

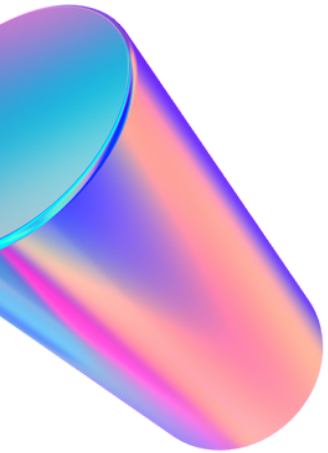
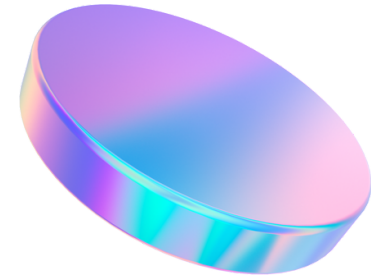
Web Scraping VS APIS

WEB SCRAPING

01

DEFINITION AND FUNCTIONALITY

- Web scraping involves harvesting data from websites.
- The data extraction process is customizable and user-friendly.



02

WEB SCRAPING TOOLS

- Software tools like Python, Ruby, PHP, Node.js are used for web scraping.
- These tools read a website's HTML, CSS, and JavaScript elements.

03

DATA ACCESS AND COLLECTION

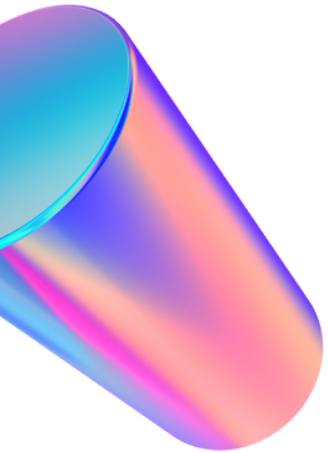
- Web scrapers can access both surface-level and deep data.
- Specific data types can be targeted from multiple websites.

WEB SCRAPING

04

PROXY USAGE

- Web scrapers use proxy servers to avoid security measures.
- Proxies hide identity and mask IP addresses to appear as normal user traffic.



05

EASE OF USE

- Web scrapers are designed to be user-friendly.
- No advanced programming or data science skills required.

06

DATA PREPARATION

- Scraped data is often converted into user-friendly formats.
- Data is packaged for easy access and download.

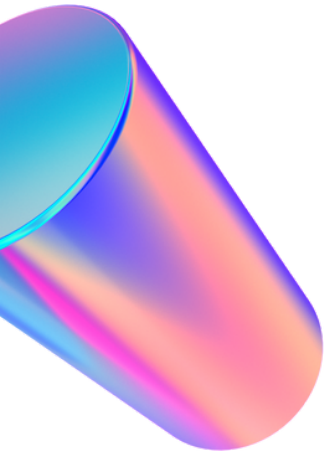
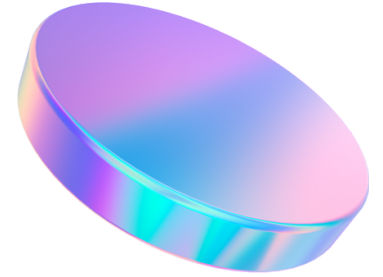
API

APPLICATION PROGRAMMING INTERFACES

01

DEFINITION AND FUNCTIONALITY

- APIs facilitate communication between websites and software.
- They enable structured data exchange according to rules.



02

API CHARACTERISTICS

- APIs enforce rules and limitations on data extraction.
- Data sources, extraction type, and request frequency are controlled.

03

QUERY LANGUAGE AND KNOWLEDGE

- APIs require familiarity with the website's query language.
- JSON is commonly used for data exchange through APIs.

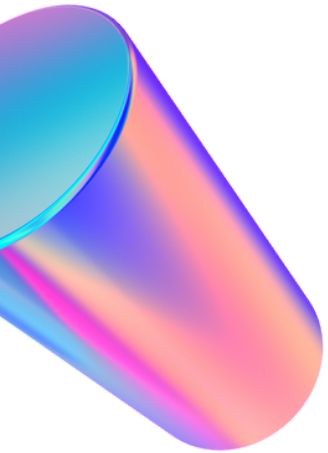
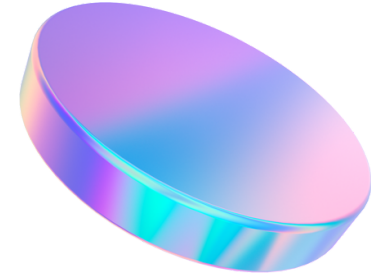
API

APPLICATION PROGRAMMING INTERFACES

04

RAW DATA AND CONVERSION

- APIs often provide raw data requiring additional processing.
- Conversion into formats like CSV or SQL may be necessary.



05

ETHICAL CONSIDERATIONS

- APIs are official tools provided by the website.
- Access is restricted to the data the owner wants to share

06

AVOIDING IP BLOCKING

- APIs eliminate concerns about proxy usage and IP blocking.
- Official communication channel minimizes these issues.



Conclusion

- **Selection Criteria**

- The choice depends on skill level, website availability, and data needs.
- Web scraping is preferred when APIs are not available or suitable.

- **Data Privacy and Ethics**

- Both methods must consider data privacy and adhere to website terms.
- Ethical considerations play a role in data collection.

- **Data Collection Strategy**

- APIs offer structured data extraction with compliance.
 - Web scraping is flexible but requires cautious implementation.
- 