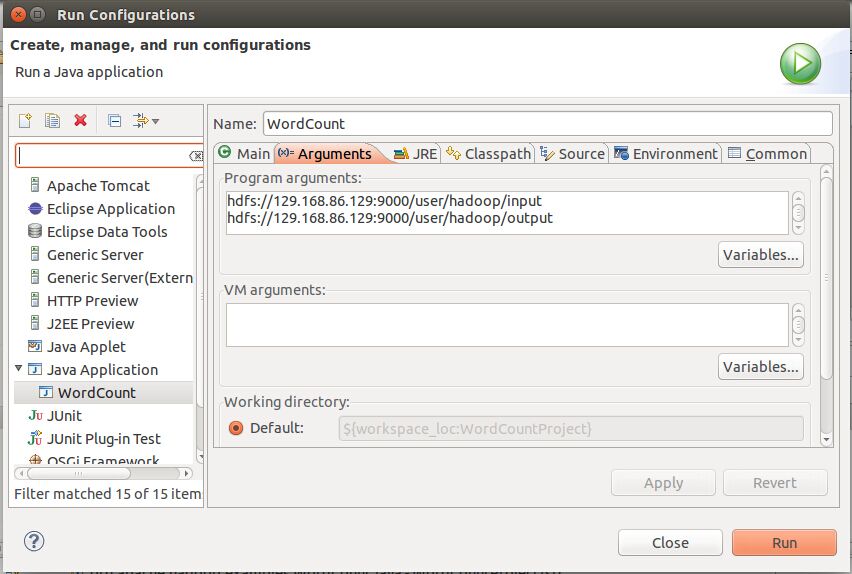
1. // String[] otherArgs = new GenericOptionsParser(conf, args).getRemainingArgs();
2. String[] otherArgs=new String[]{"input","output"}; /\* 直接设置输入参数 \*/

**接着将 /usr/local/hadoop/conf 中将有修改过的配置文件（如伪分布式需要 core-site.xml 和 hdfs-site.xml），以及 log4j.properties 复制到 WordCountProject下的 src 文件夹（workspace/WordCountProject/src）中。否则程序将无法运行，复制完成后，右键点击 WordCountProject 进行刷新，可以看到文件结构如下所示：**

****

**最后右键点击 Project Explorer 中的 WordCount.java，选择 Run As -> Run on Hadoop。可以看到运行成功的提示，刷新 DFS Location 后也能看到输出的 output 文件夹。**

**这里要注意 在运行 Run As -> Run on Hadoop之前先 Run As-> Run Configurations->Arguments填写 input Output 参数**

****

**最后有一点，在run on Hadoop 后可能出现编译的错误，按照提示the compiler compliance has been changed to 1.5 将编译器set 到1.5**