### 计划实施方案

### 基于数据分析的系统功能实现与性能分析

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 任务编号 | 任务 | 输入文档或硬件要求 | 输出文档或阶段性成果 | 执行日期 | 任务负责人 |
| 1 | 收集数据资料 | a.需求规格说明书  b.项目进展报告 | Gowalla原始用户签到数据集 |  |  |
| 2 | 整合数据并进行结构解析 | Gowalla原始用户签到数据集 | a. Gowalla用户签到数据集v1.0  b.相关数据说明文档 |  |  |
| 3 | Spark环境搭建 | Linux虚拟机、  Ubuntu 15.2  Spark1.6.1 | a.Linux系统  b.Spark开发环境 |  |  |
| 4 | Spark 开发环境测试 | 1. Spark开发环境 2. 测试用例 | 无 |  |  |
| 5 | 确定待实现的系统功能模块 | 1. 需求规格说明书 2. Spark源码 | Spark源码说明文(RDD模块和Storage模块) |  |  |
| 6(功能实现) | 确定每个UserID(用户)的签到总次数 | Gowalla用户签到数据集v1.0 | 1. Spark下的运行结果记录A 2. Spark下总执行时间统计表 |  |  |
| 1. java下的运行结果记录 2. java下总执行时间 |  |  |
| 以相同LocationID(签到地点)为分组依据，统计各分组下的签到信息 | Gowalla用户签到数据集v1.0 | 1. Spark下的运行结果记录B 2. 更新后的Spark下总执行时间统计表 |  |  |
| 推断出各用户间的亲密关系 | Spark下  的运行结果记录B | 1、Spark下  的运行结果文档C  2、更新后的Spark下总执行时间统计表 |  |  |
| 7 | 整合本阶段所有输出文档 | 上阶段输出文档及阶段性成果 | 1. Spark运行结果记录汇总文档 2. Spark性能分析表 |  |  |