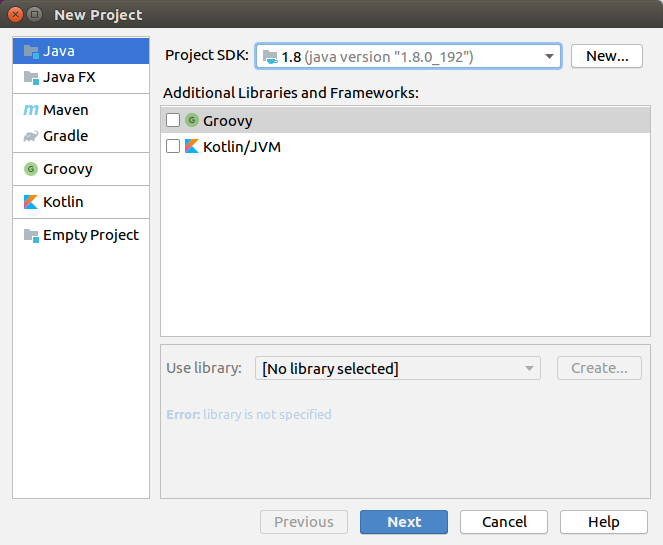
# Idea create java project:

1. [File]→[New]→[Project]

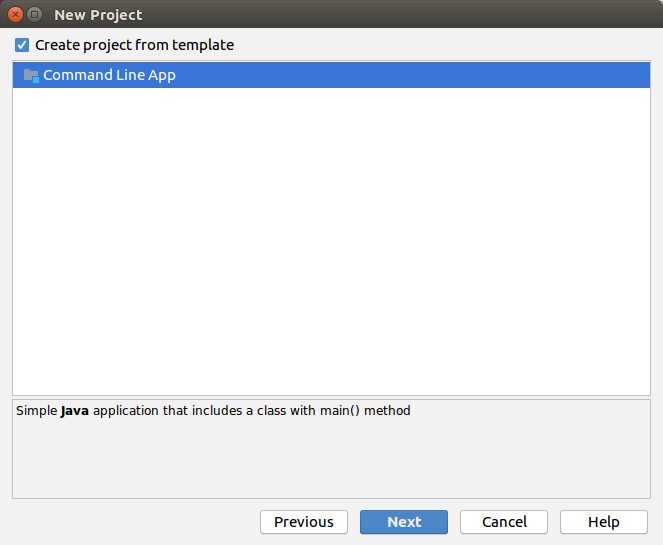
2. select Java

3.Project SDK select 1.8 (use New to locate sdk directory)

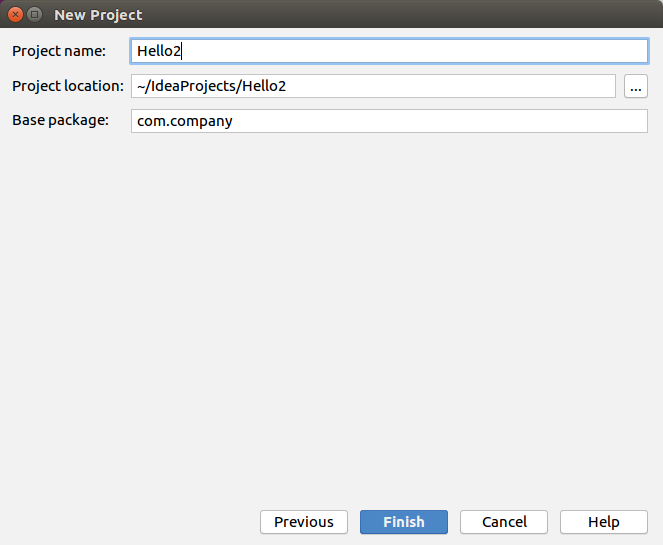
4.Next

5.check Create project from template

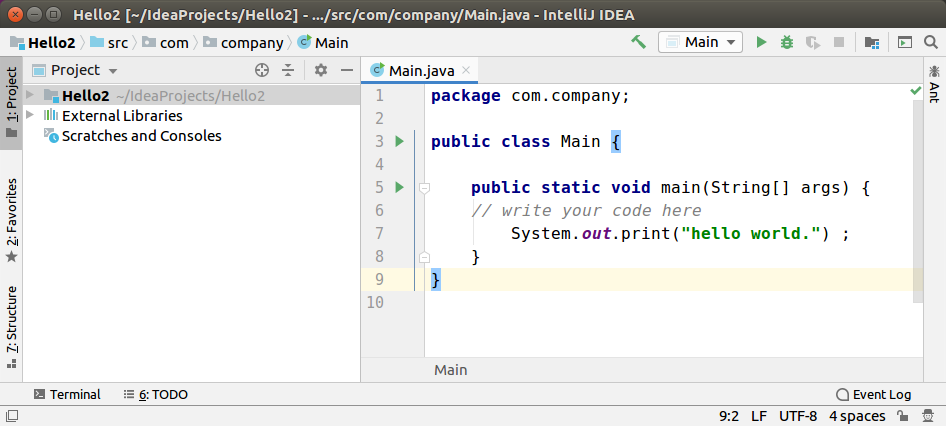
6.select Command Line App



7.input project name



8.Edit the codes



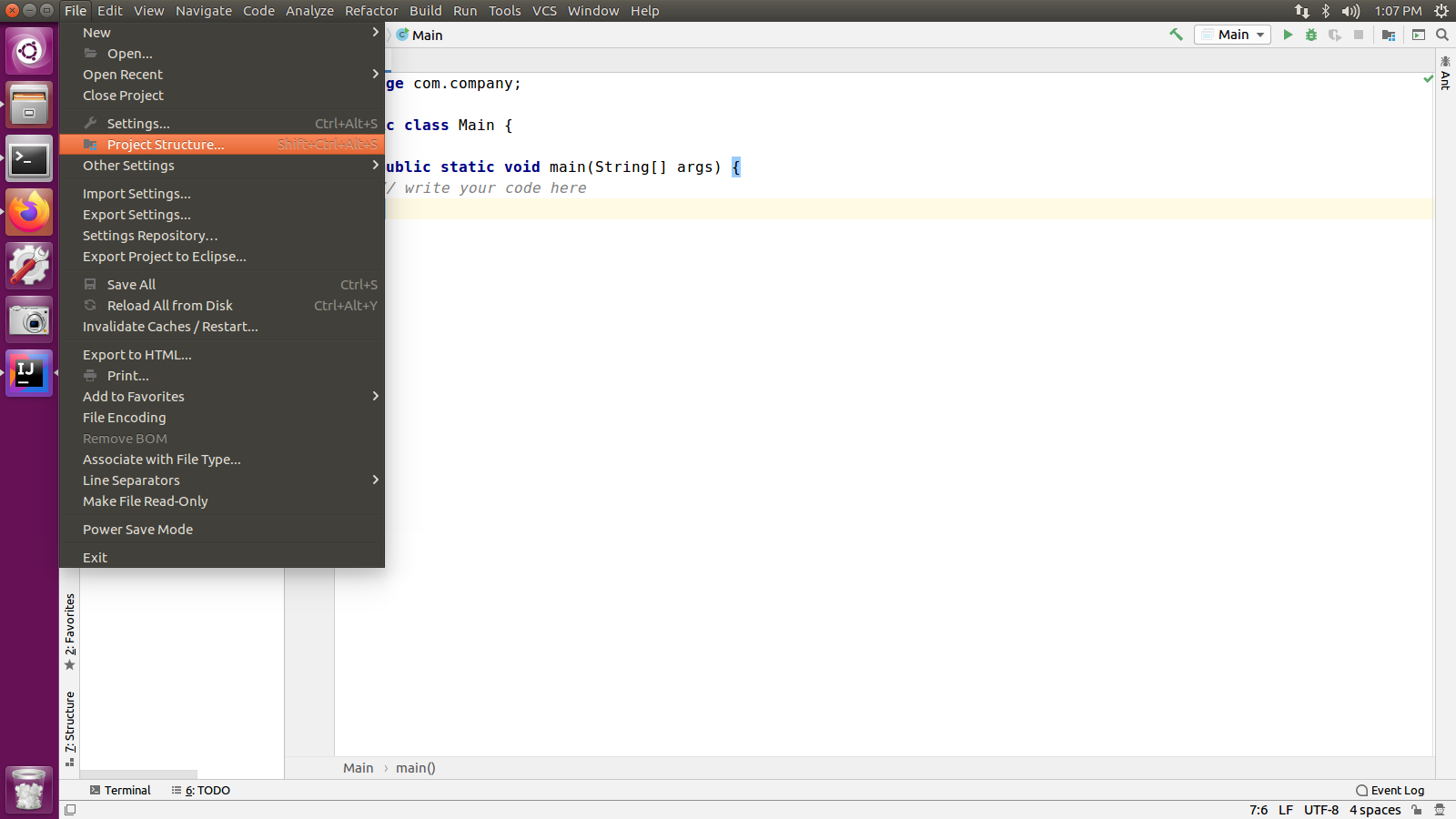
9.Run the app

menu:[Run]→[Run Main]

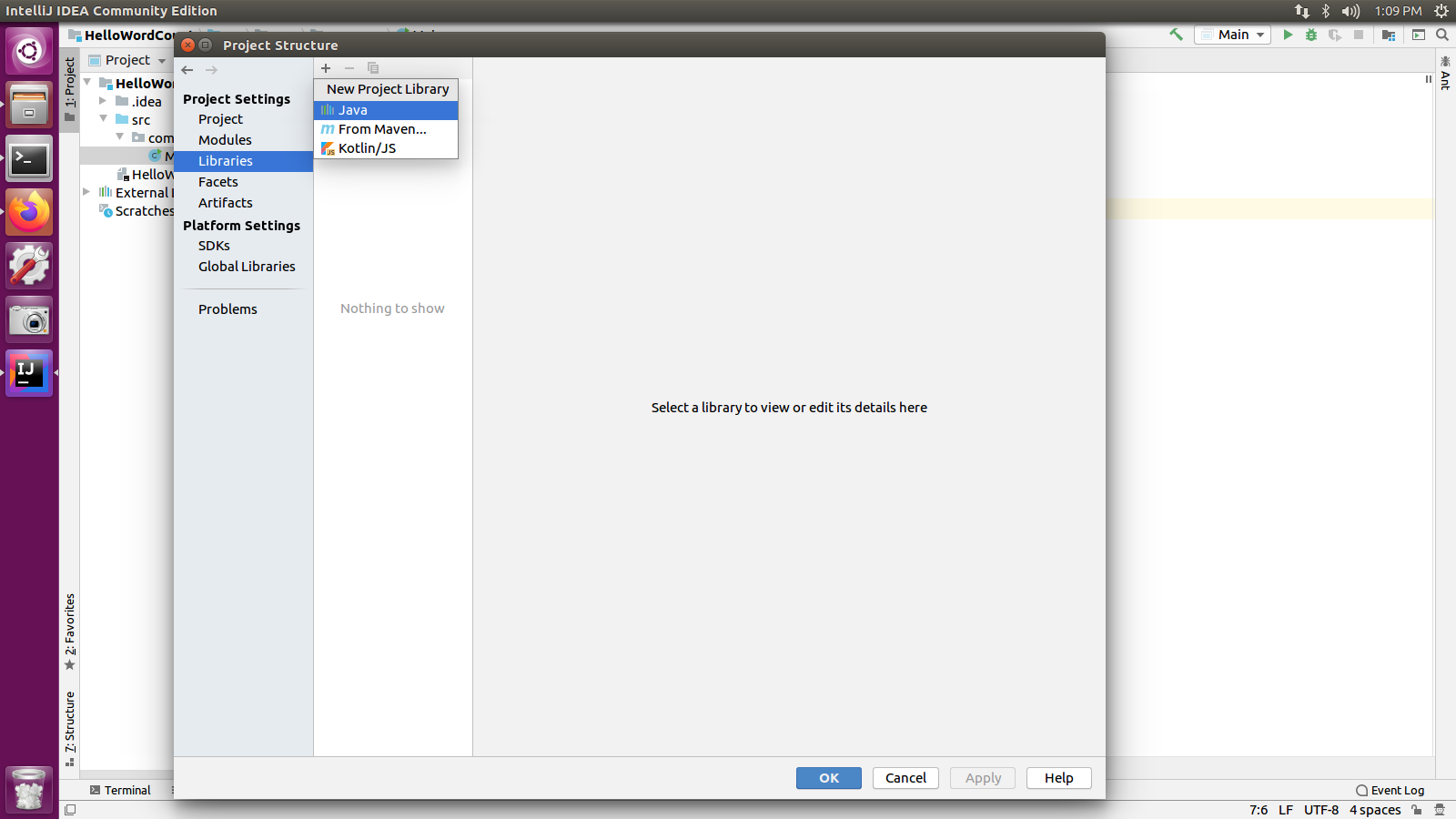
# Hadoop MapReduce API Demo Program

1.launch iDea to create a empty project **HelloWordCount**

2. [File]→[Project Structure]



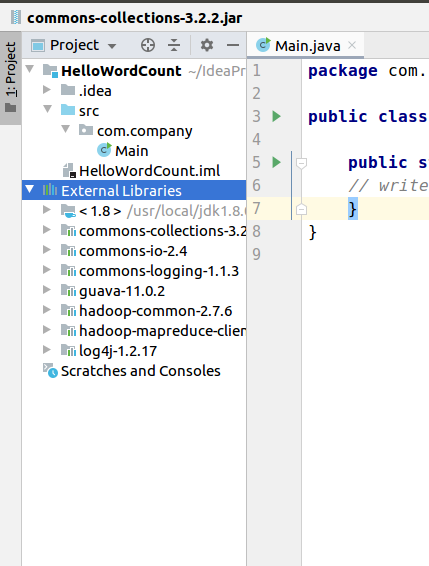
3. In Project Structure Panel->Libraries->Java



4.Add jars one by one:

|  |  |  |
| --- | --- | --- |
| /usr/local/hadoop/share/hadoop/common/hadoop-common-2.7.6.jar | | |
| /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-client-core-2.7.6.jar | | | | | |
| /usr/local/hadoop/share/hadoop/common/lib/commons-logging-1.1.3.jar | | | |
| /usr/local/hadoop/share/hadoop/common/lib/guava-11.0.2.jar | |
| /usr/local/hadoop/share/hadoop/common/lib/commons-collections-3.2.2.jar | | | | |
| /usr/local/hadoop/share/hadoop/common/lib/log4j-1.2.17.jar |

/usr/local/hadoop/share/hadoop/common/lib/commons-io.2.4.jar

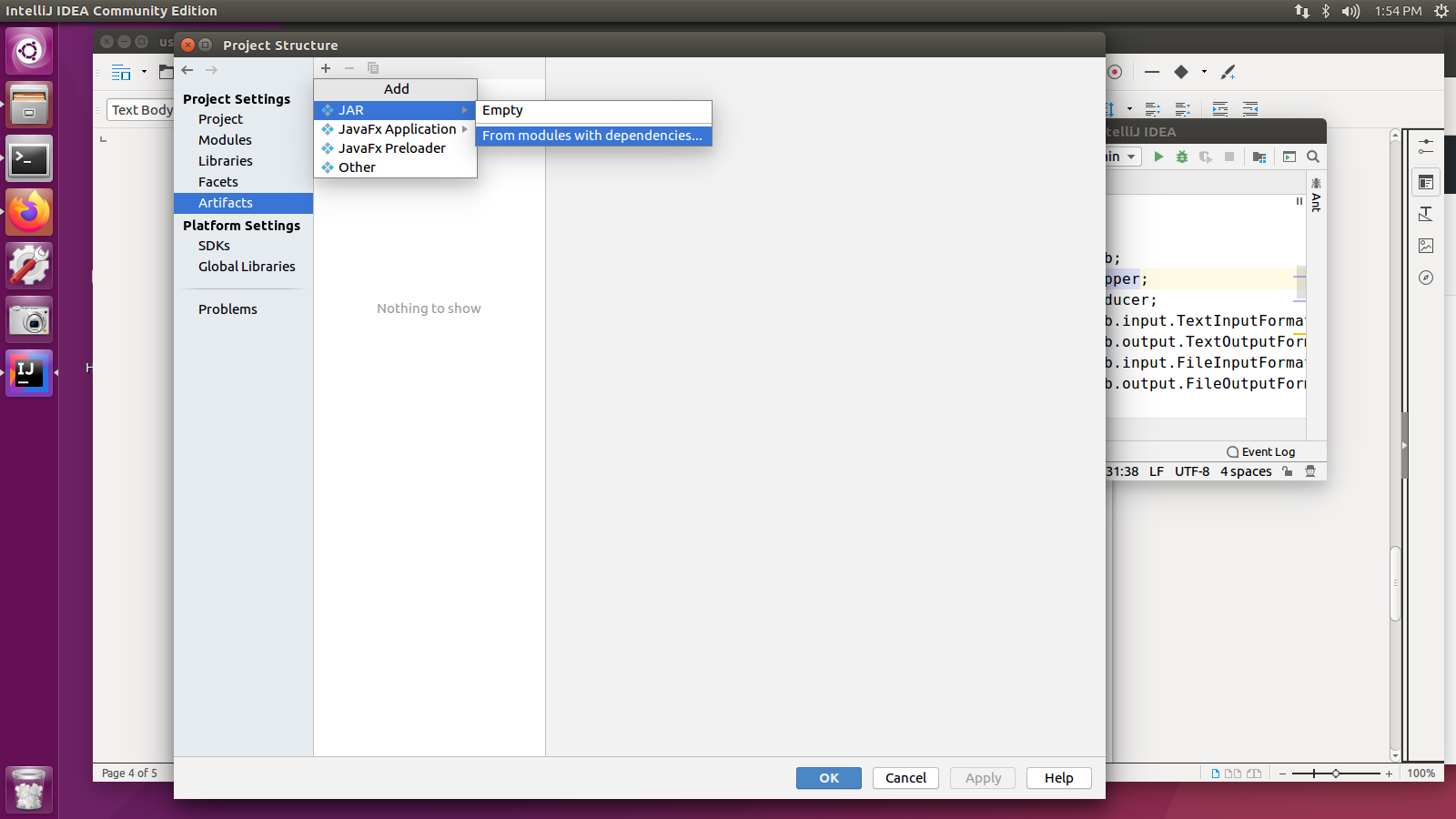


5.Write the code: copy the codes from github page.

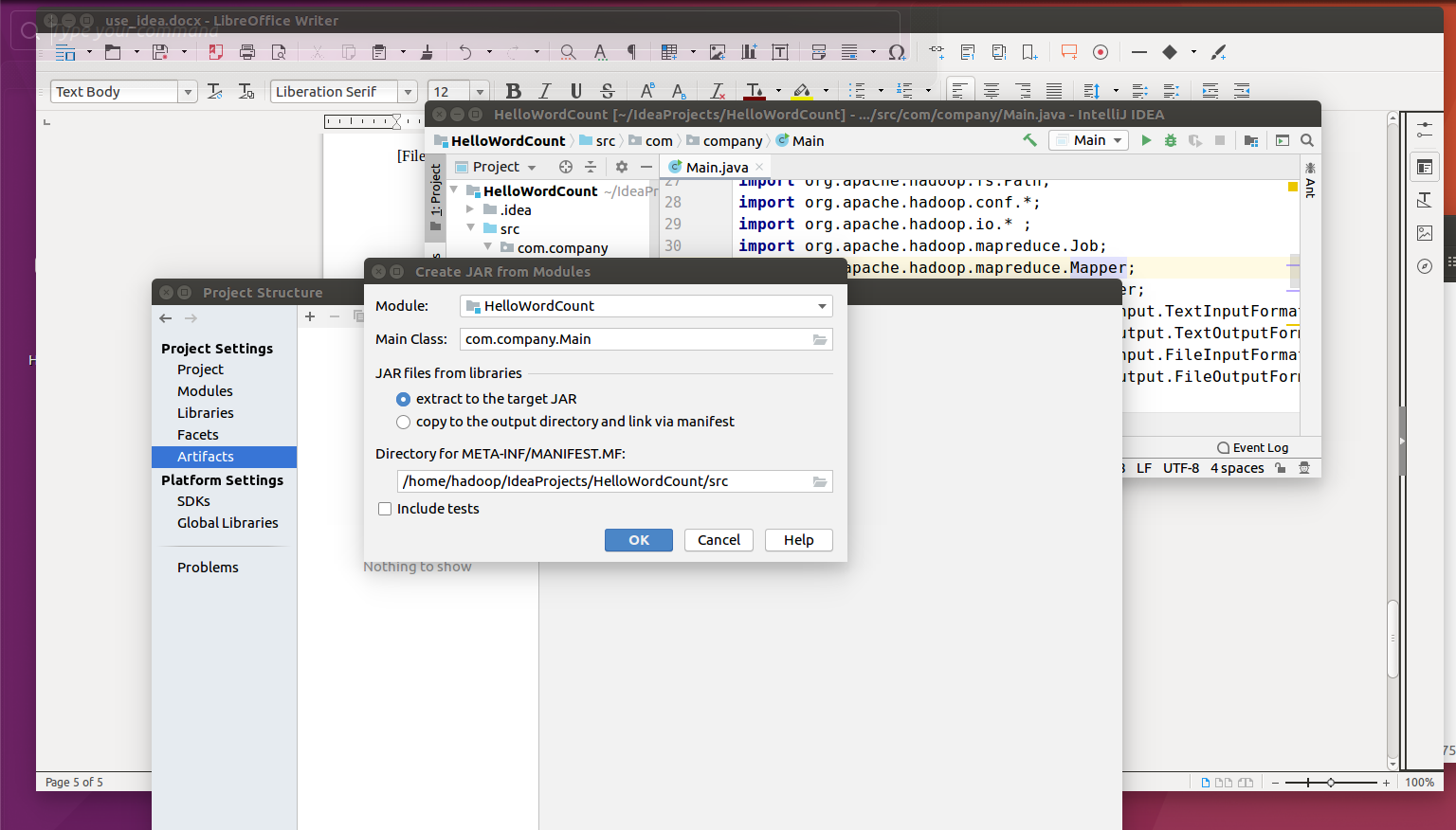
<https://github.com/wangfeng1981/hadoop-toycluster/blob/master/HelloWordCount.java>

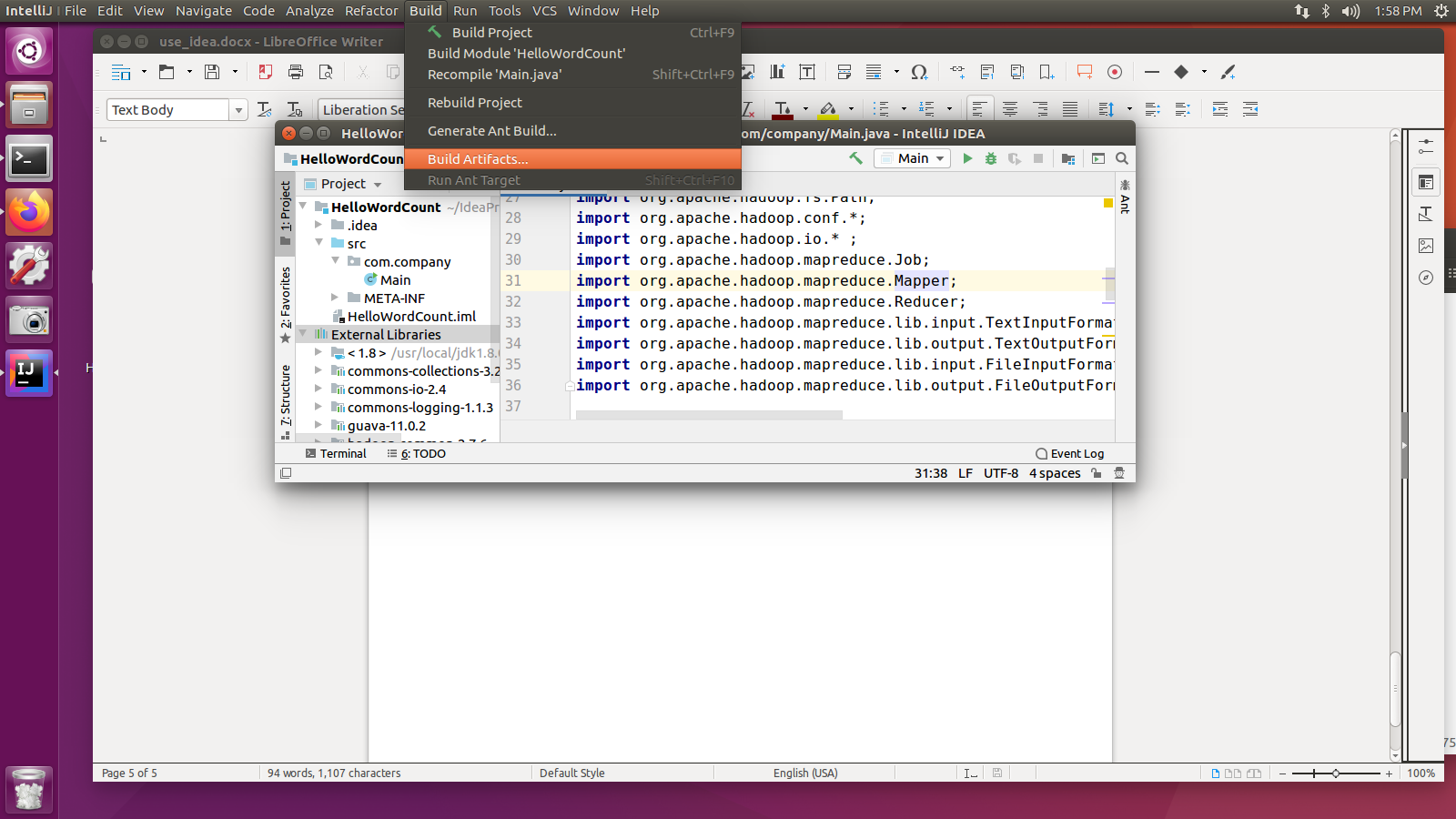
6.build the Jar.

[File]→[Project Structure]→[Artifact]



7.Pick Main Class And OK

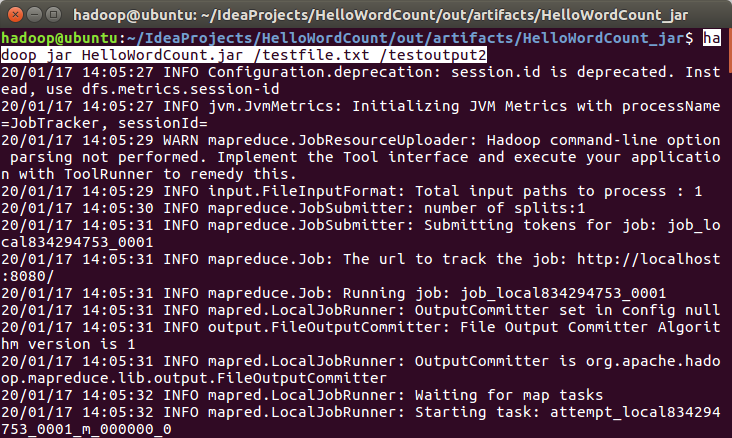
8. [Build]→[Build Artifacts]→[HelloWordcount.jar]->[Build]



9.build OK

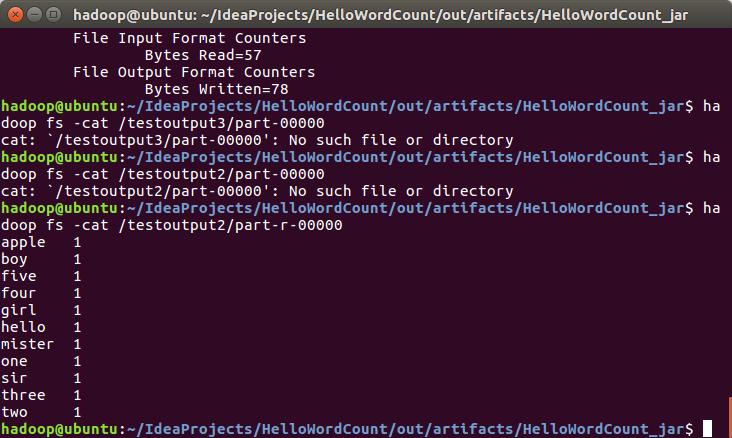
Goto the jar directory. Start a shell to test the jar.

hadoop jar HelloWordCount.jar /testfile.txt /testoutputXXX



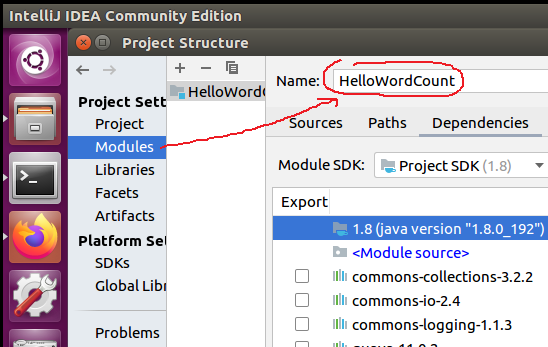
check the results:

hadoop fs -cat /testoutput2/part-r-00000

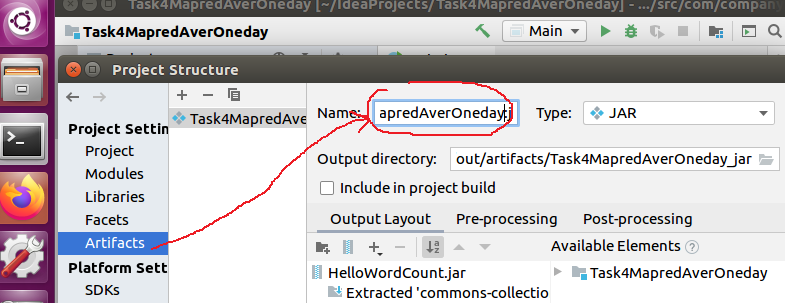


# **复制一个项目开始**

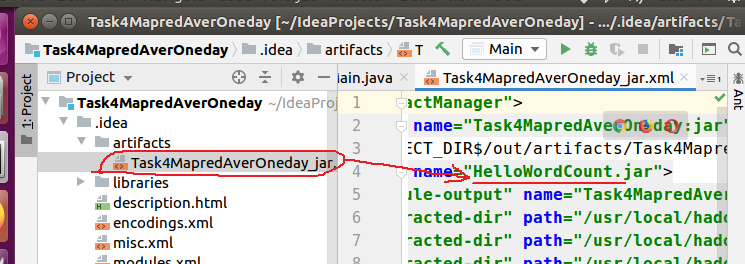
1. 复制一个类似的项目目录，重新命名该目录
2. 使用Idea打开Project，定位到这个新的目录
3. File-》Project Structure-》Modules 修改Module名称



1. File-》Project Structure-》Artifacts 修改Jar名称



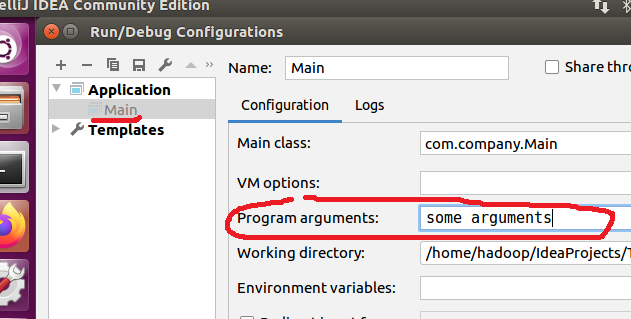
1. 修改jar.xml中的名称



1. 可以生成新的名称jar文件了。

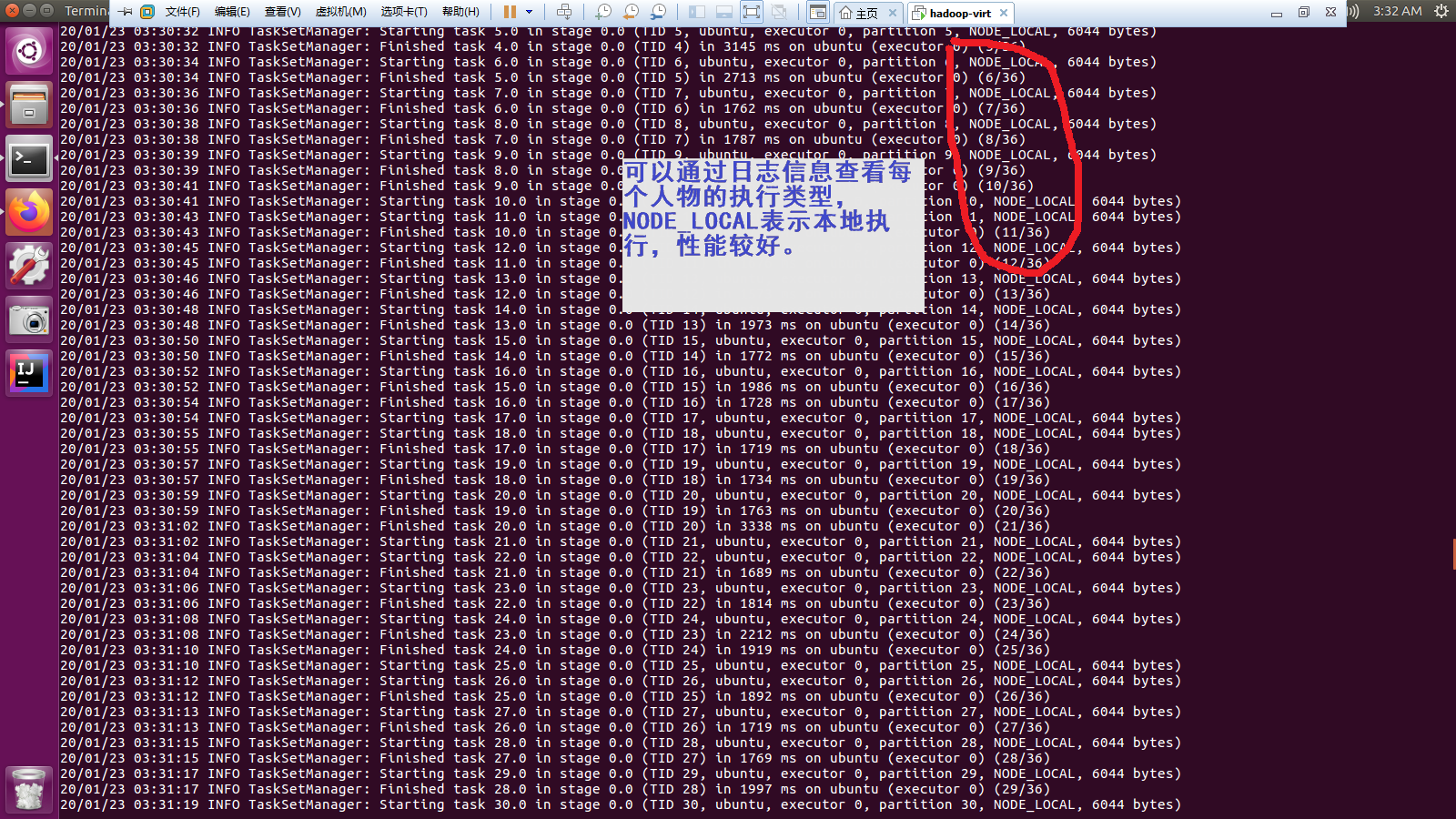
# 配置java程序启动参数

菜单-》Run-》configurations



# Spark日志中的关键内容

可以在spark运行日志（控制台中输出）查看每个任务（InputSplit）的运行类型：



比如红圈处的日志信息显示任务在节点本地运行，这样性能最优。

(done)