编码：项目级-文档编写格式规范-Version\_2.0.0-signed-20200828

文档等级：□发布级 ■项目级 □团队级

采纳标准：团队文档质量标准

网络文学大数据分析系统

**Big Data Analysis System of Network Literature**

软件需求规格说明书

**Version 2.0.0-signed**

编写人：李明霞、康艳晴

西工大-华迪实训第四小组

**All Rights Reserved**

**目 录**

**[1引言 4](#_Toc30573)**

[1.1 标识 4](#_Toc14331)

[1.2 系统概述 4](#_Toc7187)

[1.3 文档概述 4](#_Toc27062)

[1.4 基线 5](#_Toc4470)

**[2 引用文件 5](#_Toc11500)**

**[3 综合描述 5](#_Toc5215)**

[3.1项目前景 5](#_Toc6839)

[3.2用户特征 6](#_Toc21256)

[3.3运行环境 6](#_Toc949)

[3.3.1客户端 6](#_Toc29122)

[3.3.2服务器端 6](#_Toc16524)

[3.4设计和实现上的限制 6](#_Toc32200)

[3.5假设和依赖 7](#_Toc11259)

**[4外部接口需求 7](#_Toc1954)**

[4.1硬件接口 7](#_Toc14876)

[4.2软件接口 7](#_Toc10704)

[4.3通信接口 7](#_Toc32508)

**[5系统特性 8](#_Toc9617)**

[5.1需求描述 8](#_Toc18106)

[5.2数据需求 8](#_Toc20215)

[5.2.1书籍信息 8](#_Toc13253)

[5.2.2作者信息 9](#_Toc2745)

[5.2.3读者打赏信息 9](#_Toc15353)

[5.3功能需求 9](#_Toc29523)

[5.3.1优先级定义 9](#_Toc25508)

[5.3.2功能表 9](#_Toc10365)

[5.4功能描述 11](#_Toc19905)

[5.4.1题材收录占比 11](#_Toc8957)

[5.4.2月份排行 12](#_Toc13025)

[5.4.3季度排行 13](#_Toc30813)

[5.4.4年度排行 14](#_Toc7330)

[5.4.5用户阅读稳定性分析 15](#_Toc5396)

[5.4.6积分排行 16](#_Toc29097)

[5.4.7各题材top10 17](#_Toc26906)

[5.4.8长篇排行 18](#_Toc32361)

[5.4.9作者积分排行 19](#_Toc8709)

[5.4.10新晋作者排行 20](#_Toc18763)

[5.4.11稳定更新排行 21](#_Toc10011)

[5.4.12本月完结收益排行 22](#_Toc13553)

[5.4.13打赏收益排行 23](#_Toc28425)

[5.4.14读者消费排行 24](#_Toc18855)

[5.4.15读者订阅偏好 25](#_Toc17424)

[5.4.16同类网站排行 26](#_Toc1852)

**[6其他非功能需求 26](#_Toc29423)**

[6.1性能需求 26](#_Toc27362)

[6.2安全设施需求 27](#_Toc469)

[6.3安全性需求 27](#_Toc28866)

[6.4软件质量属性 27](#_Toc14363)

[6.4.1可靠性 27](#_Toc22367)

[6.4.2可获得性 27](#_Toc17561)

[6.4.3保密性 27](#_Toc23305)

[6.5其他需求 27](#_Toc32719)

**[附录：词汇表 28](#_Toc4587)**

# 1引言

## 标识

适用系统：网络文学大数据分析系统

中文名称：软件需求规格说明（SRS）

英文名称：Software Requirements Specification (SRS)

文档版本：2.0.0

文档编号：项目级-软件需求规格说明-Version\_2.0.0-signed-20200828

## 系统概述

本文档适用于“网络文学大数据分析系统”项目的开发过程。网络文学大数据分析系统项目由西北工业大学软件学院提出，由西工大-华迪实训第四小组成员负责实施，该项目编号为项目级-软件需求规格说明-Version\_2.0.0-signed-20200828，其软件产品版本号为“1.0”。

该项目基于B/S（Brower/Server）浏览器/服务器结构，采用传统的三层结构方式进行解耦开发。对当下最热门的网络文学阅读机构——“晋江文学”提供的各项数据进行可视化展示和分析。数据分析采用Hadoop和Spark。

## 文档概述

本文档依据《GB/T 8567-2006 计算机软件文档编制规范》制定，属于技术文档，仅限于项目相关人员阅读。

本文档的主要内容共分6部分:引言、引用文件、综合描述、外部接口需求、系统特性和其他非功能需求。综合描述部分主要对系统的整体结构进行了大致的介绍;系统特性部分对系统的功能需求进行了详细描述，是本文的主要部分;外部接口需求部分对软件接口、硬件接口和通信接口等进行了描述。

本文档的保密范围如下：

①开发人员

②测试阶段人员

③对本文档进行评审的人员或机构

④项目组以及其他有权需要调用本文档的人员

## 基线

基线：GB/T 8567-2006

# 2 引用文件

《GB/T 8567-2006 计算机软件文档编制规范》，国家标准

《项目级-文档编写格式规范-Version\_1.0.0-alpha-20200812》，工作室编写

《项目级-软件开发计划-Version\_1.0.0-alpha-20200812》，工作室编写

《项目级-软件配置管理计划-Version\_1.0.0-alpha-20200812》，工作室编写

《需求分析与系统设计》（机械工业出版社）作者：麦沙塞克；翻译：金芝

# 3 综合描述

## 3.1项目前景

随着计算机网络的发展，网络文学网站已经吸弓|到越来越多的用户。在此背景下，本系统通过对晋江文学网相关数据进行分析，旨在为读者推荐更优质的书籍，也为该网站的运营工作提供建议。

## 3.2用户特征

对本项目的展示数据可以快速领会；对网络文学的统计数据感兴趣；有基本的网络运用能力。

## 3.3运行环境

### 3.3.1客户端

Windows 系统。

### 3.3.2服务器端

表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 | 2020 | 官方下载 | 无 | 8/05 | 0 | 学生资格认证 |
| 虚拟机系统 | 7 | 国内镜像 | 无 | 8/06 | 0 |  |
| MySQL数据库管理 | 6.5 | 官方下载配置 | 无 | 8/11 | 0 |  |
| Redis | 5.0.9 | 官方下载配置 | 无 | 8/11 | 0 |  |

## 3.4设计和实现上的限制

1. 项目交付日期为2020年8月29日
2. 爬取数据量达到百万级别
3. 使用Hadoop集群资源管理系统管理存储数据
4. 使用Spark框架对数据清洗处理
5. 受相关法律、客户需求、使用环境(windows、Linux系统)、构建环境(Python 等)、开发技术等因素影响。
6. 硬件限制：内存储器: 8G或更高。

## 3.5假设和依赖

1. 无人员变更成本
2. 因事因病请假次数在项目期间每人不超过一次
3. 项目进度能按照进程规划如期进行

# 4外部接口需求

## 4.1硬件接口

网络文学大数据分析系统用户端需要调用NFC芯片(可选)、蓝牙芯片、移动网络芯片等一系列硬件;服务器端需要使用专用服务器。

## 4.2软件接口

网络文学大数据分析系统采用B/S结构。B/S 结构是基于 Web 技术与客户机/服务器结构的结合而提出来的一种多层结构，其中的 B 是指 Web 浏览器（Browse），S 是指应用服务器与数据服务器（Server）。本次项目采用B/S结构进行开发。

## 4.3通信接口

无特殊需求。

# 5系统特性

## 5.1需求描述

需求功能点：

1. 题材收录占比
2. 月份排行
3. 季度排行
4. 年度排行
5. 用户阅读稳定性分析
6. 积分排行
7. 各题材top10
8. 长篇排行
9. 作者积分排行
10. 新晋作者排行
11. 稳定更新排行
12. 本月完结收益排行
13. 打赏收益排行
14. 读者消费排行
15. 读者订阅偏好
16. 同类网站排行

## 5.2数据需求

### 5.2.1书籍信息

1. 书籍名称
2. 书籍题材
3. 书籍积分
4. 书籍起始更新时间
5. 书籍最后更新时间
6. 状态（完结/连载）
7. 书籍打赏数额
8. 一个月内更新天数

### 5.2.2作者信息

1. 作者名称
2. 作者积分
3. 作者入站时间

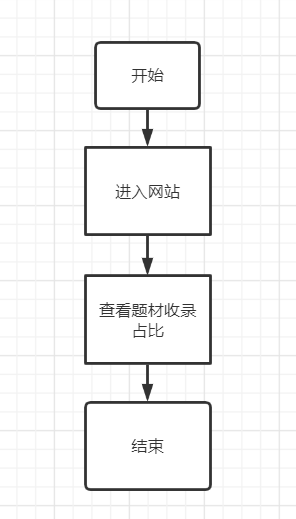
### 5.2.3读者打赏信息

1. 书籍名称
2. 读者名称
3. 打赏数额

## 5.3功能需求

## 5.4功能描述

### 5.4.1题材收录占比



题材收录占比流程图

标识：TC

简要说明：分析网站各类题材收录的图书占比。

优先级：高

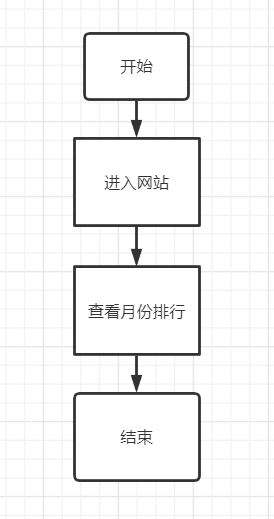
输入信息：鼠标放置在题材收录占比图的饼状分区上。

系统响应：从服务器端获取数据。

输出信息：显示该题材收录百分比及数量。

界面形式：饼状图。

### 5.4.2月份排行



月份排行流程图

标识：ML

简要说明：分析当前月份最受欢迎的图书排行。

优先级：中

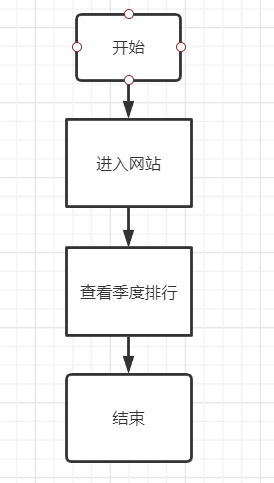
输入信息：鼠标放置在月份排行表图书柱形上。

系统响应：从服务器端获取数据。

输出信息：显示该图书的作者题材等更多信息。

界面形式：柱状图。

### 5.4.3季度排行



季度排行流程图

标识：QL

简要说明：分析当前季度最受欢迎的图书排行。

优先级：中

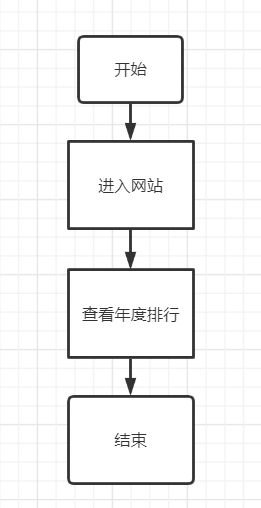
输入信息：鼠标放置在季度排行表图书柱形上。

系统响应：从服务器端获取数据。

输出信息：显示该图书的作者题材等更多信息。

界面形式：柱状图。

### 5.4.4年度排行



年度排行流程图

标识：YL

简要说明：分析当前年度最受欢迎的图书排行。

优先级：中

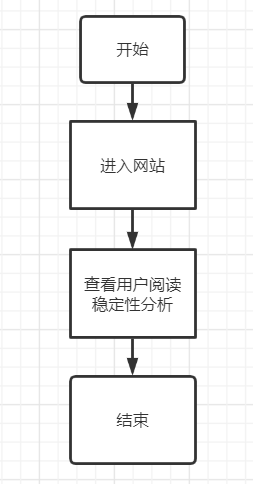
输入信息：鼠标放置在年度排行表图书柱形上。

系统响应：从服务器端获取数据。

输出信息：显示该图书的作者题材等更多信息。

界面形式：柱状图。

### 5.4.5用户阅读稳定性分析



用户阅读稳定性分析流程图

标识：RS

简要说明：通过对比月份、季度、年份排行中最受欢迎的图书题材占比，分析本网站读者阅读情况的稳定性。

优先级：中

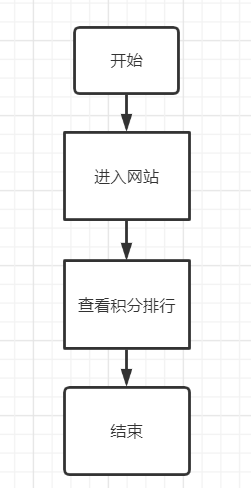
输入信息：鼠标放置在某一题材折线的对应时间点（月份、季度、年度）上。

系统响应：从服务器端获取数据。

输出信息：显示该时间点（月份、季度、年度）不同题材的热度数据。

界面形式：柱状图。

### 5.4.6积分排行



积分排行流程图

标识：CL

简要说明：通过积分分析所有书籍的受欢迎程度。

优先级：高

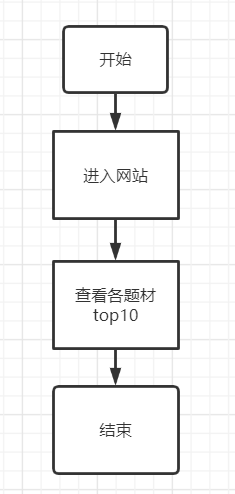
输入信息：鼠标放置在某一图书名称上。

系统响应：从服务器端获取数据。

输出信息：显示该图书积分、题材等信息。

界面形式：词云。

### 5.4.7各题材top10



各题材top10流程图

标识：TL

简要说明：各个题材受欢迎的前10名。

优先级：低

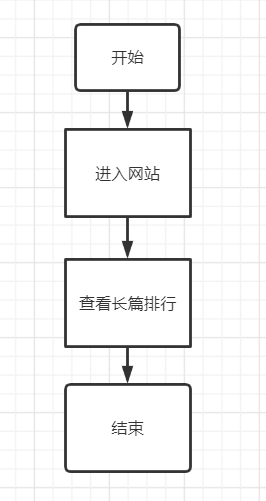
输入信息：鼠标点击题材按钮。

系统响应：从服务器端获取数据。

输出信息：显示该题材最受欢迎的图书前十信息（排名、书名、积分）。

界面形式：表格。

### 5.4.8长篇排行



长篇排行流程图

标识：NL

简要说明：连载时间超过一年，按最近一周更新字数和积分共同排序。

优先级：低

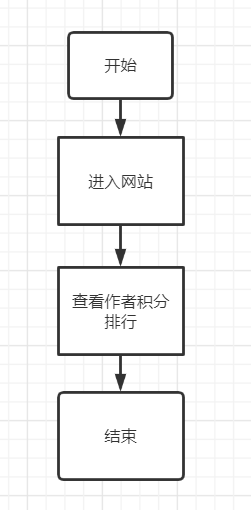
输入信息：鼠标放置在某一图书柱形上。

系统响应：从服务器端获取数据。

输出信息：显示该图书作者、积分等信息。

界面形式：柱状图。

### 5.4.9作者积分排行



作者积分排行流程图

标识：ACL

简要说明：通过作者积分分析最受欢迎的作者。

优先级：中

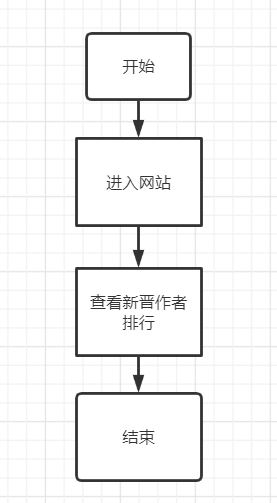
输入信息：鼠标放置在某一作者名称上。

系统响应：从服务器端获取数据。

输出信息：显示该作者的代表作信息。

界面形式：词云。

### 5.4.10新晋作者排行



新晋作者排行流程图

标识：NAL

简要说明：申请成为作者30天内积分排行。

优先级：中

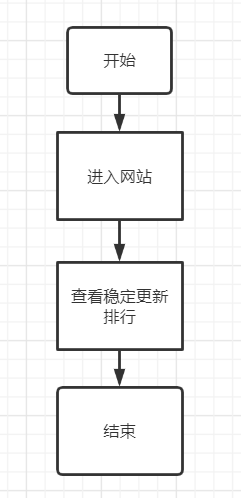
输入信息：鼠标放置在某一作者名称上。

系统响应：从服务器端获取数据。

输出信息：显示该作者创作的图书及积分信息。

界面形式：条形图。

### 5.4.11稳定更新排行



稳定更新排行流程图

标识：SUL

简要说明：一个月内稳定更新的图书排行。

优先级：低

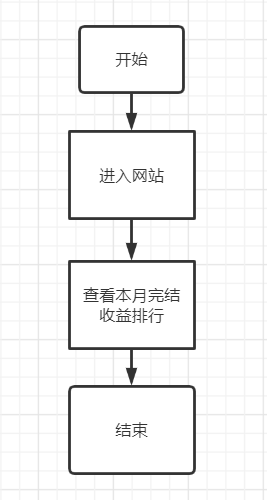
输入信息：鼠标放置在某一图书名称上。

系统响应：从服务器端获取数据。

输出信息：显示该图书更新信息。

界面形式：表格。

### 5.4.12本月完结收益排行



本月完结收益排行流程图

标识：MFEL

简要说明：图书完结30日内总收益排行。

优先级：中

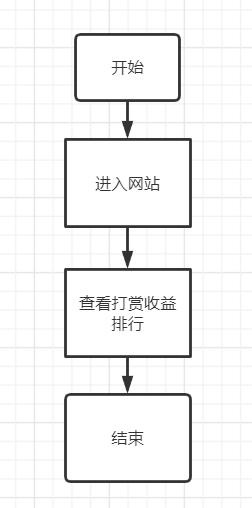
输入信息：鼠标放置在某一图书名称上。

系统响应：从服务器端获取数据。

输出信息：显示该图书积分、题材等信息。

界面形式：气泡图。

### 5.4.13打赏收益排行



打赏收益排行流程图

标识：EL

简要说明：获得打赏最多的图书排行。

优先级：高

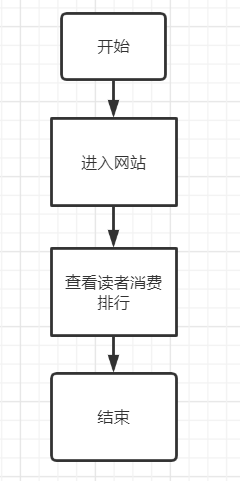
输入信息：鼠标放置在某一图书名称上。

系统响应：从服务器端获取数据。

输出信息：显示该图书打赏收益信息。

界面形式：柱状图。

### 5.4.14读者消费排行



读者消费排行流程图

标识：RCL

简要说明：给图书打赏消费最多的读者排行。

优先级：中

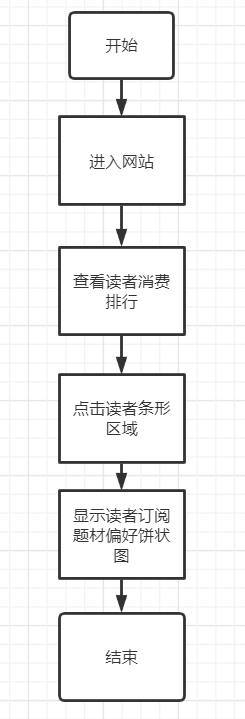
输入信息：鼠标放置在某一读者名称条形区域上。

系统响应：从服务器端获取数据。

输出信息：显示该读者打赏总额信息及偏好题材饼状图。

界面形式：条形图。

### 5.4.15读者订阅偏好



读者订阅偏好流程图

标识：RSP

简要说明：根据读者打赏书籍题材占比，分析读者阅读偏好。

优先级：中

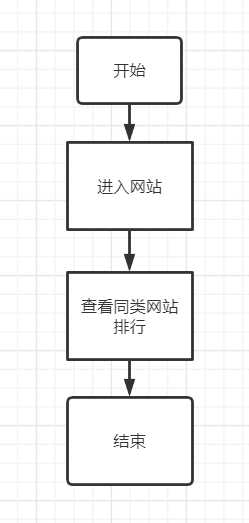
输入信息：鼠标放置在某一偏好题材饼状分区上。

系统响应：从服务器端获取数据。

输出信息：显示该读者订阅该题材的百分比。

界面形式：饼状图。

### 5.4.16同类网站排行



同类网站排行

标识：SSL

简要说明：根据用户量对比分析同类网站排行。

优先级：高

输入信息：鼠标放置在某一网站名称上。

系统响应：从服务器端获取数据。

输出信息：显示该网站用户量信息。

界面形式：折线图。

### 5.3.1优先级定义

1. 高——软件必须实现的功能，用户有明确的功能定义和要求；
2. 中——软件应该实现的功能，用户的功能定义和要求可能是模糊的、不具体的、或低约束的，但是这类功能的缺少会导致用户的不满意，因此这类功能的具体需求应当由需求分析人员诱导用户产生并明确；
3. 低——软件尽量实现的功能，并可根据开发进度进行取舍，但这类功能的实现将会增加用户的满意度。

### 5.3.2功能表

表2 功能表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 功能名称 | 功能需求标识 | 优先级 | 概述 |
| 1 | 题材收录占比 | TypeCount(TC) | 高 | 分析网站各类题材收录的图书占比 |
| 2 | 月份排行 | MonthList(ML) | 中 | 分析当前月份最受欢迎的图书排行 |
| 3 | 季度排行 | QuanterList(QL) | 中 | 分析当前季度最受欢迎的图书排行 |
| 4 | 年度排行 | YearList(YL) | 中 | 分析当前年度最受欢迎的图书排行 |
| 5 | 用户阅读稳定性分析 | ReaderStability(RS) | 中 | 通过对比月份、季度、年份排行中最受欢迎的图书题材占比，分析本网站读者阅读情况的稳定性 |
| 6 | 积分排行 | CreditList(CL) | 高 | 通过积分分析所有书籍的受欢迎程度 |
| 7 | 各题材top10 | TypeList(TL) | 低 | 各个题材受欢迎的前10名 |
| 8 | 长篇排行 | NovelList(NL) | 低 | 连载时间超过一年，按最近一周更新字数和积分共同排序 |
| 9 | 作者积分排行 | AuthorCreditList(ACL) | 中 | 通过作者积分分析最受欢迎的作者 |
| 10 | 新晋作者排行 | NewAuthorList(NAL) | 中 | 申请成为作者30天内积分排行 |
| 11 | 稳定更新排行 | StabilityUpdateList(SUL) | 低 | 一个月内稳定更新的图书排行 |
| 12 | 本月完结收益排行 | MonthFinishedEarningList(MFEL) | 中 | 图书完结30日内总受收益排行 |
| 13 | 打赏收益排行 | EarningList(EL) | 高 | 获得打赏最多的图书排行 |
| 14 | 读者消费排行 | ReaderCostList(RCL) | 中 | 给图书打赏消费最多的读者排行 |
| 15 | 读者订阅偏好 | ReaderSubscribePreference(RSP) | 中 | 根据读者打赏书籍题材占比，分析读者阅读偏好 |
| 16 | 同类网站排行 | SimilarSitesList(SSL) | 高 | 根据用户量对比分析同类网站排行 |

# 6其他非功能需求

## 6.1性能需求

用户操作软件完成，网络文学大数据分析系统会给出用户反馈，反馈在1秒内完成，软件延迟时间不超过3秒。

## 6.2安全设施需求

1. 为了系统数据的安全性，会定时对数据库的数据进行人工备份。
2. 为了防止人为破坏，实现了sql防注入攻击。
3. 为了防止没有权限的操作，系统有严格的权限控制。
4. 系统有较好的容错能力，即使发生了错误能很快恢复。

## 6.3安全性需求

确保数据安全。

## 6.4软件质量属性

可靠性优于可维护性，保密性优于可获得性。

### 6.4.1可靠性

网络文学大数据分析系统的可靠性为99%，软件运行时漏洞不超过2个。

### 6.4.2可获得性

网络文学大数据分析系统在维护期间用户无法获得软件的相关服务，其他时间可正常获得。

### 6.4.3保密性

网络文学大数据分析系统客户端打卡、申请请假、管理信息等操作，员工端和管理员端功能必须用账号密码登录后才可以使用。

## 6.5其他需求

1. 能够应用多种先进的工作器具。
2. 系统安装方便，易于维护。

# 附录：词汇表

**PyCharm**：是一种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。

**Requests**：用[Python](http://blog.csdn.net/shanzhizi" \t "_blank)语言编写，基于 urllib，采用 Apache2 Licensed 开源协议的 HTTP 库。

**Hadoop**：以分布式文件存储和MapRuduce为核心的基础架构，用户可以在不了解分布式底层细节的情况下，开发分布式程序。充分利用集群的威力进行高速运算和存储。

**Apache 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。

**B/S结构**:即Browser/Server (浏览器/服务器)结构，是随着Internet技术的兴起，对C/S结构的一种变化或者改进的结构。B/S 结构，主要是利用了不断成熟的www浏览器技术，结合浏览器的多种Script语言(VBScript、JavaScript--) 和ActiveX技术，用通用浏览器就实现了原来需要复杂专用软件才能实现的强大功能，并节约了开发成本，是一-种全新的软件系统构造技术。