

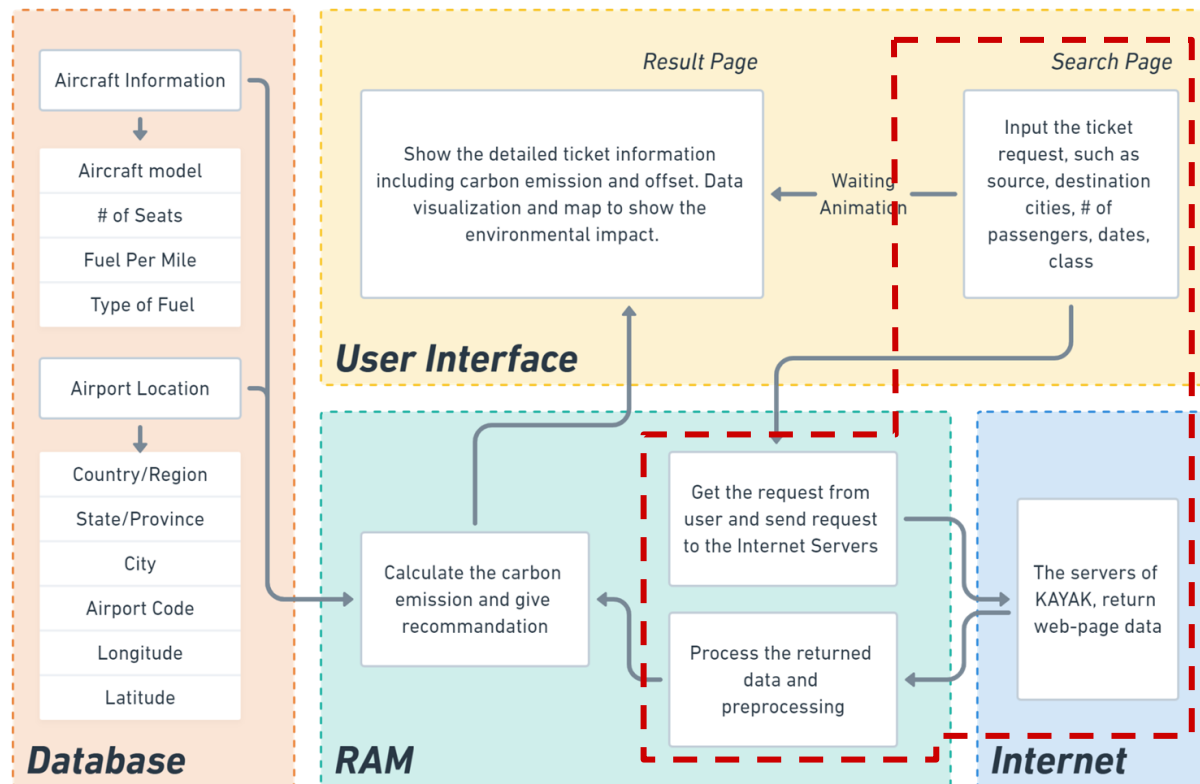
# Zero-Carbon Travel Patterns Decision-making System

University of Washington, CSE 583 Software Development for Data Scientists, Fall 2021

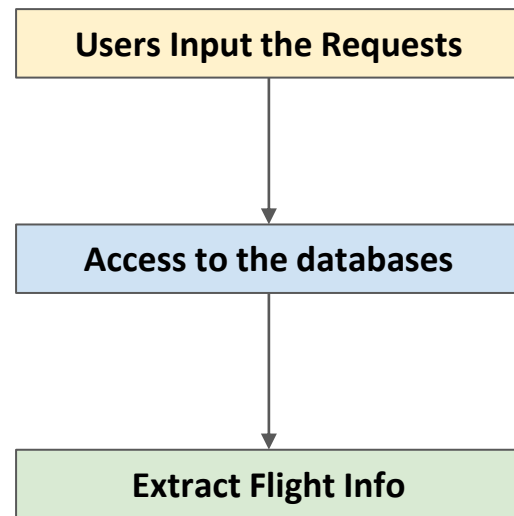
Chin-Wei Chen, Zining Cheng, Iman Haji, Haoyu Yue

Nov 16, 2021

# Background



## Today We FOCUS on How to Get Flight Ticket Data



# What technologies you considered

## Static Databases

Hard to find with prices, meaningless, huge dataset

## Existing API

Not up-to-date, do not fit our demands, limited functions for free

## Self-made Robots

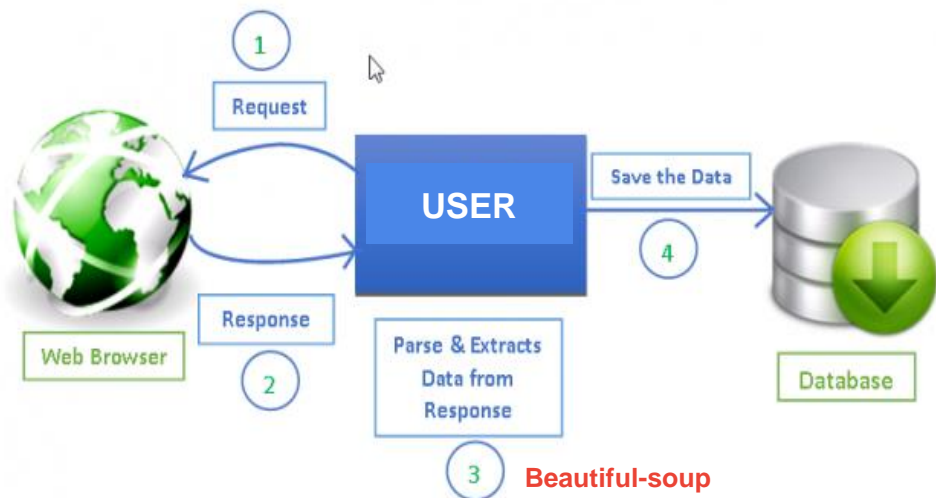
Practice chance, fit our project perfectly

**Decision ✓**

# Crawling in python: what you chose to use

**Request:** the requests module allows you to send HTTP requests using Python

**Beautifulsoup:** a Python library that is used for web scraping purposes to pull the data out of HTML files



	Scrapy	BeautifulSoup
Functionality	Complete package	Basically an HTML parser Requires additional libraries to open URLs and store the result
Learning Curve	A powerhouse for web scraping, but requires more time to learn	Relatively easy to understand for newbies
Speed and Load	Scrapy can get big jobs done very easily	Used for simple scraping jobs with efficiency, although it is slower
Extending functionality	Allow to write functions in your spider that can process your data	If the project does not require much logic

# **Crawling- Appeal of choice**

- **Achieve Automation**
- **Business Intelligence & Insights**
- **Unique and rich datasets**
- **Create applications for tools that don't have a public developer API**
- **Effective Data Management**

# Crawling- Drawbacks of choice

- Need to learn programming, use web scraping software or to pay a developer
- Websites regularly change their structure and crawlers require maintenance
- Careful about IP detection and robot.txt

# Reference

Requests: HTTP for Humans™. <https://docs.python-requests.org/en/latest/>

Scrape Beautifully With Beautiful Soup In Python. <https://analyticsindiamag.com/beautiful-soup-webscraping-python/>

The Advantages & Disadvantages of Web Scraping Data. <https://understandingdata.com/the-advantages-disadvantages-of-web-scraping-data/>

Web Crawlers in Python- A case in PCHOME website. [Mandarin].  
<https://www.youtube.com/watch?v=YHc6KvLsv20>

Making Web Crawlers Using Scrapy for Python. <https://www.datacamp.com/community/tutorials/making-web-crawlers-scrapy-python>