**课程名称大数据实训**

**实验2-hadoop 大数据开发实战**

课程名称：大数据实训

实验学时： 40 学时

综合性、设计性实验：□🗹是

学生人数：15

任务编制人：课程负责人 刘喆

# 实验目的

* 1. 掌握 Hadoop 上 java 开发的方法
  2. 熟练使用 maven/shade 这种便捷的打包方式
  3. 统一组内代码风格，使用 checkstyle 来严格规范代码风格
  4. 掌握 MapReduce 的工作原理，会写复杂的 MapRedce 作业

# 评分办法

1. 分小组在指定的时间之前完成任务，按任务的完成情况，给小组打分
2. 按每个人对于实验的贡献，包括但不限于 代码贡献/文档贡献/issue 贡献，给每个人打分
3. 每个人本次实验的最终成绩，为上面两个数字的乘积

# 实验要求

完成 “平均在线时间” 的代码编写, MR

- 成果

- 由 PM 同学汇总所有的思考和讨论， think.md

- code review 意见及修改

- 临时性的一些问题（比如机器异常等）放到 issue

- 本组的 checkstyle 文件，多少修改一些，写明修改的原因， style.xml

- 随机生成 50 亿条日志的程序放 random 目录里， 每人一个

- userId 800 万个，模拟用户登录登出，但是一定要时间上是有序的

- MR 源代码 online\_mr 目录,每个人在这个目录里一个自己写的 mvn 工程

- 所生成的日志文件的前 20 行截图

- 运行过程截图 mr\_process.png

- 运行结果的截图 mr\_result.png

# 实验环境

1. 硬件环境
2. 软件环境

# 实验内容

1. 自己贡献的源码或文档的 url
2. 自己提的 issue，附 github 连接即可
3. 其它没有记录的问题（应该都在 issue 里了，如果有落下的，写这里）