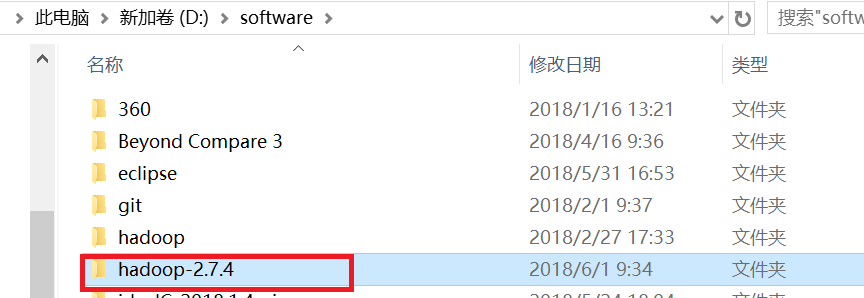
windows下搭建idea运行mapreduce环境

相关代码路径：

# 将下载的hadoop文件拷贝到windows磁盘目录下并解压,解压过程出错忽略就可以

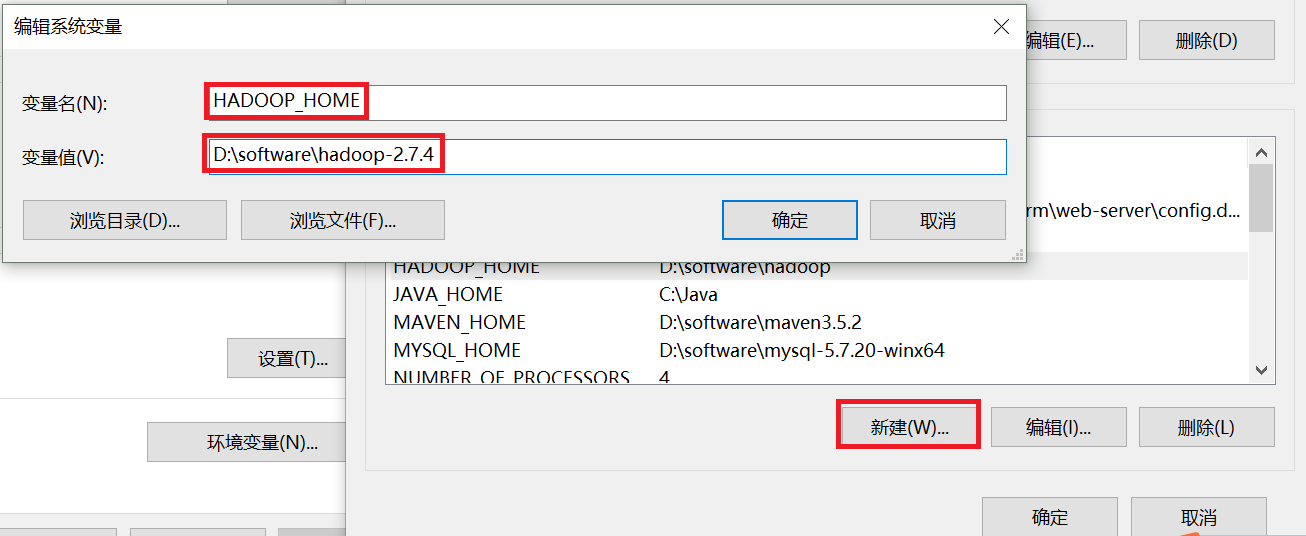


# 设置环境变量

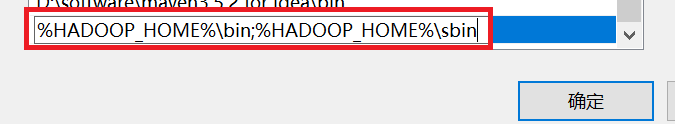
## 设置HADOOP\_HOME

变量名：HADOOP\_HOME

变量值：D:\software\hadoop-2.7.4



## 设置path路径



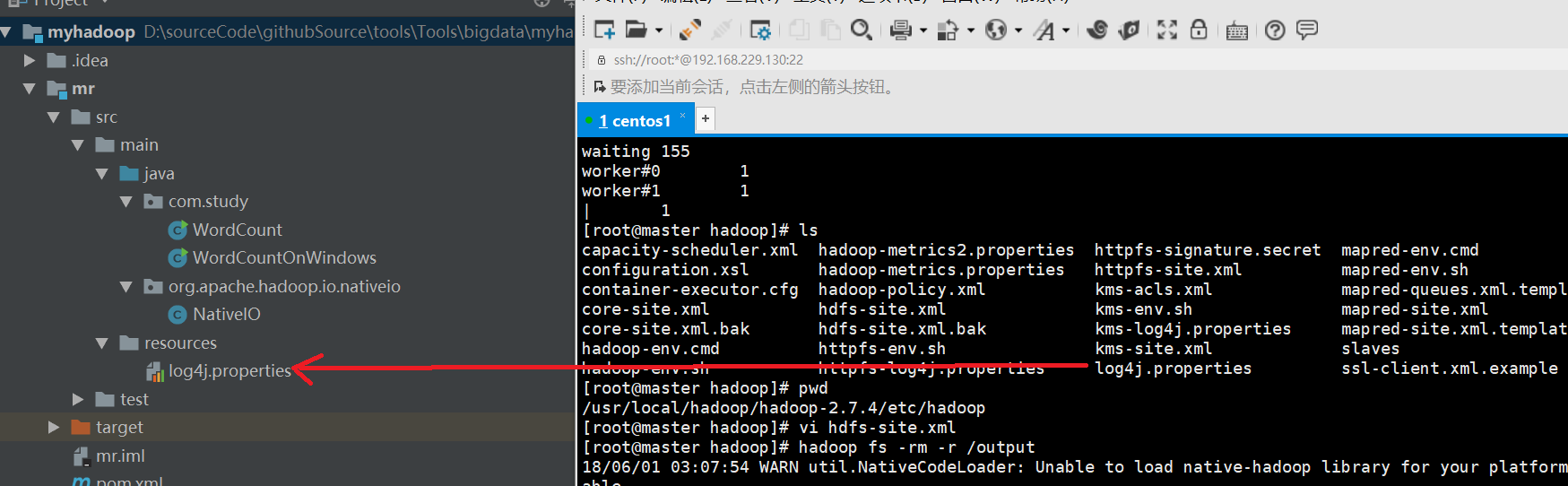
# 配置windows下运行需要的dll

需要将hadoop.dll和winutils.exe拷贝到windows下hadoop的bin目录，拷贝完成后重新启动ide



# 配置日志

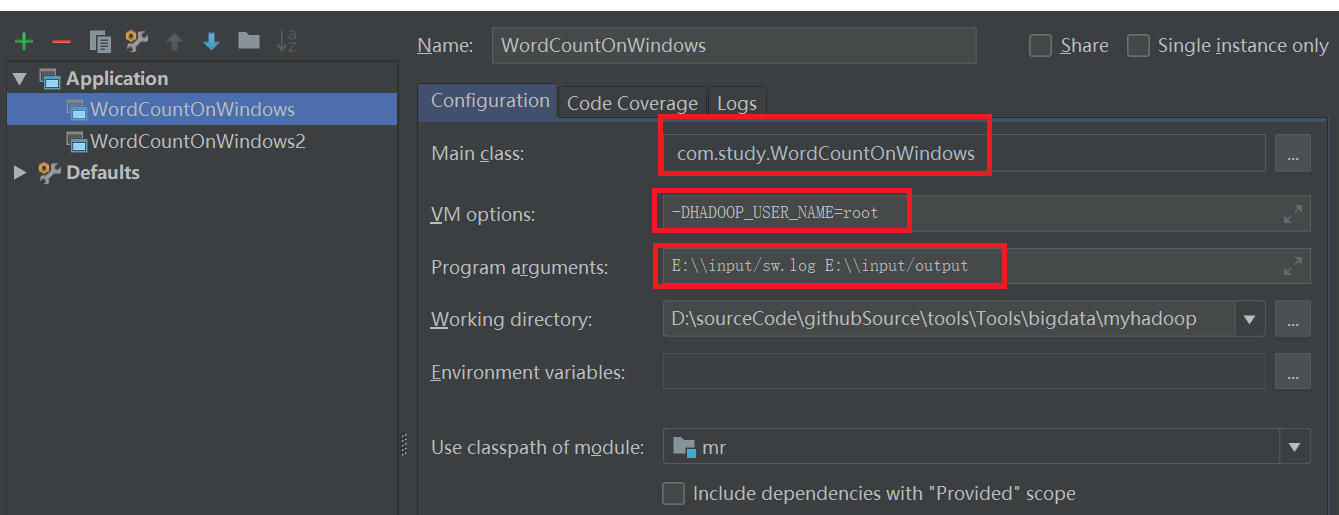
我们将hadoop安装目录配置文件下的log4j.properties 拷贝到我们的程序中



# 运行我们的mapreduce程序

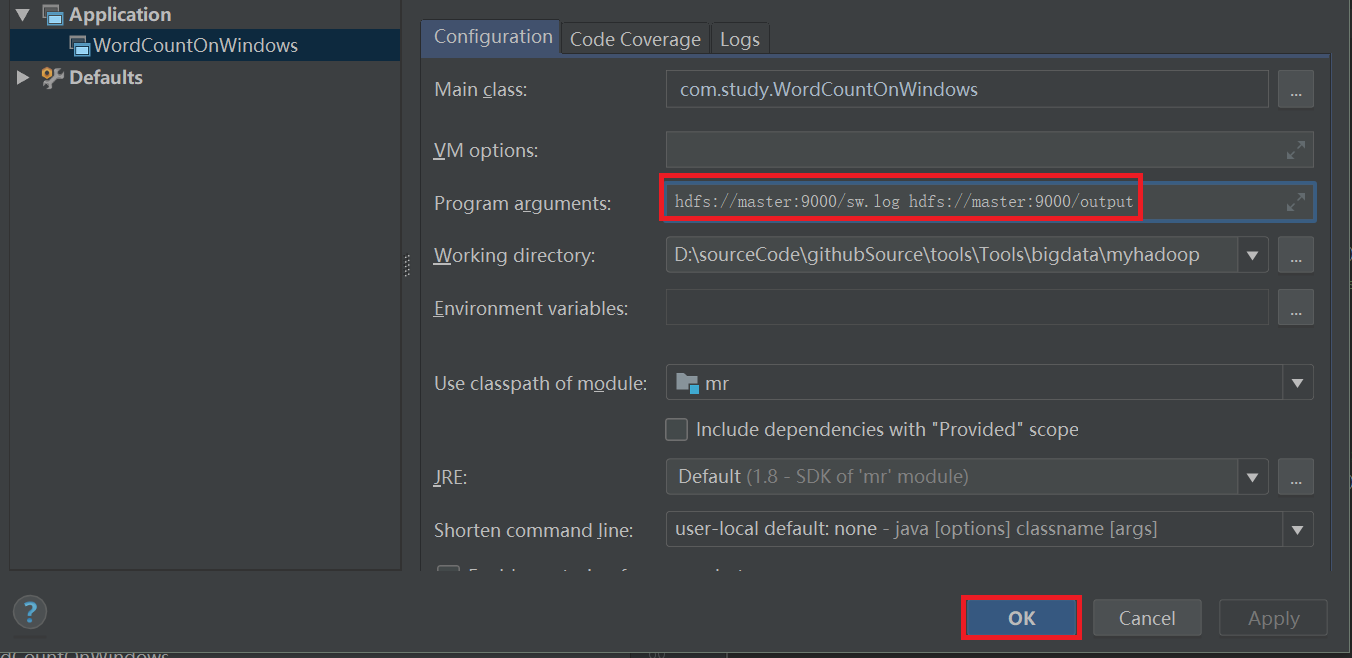
## 本地计算-本地hdfs文件

sw.log 为我们测试文件，大家随便定义



## 本地计算-远程hdfs文件

将远程的hdfs文件信息拉到本地执行mapreduce



## 远程计算-远程hdfs文件

我们将程序打成jar使用代码提交给远程的hadoop集群

修改代码，添加如下配置



# Windows下运行mapreduce常见错误

## 没有设置hdfs路径的错误

错误提示：Exception in thread "main" java.io.IOException: Cannot initialize Cluster. Please check your configuration for mapreduce.framework.name and the correspond server addresses.

at org.apache.hadoop.mapreduce.Cluster.initialize(Cluster.java:120)

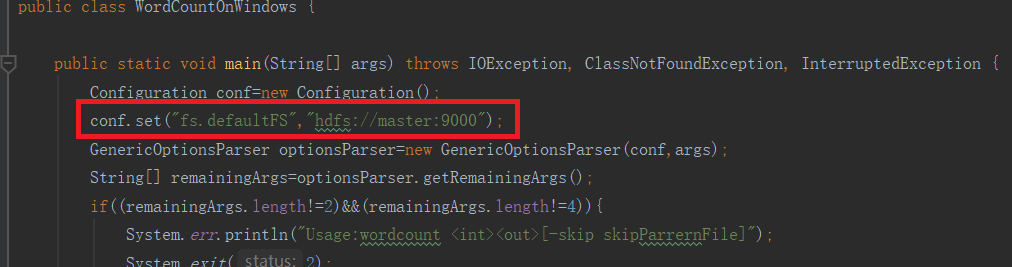
at org.apache.hadoop.mapreduce.Cluster.<init>(Cluster.java:82)

at org.apache.hadoop.mapreduce.Cluster.<init>(Cluster.java:75)

at org.apache.hadoop.mapreduce.Job$9.run(Job.java:1260)

解决办法：

设置hdfs路径：conf.set("fs.defaultFS","hdfs://master:9000");



## 没有引入hdfs相关jar的错误

错误信息：

Exception in thread "main" java.io.IOException: No FileSystem for scheme: hdfs

at org.apache.hadoop.fs.FileSystem.getFileSystemClass(FileSystem.java:2660)

at org.apache.hadoop.fs.FileSystem.createFileSystem(FileSystem.java:2667)

解决办法：

引入hdfs相关的jar包

<dependency>  
 <groupId>org.apache.hadoop</groupId>  
 <artifactId>hadoop-hdfs</artifactId>  
 <version>2.7.4</version>  
</dependency>



## 没有引入mapreduce client相关jar的错误

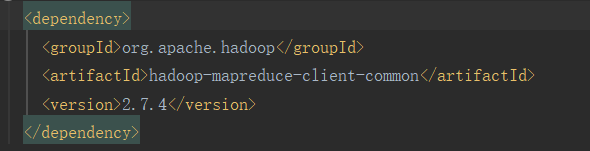
错误信息：

Exception in thread "main" java.io.IOException: Cannot initialize Cluster. Please check your configuration for mapreduce.framework.name and the correspond server addresses.

at org.apache.hadoop.mapreduce.Cluster.initialize(Cluster.java:120)

解决方法：

引入mapreduce相关的jar



## 缺少hadoop.dll

错误信息：

Exception in thread "main" java.lang.UnsatisfiedLinkError: org.apache.hadoop.io.nativeio.NativeIO$Windows.createDirectoryWithMode0(Ljava/lang/String;I)V

at org.apache.hadoop.io.nativeio.NativeIO$Windows.createDirectoryWithMode0(Native Method)

at org.apache.hadoop.io.nativeio.NativeIO$Windows.createDirectoryWithMode(NativeIO.java:524)

解决方法：

根据hadoop版本在网上找haddoop.dll 下载下来，放在我们windows机器上的hadoop/bin目录下，拷贝完成后重新启动ide

## 缺少winutils.exe

错误信息：

Exception in thread "main" java.io.IOException: (null) entry in command string: null chmod 0700 D:\tmp\hadoop-tianl\mapred\staging\tianl1595892264\.staging

at org.apache.hadoop.util.Shell$ShellCommandExecutor.execute(Shell.java:773)

at org.apache.hadoop.util.Shell.execCommand(Shell.java:869)

at org.apache.hadoop.util.Shell.execCommand(Shell.java:852)

at org.apache.hadoop.fs.RawLocalFileSystem.setPermission(RawLocalFileSystem.java:733)

解决办法：

一般在下载hadoop.dll的时候，同时也会有winutils.exe，我们将winutils.exe放在windows

Hadoop的bin目录

## 修改修改hadoop源码

错误信息：

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

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

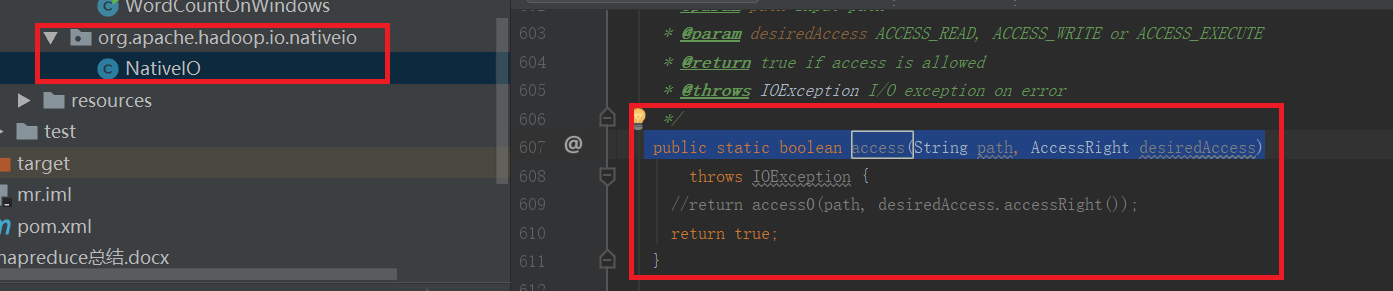
at org.apache.hadoop.io.nativeio.NativeIO$Windows.access(NativeIO.java:609)

解决办法：

* 在我们的工程中新进包：org.apache.hadoop.io.nativeio
* 在将D:\software\hadoop-2.7.4\src\hadoop-common-project\hadoop-common\src\main\java\org\apache\hadoop\io\nativeio下的NativeIO.java拷贝到

新建立的包org.apache.hadoop.io.nativeio下

* 修改NativeIO.java中函数public static boolean access(String path, AccessRight desiredAccess)内容如下：



## 没有访问hdfs的权限

错误信息：

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

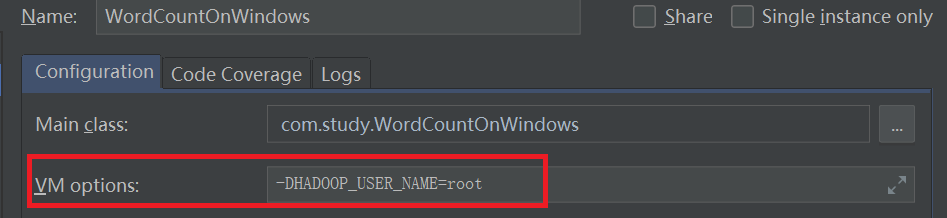
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.check(FSPermissionChecker.java:317)

at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkPermission(FSPermissionChecker.java:223)

at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkPermission(FSPermissionChecker.java:199)

解决办法一：

设置vm option ：-DHADOOP\_USER\_NAME=root



解决办法二：

访问hdfs的时候不检查权限，

* 关闭hadoop集群
* 在/usr/local/hadoop/hadoop-2.7.4/etc/Hadoop/hdfs-sit.xml中增加如下配置

<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

