**第一次需求初审意见记录**

# 基本信息

|  |  |
| --- | --- |
| **项目名称** | 流式实时分布式计算框架Spark的研究与应用 |
| **项目组成员** | 于思民、阳艳红、王铖成、武一杰 |
| **项目简介** | 我们的关注点是Apache Spark这一款流式实时分布式计算框架的研究与应用，运用软件工程的方法对Spark进行分析，并在此基础上，将Spark扩展应用。 |
| **评审时间** | 2016.3.18 晚上18：00-21：00 |
| **评审地点** | A209 |
| **评审对象** | 《流式实时分布式计算框架Spark的研究与应用》 |
| **评审方式** | 会议审查 |
| **评审员** | 刘超 教授  任健 教师  全体选课同学 |

# 评审意见

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 评审对象 | 评审意见 | 处理建议 | 小组意见 |
| 1 | 《流式实时分布式计算框架Spark的研究与应用》 | 研究每个模块在时间上是否允许？ | Spark主要分为以下几个核心模块：Schedular,Deploy,Executor,Storage.我们组每个人负责学习1-2块。 |  |
| 2 | 《流式实时分布式计算框架Spark的研究与应用》 | Spark与MapReduce的区别在于内存，内存不够怎么办？ | Spark实时的处理一部分数据。针对数据流分片，将常用的数据分片缓存，在某些情况下中间结果溢出到外存。 |  |
| 3 | 《流式实时分布式计算框架Spark的研究与应用》 | 还有另一部分数据未处理怎么办？ | 储存中间结果，进行统一处理。 |  |
| 4 | 《流式实时分布式计算框架Spark的研究与应用》 | 与A组选题相似，区别在哪里，要体现出自己小组的特点 | 侧重点不一样。从Spark开源项目出发，重新编写起需求规格说明书、测试需求规格说明书，并根据需求规格说明书与测试需求规格说明书对Spark进行模块化测试，并在此基础上拓展应用。同时与A组同学进行交流，将工作重点区分开。 | 接受 |
| 5 | 《流式实时分布式计算框架Spark的研究与应用》 | 报告环节时面向全体同学，问答环节时仔细聆听虚心接受。 |  | 接受 |