Web Scraping & Web Services

Hello!

I am Eslam Ahmed

I am a software engineer.

You can find me at jeksogsa@gmail.com

Hello!

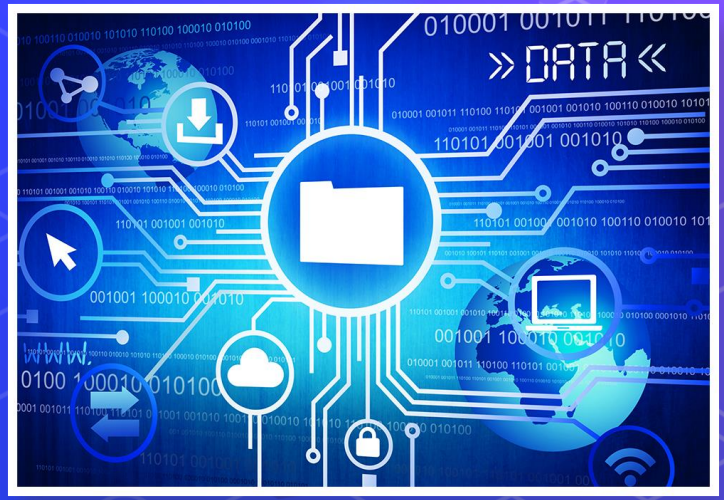
I am Eman Ehab

I am a ML research engineer.

You can find me at
emanehab.ieee@gmail.com

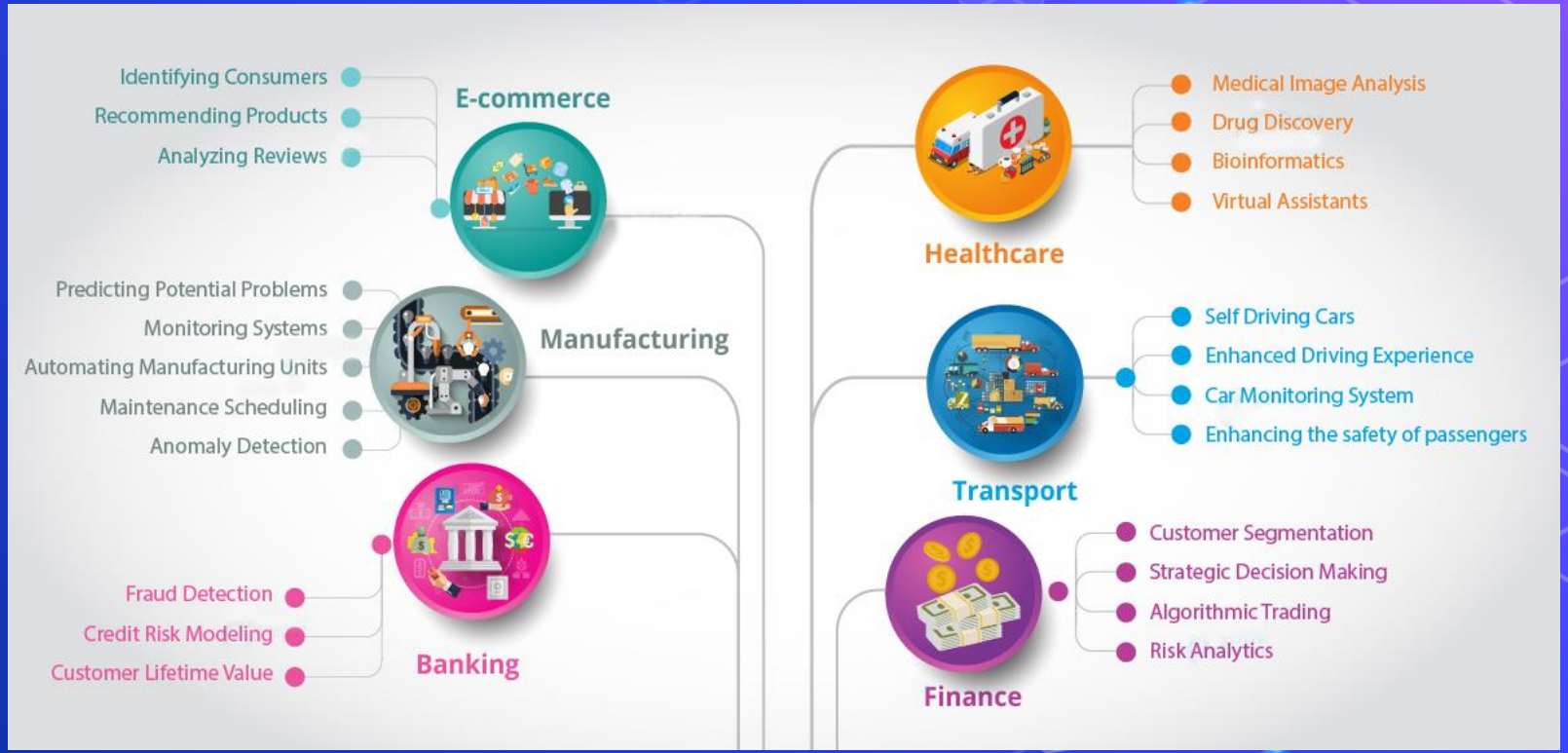
What is Data Science

Data science is a concept to unify statistics, data analysis, machine learning, domain knowledge and their related methods" in order to "understand and analyze actual phenomena with data.

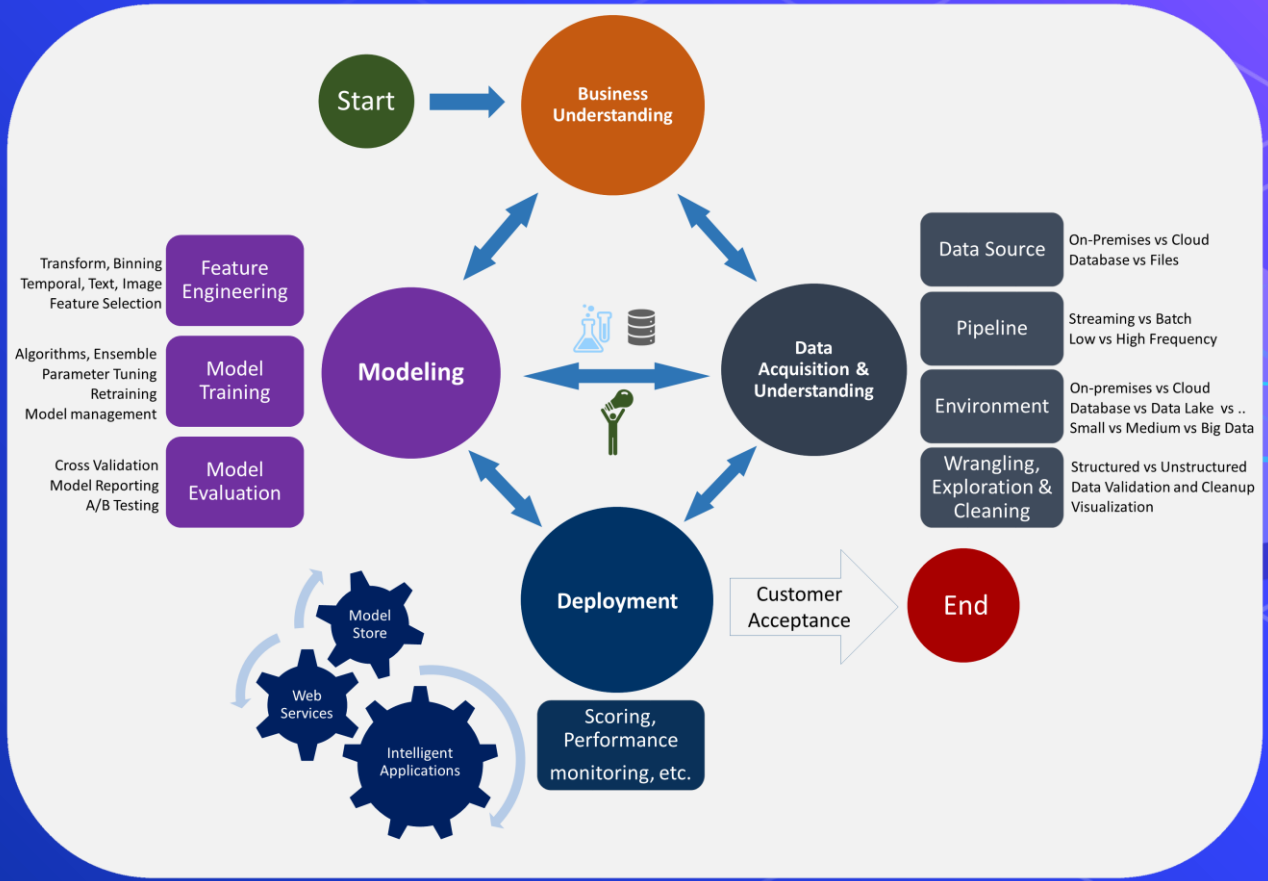


It uses techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, domain knowledge and information science.

Applications of Data Science



Data Science life cycle

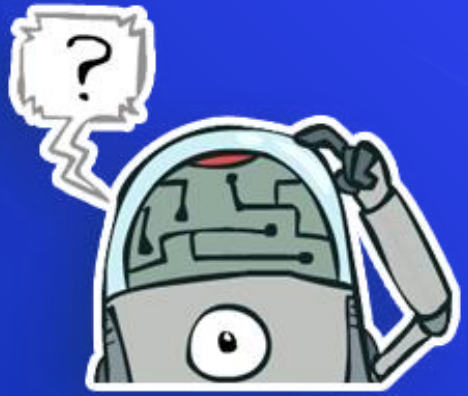


Data Collecting

Public Datasets

- https://www.kaggle.com/datasets
- http://archive.ics.uci.edu/ml/index.php
- https://data.fivethirtyeight.com
- https://github.com/BuzzFeedNews
- https://opendata.socrata.com
- https://cloud.google.com/bigquery/public-data
- https://github.com/awesomedata/awesome-public-datasets
- https://www.data.gov
- http://academictorrents.com/browse.php
- https://www.quandl.com/search
- https://tinyletter.com/data-is-plural

Data Collecting



So What we do if there is
no datasets available ?!

We will learn about Web Scraping and Rest APIs web services

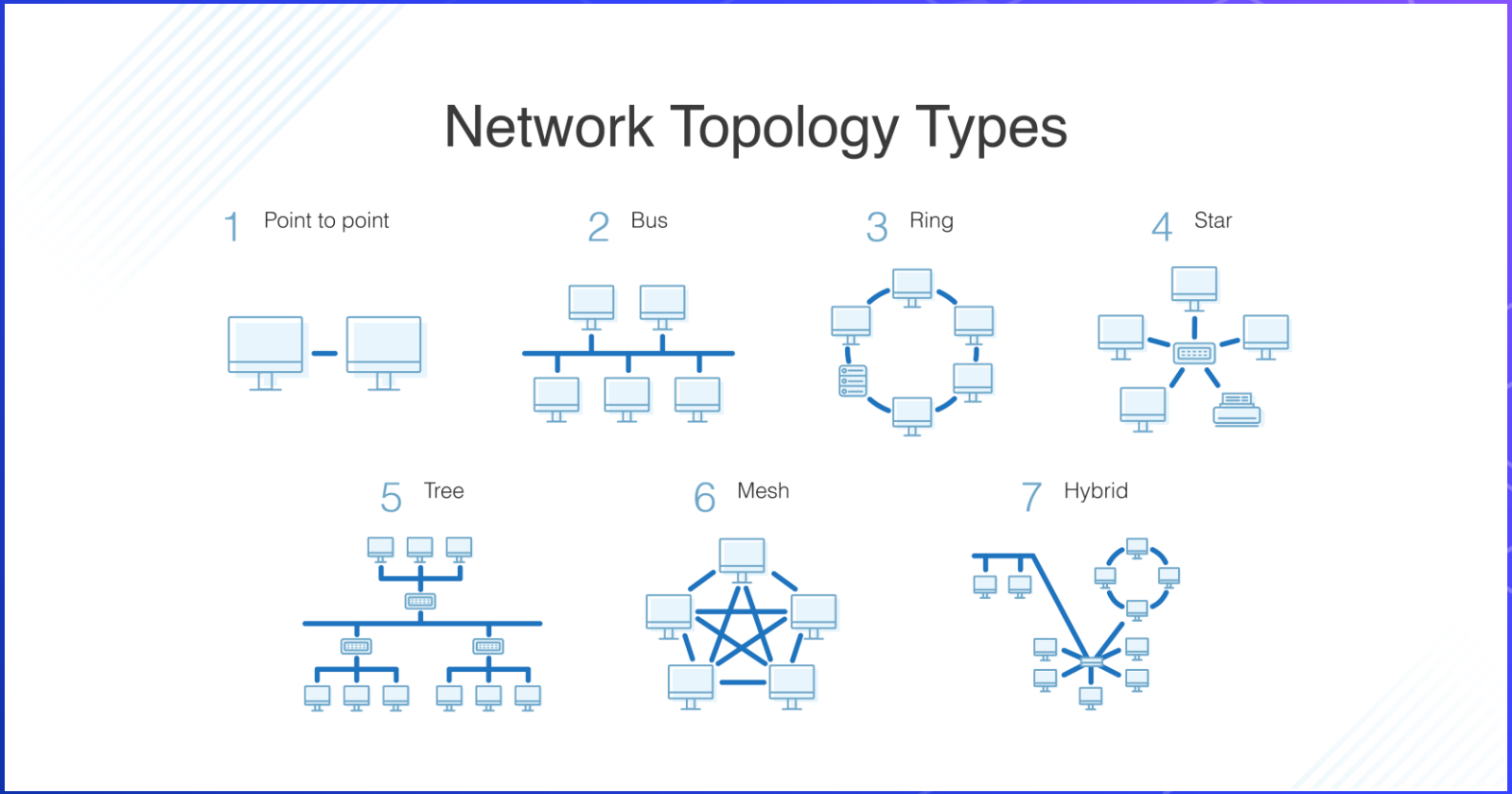
Agenda

- What is Network Topologies
- What is Internet and Web Servers
- HTTP Request/Response Cycle
- HTML
- CSS
- Scrapping Concept
- `BeautifulSoap` Library
- REST API Web Services
- Work with JSON

Agenda

- What is Network Topologies
- What is Internet and Web Servers
- HTTP Request/Response Cycle
- HTML
- CSS
- Scrapping Concept
- `BeautifulSoap` Library
- REST API Web Services
- Work with JSON

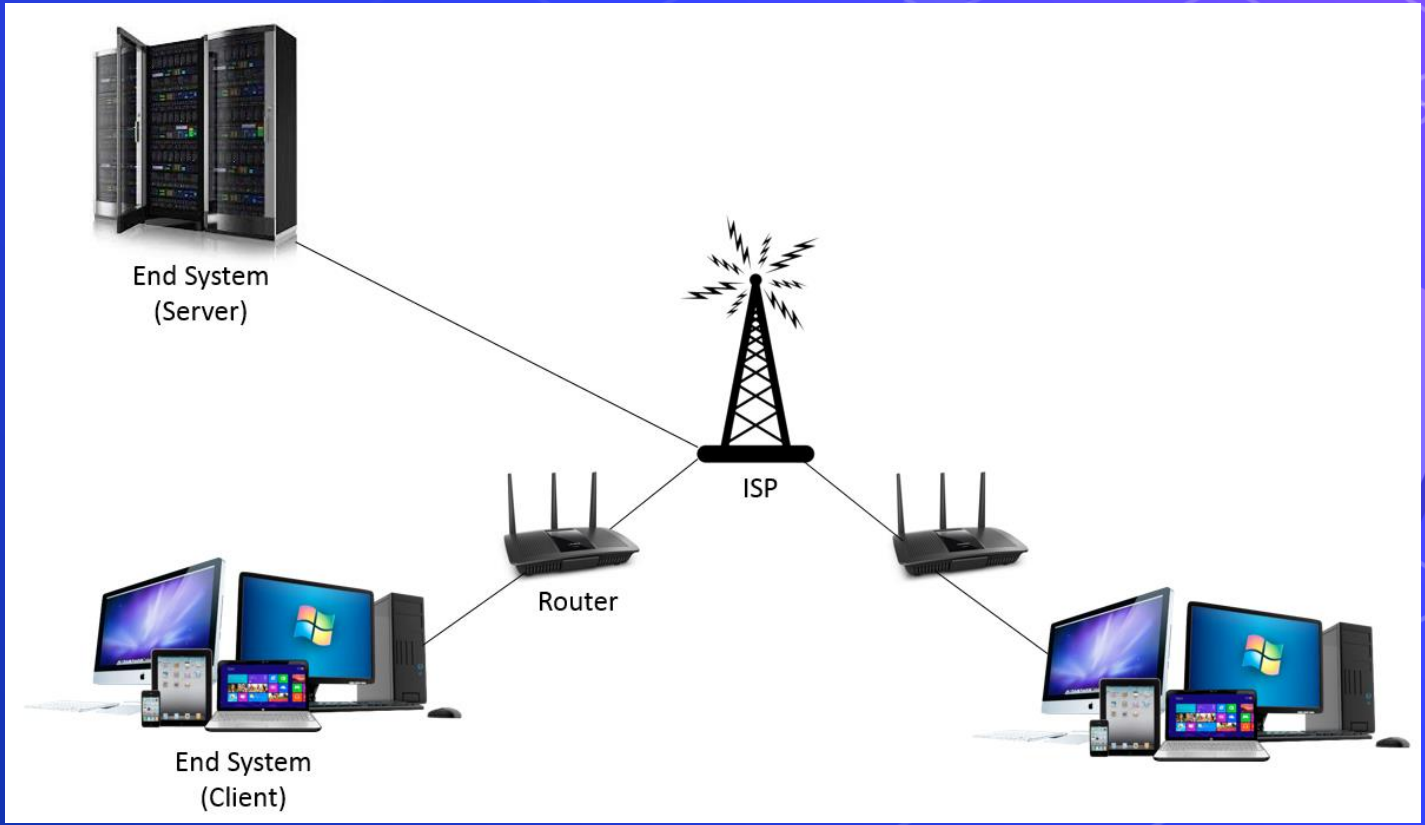
What is Network Topologies



Agenda

- What is Network Topologies
- **What is Internet and Web Servers**
- HTTP Request/Response Cycle
- HTML
- CSS
- Scrapping Concept
- `BeautifulSoap` Library
- REST API Web Services
- Work with JSON

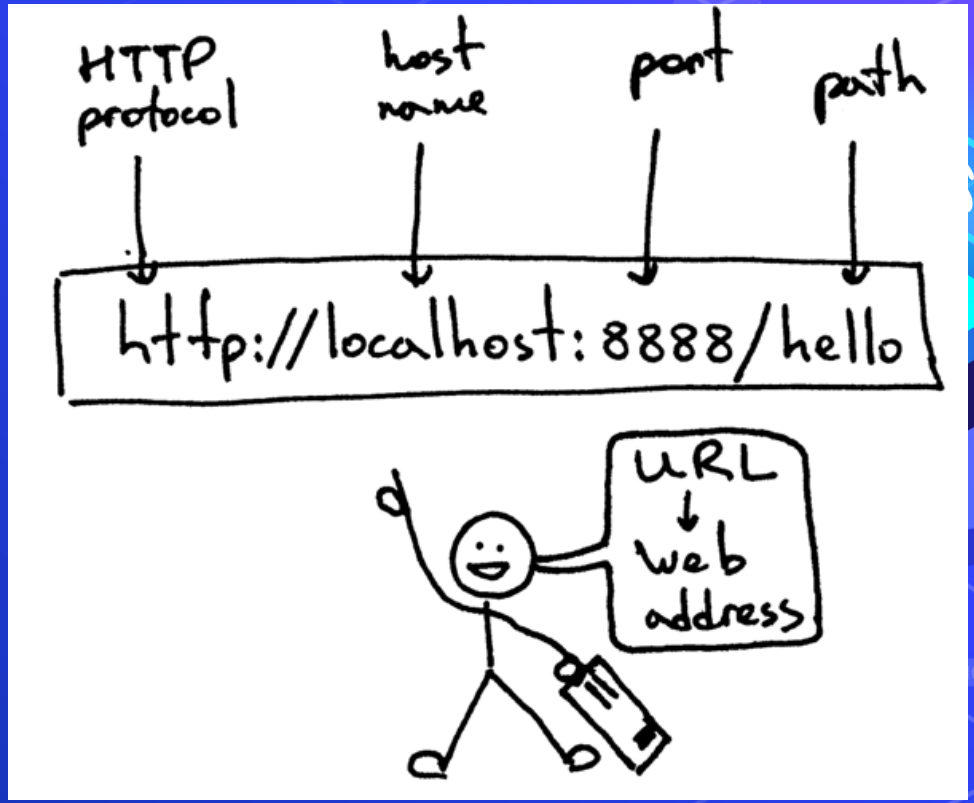
What is Internet and Web Servers



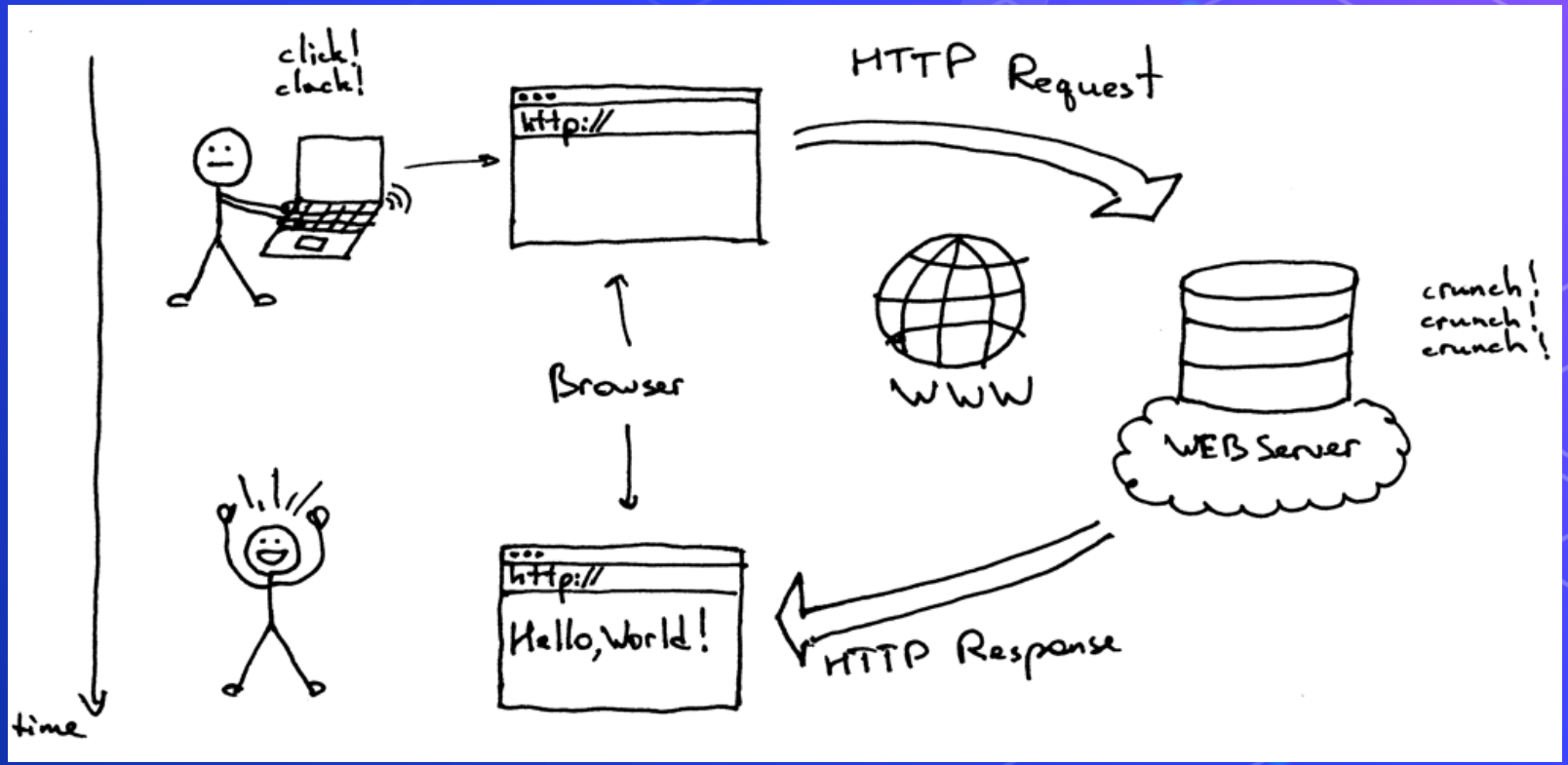
Agenda

- What is Network Topologies
- What is Internet and Web Servers
- **HTTP Request/Response Cycle**
- HTML
- CSS
- Scrapping Concept
- `BeautifulSoap` Library
- REST API Web Services
- Work with JSON

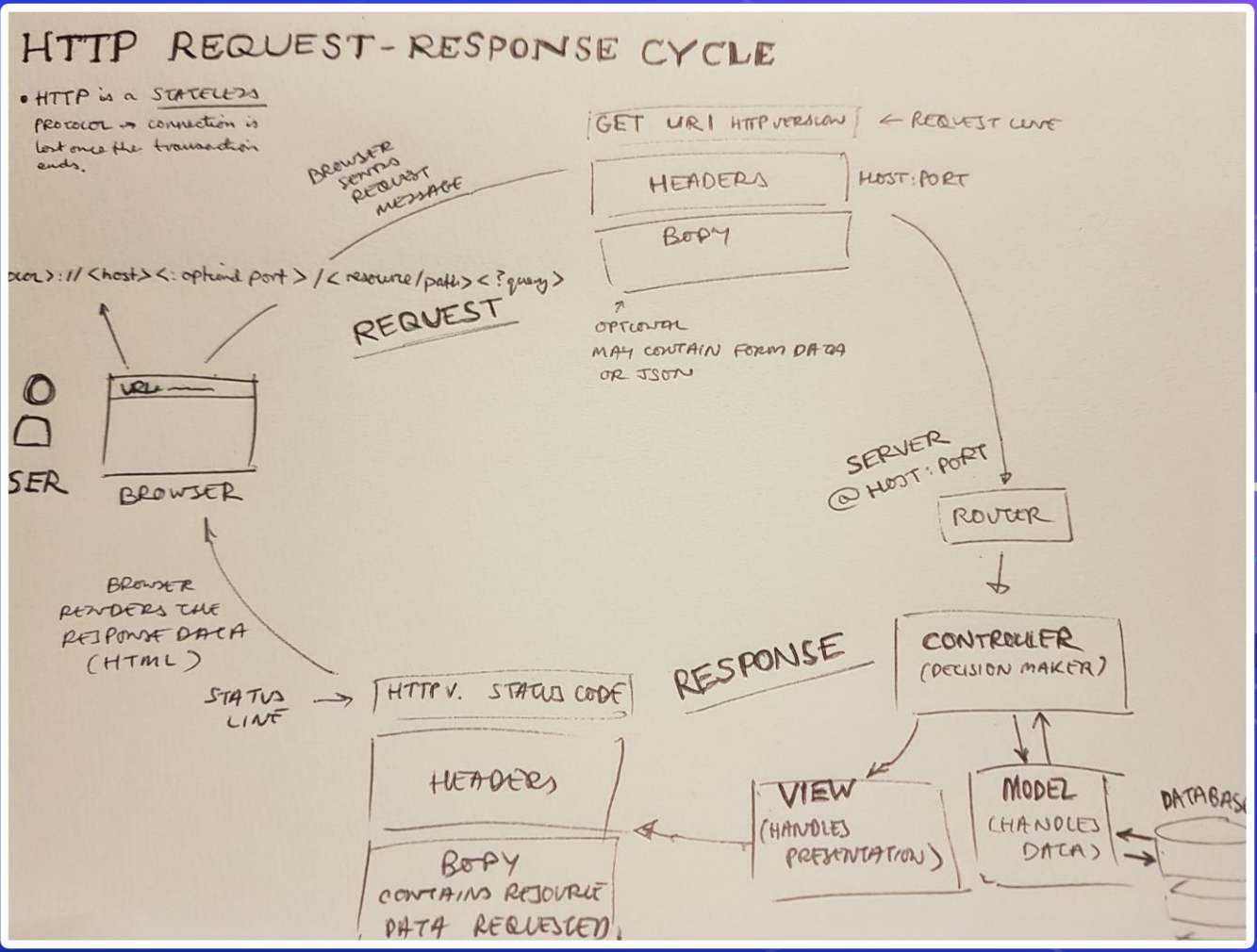
HTTP Request/Response Cycle



HTTP Request/Response Cycle



HTTP REQUEST-RESPONSE CYCLE

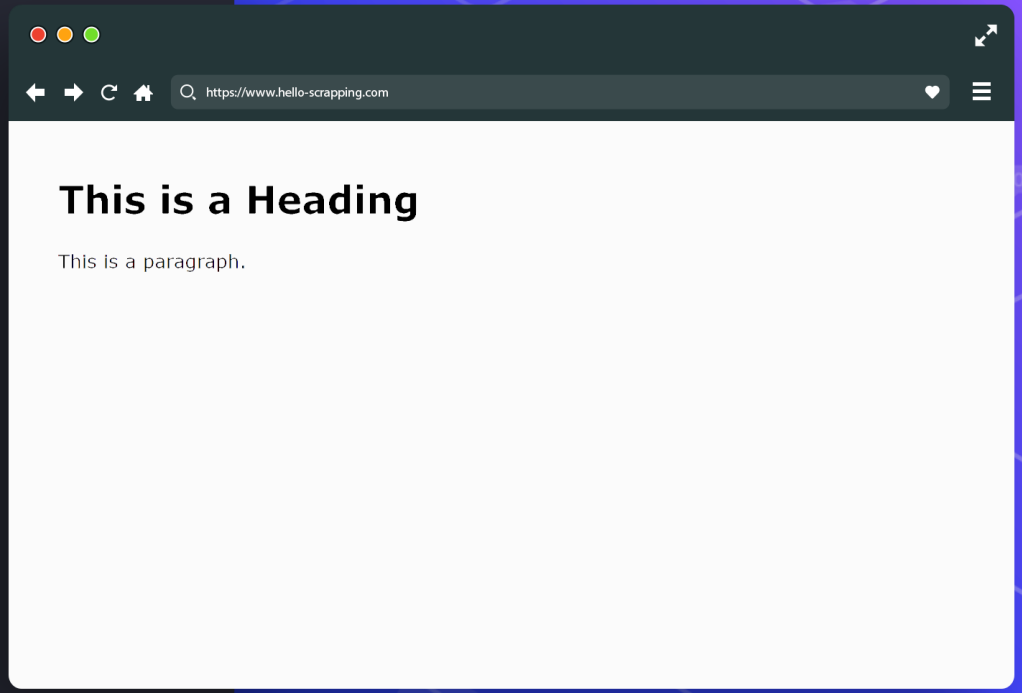


Agenda

- What is Network Topologies
- What is Internet and Web Servers
- HTTP Request/Response Cycle
- **HTML**
- CSS
- Scrapping Concept
- `BeautifulSoap` Library
- REST API Web Services
- Work with JSON

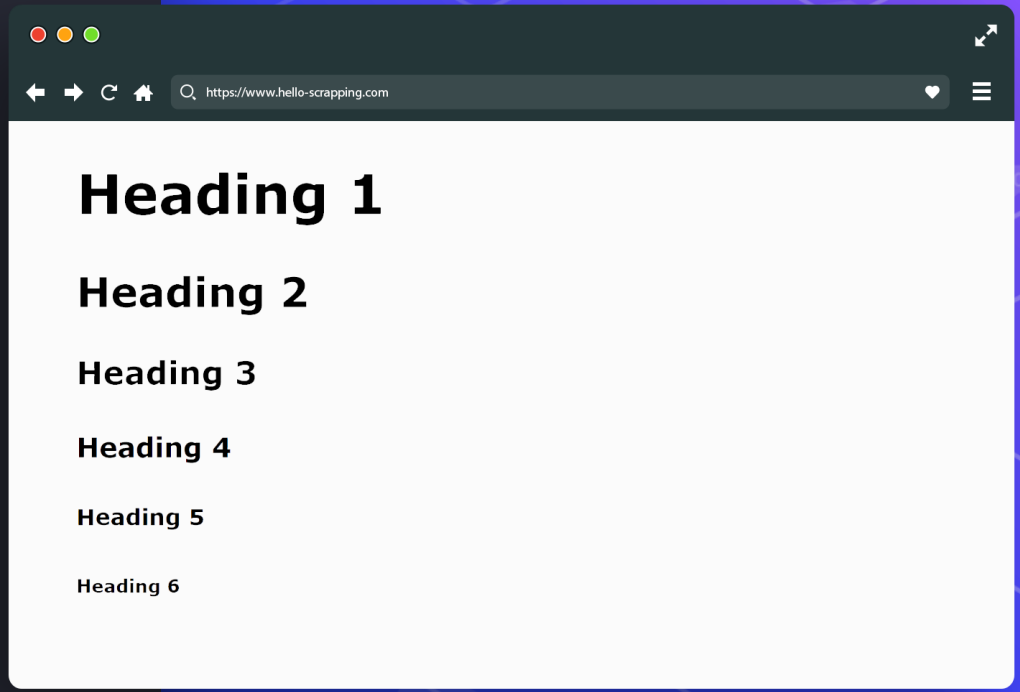
HTML (base snippet code)

```html
<!DOCTYPE html>
<html>

  <head>
    <title>Page Title</title>
  </head>


  <body>

    <h1>This is a Heading</h1>
    <p>This is a paragraph.</p>

  </body>
</html>

```



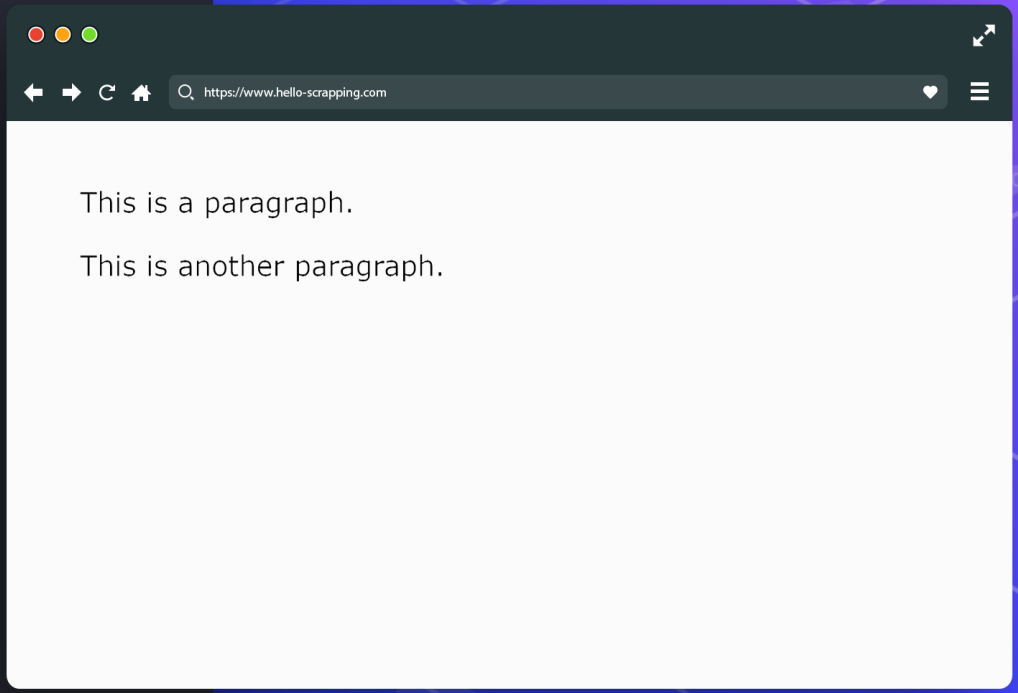
HTML (header)

```html
<!DOCTYPE html>
<html>

  <head>
    <title>Page Title</title>
  </head>


  <body>

    <h1>Heading 1</h1>
    <h2>Heading 2</h2>
    <h3>Heading 3</h3>
    <h4>Heading 4</h4>
    <h5>Heading 5</h5>
    <h6>Heading 6</h6>

  </body>
</html>

```



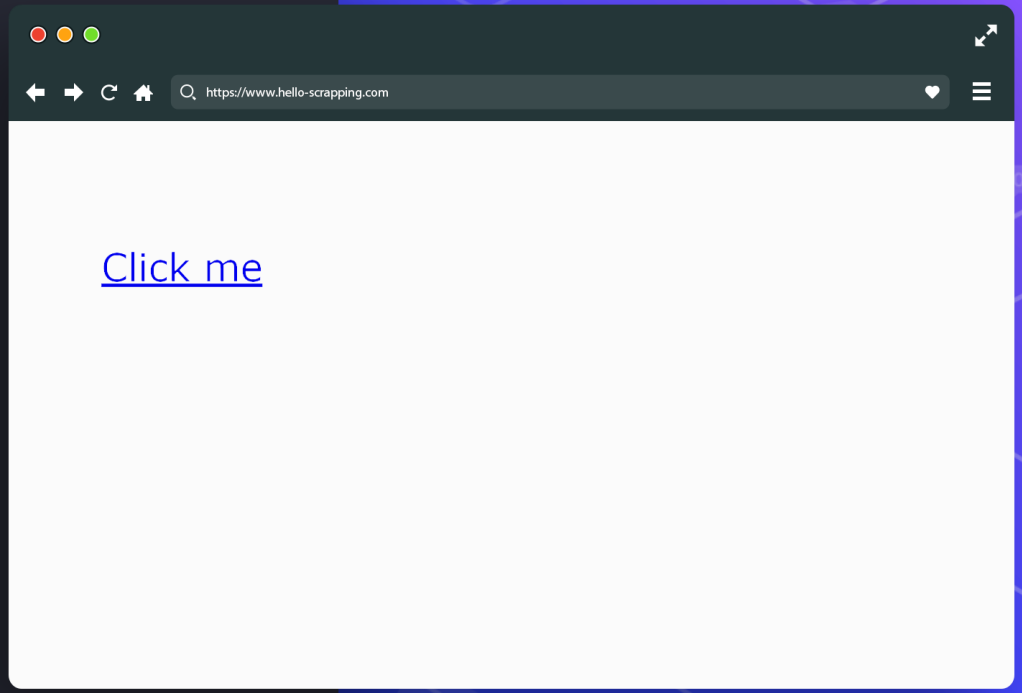
HTML (paragraph)

```html
<!DOCTYPE html>
<html>

  <head>
    <title>Page Title</title>
  </head>


  <body>

    <p>This is a paragraph.</p>
    <p>This is another paragraph.</p>

  </body>
</html>

```



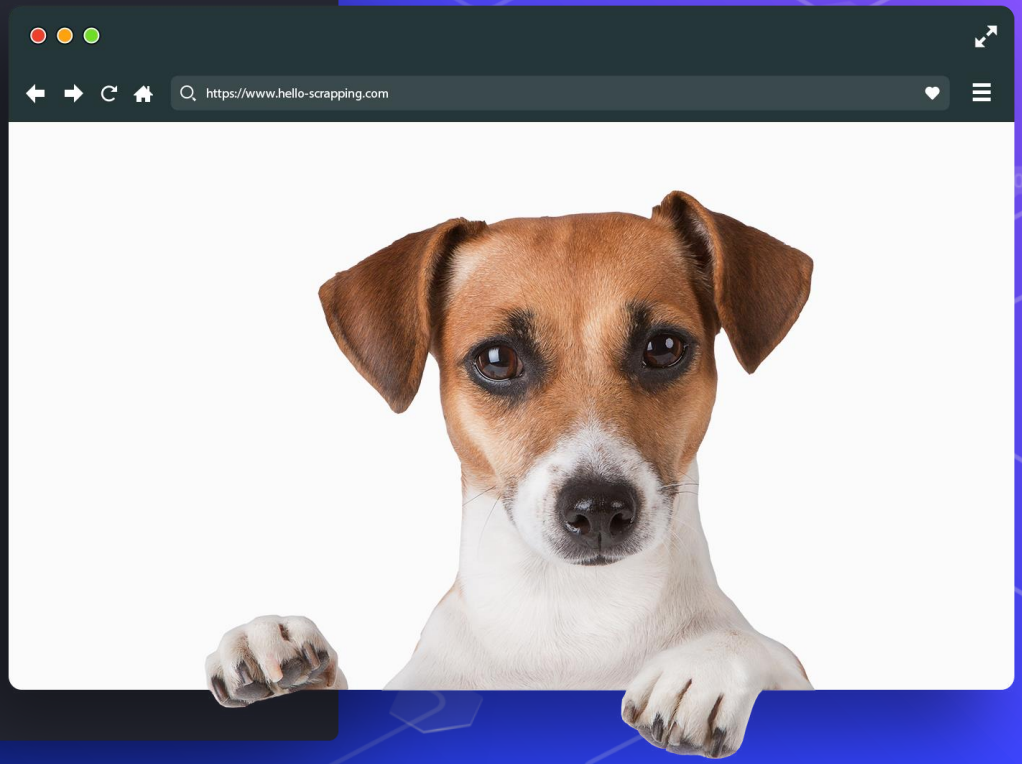
HTML (link)

```html
<!DOCTYPE html>
<html>

  <head>
    <title>Page Title</title>
  </head>


  <body>

    <a href="url">Click me</a>

  </body>
</html>

```



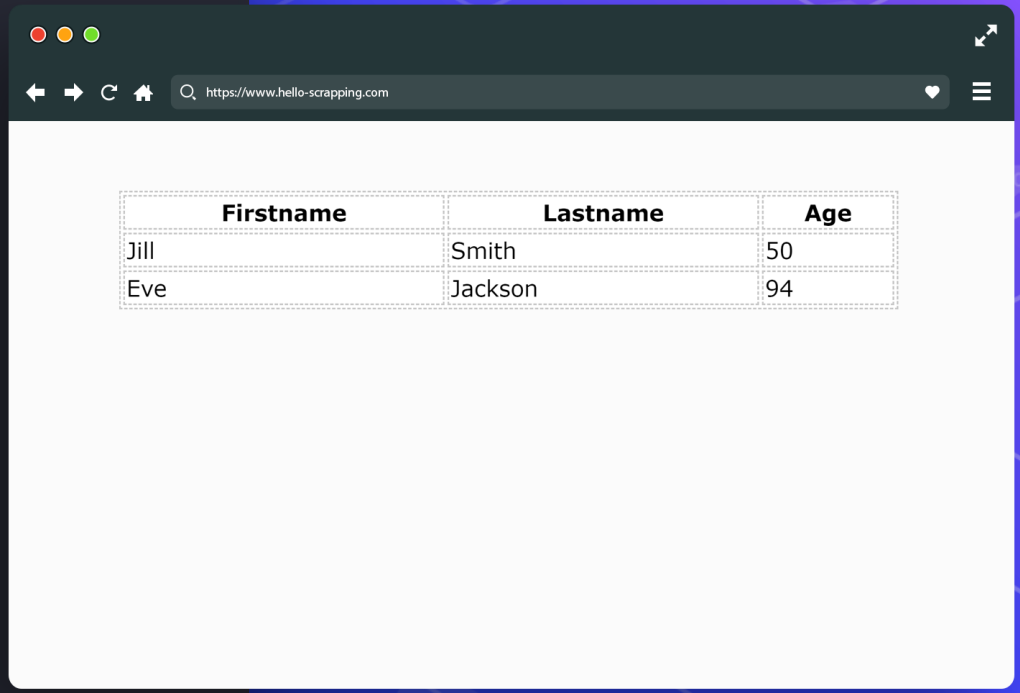
HTML (image)

```html
<!DOCTYPE html>
<html>

  <head>
    <title>Page Title</title>
  </head>


  <body>

    <img src="dog.jpg">

  </body>
</html>

```



HTML (table)

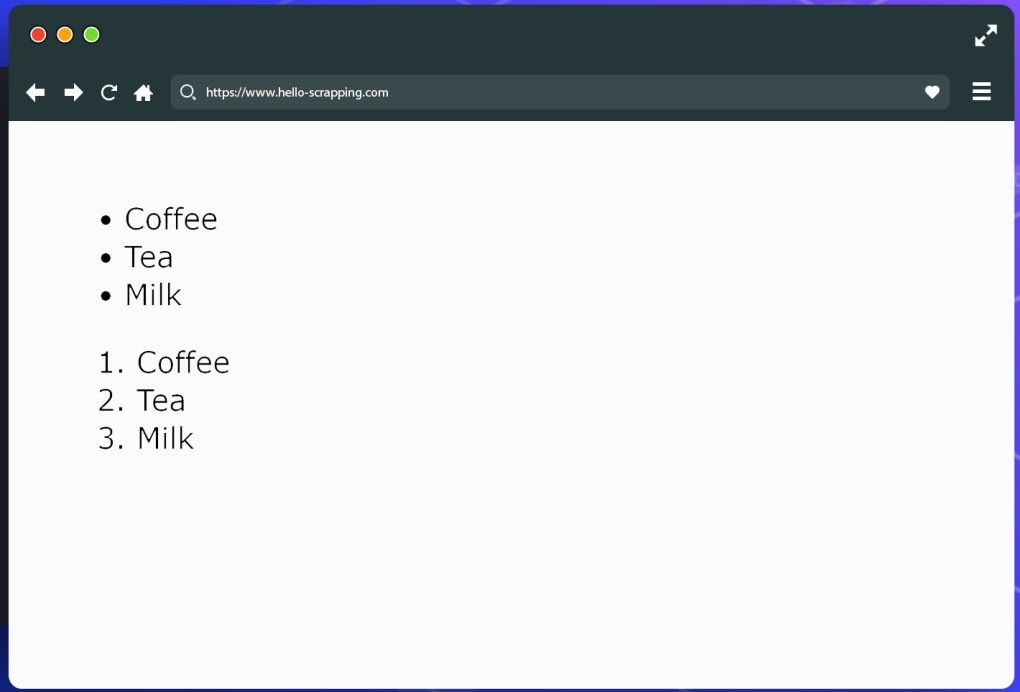
```html
<table style="width:100%">
    <tr>
        <th>Firstname</th>
        <th>Lastname</th>
        <th>Age</th>
    </tr>
    <tr>
        <td>Jill</td>
        <td>Smith</td>
        <td>50</td>
    </tr>
    <tr>
        <td>Eve</td>
        <td>Jackson</td>
        <td>94</td>
    </tr>
</table>

```



Firstname	Lastname	Age
Jill	Smith	50
Eve	Jackson	94

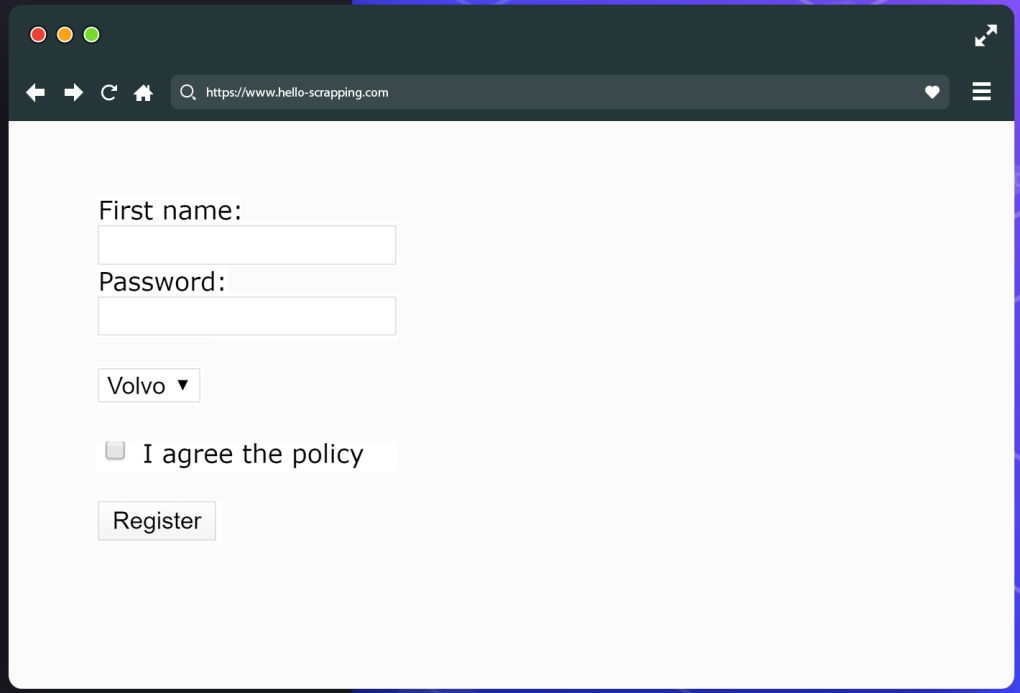
HTML (list)

```html
<ul>
    <li>Coffee</li>
    <li>Tea</li>
    <li>Milk</li>
</ul>

<ol>
    <li>Coffee</li>
    <li>Tea</li>
    <li>Milk</li>
</ol>
```



HTML (form)

```html
<form>

    First name: <input type="text">

    Passowrd: <input type="password">

    <select>
        <option>Volvo</option>
        <option>Saab</option>
        <option>Fiat</option>
        <option>Audi</option>
    </select>


    <input type="checkbox">
    I agree the policy

    <input type="submit" value="Register">

</form>
```



Other HTML Elements

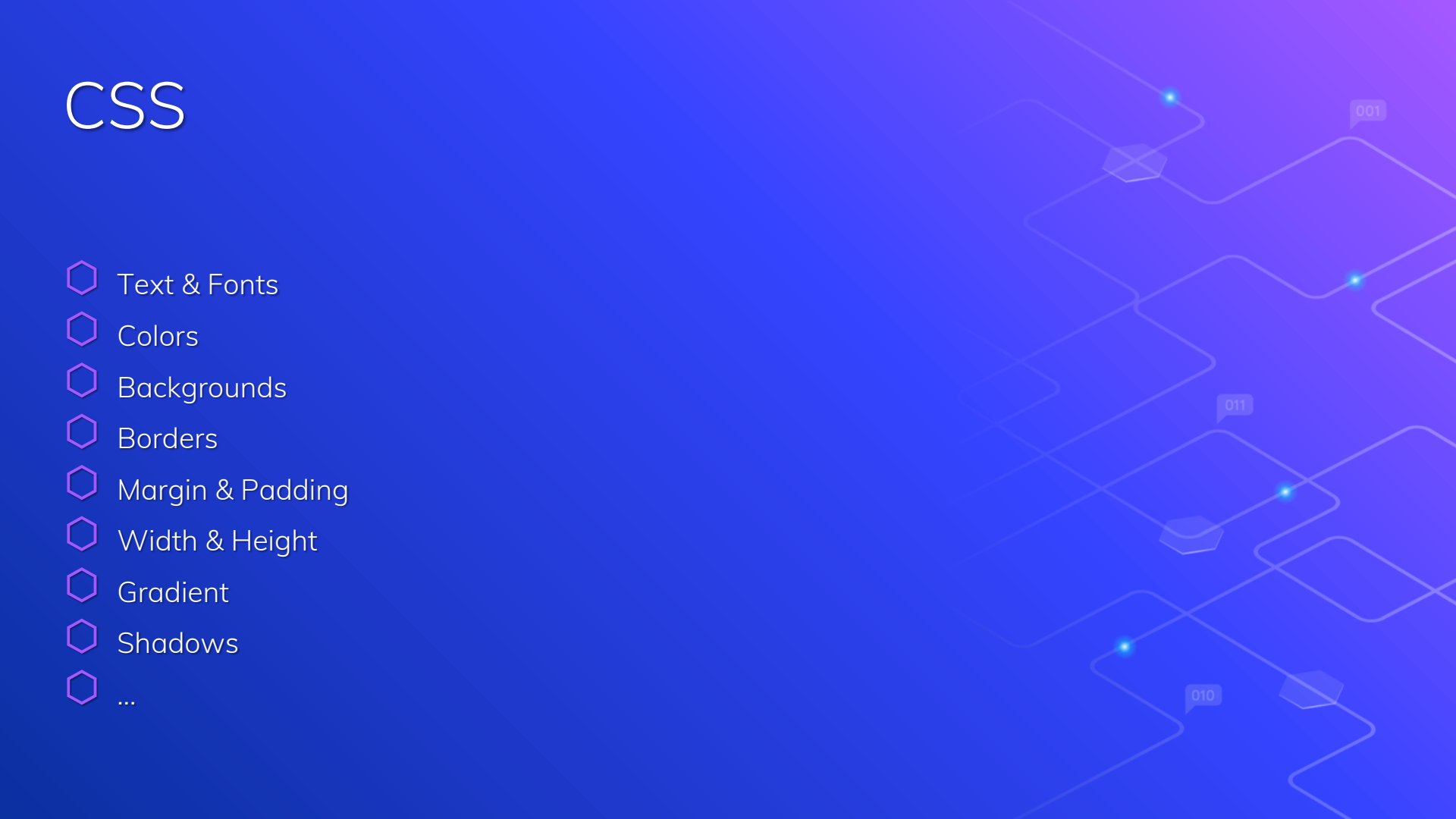
- Div
- Span
- Video
- Audio
- Iframe
- Header
- Footer
- Canvas
- ...

Agenda

- What is Network Topologies
- What is Internet and Web Servers
- HTTP Request/Response Cycle
- HTML
- CSS
- Scrapping Concept
- `BeautifulSoap` Library
- REST API Web Services
- Work with JSON

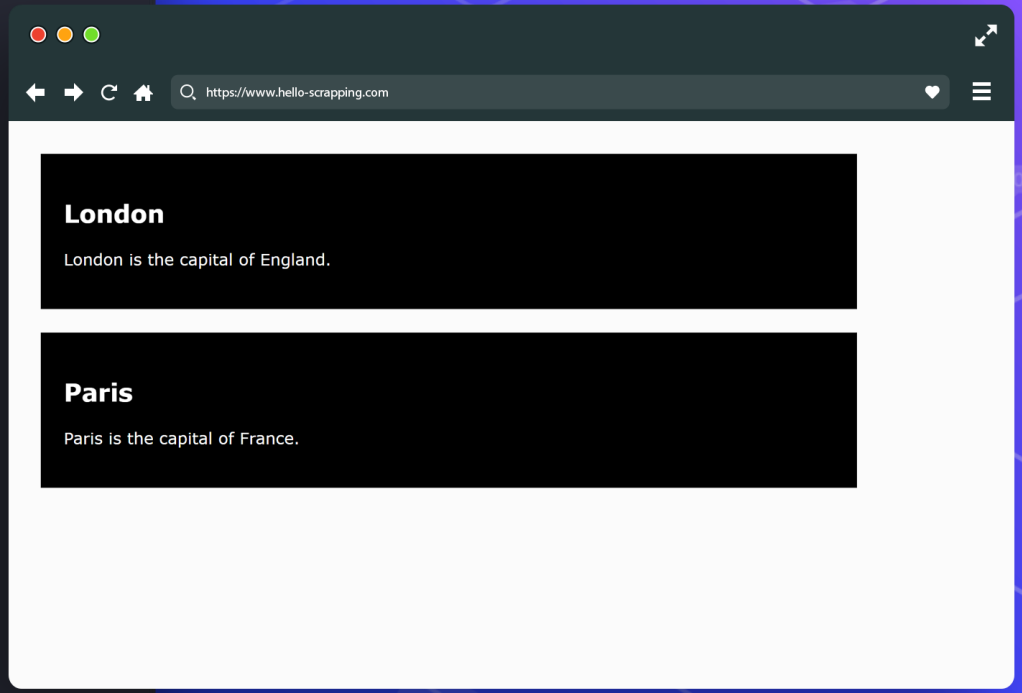
CSS

- Text & Fonts
- Colors
- Backgrounds
- Borders
- Margin & Padding
- Width & Height
- Gradient
- Shadows
- ...



Classes and ID

```html
<!DOCTYPE html>
<html>
    <head>
        <style>
            .cities
            {
                background-color: black;
                color: white;
                margin: 20px;
                padding: 20px;
            }
        </style>
    </head>
    <body>

    <div class="cities">
        <h2>London</h2>
        <p>London is the capital of England.</p>
    </div>

    <div class="cities">
        <h2>Paris</h2>
        <p>Paris is the capital of France.</p>
    </div>

    </body>
</html>

```



Check Also

- JavaScript
- React & Angular
- Ajax
- Web Sockets
- PHP: Laravel
- Python: Flask, Django
- JavaScript: Node.js
- ...

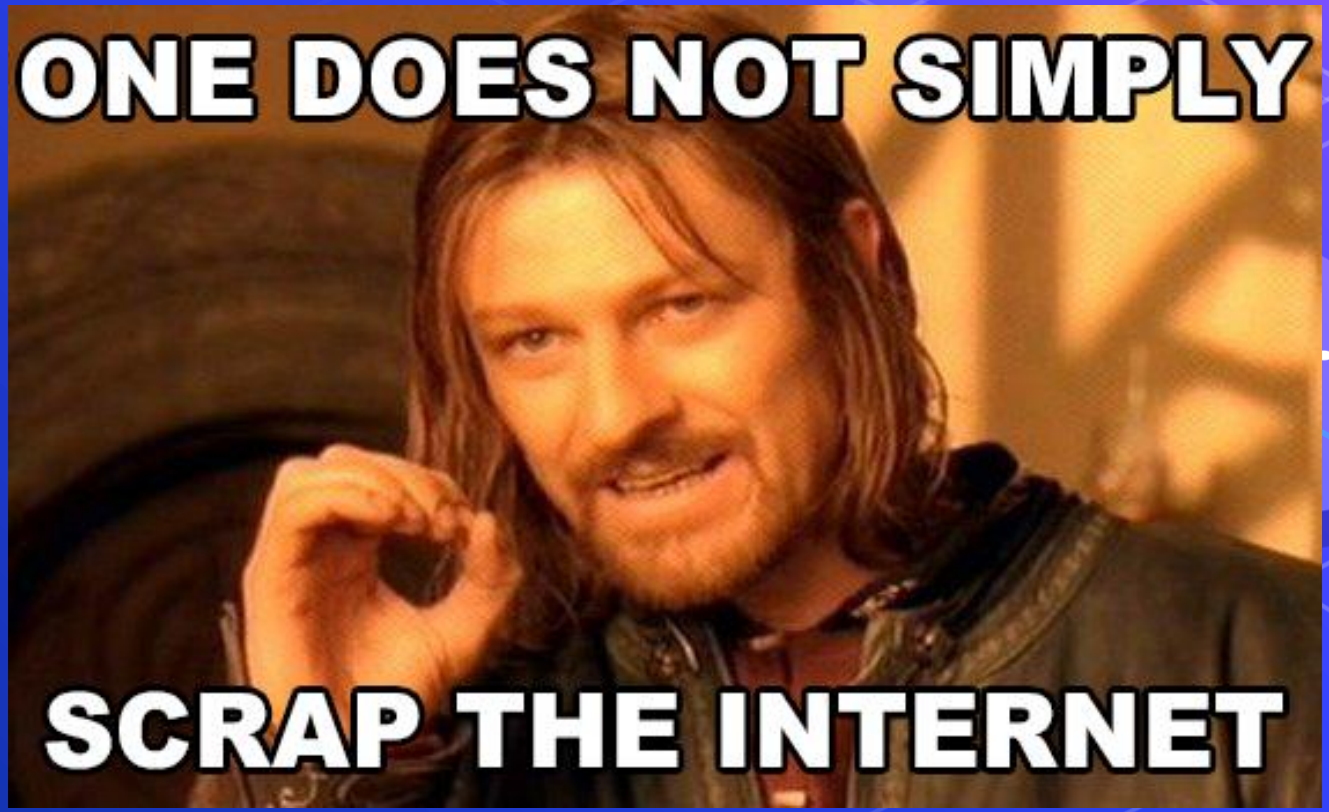
Agenda

- What is Network Topologies
- What is Internet and Web Servers
- HTTP Request/Response Cycle
- HTML
- CSS
- **Scrapping Concept**
- `BeautifulSoap` Library
- REST API Web Services
- Work with JSON

Scrapping Concept

Scrapping Concept

Scrapping Concept



Agenda

- What is Network Topologies
- What is Internet and Web Servers
- HTTP Request/Response Cycle
- HTML
- CSS
- Scrapping Concept
- **`BeautifulSoap Library`**
- REST API Web Services
- Work with JSON

BeautifulSoap Library


```
conda install beautifulsoup4
```

```python
import requests
from bs4 import BeautifulSoup

url = 'https://jekso.github.io/scrapping-example/index.html'
response = requests.get(url)

soup = BeautifulSoup(response.text, 'html.parser')

# ------------------------------------------------ #

# find first element by type
link = soup.find('a')

# find first element by type and an attribute
link = soup.find('a', attrs={'class': 'social-link'})

# ------------------------------------------------ #

# find all elements by type
all_links = soup.find_all('a')

# find all elements by type and an attribute
all_links = soup.find_all('a', attrs={'class': 'social-link'})

# ------------------------------------------------ #

# check if element has attribute
link.has_attr('class')

# get attribute
link.get('href')

# ------------------------------------------------ #

# get content
soup.find('a').get_text()

```

Check Also

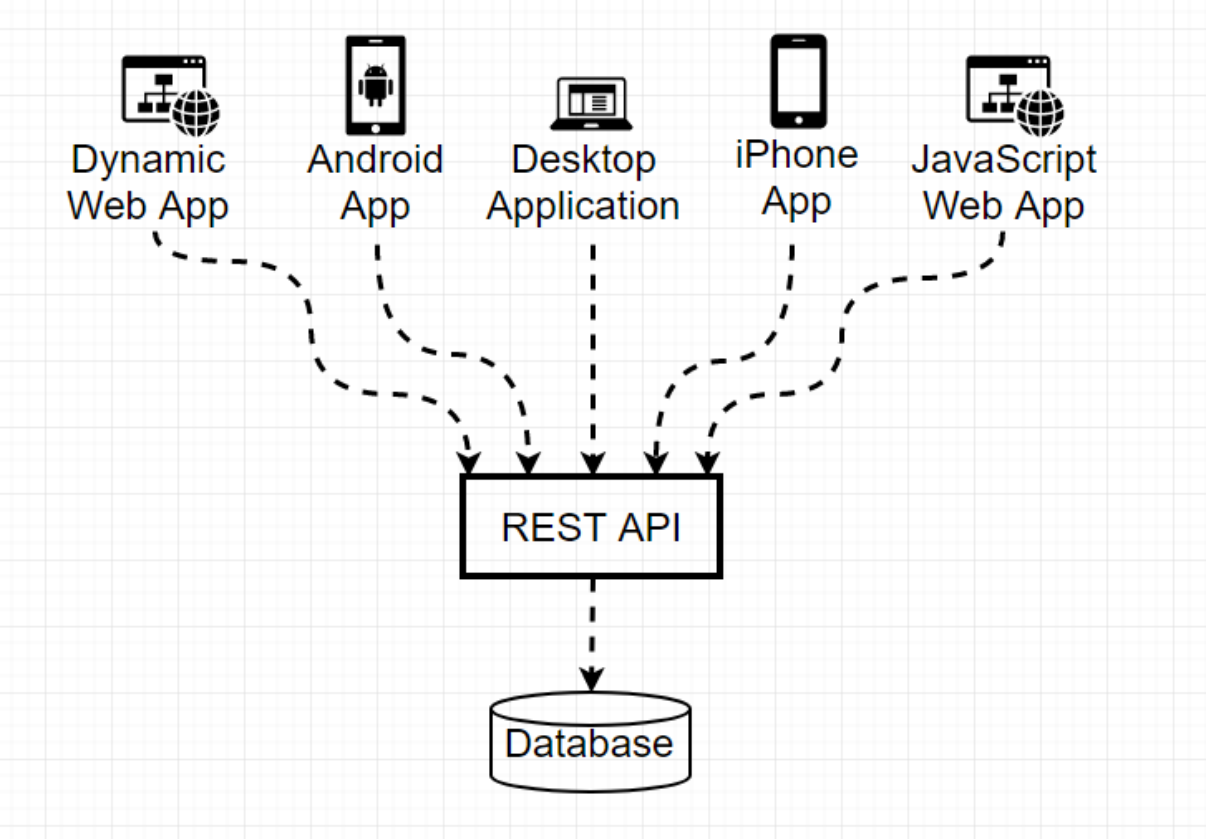
Other methods to scrap the internet.

- https://scrapy.org
- https://selenium-python.readthedocs.io
- ...

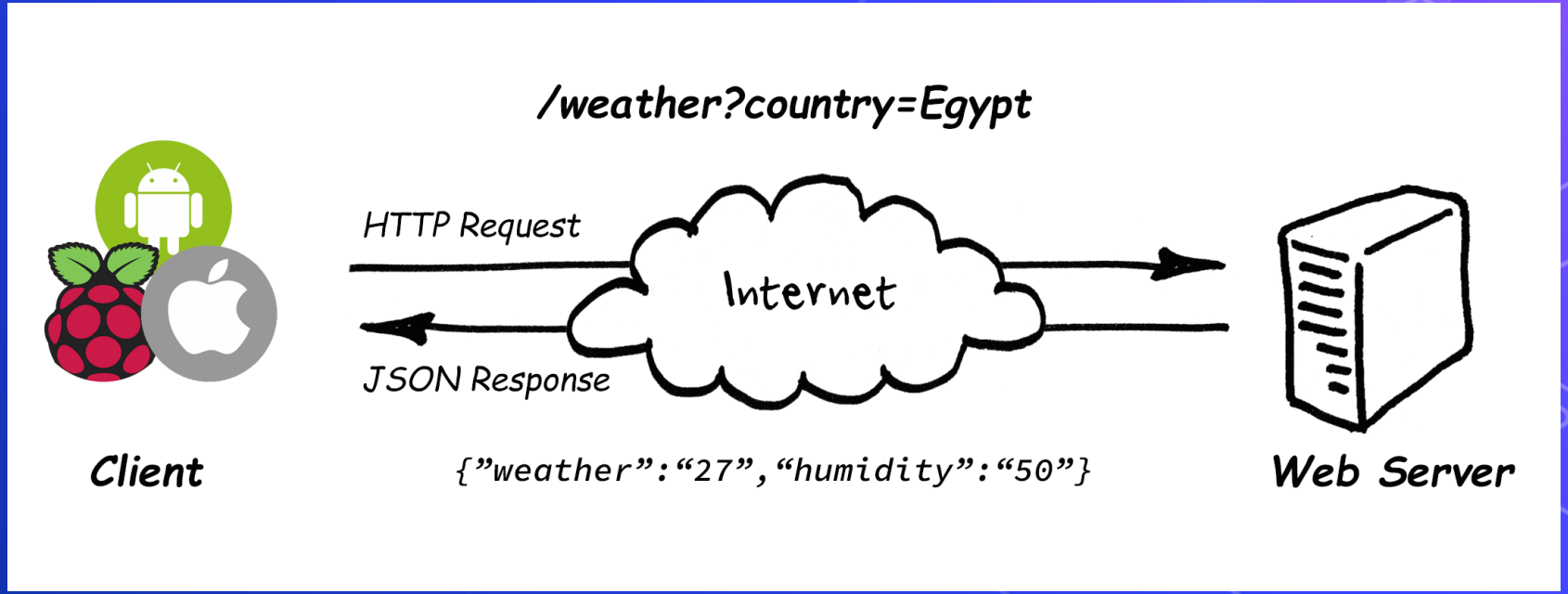
Agenda

- What is Network Topologies
- What is Internet and Web Servers
- HTTP Request/Response Cycle
- HTML
- CSS
- Scrapping Concept
- `BeautifulSoap` Library
- **REST API Web Services**
- Work with JSON

REST API Web Services

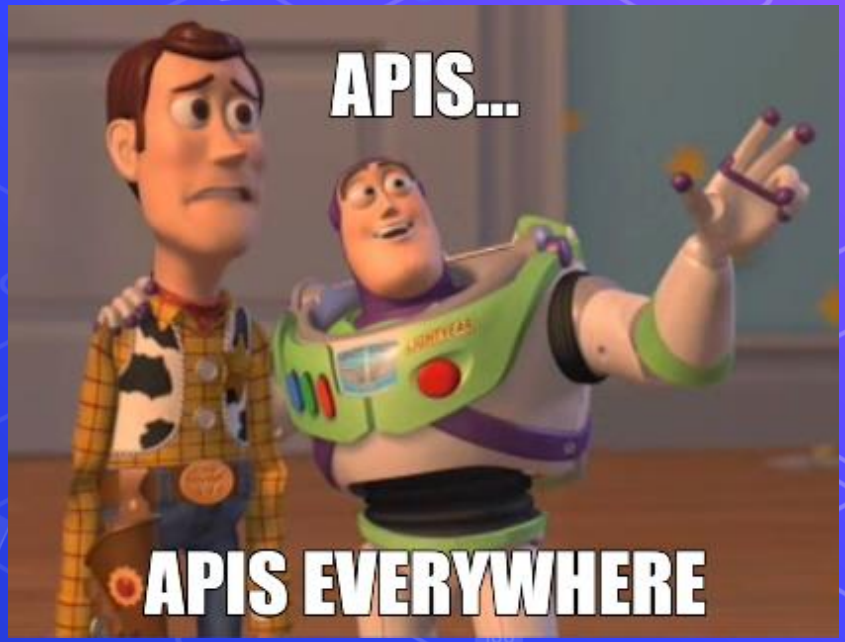


REST API Web Services



REST API Web Services

- https://any-api.com/
- https://github.com/public-apis/public-apis
- https://cloud.google.com/products/ai
- https://www.twilio.com/
- ...



Agenda

- What is Network Topologies
- What is Internet and Web Servers
- HTTP Request/Response Cycle
- HTML
- CSS
- Scrapping Concept
- `BeautifulSoap` Library
- REST API Web Services
- **Work with JSON**

Work with JSON

```json
{
    "employees":
    [
        {
          "id": "1",
          "employee_name": "Ahmed",
          "employee_salary": "320800",
          "employee_age": "61"
        },
        {
          "id": "2",
          "employee_name": "Amr",
          "employee_salary": "170750",
          "employee_age": "63"
        },
        {
          "id": "3",
          "employee_name": "Sara",
          "employee_salary": "86000",
          "employee_age": "66"
        }
    ]
}
```

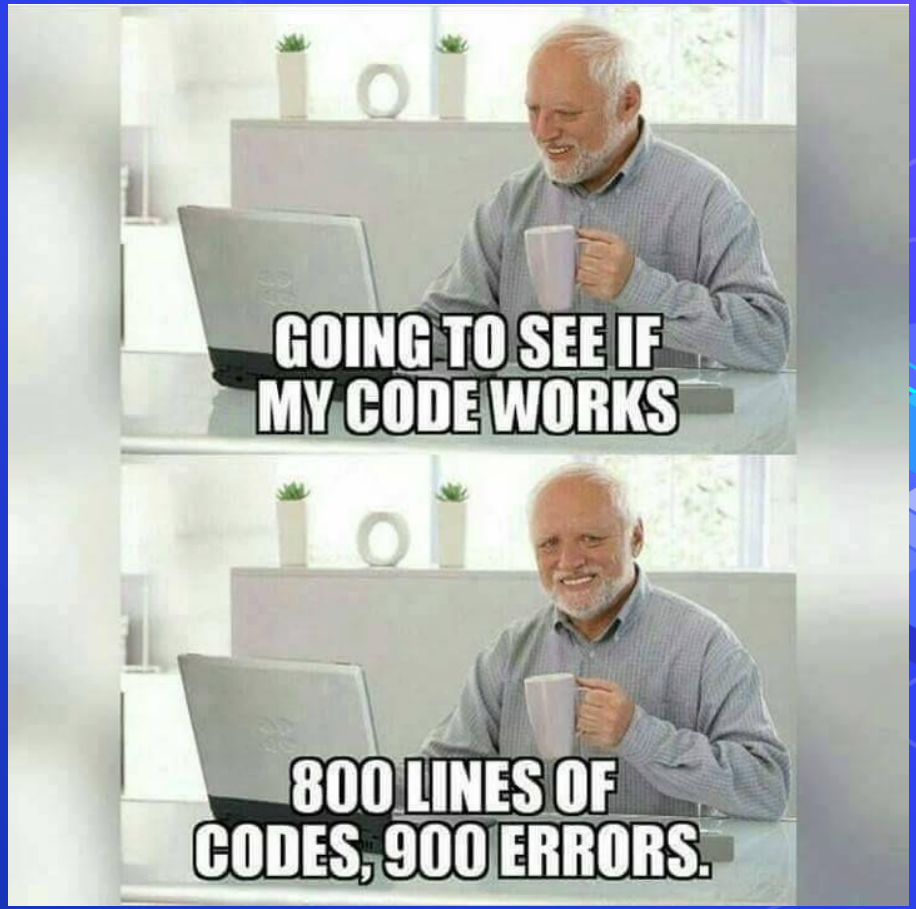
Work with JSON

Install JSON Viewer extension
For chrome or firefox

https://chrome.google.com/webstore/detail/json-viewer/gbmdgpbipfallnflgajpaliibnhdgobh

Work with JSON

```python
import json

# some JSON string:
json_text =  '{ "name":"John", "age":30, "city":"New York"}'

# parse json_text:
user = json.loads(json_text)

# the result is a Python dictionary:
print(user["age"])
```


GOING TO SEE IF
MY CODE WORKS
800 LINES OF
CODES, 900 ERRORS.

Questions ?!

Thanks!

>_ Live long and prosper