**首先说一下本人的环境:**

**Windows7  64位系统**

**Spring Tool Suite  Version: 3.4.0.RELEASE**

**Hadoop2.6.0**

# 一．简介

  Hadoop2.x之后没有Eclipse插件工具，我们就不能在Eclipse上调试代码，我们要把写好的java代码的MapReduce打包成jar然后在Linux上运行，所以这种不方便我们调试代码，所以我们自己编译一个Eclipse插件，方便我们在我们本地上调试，经过hadoop1.x的发展，编译hadoop2.x版本的eclipse插件比之前简单多了。接下来我 们开始编译Hadoop-eclipse-plugin插件，并在Eclipse开发Hadoop。

# 二．软件安装并配置

## 1.JDK配置

    1) 安装jdk

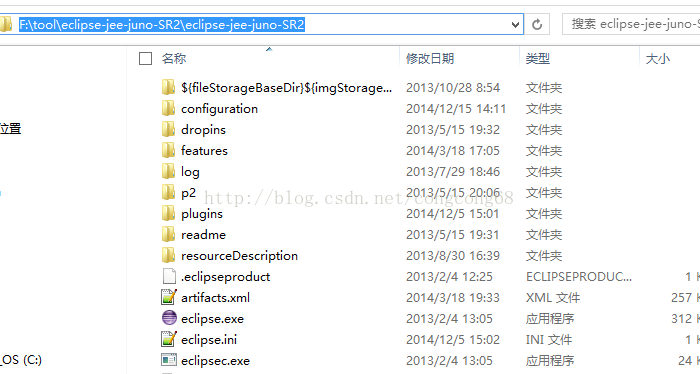
    2) 配置环境变量

      JAVA\_HOME、CLASSPATH、PATH等设置，这里就不多介绍，网上很多资料

## 2.Eclipse

   1).下载eclipse-jee-juno-SR2.rar

   2).解压到本地磁盘，如图所示：



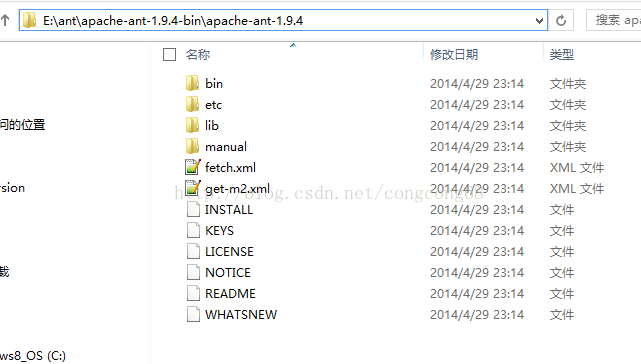
## 3.Ant

  1）下载

[http://ant.apache.org/bindownload.cgi](http://ant.apache.org/bindownload.cgi" \t "https://my.oschina.net/muou/blog/_blank)

   apache-ant-1.9.4-bin.zip

 2）解压到一个盘，如图所示：



 3）.环境变量的配置

    新建ANT\_HOME=E:\ant\apache-ant-1.9.4-bin\apache-ant-1.9.4

    在PATH后面加;%ANT\_HOME%\bin

 4）cmd 测试一下是否配置正确

    ant version   如图所示：

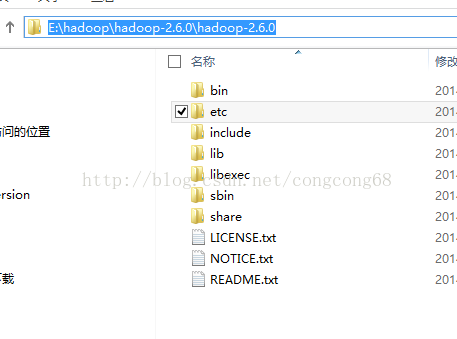
IMG_258

## 4.Hadoop

 1).下载hadoop包

    hadoop-2.6.0.tar.gz

   解压到本地磁盘，如图所示：



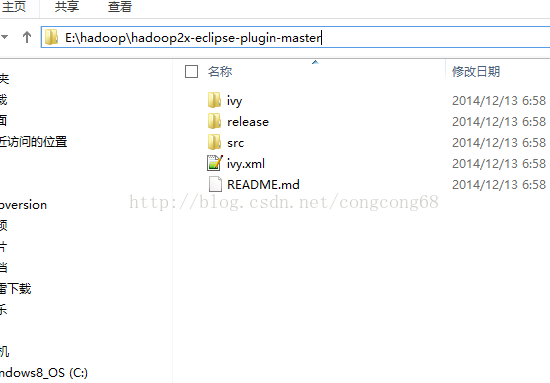
**下载hadoop2x-eclipse-plugin源代码**

 1）目前hadoop2的eclipse-plugins源代码由github脱管，下载地址是[https://github.com/winghc/hadoop2x-eclipse-plugin](https://github.com/winghc/hadoop2x-eclipse-plugin" \t "https://my.oschina.net/muou/blog/_blank)，然后在右侧的Download ZIP连接点击下载，如图所示：



2）下载hadoop2x-eclipse-plugin-master.zip

   解压到本地磁盘，如图所示：



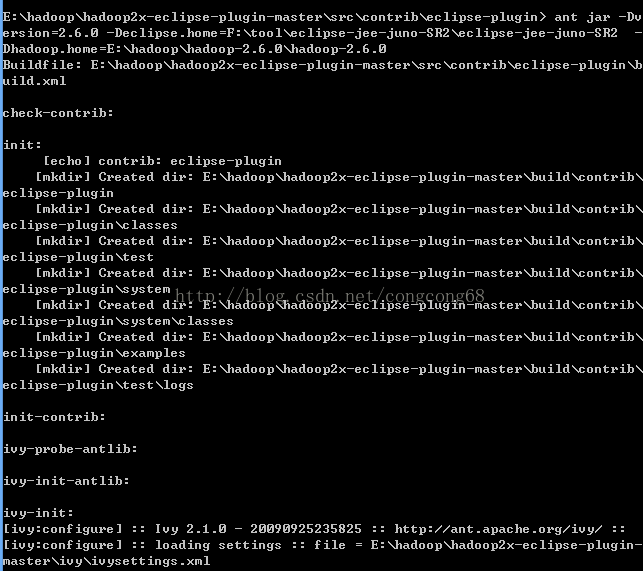
# 三．编译hadoop-eclipse-plugin插件

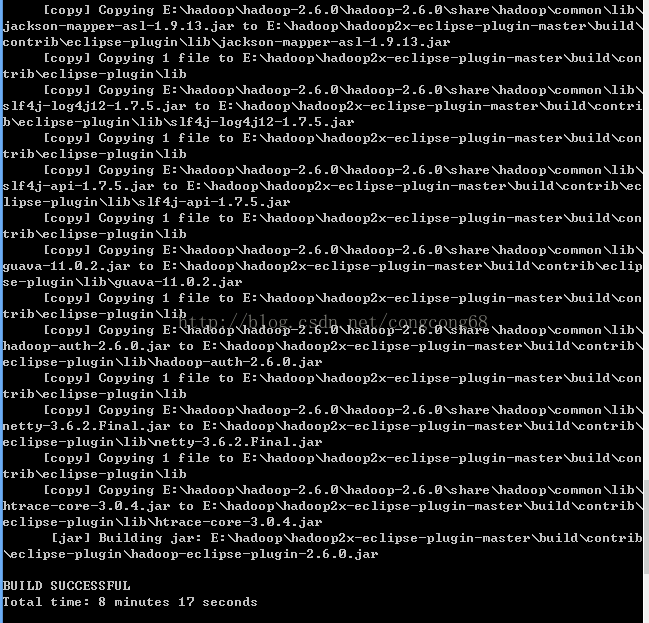
## 1.hadoop2x-eclipse-plugin-master解压在E：盘打开命令行cmd，切换到E:\hadoop\hadoop2x-eclipse-plugin-master\src\contrib\eclipse-plugin 目录，如图所示：

IMG_262

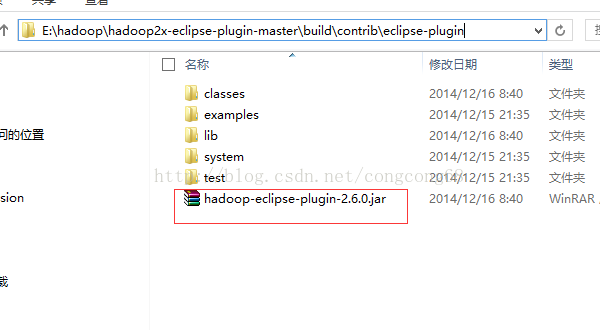
## 2.执行ant jar

 antjar -Dversion=2.6.0 -Declipse.home=F:\tool\eclipse-jee-juno-SR2\eclipse-jee-juno-SR2 -Dhadoop.home=E:\hadoop\hadoop-2.6.0\hadoop-2.6.0，如图所示：



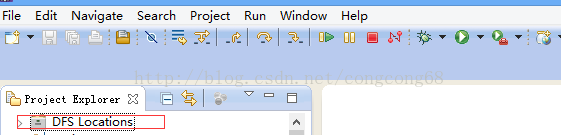


## 3.编译成功生成的hadoop-eclipse-plugin-2.6.0.jar在E:\hadoop\hadoop2x-eclipse-plugin-master\build\contrib\eclipse-plugin路径下，如图所示：

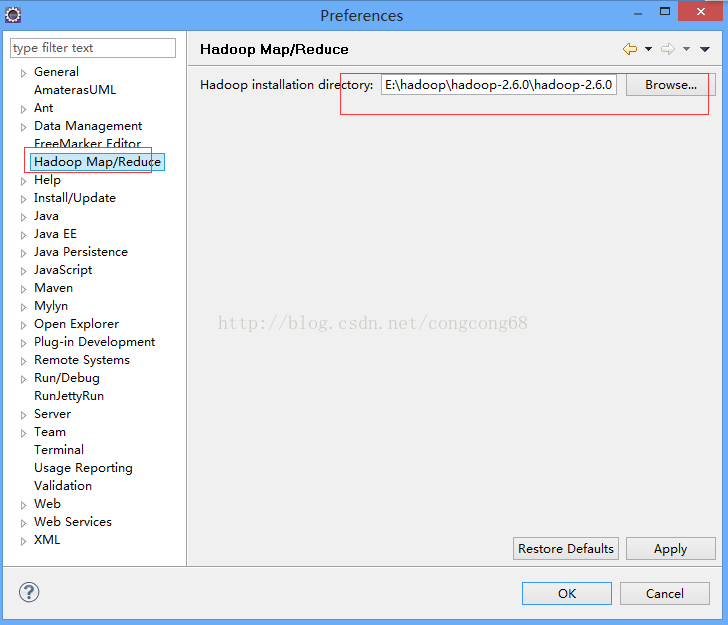


# 四．Eclipse配置hadoop-eclipse-plugin 插件

## 1.我已经把可以用的插件包上传了，[http://pan.baidu.com/s/1qWG7XxU](https://codeload.github.com/srccodes/hadoop-common-2.2.0-bin/zip/master" \t "https://my.oschina.net/muou/blog/_blank)(请注意我的环境win7 64位)把hadoop-eclipse-plugin-2.6.0.jar拷贝到F:\tool\eclipse-jee-juno-SR2\eclipse-jee-juno-SR2\plugins目录下，重启一下Eclipse，然后可以看到DFS Locations,如图所示：



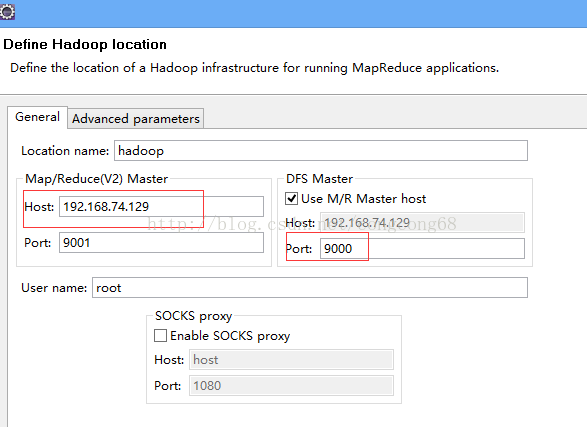
## 2.打开Window-->Preferens，可以看到Hadoop Map/Reduc选项，然后点击，然后添加hadoop-2.6.0进来，如图所示：



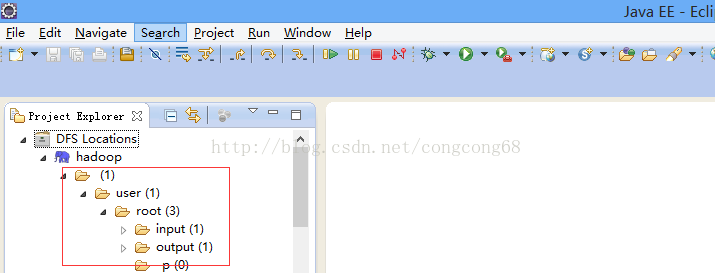
## 3.配置Map/ReduceLocations

   1）点击Window-->Show View -->MapReduce Tools  点击Map/ReduceLocation

   2）点击Map/ReduceLocation选项卡，点击右边小象图标，打开Hadoop Location配置窗口： 输入Location Name，任意名称即可.配置Map/Reduce Master和DFS Mastrer，Host和Port配置成hdfs-site.xml与core-site.xml的设置一致即可。



## 4.查看是否连接成功



# 五.新建MapReduce项目并运行

   1.右击New->Map/Reduce Project

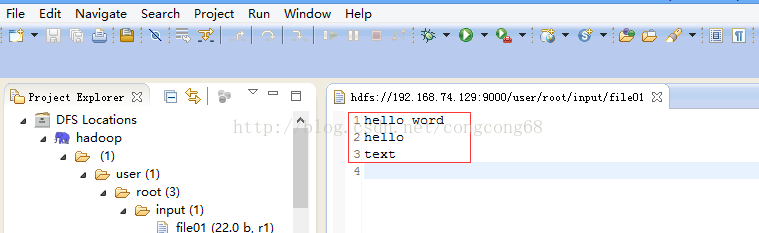
   2.新建WordCount.java（在Hadoop的share目录下找到mapreduce的案例，copy过来）

   3.在hdfs创建一个input目录（输出目录可以不用创建,运行MR是会自动创建），并上传一个file01文件(随便写几个单词)

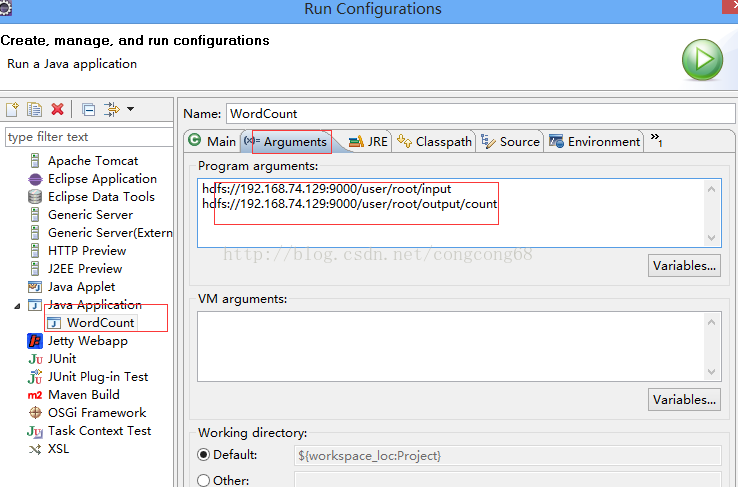
       hdfs dfs -mkdir –p  /user/root/input

       hdfs dfs -mkdir -p  /user/root/output

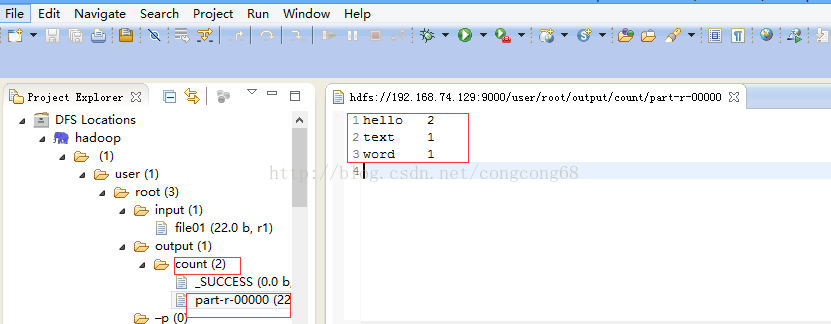
        hadoop fs -put file01 /input



 4.点击WordCount.java右击-->Run As-->Run COnfigurations   设置输入和输出目录路径，如图所示：



 5.点击WordCount.java右击-->Run As-->Run on  Hadoop

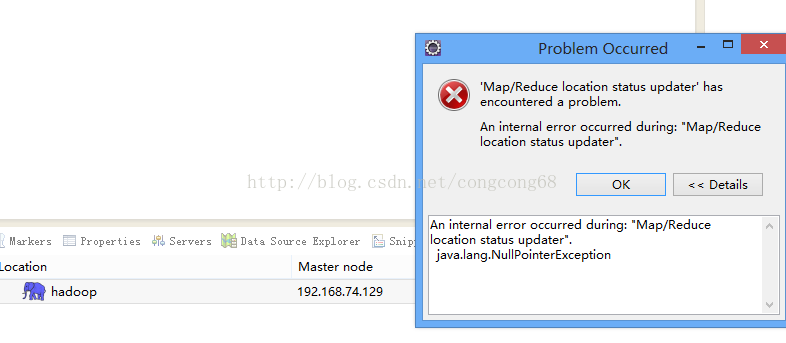


 然后到output/count目录下，有一个统计文件，并查看结果，所以配置成功。

# 六、****接下来就是可能遇到的问题：****

**问题一.An internal error occurred during: "Map/Reducelocation status updater".java.lang.NullPointerException**

   我们hadoop-eclipse-plugin-2.6.0.jar放到Eclipse的plugins目录下，我们的Eclipse目录是F:\tool\eclipse-jee-juno-SR2\eclipse-jee-juno-SR2\plugins，重启一下Eclipse，然后，打开Window-->Preferens，可以看到Hadoop Map/Reduc选项，然后点击出现了An internal error occurredduring: "Map/Reduce location status updater".java.lang.NullPointerException，如图所示：



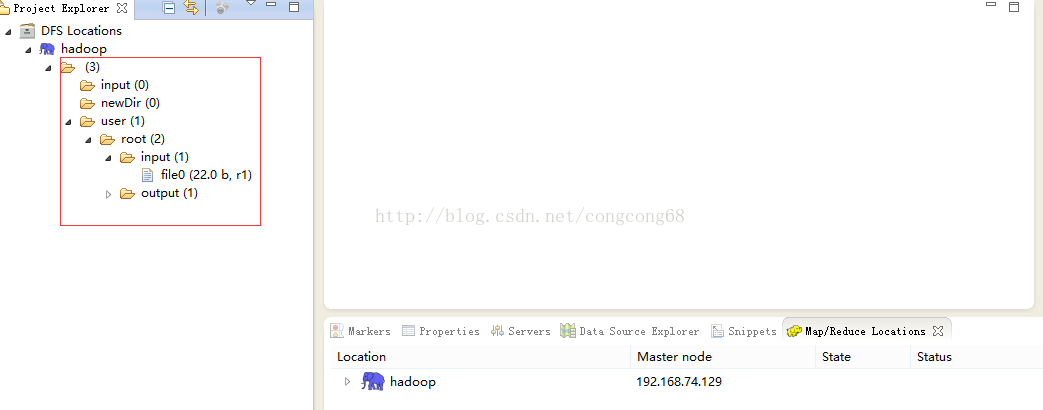
  解决：

   我们发现刚配置部署的Hadoop2还没创建输入和输出目录，先在hdfs上建个文件夹 。

   #bin/hdfs dfs -mkdir –p /user/root/input

   #bin/hdfs dfs -mkdir -p  /user/root/output

 我们在Eclipse的DFS Locations目录下看到我们这两个目录，如图所示：



**问题二.Exception in thread "main" java.lang.NullPointerException atjava.lang.ProcessBuilder.start(Unknown Source)**

运行Hadoop2的WordCount.java代码时出现了这样错误，

  log4j:WARNPlease initialize the log4j system properly.log4j:WARN Seehttp://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.

Exception in thread "main" java.lang.NullPointerException

       atjava.lang.ProcessBuilder.start(Unknown Source)

       atorg.apache.hadoop.util.Shell.runCommand(Shell.java:482)

       atorg.apache.hadoop.util.Shell.run(Shell.java:455)

       atorg.apache.hadoop.util.Shell$ShellCommandExecutor.execute(Shell.java:715)

       atorg.apache.hadoop.util.Shell.execCommand(Shell.java:808)

       atorg.apache.hadoop.util.Shell.execCommand(Shell.java:791)

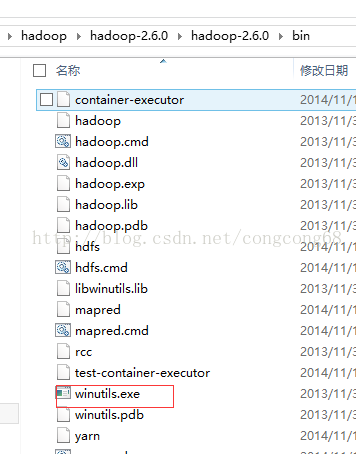
       at

分析：

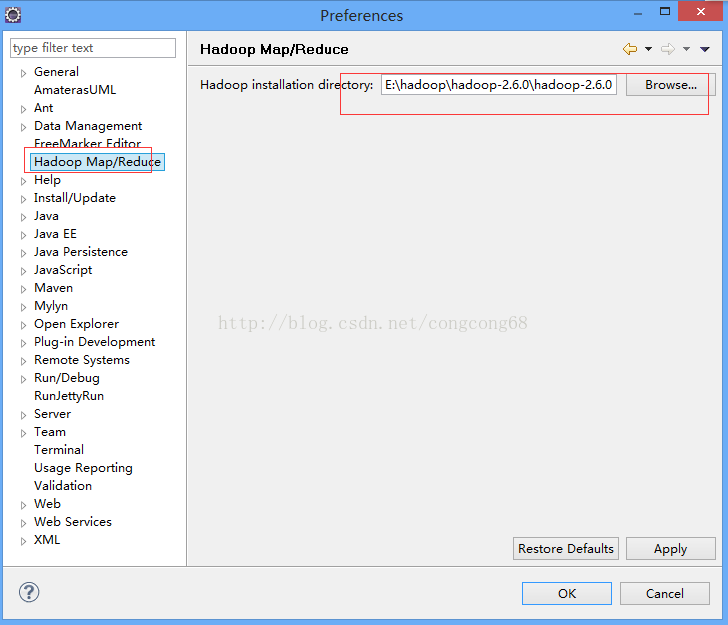
  下载Hadoop2以上版本时，在Hadoop2的bin目录下没有winutils.exe

解决：

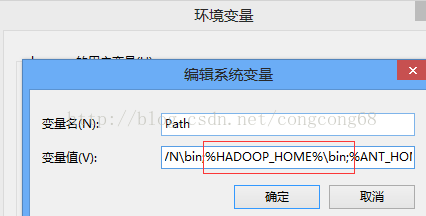
  1.下载[http://pan.baidu.com/s/1qWG7XxU](https://codeload.github.com/srccodes/hadoop-common-2.2.0-bin/zip/master" \t "https://my.oschina.net/muou/blog/_blank)下载Hadoop2.6.0-eclipse插件.zip，然后解压后，把Hadoop2.6.0-eclipse插件.zip\eclipse插件\2.4以后的目录中的winutils.exe复制Hadoop2/bin目录下。如图所示：



  2.Eclipse-》window-》Preferences 下的Hadoop Map/Peduce 把下载放在我们的磁盘的Hadoop目录引进来，如图所示：



  3.Hadoop2配置变量环境HADOOP\_HOME 和path，如图所示：



**问题三.Exception in thread "main"java.lang.UnsatisfiedLinkError:org.apache.hadoop.io.nativeio.NativeIO$Windows.access0(Ljava/lang/String;I)Z**

  当我们解决了问题三时，在运行WordCount.java代码时，出现这样的问题

log4j:WARN No appenders could be found forlogger (org.apache.hadoop.metrics2.lib.MutableMetricsFactory).log4j:WARN Please initialize the log4jsystem properly.log4j:WARN Seehttp://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.

Exception in thread "main"java.lang.UnsatisfiedLinkError:org.apache.hadoop.io.nativeio.NativeIO$Windows.access0(Ljava/lang/String;I)Z

       atorg.apache.hadoop.io.nativeio.NativeIO$Windows.access0(Native Method)

       atorg.apache.hadoop.io.nativeio.NativeIO$Windows.access(NativeIO.java:557)

       atorg.apache.hadoop.fs.FileUtil.canRead(FileUtil.java:977)

       atorg.apache.hadoop.util.DiskChecker.checkAccessByFileMethods(DiskChecker.java:187)

       atorg.apache.hadoop.util.DiskChecker.checkDirAccess(DiskChecker.java:174)

       atorg.apache.hadoop.util.DiskChecker.checkDir(DiskChecker.java:108)

       atorg.apache.hadoop.fs.LocalDirAllocator$AllocatorPerContext.confChanged(LocalDirAllocator.java:285)

       atorg.apache.hadoop.fs.LocalDirAllocator$AllocatorPerContext.getLocalPathForWrite(LocalDirAllocator.java:344)

       atorg.apache.hadoop.fs.LocalDirAllocator.getLocalPathForWrite(LocalDirAllocator.java:150)

       atorg.apache.hadoop.fs.LocalDirAllocator.getLocalPathForWrite(LocalDirAllocator.java:131)

       atorg.apache.hadoop.fs.LocalDirAllocator.getLocalPathForWrite(LocalDirAllocator.java:115)

       atorg.apache.hadoop.mapred.LocalDistributedCacheManager.setup(LocalDistributedCacheManager.java:131)

 分析：

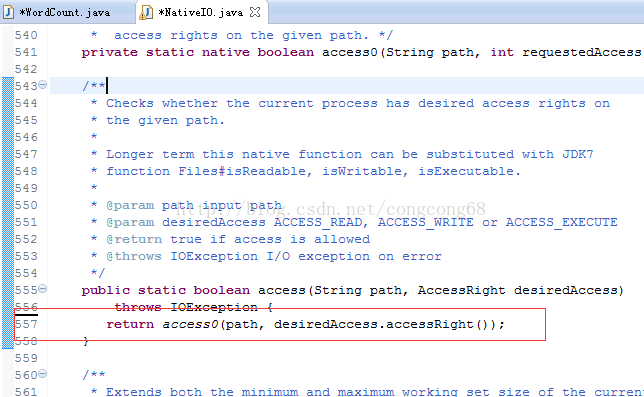
    C:\Windows\System32下缺少hadoop.dll,把这个文件拷贝到C:\Windows\System32下面即可。

 解决：

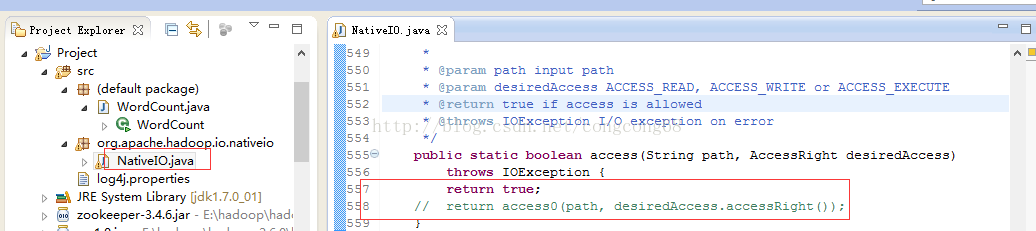
   将压缩包中的hadoop.dll放到C:\Windows\System32下，然后重启电脑，也许还没那么简单，还是出现这样的问题。如果这个还是没解决，最好在%HADOOP\_HOME%/bin目录下面也复制一份。

  我们在继续分析：

    我们在出现错误的的atorg.apache.hadoop.io.nativeio.NativeIO$Windows.access(NativeIO.java:557)我们来看这个类NativeIO的557行，如图所示：



   Windows的唯一方法用于检查当前进程的请求，在给定的路径的访问权限，所以我们先给以能进行访问，我们自己先修改源代码，return true 时允许访问。我们下载对应hadoop源代码，hadoop-2.6.0-src.tar.gz解压，hadoop-2.6.0-src\hadoop-common-project\hadoop-common\src\main\java\org\apache\hadoop\io\nativeio下NativeIO.java 复制到对应的Eclipse的project，然后修改557行为return true如图所示：



**问题四：org.apache.hadoop.security.AccessControlException: Permissiondenied: user=zhengcy, access=WRITE,inode="/user/root/output":root:supergroup:drwxr-xr-x**

  我们在执行运行WordCount.java代码时，出现这样的问题

2014-12-18 16:03:24,092  WARN (org.apache.hadoop.mapred.LocalJobRunner:560) - job\_local374172562\_0001

org.apache.hadoop.security.AccessControlException: Permission denied: user=zhengcy, access=WRITE, inode="/user/root/output":root:supergroup:drwxr-xr-x

at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkFsPermission(FSPermissionChecker.java:271)

at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.check(FSPermissionChecker.java:257)

at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.check(FSPermissionChecker.java:238)

at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkPermission(FSPermissionChecker.java:179)

at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkPermission(FSNamesystem.java:6512)

at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkPermission(FSNamesystem.java:6494)

at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkAncestorAccess(FSNamesystem.java:6446)

at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.mkdirsInternal(FSNamesystem.java:4248)

at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.mkdirsInt(FSNamesystem.java:4218)

at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.mkdirs(FSNamesystem.java:4191)

at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.mkdirs(NameNodeRpcServer.java:813)

 分析：

  我们没权限访问output目录。

解决：

    我们 在设置hdfs配置的目录是在hdfs-site.xml配置hdfs文件存放的地方，我在hadoop伪分布式部署那边有介绍过，我们在这边在复习一下，如图所示：

我们在这个etc/hadoop下的hdfs-site.xml添加

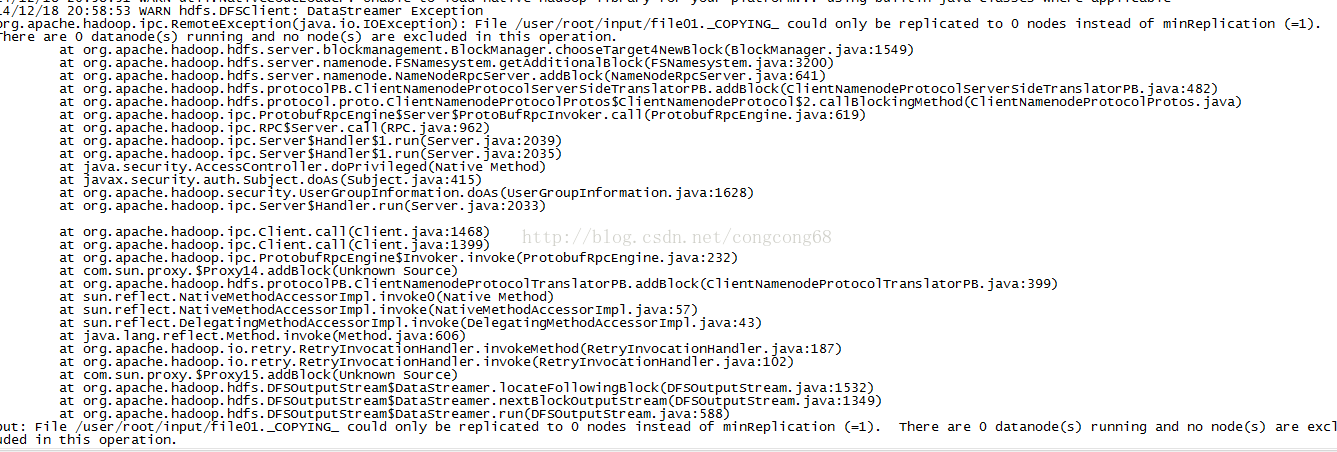
  <property>

     <name>dfs.permissions</name>   
     <value>false</value>   
  </property>

设置没有权限，不过我们在正式的 服务器上不能这样设置。

**问题五：File/usr/root/input/file01.\_COPYING\_ could only be replicated to 0 nodes instead ofminRepLication (=1) There are 0 datanode(s) running and no node(s) are excludedin this operation**

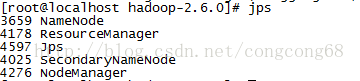
     如图所示：



  分析：

  我们在第一次执行#hadoop namenode –format 完然后在执行#sbin/start-all.sh

在执行#jps，能看到Datanode，在执行#hadoop namenode –format然后执行#jps这时看不到Datanode ,如图所示：



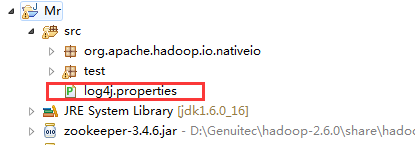
   然后我们想把文本放到输入目录执行bin/hdfs dfs -put/usr/local/hadoop/hadoop-2.6.0/test/\* /user/root/input  把/test/\*文件上传到hdfs的/user/root/input中,出现这样的问题，

 解决：

  是我们执行太多次了hadoopnamenode –format，在创建了多个，我们对应的hdfs目录删除hdfs-site.xml配置的保存datanode和namenode目录。

**问题六：在复制了hadoop.dll后，运行WordCount，发现运行一会没有任何信息输出就结束了**

**解决:可以写一个log4j日志文件，查看一下日志的输出，可能从输出的日志中发现问题。**

****

**内容写为:**

log4j.rootLogger=debug,stdout,R log4j.appender.stdout=org.apache.log4j.ConsoleAppender log4j.appender.stdout.layout=org.apache.log4j.PatternLayout log4j.appender.stdout.layout.ConversionPattern=%5p - %m%n log4j.appender.R=org.apache.log4j.RollingFileAppender log4j.appender.R.File=mapreduce\_test.log log4j.appender.R.MaxFileSize=1MB log4j.appender.R.MaxBackupIndex=1 log4j.appender.R.layout=org.apache.log4j.PatternLayout log4j.appender.R.layout.ConversionPattern=%p %t %c - %m%n log4j.logger.com.codefutures=DEBUG

**问题七：有了log4j日志输出后，查看问题就比较方便了，如果同一个MR执行两次，会出现输出文件已存在的问题**

**解决：可以删除掉存在的输出文件，也可以改代码中输出的路径**

Exception in thread "main" org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://192.168.233.11:8020/mroutput already exists

at org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileOutputFormat.java:146)

at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:562)

at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:432)

at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1296)

at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1293)

at java.security.AccessController.doPrivileged(Native Method)

at javax.security.auth.Subject.doAs(Subject.java:396)

at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1628)

at org.apache.hadoop.mapreduce.Job.submit(Job.java:1293)

at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1314)

at test.WordCount.main(WordCount.java:87)

**问题八:出现内存溢出的问题java.lang.OutOfMemoryError**

  WARN - job\_local845949011\_0001

 java.lang.Exception: java.lang.OutOfMemoryError: Java heap space

at org.apache.hadoop.mapred.LocalJobRunner$Job.runTasks(LocalJobRunner.java:462)

at org.apache.hadoop.mapred.LocalJobRunner$Job.run(LocalJobRunner.java:522)

Caused by: java.lang.OutOfMemoryError: Java heap space

at org.apache.hadoop.mapred.MapTask$MapOutputBuffer.init(MapTask.java:983)

at org.apache.hadoop.mapred.MapTask.createSortingCollector(MapTask.java:401)

at org.apache.hadoop.mapred.MapTask.access$100(MapTask.java:81)

at org.apache.hadoop.mapred.MapTask$NewOutputCollector.<init>(MapTask.java:695)

at org.apache.hadoop.mapred.MapTask.runNewMapper(MapTask.java:767)

at org.apache.hadoop.mapred.MapTask.run(MapTask.java:341)

at org.apache.hadoop.mapred.LocalJobRunner$Job$MapTaskRunnable.run(LocalJobRunner.java:243)

at java.util.concurrent.Executors$RunnableAdapter.call(Executors.java:441)

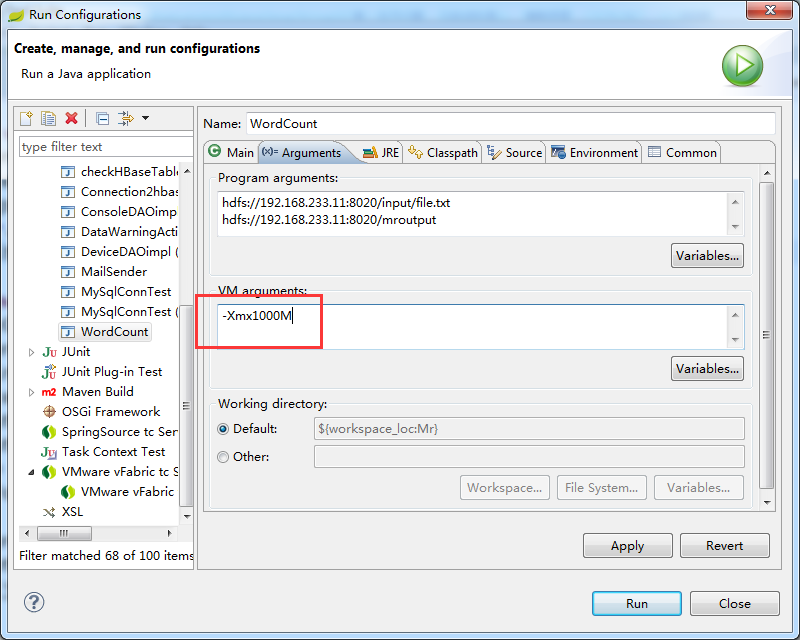
at java.util.concurrent.FutureTask$Sync.innerRun(FutureTask.java:303)

at java.util.concurrent.FutureTask.run(FutureTask.java:138)

at java.util.concurrent.ThreadPoolExecutor$Worker.runTask(ThreadPoolExecutor.java:886)

at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:908)

at java.lang.Thread.run(Thread.java:619)

**解决：右键WordCount,-->run Confi....**

感谢:（部分内容摘自下面，自己做了一些修改和补充）