**迭代评估报告**

　　　　　　　　　　　　　　　　　　评估日期：2020.11.25

|  |  |  |  |
| --- | --- | --- | --- |
| 组号 | 3 | 项目名称 | Hive反模式的检测与修复 |
| 迭代名称 | Sprint2 | 实际起止日期 | 2020.10.29 ~ 2020.11.25 |
| 任务达成情况：  阅读了反模式测试相关论文，总结了前沿技术（参数调优论文调研.docx），使用机器学习的方法，预测给定输入情况下任务巡行所需的时间，已训练出初步模型，还需要进一步调优。  完成了反模式静态检测工具的改进版本的开发与测试。  完成了第二次迭代计划中的任务，整个开发过程中每个成员较好的遵守了代码规范，关注细节，整体质量较为良好，但各部分的合作需要进一步改进。 | | | |
| 评审/测试的结果：  新增了3个Anti-pattern，寻找了Hive anti-pattern 3个可使用的sql语句数据集，对以实现的静态检测规则新增了12条测试用例，其中2条用例未通过测试，针对未通过测试的用例进行了修复。 | | | |
| 问题、变更和返工：   1. 数据集问题：希望能获取1M条数量级别的sql语句，但目前找到的数据集数量还不够，但作为demo级别的测试，已经可以满足最低要求。 2. 动态监测数据倾斜部分的复现和修复比较难实现，对集群的配置有一定要求，已经做了测试与完善。 3. 前端部分由于没有绘制原型，组员间暂时没有讨论前端如何展示，需要尽快补全原型，开发出简易的前端展示网页。 | | | |
| 经验和教训：   1. 任务分配和执行都需要组员间充分沟通。 2. 伪分布式集群的配置项设置对hive执行效率的影响。 3. 通过论文阅读浏览前沿技术，并将机器学习方法用在反模式检测上。 | | | |