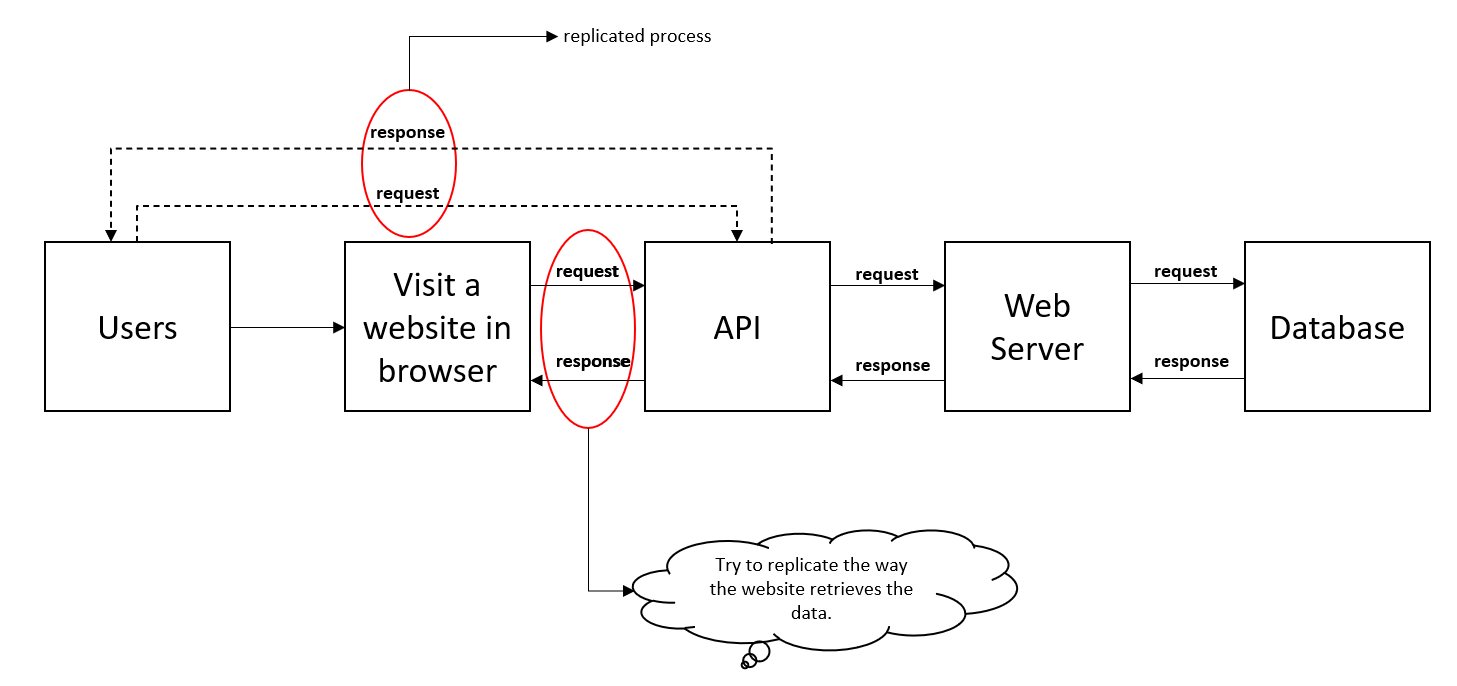
1. **Scraping HTML/XML content and “Not-Heavy” JavaScript Website**

* Recommendation package: BeautifulSoup, Selectolax
* Algorithm or Flow Process:

1. Get the HTML
2. Parse the document
3. Extract the target data
4. **Scraping by replicating requests to its API**

* The general concept:

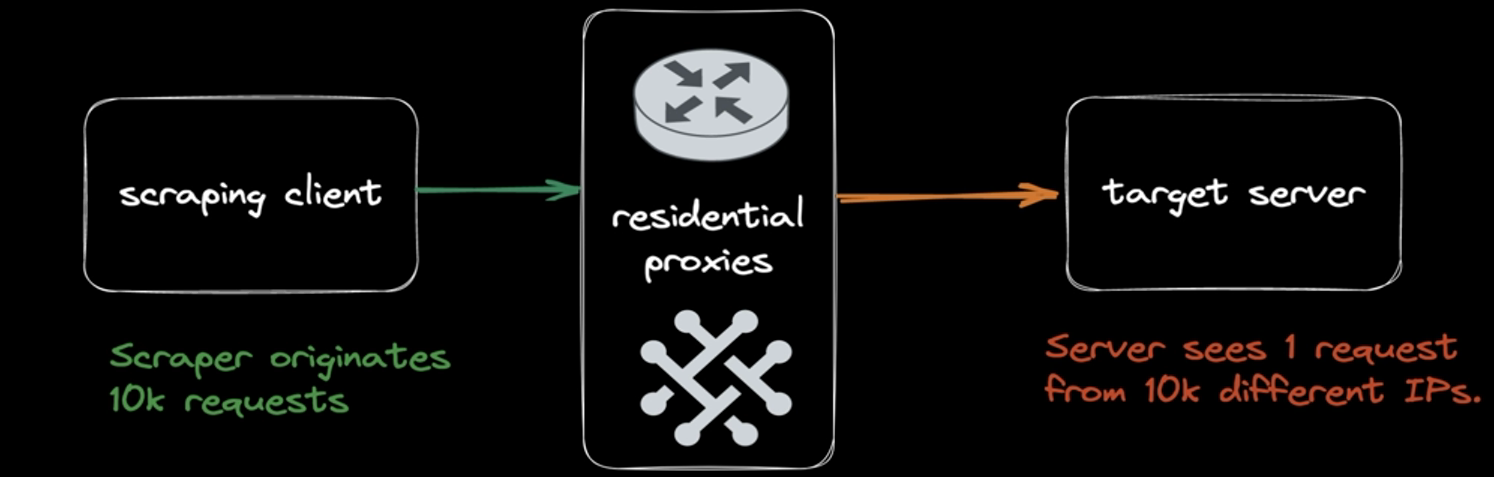
Modern websites, especially those with dynamic content, use APIs to load data onto the page. So rather than deal with HTML and CSS on website, we prefer to go straight to the server and replicate the request that the site is making directly.



* Pros:
  + The process is faster.
  + Less volatile 🡺 It means that the APIs do not change frequently as user interface.
  + Often more customizable 🡺 As long as we know the API paremeters, we can customize the data that we want to fetch.
* Cons:
  + Access restrictions and authentication 🡺 (1) Many APIs require authentication, (2) Repeated requests from the same IP address may lead to temporary or permanent blocking by the server.
  + Anti-Bot measures 🡺 Anti-bot makes the scraping process more difficult.
  + Legal and Ethical Concerns

1. **Using Proxy to Handle Anti-bot and Getting Blocked**

* The proxies are application-layer between client and server.
* The illustration of how the proxy for scraping works:

****