

## Web Scraping Architect

### Description

We are seeking a highly skilled Web Scraping Architect to join our team. The successful candidate will be responsible for designing, implementing, and maintaining web scraping processes to gather data from various online sources efficiently and accurately. As a Web Scraping Specialist, you will play a crucial role in collecting data for competitor analysis, and other business intelligence purposes.

### Responsibilities

- Scalability/Performance: Lead and provide expertise in scraping at scale e-commerce marketplaces.
- Data Source Identification: Identify relevant websites and online sources from which data needs to be scraped. Collaborate with the team to understand data requirements and objectives.
- Web Scraping Design: Develop and implement effective web scraping strategies to extract data from targeted websites. This includes selecting appropriate tools, libraries, or frameworks for the task.
- Data Extraction: Create and maintain web scraping scripts or programs to extract the required data. Ensure the code is optimized, reliable, and can handle changes in the website's structure.
- Data Cleansing and Validation: Cleanse and validate the collected data to eliminate errors, inconsistencies, and duplicates. Ensure data integrity and accuracy throughout the process.
- Monitoring and Maintenance: Continuously monitor and maintain the web scraping processes. Address any issues that arise due to website changes, data format modifications, or anti-scraping mechanisms.
- Scalability and Performance: Optimize web scraping procedures for efficiency and scalability, especially when dealing with a large volume of data or multiple data sources.
- Compliance and Legal Considerations: Stay up-to-date with legal and ethical considerations related to web scraping, including website terms of service, copyright, and privacy regulations.
- Documentation: Maintain detailed documentation of web scraping processes, data sources, and methodologies. Create clear and concise instructions for others to follow.
- Collaboration: Collaborate with other teams such as data analysts, developers, and business stakeholders to understand data requirements and deliver insights effectively.
- Security: Implement security measures to ensure the confidentiality and protection of sensitive data throughout the scraping process.

### Qualifications

- Proven experience of 7+ years as a Web Scraping Specialist or similar role, with a track record of successful web scraping projects.
- Expertise in handling dynamic content, user-agent rotation, bypass CAPTCHAs, rate limits, and utilization of proxy services.
- Knowledge on browser fingerprinting.
- Has leadership experience.
- Proficiency in programming languages commonly used for web scraping, such as Python, BeautifulSoup, Scrapy, or Selenium.

### Hiring organization

Pulse HRM

### Employment Type

Full-time

### Job Location

Remote work possible

### Date posted

April 2, 2025

### Valid through

02.05.2025

- Strong knowledge of HTML, CSS, XPath, and other web technologies relevant to web scraping and Coding.
- Knowledge and experience in best of class data storage and retrieval of large volume of scraped data.
- Understanding of web scraping best practices, including handling dynamic content, user-agent rotation, and IP address management.
- Attention to detail and the ability to handle and process large volumes of data accurately.
- Familiarity with data cleansing techniques and data validation processes.
- Good communication skills and the ability to collaborate effectively with cross-functional teams.
- Knowledge of web scraping ethics, legal considerations, and compliance with website terms of service.
- Strong problem-solving skills and the ability to adapt to changing web environments.
- Experience with cloud-based solutions and distributed web scraping systems.
- Familiarity with APIs and data extraction from non-public sources.
- Knowledge of machine learning techniques for data extraction and natural language processing is desired but not mandatory.
- Prior experience in handling large-scale data projects and working with big data frameworks.
- Understanding of various data formats such as JSON, XML, CSV, etc.
- Experience with version control systems like Git.

## **Education**

- Bachelor's degree in Computer Science, Data Science, Information Technology, or related fields