**对于项目计划书与项目计划的问题清单——第4次整理**

# 基本信息

|  |  |
| --- | --- |
| 项目名称 | 基于Apache Spark的分布式数据处理研究 |
| 项目组成员 | 孟翰、苏若、李璇、孙敏芳 |
| 项目简介 | Spark是UC Berkeley AMP lab所开源的类Hadoop MapReduce的通用的并行计算框架，Spark基于map reduce算法实现的分布式计算，拥有Hadoop MapReduce所具有的优点；但不同于MapReduce的是JobIMG_256中间输出和结果可以保存在内存中，从而不再需要读写HDFS，因此Spark能更好地适用于数据挖掘与机器学习等需要迭代的map reduce的算法。 |
| 评审对象 | 《需求规格说明书》 《项目进度报告》 |
| 评审方式 | 组间评审、老师评审、组内讨论 |
| 评审员 | 刘超老师、任健老师、全体同学 |

# 评审问题及其处理

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 评审对象 | 评审意见 | 处理意见 | 工作跟进 |
| 1 | 项目进度报告 | 对系统实施方案有何计划? | 接受 | 收集大量数据，以进行系统的功能实现和性能分析 |