**分布式爬虫系统项目管理**



**1.绘制项目的活动网络图**

分布式爬虫系统的活动网络图如图1.1所示。

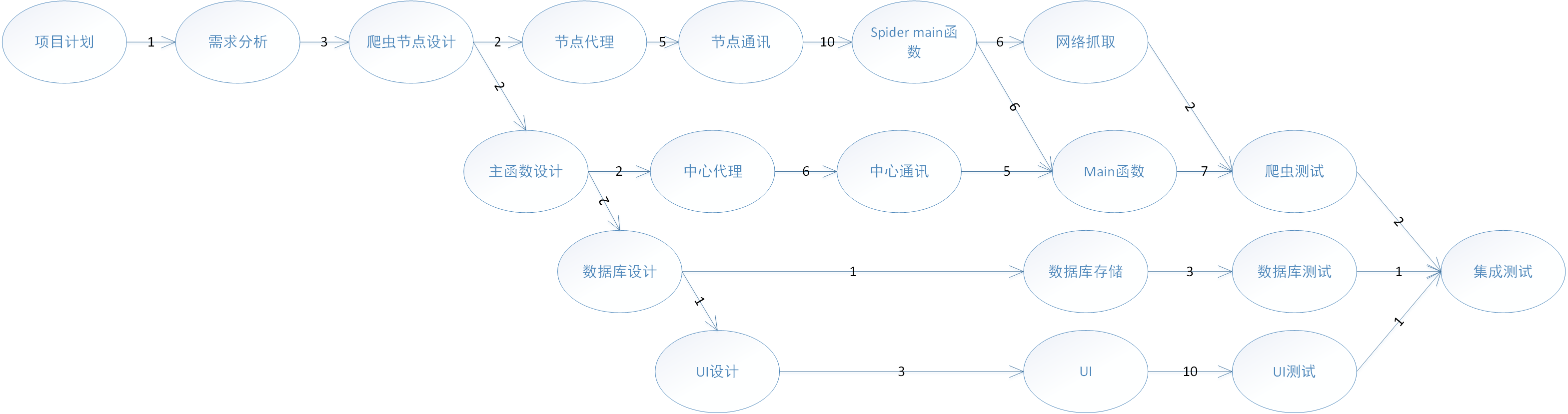


图1.1 活动网络图

图1.1完整的描述了项目设计与执行的全部流程，以项目计划为开始节点，集成测试为结束节点，分为爬虫节点、主函数、数据库和UI四个部分分别执行。每部分大体由设计、实现、测试三个小模块组成，最后经过集成测试完成对项目的总体测试。每一个活动预期需要耗费的时间都作出了标注。

**2.利用功能点法度量软件（活动）的工作量**

功能点计算方法 ：

FP = UFC×TCF = UFC × (0.65 + 0.01×Fi)

UFC (Unadjusted Function Component) : 未调整功能点计数 5个信息量的“加权和” 功能点计数, 5个信息量的 加权和

TCF (Technical Complexity Factor): 技术复杂度因子

Fi: 14个因素的“复杂性调节值” (i =1..14)

0.65, 0.01：经验常数

UFC：

|  |  |  |  |
| --- | --- | --- | --- |
| 输入文件 | 2 | URL  分析方向 | 2\*3=6 |
| 输出文件 | 3 | 分析表  聚类分析图  参数预测图 | 1\*4+2\*5=14 |
| 查询 | 1 | 查询数据库 | 1\*3=3 |
| 内部文件 | 2 | 网站信息  已爬取URL表 | 2\*7=14 |
| 外部文件 | 2 | 安全信息  政府监管信息 | 1\*7+1\*10=17 |

表2.1 UFC预测表

由表2.1可知：UFC=6+14+3+14+17=54

TCP：

|  |  |
| --- | --- |
| 系统特征 | 复杂度 |
| 数据通讯 | 2 |
| 数据分布处理 | 4 |
| 系统性能 | 5 |
| 使用配置 | 2 |
| 事务处理频率 | 2 |
| 在线数据输入 | 4 |
| 操作便利 | 1 |
| 在线更新 | 1 |
| 处理复杂度 | 2 |
| 复用性 | 2 |
| 安装难易 | 1 |
| 运行维护 | 1 |
| 多站点支持 | 1 |
| 变更支持 | 4 |

表2.2 复杂度调整表

由表2.2可知：

TCF=0.65+0.01\*（1\*5+3\*4+5\*2+5\*1）=0.87

经过复杂度调整后，可计算出调整后的功能点：

FP=UFC\*TCF=46.98

1. **利用制定项目的进度计划**

根据项目活动网络图及经复杂度调整后计算出的项目功能点，设计出图3.1所示的项目进度计划，项目从2017年4月27日开始，到6月26日完成所有测试并进入为期60个工作日的维护阶段，预计在9月28日完成整个项目。

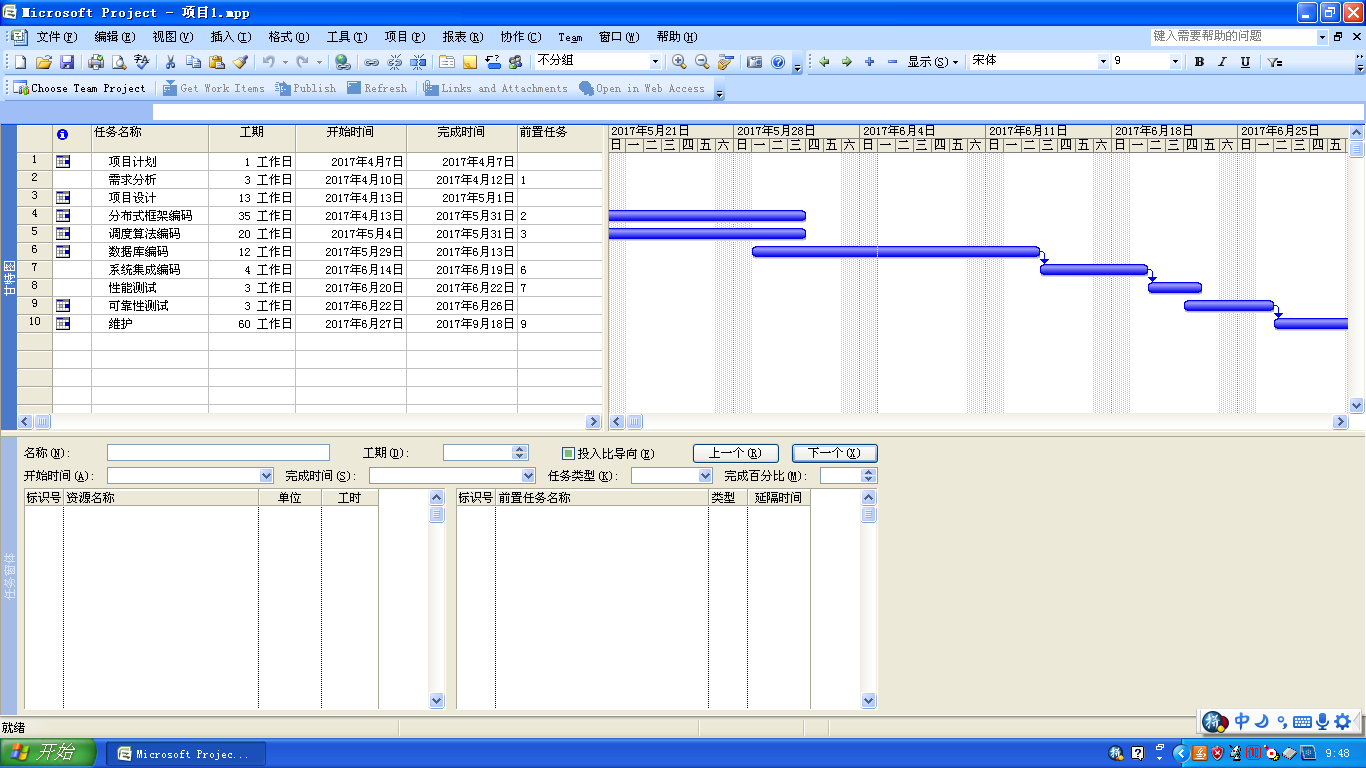


图3.1 项目进度计划

1. **计算关键路径**

根据项目进度计划，绘制了图4.1所示的关键路径图，为每个节点分配了各自完成所需的最大工作日期限，通过CPM关键路径算法得出了本项目的关键路径，即项目计划、需求分析、爬虫节点设计、结点代理、节点通讯、Spider main函数、Main函数、爬虫测试、集成测试，关键路径长度为36工作日。

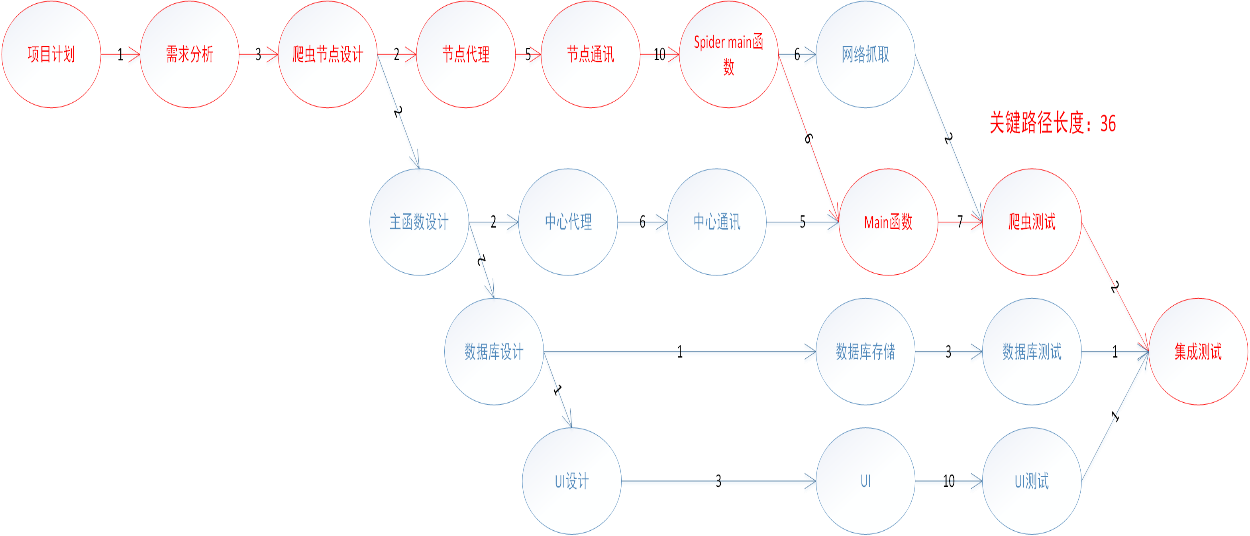


图4.1 关键路径