**对于项目计划书与项目计划的问题清单——第3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》 《项目计划》 |
| 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 | 需求规格说明书 | 缺少用例图。 | 接受 | 利用UML工具绘制用例图。 |
| 2 | 需求规格说明书 | 文档中的一些专有名词未在数据字典中给出。 | 接受 | 本组会对数据字典进行适当的补充和完善。 |
| 3 | 需求规格说明书 | 缺少针对Storage模块需求的RUCM用例描述。 | 接受 | 功能性需求中的RUCM图涵盖了RDD模块和Storage模块，为了结构更清晰，独立出Storage模块的RUCM图。 |
| 4 | 需求规格说明书 | 系统改进点不明确。 | 接受 | 本组首先会对设定的系统功能模块进行研究，在此过程中，发现切实的改进点。经过资料查询和研讨，我们暂定为四个改进方向：  1）JVM的overhead过大。  2）不同的spark app之间缺乏有效的共享内存机制。  （Tungsten，Tachyon）  3）集群有时会发生崩溃。  4）负载不均衡 |
| 5 | 需求规格说明书 | 建议添加修改记录 | 接受 | 通过征求老师和同学的意见，制定一个适用的统一的记录规范。 |
| 6 | 需求规格说明书 | 涉及到了一些设计细节 | 接受 | 更正对需求规格说明书的认识，并重新审查修改了相关内容。 |
| 7 | 需求规格说明书 | “用户可自主选择容错方式”这个算是功能上的需求，还是算是非功能的需求呢？ | 解释 | 在文档中，已经明确标明其为功能性需求。 |
| 8 | 需求规格说明书 | 参考文献是不是应该放在最后面比较合理？ | 接受 | 将参考文献放置为文档最后一节。 |
| 9 | 需求规格说明书 | 只给出了RUCM图，却未给出一些必要的用例描述。 | 解释 | RUCM图本身便是对用例的描述。 |
| 10 | 需求规格说明书 | 系统功能重点不够突出? | 解释 | 本课程设计抽取类RDD及Storage作为研究切入点，并以文档中所介绍的功能性需求和非功能性需求作为系统功能重点进行研究。 |