前提条件

本部分是为了确保项目可以运行

1. 本项目使用的工具以及版本 (其它版本没有测试)
    - hadoop3.2.1
    - spark2.2.1
    - mysql8
    - python3
    - 用到的python模块可用pip工具下载。eg. pip install pymysql
2. 必要的扩展工具
    - 请把项目中的jdbc文件放在spark安装目录的jars文件下

使用方法

第一次使用项目时

1. 进入FruitFree/flaskr目录，修改config.py文件夹，设置数据库账号密码
2. 退回FruitFree文件夹
3. 在命令行中设置临时变量（请确保在FruitFree文件夹下）
    - windows
        1. set FLASK_APP=flaskr
        2. set FLASK_ENV=development (调试模式)
        3. flask init-db (初始化数据库)
    - linux
        1. export FLASK_APP=flaskr
        2. export FLASK_ENV=development
        3. flask init-db
4. 进入FruitFree/compute目录，修改config.py文件夹，设置数据库账号密码
5. 执行所有py文件（到这里数据都在数据库中了)
6. 命令行命令：flask run
7. 浏览器输入 127.0.0.1：5000

第一次设置之后（不用再操作数据库）

1. 进入FruitFree文件夹
2. 在命令行中设置临时变量
    - windows
        1. set FLASK_APP=flaskr
        2. set FLASK_ENV=development (调试模式)
    - linux
        1. export FLASK_APP=flaskr
        2. export FLASK_ENV=development
3. 命令行命令： flask run
4. 在浏览器输入 127.0.0.1：5000

爬取数据

因为本项目数据是在2020.6爬取的，所以数据可能不再满足当下需求，可能需要重新爬取，本部分是重新爬取数据的方法（！！目标网址和网页内容、结构可能发生变化，爬虫代码2020.6之后不再维护！！）

1. 本项目涉及到的网址 （都是与水果相关的部分）
    - 淘宝 : 不是直接爬取淘宝网，而是使用的接口
    - 一亩田
    - 金投网
    - 水果交易网
2. 使用的爬虫工具
    - requests
        - 淘宝
        - 水果交易网
    - scrapy
        - 一亩田
        - 金投网
3. 使用方法：
    - requests:直接运行py文件
    - scrapy:以一亩田为例
        - 打开yimutian/spiders/yimutian_locationchart.py（spiders目录下每个py文件都是一个爬虫）
        - 发现name=yimutian
        - 回到yimutian目录
        - 命令行scrapy crawl yimutian
4. 得到csv文件后，移动到FruitFree/recource对应目录下