1 C.2 基于AI的新闻热点聚合及可视化系统技术选型

1.1 本题需求

调研进行系统开发的相关前端开发工具和数据库管理系统软件，列出相关前端+后台开发组合

1.2 前端技术选型及开发工具选型

1.2.1 前端技术选型考虑方面

1. 项目规模

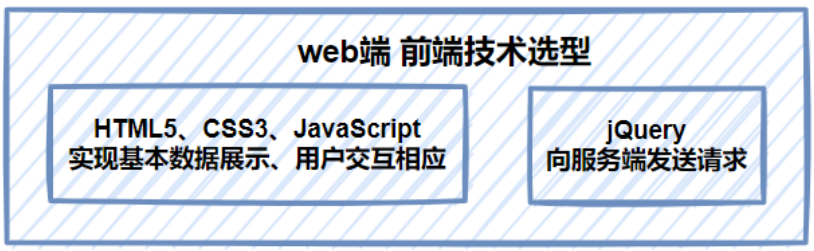
   本项目只是为课程设计准备，因此仅需完成 `version 1.0`，故可不需要过度考虑未来功能的拓展性

2. 团队因素

   本项目预计前端由两位同学共同负责，但是由于他们之前缺乏项目经验，因此预计前端他们要从零开始学，因此采用框架进行开发不是一个可行的方式，因此目前的预期是用原生的前端三件套外加 `jQuery` 做网络通讯

1.2.2 最终确定的前端技术选型

综上两个方面的考虑，因此确定前端所使用的技术栈为原生前端三件套 `HTML` 、 `CSS` 、 `JavaScript` 以及引入 `jQuery` 包做网络通讯，前端的技术选型如下图所示：



1.2.3 前端开发工具选型

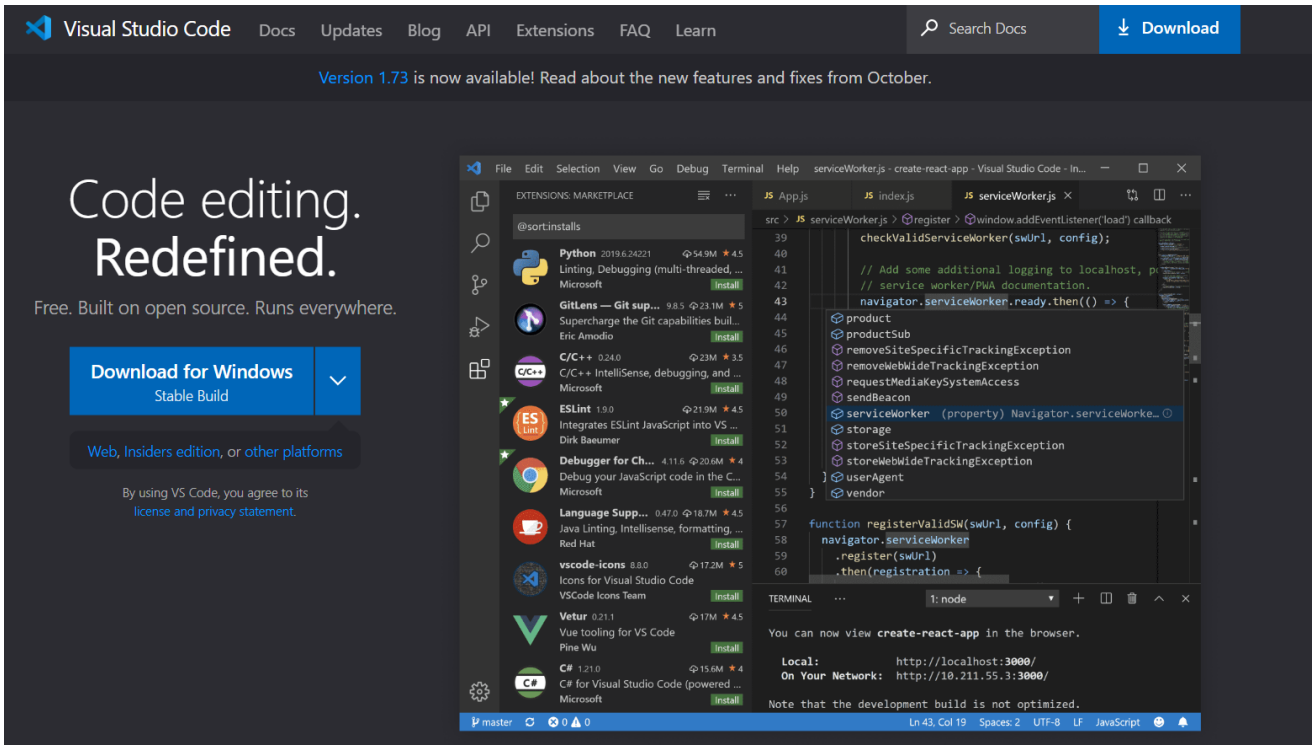
对于前端开发工具的选择，我的理想工具主要有以下两个，并对这两种工具分别进行分析

1. `WebStorm`， `jetbrains` 公司旗下的一款前端开发工具

我之所以考虑它原因是由于 `jetbrains` 公司的产品有一个共同的特点，对于开发工具的快捷键有一个统一，同时功能布局也基本相同，因此对于我这种基本依赖 `jetbrains` 公司旗下各种软件（`CLion`、`GoLand`、`Intellij IDEA`、`Pycharm`）的同学来说，可以简化很多适应编译器的过程

1. `Visual Studio Code`，Microsoft 公司旗下的一个可拓展的开发工具

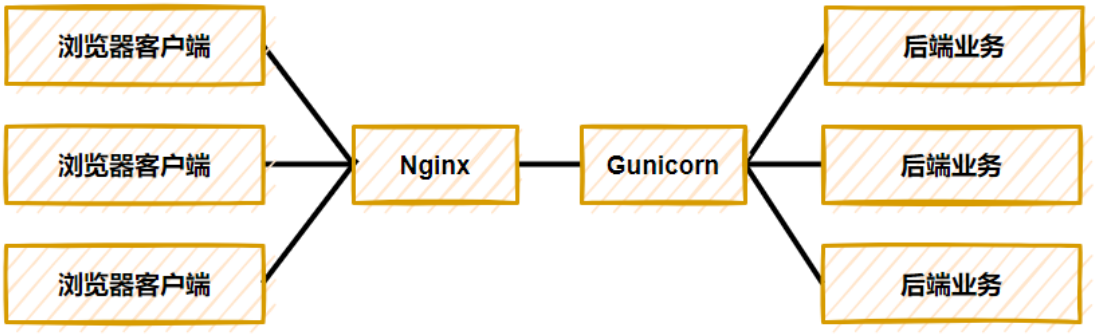
   我考虑它的原因是因为我之前做前端开发经常使用 `VS Code`，比较习惯了，同时它相比于 WebStrom 来说，优势在于轻量级，对于商务本非常友好（如果只装前端开发所需的拓展插件的话），同时它是一款免费面向所有开发者的产品，而 `jetbrains` 的产品需要付费，除非开学生认证，但是认证相对来说比较麻烦



基于以上对两种开发工具的优缺点分析，最终我对于本次数据库的课程设计前端开发工具选择 `VS Code`

1.3 后端技术选型及开发工具选型

1.3.1 网关层技术选型



网关层技术选型我经过考虑目标定位以使用 `Nginx` + `Gunicorn` 部署我的整个项目

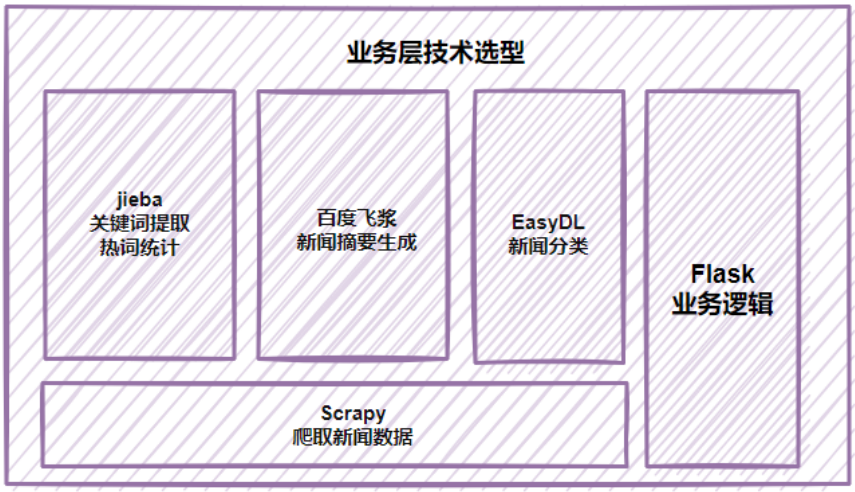
1. `gunicorn` 相当于是开启了多个进程，它具有以下优点：
2. 可以调节 `worker` 的数量，在请求较多时，自动新增 `worker`，请求较少时，自动减少 `worker`；
3. 帮我们管理 `worker`， `worker` 挂了自动重启
4. `nginx` 的优点：
    1. 负载均衡：有效的调度 `request`，而 `gunicorn` 虽然是多进程，但是没不能 主动地 对 `request` 进行调度
5. 动静分离：经过配置后， `nginx` 可以直接处理静态请求，而无需经过 `python web` 服务器，这一点 `gunicorn` 没有
    3. 缓存 `request` 和 `response`： `nginx` 可以缓存客户端的请求，收完整个请求后，转发给 `gunicorn`，等 `gunicorn` 返回 `response` 后，再转发给客户端

1.3.2 业务层技术选型

业务层技术选型主要我的选型依据参考主要是根据业务需求来进行选择，那么业务需求里面主要有两个技术需求影响了我的整个业务层技术选型

1. 爬虫
2. `NLP`

从所有语言中来说，同时对于爬虫和 `NLP` 有很好支持的可以说基本定在 `python` 了，因此语言选型就是 `python`，因此业务层只需要关注 `python` 的服务端框架选择，主要是因为我的后端技术栈是 Spring 系列，转到 python 的后端，主要的目的就是完成本次实验，因此需要找到一门简洁，而且更容易去上手的框架，而 flask 就很完美的符合我的需求，因为我肯定是做前后端分离的项目，因此我可以跳过 flask 的模板部分，学一个路由和数据库交互就可以开始进行开发，从零到上手的时间不超过一个小时（当然，不包括更进阶的拓展学习，只是简单的业务开发）



1.3.3 数据层技术选型

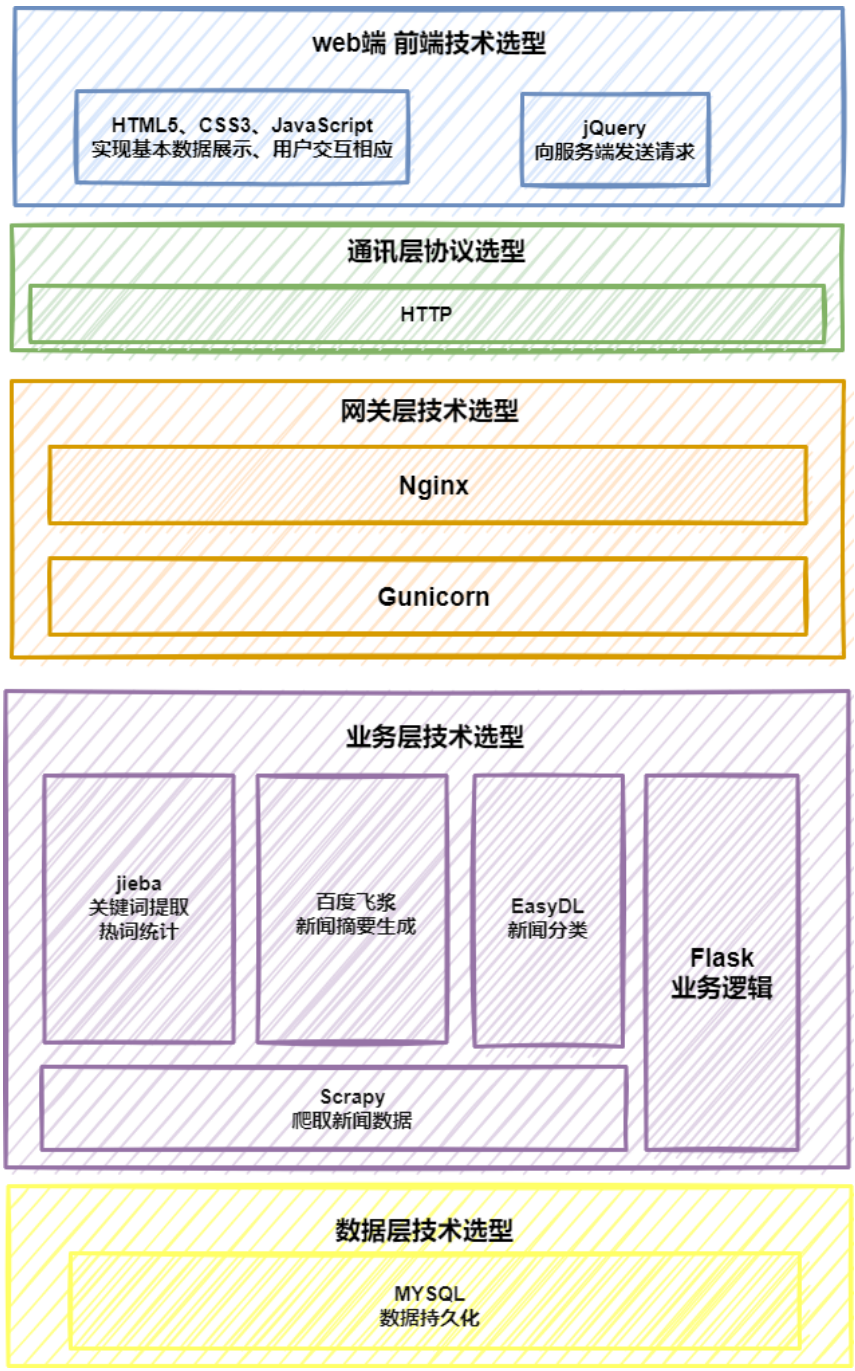
这里老师给出了一个问题，就是比较传统数据库管理软件及新型数据库云服务，我就不罗列他们的概念了，因为我对于这些数据库都有过应用，比如非关系性数据库我经常使用的就是 `Redis` 做业务中的一些数据缓存、`MongoDB` 我自己大二入门 `NOSQL` 的时候使用的第一个数据库，关系型我第一个也是最常用的就是 `MYSQL`，云数据库我第一次用的就是大二做校庆微信小程序的时候做到微信官方的云数据库

以下是我个人的一些做技术选型的时候的考虑点（但是只是个人看法，很可能出现错误）：

1. 云数据库，这个其实也分关系型和非关系型，比如微信小程序的云数据库就是类似 `MongoDB` 的数据库，而且云数据库因为官方会帮助维护，所以你只需要存入数据，然后通过云数据库提供的接口就好，因此云数据库最需要考虑的就是资金问题，是否项目值得花这么多钱，如果值得，云数据库绝对是开发体验超过一般数据库的，而很明显，我们的课程设计实验本来就是为了学数据库，所以使用云数据库相当于偷懒了，比如在服务器上的环境配置部署，本地数据库移植到服务器数据库。

   因此云数据库我一定不会去采用
2. 至于说到关系型和非关系型，我个人认为，能够结构化表现的数据表我都会采用关系型数据库，而 `MYSQL` 是目前的主流，社区建设，问题提出和解决更丰富，因此对于使用 `MYSQL` 来说，更容易接近自己遇到的问题，对于我这种小白友好

1.4 总体技术选型框架


web端 前端技术选型
HTML5、CSS3、JavaScript
实现基本数据展示、用户交互相应
jQuery
向服务端发送请求
通讯层协议选型
HTTP
网关层技术选型
Nginx
Gunicorn
业务层技术选型
jieba
关键词提取
热词统计
百度飞桨
新闻摘要生成
EasyDL
新闻分类
Flask
业务逻辑
Scrapy
爬取新闻数据
数据层技术选型
MYSQL
数据持久化