**课程名称大数据实训**

**实验2-大数据流处理技术基础**

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. 掌握大数据流处理的关键组件：kafka 和 flume
  2. 了解 java 内部的工作原理，特别是内存模型
  3. 掌握常用的 java 程序查错工具 jstat/jstack/jmap/jhat 等等

# 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

- PM

- 各种问题及解决的 issue

- RD

- 完成 kafka/flume 集群的搭建, 给出节点和角色分布表 wiw.md

- kafka 的 jstack 截图 kafka\_jstack.png, 指出哪些是正在运行的进程

- flume 的 jstat 截图 flume\_jstat.png, 说明 GC 是否合理，原因是什么

- flume 的 jmap histo 截图 flume\_histo.png, 说明是否有问题

- 完成 20 万条日志的写入，项目名 generater（每人一个）

- 自己写一个内存泄漏的 bug 程序，并用所学到的知识分析一下。

# 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）