## Hadoop 分析统计学生考试成绩

1. 开发环境：
   1. Win 10 64位 / macOS High Sierra 10.13.3
   2. Java 1.8.0\_162 64位
   3. Intellij IDEA 2017.3
   4. Maven （Intellij 已经集成的）
   5. MySQL 5.7.20

为保证正确运行，请尽量使用相同的开发环境，Java 9 的支持还有些问题。理论上支持 Linux 等其他环境。

1. 项目结构（Maven 构建，可以直接导入）：

导入项目时，在 Intellij IDEA 中直接选择打开，如果是 Eclipse，可以选择 Import maven project。之后会自动下载配置，需要等待较长一段时间，等待下载完成。

1. src 文件夹，项目代码，下面详细解释
2. target 文件夹，Maven 生成的项目 class 文件
3. pox.xml，Maven 的配置文件，自动下载各种 lib
4. hadoop.iml，Intellij 中，该项目的配置文件
5. hadoop.log 日志记录文件
6. database.sql 是建立数据库的脚本
7. input 为格式化后的 Hadoop 的输入文件，output 为输出，raw 中存放原始数据
8. 代码文件说明

介绍一下主要的项目代码文件

1. src/main/java/com/example/hadoop/db 中，DBTools.java 是关于数据库的一些操作，DBEntity.java 定义了一个数据库实体的类，方便数据相关的操作
2. src/main/java/com/example/hadoop/mapreduce 中，是 MapReduce Job 的实现，WordCount.java 是一个测试文件，与本项目无关。GradeMapper.java 是 Mapper 的设计，GradeReducer.java 是 Reducer 的设计，Grade.java 是运行这个 MapReduce Task 的入口
3. src/main/java/com/example/controllers 中，是基于 Spring Boot + Thymeleaf 渲染实现的 Web 项目，HomeController.java 是处理用户的查询请求的后台逻辑，CompareController.java 是处理用户进行成绩对比的后台逻辑
4. src/main/java/com/example/preprocess 中，完成了从 xml 文件转化成方便 Hadoop 处理的 csv 文件的工作，属于预处理阶段
5. HadoopApplication.java是整个项目运行的入口，具体使用方式可以看里面的注释，运行 Web 项目或只是本地测试。

一下是一些其他的资源文件：

1. src/main/resources一些资源文件、项目配置和 Web 项目的静态资源
   1. application.properties 是 Spring Boot 项目的相关配置和一些常量设置
   2. logback-spring.xml 是日志记录配置
   3. static 文件夹中是网页相关的 css 样式，js 脚本和一些图片 img 资源
   4. templates 文件夹中是通过 thymeleaf 渲染的 HTML 网页模板
2. 程序运行方式
   1. 准备工作

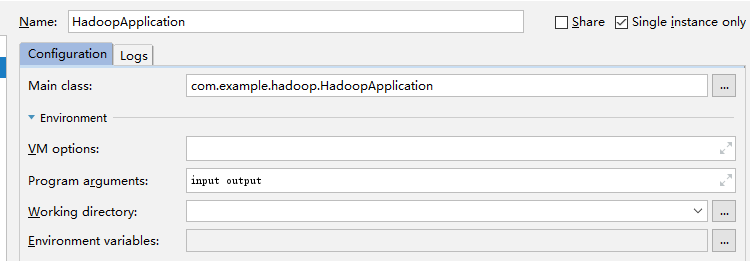
在 Intellij IDEA 或者是 (My)Eclipse 中导入该 Maven 项目，会根据 pom.xml 自动下载依赖包，时间可能有些长，如果遇到网络状况不佳，自行百度一下怎么在这些开发环境中更换 maven 源为国内的源，等待加载完毕

此外，需要安装配置 MySQL，根据提供的 SQL 脚本创建数据库和表

* 1. 运行 HadoopApplication 本地测试程序

根据注释说明，按步骤运行

* 首先是预处理，从 raw 中读取 xls 文件，按照 csv 格式写入 input 中
* 使用 Hadoop 对成绩进行统计，这里需要设置命令行参数为 input 和 output，前者是 csv 文件的路径，后者是 Hadoop 任务结束后，将结果输出的路径



* 1. 运行网页程序

只需要执行 SpringApplication.run(HadoopApplication.class, args); 当项目构建好后，会有如下类似的提示，实际可能略有偏差，

`com.example.spam.HadoopApplication: Started HadoopApplication in 2.879 seconds (JVM running for 4.537) `

就可以从浏览器访问 localhost:8080 进行操作