

# 工实验个人总结

SY1506412 于思民

## 1. 项目简介

Apache Spark 是一种快速、通用、可扩展的大数据分析引擎。它是不断壮大的大数据分析解决方案家族中备受关注的明星成分，为分布式数据集的处理提供了一个有效框架，并以高效的方式处理分布式数据集。Spark 集批处理、实时流处理、交互式查询与图计算于一体，避免了多种运算场景下需要部署不同集群带来的资源浪费。

## 2. 项目角色及任务

由于我个人相对于对分布式计算框架较为熟悉，所以由我担任本组的组长，协调组员共同完成各次实验的任务。

我在本学期的实验中的具体任务主要由以下几点：

- 搜集实验所需的资料
- 根据组员能力分配任务
- 组织组员开小组会议进行沟通、交流
- 撰写需求文档、测试文档、测试报告等文档
- 制作课堂各次实验报告的 PPT
- 搭建实验运行环境、做实验

## 3. 项目中的问题及解决办法

由于所有组员对整个软件工程过程不熟悉，导致在各次实验中，撰写的文档从内容上出现各类问题。主要有以下几点：

- 项目需求用户不明确
- 项目需求用例描述不清楚

- 项目测试用例与需求用例不一致
- 项目测试用例覆盖度不充分

虽然在实验过程中遇到了以上的任务，但是经过刘老师、任老师以及各位同学的课堂评审与论坛评审意见指导，各类问题得到的一一解决。

还有一大类问题就是，由于对于 Spark 系统不熟悉，在使用 Spark 的过程中也出现以下几类问题：

- 集群节点连接不上
- 应用程序读取中文文件乱码
- 系统配置错误

虽然在配置、使用系统做实验的过程中出现各类问题，但是经过与组员之间的沟通、交流以及从网上获取帮助，都一一得到解决。

## 4. 个人心得体会

通过本学期的实验课程，我主要有以下两方面的收获。

第一方面，就是软件工程方面的收获。由于研究生未曾选修《高等软件工程》课程，导致在软件工程实验中，对于理解需求有很多不足之处。但是经过本学期的实验课程，使得我在接触一个新的系统时，应该以需求的角度、使用者的角度来审视系统，这样才能更快、更准的理解系统。还有一点就是，软件工程不是一个人的工程，而是一个团队的工程。本次实验课同时让我体会到了一个人的强大不算什么，一个团队的齐心协力才是真正的强大。这也验证了“众人拾柴火焰高”的理念。

第二方面，就是对 Spark 系统更加熟悉。Spark 是一个很复杂的系统，通过本次实验课程，让我对 Spark 有了新的认识。虽然，可能我们对 Spark 的还只是九牛一毛，但是经过此次课程学习，我更加懂得以软件工程的角度来审视 Spark，从而牵“一毛”而动全身。虽然此次课程结束了，但是软件工程却不会结束，对于 Spark 的学习不会结束，二者才是刚刚开始。

## 5. 对本学期实验的总结及建议

本学期的实验课程是一门注重团队协作、着重软件工程实际的课程。本次实验课程不仅有老师与同学的交流、也有小组间、小组内部的交流。这种交流、协作是软件工程作为一门

工程实践所必须，而且十分重要。

但是这里我也想给出一些个人的建议，希望能对课程的提升有所帮助。我认为，选取该实验课程的同学必须要有一定的理论知识。希望将《高等软件工程》、《软件体系结构》等课程与本实验课程结合起来，这样同学们才能理论指导实践。

谢谢刘老师、任老师在各次实验中的悉心指导，祝工作顺利、生活愉快！