**数字货币大数据分析平台**

# 项目开发计划文档

## 项目名称：数字货币大数据分析平台

**项目组：数字货币大数据分析平台开发组**

**项目成员：肖文斌（项目经理）、文杰、陈楚凡、景南瑞、吴后望、陈修鸿**

**2021 年 07 月 05 日**



**目 录**

[1. 引言 1](#_Toc77598433)

[1.1 文档编写目的 1](#_Toc77598434)

[1.2 术语与缩略语 1](#_Toc77598435)

[2. 项目概述 1](#_Toc77598436)

[2.1 项目背景 1](#_Toc77598437)

[2.2 项目概况 2](#_Toc77598438)

[2.4 交付的产品 2](#_Toc77598439)

[3. 开发成员 3](#_Toc77598440)

[4. 项目开发计划 3](#_Toc77598441)

[4.1 需求确定阶段 3](#_Toc77598442)

[4.2 数据爬取阶段 3](#_Toc77598443)

[4.3 数据分析阶段 4](#_Toc77598444)

[4.4 数据可视化阶段 4](#_Toc77598445)

[5. 项目所需技能和培训计划 4](#_Toc77598446)

[5.1 项目所需技能 4](#_Toc77598447)

[5.2 项目培训计划 4](#_Toc77598448)

[6. 开发计划 5](#_Toc77598449)

[6.1 项目监控记录 5](#_Toc77598450)

[6.2 风险管理计划 5](#_Toc77598451)

[6.3 度量与分析计划 5](#_Toc77598452)

[6.4 质量保证计划 6](#_Toc77598453)

[6.5 配置管理计划 6](#_Toc77598454)

[7. 工作环境 7](#_Toc77598455)

[7.1 开发环境 7](#_Toc77598456)

[7.2 测试环境 7](#_Toc77598457)

|  |  |
| --- | --- |
| **文档信息** | |
| **标题** | 项目开发计划文档 Project Development Plan Document |
| **作者** | 景南瑞 |
| **创建日期** | 2021.07.05 |
| **更新日期** | 2021.07.19 |
| **版本** | V1.1 |
| **部门名称** | 数字货币大数据分析平台项目开发组-重庆大学实训大数据二组 |

|  |  |  |  |
| --- | --- | --- | --- |
| **文档更新记录** | | | |
| **版本号** | **更新内容** | **操作日期** | **操作人员** |
| **1.0** | 初步拟定文档 | 2021.07.05 | 景南瑞 |
| **1.1** | 按照时间顺序更新项目监控记录 | 2021.07.19 | 景南瑞 |

**项目开发计划文档**

**Project Development Plan Document**

# 引言

## 文档编写目的

制定本文档的目的，包括工作任务分解、估算工作产品和活动的规模、工作量、成本以及所需资源；制定开发计划、风险管理计划、质量保证计划、配置管理计划、度量计划、测试计划等，经评审、批准后作为项目实施和完成的依据，项目跟踪和监控的基础。

## 术语与缩略语

①PyCharm：一种 Python IDE，带有一整套可以帮助用户在使用 Python

语言开发时提高其效率的工具;

②vscode：全称Visual Studio Code，一种免费跨平台的编程工具，支持各种自定义及丰富的插件;

③MySQL：一个关系型数据库管理系统，是最流行的关系型数据库管理系统之一;

④Git：一个开源的分布式版本控制系统，用以有效、高速的处理从很小到非常大的项目版本管理。支持克隆/下载等团队合作功能;

⑤Apache Spark：是专为大规模数据处理而设计的快速通用的计算引擎。

# 项目概述

本项目为基于 spark 计算引擎的数据货币大数据分析平台，数十万级数据经由框架存储、整合、清洗并上传前端界面实现数据可视化的目标，通过前端可视化给用户提供以echart实现的动态数据，并提供后台管理系统给需求为文本化数据的用户，除此之外还提供了有关数字货币热点新闻的抓取，以此观察热点事件与货币涨跌趋势的某种关联。

## 项目背景

数字货币概念自2008年金融危机中诞生，经过十余年发展，在2015年、2020年迎来爆炸性增长，参与数字货币挖掘的个人、企业人数激增。面对这样一个新兴且热门的事物，掌握其具体的发展数据，对其进行大数据分析是有必要的。本项目开发的软件旨在基于互联网中数字货币交易所数据，对其中热门数字货币的历史数据进行收集，对其中的数字货币进行大数据可视化，并运用股票领域工业模型对数据进行预测，使用机器学习算法对数字货币间的关联性进行挖掘；同时基于数字货币门户网站开发数据，建立实时交易数据展示；爬取互联网中与数字货币市场关联度高的新闻，在本平台进行精简展示。达成以上目标后，用户可以清晰直观获得数字货币市场的历史变化，便捷获取数字货币市场实时交易详情，了解当前全球数字货币市场的舆论情况，帮助用户对数字货币市场做出及时准确的判断。

## 项目概况

|  |  |
| --- | --- |
| **项目名称** | 数字货币大数据分析平台 |
| **任务提出者** | 华迪公司 |
| **用户** | 华迪公司 |
| **开发者** | 数字货币大数据分析平台开发组 |

* 1. **项目范围**

统计各类数字货币每日交易基本信息，以及有关数字货币的热点事件，存入数据库随后基于这些信息进行分类和预测，并反馈给用户作为参考信息。

## 交付的产品

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **序号** | **产品名称（包括服务）** | **介质** | **交付日期** | **接收方** | **接受准则** |
| **1** | 数字货币大数据分析平台 | 前端网站 | 2021.07.25 | 华迪公司 |  |

# 开发成员

|  |  |  |  |
| --- | --- | --- | --- |
| **序号** | **角色** | **成员姓**  **名** | **职责** |
| 1 | 项目经理、爬虫工程师、前端工程师、后端工程师 | 肖文斌 | 负责项目的整体协调工作，项目需求分析文档编写，文档审查与修改，原始数据集的爬取；  负责登录、注册、忘记密码、后台管理等功能的前端、后端功能设计及实现。 |
| 2 | 前端工程师、后端工程师、测试工程师 | 文杰 | 负责项目展示大面板的前端、后端功能设计及实现，协助登录、后台管理部分功能实现及Bug修复；  负责平台功能测试工作，编写测试结果文档。 |
| 3 | 爬虫工程师、前端工程师、后端工程师 | 陈楚凡 | 负责实时大数据的数据采集、分析、展示，项目系统架构文档编写；  负责实时大数据展示前端、后端功能设计及实现。 |
| 4 | 前端工程师、后端工程师、测试工程师 | 景南瑞 | 负责登录、注册、忘记密码等功能的前端、后端功能设计及实现，大数据展示面板透明化设计及实现。  负责项目开发计划文档编写，负责项目测试计划设计及相关文档编写。 |
| 5 | 前端工程师、后端工程师、测试工程师 | 吴后望 | 负责项目展示大面板的前端、后端功能设计及实现；  负责项目测试用例设计及相关文档编写，负责平台安全、前端、压力测试工作。 |
| 6 | 数据库工程师、数据分析师、  算法工程师 | 陈修鸿 | 负责项目所使用的数据库设计及实现，对原始数据进行数据挖掘、分析，编写相关数据库设计文档；  负责项目使用机器学习算法的设计，实现预测、关联数据的计算获取。 |

# 项目开发计划

## 需求确定阶段

对数字货币大数据分析平台进行需求分析，确定该项目需要实现的所有具体功能及其实现方法。随后对整个流程进行分工，撰写项目需求分析报告。

## 数据爬取阶段

根据上一阶段敲定的功能确定需要爬取的数据内容，编写爬虫程序，进入目标网站进行数据爬取。由于数据量非常大，数据爬取过程可能会在整个开发过程中持续，结果按照《数据库设计文档》规定进行存储。

## 数据分析阶段

将已爬取的数据进行分析。首先进行数据清洗，清除无关数据与各种符号， 并将互相有关联的数据整合，得到若干表。随后采用机器学习相关算法对数据进行预测或分类。

## 数据可视化阶段

根据具体功能和 Web 页面的布局，将各数据、图像等展示在合适的位置， 满足《项目需求分析文档》4.4 节中的界面需求，达到良好的用户交互。

# 项目所需技能和培训计划

## 项目所需技能

管理技能：项目负责人对整体进度的规划调整，和对人力资源进行合理分配的技能；

技术技能：网站爬虫、前端后端实现、数据库交互、分布式集群计算等技能； 业务技能：团队成员相互沟通的能力，解决开发过程中遇到的各种问题的能

力等。

## 项目培训计划

①Python 语言基础运用

②Requests、Scrapy、Selenium 等爬虫框架实例教学

③虚拟机分布式集群环境搭建教程

④前端框架构建教程

⑤git团队项目合作的运用

⑥spark机器学习对数据处理的运用

⑦MySQL、HDFS 等各数据库的原理与运用

# 开发计划

## 项目监控记录

|  |  |  |  |
| --- | --- | --- | --- |
| **跟踪活动** | **日期** | **形式或工**  **具** | **结果或记录** |
| 进度跟踪 | 07.13 | 人工检查 | 1. 项目立项 2. 拟定需求文档，添加数据说明 3. Gitee上建立项目初始结构 |
| 进度跟踪 | 07.14 | 人工检查 | 1. Web框架flask和Django试用 2. 前端图标及界面设计 |
| 进度跟踪 | 07.15 | 人工检查 | 1. 爬虫部分基本完成，部分细节之后进行调整 2. 机器学习进行中，导出部分样例数据 3. 确定Web框架使用flask进行设计 |
| 进度跟踪 | 07.16 | 人工检查 | 1. 用户登录注册忘记密码功能实现 2. 前端Echart绘图，涨跌图，排行榜实现 3. 后台管理界面初步设计 4. 添加导航栏 5. 数据库属性重构 |
| 进度跟踪 | 07.17 | 人工检查 | 1. 前端Echart绘图，主要的k线图、饼图、周涨幅图完善 2. 添加404，500错误页面 3. 注册界面用户名及密码正则验证实现 |
| 进度跟踪 | 07.19 | 人工检查 | 1. Spark部分上传币种相关性的信息 2. 前端样式修改 3. 后端代码优化修改 |

## 风险管理计划

①人员流动与请假风险：无流动可能，事假和病假每人每周控制在 8 小时；

②项目要求变更风险：采取敏捷开发模式，将产品功能模块化，高内聚低耦合，方便追加功能；

③技术风险计划：尽可能选用开发社区活跃、开发体系完整成熟的语言进行项目开发，并统一组内成员的开发环境

## 度量与分析计划

|  |  |  |  |
| --- | --- | --- | --- |
| **度量周期** | **度量任务描述** | **负责人** | **参与成员** |
| 每周 | 描述一周内工作执行情况与完成度 | 项目经理 | 全部小组成员 |
| 每周 | 各组员按照自己工作内容完成周报 | 各位成员 | 全部小组成员 |
| 系统测试完成后一天 | 测试人员生成《系统测试报告》进  行反馈 | 项目经理 | 测试人员 |

## 质量保证计划

①每阶段任务结束前进行质量保证测试，根据测试结果优化软件产品；

②质量保障内容：可修改性、易用性、可移植性。

## 配置管理计划

①配置管理内容：源代码版本控制、开发环境配置、数据整合控制；

②配置管理方案：

* + - 定义各类配置项入库、出库的准则和操作流程；
    - 定义基线变更的准则和操作流程；
    - 明确配置库的备份及维护方法，当出现异常后如何恢复的预案等；
    - 版本发布的准则、发布流程及发布计划，如测试版本、β版本等；

③管理报告内容：基线库的基线项的清单、基线项的名称、版本、存放位置等。

* 1. **系统测试计划**见测试计划文档

# 工作环境

## 开发环境

* + 1. *硬件设备*

本项目在 PC 环境下开发，要求各计算机配备足够的存储空间。

* + 1. *支持工具和软件环境*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **软件／工具** | **版本** | **获取方式** | **负责人** | **到位时间** | **费用预算（万元）** | **备注** |
| JetBrains PyCharm | 2018 | 社区版 | 无 | / | 0 | 非商用，免  费 |
| CentOS7 虚  拟机系统 | 7 | 国内镜像 | 无 | / | 0 |  |
| MySQL 数  据库 | 6.5 | 官方下载  配置 | 无 | / | 0 |  |

## 测试环境

* + 1. *硬件设备*

个人计算机

* + 1. *支持工具和软件环境*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **软件／工具** | **版本** | **获取方式** | **负责人** | **到位时间** | **费用预算（万元）** | **备注** |
| JetBrains PyCharm | 2018 | 社区版 | 无 | / | 0 | 非商用，免费 |
| 谷歌浏览器 | 72.0 | 官方下载 | 无 | / | 0 |  |

## 